Journal of Multivariate Analysis 173 (2019) 70–84

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress

Lea Petrella [a], Valentina Raponi [b,a,*]

[a] MEMOTEF, Sapienza, University of Rome, Rome, Italy
[b] Imperial College Business School, Imperial College London, London, United Kingdom

ARTICLE INFO

Article history:
Received 16 May 2018
Received in revised form 14 February 2019
Accepted 14 February 2019
Available online 20 February 2019

MSC:
62F10
62F30
62H12
60E05

Keywords:
EM algorithm
Maximum likelihood
Multivariate asymmetric Laplace distribution
Multiple quantiles
Multivariate response variables
Quantile regression

ABSTRACT

This paper proposes a maximum likelihood approach to jointly estimate marginal conditional quantiles of multivariate response variables in a linear regression framework. We consider a slight reparameterization of the multivariate asymmetric Laplace distribution proposed by Kotz et al. (2001) and exploit its location–scale mixture representation to implement a new EM algorithm for estimating model parameters. The idea is to extend the link between the asymmetric Laplace distribution and the well-known univariate quantile regression model to a multivariate context, i.e., when a multivariate dependent variable is concerned. The approach accounts for association among multiple responses and studies how the relationship between responses and explanatory variables can vary across different quantiles of the marginal conditional distribution of the responses. A penalized version of the EM algorithm is also presented to tackle the problem of variable selection. The validity of our approach is analyzed in a simulation study, where we also provide evidence on the efficiency gain of the proposed method compared to estimation obtained by separate univariate quantile regressions. A real data application examines the main determinants of financial distress in a sample of Italian firms.

1. Introduction

Quantile regression has become a widely used technique in many empirical applications, since the seminal work of Koenker and Basset [30]. It provides a way to model the conditional quantiles of a response variable with respect to a set of covariates in order to have a more complete picture of the entire conditional distribution than the ordinary least squares regression. This approach is quite suitable to be used in all the situations where specific features, like skewness, fat-tails, outliers, truncation, censoring and heteroscedasticity arise. In fact, unlike standard linear regression models, which only consider the conditional mean of a response variable, quantile regression allows one to assume that the relationship between the response and explanatory variables can vary across the conditional distribution of the dependent variable.

Many univariate quantile regression methods are now well consolidated in the literature and have been implemented in a wide range of different fields, like medicine [14,39,44], survival analysis [31], financial and economic research [4,24,42], and environmental modeling; see, e.g., [24,41] for a discussion. Koenker [28] provides an overview of the most used quantile regression techniques in a classical setting. In longitudinal studies, quantile regression models with random effects are also analyzed, in order to account for the dependence between serial observations on the same subject;

* Correspondence to: Imperial College Business School, South Kensington Campus, United Kingdom.
E-mail addresses: lea.petrella@uniroma1.it (L. Petrella), v.raponi13@imperial.ac.uk (V. Raponi).

https://doi.org/10.1016/j.jmva.2019.02.008
0047-259X/

see, e.g., [21,27,29,38] for references. Bayesian versions of quantile regression have also been extensively proposed; see [6,34,35,53].

It is well known in the literature that the univariate quantile regression approach has a direct link with the asymmetric Laplace (AL) distribution. In fact, while the frequentist quantile regression framework relies on the minimization of the asymmetric loss function introduced by Koenker and Basset [30], the Bayesian approach introduces the asymmetric Laplace (AL) distribution as an inferential tool to estimate model parameters; see the seminal work [53]. The two approaches are justified by the well established relationship between the loss function and the AL density. That is, the loss function minimization problem is equivalent (in terms of parameter estimates) to the maximization of the likelihood associated with the AL density; see, e.g., [37]. Therefore, the AL distribution could offer a convenient device to implement a likelihood-based inferential approach when dealing with quantile regression analysis.

Still in the context of univariate regression framework, part of the literature is concerned with estimation of multiple quantiles, say $\mathcal{Q}_Y(\tau_1), \ldots, \mathcal{Q}(\tau_p)$, of a given response variable $Y \in \mathbb{R}$. In this case, joint estimation of these p quantiles provides a gain in efficiency compared to traditional sequential estimation of multiple quantiles. Cho et al. [13], for example, show that the asymptotic covariance obtained from the multiple quantile regression model is always smaller than the one obtained from the single quantile regression model. Even though only one quantile regression is of particular interest, their proposed simultaneous multiple quantile estimation may also be desirable when simultaneous estimation is supported by reasonably sized samples. For example, a more efficient median regression estimator could be obtained by estimating the 25th and 75th quantile regression estimators together. Hence, a multiple quantile estimation approach is more efficient than either ignoring the inter-quantile correlation or estimating regression parameters individually.

When multivariate response variables are concerned, the existing literature on quantile regression is less extensive. The multivariate quantile problem has the goal to estimate the multivariate quantile, $\mathcal{Q}_{\boldsymbol{Y}}(\tau)$, of a multivariate response variable $\boldsymbol{Y} \in \mathbb{R}^p$, with $p > 1$ and where the index τ is a scalar. In this case, the main challenge concerns the definition of a multivariate quantile, given that there is no natural ordering in a p-dimensional space [11,23,32]. Other attempts and extensions can be also found in [2,8,33,40,46].

The main goal of the present paper is to extend the univariate linear quantile regression methodology to a multivariate context. In particular we want to generalize the inferential approach based on the AL distribution to a multivariate framework, by using the multivariate asymmetric Laplace (MAL) distribution introduced by Kotz et al. [36]. In this way we are not concerned to define a multivariate quantile. Instead, we are conducting a simultaneous inference on the marginal conditional quantiles of a multivariate response variable, taking also into account for the possible correlation among marginals.

This need could arise in many situations where different responses might have similar distributions or might be affected by the same set of covariates at different parts of their distributions other than the mean. Hence, by jointly modeling them, we can borrow information across responses and conduct joint inference at the marginal quantiles level, rather than defining a central point for their distributions.

Similar attempts in the literature have been proposed by Jun and Pinkse [26], who introduced a seemingly unrelated quantile regression approach which entails a nonparametric estimation of a set of moment conditions for the conditional quantiles of interest. Waldmann and Kneib [49] proposed a bivariate quantile regression using a Bayesian approach and showed how to estimate the conditional correlations between the two response variables. Their work, however, has not been generalized to the multivariate case.

Our paper offers a likelihood-based approach for modeling and estimating conditional marginal quantiles jointly, by using the MAL distribution as working likelihood in a linear regression framework. Specifically, we propose a slight reparameterization of the MAL distribution, subject to some specific constraints, which allows us to estimate regression coefficients via maximum likelihood (ML), accounting for the possible association among the responses. The inferential problem is solved by developing a suitable Expectation–Maximization (EM) algorithm, which exploits the mixture representation of the MAL distribution; see also [3].

Using simulation exercises, we assess the validity and the robustness of our approach by considering different model distributional settings. We find that the estimation of the regression coefficients is not highly affected by the MAL distributional assumption. Moreover, the estimation efficiency of our multivariate approach is higher than the one obtained by running separate single quantile regression models on the marginals when estimating model parameters. That is, taking into account for the potential association among the response variables can significantly reduce the root mean square error of the estimated coefficients, hence improving the precision of the estimates.

When dealing with multivariate regression, the high dimensionality setting is an intrinsic part of the model building problem. In order to gain in parsimony and to conduct a variable selection procedure, we consider the penalized Least Absolute Shrinkage and Selecting Operator (LASSO) approach proposed by Tibshirani [47]. In particular, we propose a penalized version of the EM algorithm (PEM) accounting for an ℓ_1 penalty term. We evaluate the estimation performance of the proposed approach through a simulation exercise, where we compute the bias and the root mean square error of the estimated parameters at different quantile levels.

The relevance of our approach is also shown empirically, contributing to the increasingly widespread literature that uses quantiles as measures of risk. In the recent years, due the financial and economic crises, particular attention has been devoted to measuring and quantifying the level of financial risk and financial distress within a firm or investment portfolios. In this respect, many risk measures developed in literature are based on quantile values, like for example the Value-at-Risk; see [17,50]. Moreover, quantile regression methods turn out to be very helpful to quantify either the magnitude and the causes of riskiness; see, e.g., [18,51].

In this paper, we implement the proposed quantile regression approach to investigate the main determinants of financial distress on a sample of 2020 Italian firms. In particular, we use the definition of financial distress adopted in [5], which classifies a firm as financially distressed if its earnings before interest and taxes depreciation and amortization (EBITDA) are under the first quartile of the sample or if its leverage is above the third quartile of the sample. Hence, starting from this definition, we apply our methodology to analyze the relationships between financial distress and firms' characteristics and evaluate how it may vary when considering different (more extreme) quantiles of the distribution of leverage and EBITDA. In this way we are able to assess not only what are the main determinants for a firm's risk of financial distress, but also how these factors matter as more serious levels of distress are considered.

The rest of the paper is organized as follows. In Section 2, we introduce the main notation and briefly revise the univariate quantile regression model. Section 3 introduces the joint quantile regression framework, while Section 4 proposes the EM-based maximum likelihood approach and the related penalized EM (PEM) algorithm to estimate model parameters. In Section 5 we provide simulation results, while the empirical application is presented in Section 6. Section 7 summarizes our conclusions.

