Copula-based pairwise estimator for quantile regression with hierarchical missing data

Peer-reviewed author version

VERHASSELT, Anneleen; FLOREZ POVEDA, Alvaro; MOLENBERGHS, Geert & VAN KEILEGOM, Ingrid (2024) Copula-based pairwise estimator for quantile regression with hierarchical missing data. In: Statistical modelling, 25 (2), p. 129-149.

DOI: 10.1177/1471082X231225806
Handle: http://hdl.handle.net/1942/42649

Copula-based pairwise estimator for quantile regression with hierarchical missing data

Anneleen Verhasselt [1], Alvaro J. Flórez [1,2], Geert Molenberghs [1,3], and Ingrid Van Keilegom [4]

[1] Data Science Institute, I-Biostat, Universiteit Hasselt, Diepenbeek, Belgium

[2] School of Statistics, Universidad del Valle, Cali, Colombia

[3] I-BioStat, KU Leuven, Leuven, Belgium

[4] ORSTAT, KU Leuven, Leuven, Belgium

Address for correspondence: Alvaro José Flórez, School of Statistics, Faculty of Engineering, Universidad del Valle, Edificio E56, Ciudad Universitaria - Meléndez, Calle 13 # 100-00, Cali, Colombia.

E-mail: alvaro.florez@correounivalle.edu.co.

Phone: (+57) 319 523 09 60.

Fax: (+572) 321 21 00.

Abstract: Quantile regression can be a helpful technique for analyzing clustered (such as longitudinal) data. It can characterize the change in response over time without making distributional assumptions and is robust to outliers in the response. A quantile regression model using a copula-based multivariate asymmetric Laplace distribution for addressing correlation due to clustering is introduced. Furthermore,

we propose a pairwise estimator for the parameters of the model. Since it is based on pseudo-likelihood, it needs to be modified to avoid bias in presence of missingness. Therefore, we enhance the model with inverse probability weighting. In this way, our proposal is unbiased under the missing at random assumption. Based on simulations, the estimator is efficient and computationally fast. Finally, the methodology is illustrated using a study in ophthalmology.

Key words: Asymmetric Laplace distribution; copulas; inverse probability weighting; quantile regression; longitudinal data; missing data; pairwise estimator.

1 Introduction

An important feature of clustered data, such as longitudinal data, is that observations within clusters can be correlated, but measurements of different clusters are independent. This correlation needs to be accommodated in the modeling. Furthermore, in many settings, missing data are commonly encountered. For instance, in longitudinal studies, not all planned measurements of a subject's outcome are observed. Therefore, it is necessary to address missingness into the modeling.

Quantile regression (QR; Koenker and Bassett, 1978;Koenker, 2005) is an appealing technique to summarize the effect of a set of covariates on different quantiles of a response variable. When analyzing growth curves, interest can be in the lower tails of the distribution, e.g., the lower weights in babies' growth charts (Abrevaya and Dahl, 2008). Moreover, it makes no distributional assumptions on the outcomes, rendering it useful when the distribution of the response is skewed or heavy-tailed.

The properties of QR methods are well established for independent (cross-sectional) settings (see Koenker, 2005), and in the last few years, several approaches to extend QR to clustered data have been proposed and discussed. These can be classified along a conditional versus marginal modeling dichotomy. Within the former, dependence is accounted for by including cluster-specific random effect into the linear QR (Geraci and Bottai, 2006, 2013; Marino et al., 2018; Geraci, 2019; Merlo et al., 2022). On the other hand, within a marginal modeling framework, the correlation is addressed via generalized estimating equations (He et al., 2003; Yi and He, 2008; Fu and Wang, 2012) and likelihood-based methods on multivariate extensions of the asymmetric Laplace (AL) distribution (De Backer et al., 2017; Wang et al., 2019; Petrella and Raponi, 2019; Flórez et al., 2022). Petrella and Raponi (2019) implemented an expectation-maximization (EM) estimator based on the multivariate asymmetric Laplace (MAL) distribution introduced by Kotz (2001). Subsequently, to solve some mathematical and computational issues, Flórez et al. (2022) proposed a maximum likelihood estimator (MLE) based on the family of multivariate generalized hyperbolic (MGH) distributions, of which the MAL distribution is a limiting case. De Backer et al. (2017) and Wang et al. (2019) consider semi-parametric QR models for different settings in which the data dependence is addressed via copulas (Nelsen, 2007). We refer to Galvao and Kato (2017) and Marino and Farcomeni (2015) for a review of the literature on QR for correlated data.

In this paper, we consider a QR model based on a multivariate AL distribution constructed through copulas. In this fashion, the marginal quantiles are modeled by AL distributions, and a copula captures the within-cluster association. Given that we can choose different types of copula structures, this approach permits a flexible and wide range of dependence. On the contrary, other multivariate representations of the

AL distributions can be restricted to some correlation structures. For example, the MAL distribution (Kotz, 2001) is mildly restricted to positive correlations when the skewness parameter (τ) is near 0.5. However, it is heavily restricted when τ is close to the boundaries (0 or 1).

Likelihood-based estimation is a challenging task. Firstly, the log-likelihood function based on the AL distribution is not differentiable with respect to the QR coefficients. Secondly, it is computationally burdensome as the number of repeated measurements increases. Wang et al. (2019) addressed the former by induced smoothing methods and the latter by cycling through calculating the dependence parameters and quantile coefficients in a three-step estimator. Our proposal handles these issues differently.

For the non-differentiability of the log-likelihood, a perturbation of the AL density function can be employed to make it smooth and, therefore, differentiable. Algorithms based on smoothing non-differentiable QR objective functions have been shown to provide good results (Hunter and Lange, 2000; Chen, 2007). Moreover, we propose a pairwise estimator based on the pseudo-likelihood (PL) framework (Molenberghs et al., 2011) for handling the computational complexity. The idea of PL is to replace a complicated joint likelihood function with a simple and computationally tractable one, in this case, the product of the pairwise bivariate density functions. The PL methodology leads to consistent and asymptotically normal estimators; however, with an often slight reduction in precision (Arnold and Strauss, 1991; Aerts et al., 2002).

Depending on its nature, missing data can lead to substantially biased estimates. The statistical toolkit for handling missingness is vast, where the most common techniques are multiple imputation (MI; Rubin, 1987; van Buuren, 2012; Carpenter and Kenward, 2013) and inverse probability weighting (IPW; Robins et al., 1995; Fitz-

maurice, 2015). MI techniques follow three phases: (1) replacing the missing values by a collection of several plausible values based on a predictive model using the observed data, (2) fitting each multiply imputed dataset with complete data techniques, and (3) applying the combination rules. In the IPW approach, bias is corrected by weighting the observed measures by the reciprocal of their probability of being observed. When both techniques are implemented correctly, the MI is generally more efficient than the IPW approach. Nevertheless, it is technically and computationally more complex (Seaman and White, 2011; Little et al., 2022). For these reasons, we opt for IPW, modeling the probability of data being missing using a logistic model for computing the weights for each pair of the pairwise pseudo-likelihood. A general framework for handling incomplete data within pseudo-likelihood settings can be found in Molenberghs et al. (2011). Other alternatives have been implemented in the QR framework to informative missing data (Farcomeni and Viviani, 2015; Yu et al., 2023).

The rest of the paper is organized as follows. In Section 2, an overview of QR is given. The copula-based asymmetric Laplace model is introduced in Section 3. The MLE and the pairwise estimator are described in Section 4. The missing data problem and the inverse probability weighting approach are presented in Section 5. In Section 6, a simulation study for evaluating the performance of our proposal is performed. The analysis of a dataset in ophthalmology is shown in Section 7. Finally, some concluding remarks and future ideas are reserved for Section 8.

2 Quantile regression

We consider a balanced longitudinal design in which N individuals are planned to be observed on n occasions at fixed times. Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$ be an n-dimensional response vector for the i-th individual, for $i = 1, \ldots, N$, and $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{in})'$ its corresponding $(n \times p)$ matrix of covariates. Consider the following linear QR model:

$$Q_{\mathbf{Y}_i}(\tau|\mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\beta}^\tau,$$

where $\boldsymbol{\beta}^\tau = (\beta_0^\tau, \beta_1^\tau, \ldots, \beta_{p-1}^\tau)'$ is the vector of QR coefficients and $Q_{\mathbf{Y}_i}(\tau|\mathbf{X}_i)$ is the elementwise τ-th conditional quantile of $\mathbf{Y}_i$, for $0 < \tau < 1$.

