Quantile regression for longitudinal data via the multivariate generalized hyperbolic distribution

Alvaro J. Flórez¹, Ingrid Van Keilegom², Geert Molenberghs³ and Anneleen Verhasselt⁴

¹DSI, I-BioStat, Universiteit Hasselt, Belgium and School of Statistics, Universidad del Valle, Colombia

²ORSTAT, KU Leuven, Belgium

³DSI, I-BioStat, Universiteit Hasselt, Belgium and I-BioStat, KU Leuven, Belgium

⁴DSI, I-BioStat, Universiteit Hasselt, Belgium

Abstract: While extensive research has been devoted to univariate quantile regression, this is considerably less the case for the multivariate (longitudinal) version, even though there are many potential applications, such as the joint examination of growth curves for two or more growth characteristics, such as body weight and length in infants. Quantile functions are easier to interpret for a population of curves than mean functions. While the connection between multivariate quantiles and the multivariate asymmetric Laplace distribution is known, it is less well known that its use for maximum likelihood estimation poses mathematical as well as computational challenges. Therefore, we study a broader family of multivariate generalized hyperbolic distributions, of which the multivariate asymmetric Laplace distribution is a limiting case. We offer an asymptotic treatment. Simulations and a data example supplement the modelling and theoretical considerations.

Key words: asymptotics, longitudinal data, maximum likelihood, pseudo-likelihood, quantile regression

Received 7 September 2020; revised 10 February 2021; accepted 19 March 2021

1 Introduction

Since the pioneering work of Koenker and Bassett (1978), quantile regression has taken a prominent role in both theoretical and applied statistics, as alternative to classical mean regression. While mean regression models solely grasp the central behaviour of the data, quantile regression allows to examine the effect of a set of covariates on different quantiles of a response variable. Quantile regression has also been shown useful when the distribution of the response is skewed rather than symmetric, when the data contain outliers, or when flexibility to the error distribution is important. For non-crossing curves, the mean curve is typically different in shape from all subject curves (Molenberghs and Verbeke, 2005), often causing confusion,

Address for correspondence: Ingrid Van Keilegom, ORSTAT, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium.

E-mail: ingrid.vankeilegom@kuleuven.be

whereas this is not the case for median and other quantile curves. Furthermore, growth curves or other profiles are often examined together, such as body length and body height in growing infants, or blood pressure and other vital parameters in patients. We refer to Koenker (2005) for a book length treatment of the methodology of quantile regression.

Whereas classical means are obtained by minimizing the quadratic loss function, the quantile m_τ of order τ ($0 < \tau < 1$) of a variable Y is defined as

$$m_\tau = \operatorname{argmin}_m E[\rho_\tau(Y - m)],$$

where $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is the so-called check function, and $I(\cdot)$ is the usual indicator function. When i.i.d. data $Y_1, \dots, Y_n$ are available, m_τ is estimated by its empirical counterpart $\hat{m}_\tau = \operatorname{argmin}_m \sum_{i=1}^n \rho_\tau(Y_i - m)$. In the special case where $\tau = 1/2$, $\rho_\tau(u) = 0.5|u|$, showing that median regression is based on the L_1 -norm as opposed to the L_2 -norm in the case of mean regression. An enormous literature on quantile regression exists in a wide variety of areas of statistics. Moreover, extensions of the ‘basic’ check function approach to complex data have been proposed, see, for example, Wang and Wang (2009), De Backer et al. (2019) and the references therein for the extension to censored data, Verhasselt et al. (2018) for the extension to missing data, Wang et al. (2018) for the case of variable selection, etc.

We focus on multivariate and/or longitudinal data, in this article. The methodology is illustrated on diabetes data in Section 6. The dataset consists of information of 2495 patients with type 2 diabetes mellitus that were enrolled in the Leuven diabetes project. It contains several variables that were measured at the beginning of the programme and one year later. Among these, LDL-cholesterol (low-density lipoprotein cholesterol, mg/dl) and HbA1c (glycosylated haemoglobin, %) are some of the most important measures to evaluate how well diabetes is controlled. A reduction of these endpoints indicates an improvement in the physical condition of the patients due to the treatment. Therefore, we are interested in studying the joint evolution of these outcomes in time, and to explore the effect of some potentially disease-related characteristics of the patients. The joint modelling allows to address the association between both responses in the estimation process. The data are, therefore, multivariate and longitudinal in nature.

In this article, we consider a quantile regression model with multivariate and/or longitudinal data, and we propose a novel way to take the longitudinal structure of the data into account in the estimation of the model. To do this, first recall the well-known equivalence in the univariate case between the minimization of the check function and the maximization of the likelihood based on an asymmetric Laplace distribution (see Kotz et al., 2001). Indeed, in the univariate case, using the parametrization of Poiraud-Casanova and Thomas-Agnan (2000), if the density of Y is the asymmetric Laplace density, given by

$$f_\tau(y|m) = \begin{cases} \tau(1 - \tau) \exp[-(1 - \tau)|y - m|] & \text{for } y < m \\ \tau(1 - \tau) \exp[-\tau|y - m|] & \text{for } y \geq m, \end{cases} \quad (1.1)$$

then it is easily seen that $m_\tau = \operatorname{argmin}_m E[\rho_\tau(Y - m)] = \operatorname{argmax}_m E[f_\tau(Y|m)]$. Motivated by this equivalence, it seems natural at first sight to estimate the quantiles in the multivariate case by maximizing the likelihood based on a multivariate extension of the asymmetric Laplace distribution. Note that the multivariate distribution allows to take the dependence between the longitudinal observations into account, whereas the classical univariate asymmetric Laplace distribution ignores this dependence. However, as we will see later, the multivariate asymmetric Laplace (MAL) distribution suffers from serious problems, the most important one being the fact that it leads to a likelihood that can have a multiplicity of spikes (or asymptotes) that lead to severe problems. For that reason, we will propose to work with a generalization of this MAL distribution, called the multivariate generalized hyperbolic (MGH) distribution, that does not suffer from these problems.

The article is organized as follows. In the next section, we give an overview of the existing literature on quantile regression with longitudinal data. In Section 3, we define our multivariate quantile regression model, and propose an estimator of the model parameters via the maximization of the MGH likelihood. We also show the asymptotic normality of the proposed estimator. Section 4 describes some of the computational issues related to the proposed estimator, whereas a detailed simulation study and the analysis of diabetes data are studied in Sections 5 and 6, respectively. Some conclusions and ideas for future research are given in Section 7. Finally, additional simulation results and technical details are exhibited in the Supplementary Materials available online.