2. Preliminaries on univariate quantile regression and the AL distribution

To better explain the link between the MAL distribution and joint quantile regression, we briefly revise the univariate quantile regression model and its direct connection with the AL density. As argued in [37,53], we say that a random variable Y is distributed as an AL with location parameter μ, scale parameter $\delta > 0$, and skewness parameter $\tau \in (0,1)$, i.e., $AL(\mu,\delta,\tau)$, if its probability density function is of the form

$$f_{AL}(y;\mu,\delta,\tau) = \tau(1-\tau)\delta^{-1}\exp[-\rho_\tau\{(y-\mu)/\delta\}], \tag{1}$$

where ρ_τ denotes the so-called loss (or check) function defined by $\rho_\tau(x) = x\{\tau - \mathbf{1}(x<0)\}$, with $\mathbf{1}$ being the indicator function and where the quantity $\rho_\tau\{(y-\mu)/\delta\}$ follows an exponential distribution with rate parameter equal to $1/\delta$. Kotz et al. [36] show that the AL distribution in (1) admits a Gaussian mixture representation. In particular, if $Y \sim AL(\mu,\delta,\tau)$, then Y can be also written as

$$Y = \mu + \xi_\tau U + \theta_\tau \sqrt{\delta U} Z,$$

where U follows an exponential distribution with rate parameter $1/\delta$ and Z is a $\mathcal{N}(0,1)$ random variable. Moreover, in order to guarantee that the parameter μ coincides with the quantile of Y at a chosen level τ, the following conditions on ξ_τ and θ_τ must be satisfied:

$$\xi_\tau = (1-2\tau)/\{\tau(1-\tau)\}, \quad \theta_\tau = 2/\{\tau(1-\tau)\}.$$

Now for all $i \in \{1,\ldots,n\}$, let y_i be a response variable of interest and let $\boldsymbol{x}_i$ be a $k\times 1$ vector of covariates associated with the ith observation. Let $\mathcal{Q}_{y_i}(\tau|\boldsymbol{x}_i)$ denote the quantile regression function of y_i given $\boldsymbol{x}_i$ at a given level $\tau \in (0,1)$, and assume that the relationship between $\mathcal{Q}_{y_i}(\tau|\boldsymbol{x}_i)$ and $\boldsymbol{x}_i$ can be modeled as

$$\mathcal{Q}_{y_i}(\tau|\boldsymbol{x}_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau, \tag{2}$$

where $\boldsymbol{\beta}_\tau$ is a $k\times 1$ vector of regression coefficients. Notice that the relationship in (2) implies the linear quantile regression model in which, for all $i \in \{1,\ldots,n\}$,

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau + \epsilon_i, \tag{3}$$

where the error term ϵ_i is such that its distribution is restricted to have the τ-quantile equal to zero. If the distribution of the error term is left unspecified, then the parameter estimation proceeds by minimizing the objective function

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}\in\mathbb{R}^k}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}_\tau). \tag{4}$$

As the loss function ρ_τ is not differentiable at zero, explicit solutions for $\hat{\beta}_\tau$ cannot be derived and direct optimization is typically applied. As shown in [53], the AL distribution provides a direct connection between the minimization problem in (4) and maximum likelihood (ML) estimation. In fact, if we use the AL density as likelihood tool in (3), we have

$$\mathcal{L}\left(\boldsymbol{\beta}_\tau, \delta|\boldsymbol{y}\right) = \tau^n(1-\tau)^n\delta^{-n}\exp\left[-\sum_{i=1}^n \rho_\tau\{(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}_\tau)/\delta\}\right]. \tag{5}$$

for a given τ, and with $\delta > 0$. It is easy to verify that the minimization of the objective function in (4) with respect to the parameter $\boldsymbol{\beta}_\tau$ is equivalent to the maximization of the likelihood in (5). Therefore, the AL distribution offers a valid tool to set up the quantile regression model in a likelihood framework.

In the next section we extend such link between the AL and the quantile regression to a multivariate framework.

3. Joint quantile regression and the MAL distribution

Extending the results of the previous section, we now show how to use the MAL distribution [36] for jointly modeling marginal conditional quantiles of a multivariate response variable. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ be a p-variate response variable for each individual $i \in \{1, \ldots, n\}$ and assume that the τ_j-quantile of each of the jth components of $\boldsymbol{Y}_i$ can be modeled as a function of some k independent variables, for $j \in \{1, \ldots, p\}$.

Let $\boldsymbol{X}_i$ be a $k \times 1$ vector of regressors for the ith observation and let $\boldsymbol{\beta}_\tau = (\boldsymbol{\beta}_{\tau_1}, \ldots, \boldsymbol{\beta}_{\tau_p})^\top$ be a $p \times k$ matrix of unknown parameters, with $\boldsymbol{\beta}_{\tau_j} = (\beta_{1,\tau_j}, \ldots, \beta_{k,\tau_j})$. Then, assume that for each $i \in \{1, \ldots, n\}$, the relationship between $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ can be modeled as

$$\begin{bmatrix} \mathcal{Q}_{Y_{i1}}(\tau_1|\boldsymbol{X}_i) \\ \vdots \\ \mathcal{Q}_{Y_{ip}}(\tau_p|\boldsymbol{X}_i) \end{bmatrix} = \boldsymbol{\beta}_\tau \boldsymbol{X}_i, \tag{6}$$

where $\mathcal{Q}_{Y_{ij}}(\tau_j|\boldsymbol{X}_i)$ denotes the τ_j-level quantile regression function of Y_{ij} given $\boldsymbol{X}_i$.

Our objective is to provide joint estimation of the p marginal conditional quantiles of $\boldsymbol{Y}_i \in \mathbb{R}^p$ for each $i \in \{1, \ldots, n\}$. The representation in (6) implies the multivariate linear regression model

$$\boldsymbol{Y}_i = \boldsymbol{\beta}_\tau \boldsymbol{X}_i + \boldsymbol{\epsilon}_i \tag{7}$$

where $\boldsymbol{\epsilon}_i$ denotes a $p \times 1$ vector of error terms with univariate component-wise quantiles (at fixed levels $\tau_1, \ldots, \tau_p$, respectively) equal to zero.

For the regression model in (7), consider now the following $\mathrm{MAL}_p(\boldsymbol{\beta}_\tau \boldsymbol{X}_i, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$ distribution with density function [36]

$$f_Y(\boldsymbol{y}_i|\boldsymbol{\beta}_\tau \boldsymbol{X}_i, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}) = \frac{2\exp\{(\boldsymbol{y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top \boldsymbol{D}^{-1}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\xi}}\}}{(2\pi)^{p/2}|\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}|^{1/2}} \left(\frac{\tilde{m}}{2+\tilde{d}}\right)^{\nu/2} K_\nu\left\{\sqrt{(2+\tilde{d})\tilde{m}}\right\}, \tag{8}$$

where $\beta_\tau \boldsymbol{X}_i$ is the location parameter vector, $\boldsymbol{D}\tilde{\boldsymbol{\xi}} \in \mathbb{R}^p$ is the scale (or skew) parameter, with $\boldsymbol{D} = \mathrm{diag}(\delta_1, \ldots, \delta_p)$, $\delta_j > 0$ and $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_p)^\top$, having generic element $\tilde{\xi}_j = \left(1 - 2\tau_j\right)/\{\tau_j(1-\tau_j)\}$.

Here, $\tilde{\boldsymbol{\Sigma}}$ is a $p \times p$ positive definite matrix such that $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}$, with $\boldsymbol{\Psi}$ being a correlation matrix and $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_p)$, with $\tilde{\sigma}_j^2 = 2/\{\tau_j(1-\tau_j)\}$ for all $j \in \{1, \ldots, p\}$. Moreover,

$$\tilde{m} = (\boldsymbol{y} - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})^{-1}(\boldsymbol{y} - \boldsymbol{\beta}_\tau \boldsymbol{X}_i), \quad \tilde{d} = \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\xi}},$$

and K_ν denotes the modified Bessel function of the third kind with index parameter $\nu = (2-p)/2$.

Using (7) and (8), and following [36], the $\mathrm{MAL}_p(\boldsymbol{\beta}_\tau \boldsymbol{X}_i, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$ can be written as a location–scale mixture, having the representation

$$\boldsymbol{Y}_i = \boldsymbol{\beta}_\tau \boldsymbol{X}_i + \boldsymbol{D}\tilde{\boldsymbol{\xi}} W + \sqrt{W}\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z}, \tag{9}$$

where $\boldsymbol{Z} \sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{I}_p)$ denotes a p-variate standard Normal distribution and $W \sim \mathcal{E}(1)$ has a standard exponential distribution, with $\boldsymbol{Z}$ being independent of W.

It is worth noticing that the following constraints, imposed for all $j \in \{1, \ldots, p\}$,

$$\tilde{\xi}_j = (1-2\tau_j)/\{\tau_j(1-\tau_j)\} \quad \text{and} \quad \tilde{\sigma}_j^2 = 2/\{\tau_j(1-\tau_j)\}, \tag{10}$$

represent necessary conditions to guarantee that the model in (6) holds, as shown in the proposition below, whose proof is given, as all others, in Appendix.

Proposition 1. *Let $\boldsymbol{Y} \sim \mathrm{MAL}_p(\boldsymbol{\mu}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$ as defined in (9), and let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)^\top$ be a fixed $p \times 1$ vector, such that $\tau_1, \ldots, \tau_p \in (0, 1)$, then $\Pr(Y_j < \mu_j) = \tau_j$ if and only if*

$$\tilde{\xi}_j = (1-2\tau_j)/\{\tau_j(1-\tau_j)\} \quad \textit{and} \quad \tilde{\sigma}_j^2 = 2/\{\tau_j(1-\tau_j)\}. \tag{11}$$

In addition, $Y_j \sim \mathrm{AL}(\mu_j, \tau_j, \delta_j)$.

Notice that, the representation in (8), under (10), is a suitable reparameterization of [36]. Indeed the introduction of the diagonal matrix $\boldsymbol{D}$ ensures that each δ_j represents the scale parameter of the marginal AL distribution of Y_j, for every $j \in \{1, \ldots, p\}$; see the proof of Proposition 1 in Appendix for further details.

It is also worth adding a brief comment on parameters identifiability of the MAL distribution, as one could reasonably ask whether $\boldsymbol{D}\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}$ are uniquely identified. In the next proposition, we argue that the constraints (10) in (9) are necessary conditions for model identifiability, for any fixed quantile level $\tau_1, \ldots, \tau_p$.