Koenker and Bassett (1978) proposed to estimate $\boldsymbol{\beta}^\tau$ by:

$$\widehat{\boldsymbol{\beta}}^\tau = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N}\sum_{j=1}^{n} \rho_\tau(Y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}), \tag{2.1}$$

where $\rho_\tau(u) = u[\tau - I(u < 0)]$ is the check-loss function and $I(\cdot)$ is the indicator function. Although there is no analytical solution for (2.1), $\widehat{\boldsymbol{\beta}}^\tau$ can be found easily using linear programming methods, such as the Barrodale-Roberts algorithm (Koenker and D'Orey, 1987), or interior point methods, like Frisch–Newton (Portnoy and Koenker, 1997). More details of these methods can be found in Furno and Vistocco (2018), and Koenker (2005).

There is an equivalence between minimization of (2.1) and maximization of the likelihood based on an AL distribution (see Kotz, 2001). Therefore, we can consider that $Y_{ij} \sim \text{AL}(\mu_{ij}, \phi, \tau)$, with the following density function:

$$f_{Y_{ij}}(y_{ij}|\mu_{ij}, \phi, \tau) = \frac{\tau(1-\tau)}{\phi} \exp\left[-\rho_\tau\left(\frac{y_{ij} - \mu_{ij}}{\phi}\right)\right], \tag{2.2}$$

where τ is a skewness parameter, $\phi > 0$ and $\mu_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta}^\tau$ are scale and location

parameters. Then, assuming that the observations are independent, the likelihood function is:

$$L(\boldsymbol{\beta}^{\tau}, \phi) = \prod_{i=1}^{N} \left[\frac{\tau(1-\tau)}{\phi}\right]^{n} \exp\left[-\sum_{j=1}^{n} \rho_{\tau}\left(\frac{Y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}^{\tau}}{\phi}\right)\right]. \tag{2.3}$$

In our case, τ is defined by the investigator, and ϕ is considered a nuisance parameter. The equivalence between the maximization problem (2.3) and the minimization problem (2.1) allows us to use likelihood-based techniques for QR, such as inference, goodness-of-fit (Koenker and Machado, 1999) and Bayesian estimation (Yu and Moyeed, 2001). For more details on the methodology of QR, we refer to Koenker (2005).

In longitudinal settings, the measurements of the same individual are expected to be correlated. Therefore, estimator (2.1) is not optimal. A natural extension of a QR estimator for correlated data is to consider a multivariate representation of the AL distribution and to find an MLE. In this fashion, we can take the dependence between the observations into account.

3 Copula-based asymmetric Laplace model

We construct a multivariate AL distribution through copulas (Joe, 1997). In this methodology, the marginal distributions of a collection of random variables are connected by a copula that accommodates dependency. We consider the following multivariate density of $\mathbf{Y} = (Y_1, \ldots, Y_n)'$:

$$f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta}) = c\left[F_{Y_1}(y_1; \boldsymbol{\eta}_1), \ldots, F_{Y_n}(y_n; \boldsymbol{\eta}_n); \boldsymbol{\theta}_C\right] \prod_{j=1}^{n} f_{Y_j}(y_j; \boldsymbol{\eta}_j),$$

where f_{Y_j} is an AL probability density function (2.2), with $\boldsymbol{\eta}_j = (\mu_j, \phi_j, \tau)'$; F_{Y_j} is the AL cumulative distribution function. Finally, $c(\mathbf{z}; \boldsymbol{\theta}_C)$ is the density of a copula, and

$\boldsymbol{\theta}_C$ is the vector of its parameters. Therefore, $\boldsymbol{\theta} = (\boldsymbol{\theta}_C, \boldsymbol{\phi}, \boldsymbol{\beta}^\tau)$, with $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$.

We consider a Gaussian copula defined as:

$$c(z_1, \ldots, z_n; \mathbf{R}) = |\mathbf{R}|^{-1/2} \exp\left[\frac{1}{2}\mathbf{q}'(\mathbf{I} - \mathbf{R}^{-1})\mathbf{q}\right],$$

where $\mathbf{q} = (q_1, \ldots, q_n)'$ with $q_j = \Phi^{-1}(z_j)$, Φ is the standard normal cumulative distribution function, and $\mathbf{R}$ is a $n \times n$ correlation matrix.

For simplicity, we are assuming a fixed n for all individuals. However, it can easily be extended for different cluster sizes. For these cases, we can impose a structure for the copula, e.g., compound-symmetry or autoregressive.

4 Estimation

Estimation of model (2.2) can be accomplished by maximizing the likelihood function over all parameters simultaneously. Under a correct model specification, the MLE is asymptotically unbiased and the most efficient estimator (White, 1982). However, the computational complexity increases as the dimensionality gets larger, making it impractical when we have several measurements per individual. As an alternative, we implement a pairwise estimator (PWE) using pseudo-likelihood (PL; Molenberghs et al., 2011). Here, the likelihood function is replaced by the product of each pairwise likelihood. In this fashion, the computational burden is reduced. Both methods are introduced below.

4.1 Maximum likelihood estimator

Considering a Gaussian copula, $\boldsymbol{\theta}$ is a $[p + n(1 + (n-1)/2)]$-dimensional vector in which $\boldsymbol{\theta}_C$ consists of the upper-diagonal elements of $\mathbf{R}$. The MLE maximizes the log-likelihood:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \left\{ \log c\left[F_{Y_1}(Y_{i1};\boldsymbol{\eta}_1), \ldots, F_{Y_n}(Y_{in};\boldsymbol{\eta}_n); \boldsymbol{\theta}_C\right] + \sum_{j=1}^{n} \left[\log \frac{\tau(1-\tau)}{\phi_j} - \rho_\tau\left(\frac{Y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}^\tau}{\phi_j}\right)\right] \right\}. \tag{4.1}$$

However, (4.1) is not differentiable with respect to $\boldsymbol{\beta}$ at all points, but it is differentiable with respect to the other parameters ($\boldsymbol{\phi}$ and $\boldsymbol{\theta}_C$). Following Hunter and Lange (2000), we can approximate the log-likelihood by perturbing the check function in (4.1) as:

$$\rho_\tau^\epsilon(u_{ij}) = \rho_\tau(u_{ij}) - \frac{\epsilon}{2}\log\left(\epsilon + |u_{ij}|\right),$$

where $\epsilon > 0$. By this approximation, the (modified) log-likelihood is smooth and a derivative-based algorithm can be implemented to obtain the MLE and its precision. This can be also done by Huber-type functions (Huber, 1973; Chen, 2007).

Finding the maximum of (4.1) should lead to a solution in which $\phi_j > 0$, for $j = 1, \ldots, n$. Furthermore, the constraints on $\boldsymbol{\theta}_C$ depend on the copula's structure. For the Gaussian copula, $\boldsymbol{\theta}_C$ contains the pairwise correlations of a positive-definite (PD) matrix $\mathbf{R}$.

For estimation purposes, we consider the following transformations. Firstly, for ϕ_j, we use $\phi_j^* = \log\phi_j$. Secondly, we reparameterize $\mathbf{R}$ through its hyperspherical coordinates (Rousseeuw and Molenberghs, 1993; Pourahmadi and Wang, 2015). That is, a PD correlation matrix $\mathbf{R} = (r_{jk})_{j,k=1}^n$ can be factorized as $\mathbf{R} = \mathbf{B}\mathbf{B}'$ where

$\mathbf{B} = (b_{jk})_{j,k=1}^{n}$ is a lower triangular matrix with $b_{11} = 1$, $b_{i1} = \cos\,\psi_{i1}$, for $i = 2, \ldots, n$, and

$$b_{jk} = \begin{cases} \prod_{l=1}^{k-1} \sin\,\psi_{jl} & \text{for } j = k, \\ \cos\,\psi_{jk} \prod_{l=1}^{k-1} \sin\,\psi_{jl} & \text{for } j = k+1, \ldots, n, \end{cases}$$

for $k = 2, \ldots, n$, and where $\psi_{jk}, j > k$, are some angles. The matrix $\mathbf{B}$ is unique if the angles are restricted to the interval $(0, \pi)$. An unconstrained set of parameters ψ_{jk}^{*} is obtained by transforming ψ_{jk} as follows:

$$\psi_{jk}^{*} = \log \frac{\psi_{jk}}{\pi - \psi_{jk}}.$$

Therefore, maximizing the unconstrained angles ensures positive-definiteness of the corresponding solution of $\mathbf{R}$.