2 Literature overview

The existing literature on quantile regression for multivariate and/or longitudinal data can be roughly classified into two groups of articles, namely those articles that only posit a model for the quantile of interest, and those that impose a model on the entire distribution. Only assuming a model for the quantile of interest is of course a less restrictive assumption, but the existing approaches suffer from one or several possible drawbacks. For example, in some articles the independence working model is assumed, in which case conventional quantile regression techniques can be used. They lead to consistent, yet inefficient estimators, due to the incorrect specification of the variance structure. So, while an independence working assumption does not in itself imply that zero correlation is assumed to be correct, the intrinsic corrections that take place in such a semi-parametric approach and that ensure consistency, come at the price of considerable efficiency loss. We refer the reader to Lipsitz et al. (1997) and Chen et al. (2004), among others. These authors follow a semi-parametric approach whereby Lipsitz et al. (1997) also allows for incomplete data. Parente and Santos Silva (2016) examine the performance of the univariate model even when data are clustered. Another option is to add subject-specific fixed effects to the usual quantile regression model. This approach has the drawback that it potentially leads to a large number of parameters in the model. This is especially the case when the

number of subjects is large compared to the number of repeated measures, in which case the estimation of these fixed effects is inconsistent. To bypass these problems, Koenker (2004) and Lamarche (2010) proposed penalized estimators, where the penalty is necessary to control the large number of fixed effects. An overview of this and other approaches in the econometrics literature, mainly based on fixed effects, can be found in Galvao and Kato (2017).

The second category of articles on quantile regression for longitudinal data contains articles that impose a model not only for the quantile of interest, but for the entire distribution. The proposed approaches are mostly likelihood based. One possibility is to replace the subject-specific fixed effects mentioned above by random effects. In that case, one essentially binds together univariate Laplace distributions by using normal or other types of random effects; see, for example, Geraci and Bottai (2007), Geraci and Bottai (2014), Liu and Bottai (2009), Yuan and Yin (2010), Lee and Neocleous (2010), Farcomeni (2012) and references therein. Note, however, that random effects are formulated at the subject level. In case one is interested in marginal functions as in our quantile regression case, often cumbersome integration over such random effects is needed. If these random-effects enter the model in a non-linear fashion, integration over them is not straightforward and, importantly, often does not provide closed forms. Moreover, the so obtained functions may be cumbersome to interpret in the light of the population under study, as it may not correspond to an observable quantile function.

Instead of specifying two distributions (the univariate distribution and the random effects distribution), it is therefore a valid and very sensible alternative to assume a marginal parametric model for the multivariate distribution of the response vector for a given subject, in which case the specification of the dependence between the observations is effectuated via the multivariate distribution. This facilitates a more direct and easier to interpret specification of the model when compared with the random-effects model. Waldmann and Kneib (2015) and Petrella and Raponi (2019) contributed to this research line. In the former article, the authors considered a Bayesian bivariate quantile regression model in which the asymmetric Laplace distribution is an auxiliary error distribution, whereas in the latter article the asymmetric Laplace distribution was replaced by a multivariate extension of it, proposed and studied by Kozubowski and Podgórski (2000) and Kotz et al. (2001). Petrella and Raponi (2019) showed via simulations that, in spite of the peaks and non-differentiability problems that are inherent to the latter distribution, it is possible to estimate the model correctly. They use an intuitive approach to this problem, by maximizing the likelihood as if it is a classical, well behaved likelihood. However, as Kotz et al. (2001) point out, there are some issues with the multivariate asymmetric Laplace (MAL) distribution, that require special treatment and attention. In particular, a peculiarity of this distribution is that it has an asymptote at the origin, the implications of which have so far not been carried forward to multivariate data analysis. One of the special features of the likelihood surface is that in certain cases it contains a ‘minefield’ of spikes, leading to problems in the maximization of the likelihood and even more in the calculation of the gradient and Hessian matrix. We solve these issues using the link between the MAL distribution and the multivariate

generalized hyperbolic (MGH) distribution. Furthermore, we show that the estimator is asymptotically normal and that standard errors can be computed via this minor modification in the likelihood function.

3 Model and methodology

3.1 Multivariate generalized hyperbolic distribution

Suppose that $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in})'$ is an n -dimensional response vector for subject $i = 1, \dots, N$. Throughout, we consider the setup where N is large (and goes to infinity in the asymptotics), whereas n is fixed. While the methodology and theory would also hold when n were allowed to vary from subject to subject, at the condition that it is bounded from above, for simplicity we will assume n to be fixed. Consider the multivariate regression model

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad (3.1)$$

where $\mathbf{X}_i$ is a $(n \times p)$ -design matrix of covariates, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a vector of regression coefficients, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in})'$ is a vector of error terms.

This formulation is generic and encompasses several special cases. To see this, write $n = T \cdot q$, then q can be taken to signify a number of repeated-measures sequences, all of length T . The univariate longitudinal setting follows from setting $q = 1$, whereas the (cross-sectional) multivariate setting follows from specifying $T = 1$. The general setting obviously ensues when $T > 1$ and $q > 1$. The modeller has considerable flexibility in specifying the effect of covariates. For example, a given covariate can affect one, several, or all of the multivariate sequences. Its effect across time can be constant or time-varying. Moreover, some covariates can be measured once per subject (e.g., baseline age), while others can themselves be measured repeatedly over time (e.g., weekly summaries of dietary intake). Examples are given in Section A of the Supplementary Materials. In the diabetes data (introduced in Section 1), $T = 2$ refers to the occasions at which we have measurements and $q = 2$ to the outcomes of interest LDL-cholesterol and HbA1c. The covariates gender, age, diabetes duration and insulin are baseline covariates, while the BMI is measured over time.

We are interested in conditional quantiles of order τ for each Y_{ij} , $j = 1, \dots, n$, taking the possible correlation between $Y_{i1}, \dots, Y_{in}$ into account. Since in the univariate case ($n = 1$) quantiles can be obtained by maximizing the asymmetric Laplace likelihood (see Kotz et al., 2001), a natural extension to the multivariate case exists in maximizing a multivariate version of this asymmetric Laplace likelihood. This is exactly what Petrella and Raponi (2019) did. More precisely, they proposed to work with the so-called multivariate asymmetric Laplace (MAL) distribution, which has been proposed and studied by Kozubowski and Podgórski (2000) and Kotz et al. (2001). In the Supplementary Material (Section G) we give the precise definition of this MAL distribution.

