Marginal M-quantile regression for multivariate dependent data

Luca Merlo[1], Lea Petrella[2], Nicola Salvati[3], and Nikos Tzavidis[4]

[1]Department of Statistics, Sapienza University of Rome
[2]MEMOTEF Department, Sapienza University of Rome
[3]Department of Economics and Management, University of Pisa
[4]Department of Social Statistics and Demography, Southampton Statistical Sciences Research Institute, University of Southampton

December 31, 2021

Abstract

This paper develops an M-quantile regression model for the analysis of multiple dependent outcomes by introducing the notion of directional M-quantiles for multivariate responses. In order to incorporate the correlation structure of the data into the estimation framework, we propose a robust marginal M-quantile model extending the well-known generalized estimating equations approach to the case of regression M-quantiles with Huber's loss function. We discuss the estimation of the model and derive the asymptotic properties of estimators. In addition, we introduce the idea of M-quantile contours that can be used to describe the dependence between the response variables and to investigate the effect of covariates on the location, spread and shape of the distribution of the responses. To examine their variability, we build confidence envelopes via nonparametric bootstrap. The validity of the proposed methodology is explored both by means of simulation studies and through an application to educational data.

Keywords: Asymptotic properties, Correlated data, Directional M-quantile, Generalized M-Quantile Estimating Equations, M-quantile contour

1 Introduction

Quantile regression has attracted considerable interest in many empirical studies since its introduction in the seminal paper of Koenker & Bassett Jr (1978). It provides a way to model the conditional quantiles of a response as a function of explanatory variables in order to have a more complete picture of the entire conditional distribution compared to the classical mean regression. For this reason quantile regression methods have become widely used in the literature especially in those situations where skewness, heavy-tails, outliers, truncation, censoring

and heteroscedasticity arise. For a detailed review and list of references, Koenker (2005) and Koenker et al. (2017) provide an overview of the most used quantile regression techniques.
In real data applications, observations are often correlated with each other across time, space, or other dimensions, like groups, and their analysis requires specific data analysis tools which have received considerable attention over the years (Diggle et al. 2002, Molenberghs & Verbeke 2006, Fitzmaurice et al. 2012, Goldstein 2011). In particular, dependency of observations may be seen as a clustering effect (Bergsma et al. 2009) which arises in a number of sampling designs, including clustered, multilevel, spatial, and repeated measures (Heagerty et al. 2000, Bergsma et al. 2009, Geraci & Bottai 2014). In this context, quantile methods for modeling dependent-type data have been considered in a wide range of different applications spanning from medicine (Smith et al. 2015, Farcomeni 2012, Alfò et al. 2017, Marino et al. 2018, Merlo, Maruotti & Petrella 2021), social inequality (Heise & Kotsadam 2015), economics (Bassett & Chen 2002, Kozumi & Kobayashi 2011, Bernardi et al. 2015, 2018, Giovannetti et al. 2018, Merlo, Petrella & Raponi 2021), environmental modeling (Hendricks & Koenker 1992, Pandey & Nguyen 1999, Reich et al. 2011) and education (Kelcey et al. 2019).
When the interest of the research is on the entire conditional distribution, in addition to the classical quantile regression, a possible alternative approach is to consider the M-quantile regression proposed by Breckling & Chambers (1988). This method provides a "quantile-like" generalization of the mean regression based on influence functions, combining in a common framework the robustness and efficiency properties of quantiles and expectiles (Newey & Powell 1987), respectively. In fact, M-quantiles extend the ideas of M-estimation of Huber (1964) and Huber & Ronchetti (2009) by introducing a class of asymmetric influence functions to model the entire conditional distribution of the response given the covariates. Depending on the type of influence function used, M-quantiles may reduce to standard quantiles or expectiles. Although M-quantiles have a less intuitive interpretation than standard quantiles (Jones 1994), they offer additional substantial benefits. More precisely, they allow for robust estimation in the presence of influential data and they can trade robustness for efficiency. From a computational perspective M-quantile regression ensures uniqueness of the Maximum Likelihood solutions, and it offers greater stability as a wide range of continuous influence functions can be employed (see Tzavidis et al. 2016 and Bianchi et al. 2018). The most frequently used function is the popular Huber loss (Huber 1964) which utilises a tuning constant that can adjust the robustness of the estimator in the presence of outliers and it is, henceforth, assumed throughout our paper.
In the literature, M-quantiles have been implemented in a broad range of disciplines spanning from multilevel modeling (Tzavidis et al. 2016, Alfò et al. 2017), small area estimation (Chambers & Tzavidis 2006, Chambers et al. 2014, Salvati et al. 2021), poverty mapping (Tzavidis et al. 2008) and longitudinal studies (Alfò et al. 2017, Borgoni et al. 2018, Alfò et al. 2021).
Most of those proposals are, however, designed for a univariate framework. When the purpose of the matter being investigated lies in describing the distribution of a multivariate response, since there does not exist a natural ordering in a p-dimensional space, $p > 1$, the univariate notion of M-quantile does not straightforwardly extend to higher dimensions. Originally, Breckling & Chambers (1988) addressed the problem of defining a multivariate M-quantile by introducing a direction vector in the Euclidean p-dimensional space to establish a suitable ordering procedure for multivariate observations. The multivariate M-quantile along a specified direction is then obtained by minimizing a multidimensional Huber loss function (Huber & Ronchetti 1981). Subsequently, Kokic et al. (2002) generalized their definition by introducing a class of multivariate M-quantiles based on weighted estimating equations. More recently,