Proposition 2. *Let $\boldsymbol{Y} \in \mathbb{R}^p$ be distributed as a* $\mathrm{MAL}_p(\boldsymbol{\mu}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$*, where* $\boldsymbol{D} = \mathrm{diag}(\delta_1, \ldots, \delta_p)$*,* $\delta_j > 0$*,* $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_p)^\top$*, with* $\tilde{\xi}_j = (1 - 2\tau_j)/\{\tau_j(1-\tau_j)\}$ *being known for any fixed value of* τ_j*. Furthermore,* $\tilde{\boldsymbol{\Sigma}}$ *is a* $p \times p$ *positive definite matrix such that* $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}$*, with* $\boldsymbol{\Psi}$ *being an unknown correlation matrix and* $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}(\tilde{\sigma}_1, \ldots, \tilde{\sigma}_p)$*, with fixed element* $\tilde{\sigma}_j^2 = 2/\{\tau_j(1-\tau_j)\}$*, for every* $j \in \{1, \ldots, p\}$*. Then, the parameters* $\boldsymbol{D}$ *and* $\boldsymbol{\Psi}$ *are uniquely identified.*

In other words, the constraints in (10) represent essential conditions not only to retrieve the joint quantile regression model in (6), as discussed in Proposition 1, but also to ensure that the model does not suffer from identifiability problems; see Proposition 2.

4. Maximum likelihood estimation

As shown in the previous sections, the MAL density represents a convenient tool to jointly model marginal conditional quantiles of a multivariate response variable in a quantile regression framework. In this section, we introduce a maximum likelihood approach to estimate and make inference on model parameters. We propose a suitable likelihood-based EM algorithm [16], showing that model parameters can be easily obtained in closed form, hence facilitating the computational burden of the algorithm compared to the direct maximization approach. Moreover, as mentioned in the Introduction, given the possible high dimensionality problem in multivariate settings, we also propose a penalized version (PEM) of the EM algorithm by considering a LASSO regularization approach [47]. The procedure essentially modifies the M-step of the EM algorithm by introducing a penalty term and provides a data-driven procedure for variable selection.

4.1. The EM algorithm

In this section we propose a maximum likelihood-based approach to estimate the parameters of the quantile regression models defined in (6). Specifically, we derive a new EM algorithm by exploiting the Gaussian location–scale mixture representation (9) of the MAL distribution, under the constraints in (10). Although not in the context of joint quantile regression, the EM algorithm has been intensively exploited in the context of mixture models. For example, Franczak et al. [19] derive an EM algorithm for mixture of Shifted Asymmetric Laplace (SAL) distributions. Other extensions to generalized hyperbolic mixtures have been also proposed in [48] and references therein.

The EM algorithm essentially alternates between performing an expectation (E) step, which defines the expectation of the complete log-likelihood function evaluated using the current estimate for 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 estimators are given below in the following two propositions. All the proofs are collected in Appendix.

For sake of clarity, in the following we introduce the notation $\boldsymbol{D}_{(\delta)}$ and $\tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})}$ to make clear that the matrices $\boldsymbol{D}$ and $\tilde{\boldsymbol{\Sigma}}$ depend on the parameters $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^\top$ and $\boldsymbol{\Psi}$, respectively.

Proposition 3. *For a given vector* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)^\top$*, let* $\boldsymbol{D}_{(\delta)} = \mathrm{diag}(\boldsymbol{\delta})$*, with* $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^\top$ *and let* $\tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} = \tilde{\boldsymbol{\Lambda}}\Psi\tilde{\boldsymbol{\Lambda}}$ *with* $\tilde{\boldsymbol{\Lambda}}$ *subject to the constrains in (11). For every* $i \in \{1, \ldots, n\}$*, let*

$$\tilde{m}_i = (\boldsymbol{y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_{(\delta)})^{-1} (\boldsymbol{y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i),$$

and define $\tilde{d} = \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \tilde{\boldsymbol{\xi}}$*. Then, the expected complete log-likelihood function (up to additive constants) is*

$$\mathrm{E}\{\ell_c(\boldsymbol{\beta}_\tau, \boldsymbol{D}_{(\delta)}, \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} | \boldsymbol{Y}_i, \hat{\boldsymbol{\beta}}_\tau, \boldsymbol{D}_{(\hat{\delta})}, \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})})\} \tag{12}$$

$$= -\frac{n}{2} \ln |\boldsymbol{D}_{(\hat{\delta})} \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})} \boldsymbol{D}_{(\hat{\delta})}| + \sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau X_i)^\top \boldsymbol{D}_{(\hat{\delta})}^{-1} \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})}^{-1} \tilde{\boldsymbol{\xi}} \tag{13}$$

$$- \frac{1}{2} \sum_{i=1}^{n} z_i (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\hat{\delta})} \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})} \boldsymbol{D}_{(\hat{\delta})})^{-1} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i) \tag{14}$$

$$- \frac{1}{2} \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})} \tilde{\boldsymbol{\xi}} \sum_{i=1}^{n} u_i, \tag{15}$$

where

$$u_i = \left(\frac{\hat{\tilde{m}}_i}{2 + \hat{\tilde{d}}} \right)^{1/2} \frac{K_{\nu+1}\left\{\sqrt{(2 + \hat{\tilde{d}})\hat{\tilde{m}}_i}\right\}}{K_\nu\left\{\sqrt{(2 + \hat{\tilde{d}})\hat{\tilde{m}}_i}\right\}}, \qquad z_i = \left(\frac{2 + \hat{\tilde{d}}}{\hat{\tilde{m}}_i} \right)^{1/2} \frac{K_{\nu+1}\left\{\sqrt{(2 + \hat{\tilde{d}})\hat{\tilde{m}}_i}\right\}}{K_\nu\left\{\sqrt{(2 + \hat{\tilde{d}})\hat{\tilde{m}}_i}\right\}} - \frac{2\nu}{\hat{\tilde{m}}_i}$$

with

$$\hat{\tilde{m}}_i = (\boldsymbol{y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\hat{\delta})} \tilde{\boldsymbol{\Sigma}}_{\hat{\Psi}} \boldsymbol{D}_{(\hat{\delta})})^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i), \quad \hat{\tilde{d}} = \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\hat{\Psi})} \tilde{\boldsymbol{\xi}}.$$

For a given $\boldsymbol{\tau}$, maximizing the expectation of the complete data log-likelihood in Eqs. (13)–(15) with respect to the parameters $\boldsymbol{\beta}_\tau$, $\boldsymbol{\Psi}$, and $\boldsymbol{\delta}$ leads to the following M-step updates.

Proposition 4. *Given the vector $\boldsymbol{\tau}$, the values of $\boldsymbol{\beta}_\tau$, $\tilde{\boldsymbol{\Sigma}}_{(\Psi)}$ and $\boldsymbol{\delta}$ maximizing (13)–(15) are*

$$\hat{\boldsymbol{\beta}}_\tau^\top = \left(\sum_{i=1}^n z_i \boldsymbol{X}_i \boldsymbol{X}_i^\top \right)^{-1} \left(\sum_{i=1}^n z_i \boldsymbol{X}_i \boldsymbol{Y}_i^\top - \sum_{i=1}^n \boldsymbol{X}_i \tilde{\boldsymbol{\xi}}^\top \boldsymbol{D}_{(\hat{\delta})} \right), \tag{16}$$

and

$$\tilde{\boldsymbol{\Sigma}}_{(\hat{\Psi})} = \frac{1}{n} \sum_{i=1}^n z_i \boldsymbol{D}_{(\hat{\delta})}^{-1} (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i)^\top \boldsymbol{D}_{(\hat{\delta})}^{-1} + \frac{1}{n} \sum_{i=1}^n u_i \tilde{\boldsymbol{\xi}} \tilde{\boldsymbol{\xi}}^\top - \frac{2}{n} \boldsymbol{D}_{\hat{\delta}}^{-1} \sum_{i=1}^n (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i)^\top \tilde{\boldsymbol{\xi}}^\top, \tag{17}$$

while the estimation of $\boldsymbol{\delta}$ is obtained through a numerical optimization, by solving the non linear first order condition

$$\sum_{i=1}^n (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i) \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\hat{\Psi})} + n \boldsymbol{D}_{(\hat{\delta})} - \sum_{i=1}^n Z_i (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i)^\top \boldsymbol{D}_{(\hat{\delta})}^{-1} \tilde{\boldsymbol{\Sigma}}_{(\hat{\Psi})} = 0_{p \times p} \tag{18}$$

Therefore, the EM algorithm can be implemented as follows:

E-step: Set the iteration number $h = 1$. Fix the vector $\boldsymbol{\tau}$ at the chosen quantile levels $\tau_1, \ldots, \tau_p$ of interest and initialize the parameter $\boldsymbol{\theta} = (\boldsymbol{\beta}_\tau, \Psi, \boldsymbol{\delta})$, deriving $\tilde{\boldsymbol{\Sigma}}_{(\Psi)}$ and $\boldsymbol{D}_{(\delta)}$. Then, given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(h)}$, calculate the weights

$$u_i^{(h)} = \left(\frac{m_i^{(h)}}{2 + \tilde{d}^{(h)}} \right)^{1/2} \frac{K_{\nu+1} \left\{ \sqrt{(2 + \tilde{d}^{(h)}) \tilde{m}_i^{(h)}} \right\}}{K_\nu \{ \sqrt{(2 + \tilde{d}^{(h)}) \tilde{m}_i^{(h)}} \}} \tag{19}$$

$$z_i^{(h)} = \left(\frac{2 + \tilde{d}^{(h)}}{\tilde{m}_i^{(h)}} \right)^{1/2} \frac{K_{\nu+1} \left\{ \sqrt{(2 + \tilde{d}^{(h)}) \tilde{m}_i^{(h)}} \right\}}{K_\nu \left\{ \sqrt{(2 + \tilde{d}^{(h)}) \tilde{m}_i^{(h)}} \right\}} - \frac{2\nu}{\tilde{m}_i^{(h)}} \tag{20}$$

where

$$\tilde{m}_i^{(h)} = (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau^{(h)} \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\delta)}^{(h)} \tilde{\boldsymbol{\Sigma}}_{(\Psi)}^{(h)} \boldsymbol{D}_{(\delta)}^{(h)})^{-1} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau^{(h)} \boldsymbol{X}_i),$$

and $\tilde{d}^{(h)} = \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\Psi)}^{(h)} \tilde{\boldsymbol{\xi}}$ for all $i \in \{1, \ldots, n\}$.