An important property of the MAL distribution is that its marginals are univariate asymmetric Laplace distributions (see Kotz et al., 2001), and hence this distribution is a natural extension of the univariate setting, while at the same time taking the dependence structure into account. Despite this important property, there are also serious concerns and drawbacks related to this model, which we explain in Section G of the Supplementary Material. They are all related to the fact that for $n \geq 2$ the MAL density is unbounded, and in particular it diverges to infinity when $\mathbf{y}$ tends to $\mathbf{X}_i\boldsymbol{\beta}$. The discussion of these drawbacks shows that there are some peculiar differences between the shape of the asymmetric Laplace density in the univariate case ($n = 1$) and the multivariate case ($n \geq 2$), and that there are severe problems with this density when it comes to $n \geq 2$. An alternative density that does not suffer from the above problems is therefore required. We will slightly modify the MAL density, so that it becomes a well-behaved, smooth and bounded density. To do so, we employ the fact that the MAL density is a special case of the multivariate generalized hyperbolic (MGH) density (Barndorff-Nielsen and Kendall, 1977; Barndorff-Nielsen and Blaesild, 1981), defined as follows. Suppose that

$$\mathbf{Y}_i \sim \text{MGH}_n(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Delta}\boldsymbol{\xi}, \boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta}, \epsilon),$$

with density

$$f_\epsilon(\mathbf{y}|\mathbf{X}_i, \boldsymbol{\theta}) = \frac{\sqrt{2} \exp[(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta})' \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\xi}]}{\sqrt{\epsilon}(2\pi)^{n/2} |\boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta}|^{1/2} K_1(\sqrt{2\epsilon})} \left(\frac{\epsilon + m_i}{2 + d} \right)^{\nu/2} K_\nu \left[\sqrt{(2 + d)(\epsilon + m_i)} \right], \quad (3.2)$$

for some $\epsilon > 0$, where $\mathbf{y} = (y_1, \dots, y_n)'$, $(\mathbf{X}_i\boldsymbol{\beta})_j = Q_{Y_j}(\tau|\mathbf{X}_i)$ is the τ th conditional quantile of Y_j (the location parameter vector, for $j = 1, \dots, n$), $\boldsymbol{\Delta}\boldsymbol{\xi}$ is the scale (or skewness) parameter vector and $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}$ a positive definite matrix. Here, we use the following notation: $\boldsymbol{\Delta} = \text{diag}(\delta_1, \dots, \delta_n)$, $\delta_j > 0$ (for $j = 1, \dots, n$), $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)'$, $\xi_j = \frac{1-2\tau}{\tau(1-\tau)}$ for $j = 1, \dots, n$, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\lambda_j^2 = \frac{2}{\tau(1-\tau)}$ for $j = 1, \dots, n$, and $\boldsymbol{\Psi}$ is a correlation matrix. Further $m_i = (\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta})'(\boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta})^{-1}(\mathbf{y} - \mathbf{X}_i\boldsymbol{\beta})$, $d = \boldsymbol{\xi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}$, K_ν is the modified Bessel function of the third kind with index parameter $\nu = (2 - n)/2$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{\Psi})$ is the vector of unknown parameters.

Note that this density is bounded for $\mathbf{y}$ in a neighbourhood of $\mathbf{X}_i\boldsymbol{\beta}$, since $K_\nu[\sqrt{(2 + d)(\epsilon + m_i)}] \leq K_\nu(\epsilon) < \infty$ provided $\epsilon > 0$. Using the fact that $K_\nu(x) \sim \Gamma(\nu)2^{\nu-1}x^{-\nu}$ ($\nu > 0$) for $x \rightarrow 0$, it is easily seen that for $\epsilon = 0$ the MGH density reduces to the MAL density (Kotz et al., 2001). Furthermore, the MGH distribution is closed under marginalization (McNeil et al., 2005). This shows that for small values of ϵ the marginals of the MGH distribution are close to the marginals of the univariate Laplace distribution, and hence this distribution is a natural extension of the univariate setting, since for $n = 1$ and $\epsilon \rightarrow 0$ we get the univariate quantiles back while at the same time taking the dependence structure into account and working with a multivariate distribution that is bounded and well-behaved.

If $\mathbf{Y}_i \sim \text{MGH}_n(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Delta}\boldsymbol{\xi}, \boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta}, \epsilon)$, then $E(\mathbf{Y}_i) = \mathbf{X}_i\boldsymbol{\beta} + a_1\boldsymbol{\Delta}\boldsymbol{\xi}$ and

$$V(\mathbf{Y}_i) = \boldsymbol{\Delta}(a_1\boldsymbol{\Sigma} + a_2\boldsymbol{\xi}\boldsymbol{\xi}')\boldsymbol{\Delta}', \quad (3.3)$$

where $a_1 = \sqrt{\frac{\epsilon}{2}}R_1(\sqrt{2\epsilon})$, $a_2 = \sqrt{\frac{\epsilon}{2}}R_2(\sqrt{2\epsilon}) - a_1^2$, and $R_i(x) = \frac{K_{1+i}(x)}{K_1(x)}$. As $\epsilon \rightarrow 0$, we get $a_1 = a_2 = 1$. Note in (3.3) that the correlation structure of $\mathbf{Y}$ is partially determined by τ through $\boldsymbol{\xi}$ and $\boldsymbol{\Lambda}$. For $\tau = 0.5$, we have that $\boldsymbol{\xi} = \mathbf{0}$ and the correlation is completely determined by $\boldsymbol{\Sigma}$. On the other hand, as τ departs from 0.5, the correlation structure is more restricted to positive values.

We are interested in doing inference for $\boldsymbol{\theta}$ and in particular for $\boldsymbol{\beta}$, since the latter is crucial for the estimation of the quantile. The estimation of $\boldsymbol{\theta}$ is done by maximizing the log-likelihood:

$$\begin{aligned} \ell(\boldsymbol{\theta}) \sim \sum_{i=1}^N \left\{ -\frac{1}{2} \log |\boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta}| + (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})'\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\xi} + \frac{\nu}{2} [\log(\epsilon + M_i) - \log(2 + d)] \right. \\ \left. + \log K_\nu \left[\sqrt{(2 + d)(\epsilon + M_i)} \right] \right\}, \end{aligned} \quad (3.4)$$

where $M_i = (\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\boldsymbol{\Delta}\boldsymbol{\Sigma}\boldsymbol{\Delta})^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta})$. Let

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Delta}}, \hat{\boldsymbol{\Psi}}) = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}) \quad (3.5)$$

be the maximum likelihood estimator (MLE), where Θ is a compact subset of $\mathbf{R}^k$ ($k = \dim(\boldsymbol{\theta})$).

3.2 Asymptotic properties of the maximum likelihood estimator

In this subsection, we will develop the asymptotic theory for the MLE $\hat{\boldsymbol{\beta}}$ defined in (3.5). A crucial difference exists between the univariate asymmetric Laplace density and the MGH density, which has an impact on the asymptotic properties. In the univariate case, it is well known that $f(y|\mathbf{X}_i; \boldsymbol{\theta})$ is non-smooth in y for $y = \mathbf{X}_i\boldsymbol{\beta}$. For that reason, classical results on the consistency and asymptotic normality of maximum likelihood estimators (as in, e.g., White, 1982) cannot be applied, and more elaborate techniques have therefore been used to show that the quantile estimator in the univariate case is asymptotically normal (see, e.g., Koenker, 2005). However, surprisingly, the MGH density $f_\epsilon(\mathbf{y}|\mathbf{X}_i; \boldsymbol{\theta})$ is smooth in all arguments, and we will show below that the conditions that are required for applying the classical theory for maximum likelihood estimators developed by White (1982), are all satisfied. As a consequence, the asymptotic theory for the MGH density is substantially easier than for the univariate asymmetric Laplace density. Section H of the Supplementary Material contains the regularity conditions (C1)–(C4) mentioned below as well as the proof of the theorem.