Alfò et al. (2021) proposed an M-quantile regression for multivariate longitudinal data where, however, they sidestep the problem of defining a multivariate M-quantile. The authors consider, in fact, univariate M-quantile regression models with specific random effects for each outcome and dependence between outcomes is introduced by assuming that the random effects in the univariate models are dependent.

In the present paper we approach the problem of M-quantile regression for the analysis of multivariate dependent structured data. We rely on the notion of directional quantile proposed by Kong & Mizera (2012) which consider the quantiles of projections of random vectors onto unit norm directions. We extend their approach to the M-quantile framework by using the Huber's influence function in Huber (1964). In this context, directional M-quantiles, obtained from the projection of the original data onto the real line along a specified direction, inherit robustness properties of standard univariate M-quantiles where the corresponding direction assigns a relative weight to each marginal of the response involved in the regression problem. The main advantage of the projection-based definition is that it allows for a solution to an easier problem than the multivariate one but, at the same time, it condenses valuable information about the dependence embedded in multivariate data. The validity of this directional approach is also proven by the continuously growing literature on the subject (see Hallin et al. 2010, Paindaveine & Šiman 2011, Kong & Mizera 2012, Geraci et al. 2020, Farcomeni et al. 2020 and Cascos & Ochoa 2021).

In order to estimate directional M-quantiles as function of the covariates while capturing within cluster correlations, we develop a Marginal M-quantile (MMQ) model. The marginal approach refers to a general class of statistical methods that are used to model dependent data where observations within a cluster are correlated with each other (Liang & Zeger 1986, Lindsey 1999, Heagerty et al. 2000, Diggle et al. 2002, Goldstein 2011). When fitting marginal models, the interest focuses on the relationship between the response and explanatory variables while, at the same time, acknowledging dependencies in the data. A popular estimation procedure for estimating the marginal model parameters is the Generalized Estimating Equations (GEE) approach introduced by Liang & Zeger (1986) and Zeger & Liang (1986). Because the true correlation structure is unknown, the GEE formulates a "working covariance matrix" to capture the dependence between observations and incorporate that structure into the model. This method provides consistent estimates of the regression coefficients in the presence of misspecification of the postulated correlation matrix (Zeger et al. 1988) and has been adapted to quantile regression by Fu & Wang (2012) and Lu & Fan (2015). Related literature on the use of quantile regression and marginal models includes Lipsitz et al. (1997), Yang et al. (2017), Zhao et al. (2020) and Lin et al. (2020), for example.
In our paper we introduce a generalization of the GEE approach of Liang & Zeger (1986) by using the Huber's loss function. We define a new robust estimator based on the Generalized M-quantile Estimating Equations (GMQEE) and establish its asymptotic properties using the Bahadur representation (Bahadur 1966). The proposed method is robust to influential observations in the data and improves the estimation efficiency by taking into account the correlation between linear combinations of the outcomes within each cluster.
Moreover, when theoretically all directions are investigated simultaneously, the proposed directional M-quantiles generate centrality regions and contours which allow us to assess the effect of covariates on the location, spread and shape of the entire distribution of the responses. In this case, M-quantile contours are represented by contour lines with constant quantile level dividing the responses in two groups. In particular, the points that lie outside can be classified as jointly abnormal compared to those that fall within the contour, condi-