M-step: Use $u_i^{(h)}$ and $z_i^{(h)}$ to maximize $\mathrm{E}\{\ell_c(\boldsymbol{\theta}|\boldsymbol{\theta}^{(h)})\}$ with respect to $\boldsymbol{\theta}$, and obtain the new parameter estimates as

$$\hat{\boldsymbol{\beta}}_\tau^{\top(h+1)} = \left(\sum_{i=1}^n z_i^{(h)} \boldsymbol{X}_i \boldsymbol{X}_i^\top \right)^{-1} \left(\sum_{i=1}^n z_i^{(h)} \boldsymbol{X}_i \boldsymbol{Y}_i^\top - \sum_{i=1}^n X_i \tilde{\boldsymbol{\xi}}^\top \boldsymbol{D}_{(\hat{\delta})}^{(h)} \right) \tag{21}$$

and

$$\tilde{\boldsymbol{\Sigma}}_{(\hat{\Psi})}^{(h+1)} = \frac{1}{n} \sum_{i=1}^n z_i^{(h)} \boldsymbol{D}_{(\hat{\delta})}^{-1(h)} (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau^{(h+1)} \boldsymbol{X}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau^{(h+1)} \boldsymbol{X}_i)^\top \boldsymbol{D}_{(\hat{\delta})}^{-1(h)} \tag{22}$$

$$+ \frac{1}{n} \sum_{i=1}^n u_i^{(h)} \tilde{\boldsymbol{\xi}} \tilde{\boldsymbol{\xi}}^\top - \frac{2}{n} \boldsymbol{D}_{(\hat{\delta})}^{-1(h)} \sum_{i=1}^n (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau^{(h+1)} \boldsymbol{X}_i)^\top \tilde{\boldsymbol{\xi}}^\top \tag{23}$$

while $\boldsymbol{D}_{(\delta)}^{(h+1)}$ is obtained as the solution of the equation

$$\sum_{i=1}^n (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau^{(h+1)} \boldsymbol{X}_i) \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\hat{\Psi})}^{(h+1)} + n \boldsymbol{D}_{(\delta)}$$

$$- \sum_{i=1}^n z_i^{(h)} (\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau^{(h+1)} \boldsymbol{X}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\beta}}_\tau^{(h+1)} \boldsymbol{X}_i)^\top \boldsymbol{D}_{(\delta)}^{-1} \tilde{\boldsymbol{\Sigma}}_{(\hat{\Psi})}^{(h+1)} = \boldsymbol{0}_{p \times p}. \tag{24}$$

The procedure is iterated until convergence, i.e., when the difference between the likelihood function evaluated at two consecutive iterations is small enough. In our paper, we set this convergence criterion equal to 10^{-6}. In order to avoid possible premature convergence of the algorithm, we also implement the procedure of [9] as alternative stopping rule. Since in our case all the parameters are available in closed form, in both the cases the convergence of the algorithm is very fast and estimation results remain unchanged.

Notice that all the parameter estimates in Eqs. (21)–(24) account for the multivariate structure of the data through the weights u_i and z_i in (19) and (20), which depend on the index ν as a function of p. In the univariate case, $\nu = 1/2$ and the estimators reduce to the case of [45].

4.2. Variable selection and the penalized EM (PEM) algorithm

As mentioned before, when dealing with high dimensional statistical problems, Lasso penalized procedures represent possible solutions to detect significant predictors from a large pool of candidate variables. Therefore, in this section we introduce a penalized version of the EM algorithm described in Section 4.1. The PEM algorithm was originally proposed by Green [22] to allow for the maximization of a difficult-to-calculate penalized likelihood. Compared to the EM, the PEM algorithm leaves the E-step unchanged and modifies the M-step with a penalty function introduced to achieve shrinkage (e.g., ridge regression, as in [25]), variable selection (e.g., Lasso, as in [47]) or simultaneous shrinkage and variable selection (e.g., elastic net, as proposed in [54]).

For a chosen level $\boldsymbol{\tau}$, let us denote by $\boldsymbol{\theta} = (\boldsymbol{\beta}_\tau, \boldsymbol{D}_{(\delta)}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\Psi}})$ the parameter set. Then, consider the penalized-M (PM) step

$$\boldsymbol{\theta}^{(h+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left\{ Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(h)}) - \lambda \sum_{j=1}^{p} \sum_{s=1}^{k} |\beta_{js,\tau_j}| \right\}, \tag{25}$$

where

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(h)}) = -\frac{n}{2} \ln |\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_{(\delta)}| + \sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau X_i)^\top \boldsymbol{D}_{(\delta)}^{-1} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})}^{-1} \tilde{\boldsymbol{\xi}} \tag{26}$$

$$- \frac{1}{2} \sum_{i=1}^{n} z_i (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_{(\delta)})^{-1} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i) \tag{27}$$

$$- \frac{1}{2} \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \tilde{\boldsymbol{\xi}} \sum_{i=1}^{n} u_i. \tag{28}$$

with all the quantities defined as in Section 4.1. In (25), the quantity

$$\lambda \sum_{j=1}^{p} \sum_{s=1}^{k} |\beta_{js,\tau_j}|$$

represents a convex penalty function, where λ is a tuning parameter that regulates the strength of the penalization assigned to the coefficients in the model. The choice of λ is made by performing cross-validation techniques which allow us to consider λ as a data-driven parameter. In particular, we compute the solutions for a decreasing sequence of values for λ, starting from the smallest value $\lambda_{\max}$ for which the entire vector $\hat{\boldsymbol{\beta}}_k = 0$. We then select a minimum value $\lambda_{\min} = \epsilon \lambda_{\max}$ and construct a sequence of m values of λ, decreasing from $\lambda_{\max}$ to $\lambda_{\min}$ on the log scale. This would lead to a more stable algorithm; see, e.g., [20].

Notice that, even though the algorithm penalizes only the M-step, the expressions of the weights u_i and z_i in the E-step will be indirectly affected by the new (penalized) estimates in the M-step at each iteration.

In the next section we assess the performance of the EM and PEM algorithms using a simulation exercise.

5. Simulation study

In the following sections we conduct a simulation study to evaluate the small sample properties of the proposed methods. The idea of this exercise is to show that both the EM and PEM algorithms represent valid procedures to estimate the quantile regression coefficients, regardless of the true data generation process.

5.1. Joint quantile regression

In this section we asses the performance of our estimation procedure by simulating a joint quantile regression model as described in Section 3. For this purpose, we consider a simple case of $n = 1000$ units, of dimension $p = 3$ and two explanatory variables. The observations are generated using the following data generating process. For each $i \in \{1, \ldots, n\}$,

$$\boldsymbol{Y}_i = \boldsymbol{\beta}_{\tau,0} + \boldsymbol{\beta}_{\tau,1} X_{i2} + \boldsymbol{\beta}_{\tau,2} X_{i3} + \boldsymbol{\epsilon}_i.$$

Table 1
Parameter estimates at different quantile levels: Point estimates and relative bias.

	$\hat{\beta}_{01}$	$\hat{\beta}_{02}$	$\hat{\beta}_{03}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{13}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$	$\hat{\beta}_{23}$
	Panel A: $\boldsymbol{\tau} = (0.50, 0.50, 0.50)^\top$								
$\boldsymbol{\epsilon}_i \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\xi}\tilde{\xi}^\top \boldsymbol{D} + \boldsymbol{D}\tilde{\Sigma}\boldsymbol{D})$									
Estimate	−0.382	1.984	0.667	−0.374	0.653	1.083	0.713	0.771	0.586
Bias (%)	−0.063	−0.423	−0.325	0.560	0.443	0.423	−0.199	0.871	0.305
$\boldsymbol{\epsilon}_i \sim t_3(\boldsymbol{D}\tilde{\Sigma}\boldsymbol{D}, \boldsymbol{D}\tilde{\xi})$									
Estimate	−0.379	1.980	0.666	−0.369	0.646	1.072	0.706	0.756	0.508
Bias (%)	0.632	0.105	0.698	0.749	0.367	0.452	0.778	0.056	0.898
	Panel B: $\boldsymbol{\tau} = (0.25, 0.50, 0.75)^\top$								
$\boldsymbol{\epsilon}_i \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\xi}\tilde{\xi}^\top \boldsymbol{D} + \boldsymbol{D}\tilde{\Sigma}\boldsymbol{D})$									
Estimate	−0.394	1.992	0.697	−0.371	0.647	1.083	0.714	0.768	0.590
Bias (%)	3.141	−0.030	4.030	−0.166	−0.478	0.405	−0.175	0.609	1.189
$\boldsymbol{\epsilon}_i \sim t_3(\boldsymbol{D}\tilde{\Sigma}\boldsymbol{D}, \boldsymbol{D}\tilde{\xi})$									
Estimate	−.0418	2.023	0.709	−0.373	0.655	1.086	0.714	0.764	0.589
Bias (%)	7.515	1.501	4.776	0.337	0.817	0.664	0.510	0.448	0.956
	Panel C: $\boldsymbol{\tau} = (0.10, 0.50, 0.90)^\top$								
$\boldsymbol{\epsilon}_i \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\xi}\tilde{\xi}^\top \boldsymbol{D} + \boldsymbol{D}\tilde{\Sigma}\boldsymbol{D})$									
Estimate	−0.431	2.006	0.731	−0.363	0.651	1.065	0.771	0.764	0.576
Bias (%)	12.821	0.669	9.104	−1.413	0.181	−1.218	−0.539	0.051	−1.271
$\boldsymbol{\epsilon}_i \sim t_3(\boldsymbol{D}\tilde{\Sigma}\boldsymbol{D}, \boldsymbol{D}\tilde{\xi})$									
Estimate	−0.436	2.012	0.755	−0.386	0.659	1.100	0.717	0.764	0.590
Bias (%)	14.222	0.974	12.694	2.748	1.407	1.987	1.013	0.435	0.987

The covariates are randomly drawn from a standard Normal distribution. The true value of $\boldsymbol{\beta}_\tau$ is set equal to

$$\boldsymbol{\beta}_\tau = \begin{bmatrix} -0.382 & -0.372 & 0.715 \\ 1.993 & 0.650 & 0.764 \\ 0.670 & 1.079 & 0.584 \end{bmatrix}. \tag{29}$$

We analyze three different quantile vectors. In the first case, we assume $\boldsymbol{\tau} = (0.50, 0.50.0.50)^\top$, which implies that $\tilde{\boldsymbol{\xi}} = (0, 0, 0)^\top$ and $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}(2.828, 2.828, 2.828)$. In the second scenario, we set $\boldsymbol{\tau} = (0.25, 0.50, 0.75)^\top$ and, consequently, $\tilde{\boldsymbol{\xi}} = (2.667, 0, -2.667)^\top$ and $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}(3.266, 2.828, 3.266)$. Finally, we consider a more extreme case with $\boldsymbol{\tau} = (0.90, 0.50, 0.10)^\top$, $\tilde{\boldsymbol{\xi}} = (8.889, 0, -8.889)^\top$ and $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}(4.714, 2.828, 4.714)$. For each of the three cases, the parameter vector is represented by $\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \delta_1, \delta_2, \delta_3, \rho_{12}, \rho_{13}, \rho_{23})$, where the true values of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, are defined by the columns of $\boldsymbol{\beta}_\tau$ in (29) and where we set $\delta_1 = 0.13$, $\delta_2 = 0.30$, $\delta_3 = 0.23$, with $\rho_{12} = 0.50$, $\rho_{13} = 0.30$ and $\rho_{23} = 0.40$.