Theorem 1. *Suppose (C1)-(C4). Then,*

$$N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_*) \xrightarrow{\mathcal{D}} N(0, \mathbf{V}_{\boldsymbol{\beta}\boldsymbol{\beta}}), \quad \text{as } N \rightarrow \infty,$$

where $\mathbf{V}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ is the upper left submatrix of dimension $p \times p$ of the matrix

$$\mathbf{V} = \{\mathbf{A}(\boldsymbol{\theta}_*)\}^{-1} \mathbf{B}(\boldsymbol{\theta}_*) \{\mathbf{A}(\boldsymbol{\theta}_*)\}^{-1},$$

and where $\mathbf{A}(\boldsymbol{\theta}_*)$ and $\mathbf{B}(\boldsymbol{\theta}_*)$ are defined in assumption (C4)(b)–(c) and $\boldsymbol{\theta}_*$ is defined in assumption (C2).

Note that if the model is correctly specified, the estimator $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased, that is, $\boldsymbol{\beta}_* = \boldsymbol{\beta}$ in that case.

4 Computational aspects

The model and algorithm described in the previous section can be summarized as follows. First, in order to overcome the problems created by the MAL likelihood we propose to work with the MGH density defined in (3.2) above. Unlike the MAL density, the MGH density has no spikes and its log-likelihood given in (3.4) is therefore smooth in all its parameters for positive values of ϵ . We first specify a value for τ and ϵ , then the likelihood is maximized with respect to $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$ and $\boldsymbol{\Psi}$. The maximization of this log-likelihood is standard. We proceed with explaining the details and technicalities that are important when maximizing the likelihood in practice.

Since the maximization of the log-likelihood cannot be achieved analytically, the maximum likelihood estimator is obtained iteratively. To do so, the estimator is implemented in the `optim` function in R utilizing the Broyden–Fletcher–Goldfarb–Shanno (Fletcher, 2000) algorithm and using (central) finite-differences for computing gradients and Hessian matrices.

To guarantee that the solution of $\boldsymbol{\Psi}$ is positive definite (PD) in each step of the maximization process, we reparameterize $\boldsymbol{\Psi}$ through its hyperspherical coordinates (Rousseeuw and Molenberghs, 1993; Pourahmadi and Wang, 2015). That is, a PD correlation matrix $\mathbf{R} = (r_{ij})_{i,j=1}^n$ can be factorized as $\mathbf{R} = \mathbf{U}\mathbf{U}'$ where $\mathbf{U} = (u_{ij})_{i,j=1}^n$ is a lower triangular matrix with $u_{11} = 1$, $u_{i1} = \cos \phi_{i1}$, for $i = 2, \dots, n$, and

$$u_{ij} = \begin{cases} \prod_{k=1}^{j-1} \sin \phi_{ik} & \text{for } i = j, \\ \cos \phi_{ij} \prod_{k=1}^{j-1} \sin \phi_{ik} & \text{for } i = j + 1, \dots, n, \end{cases}$$

for $j = 2, \dots, n$, and where ϕ_{ij} , $i > j$, are some angles. The matrix $\mathbf{U}$ is unique if the angles are restricted to the range $(0, \pi)$. Furthermore, the transformation from $\mathbf{R}$ to $\boldsymbol{\Phi} = (\phi_{ij})_{i,j=1}^n$ is one-to-one, where $\phi_{ij} = 0$ for $i \leq j$. Evidently, this transformation is only needed when $n > 2$. If $n = 2$, the only constraint is $-1 < \psi_{12} < 1$.

The initial values are obtained using a method-of-moments estimator (MME). First, we estimate the diagonal elements of $\mathbf{\Delta}^{(0)}$ as:

$$\delta_j^0 = \sqrt{\frac{\tilde{s}_{jj}}{\xi_j^2 + \lambda_j^2}}.$$

where $\tilde{s}_{jk} = N^{-1} \sum_{i=1}^N (Y_{ij} - \bar{Y}_j)(Y_{ik} - \bar{Y}_k)$ and $\bar{Y}_j = N^{-1} \sum_{i=1}^N Y_{ij}$, $j, k = 1, \dots, n$. Second, using $\mathbf{\Delta}^{(0)}$, we get $\mathbf{\Psi}^{(0)}$ and $\mathbf{\beta}^{(0)}$ as:

$$\mathbf{\Psi}^{(0)} = (\mathbf{\Delta}^{(0)})^{-1} \mathbf{\Lambda}^{-1} (\tilde{\mathbf{S}} - \mathbf{\Delta}^{(0)} \mathbf{\xi} \mathbf{\xi}' \mathbf{\Delta}^{(0)}) \mathbf{\Lambda}^{-1} (\mathbf{\Delta}^{(0)})^{-1},$$

and

$$\mathbf{\beta}^{(0)} = N^{-1} \sum_{i=1}^N (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i (\mathbf{Y}_i - \mathbf{\Delta}^{(0)} \mathbf{\xi}),$$

respectively, where $\tilde{\mathbf{S}} = \{\tilde{s}_{jk}\}_{j,k=1}^n$.

The covariance matrix of $\hat{\boldsymbol{\theta}}$ is calculated using the so-called ‘sandwich’ estimator (Welsh, 1996) plugging-in the MLE in the expression of $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$, that is, $V(\hat{\boldsymbol{\theta}}) = \{\mathbf{A}(\hat{\boldsymbol{\theta}})\}^{-1} \mathbf{B}(\hat{\boldsymbol{\theta}}) \{\mathbf{A}(\hat{\boldsymbol{\theta}})\}^{-1}$ where

$$\mathbf{A}(\hat{\boldsymbol{\theta}}) = N^{-1} \sum_{i=1}^N \frac{\partial^2}{\partial \boldsymbol{\theta}' \partial \boldsymbol{\theta}} f_{\epsilon}(\mathbf{Y}_i | \mathbf{X}_i; \hat{\boldsymbol{\theta}}),$$

and

$$\mathbf{B}(\hat{\boldsymbol{\theta}}) = N^{-1} \sum_{i=1}^N \left[\frac{\partial}{\partial \boldsymbol{\theta}} f_{\epsilon}(\mathbf{Y}_i | \mathbf{X}_i; \hat{\boldsymbol{\theta}}) \right] \left[\frac{\partial}{\partial \boldsymbol{\theta}} f_{\epsilon}(\mathbf{Y}_i | \mathbf{X}_i; \hat{\boldsymbol{\theta}}) \right]',$$

respectively.

Note that the maximization is done using the MGH density after adding ϵ , that is, $f_{\epsilon}(\mathbf{y} | \mathbf{X}_i; \boldsymbol{\theta})$. In this way, we not only prevent that the algorithm lands in a local maximum, but it also allows us to compute the variance matrix of $\hat{\boldsymbol{\theta}}$ correctly.