tional on the covariates. M-quantile contours adapt to the shape of the distribution of interest and summarize the information carried by directional M-quantiles describing the dependence between the responses and specific features of multivariate data. To analyse their shapes and study the sampling variability of the M-quantile estimator of the contours, we explore the use of a bootstrap approach to build confidence envelopes.

Using simulations, we illustrate the finite sample performance and the improvement in the estimation efficiency under the approach introduced compared to the case where clustering is ignored, and study the behaviour of the proposed robust estimator in the presence of outliers. From an empirical standpoint, we exploit the proposed MMQ regression model to analyse the Tennessee's Student/Teacher Achievement Ratio (STAR) experiment (see Word et al. 1990 and Finn & Achilles 1990). Educational data often have a natural dependency structure, namely pupils are nested within schools, which induces correlation between students belonging to the same school. We develop a MMQ regression to jointly model students' mathematics and reading scores as a function of classroom size and teacher's experience in kindergarten. This model might be of great interest since it allows us to investigate the potentially differential impact of covariates on the joint distribution of the response variables.

The rest of the paper is organized as follows. In Section 2 we introduce the main notation and briefly review the M-quantile regression model. Section 3 describes the proposed methodology and Section 4 discusses the estimation procedure and develops the asymptotic theory. Section 5 presents the simulation study and the results. The application is presented in Section 6 and Section 7 presents some concluding remarks.

2 Preliminaries on M-quantile regression

M-quantile regression is a "quantile-like" generalization of regression based on influence functions. It extends the ideas of M-regression (Huber 1964) to model the relationship between the dependent variable and its predictors at different parts of the conditional distribution. In particular, this method provides a procedure which can be varied smoothly so as to capture the effect of explanatory variables on the response either in the center of the sample or in the tails by using continuous influence functions (Breckling & Chambers 1988).

Formally, the M-quantile of order $\tau \in (0,1)$ of a continuous scalar response Y given the k-dimensional vector of covariates $\mathbf{X} = \mathbf{x}$, is defined as the solution $\theta_{\mathbf{x}}(\tau)$ of the following estimating equation:

$$\int \psi_\tau(y - \theta_{\mathbf{x}}(\tau)) dF_{Y|\mathbf{X}}(y \mid \mathbf{x}) = 0, \tag{1}$$

where $F_{Y|\mathbf{X}}(\cdot \mid \mathbf{x})$ is the conditional distribution function of Y, $\psi_\tau(u) = \mid \tau - \mathbf{1}_{(u<0)} \mid \psi(u/\sigma_\tau)$ with $\psi(\cdot)$ being the first derivative of a convex loss function $\rho(\cdot)$ and σ_τ is a suitable scale parameter.

In a regression framework, for a given τ and $\psi(\cdot)$, a linear M-quantile regression model is defined as follows:

$$\theta_{\mathbf{x}}(\tau) = \mathbf{x}'\boldsymbol{\beta}(\tau), \tag{2}$$

where $\boldsymbol{\beta}(\tau)$ is the k-dimensional regression parameter vector.

In this work, the influence function $\psi(\cdot)$ in (1) is chosen to be the well-known Huber influence function (Huber 1964):

$$\psi(u) = u\mathbf{1}_{(|u| \le c)} + c\,\mathrm{sign}(u)\mathbf{1}_{(|u|>c)}, \tag{3}$$