Then, the set of unconstrained parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\tau}, \boldsymbol{\phi}^{*}, \boldsymbol{\Psi}^{*})$, where $\boldsymbol{\phi}^{*} = (\phi_1^{*}, \ldots, \phi_n^{*})'$ and $\boldsymbol{\Psi}^{*} = (\psi_{ij}^{*})_{i,j=1}^{n}$, can be used for iterative maximization.

Since $\widehat{\boldsymbol{\theta}}$ depends on ϵ, we recommend performing a sensitivity analysis checking the estimates, their standard error, gradients, and the Hessian matrix for several values of ϵ. If this value is too small, the perturbed log-likelihood function is not smooth enough, and the solution may lead to a non-positive-definite Hessian matrix, making variance calculation impossible. On the other hand, if ϵ is too large, the estimates can be biased. An adequate solution requires gradients close to zero and a positive-definite Hessian matrix. In this way, we have a log-likelihood function that is sufficiently smooth, and the standard errors are well calculated. In Section B of the Supplementary Materials, we illustrate this sensitivity analysis using a real dataset.

4.2 Pairwise estimator

Let S be the set of all $\frac{n!}{2!(n-2)!}$ vectors of length n consisting of zeros and ones, with each vector having exactly two non-zero entries (these correspond to all possible pairs). Denote by $\mathbf{Y}_i^{(s)}$ the subvector of $\mathbf{Y}_i$ corresponding to the components of s that are non-zero with associated joint density $f_{\mathbf{Y}^{(s)}}(\mathbf{y}_i^{(s)}; \boldsymbol{\theta}^{(s)})$. Then, the pairwise estimator maximizes the following pseudo log-likelihood (PL):

$$p\ell\,(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{s \in S} \log f_{\mathbf{Y}^{(s)}}(\mathbf{y}_i^{(s)}; \boldsymbol{\theta}^{(s)}). \tag{4.2}$$

Generally, the PL does not lead to a properly defined probability density function. However, under suitable regularity conditions (Molenberghs et al., 2011), it can be shown that maximizing (4.2) produces consistent and asymptotically normally distributed estimators for $\boldsymbol{\theta}$. We have that, asymptotically as $N \to \infty$,

$$N^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*) \xrightarrow{d} N\left[\mathbf{0}, V(\boldsymbol{\theta}_*)\right],$$

where $\boldsymbol{\theta}_*$ is the parameter vector which minimizes the Kullback-Leibler Information criterion (White, 1982), and:

$$V(\widehat{\boldsymbol{\theta}}) = I_0(\boldsymbol{\theta}_*)^{-1} I_1(\boldsymbol{\theta}_*) I_0(\boldsymbol{\theta}_*)^{-1}, \tag{4.3}$$

where $I_0(\boldsymbol{\theta}_*)$ and $I_1(\boldsymbol{\theta}_*)$ are $(q \times q)$ symmetric matrices with:

$$\{I_0(\boldsymbol{\theta}_*)\}_{kl} = -\sum_{s \in S} E_{\boldsymbol{\theta}} \left[\frac{\partial^2 \log f_{\mathbf{Y}^{(s)}}(\mathbf{Y}^{(s)}; \boldsymbol{\theta}_*^{(s)})}{\partial \theta_k \partial \theta_l} \right],$$

and

$$\{I_1(\boldsymbol{\theta}_*)\}_{kl} = \sum_{s \in S} E_{\boldsymbol{\theta}} \left[\frac{\partial \log f_{\mathbf{Y}^{(s)}}(\mathbf{Y}^{(s)}; \boldsymbol{\theta}_*^{(s)})}{\partial \theta_k} \frac{\partial \log f_{\mathbf{Y}^{(s)}}(\mathbf{Y}^{(s)}; \boldsymbol{\theta}_*^{(s)})}{\partial \theta_l} \right], \tag{4.4}$$

respectively. Since the partial derivatives of the log-likelihood with respect to $\boldsymbol{\beta}$ do not always exist in (4.4), these should be understood as the average of the left-hand

derivative and the right-hand derivative. In such a way, there is no problem to first take the derivative and then the expected value to compute $\{I_1(\boldsymbol{\theta}_*)\}_{kl}$.

The main reason to choose a PL approach is its computational convenience. Given that it is based on the sum of bivariate distributions, the PL, and its derivatives, are easier to evaluate than the same quantities for the whole multivariate distribution. Alternatively, we can fit each pair in (4.2) separately using parallel computing and adequately assemble the results to obtain a single estimate for each parameter. In such a way, the computational efficiency can be improved.

5 Missing data problem

Using standard terminology (Little and Rubin, 2002), there are three missing data mechanisms: (1) *missing completely at random* (MCAR) occurs when the non-responses are independent of both unobserved and observed data, (2) *missing at random* (MAR) arises when, conditional on the observed data, the missingness is independent of the unobserved measurements, and finally, (3) *missing not at random* (MNAR) when the process is neither MCAR nor MAR.

Under the MAR assumption, and mild regularity conditions (separability), likelihood-based methods are valid, meaning that they maintain the same properties encountered with full data (Rubin, 1976; Little and Rubin, 2002). On the other hand, the PWE is based on a modification of the likelihood. Therefore, it does not enjoy the same properties. Consequently, it is generally not valid under MAR.

For maximizing (4.2) with incomplete data, we can consider: (1) the complete cases

(CC; considering only the subjects with the planned measurements fully observed), (2) the complete pairs (CP; including in the analysis all the pairs of observations that are fully observed), and (3) the available cases (AC; adding the single information of incomplete pairs, also called the widow factors). However, in all these cases, solving the estimating equations would lead to biased estimations under MAR, except for a few special cases, such as full exchangeability.

Generally, there are two approaches for dealing with missing data, multiple imputation (MI) techniques and inverse probability weighting (IPW) methods. Molenberghs et al. (2011) proposed to employ inverse probability weighting (IPW) or double robustness ideas to overcome this problem. For the former, a weight model is added to account for a subject's contribution to the pseudo-likelihood. For the latter, a predictive model for the unobserved outcomes given the observed ones is included in addition of the weights. Here, we opt for the first approach.

Inverse probability weighting

We are considered the IPW approach in its standard version which is restricted to monotone missingness patterns. In this fashion, the vector of missingness is defined as $\mathbf{G}_i = (G_{i1}, G_{i2}, \ldots, G_{in})'$, with $G_{ij} = 1$ if Y_{ij} is observed, and $G_{ij} = 0$ otherwise. For dropouts, the time at which subject i withdraws is denoted by the dropout indicator $D_i = 1 + \sum_{j=1}^{n} G_{ij}$.

For the pairwise approach, we add weights to each pair of log-pseudo-likelihood, as follows:

$$p\ell\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{N} \sum_{s \in S} \frac{G_i^{(s)}}{\pi_i^{(s)}} \log f_{\mathbf{Y}^{(s)}}(\mathbf{y}_i^{(s)}; \boldsymbol{\theta}^{(s)}), \tag{5.1}$$

where $G_i^{(s)}$ is the indicator for observing the corresponding sub-vector $\mathbf{Y}_i^{(s)}$, and $\pi_i^{(s)}$ is the probability of it being observed.