Two different distributions for the error term generating process are considered in each simulation study:

(a) a multivariate Normal random variable with zero mean and a variance–covariance matrix equal to $(\boldsymbol{D}\tilde{\xi}\tilde{\xi}^\top D + \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$, and
(b) a multivariate Student t distribution with 3 degrees of freedom, scale parameter $\boldsymbol{D}\tilde{\Sigma}D$ and non centrality parameter equal to $\boldsymbol{D}\tilde{\xi}$.

For each distribution of the error term, we carry out $B = 500$ Monte Carlo replications and report the relative bias and the root mean square error (RMSE), averaged across the 500 simulations, for each parameter value in $\boldsymbol{\theta}$. The results are shown in Tables 1 and 2.

Table 1 analyzes the regression coefficient estimates for the three quantile levels described in Panel A, Panel B and Panel C. For each of the three panels, the two error term distributions are considered and the corresponding point estimates and percentage bias are reported. As can be easily inferred, the bias effect is quite small when we analyze the median levels (see Panel A). As the quantile levels become more extreme (see Panels B and C), the bias slightly increases but it still remains reasonably small.

In Table 2 we report the Root Mean Square Error (RMSE) of the regression coefficients for the same $\boldsymbol{\tau}$-levels under the same error distributions of Table 1, and compare the results obtained by running both the proposed joint quantile regression (*Joint QR RMSE* in the table) and the univariate quantile regressions (*Univariate QR RMSE* in the table) separately for each marginal Y_j. In this way we want to highlight the added value of using the MAL distribution for multiple quantile regression purpose, which accounts for potential correlation among the responses. It is worth noting that in all the simulation situations, the effective gain of the proposed joint approach is reflected in the smaller RMSE of the estimates compared to those in the univariate case.

Table 2
Joint quantile estimation versus univariate quantile regressions: RMSEs.

	$\hat\beta_{01}$	$\hat\beta_{02}$	$\hat\beta_{03}$	$\hat\beta_{11}$	$\hat\beta_{12}$	$\hat\beta_{13}$	$\hat\beta_{21}$	$\hat\beta_{22}$	$\hat\beta_{23}$
	Panel A: $\boldsymbol{\tau} = (0.50, 0.50, 0.50)^\top$								
$\boldsymbol{\epsilon}_i \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\top\boldsymbol{D} + \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$									
Joint QR RMSE	0.022	0.046	0.038	0.024	0.049	0.040	0.020	0.048	0.032
Univariate QR RMSE	0.023	0.049	0.039	0.029	0.053	0.041	0.020	0.049	0.034
$\boldsymbol{\epsilon}_i \sim t_3(\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}, \boldsymbol{D}\tilde{\boldsymbol{\xi}})$									
Joint QR RMSE	0.028	0.060	0.050	0.027	0.067	0.049	0.030	0.051	0.034
Univariate QR RMSE	0.034	0.074	0.058	0.030	0.077	0.044	0.027	0.067	0.059
	Panel B: $\boldsymbol{\tau} = (0.25, 0.50, 0.75)^\top$								
$\boldsymbol{\epsilon}_i \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\top\boldsymbol{D} + \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$									
Joint QR RMSE	0.467	0.049	0.283	0.029	0.049	0.049	0.031	0.044	0.049
Univariate QR RMSE	0.470	0.055	0.324	0.029	0.049	0.051	0.038	0.048	0.051
$\boldsymbol{\epsilon}_i \sim t_3(\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}, \boldsymbol{D}\tilde{\boldsymbol{\xi}})$									
Joint QR RMSE	0.428	0.059	0.292	0.053	0.049	0.096	0.056	0.106	0.041
Univariate QR RMSE	0.491	0.061	0.572	0.056	0.049	0.103	0.061	0.196	0.049
	Panel C: $\boldsymbol{\tau} = (0.10, 0.50, 0.90)^\top$								
$\boldsymbol{\epsilon}_i \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\top\boldsymbol{D} + \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$									
Joint QR RMSE	0.981	0.055	0.835	0.063	0.039	0.130	0.062	0.045	0.131
Univariate QR RMSE	1.063	0.05	3.274	0.087	0.048	0.140	0.075	0.047	0.142
$\boldsymbol{\epsilon}_i \sim t_3(\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}, \boldsymbol{D}\tilde{\boldsymbol{\xi}})$									
Joint QR RMSE	0.997	0.101	0.936	0.093	0.241	0.165	0.065	0.048	0.152
Univariate QR RMSE	0.692	0.135	1.228	0.097	0.243	0.182	0.071	0.053	0.169

Table 3
The performance of the PEM algorithm: True Positive Rate (TPR).

	$\hat\beta_{12}$	$\hat\beta_{14}$	$\hat\beta_{23}$	$\hat\beta_{24}$	$\hat\beta_{32}$	$\hat\beta_{33}$
$\boldsymbol{\epsilon}_i \sim \mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\top\boldsymbol{D} + \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$						
$\boldsymbol{\tau} = (0.50, 0.50, 0.50)^\top$	89.122	92.475	84.341	89.342	91.753	89.134
$\boldsymbol{\tau} = (0.25, 0.50, 0.75)^\top$	87.221	83.453	79.978	86.397	83.641	87.550
$\boldsymbol{\tau} = (0.10, 0.50, 0.90)^\top$	83.341	82.512	80.123	86.101	82.308	86.761
$\boldsymbol{\epsilon}_i \sim t_3(\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}, \boldsymbol{D}\tilde{\boldsymbol{\xi}})$						
$\boldsymbol{\tau} = (0.50, 0.50, 0.50)^\top$	88.324	89.546	84.209	86.008	88.331	88.198
$\boldsymbol{\tau} = (0.25, 0.50, 0.75)^\top$	86.130	83.346	80.001	87.121	82.453	85.345
$\boldsymbol{\tau} = (0.10, 0.50, 0.90)^\top$	83.007	81.978	81.321	86.004	80.423	84.121

5.2. Penalized joint quantile regression

In this section, a simulation study is proposed to evaluate the performance of the penalized joint quantile regression, which uses the PEM algorithm proposed in Section 4.2. We analyze a simple case with $n = 1000$, $p = 3$ and a set of four explanatory variables using the same data generating process of the previous section. For a fixed $\boldsymbol{\tau}$-level, the matrix of regression coefficients $\boldsymbol{\beta}_\tau$ contains 15 elements, where we set six of them (namely, $\beta_{12}, \beta_{14}, \beta_{23}, \beta_{24}, \beta_{32}, \beta_{33}$) equal to zero.

As in the previous section, we analyze three different quantile vectors, i.e., $\boldsymbol{\tau} = (0.50, 0.50, 0.50)^\top$, $\boldsymbol{\tau} = (0.25, 0.50, 0.75)^\top$, and $\boldsymbol{\tau} = (0.10, 0.50, 0.90)$. For each of the three cases we perform 100 Monte Carlo simulations, under either the $\mathcal{N}_3(\mathbf{0}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}\tilde{\boldsymbol{\xi}}^\top\boldsymbol{D} + \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D})$ and the $t_3(\boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}, \boldsymbol{D}\tilde{\boldsymbol{\xi}})$ distributions as possible data generating process. Then, for each case, we estimate the model parameters using the penalized objective function defined in (25). The estimation of the tuning parameter λ is obtained using a 10-fold cross validation method, where the initial grid of the possible values for $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ has been described in Section 4.2.

Table 3 reports the true positive rate (TPR) for each of the true coefficients initially set equal to zero. The TPR gives a measure of how sensitive a given method is at discovering non-zero entries and we calculate it as the ratio between the number of simulations that correctly identify the parameter as a zero value, over the total number of simulations (i.e., the number of true zeros for each coefficient).

The results show that the PEM method performs quite well, with an average TPR of more than the 80% across the three simulation scenarios and regardless of the distributional assumption of the errors.

Table 4
Estimated coefficients at different quantile levels and standard errors.