where c denotes a tuning constant bounded away from zero. The function in (3) down weights residuals exceeding the selected value of c and remains bounded to ensure that $\theta_{\mathbf{x}}(\tau)$ will not be distorted by arbitrarily large observations. The use of the Huber influence function is chosen for several reasons. The tuning constant c can be used to trade robustness for efficiency with increasing robustness when c is chosen to be positive and close to 0 and increasing efficiency when c is chosen to be large and positive. If $c \to 0$, $\psi(u) = \text{sign}(u)$, one obtains the quantile regression (Koenker & Bassett Jr 1978); on the other hand, if $c \to \infty$, $\psi(u) = u$, M-quantile regression reduces to expectile regression (Newey & Powell 1987). Secondly, as described in Street et al. (1988), the regression parameters $\boldsymbol{\beta}(\tau)$ in (2) can be estimated by Iterative Reweighted Least Squares (IRLS) or using the Newton-Raphson algorithm developed in Bianchi et al. (2018). In contrast to algorithms used for fitting quantile regression models, the use of a continuous monotone influence function, as it is the case for the Huber function, guarantees convergence to a unique solution (Kokic et al. 1997). Proofs of consistency, asymptotic normality and estimators of the variance of the M-quantile regression coefficients are established in Bianchi & Salvati (2015). These properties make the M-quantile regression versatile and computationally appealing.

When it comes to a multivariate adaption of univariate M-quantiles, the main difficulty is that there does not exist a natural ordering in p dimensions, $p > 1$ (Breckling & Chambers 1988). The papers by Breckling & Chambers (1988) and Kokic et al. (2002) addressed this problem by introducing a direction vector in the Euclidean p-dimensional space to establish a suitable ordering procedure for multivariate observations. In both the univariate and multivariate cases, the available definitions of M-quantile assume independent observations and do not allow for the analysis of dependent data. In the next section we will consider a different approach to multivariate M-quantiles based on directional M-quantiles accounting for the possible correlation between observations that belong to the same cluster.

3 Marginal M-quantile model for multivariate dependent data

In this section we introduce a new definition of multivariate M-quantiles based on directional M-quantiles by extending the idea of Kong & Mizera (2012). In order to account for dependencies in the data, we develop a Marginal M-quantile (MMQ) regression model for directional M-quantiles, which incorporates a correlation matrix to handle within-cluster correlation. We then summarize the information contained in directional M-quantiles by describing the dependence between the outcome variables, and the location, shape and spread of the distribution of the responses conditional on different values of the covariates.

Suppose we have data on an absolutely continuous p-variate response variable $\mathbf{Y}_{ij} = (Y_{ij}^{(1)}, \dots, Y_{ij}^{(p)})'$ with $\mathbf{y}_{ij}$ being the corresponding observed value and let $\mathbf{X}_{ij} = (X_{ij}^{(1)}, \dots, X_{ij}^{(k)})$ be a k-dimensional vector of explanatory variables recorded for the i-th unit in the j-th cluster of size n_j, for $j = 1, ..., d$ and $i = 1, ..., n_j$ with $n = \sum_{j=1}^{d} n_j$. To simplify the notation, we stack up the projected responses on $\mathbf{u}$ to the n_j dimensional vector $\tilde{\mathbf{Y}}_j = (\mathbf{u}'\mathbf{Y}_{1j}, \dots, \mathbf{u}'\mathbf{Y}_{n_j j})'$, while $\mathbf{X}_j = (\mathbf{X}_{ij}, \dots, \mathbf{X}_{n_j j})$ is a $n_j \times k$ matrix collecting the covariates for group j.

Definition 1. *Let* $\mathbf{Y}$ *be a continuous* p*-dimensional random vector with absolutely continuous distribution function and let* $\psi(\cdot)$ *denote the Huber influence function in* (3). *For any* $\tau \in (0, 1)$

and direction $\mathbf{u} \in \mathcal{S}^{p-1}$*, the directional M-quantile of order* τ*, in the direction* $\mathbf{u}$*,* $\theta_{\mathbf{u}}(\tau)$*, is the* τ*-th M-quantile of the corresponding projection of the distribution of* $\mathbf{Y}$*.*