Let us define π_{ij} as the probability of being observed up to and including time j. Therefore, if the subject drops out at occasion j:

$$\pi_{ij} = p_{ij} \prod_{l=2}^{j-1} (1 - p_{il}),$$

and if the subject does not drop out at occasion j:

$$\pi_{ij} = \prod_{l=2}^{j} (1 - p_{il}),$$

where $p_{il} = P(D_i = l | D_i \geq l, \bar{\mathbf{Y}}_{i\bar{l}}, \mathbf{X}_i)$, i.e., the probability of dropping out at occasion l, given the subject is still in the study, and the outcome history, $\bar{\mathbf{Y}}_{i\bar{l}} = (Y_{i1}, \ldots, Y_{i,l-1})$. These probabilities (p_{il}) can be estimated by logistic regression.

Therefore, if the pair $\mathbf{Y}_i^{(s)} = (Y_{ij}, Y_{ik})$, for $j < k$, is fully observed (i.e., dropout occurs after time k), we have that:

$$G_i^{(s)} = 1 \text{ and } \pi_i^{(s)} = \prod_{l=2}^{k} (1 - p_{il}),$$

and in the case that only Y_{ij} is observed (i.e., dropout happens at occasion r, with $j < r < k$):

$$G_i^{(s)} = 1 \text{ and } \pi_i^{(s)} = p_{ir} \prod_{l=2}^{r-1} (1 - p_{il}),$$

otherwise, $G_i^{(s)} = 0$.

Generally, the IPW approach provides consistent estimators if the missing data model is correctly specified (Molenberghs et al., 2011). Nevertheless, instability in the estimates can be encountered when some of the calculated weights are large (Zubizarreta, 2015). This problem can be solved by truncating the weights to a maximum value.

Although it may introduce some bias, it is reasonable when this problem is due to a misspecification of the missing data model (Seaman and White, 2011).

6 Simulations

6.1 Settings

A longitudinal setting with three and five measures per subject, $n = \{3, 5\}$. The data-generating model is:

$$Y_{ij} = \beta_0 + x_j\beta_1 + x_j z_i \beta_2 + (\gamma_0 + x_j\gamma_1 + x_j z_i \gamma_2)\varepsilon_{ij},$$

for $i = 1, \ldots, N$, and $j = 1, \ldots, N$, where $x_j = (j-1)/(n-1)$ indicating the measurement time, z_i is a binary variable (generated by a Bernoulli distribution with success probability of 0.5). For the vector of error terms, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in})'$, we consider two multivariate distributions. First, a multivariate normal, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 1 & 0.50 & 0.70 \\ & 1 & 0.80 \\ & & 1 \end{pmatrix},$$

for $n = 3$, and

$$\boldsymbol{\Sigma}_5 = \begin{pmatrix} 1 & 0.70 & 0.49 & 0.34 & 0.24 \\ & 1 & 0.70 & 0.49 & 0.34 \\ & & 1 & 0.70 & 0.49 \\ & & & 1 & 0.70 \\ & & & & 1 \end{pmatrix},$$

for $n = 5$ (autorregresive (AR) structure). Second, a multivariate Student-t with four degrees of freedom, $\boldsymbol{\varepsilon}_i \sim t_4(\mathbf{0}, 1/2\boldsymbol{\Sigma}_n)$.

Therefore, the (element-wise) τ-th conditional quantile of $\mathbf{Y}_i$ is given by:

$$\begin{aligned} Q_{Y_{ij}}(\tau|x_j, z_i) =& \beta_0 + \gamma_0 F^{-1}(\tau) + x_j \left[\beta_1 + \gamma_1 F^{-1}(\tau)\right] + x_j z_i \left[\beta_2 + \gamma_2 F^{-1}(\tau)\right] \\ =& \beta_0^\tau + x_j \beta_1^\tau + x_j z_i \beta_2^\tau, \end{aligned} \tag{6.1}$$

where $F^{-1}(\cdot)$ is the inverse of the cumulative distribution of the errors. We set $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)' = (10, 5, 2)'$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_3)' = (1, 1.5, 0)'$. Moreover, we consider a sample size of $N = \{100, 200, 1000\}$, and $\tau = \{0.1, 0.25, 0.5, 0.75, 0.9\}$.

Missing data setting

In the simulations, we consider an MAR mechanism with dropouts at $j = 2, \ldots, n$. Then, the probability that Y_{ij} is missing is given by:

$$P\left(D_i = j | Y_{i,j-1}, z_i\right) = \frac{\exp(\alpha_{0j} + Y_{i,j-1}\alpha_{1j})}{1 + \exp(\alpha_{0j} + Y_{i,j-1}\alpha_{1j})}. \tag{6.2}$$

We set $\boldsymbol{\alpha}_j = (\alpha_{0j}, \alpha_{1j})'$ in order to obtain around 22% of missingness. Finally, $M = 500$ datasets are generated for each scenario.

6.2 Estimators and parameters of interest

For estimating the QR model, we implement the PWE using $\epsilon = 0.1$. This value is selected based on the evaluation of different values through extra simulations (see Section A.1 in the Supplementary Materials). Initial values are obtained using a method-of-moments estimator. The approximated (pseudo) log-likelihood is maximized by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Fletcher, 2000).

The variance of $\widehat{\boldsymbol{\beta}}^{\tau}$ is calculated using the empirically corrected estimator, sometimes also called sandwich estimator (4.3), in which the gradient vectors and Hessian matrices are computed using (central) finite-differences.

For comparison, we also implement the univariate quantile regression (UQR), i.e., without taking the correlation into account, and the MLE based on the MGH distribution (Flórez et al., 2022). Finally, for the IPW approaches, the weights are computed using a logistic model using the previously observed response and the treatment effect as covariates.

We are interested in the QR coefficients, i.e., $\boldsymbol{\beta}^{\tau}$. The rest of the parameters of the model $(\boldsymbol{\phi}, \boldsymbol{\theta}_C)$ are considered nuisance parameters. The estimators are compared using the relative bias (RB) and the root mean squared error (RMSE) defined as:

$$RB_{\text{Estimator}}\left(\widehat{\beta}_j^{\tau}\right) = \frac{1}{M}\sum_{k=1}^{M}\frac{\widehat{\beta}_{jk}^{\tau} - \beta_j^{\tau}}{\beta_j^{\tau}},$$

and

$$RMSE_{\text{Estimator}}(\widehat{\beta}_j^{\tau}) = \sqrt{\frac{1}{M}\sum_{k=1}^{M}\left(\widehat{\beta}_{jk}^{\tau} - \beta_j^{\tau}\right)^2},$$

respectively. Here, $\widehat{\boldsymbol{\beta}}_{jk}^{\tau}$ refers to the estimation of $\boldsymbol{\beta}_j^{\tau}$ in the k-th sample. Furthermore, we evaluate the coverage of the 95% confidence interval (CI) of the pairwise estimator.

6.3 Results

In this section, we present the results of the normal scenarios with $n = 3$. The rest of the cases are reserved for Section A.2 of the Supplementary Materials.

Tables 1 and 2 displays the RB and RMSE for each QR coefficient using the proposed methods for the scenarios with full and MAR normal data, respectively. Without

Table 1: Normal scenarios with $n = 3$. Relative bias of the estimators with full data (no missing values), and MAR data using available cases with/without inverse probability weighting.