The choice of ϵ can be done by checking the gradients and Hessian matrix evaluated at $\hat{\boldsymbol{\theta}}$. When the solution of the algorithm falls into a peak (local maximum), these quantities are considerably large (also, the Hessian matrix is possibly non-PD), leading to a volatile estimation of $V(\hat{\boldsymbol{\theta}})$. Therefore, we can proceed increasing ϵ until the gradients are close to zero, and the Hessian matrix and $V(\hat{\boldsymbol{\theta}})$ are stable.

The algorithm does not always get trapped in a peak. Nevertheless, we recommend in any case to conduct a sensitivity analysis checking the estimates and standard errors for several values of ϵ .

When $\epsilon = 0$, the likelihood function (3.4) is not differentiable with respect to β to all points, but it is differentiable with respect to the other parameters (Δ, Ψ) . Therefore, the maximization (when $\epsilon = 0$) can be done cycling between two maximization algorithms. Firstly, maximizing (3.4) with respect to (Δ, Ψ) using the BFGS algorithm. Secondly, maximizing (3.4) with respect to β using the derivative-free Nelder-Mead algorithm (Nelder and Mead, 1965). The procedure stops when a convergence criterion is satisfied.

5 Simulations

The main objective of the simulations is to evaluate the performance of our proposal not only for estimating the parameters of the multivariate asymmetric Laplace distribution, but most importantly, for estimating the coefficients of the quantile regression with correlated data.

Therefore, we consider two simulation settings. First, we consider a bivariate asymmetric Laplace distribution as data-generating model for the outcome. Thereby, we illustrate that our estimator provides efficient estimates for the parameters of the MAL distribution. The main results of this simulation study are shown in Section B of the Supplementary Materials.

Second, we contemplate settings in the quantile regression context. Here, we compare our proposal with the univariate quantile regression estimator (UQR). The data-generating models and results of these simulations are presented in Sections 5.1 and 5.2, respectively.

5.1 Settings

The settings resemble longitudinal data with two measurements per subject. The bivariate data-generating model is:

$$Y_{ij} = \alpha_0 + t_j \alpha_1 + t_j T_i \alpha_2 + (\gamma_0 + t_j \gamma_1 + t_j T_i \gamma_2) \epsilon_{ij}, \text{ for } i = 1, \dots, N, \text{ and } j = 1, 2,$$

with $t_j = j - 1$ indicating the measuring time, and T_i representing a Bernoulli variable with success probability $\pi = 0.5$. For the vector of error terms, $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2})'$, we consider three different distributions. These are:

$$\epsilon_i \sim N(0, S), \epsilon_i \sim t_3\left(0, \frac{1}{3}S\right), \text{ and } \epsilon_i \sim \text{Cauchy}(0, S),$$

where the covariance matrix S is defined as:

$$S = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

We are aiming to estimate the τ -th conditional quantile of $\mathbf{Y}_{ij}$. This is then given by:

$$\begin{aligned} Q_Y(\tau|\mathbf{x}_{ij}) &= \alpha_0 + \gamma_0 Q_\varepsilon(\tau) + t_j [\alpha_1 + \gamma_1 Q_\varepsilon(\tau)] + T_i t_j [\alpha_2 + \gamma_2 Q_\varepsilon(\tau)] \\ &= \beta_0 + t_j \beta_1 + T_i t_j \beta_2, \end{aligned}$$

where $\mathbf{x}_{ij} = (1, t_j, T_i t_j)'$, and $Q_\varepsilon(\cdot)$ is the quantile function of the error distribution.

For these scenarios, we set $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)' = (4, 2, 1)'$. The number of subjects, the quantile level, and the correlation are set to $N = \{50, 100, 200\}$, $\tau = \{0.25, 0.5, 0.9\}$, and $\rho = \{0.5, 0.9\}$, respectively, leading to 18 scenarios for each distribution of the errors. Furthermore, 1 000 datasets are simulated in each case.

5.1.1 Parameters of interest

The parameter of interest is $\boldsymbol{\beta}$ which depends on τ and the distribution of the error vectors. This takes the form:

$$\boldsymbol{\beta} = \boldsymbol{\alpha} + Q_\varepsilon(\tau)\boldsymbol{\gamma},$$

where $\boldsymbol{\gamma}$ is a vector of the same dimension as $\boldsymbol{\alpha}$. For the bivariate normal, Student- t and Cauchy cases, $\boldsymbol{\gamma}$ is $(1, 1, 0)$, $(\sqrt{1/3}, \sqrt{1/3}, 0)$, and $(1, 1, 0)$, respectively. Note that ε_{ij} is independent of T_i .

5.1.2 Estimators

We estimate the parameter of interest using the MLE with $\epsilon = 0.01$. This value was chosen based on checking simulated data. Overall, this quantity provides good estimates for the standard errors. However, this does not mean that this value is optimal in all cases. Depending on the setting, it can be smaller or larger. See also the end of Section 4 for more guidelines on how to choose ϵ in practice.

Although the inclusion of a small value for ϵ leads to some bias, this is negligible. Furthermore, simulations show that the MLE with $\epsilon = 0.01$ is more efficient than the MLE with $\epsilon = 0$ in all cases. These results are presented in Section E of the Supplementary Materials.

As a reference, we compare the MLE with the univariate quantile regression estimator (UQR). The comparison is done using the relative bias (RB) and the relative efficiency (RE) of $\hat{\beta}_j$, $j = 1, \dots, p$. The latter is defined as:

$$RE_j = \frac{\text{MSE}_{\text{UQR}}(\hat{\beta}_j)}{\text{MSE}_{\text{MLE}}(\hat{\beta}_j)}.$$

If $RE_j > 1$, it indicates that the MLE is more efficient than the reference estimator. Furthermore, we consider the coverage of 95% confidence intervals for β_j based on the MLE with $\epsilon = 0.01$, i.e., the proportion of samples for which the true parameter β_j is contained in the confidence interval.

5.2 Results

In this section, we present the main findings of the estimators of β for the bivariate normal and Student- t scenarios. The results for the Cauchy cases are exhibited in Section C of the Supplementary Materials. Furthermore, the settings with bivariate normal and Student- t distributions are extended to two longitudinal variables with two measurements each. These simulations are displayed in Section D of the Supplementary Materials.

The relative bias and efficiency of the MLE for each element of β in the bivariate normal and Student- t cases are displayed in Table 1. In both cases, similar results are found. The MLE seems to be unbiased (all relative bias are smaller or around 5%). Generally, the MLE provides better estimates than the UQR. Its efficiency depends mostly on the correlation. When the outcomes are strongly correlated ($\rho = 0.9$), the MLE provides considerably more efficient estimates for β_1 and β_2 . However, it is slightly less efficient for β_0 . Furthermore, the REs are larger in the Student- t cases.

The coverage of the 95% confidence intervals of each element of β based on the MLE with $\epsilon = 0.01$ for the bivariate normal and Student- t cases is presented in Table 2. For all parameters in both scenarios, the coverage gets closer to 0.95 as N gets bigger. Nevertheless, this convergence to the nominal value is slower for $\tau = 0.9$.