The proposed directional M-quantile is real-valued and it corresponds to the univariate τ-th M-quantile of the distribution of $\mathbf{u}'\mathbf{Y}$, where the direction $\mathbf{u}$ can be interpreted as a weight vector for each marginal distribution of $\mathbf{Y}$ involved in the regression problem. In addition, directional M-quantiles inherit the computational advantages, robustness and efficiency properties of standard univariate M-quantiles described in Section 2. Specifically, by varying the tuning constant c in (3), directional M-quantiles reduce to directional quantiles of Kong & Mizera (2012) when $c \to 0$ and reduce to directional expectiles for c large. Clearly, Definition 1 includes the traditional notion of univariate M-quantile. For $p = 1$, indeed, $\mathcal{S}^0$ simply reduces to two end-points, $\{-1, 1\}$, and $\theta_u(\tau)$ to the classical univariate M-quantile. The direction $\mathbf{u}$ is often selected depending on the empirical problem in order to produce meaningful results (see Paindaveine & Šiman 2011, Kong & Mizera 2012, Geraci et al. 2020 and Farcomeni et al. 2020). A further possibility is to use the principal component of a Principal Component Analysis by maximizing the variance of the projected data $\mathbf{u}'\mathbf{Y}$ as discussed in Korhonen & Siljamäki (1998) and in Geraci et al. (2020).

In the regression context, the proposed definition can be easily extended to conditional distributions when covariates are available. For a given $\tau \in (0, 1)$ and $\mathbf{u} \in \mathcal{S}^{p-1}$, the conditional directional M-quantile is defined as:

$$\theta_{\mathbf{u},\mathbf{x}}(\tau) = \mathbf{x}'_{ij}\boldsymbol{\beta}(\tau), \qquad i = 1, ..., n_j \text{ and } j = 1, ..., d, \tag{4}$$

where $\mathbf{x}_{ij}$ is the covariates vector for the i-th subject in the j-th group and $\boldsymbol{\beta}(\tau)$ is the k-dimensional vector of regression coefficients.

In the literature there have been numerous approaches proposed to account for the dependence structure of the data (see for instance Liang & Zeger 1986, Heagerty et al. 2000, Diggle et al. 2002, Goldstein 2011 and the references therein). One possible solution is to consider the so called marginal modeling framework (see Liang & Zeger 1986, Lindsey 1999, Heagerty et al. 2000, Bergsma et al. 2009) and estimate the parameters using the GEE approach of Liang & Zeger (1986). To account for the dependence structure which arises because of the clustered observations, we introduce a suitable correlation matrix $\mathbf{C}_j(\mathbf{r}_j)$ of size n_j indexed by the s_j-dimensional vector $\mathbf{r}_j$ which fully characterizes the correlation between groups, $j = 1, ..., d$. This "working" correlation matrix $\mathbf{C}_j(\mathbf{r}_j)$ is able to capture within group dependence and enhance the efficiency of the regression coefficients estimator (see also Liang & Zeger 1986, Zeger & Liang 1986 and Zeger et al. 1988).

Following Sinha & Rao (2009) and Liang & Zeger (1986), for a given τ and direction $\mathbf{u}$, we define the estimator $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ as the solution of the following Generalized M-quantile Estimating Equations (GMQEE):

$$\mathbf{U}(\boldsymbol{\beta}(\tau)) = \sum_{j=1}^{d} \mathbf{U}_j(\boldsymbol{\beta}(\tau)) = \sum_{j=1}^{d} \mathbf{X}'_j \boldsymbol{\Sigma}_j^{-1}(\mathbf{r}_j)\mathbf{V}_j^{\frac{1}{2}} \psi_\tau(\mathbf{z}_j) = \mathbf{0}, \tag{5}$$

where $\mathbf{z}_j = \mathbf{V}_j^{-\frac{1}{2}}(\tilde{\mathbf{Y}}_j - \mathbf{X}_j\boldsymbol{\beta}(\tau))$ denotes the n_j-dimensional vector of standardized residuals, $\mathbf{V}_j$ is the diagonal matrix of size n_j which contains the scale parameter σ_τ^2 for the residuals' distribution $\tilde{\mathbf{Y}}_j - \mathbf{X}_j\boldsymbol{\beta}(\tau)$, $\psi_\tau(\cdot)$ is the influence function in (3) and

$$\boldsymbol{\Sigma}_j(\mathbf{r}_j) = \mathbf{V}_j^{\frac{1}{2}}\mathbf{C}_j(\mathbf{r}_j)\mathbf{V}_j^{\frac{1}{2}}, \tag{6}$$