			$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.90$		
data	N	est*	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ
Full	100	UQR	-0.02	0.45	-1.68	0.02	-0.10	-0.99	-0.03	-0.39	0.77	-0.05	0.01	-1.52	-0.02	-0.08	-0.41
		PWE	-1.02	-4.35	0.31	-0.03	0.14	-0.73	-0.01	-0.33	0.85	-0.06	-0.13	-0.55	0.68	1.72	-0.01
		MLE	-1.37	-7.54	0.05	-0.38	-1.67	0.02	-0.05	-0.34	0.02	0.28	0.75	0.26	0.99	2.97	0.30
	200	UQR	-0.03	0.45	-2.53	-0.00	0.13	-1.42	-0.04	-0.06	-1.15	-0.01	-0.11	-0.68	0.01	-0.08	-0.66
		PWE	-0.93	-4.19	-1.42	-0.02	0.17	-1.16	-0.06	-0.13	-0.98	-0.01	-0.12	-0.73	0.64	1.72	-0.77
		MLE	-1.42	-7.50	-1.21	-0.42	-1.49	-0.59	-0.05	-0.13	-0.38	0.28	0.89	-0.33	1.07	3.38	-0.54
	1000	UQR	0.02	-0.39	0.59	-0.01	-0.27	0.71	-0.03	-0.13	0.60	-0.01	-0.22	0.59	-0.01	-0.14	0.89
		PWE	-0.75	-3.99	0.56	-0.00	-0.12	0.55	-0.02	-0.18	0.67	-0.03	-0.27	0.51	0.59	1.56	0.53
		MLE	-1.36	-7.94	0.57	-0.36	-1.82	0.63	-0.01	-0.22	0.56	0.30	0.85	0.53	1.04	3.24	0.52
MAR	100	UQR	0.38	-14.29	-7.07	0.50	-13.24	-6.69	0.48	-12.67	-4.94	0.46	-11.24	-5.74	0.60	-11.60	-0.79
		PWE	-0.47	-7.87	-1.88	0.04	-4.80	-3.14	0.14	-4.94	-1.42	0.29	-3.67	-2.33	0.95	-1.07	-0.64
		MLE	-1.12	-8.82	0.47	-0.34	-2.11	0.19	-0.05	-0.39	-0.04	0.32	0.55	0.38	0.93	2.14	0.88
	1000	UQR	0.46	-16.26	-4.31	0.48	-14.03	-4.38	0.50	-12.59	-3.81	0.58	-11.91	-3.28	0.63	-11.27	-2.64
		PWE	-0.28	-8.28	-1.73	0.06	-5.22	-1.67	0.14	-4.62	-1.16	0.34	-3.73	-1.08	0.89	-0.51	-2.72
		MLE	-1.07	-9.71	0.84	-0.32	-2.32	0.76	-0.03	-0.13	0.58	0.37	0.79	0.35	1.00	2.47	0.08
MAR+IPW	100	UQR	-0.13	1.12	-0.19	-0.05	0.76	-0.23	-0.10	0.61	2.27	-0.10	0.04	2.66	0.19	-1.12	-5.22
		PWE	-0.11	-5.09	0.00	-0.17	-0.41	0.18	0.04	-0.22	1.53	0.17	-0.14	-0.34	0.52	0.20	-1.55
	200	UQR	-0.03	0.20	-2.08	-0.03	0.28	-1.02	-0.05	0.24	-1.82	-0.03	0.40	-1.86	0.14	-0.67	-1.91
		PWE	-0.37	-4.32	-2.46	-0.08	-0.28	-1.18	0.04	0.00	-1.77	0.18	0.23	-2.16	0.61	1.61	-3.75
	1000	UQR	0.02	-0.51	0.96	-0.01	-0.25	0.64	-0.04	-0.10	1.05	-0.01	-0.33	1.91	-0.02	-0.05	1.96
		PWE	-0.48	-3.59	0.56	-0.20	-0.27	0.35	-0.03	-0.09	0.83	0.13	0.00	0.96	0.43	1.75	0.61

* UQR: univariate quantile regression; PWE: pairwise estimator based on a Gaussian copula with $\epsilon = 0.1$, MLE: maximum likelihood estimator based on the multivariate hyperbolic distribution with $\epsilon = 0.05$

missingness, all estimators provide reasonably unbiased estimates for the parameters at each quantile level, except from a small bias of the MLE and PWE when $\tau = 0.1$ and $\tau = 0.9$. Furthermore, the PWE provides more efficient estimates in most scenarios (i.e., lowest RMSE). Nevertheless, the efficiency gets smaller as τ departs from 0.5. As expected, the PWE and UQR are biased under MAR, with a smaller bias for the former. This is more noticeable in the parameter associated with the slope (β_1^τ). However, it is corrected by using inverse probability weighting with the cost of large variability. Note that the RMSE increment is lesser for the PWE.

Table 2: Normal scenarios with $n = 3$. Square root mean squared error of the estimators with full data (no missing values), and MAR data using available cases with/without inverse probability weighting.

			$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.90$		
data	N	est*	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ
Full	100	UQR	0.16	0.57	0.81	0.13	0.42	0.64	0.11	0.41	0.62	0.12	0.44	0.67	0.16	0.57	0.81
		PWE	0.21	0.50	0.59	0.12	0.34	0.49	0.11	0.32	0.45	0.12	0.36	0.49	0.20	0.53	0.63
		MLE	0.19	0.44	0.40	0.12	0.32	0.39	0.11	0.31	0.41	0.12	0.32	0.40	0.19	0.44	0.42
	200	UQR	0.12	0.40	0.58	0.09	0.31	0.46	0.08	0.28	0.42	0.09	0.32	0.46	0.11	0.41	0.60
		PWE	0.15	0.38	0.44	0.09	0.25	0.34	0.08	0.22	0.30	0.09	0.25	0.33	0.14	0.36	0.45
		MLE	0.16	0.35	0.29	0.09	0.23	0.28	0.07	0.21	0.28	0.09	0.24	0.29	0.16	0.36	0.29
	1000	UQR	0.05	0.18	0.26	0.04	0.15	0.22	0.04	0.13	0.20	0.04	0.14	0.21	0.05	0.17	0.26
		PWE	0.09	0.20	0.20	0.04	0.12	0.16	0.03	0.10	0.15	0.04	0.12	0.16	0.09	0.19	0.20
		MLE	0.13	0.28	0.13	0.05	0.13	0.13	0.03	0.10	0.14	0.05	0.12	0.13	0.13	0.26	0.13
MAR	100	UQR	0.16	0.82	1.00	0.14	0.76	0.80	0.13	0.79	0.73	0.14	0.86	0.82	0.19	1.06	1.00
		PWE	0.17	0.61	0.76	0.12	0.47	0.61	0.11	0.46	0.56	0.13	0.51	0.64	0.21	0.68	0.83
		MLE	0.18	0.52	0.53	0.12	0.39	0.52	0.11	0.38	0.54	0.12	0.43	0.54	0.19	0.56	0.57
	200	UQR	0.12	0.70	0.73	0.10	0.68	0.58	0.10	0.71	0.52	0.11	0.79	0.54	0.14	0.90	0.70
		PWE	0.12	0.47	0.56	0.09	0.37	0.44	0.08	0.35	0.40	0.10	0.37	0.42	0.17	0.46	0.58
		MLE	0.14	0.43	0.39	0.09	0.28	0.37	0.08	0.26	0.37	0.09	0.30	0.37	0.16	0.41	0.39
	1000	UQR	0.07	0.55	0.34	0.06	0.59	0.27	0.06	0.65	0.25	0.08	0.74	0.26	0.09	0.81	0.30
		PWE	0.06	0.31	0.25	0.04	0.25	0.20	0.04	0.26	0.19	0.05	0.27	0.20	0.11	0.20	0.26
		MLE	0.10	0.33	0.18	0.05	0.15	0.18	0.03	0.12	0.18	0.06	0.15	0.18	0.12	0.24	0.17
MAR+IPW	100	UQR	0.16	0.63	0.91	0.13	0.53	0.82	0.13	0.61	0.95	0.14	0.83	1.32	0.19	1.12	1.59
		PWE	0.31	0.76	0.97	0.25	0.62	0.78	0.16	0.43	0.67	0.16	0.58	0.88	0.21	0.79	1.10
	200	UQR	0.12	0.45	0.64	0.09	0.39	0.57	0.09	0.39	0.60	0.10	0.55	0.90	0.14	0.81	1.33
		PWE	0.23	0.48	0.60	0.15	0.33	0.46	0.11	0.29	0.44	0.11	0.37	0.58	0.16	0.62	0.87
	1000	UQR	0.05	0.20	0.30	0.04	0.17	0.27	0.04	0.18	0.28	0.05	0.23	0.38	0.06	0.35	0.64
		PWE	0.11	0.24	0.26	0.07	0.15	0.21	0.05	0.14	0.21	0.05	0.17	0.26	0.07	0.29	0.43

* UQR: univariate quantile regression; PWE: pairwise estimator based on a Gaussian copula with $\epsilon = 0.1$, MLE: maximum likelihood estimator based on the multivariate hyperbolic distribution with $\epsilon = 0.05$.