6 Leuven diabetes project

The Leuven diabetes project (LDP) is a randomized trial with before/after measurements and two intervention arms, conducted from January 2005 until December 2006. It aimed to evaluate the effectiveness of a structured model for chronic diabetes care based on clinical outcomes of the patient and to create bases for the development of a national diabetes care programme (Borgermans et al., 2009). The study is discussed in detail in Borgermans et al. (2008) and Borgermans et al. (2009), and it is further analysed by Ivanova et al. (2015).

Figure 1 displays the relationship between the LDL-cholesterol and HbA1c at the beginning of the programme and one year later. There is a moderate positive correlation (around 0.5) between each variable at both time points. However, the correlations between variables are considerably low. Furthermore, both variables show a positive asymmetric behaviour in their densities with a larger asymmetry for the HbA1c.

In the analysis, we consider the following covariates: gender, age, use of insulin, diabetes duration (in decades), and body mass index (BMI) of the patient. Unfortunately, there are missing values in both outcomes, roughly 15% in LDL-cholesterol and 6% in HbA1c, and in all covariates, ranging between 2% and 11%. For the analysis, the complete cases are considered, that is, patients without missing values neither in the outcomes and covariates in both measurement points. Therefore, the dataset is reduced to 1 562 patients.

Table 1 Bivariate normal and Student-*t* cases. Relative bias (%) and efficiency of the MLE with $\epsilon = 0.01$ for different values of τ , ρ , and N

distr.	τ	parm	Relative Bias (%)						Relative Efficiency					
			$\rho = 0.50$			$\rho = 0.90$			$\rho = 0.50$			$\rho = 0.90$		
			$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
normal	0.25	β_0	-0.39	-0.51	-0.35	-0.50	-0.35	-0.30	1.02	1.00	1.02	0.99	1.04	1.06
		β_1	-1.71	-1.35	-1.30	-1.21	-0.76	-0.89	1.31	1.46	1.41	2.98	3.13	3.13
		β_2	0.27	1.40	0.99	-0.29	0.24	0.59	1.44	1.57	1.64	5.96	6.32	7.04
	0.50	β_0	0.09	-0.08	-0.01	0.06	-0.07	-0.01	1.02	1.08	1.01	1.04	1.10	1.06
		β_1	-0.24	-0.06	-0.36	-0.08	-0.12	-0.15	1.31	1.38	1.39	2.87	3.03	3.20
		β_2	1.34	-0.36	1.58	0.28	-0.10	0.76	1.20	1.37	1.37	4.64	5.31	5.50
Student- <i>t</i>	0.90	β_0	1.31	1.38	1.49	0.77	0.79	0.88	0.99	0.91	0.75	1.11	1.04	0.99
		β_1	1.87	1.99	2.21	1.06	1.34	1.38	1.61	1.54	1.57	3.45	3.40	3.26
		β_2	-1.17	0.43	1.24	0.04	-0.16	0.77	2.10	2.18	2.17	9.43	9.74	9.14
	0.25	β_0	0.40	0.51	0.72	0.47	0.58	0.80	1.09	1.12	1.03	1.09	1.14	1.00
		β_1	2.08	1.86	1.25	1.69	1.60	1.75	1.50	1.70	1.66	3.58	3.83	3.54
		β_2	-2.48	-0.45	0.22	-1.38	0.07	-0.13	1.89	2.05	1.97	7.78	8.79	8.23
	0.50	β_0	-0.02	-0.06	0.02	-0.02	-0.06	0.02	1.11	1.12	1.15	1.11	1.12	1.15
		β_1	1.09	0.15	0.18	0.51	-0.02	0.13	1.39	1.51	1.44	3.00	3.50	2.93
		β_2	-2.48	-0.07	-0.75	-1.25	-0.03	-0.38	1.42	1.53	1.48	5.04	6.05	5.51
	0.90	β_0	0.46	0.02	0.03	-0.36	-0.67	-0.62	1.45	1.38	1.34	1.60	1.38	1.34
		β_1	1.71	0.37	0.07	-0.02	-1.07	-1.02	2.85	2.22	2.52	5.89	4.14	4.93
		β_2	-2.30	-0.96	-0.74	-1.52	-0.06	-0.48	5.00	4.43	4.78	24.20	22.23	23.37

Table 2 Bivariate normal and Student-*t* cases. Coverage of the 95% confidence intervals based on the MLE with $\epsilon = 0.01$ for different values of τ , ρ , and N

distr.	τ	parm	$\rho = 0.50$			$\rho = 0.90$		
			$N = 50$	$N = 100$	$N = 200$	$N = 50$	$N = 100$	$N = 200$
normal	0.25	β_0	89.7	91.8	93.7	89.5	94.1	93.8
		β_1	85.7	89.0	92.4	87.3	91.4	93.0
		β_2	89.7	93.7	93.4	89.9	93.7	94.4
	0.50	β_0	90.1	93.3	93.3	89.0	92.2	93.0
		β_1	87.2	90.0	93.9	89.4	91.9	95.1
		β_2	89.6	92.5	93.8	90.1	93.0	94.0
	0.90	β_0	83.0	88.8	88.8	82.1	88.2	92.5
		β_1	83.7	90.1	92.1	83.9	89.6	91.9
		β_2	91.1	94.1	91.9	90.7	93.6	94.3
Student- <i>t</i>	0.25	β_0	87.0	93.0	89.9	87.1	93.3	88.6
		β_1	85.6	90.5	91.6	87.3	91.5	91.1
		β_2	91.5	94.4	94.2	92.7	94.2	93.0
	0.50	β_0	92.5	93.3	93.7	92.5	93.3	93.7
		β_1	87.7	91.1	91.2	89.2	92.3	92.1
		β_2	89.2	92.6	92.8	89.2	92.6	92.8
	0.90	β_0	84.4	89.5	91.4	82.8	85.6	89.6
		β_1	86.9	91.5	92.8	87.0	88.4	91.2
		β_2	93.7	93.5	95.4	92.9	93.5	95.4

Let $Y_{1,i}$ and $Y_{2,i}$ be the LDL-cholesterol and HbA1c measured in the i th patient ($i = 1, \dots, 1562$) at occasion t_j ($j = 1, 2$), respectively. For both of them, we consider the same covariance structure. Then, for the l th outcome ($l = 1, 2$), the quantile model takes the form:

$$\begin{aligned} Q_{Y_l}(\tau | \text{gender}_i, \dots, \text{BMI}_{ij}) = & \beta_{l,0} + \text{gender}_i \beta_{l,1} + \text{age}_i \beta_{l,2} + \text{insulin}_i \beta_{l,3} \\ & + \text{Diab.dur}_i \beta_{l,4} + t_j \beta_{l,5} + \text{BMI}_{ij} \beta_{l,6}, \end{aligned} \quad (6.1)$$

where $t_1 = 0$ and $t_2 = 1$, $\text{gender}=1$ if the patient is female ($\text{gender}=0$ otherwise), $\text{insulin}=1$ if the patient uses insulin ($\text{insulin}=0$ otherwise). Furthermore, the continuous covariates (BMI, age, and diabetes duration) were centred at zero to reduce collinearity and possible convergence issues.