	Leverage	EBITDA	Leverage	EBITDA
	Panel A: $\boldsymbol{\tau} = (0.75, 0.25)^\top$		Panel B: $\boldsymbol{\tau} = (0.90, 0.10)^\top$	
Constant	−**4.472** (0.501)	−**0.321** (0.105)	−**4.359** (0.557)	−**0.274** (0.087)
Profit	−**0.022** (0.003)	−**0.024** (0.001)	−**0.022** (0.003)	−**0.024** (0.005)
Finexp	**13.598** (0.538)	**0.451** (0.107)	**15.849** (0.567)	**0.312** (0.091)
Earnings	**1.118** (0.509)	**0.748** (0.110)	**0.984** (0.571)	**0.621** (0.092)
Employee	−0.006 (0.005)	−0.001 (0.001)	−0.003 (0.005)	−**0.002** (0.001)
Fixasset	**2.445** (0.132)	**0.187** (0.027)	**2.907** (0.119)	**0.151** (0.020)
Netincome	−0.371 (1.654)	0.347 (0.337)	−**3.816** (1.658)	0.345 (0.245)
Current debt	**0.107** (0.008)	−**0.009** (0.002)	**0.122** (0.008)	−**0.007** (0.001)
Cashflow	−**4.089** (1.376)	**6.251** (0.251)	−**4.704** (1.256)	**4.407** (0.212)
δ_j	**0.952** (0.018)	**0.190** (0.003)	**0.800** (0.012)	**0.138** (0.002)
ρ_{12}		−**0.131** (0.027)	−0.005 (0.032)	
n	2020	2020	2020	2020

6. Empirical application

As stated in the Introduction, recent financial and economic crises have put the need for a thorough analysis of the causes and the effects of the financial distress. In this context, quantile regression has turned out to be an effective framework to study and evaluate financial stability of systems. In line with the recent literature that links quantile regression with measures of financial risks (see, e.g., [1,15,18,52]), in this section we use our methodology to identify the main determinants for a risk of financial distress. We use data on 2020 private limited non-financial Italian firms from the Amadeus Bureau van Dijk dataset, with reference year 2015. Following [5,43], among others, we adopt a definition of financial distress that evaluates the firm's capacity to satisfy its financial obligations. Specifically, they classify a company as financially distressed not only when it files for bankruptcy but also when the two following events occur: (1) its earnings before interest and taxes depreciation and amortization (EBITDA) are under the first quartile of the sample or (2) the firm's leverage is above the third quartile of the sample.

The idea of this real data application is to use our joint estimation approach to study the entire distribution of both firm's leverage and EBITDA and assess how the impact of firm's characteristics (such as profitability, financial expenses and earnings) varies with different quantile levels. This allows us to identify not only the main determinants of firm's risk of financial distress, but also how the effect of these factors may vary depending on the severity of the distress a firm is facing.

As a measure of firm's leverage (*leverage*) we use the ratio between firm's total asset and equity. As explanatory variables, we consider indexes of firm's profitability (*profit*), financial expenses (*finexp*), and retained earnings (*earnings*), as suggested in [43]. We also consider the impact of short-term debt (*current debt*) and the ratio between firm's cashflow and total assets (*cashflow*), and control for other firm's specific characteristics such as firms' total fixed assets over total asset ratio (*fixassets*, which can be interpreted as an indicator of firm's collaterals), firm's net income scaled by total assets (*netincome*, as a proxi of the activity level of the firm), and size, measured by the number of employees in the firm (*employees*).

The results for the two cases when $\boldsymbol{\tau} = (0.75, 0.25)^\top$ and $\boldsymbol{\tau} = (0.90, 0.10)^\top$ are reported in Table 4, Panel A and Panel B, respectively. In the table, parameter estimates are displayed in boldface when significant at the standard 5% level. Standard errors of the estimates are computed using nonparametric bootstrap (see, e.g., [21]) and are reported in brackets.

We find a negative relationship between profitability and financial distress, as the coefficient of profitability (*profit*) shows a significant and negative impact on both EBITDA and leverage. This is in line with the results in [10,43], who argue that firms that face financial distress are most likely unable to fulfill their financial obligations. This effect remains constant also at more extreme quantile levels of the distribution, as shown in Panel B of Table 4.

We also find a positive effect of financial expenses, whose magnitude is amplified especially for the leverage component. This effect is still consistent with the findings in [43] and confirms the expectations that the risk of financial distress increases as the firm's risk of not being able to comply with its financial obligations rises. This effect captures the firm's financial vulnerability, which increases when considering more extreme (riskier) values of leverage.

In analyzing the effect of retained earnings on financial distress, we find evidence of a positive impact on both EBITDA and leverage, even though it decreases as riskier quantiles are analyzed. This is in contrast with the expected results in the literature, where a negative relationship with financial distress likelihood is typically postulated, as a firm should have a lower capacity of self-financing during periods of higher financial stress. However, several papers have documented the relationship between financial distress and earnings management practices in economy like Italy, and find evidence that private companies experiencing financial distress tend to manipulate their earnings to portray better financial performance and obtain bank financing; see, e.g., [7]. In this case, earnings should not be considered a very informative determinant of firm's financial distress.

Table 5
LASSO parameter estimates.

	EBITDA	Leverage	EBITDA	Leverage
	Panel A: $\boldsymbol{\tau} = (0.75, 0.25)^\top$		Panel B: $\boldsymbol{\tau} = (0.90, 0.10)^\top$	
Constant	−3.085	−0.408	−4.566	−0.364
Profit	−0.012	−0.343	−0.014	−0.038
Finexp	12.624	0.019	16.222	0.014
Earnings	0.478	1.422	1.376	0.389
Employee	–	–	–	–
Fixasset	2.038	0.307	1.067	0.235
Netincome	−0.676	–	−1.267	–
Current debt	0.721	−0.330	0.220	−0.107
Cashflow	−3.035	5.606	−2.072	5.236
δ_j	0.937	0.831	0.827	0.135
ρ_{12}	−0.114		−0.006	
λ	2.259		5.612	
n	2020	2020	2020	2020

Another important factor in explaining financial distress is the ratio between firm's fixed asset over total asset. The claim is typically that tangible assets tend to reduce the financial distress costs because of the liquidation possibility in case of default; see, e.g., [12]. Hence, the higher is the risk of a financial distress, the higher is the level of tangible assets a firm will have on its balance sheet. Our results confirm this evidence, showing a positive effect of *fixasset* on both EBITDA and leverage.

Firm's cash flow over total asset ratio is also a significant predictor of financial distress. Its impact on leverage is highly negative, while it is found to be positive on the EBITDA component, with a sensible decrease when moving towards more extreme quantile. This is could be due to the fact that firms with low cash flow are less likely to make leverage adjustments, especially during period of financial stress. Predictive power is also shown by the firm's short-term debt. Finally, no significant effect is found in terms of size (measured by the number of employees) and activity level (*netincome*).

In Table 5 we report the LASSO estimates computed on the same firms' sample and on the same quantile levels, which essentially confirm the above findings.

To conclude this section, we illustrate the same example which includes a higher number of regressors, to highlight the main advantages of using LASSO. In particular, we consider the same response variables as in the example above (i.e., EBITDA and firm's leverage), but we now include two lagged variables for each of the regressors. This analysis could be extremely relevant when assessing systemic risk within a firm, as it allows to identify leading indicators of financial distress. In this case LASSO should help in increasing the model interpretability by eliminating irrelevant variables that are not associated with the response variables, and by reducing overfitting issues.

The results are reported in Table 6, for the two $\boldsymbol{\tau}$-levels of $(0.75, 0.25)$ (Panel A) and $(0.90, 0.10)$ (Panel B). The estimated coefficients for the coincident variables (in 2015) maintain the same sign and order of magnitude of the previous case. However, several variables exhibit also a significant leading property that could signal the risk of financial distress in advance. In particular, variables like financial expenses, net income and the ratio between cashflow over total assets show a significant coefficient up to a lag of order 2, for both EBITDA and leverage. This calls for these factors to be monitored on a systematic basis, as they could provide early warnings of impending financial distress.

7. Conclusions

This paper proposes a new likelihood-based method to jointly estimate marginal conditional quantiles of a multivariate response variable in a linear regression framework. We use a suitable reparameterization of the multivariate asymmetric Laplace distribution of Kotz et al. [36], whose mixture representation allows us to implement a new EM algorithm. Using this procedure, we show that the regression parameters can be easily estimated in closed form, hence avoiding direct maximization procedures. A penalized version of the algorithm is also proposed as an automatic data-driven procedure to perform variable selection. The good performance of the two methods is evaluated using a simulation exercise, where extreme quantiles are also considered as possible simulation scenarios. An empirical application to financial distress analysis on a sample of Italian firms is finally presented.

As quantile regression is widely used in the literature, several extensions of the results obtained in this paper can be analyzed in future research. Longitudinal settings and/or random effects models are of particular interest in this context, with the goal of characterizing the change in the response variables over time and accounting also for the dependence between serial observations on the same subject. Other forms of penalized algorithm might also be promising, as the simultaneous regularization of the parameters could perform better when a large set of (possibly correlated) explanatory variables is used in the application.

Table 6
LASSO estimates using lagged variables in the model.

	EBITDA	Leverage	EBITDA	Leverage
	Panel A: $\boldsymbol{\tau} = (0.75, 0.25)^\top$		Panel B: $\boldsymbol{\tau} = (0.90, 0.10)^\top$	
Constant	−2.078	−0.573	−3.862	−0.581
profit_{2015}	−0.027	−0.034	−0.015	−0.028
profit_{2014}	−0.049	–	−0.001	–
profit_{2013}	–	–	–	–
finexp_{2015}	17.934	0.291	18.647	0.011
finexp_{2014}	1.386	0.013	0.395	–
finexp_{2013}	0.205	0.025	0.463	–
earnings_{2015}	0.870	0375	1.010	0.200
earnings_{2014}	0.865	–	0.162	–
earnings_{2013}	–	–	0.665	–
employee_{2015}	0.005	–	−0.088	–
employee_{2014}	–	–	–	–
employee_{2013}	–	–	–	–
fixasset_{2015}	1.075	0.283	1.110	0.236
fixasset_{2014}	0.010	–	0.133	–
fixasset_{2013}	–	–	0.725	–
netincome_{2015}	−2.011	0.974	−1.340	−0.316
netincome_{2014}	−0.489	0.020	–	–
netincome_{2013}	−0.313	–	−0.288	–
current debt_{2015}	0.396	−0.657	0.940	−0.519
current debt_{2014}	0.544	−0.526	0.162	–
current debt_{2013}	–	–	0.094	–
cashflow_{2015}	−2.017	2.132	−2.029	3.081
cashflow_{2014}	−0.245	0.088	−0.429	1.075
cashflow_{2013}	−0.220	–	–	–
δ_j	0.820	0.168	0.595	0.196
ρ_{12}	−0.144		−0.160	
λ	1.270		2.245	
n	1576	1576	1576	1576

Appendix. Proofs of Propositions

Proof of Proposition 1. Using the result in (9), where we denote $\boldsymbol{\beta}_\tau \boldsymbol{X}_i = \boldsymbol{\mu}$ without loss of generality, for each component Y_j the following holds

$$Y_j = \mu_j + \delta_j \xi_j W + \delta_j \sqrt{W} \sigma_j Z_j, \qquad \forall j = 1, \ldots, p \tag{A.1}$$

where $Z_j \sim \mathcal{N}(0, 1)$ represents the jth component of Z. Then, defining $V = \delta_j W \sim \mathcal{E}(1/\delta)$, the representation in (A.1) can be also written as

$$Y_j = \mu_j + \xi_j V + \sigma_j \sqrt{\delta_j V} Z_j. \tag{A.2}$$

Following [36,37], and imposing (11), the result follows since (A.2) represents a univariate AL distribution with location, skewness and scale parameter equal to μ_j, τ_j and δ_j, respectively. □

Proof of Proposition 2. Remember that $\boldsymbol{D} = \text{diag}(\delta_1, \ldots, \delta_j)$ is an unknown matrix with $\delta_j > 0$ for all $j \in \{1, \ldots, p\}$. Moreover $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Lambda}} \boldsymbol{\Psi} \tilde{\boldsymbol{\Lambda}}$, where $\tilde{\boldsymbol{\Lambda}}$ is a known $p \times p$ diagonal matrix with (j, j)th element equal to $\tilde{\sigma}_j = \sqrt{2/\{\tau_j(1 - \tau_j)\}}$ and where Ψ is an unknown correlation matrix to be estimated.