The coverage of the 95% confidence intervals for the PWE is shown in Table 3. In most cases, the CI coverage is around the nominal value, indicating that the variance matrix of $\widehat{\boldsymbol{\beta}}^\tau$ is properly estimated. With $\tau = 0.1$ and $\tau = 0.9$, there is a noticeable lower coverage. However, it gets larger as the sample sizes increases. As with the bias, the IPW approach corrects the CI coverage (showing similar results to the ones obtained with full data).

Table 3: Normal scenarios with $n = 3$. 95% confidence interval coverage of the pairwise estimator with full data (no missing values), and MAR data using available cases with/without inverse probability weighting.

		$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.90$		
data	N	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ	β_0^τ	β_1^τ	β_2^τ
FULL	100	0.76	0.78	0.87	0.91	0.90	0.93	0.93	0.91	0.95	0.88	0.89	0.92	0.76	0.76	0.86
	200	0.80	0.82	0.90	0.93	0.92	0.95	0.93	0.94	0.96	0.92	0.92	0.94	0.81	0.83	0.88
	1000	0.77	0.86	0.94	0.94	0.93	0.95	0.94	0.94	0.93	0.93	0.95	0.95	0.78	0.89	0.94
MAR	100	0.79	0.75	0.82	0.90	0.84	0.92	0.92	0.83	0.92	0.90	0.82	0.91	0.74	0.69	0.78
	200	0.87	0.78	0.89	0.92	0.83	0.92	0.93	0.83	0.93	0.91	0.82	0.94	0.75	0.81	0.88
	1000	0.90	0.68	0.94	0.94	0.64	0.94	0.94	0.54	0.93	0.84	0.63	0.94	0.56	0.90	0.93
MAR+IPW	100	0.58	0.75	0.81	0.81	0.89	0.92	0.89	0.92	0.92	0.88	0.83	0.86	0.78	0.60	0.64
	200	0.70	0.86	0.88	0.90	0.92	0.93	0.93	0.93	0.94	0.91	0.88	0.91	0.85	0.73	0.75
	1000	0.90	0.89	0.94	0.94	0.94	0.95	0.96	0.92	0.94	0.95	0.93	0.95	0.88	0.89	0.92

The computational complexity is based on CPU time needed for estimating the parameters and obtaining standard errors. Table 4 displays the median time (in seconds) for the MLE and PWE for each scenario. Even with the largest number of individuals, the median computing time of the PWE is less than five seconds. Furthermore, it is faster than MLE in all scenarios. Of course, computational performance is machine-dependable. These times were taken from a desktop computer with an AMD Ryzen(TM)-7 3700X 8-core processor at 3.59GHz and 16GB of RAM. Nevertheless, these results show that the algorithms are computationally affordable. Furthermore, with Student-t data, the PWE exhibits similar performance to the ones presented above.

Extra simulations (Student-t errors and five observations per subject) display similar results. The PWE shows good performance compared with the UQR and MLE. Furthermore, it is still computationally efficient (see Section A.2 and A.3 in the Supplementary Materials).

Table 4: Normal scenarios. Median computation time (in seconds) for fitting the quantile regression model using the pairwise estimator based on a Gaussian copula with $\epsilon = 0.1$.

		FULL data					MAR data					MAR data + IPW				
		τ					τ					τ				
N	est*	0.10	0.25	0.50	0.75	0.90	0.10	0.25	0.50	0.75	0.90	0.10	0.25	0.50	0.75	0.90
100	PWE	0.56	0.53	0.51	0.53	0.56	0.58	0.54	0.52	0.55	0.69	0.57	0.54	0.53	0.55	0.58
	MLE	1.02	0.78	0.75	0.77	0.92	1.34	0.97	0.95	0.97	1.16					
200	PWE	0.94	0.87	0.87	0.88	0.94	0.93	0.90	0.89	0.91	0.96	0.98	0.94	0.91	0.94	1.20
	MLE	1.70	1.35	1.33	1.35	1.55	2.22	1.66	1.61	1.66	2.03					
1000	PWE	4.08	3.88	3.96	3.82	4.00	3.99	4.12	4.09	3.97	4.11	4.28	3.93	4.00	4.00	4.99
	MLE	7.36	6.22	6.05	6.07	6.59	9.20	7.20	7.14	7.21	7.68					

PWE: pairwise estimator based on a Gaussian copula with $\epsilon = 0.1$, MLE: maximum likelihood estimator based on the multivariate hyperbolic distribution with $\epsilon = 0.05$.

7 A trial in opthalmology

Age-related macular degeneration (ARMD) is a medical condition in which individuals progressively lose sight. The ARMD study is a randomized multicenter double-blinded trial aiming to compare an experimental treatment (interferon-α) to a corresponding placebo for treating patients with ARMD (more details in Pharmacological Therapy for Macular Degeneration Study Group, 1997). The dataset, available in the R package `nlmeU` (Galecki and Burzykowski, 2013), contains information of 240 individuals in two of the groups: placebo and the highest dose (6 million units daily) of interferon-α. This dataset has been widely analyzed under hierarchical modeling (Gałecki and Burzykowski, 2013) and surrogate marker evaluation (Buyse et al., 2000).

The primary outcome is visual acuity, measured as the ability to read lines of letters on standardized vision charts, observed at five different time points (baseline, week

4, week 12, week 24, and week 52). However, some individuals were unavailable for follow-up at all measurement times. In particular, 78% of the individuals have full observations, 19% corresponds to patients that dropout at different time points, and 3% to subjects with intermittent missing values. For our analysis, the latter group is removed from the dataset.

Let Y_{ij} be the visual acuity loss measured at patient i in week t_j, for $t_j = \{0, 4, 12, 24, 52\}$. We are interested in fitting the following quantile model:

$$Q_{Y_{ij}}(\tau|T_i, t_j) = \beta_0^\tau + T_i\beta_1^\tau + t_j\beta_2^\tau + T_i t_j \beta_3^\tau, \tag{7.1}$$

where $T_i = 1$ if patient i is in the treatment arm, and $T_i = 0$ otherwise.

For the QR-based IPW approach, we propose the following dropout model:

$$\begin{aligned}\text{logit}\left[P\left(D_i = j|D_i \geq j, \bar{\mathbf{Y}}_{i\bar{j}}, T_i, t_j\right)\right] =& \alpha_0 + Y_{i,j-1}\alpha_1 + T_i\alpha_2 + t_j\alpha_3 + \\ & L_{1i}\alpha_4 + L_{2i}\alpha_5 + L_{3i}\alpha_6,\end{aligned}$$

where $Y_{i,j-1}$ is the outcome at the previous time (t_{j-1}), and L_{1i}, L_{2i}, and L_{3i} are the three dummy variables associated to the level of lesion (classic, occult, recurrent, or hemorrhage) at baseline.

For fitting model (7.1), we consider the PWE based on Gaussian copulas (setting $\epsilon = 0.1$). The value of ϵ was chosen based on a sensitivity analysis checking the estimates, standard error, and gradients at the solution, with different values of ϵ (see Section B of the Supplementary Materials). For comparison, the estimates based on the UQR are also included. For this method, the standard errors were computed using a clustered bootstrap (Hagemann, 2017).

Table 5 displays the parameter estimates of (7.1) for different values of τ using the PWE. Furthermore, results by UQR are also presented for comparison. Comparing

UQR-IPW and PWEs-IWP, the estimates of β_0^τ and β_2^τ are larger for the latter as τ increases. For the remaining parameters, there are no substantial differences. The treatment effect ($\beta^\tau{}_3^\tau$) is not significant for any τ. Furthermore, there is a significant negative time effect for all quantile levels. These results are in line with the previous analysis using hierarchical models. Nevertheless, visual acuity deteriorates differently depending on the current eyesight. It is faster for patients with less vision ($\widehat{\beta}_1^\tau$ decreases with τ). Note that β_1^τ is not significant for $\tau = 0.9$, showing that eyesight degradation does not happen to the 10% of the patients with most acute sight.