The selection of $\epsilon = 0.01$ is based on a sensitivity analysis in which the estimates, standard errors and gradients are computed for a wide range of values of ϵ (see Section F of the Supplementary Materials). Here, we observed that none of these quantities are considerably affected by ϵ .

Table 3 displays the parameter estimates for quantile levels 0.25, 0.5, and 0.9 of model (6.1) using the MLE with $\epsilon = 0.01$. For comparison, the estimates based on the univariate quantile regression (UQR) are also included. For UQR, the standard errors were computed using cluster bootstrap (Hagemann, 2017).

Figure 1 LDP data. Scatter plot matrix of the LDL-cholesterol and HbA1c at baseline (LDL.0 and Hba1c.0), and after one year on the programme (LDL.1 and Hba1c.1). The entries above the main diagonal displays the correlations, the entries below the main diagonal the scatter plots, and in the main diagonal the densities

The results in Table 3 show that both estimators provide similar estimates and standard errors. Nevertheless, some differences can be found. For both endpoints, there is a significant negative effect of time in all quantile levels indicating that the treatment successfully improves patients' conditions. However, this effect is consistently smaller for the MLE with a lower standard error. Similarly, the effect of the use of insulin is significantly negative for LDL-cholesterol, but positive for HbA1c. The former is stronger for the MLE, with a larger standard error, for quantiles 0.25

Table 3 LDP data. Parameter estimates, standard errors, and p -values for the maximum likelihood estimator (MLE) with $\epsilon = 0.01$, and the univariate quantile regression (UQR) for different quantile levels (τ). Important differences between the p -values for the two estimators are in bold

τ	effect	LDL-cholesterol				HbA1c			
		UQR		MLE		UQR		MLE	
		est (s.e.)	p-val	est (s.e.)	p-val	est (s.e.)	p-val	est (s.e.)	p-val
0.25	Intercept	85.07 (1.521)	0.000	84.05 (1.380)	0.000	6.29 (0.032)	0.000	6.27 (0.028)	0.000
	Gender	4.23 (1.608)	0.009	4.92 (1.717)	0.004	0.04 (0.038)	0.313	0.07 (0.036)	0.064
	Age	0.15 (0.064)	0.022	0.16 (0.089)	0.077	-0.01 (0.002)	0.004	-0.01 (0.002)	0.001
	Insulin	-10.37 (2.345)	0.000	-13.06 (2.113)	0.000	0.38 (0.065)	0.000	0.20 (0.056)	0.000
	Diab.dur.	-1.95 (1.695)	0.251	-1.40 (1.435)	0.331	0.13 (0.039)	0.001	0.23 (0.033)	0.000
	Time	-13.12 (1.029)	0.000	-9.54 (0.772)	0.000	-0.19 (0.025)	0.000	-0.14 (0.017)	0.000
	BMI	-0.17 (0.158)	0.285	-0.01 (0.177)	0.962	0.01 (0.003)	0.003	0.01 (0.003)	0.000
0.50	Intercept	103.88 (1.302)	0.000	102.81 (1.284)	0.000	6.76 (0.039)	0.000	6.74 (0.035)	0.000
	Gender	6.34 (1.518)	0.000	5.21 (1.702)	0.002	0.02 (0.040)	0.629	0.03 (0.039)	0.414
	Age	0.09 (0.078)	0.251	0.11 (0.080)	0.162	-0.01 (0.002)	0.000	-0.01 (0.002)	0.000
	Insulin	-10.98 (2.221)	0.000	-10.00 (2.571)	0.000	0.47 (0.057)	0.000	0.30 (0.065)	0.000
	Diab.dur.	-2.36 (1.680)	0.160	-3.46 (1.440)	0.016	0.24 (0.035)	0.000	0.25 (0.033)	0.000
	Time	-11.91 (1.148)	0.000	-9.29 (0.822)	0.000	-0.24 (0.025)	0.000	-0.22 (0.018)	0.000
	BMI	0.05 (0.189)	0.774	0.20 (0.159)	0.204	0.01 (0.004)	0.063	0.02 (0.003)	0.000
0.90	Intercept	148.17 (2.476)	0.000	147.05 (2.092)	0.000	8.34 (0.129)	0.000	8.24 (0.077)	0.000
	Gender	4.76 (3.197)	0.136	3.91 (2.576)	0.129	-0.09 (0.115)	0.447	-0.04 (0.077)	0.589
	Age	0.02 (0.147)	0.905	0.05 (0.119)	0.677	-0.02 (0.005)	0.003	-0.01 (0.003)	0.001
	Insulin	-4.48 (5.077)	0.377	-7.45 (2.801)	0.008	0.58 (0.172)	0.001	0.33 (0.083)	0.000
	Diab.dur.	-6.70 (2.444)	0.006	-3.02 (2.161)	0.162	0.31 (0.096)	0.001	0.30 (0.066)	0.000
	Time	-13.97 (2.314)	0.000	-12.28 (1.421)	0.000	-0.67 (0.106)	0.000	-0.55 (0.057)	0.000
	BMI	-0.17 (0.320)	0.603	0.63 (0.243)	0.009	0.03 (0.015)	0.037	0.03 (0.008)	0.000

and 0.9. The gender of the patient has significant effect on LDL-cholesterol with $\tau = 0.25$ and $\tau = 0.5$, but not at $\tau = 0.9$. Furthermore, BMI is only significant for HbA1c.

7 Conclusion and perspectives

Summary. In this article, we proposed a new quantile regression model that takes the dependence structure in multivariate or longitudinal data into account. The model is based on the MGH distribution. The proposed estimation method is studied both from an asymptotic and a finite sample point of view.

Limitations and further research. Given the peculiarities of the MGH distribution, our model is particularly suitable for positively correlated data, which is typical for data that are collected repeatedly over time on the same subject.

The new methodology opens doors to many extensions, improvements and adaptations, that will be studied in the future. For instance, we can extend the current methodology to a pairwise pseudo-likelihood approach (see, e.g., Hermans

et al., 2018). This will reduce the multivariate log-likelihood to a bivariate (pseudo) log-likelihood, allowing us to use the methodology developed in this article for the special case of $n = 2$. This pseudo-likelihood is also very appealing in the context of large to very large numbers of repetitions.

A further extension exists in allowing for more flexibility in the model, by modelling and estimating the location parameter of the MGH distribution non-parametrically via for example (penalized) B-splines (see, e.g., Bollaerts et al., 2006) or local polynomial modelling (see, e.g., Gijbels et al., 2019). These nonparametric specifications of the location parameter can then also be used for model diagnostics, by comparing them with the linear fit by means of a suitable test statistic that measures the distance between the two fits.