is the "working" covariance matrix. Several remarks are noteworthy regarding the methodology introduced above. First, when $\mathbf{C}_j(\mathbf{r}_j) = \mathbf{I}_{n_j}$, with $\mathbf{I}_{n_j}$ being the identity matrix of size n_j, (5) reduces to:

$$\mathbf{U}_I(\boldsymbol{\beta}(\tau)) = \sum_{j=1}^{d} \mathbf{X}_j' \mathbf{V}_j^{-\frac{1}{2}} \psi_\tau(\mathbf{z}_j) = \mathbf{0}, \tag{7}$$

where independence between clustered observations is assumed. In this case, we denote $\widehat{\boldsymbol{\beta}}_I(\tau)$ the estimator of $\boldsymbol{\beta}(\tau)$ as the solution of (7). Second, contrary to the well known GEE estimator of Liang & Zeger (1986), using the Huber's function ensures that $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ behaves robustly against outliers for finite values of c. Furthermore, by focusing on linear combinations of $\mathbf{Y}$, inference on $\boldsymbol{\beta}_{MMQ}(\tau)$ accounts for the possible correlation between the outcomes through the working correlation structure in (6). Finally, it should also be pointed out that, when the p directions forming the standard basis of $\mathbb{R}^p$ are considered, our methodology reduces to p component-wise univariate MMQ regressions as a by-product. To the best of our knowledge, this is the first time a marginal M-quantile regression model is being introduced in the literature.

As stated in Liang & Zeger (1986) and Zeger et al. (1988), (5) gives consistent estimates of the regression parameters and of their variances, and when the correlation structure of the data is appropriately incorporated, it improves the efficiency of parameter estimation relative to $\widehat{\boldsymbol{\beta}}_I(\tau)$ (Liang & Zeger 1986, Crowder 1995, Wang & Carey 2003, 2004 and Hin & Wang 2009). We present the asymptotic properties of the proposed estimator $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ in Section 4.1.

Several choices for $\mathbf{C}_j(\mathbf{r}_j)$ have been proposed in the related literature, such as the exchangeable correlation structure $[\mathbf{C}_j(\mathbf{r}_j)]_{ik} = r$ for all units i and k, $i \neq k$, in the j-th group, or the AR1 structure $[\mathbf{C}_j(\mathbf{r}_j)]_{ik} = r^{|i-k|}$ where the correlation decreases geometrically with separation as in autoregressive schemes; or the totally unspecified structure $[\mathbf{C}_j(\mathbf{r}_j)]_{ik} = r_{ik}$, where $[\mathbf{C}_j(\mathbf{r}_j)]_{ik}$ denotes the (i,k)-th element of $\mathbf{C}_j(\mathbf{r}_j)$. Their specification and the parameters interpretation depend on the application under investigation. For example, the exchangeable correlation structure occurs in clustered data while the AR1 structure can be a suitable choice to take into account time dependence among repeated measurements in longitudinal data.

3.1 M-quantile regions and contours

In the previous sections we described the MMQ regression model when a fixed direction in $\mathcal{S}^{p-1}$ is considered. To provide a full description of the dependence of the responses $\mathbf{Y}$ on the regressors $\mathbf{X}$, we investigate how directional M-quantiles can provide a summary when, theoretically, all directions over the $(p-1)$-dimensional unit sphere $\mathcal{S}^{p-1}$ are investigated simultaneously, for fixed τ.

Let $\mathbf{y}$ denote the realization of the random vector $\mathbf{Y}$. For a given $\tau \in (0,1)$ and $\mathbf{u} \in \mathcal{S}^{p-1}$, we first define the τ-th directional M-quantile regression hyperplane:

$$\pi_{\mathbf{u},\mathbf{x}}(\tau) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{u}'\mathbf{y} = \theta_{\mathbf{u},\mathbf{x}}(\tau)\}, \tag{8}$$