Suppose we have two sets of parameters (say, $\boldsymbol{D}$ and $\boldsymbol{D}^*$, and $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}^*$) such that

$$\boldsymbol{D} \tilde{\boldsymbol{\Lambda}} \boldsymbol{\Psi} \tilde{\boldsymbol{\Lambda}} \boldsymbol{D} = \boldsymbol{D}^* \tilde{\boldsymbol{\Lambda}} \boldsymbol{\Psi}^* \tilde{\boldsymbol{\Lambda}} \boldsymbol{D}^*. \tag{A.3}$$

Then, identifiability analysis asks whether it is possible to identify both $\boldsymbol{D}$ and Ψ uniquely. That is, if (A.3) holds for some $\boldsymbol{D} \neq \boldsymbol{D}^*$ and some $\boldsymbol{\Psi} \neq \boldsymbol{\Psi}^*$, then the model is not identifiable, as two different sets of parameters give the same MAL distribution. Now, notice that (A.3) implies that

$$\text{diag}(\boldsymbol{D} \tilde{\boldsymbol{\Lambda}} \boldsymbol{\Psi} \tilde{\boldsymbol{\Lambda}} \boldsymbol{D}) = \text{diag}(\boldsymbol{D}^* \tilde{\boldsymbol{\Lambda}} \boldsymbol{\Psi}^* \tilde{\boldsymbol{\Lambda}} \boldsymbol{D}^*). \tag{A.4}$$

Since both $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}^*$ are correlation matrices then, for each $j \in \{1, \ldots, p\}$, we have $\Psi_{jj} = \Psi^*_{jj} = 1$. This implies that (A.4) can be rewritten, for each $j \in \{1, \ldots, p\}$, as

$$(\delta_j)^2 \sigma_j^2 = (\delta_j^*)^2 \sigma_j^2. \tag{A.5}$$

But then, since by assumption each δ_j (and δ_j^*) is greater than zero, the relationship in (A.5) is satisfied if and only if $\delta_j = \delta_j^*$, for all $j \in \{1, \ldots, p\}$, or equivalently $\boldsymbol{D} = \boldsymbol{D}^*$. Given this, the relationship in (A.3) reduces to

$$\boldsymbol{D}\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}\boldsymbol{D} = \boldsymbol{D}\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}^*\tilde{\boldsymbol{\Lambda}}\boldsymbol{D}.$$

Finally, since both $\boldsymbol{D}$ and $\tilde{\Lambda}$ are squared diagonal matrices (hence invertible), we have

$$(\boldsymbol{D}\tilde{\boldsymbol{\Lambda}})^{-1}\boldsymbol{D}\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}\boldsymbol{D}(\boldsymbol{D}\tilde{\boldsymbol{\Lambda}})^{-1} = (\boldsymbol{D}\tilde{\boldsymbol{\Lambda}})^{-1}\boldsymbol{D}\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}^*\tilde{\boldsymbol{\Lambda}}\boldsymbol{D}(\boldsymbol{D}\tilde{\boldsymbol{\Lambda}})^{-1},$$

which implies that $\boldsymbol{\Psi} = \boldsymbol{\Psi}^*$. Therefore, both $\boldsymbol{D}$ and $\boldsymbol{\Psi}$ are just identified, as (A.3) holds if and only if $\boldsymbol{D} = \boldsymbol{D}^*$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}^*$. □

Proof of Proposition 3. Notice that, under the constraints in (11), the representation in (9) implies that

$$\boldsymbol{Y}_i | W_i = w_i \sim \mathcal{N}_p(\boldsymbol{\beta}_\tau \boldsymbol{X}_i + \boldsymbol{D}\tilde{\boldsymbol{\xi}} w_i\,,\; w_i \boldsymbol{D}\tilde{\boldsymbol{\Sigma}}\boldsymbol{D}), \quad W_i \sim \mathcal{E}(1)$$

This implies that the joint density function of $\boldsymbol{Y}$ and W is

$$f_{Y,W}(\boldsymbol{y}, w) = \frac{\exp\left\{(\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{D}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}\right\}}{(2\pi)^{p/2}|\boldsymbol{S}|^{1/2}} \left[w^{-p/2} \exp\left\{ -\frac{1}{2}\frac{m}{w} - \frac{1}{2} w(d+2) \right\} \right]. \tag{A.6}$$

Then, using (A.6), the complete log-likelihood function (up to additive constant terms) can be written as

$$\begin{aligned}
\ell_c(\boldsymbol{\beta}_\tau, \boldsymbol{D}_{(\delta)}, \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} | \boldsymbol{Y}_i, W_i) = {} & \sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top \boldsymbol{D}_{(\delta)}^{-1} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})}^{-1} \tilde{\boldsymbol{\xi}} - \frac{n}{2} \ln |\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_{(\delta)}| && \text{(A.7)}\\
& - \frac{1}{2} \sum_{i=1}^{n} \frac{1}{W_i} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_\delta)^{-1} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i) && \text{(A.8)}\\
& - \frac{1}{2} \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \tilde{\boldsymbol{\xi}} \sum_{i=1}^{n} W_i, && \text{(A.9)}
\end{aligned}$$

where we use the notation $\boldsymbol{D}_{(\delta)}$ and $\tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})}$ to express the matrices $\boldsymbol{D}$ and $\tilde{\boldsymbol{\Sigma}}$ as a function of their parameters, $\boldsymbol{\delta}$ and $\boldsymbol{\Psi}$, respectively. In practice, W_i is a latent variable and, hence, not observable. For this reason, the E-step of the EM algorithm considers the conditional expectation of the complete log-likelihood function, given the observed data $\boldsymbol{Y}_i$ and the parameter estimates $(\hat{\boldsymbol{\beta}}_\tau, \boldsymbol{D}_{(\hat{\delta})}, \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})})$. That is,

$$\begin{aligned}
& \mathrm{E}\{\ell_c(\boldsymbol{\beta}_\tau, \boldsymbol{D}_{(\delta)}, \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} | \boldsymbol{Y}_i, \hat{\boldsymbol{\beta}}_\tau, \boldsymbol{D}_{(\hat{\delta})}, \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})})\}\\
& \quad = \sum_{i=1}^{n} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top \boldsymbol{D}_\delta^{-1} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})}^{-1} \tilde{\boldsymbol{\xi}} - \frac{n}{2} \ln |\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_{(\delta)}| && \text{(A.10)}\\
& \quad - \frac{1}{2} \sum_{i=1}^{n} \mathrm{E}(W_i^{-1}) (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_\delta)^{-1} (\boldsymbol{Y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i) && \text{(A.11)}\\
& \quad - \frac{1}{2} \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \tilde{\boldsymbol{\xi}} \sum_{i=1}^{n} \mathrm{E}(W_i), && \text{(A.12)}
\end{aligned}$$

where $\mathrm{E}(W_i|\cdot)$ and $\mathrm{E}(W_i^{-1}|\cdot)$ denote the conditional expectations of W_i and W_i^{-1} conditional on $(\cdot)$, respectively.

Let $\tilde{m}_i = (\boldsymbol{y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\delta)} \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \boldsymbol{D}_{(\delta)})^{-1} (\boldsymbol{y}_i - \boldsymbol{\beta}_\tau \boldsymbol{X}_i)$, for every $i \in \{1, \ldots, n\}$, and define $\tilde{d} = \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})} \tilde{\boldsymbol{\xi}}$. Using the joint distribution of $\boldsymbol{Y}$ and W derived in (A.6) and the pdf of $\boldsymbol{Y}$ given in (8), we have

$$f_{W|Y}(W_i | \boldsymbol{Y}_i = \boldsymbol{y}_i) = \frac{f_{W,Y}(w_i, \boldsymbol{y}_i)}{f_Y(\boldsymbol{y}_i)} = \frac{w_i^{-p/2} \{(2+\tilde{d})/\tilde{m}_i\}^{\nu/2} \exp\{-\tilde{m}_i/(2w_i) - w_i(2+\tilde{d})/2\}}{2K_\nu\left\{\sqrt{(2+\tilde{d})\tilde{m}_i}\right\}}$$

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

$$f_{W|Y}(W_i | \boldsymbol{Y}_i = \boldsymbol{y}_i) \sim \mathrm{GIG}(\nu, \tilde{d}+2, \tilde{m}_i).$$