Acknowledgements

The authors would like to thank Michael G. Kenward and Geert Verbeke for helpful suggestions.

Supplementary materials

The R-code for executing the simulations and the data analysis is available at <http://www.statmod.org/smij/archive.html>.

Additional results and technical details are exhibited in the Supplementary Materials available online (<http://www.statmod.org/smij/archive.html>). In Section A, an example of a multivariate longitudinal setting is introduced. Sections B-E show additional results of the simulation study. Finally, a sensitivity analysis of the MLE for selecting ϵ using the LDP and simulated data is presented in Section F. Sections G and H are related to the MAL distribution and to the asymptotic theory for the proposed estimator, respectively.

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 disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The authors gratefully acknowledge Special Research Fund (Bijzonder Onderzoeksfonds) of Hasselt University [BOF14NI06], and the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409).

References

- Barndor-Nielsen O and Blaesild P (1981) Hyperbolic distributions and ramifications: Contributions to theory and application. In *Statistical Distributions in Scientific Work*, edited by C Taillie, GP Patil and BA Baldessari. Vol. 79, pages 19–44. Dordrecht: Springer.
- Barndor-Nielsen O and Kendall DG (1977) Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. A: Mathematical and Physical Sciences*, 353, 401–19.
- Bollaerts K, Eilers PHC and Aerts M (2006) Quantile regression with monotonicity restrictions using P-splines and the L1-norm. *Statistical Modelling*, 6, 189–207.
- Borgermans L, Goderis G, Broeke CVD, Mathieu C, Aertgeerts B, Verbeke G, Carbonez A, Ivanova A, Grol R and Heyrman J (2008) A cluster randomized trial to improve adherence to evidence-based guidelines on diabetes and reduce clinical inertia in primary care physicians in Belgium: Study protocol [NTR 1369]. *Implementation Science*, 3. doi: 10.1186/1748-5908-3-42.
- Borgermans L, Goderis G, Broeke CVD, Verbeke G, Carbonez A, Ivanova A, Mathieu C, Aertgeerts B, Heyrman J and Grol R (2009) Interdisciplinary diabetes care teams operating on the interface between primary and specialty care are associated with improved outcomes of care: Findings from the Leuven diabetes project. *BMC Health Services Research*, 9(1). doi: 10.1186/1472-6963-9-179.
- Chen L, Wei L-J and Parzen MI (2004) Quantile regression for correlated observations. In *Proceedings of the Second Seattle Symposium in Biostatistics*, edited by Danyu Lin and PJ Heagerty, pages 51–69. New York, NY: Springer.
- De Backer M, El Ghouh A and Van Keilegom I (2019) An adapted loss function for censored quantile regression. *Journal of the American Statistical Association*, 114, 1126–37.
- Farcomeni A (2012) Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Statistics and Computing*, 22, 141–52.
- Fletcher R (2000) *Practical Methods of Optimization*, 2nd edition. Hoboken, NJ: John Wiley & Sons.
- Galvao AF and Kato K (2017) Quantile regression methods for longitudinal data. In *Handbook of Quantile Regression*, edited by Roger Koenker, Victor Chernozhukov, Xuming He and Limin Peng, pages 363–80. Boca Raton, FL: Chapman and Hall/CRC.
- Geraci M and Bottai M (2007) Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, 8, 140–54.
- (2014) Linear quantile mixed models. *Statistics and Computing*, 24, 461–79.
- Gijbels I, Karim R and Verhasselt A (2019). Quantile estimation in a generalized asymmetric distributional setting. In *Springer Proceedings in Mathematics and Statistics, Proceedings of SMSA 2019, 14th Workshop on Stochastic Models, Statistics and their Application*, Dresden, edited by A. Steland, E. Rafajlowicz and O. Okhrin (to appear). Berlin: Springer.
- Hagemann A (2017) Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association*, 112, 446–56.
- Hermans L, Nassiri V, Molenberghs G, Kenward M, Van der Elst W, Aerts M and Verbeke G (2018) Fast, closed-form, and efficient estimators for hierarchical models with ar(1) covariance and unequal cluster sizes. *Communications in Statistics: Simulation and Computation*, 47, 1492–1505.
- Ivanova A, Molenberghs G and Verbeke G (2015) Fast and highly efficient pseudo-likelihood methodology for large and complex ordinal data. *Statistical Methods in Medical Research*, 26, 2758–779.

- Koenker R (2004) Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74–89.
- (2005) *Quantile Regression*. Cambridge: Cambridge University Press.
- Koenker R and Bassett G (1978). Regression quantiles. *Econometrica*, **46**, 33–50.
- Kotz S, Kozubowski TJ and Podgorski K (2001) *The Laplace Distribution and Generalizations*. Secaucus, NJ: Springer Science/Business Media.
- Kozubowski TJ and Podgorski K (2000) A multivariate and asymmetric generalization of Laplace distribution. *Computational Statistics*, **15**, 531–40.
- Lamarche C (2010) Robust penalized quantile regression estimation for panel data. *Journal of Econometrics*, **157**, 396–408.
- Lee D and Neocleous T (2010) Bayesian quantile regression for count data with application to environmental epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **59**, 905–20.
- Lipsitz SR, Fitzmaurice GM, Molenberghs G and Zhao LP (1997) Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **46**, 463–76.
- Liu Y and Bottai M (2009). Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics*, **5**(1).
- McNeil A, Frey R and Embrechts P (2005) *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.
- Molenberghs G and Verbeke G (2005) *Models for Discrete Longitudinal Data*. New York, NY: Springer.
- Nelder JA and Mead R (1965) A simplex method for function minimization. *The Computer Journal*, **7**, 308–13.
- Parente PM and Santos Silva JM (2016) Quantile regression with clustered data. *Journal of Econometric Methods*, **5**(1), 1–15.
- 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.
- Poiraud-Casanova S and Thomas-Agnan C (2000) About monotone regression quantiles. *Statistics and Probability Letters*, **48**, 101–4.
- Pourahmadi M and Wang X (2015) Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor. *Statistics & Probability Letters*, **106**, 5–12.
- Rousseeuw PJ and Molenberghs G (1993) Transformation of non positive semidefinite correlation matrices. *Communications in Statistics, Theory and Methods*, **22**, 965–84.
- Verhasselt A, Florez AJ, Molenberghs G, Nassiri V, Mamouris P and Vaes B (2018) Multiple imputation for quantile regression with missing response. Submitted, 000–000.
- Waldmann E and Kneib T (2015) Bayesian bivariate quantile regression. *Statistical Modelling*, **15**, 326–44.
- Wang HJ and Wang L (2009) Locally weighted censored quantile regression. *Journal of the American Statistical Association*, **104**, 1117–28.
- Wang L, Van Keilegom I and Maidman A (2018) Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors. *Biometrika*, **105**, 859–72.
- Welsh AH (1996) *Aspects of Statistical Inference*. New York, NY: Wiley-Interscience.
- White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.
- Yuan Y and Yin G (2010) Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, **66**, 105–114.