Table 5: ARMD data. Parameter estimates, their standard error (se), and significance for the quantile regression with $\tau = \{0.1, 0.25, 0.5, 0.75, 0.9\}$ using available cases with inverse probability weighting.

univariate quantile regression with inverse probability weighting																
		$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$			$\tau = 0.90$		
effect	parm	est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value
intercept	β_0	45.00	2.20	0.0000	45.00	2.28	0.0000	55.14	1.79	0.0000	65.50	1.74	0.0000	75.20	2.55	0.0000
t	β_1	-0.33	0.07	0.0000	-0.33	0.07	0.0000	-0.21	0.05	0.0000	-0.12	0.05	0.0082	-0.10	0.05	0.0360
treat	β_2	-3.00	3.40	0.3772	-3.00	3.52	0.3943	-1.14	2.91	0.6942	-0.50	2.42	0.8365	-3.91	3.35	0.2435
t×treat	β_3	-0.06	0.10	0.5486	-0.06	0.10	0.5677	-0.11	0.08	0.1798	-0.07	0.07	0.3121	-0.08	0.09	0.4060
pairwise estimator with inverse probability weighting																
		$\tau = 0.10$			$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$			$\tau = 0.90$		
effect	parm	est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value	est	se	p-value
intercept	β_0	37.19	5.50	0.0000	47.28	5.25	0.0000	60.09	8.30	0.0000	70.91	6.49	0.0000	80.16	0.40	0.0000
t	β_1	-0.34	0.11	0.0029	-0.32	0.07	0.0000	-0.23	0.06	0.0004	-0.13	0.04	0.0011	-0.00	0.19	0.9851
treat	β_2	1.08	5.83	0.8527	-2.75	5.38	0.6087	-3.60	8.29	0.6642	-8.31	6.84	0.2248	-9.77	9.02	0.2785
t×treat	β_3	-0.19	0.12	0.1181	-0.06	0.09	0.4729	-0.13	0.08	0.1131	-0.05	0.06	0.4202	-0.12	0.19	0.5130

UQR-IPW: univariate quantile regression using inverse probability weighting; PWE: pairwise estimator based on a Gaussian copula with $\epsilon = 0.1$ and using inverse probability weighting.

The QR estimates for the different τ levels using the proposed techniques are shown graphically in Figure 1. Here, we can see that all methods provide similar estimates for the quantiles for both treatment and placebo arms. Nonetheless, the PWE-IPW provides consistently larger estimates for the visual acuity than UQR, for all quantile

levels. This is more noticeable for $\tau = 0.9$ in which more differences in the estimates of β_0^τ and β_1^τ are encountered. Furthermore, in simulation results, the PWE provides somewhat biased estimates for the intercept when τ is close to one.

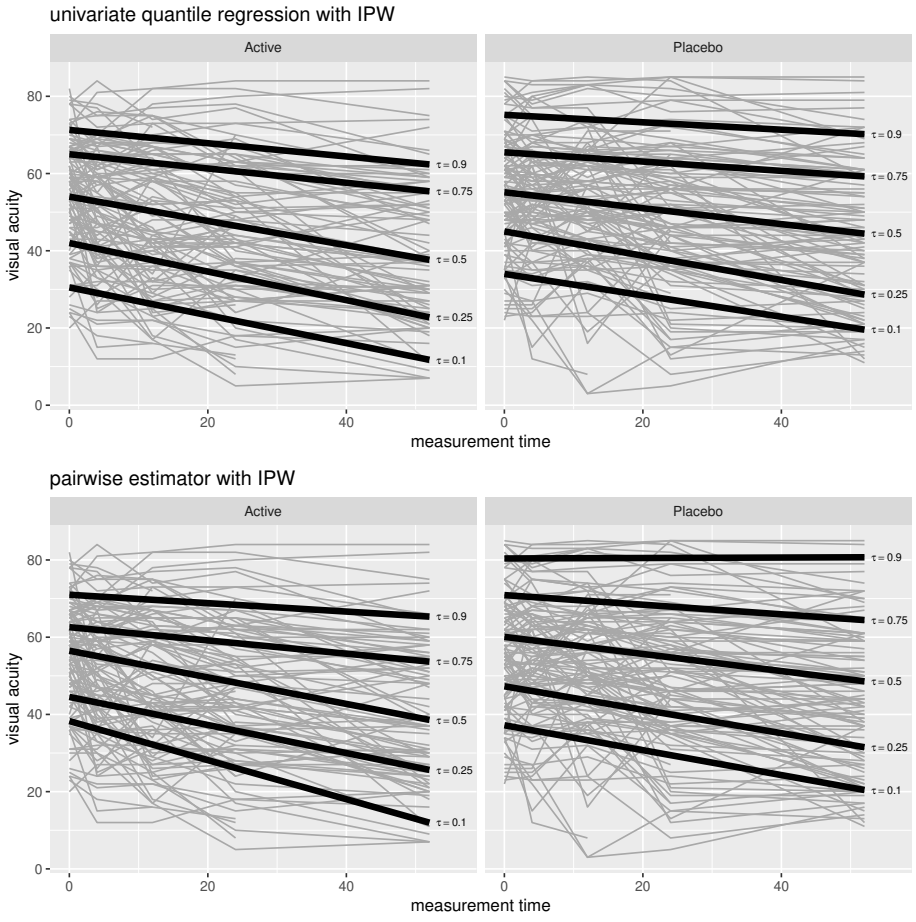

Figure 1: ARMD data. Quantile regression estimates for $\tau = \{0.1, 0.25, 0.5, 0.75, 0.9\}$ using univariate quantile regression with IPW (upper plots) and pairwise estimator with IPW (lower plots). Light-grey lines represent a patients' visual acuity trajectories.

8 Final remarks

We proposed a QR model for hierarchical data in the presence of dropouts. Our proposal fits the marginal quantiles using an Asymmetric Laplace distribution, and a copula is used for modeling the dependence structure. A pairwise estimator (PWE) based on pseudo-likelihood (PL) is introduced to reduce computational complexity. Both copula modeling and pairwise estimating are flexible methodologies that permit fitting QR models for clustered data with different types of dependence structures. Since the PWE is not a likelihood-based estimator, it can be biased when the missing data mechanism is missing at random. Therefore, for bias correction, an inverse probability weighting (IPW) methodology is implemented.

Based on simulations, the PWE is asymptotically unbiased and efficient. In some cases, it shows slightly better results than the maximum likelihood estimator (MLE) based on the multivariate generalized hyperbolic (MGH) distribution, especially when τ is close to the parameter space boundaries. This can be explained by two factors. Firstly, the MGH's correlation structure is more restricted in these cases (Flórez et al., 2022). On the other hand, modeling through copulas does not impose any restriction on the correlations, thus making it more flexible. Secondly, PL techniques are robust against model misspecification. Contrary to MLEs, they only require pairwise marginal distributions and not the correct specification of the whole multivariate density function. The smooth approximation of the AL log-likelihood provides an adequate solution for the non-differentiable objective function with the cost of small bias. Finally, the IPW method successfully corrects the bias due to missingness at the cost of increased variance. Note that this increment is more noticeable for the univariate quantile regression. In terms of computation speed, the copula-based PWE

is fast.

In this paper, we allow for dropouts only. The IPW approach can be extended to non-monotone missing patterns. However, the calculation of the weights is not trivial. Another alternative is using multiple imputation (MI), making the data monotone, but without imputing values after dropout. Once this has been done, the IPW method can be employed. Of course, Rubin's rules then have to be applied to combine the multiple inferences into a single one. Although most MI approaches specify a distribution for the missing data, e.g., Gaussian, they are robust to model misspecification. Furthermore, some imputation models are also QR-based (Bottai and Zhen, 2013; Geraci, 2016).

While our approach is valid under MAR, it is evident that the operation of an MNAR mechanism cannot be excluded (Molenberghs et al., 2008). This naturally leads to the concept of sensitivity analysis, in which the impact of the non-verifiable assumptions about the missing data mechanism on inferences, often by considering a range of options under MNAR, are explored. A variety of sensitivity analysis tools are described in Molenberghs and Kenward (2007) and Molenberghs et al. (2014). Practically, it is convenient to apply multiple imputation under a variety of MNAR scenarios, as described in, for example, Carpenter et al. (2013). Their scenarios are formulated in a pattern-mixture format. Once data are imputed, the methods described in this paper can be applied without any problem, to every one of the imputed datasets. Rubin's rules can then be applied to combine the multiple inferences into a single one.