It follows that

$$\mathrm{E}\{W_i | \boldsymbol{Y}_i, \hat{\boldsymbol{\beta}}_\tau, \boldsymbol{D}_{(\hat{\delta})}, \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})}\} = \left(\frac{\hat{\tilde{m}}_i}{2+\hat{\tilde{d}}}\right)^{1/2} \frac{K_{\nu+1}\left\{\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_i}\right\}}{K_\nu\left\{\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_i}\right\}} \tag{A.13}$$

and

$$\mathrm{E}\{W_i^{-1}|\boldsymbol{Y}_i, \hat{\boldsymbol{\beta}}_\tau, \boldsymbol{D}_{(\hat{\delta})}, \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})}\} = \left(\frac{2+\hat{\tilde{d}}}{\hat{\tilde{m}}_i}\right)^{1/2} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_i}\right)}{K_\nu\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_i}\right)} - \frac{2\nu}{\hat{\tilde{m}}_i} \tag{A.14}$$

where

$$\hat{\tilde{m}}_i = (\boldsymbol{y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i)^\top (\boldsymbol{D}_{(\hat{\delta})} \tilde{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\Psi}}} \boldsymbol{D}_{(\hat{\delta})})^{-1} (\boldsymbol{y}_i - \hat{\boldsymbol{\beta}}_\tau \boldsymbol{X}_i), \quad \hat{\tilde{d}} = \tilde{\boldsymbol{\xi}}^\top \tilde{\boldsymbol{\Sigma}}_{(\hat{\boldsymbol{\Psi}})} \tilde{\boldsymbol{\xi}}.$$

Denoting the two conditional expectations in (A.13) and (A.14) by u_i and z_i respectively, concludes the proof. □

Proof of Proposition 4. Imposing the first order conditions on (13)–(15) with respect to $\boldsymbol{\beta}_\tau$ and $\tilde{\boldsymbol{\Sigma}}_{(\boldsymbol{\Psi})}$ gives the parameter estimates in (16), (17) and (18). □

References

[1] T. Adrian, M.K. Brunnermeier, CoVaR, Amer. Econ. Rev. 106 (2016) 1705–1741.
[2] M. Alfo, M.F. Marino, M.G. Ranalli, N. Salvati, N. Tzavidis, M-quantile regression for multivariate longitudinal data: Analysis of the millennium cohort study data, 2016, ArXiv preprint arXiv:161208114.
[3] O. Arslan, An alternative multivariate skew Laplace distribution: Properties and estimation, Stat. Pap. 51 (2010) 865–887.
[4] G.W. Bassett, H.L. Chen, Portfolio style: Return-based attribution using quantile regression, in: B. Fitzenberger, R. Koenker, J.A.F. Machado (Eds.), Economic Applications of Quantile Regression, Physica-Verlag, Heidelberg, 2002, pp. 293–305.
[5] R. Bastos, J. Pindado, Trade credit during a financial crisis: A panel data analysis, J. Bus. Res. 66 (2013) 614–620.
[6] M. Bernardi, G. Gayraud, L. Petrella, Bayesian tail risk interdependence using quantile regression, Bayesian Anal. 10 (2015) 553–603.
[7] M. Bisogno, R. De Luca, Financial distress and earnings manipulation: Evidence from italian SMEs, J. Acc. Finance 4 (2015) 42–51.
[8] P. Boček, M. Šiman, On weighted and locally polynomial directional quantile regression, Comput. Stat. 32 (2017) 929–946.
[9] D. Böhning, E. Dietz, R. Schaub, P. Schlattmann, B.G. Lindsay, The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family, Ann. Inst. Statist. Math. 46 (1994) 373–388.
[10] J.Y. Campbell, J. Hilscher, J. Szilagyi, In search of distress risk, J. Finance 63 (2008) 2899–2939.
[11] B. Chakraborty, On multivariate quantile regression, J. Statist. Plann. Inference 110 (2003) 109–132.
[12] E. Charalambakis, D. Psychoyios, What do we know about capital structure? Revisiting the impact of debt ratios on some firm-specific factors, Appl. Financ. Econom. 22 (2012) 1727–1742.
[13] H. Cho, S. Kim, M.-O. Kim, Multiple quantile regression analysis of longitudinal data: Heteroscedasticity and efficient estimation, J. Multivariate Anal. 155 (2017) 334–343.
[14] T.J. Cole, P.J. Green, Smoothing reference centile curves: The LMS method and penalized likelihood, Stat. Med. 11 (1992) 1305–1319.
[15] F.B. Covas, B. Rump, E. Zakrajšek, Stress-testing us bank holding companies: A dynamic panel quantile regression approach, Int. J. Forecast. 30 (2014) 691–713.
[16] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B Stat. Methodol. (1977) 1–38.
[17] P. Embrechts, R. Frey, A.J. McNeil, Quantitative Risk Management: Concepts, Techniques and Tools, Princeton University Press, Princeton, NJ, 2005.
[18] R.F. Engle, S. Manganelli, Caviar: Conditional autoregressive value at risk by regression quantiles, J. Bus. Econom. Statist. 22 (2004) 367–381.
[19] B.C. Franczak, R.P. Browne, P.D. McNicholas, Mixtures of shifted asymmetric Laplace distributions, IEEE Trans. Pattern Anal. Mach. Intel. 36 (2014) 1149–1157.
[20] J. Friedman, T. Hastie, R.J. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (2010) 1.
[21] M. Geraci, M. Bottai, Quantile regression for longitudinal data using the asymmetric Laplace distribution, Biostatistics 8 (2006) 140–154.
[22] P.J. Green, On use of the EM for penalized likelihood estimation, J. R. Stat. Soc. Ser. B Stat. Methodol. 52 (1990) 443–452.
[23] M. Hallin, D. Paindaveine, M. Šiman, Y. Wei, R. Serfling, Y. Zuo, L. Kong, I. Mizera, Multivariate quantiles and multiple-output regression quantiles: from l1 optimization to half space depth, Ann. Statist. 38 (2010) 635–703.
[24] W. Hendricks, R. Koenker, Hierarchical spline models for conditional quantiles and the demand for electricity, J. Amer. Statist. Assoc. 87 (1992) 58–68.
[25] A.E. Hoerl, R.W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics 12 (1970) 55–67.
[26] S.J. Jun, J. Pinkse, Efficient semiparametric seemingly unrelated quantile regression estimation, Econom. Theory 25 (2009) 1392–1414.
[27] R. Koenker, Quantile regression for longitudinal data, J. Multivariate Anal. 91 (2004) 74–89.
[28] R. Koenker, Quantile Regression, Cambridge University Press, Cambridge, 2005.
[29] R. Koenker, Quantile regression: 40 years on, Ann. Rev. Econom. 9 (2017) 155–176.
[30] R. Koenker, G. Bassett Jr, Regression quantiles, Econometrica (1978) 33–50.
[31] R. Koenker, O. Geling, Reappraising medfly longevity: A quantile regression survival analysis, J. Amer. Statist. Assoc. 96 (2001) 458–468.
[32] L. Kong, I. Mizera, Quantile tomography: Using quantiles with multivariate data, Statist. Sinica 22 (2012) 1589–1610.
[33] L. Kong, H. Shu, G. Heo, Q.C. He, Estimation for bivariate quantile varying coefficient model, arXiv preprint arXiv:151102552, 2015.
[34] A. Kottas, A.E. Gelf, Bayesian semiparametric median regression modeling, J. Amer. Statist. Assoc. 96 (2001) 1458–1468.
[35] A. Kottas, M. Krnjajić, Bayesian semiparametric modelling in quantile regression, Scand. J. Stat. 36 (2009) 297–319.
[36] S. Kotz, T. Kozubowski, K. Podgorski, The Laplace Distribution and Generalizations: A Revisit with Applications To Communications, Economics, Engineering, and Finance, Birkhäuser, Basel, 2001.
[37] H. Kozumi, G. Kobayashi, Gibbs sampling methods for Bayesian quantile regression, J. Stat. Comput. Simul. 81 (2011) 1565–1578.
[38] M.F. Marino, A. Farcomeni, Linear quantile regression models for longitudinal experiments: An overview, Metron 73 (2015) 229–247.
[39] M.F. Marino, N. Tzavidis, M. Alfò, Mixed hidden Markov quantile regression models for longitudinal data with possibly incomplete sequences, Stat. Methods Med. Res. 27 (2018) 2231–2246.
[40] D. Paindaveine, M. Šiman, Computing multiple-output regression quantile regions from projection quantiles, Comput. Stat. 27 (2012) 29–49.

[41] G.R. Pandey, V.-T.-V. Nguyen, A comparative study of regression based methods in regional flood frequency analysis, J. Hydrol. 225 (1999) 92–101.
[42] L. Petrella, A.G. Laporta, L. Merlo, Cross-country assessment of systemic risk in the European stock market: Evidence from a covar analysis, Soc. Indic. Res. (2018) 1–18.
[43] J. Pindado, L. Rodrigues, C. de la Torre, Estimating financial distress likelihood, J. Bus. Res. 61 (2008) 995–1003.
[44] P. Royston, D.G. Altman, Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling, Appl. Stat. 43 (1994) 429–467.
[45] B. Sánchez, H. Lachos, V. Labra, Likelihood based inference for quantile regression using the asymmetric Laplace distribution, Working Paper, Universidade Estadual de Campinas, Brazil, 2013.
[46] P. Stolfi, M. Bernardi, L. Petrella, The sparse method of simulated quantiles: An application to portfolio optimization, Stat. Neerl. 72 (2018) 375–398.
[47] R.J. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996) 267–288.
[48] C. Tortora, P.D. McNicholas, R.P. Browne, A mixture of generalized hyperbolic factor analyzers, Adv. Data Anal. Classif. 10 (2016) 423–440.
[49] E. Waldmann, T. Kneib, Bayesian bivariate quantile regression, Stat. Model. 15 (2015) 326–344.
[50] E. Wipplinger, Philippe. Jorion, Value at risk – the new benchmark for managing financial risk, Financial Mark. Portfolio Manag. 21 (2007) 397.
[51] C.M. Wong, L.L.O. Ting, A quantile regression approach to the multiple period value at risk estimation, J. Econom. Manag. 12 (2016) 1–35.
[52] Z. Xiao, H. Guo, M.S. Lam, Quantile Regression and Value At Risk, in: Handbook of Financial Econometrics and Statistics, Springer, 2015, pp. 1143–1167.
[53] K. Yu, R.A. Moyeed, Bayesian quantile regression, Statist. Probab. Lett. 54 (2001) 437–447.
[54] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2005) 301–320.