Finally, our proposal is susceptible to the non-monotonicity or crossing problem in finite samples. By definition, conditional QR should be an increasing function in τ. However, with small samples, an estimated conditional quantile at some τ may be

smaller than at a lower quantile. Some techniques available in the literature can be implemented with our method to prevent this crossing from happening. A two-step procedure can be proposed by firstly estimating the QR for each τ, and secondly, monotonizing the quantile curves (Mammen, 1991). Following Wu and Liu (2009), a stepwise estimator can be implemented by starting from a particular quantile and moving on to estimating the higher (or lower) order imposing constraints to ensure non-crossing. Other alternatives can be found in Chernozhukov et al. (2009), Schnabel and Eilers (2013) and Andriyana et al. (2018).

Supplementary Materials

Additional results are exhibited in the Supplementary Materials available online. Section A exhibits additional simulation results about the effect of smoothing approximations of the checked function different distributions of the error term and number of observations per subject. Section B shows a sensitivity analysis of the impact of ϵ on the PWE using the ARMD data.

References

Abrevaya, J. and Dahl, C. M. (2008). The effects of birth inputs on birthweight. *Journal of Business & Economic Statistics*, **26**(4), 379–397.

Aerts, M., Molenberghs, G., Geys, H., and Ryan, L. (2002). *Topics in Modelling of Clustered Data*. Chapman & Hall.

Andriyana, Y., Gijbels, I., and Verhasselt, A. (2018). Quantile regression in varying-coefficient models: non-crossing quantile curves and heteroscedasticity. *Statistical Papers*, **59**(4), 1589–1621.

Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood estimation: Some examples. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, **53**(2), 233–243.

Bottai, M. and Zhen, H. (2013). Multiple imputation based on conditional quantile estimation. *Epidemiology, Biostatistics and Public Health*, **10**(1).

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., and Geys, H. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, **1**(1), 49–67.

Carpenter, J. R., Roger, J. H., and Kenward, M. G. (2013). Analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics*, **23** (6), 1352–1371.

Carpenter, J. and Kenward, M. (2013). *Multiple Imputation and Its Application*. John Wiley & Sons.

Chen, C. (2007). A finite smoothing algorithm for quantile regression. *Journal of Computational and Graphical Statistics*, **16**(1), 136–164.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, **96**(3), 559–575.

De Backer, M., El Ghouch, A., and Van Keilegom, I. (2017). Semiparametric copula quantile regression for complete or censored data. *Electronic Journal of Statistics*, **11**(1).

Farcomeni, A. and Viviani, S. (2015). Longitudinal quantile regression in the presence of informative dropout through longitudinal–survival joint modeling. *Statistics in Medicine*, **34**(7), 1199–1213.

Fitzmaurice, G. M. (2015). Semiparametric methods: Introduction and overview. In *Handbook of Missing data*, chapter 7, pages 139–148. Chapman and Hall/CRC, Boca Raton, FL.

Fletcher, R. (2000). *Practical Methods of Optimization*. John Wiley & Sons, 2 edition.

Flórez, A. J., Van Keilegom, I., Molenberghs, G., and Verhasselt, A. (2022). Quantile regression for longitudinal data via the multivariate generalized hyperbolic distribution. *Statistical Modelling*, **22**(6), 566–584.

Fu, L. and Wang, Y.-G. (2012). Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, **56**(8), 2526–2538.

Furno, M. and Vistocco, D. (2018). *Quantile Regression: Estimation and Simulation*, volume 2. Wiley.

Galecki, A. and Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*. Springer, New York, first edition.

Gałecki, A. and Burzykowski, T. (2013). *Linear Mixed-Effects Models Using R*. Springer New York.

Galvao, A. and Kato, K. (2017). Quantile regression methods for longitudinal data. In *Handbook of Quantile Regression*, chapter 19, pages 369–380. Chapman and Hall/CRC, Boca Raton, FL.

Geraci, M. and Bottai, M. (2006). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**(1), 140–154.

Geraci, M. (2016). Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants. *Statistical Methods in Medical Research*, **25**(4), 1393–1421.

Geraci, M. (2019). Modelling and estimation of nonlinear quantile regression with clustered data. *Computational Statistics & Data Analysis*, **136**, 30–46.

Geraci, M. and Bottai, M. (2013). Linear quantile mixed models. *Statistics and Computing*, **24**(3), 461–479.

Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association*, **112**, 446–456.

He, X., Fu, B., and Fung, W. K. (2003). Median regression for longitudinal data. *Statistics in Medicine*, **22**(23), 3655–3669.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics*, **1**(5).

Hunter, D. R. and Lange, K. (2000). Quantile regression via an MM algorithm. *Journal of Computational and Graphical Statistics*, **9**(1), 60.

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London New York.

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**(1), 33.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**(448), 1296–1310.

Koenker, R. W. and D'Orey, V. (1987). Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **36** (3), 383–393.

Kotz, S. (2001). *The Laplace distribution and generalizations : a revisit with applications to communications, economics, engineering, and finance*. Birkhäuser, Boston.

Little, R. J., Carpenter, J. R., and Lee, K. J. (2022). A comparison of three popular methods for handling missing data: Complete-case analysis, inverse probability weighting, and multiple imputation. *Sociological Methods & Research*.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc.

Mammen, E. (1991). Estimating a smooth monotone regression function. *The Annals of Statistics*, **19**(2), 724–740.

Marino, M. F., Tzavidis, N., and Alfò, M. (2018). Mixed hidden markov quantile regression models for longitudinal data with possibly incomplete sequences. *Statistical Methods in Medical Research*, **27**(7), 2231–2246.

Marino, M. F. and Farcomeni, A. (2015). Linear quantile regression models for longitudinal experiments: an overview. *METRON*, **73**(2), 229–247.

Merlo, L., Maruotti, A., and Petrella, L. (2022). Two-part quantile regression models for semi-continuous longitudinal data: A finite mixture approach. *Statistical Modelling*, **22**(6), 485–508.

Molenberghs, G. and Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Wiley.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **70**(2), 371–388.

Molenberghs, G., Kenward, M. G., Verbeke, G., and Birhanu, T. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica*, **21**(1), 187–206.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., and Verbeke, G., editors (2014). *Handbook of Missing Data Methodology*. Chapman and Hall/CRC.

Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer-Verlag GmbH.

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.

Pharmacological Therapy for Macular Degeneration Study Group (1997). Interferon alfa-2a is ineffective for patients with choroidal neovascularization secondary to age-related macular degeneration. *Archives of Ophthalmology*, **115**(7), 865–870.

Portnoy, S. and Koenker, R. (1997). The gaussian hare and the laplacian tortoise: Computability of squared- error versus absolute-error estimators. *Statistical Science*, **12**(4), 279–296.

Pourahmadi, M. and Wang, X. (2015). Distribution of random correlation matrices: Hyperspherical parameterization of the cholesky factor. *Statistics & Probability Letters*, **106**, 5–12.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**(429), 106–121.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics, Theory and Methods*, **22** (4), 965–984.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**(3), 581–592.

Rubin, D. B., editor (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc.

Schnabel, S. K. and Eilers, P. H. C. (2013). Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, **97**(1), 77–87.

Seaman, S. R. and White, I. R. (2011). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, **22**(3), 278–295.

van Buuren, S. (2012). *Flexible imputation of missing data*. CRC Press, Boca Raton.

Wang, H. J., Feng, X., and Dong, C. (2019). Copula-based quantile regression for longitudinal data. *Statistica Sinica*, **29**, 245–264.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

Wu, Y. and Liu, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and Its Interface Volume*, **2**(3), 299–310.

Yi, G. Y. and He, W. (2008). Median regression models for longitudinal data with dropouts. *Biometrics*, **65**(2), 618–625.

Yu, A., Zhong, Y., Feng, X., and Wei, Y. (2023). Quantile regression for nonignorable missing data with its application of analyzing electronic medical records. *Biometrics*, **79**(3), 2036–2049.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, **54**(4), 437–447.

Zubizarreta, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, **110** (511), 910–922.