where $\theta_{\mathbf{u},\mathbf{x}}(\tau)$ is defined in (4). For example, when $p = 2$, the hyperplanes in (8) amount to lines which indicate how directional M-quantiles divide the data. Each hyperplane $\pi_{\mathbf{u},\mathbf{x}}(\tau)$ characterizes a lower (open) and an upper (closed) M-quantile regression halfspace $H^-_{\mathbf{u},\mathbf{x}}(\tau) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{u}'\mathbf{y} < \theta_{\mathbf{u},\mathbf{x}}(\tau)\}$ and $H^+_{\mathbf{u},\mathbf{x}}(\tau) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{u}'\mathbf{y} \geq \theta_{\mathbf{u},\mathbf{x}}(\tau)\}$, respectively. M-quantile centrality regions and contours of order τ are obtained by taking the "upper envelope" of the τ-th directional M-quantile hyperplanes in (8). If the distribution of $\mathbf{Y}$ is absolutely continuous, we may restrict to $\tau \in (0, \frac{1}{2}]$ and define the τ-th M-quantile region conditional on $\mathbf{X} = \mathbf{x}$, $R_{\mathbf{x}}(\tau) \subset \mathbb{R}^p$, as:

$$R_{\mathbf{x}}(\tau) = \bigcap_{\mathbf{u} \in \mathcal{S}^{p-1}} H^{+}_{\mathbf{u},\mathbf{x}}(\tau). \tag{9}$$

The region defined in (9) is convex, compact and bounded (Hallin et al. 2010, Kong & Mizera 2012), and the corresponding conditional M-quantile contour of order τ is defined as the boundary $\partial R_{\mathbf{x}}(\tau)$ of $R_{\mathbf{x}}(\tau)$. Such quantities are of crucial interest as they are able to detect covariate-dependent features of the distribution of the responses given $\mathbf{X}$, while ensuring robustness to outlying data. Specifically, for fixed τ, when the tuning constant of the Huber loss function c in (3) goes to zero, M-quantile contours reduce to directional quantile envelopes illustrated in Kong & Mizera (2012); on the other hand, when $c \to \infty$ our methodology allows us to introduce the definition of expectile contours as a particular case. Meanwhile, for a given c, the contours are nested as τ increases. As $\tau \to 0$, the M-quantile contour of order τ approaches the convex hull of the sample data providing valuable information about the extent of extremeness of the points.

4 Estimation and inference

In this section we provide the algorithm to compute an estimate of the robust estimator $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ in (5), for fixed τ and $\mathbf{u}$. Then, holding τ fixed, when $\mathbf{u}$ ranges over a subset of $\mathcal{S}^{p-1}$ we present the estimation procedure to obtain $\partial R_{\mathbf{x}}(\tau)$ and construct confidence envelopes. We conclude this section by deriving the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$.

In order to estimate $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ and the corresponding covariance matrix $\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau))$ we propose to use the iterative Newton-Raphson algorithm to solve the GMQEE in (5). The elements $\mathbf{r}_j$ of the correlation matrix $\mathbf{C}_j(\mathbf{r}_j)$, $j = 1, ..., d$ are obtained by exploiting the method of moments (see Liang & Zeger 1986, Fu & Wang 2012, Marino & Farcomeni 2015, Lu & Fan 2015 and Barry et al. 2018). As mentioned before, the choice of $\mathbf{C}_j(\mathbf{r}_j)$ depends on the empirical problem at hand. If, for example, we assume an exchangeable structure, we have that the correlation parameter $\mathbf{r}_j = r$ can be computed by using the following formula:

$$r = \frac{\sum_{j=1}^{d} \sum_{i<i'}^{n_j} \psi_\tau(z_{ij}) \psi_\tau(z_{i'j})}{\phi(\sum_{j=1}^{d} \frac{1}{2} n_j(n_j - 1) - k)} \qquad \textit{Exchangeable}, \tag{10}$$

while, for the first-order autoregressive working strucutre, r can be estimated by:

$$r = \frac{\sum_{j=1}^{d} \sum_{i=1}^{n_j - 1} \psi_\tau(z_{ij}) \psi_\tau(z_{i+1j})}{\phi(\sum_{j=1}^{d} n_j(n_j - 1) - k)} \qquad \textit{Autoregressive}. \tag{11}$$

Alternatively, when a completely general correlation matrix is considered, we have the unstructured case, i.e.:

$$r_{ii'} = \frac{\sum_{j=1}^{d} \psi_\tau(z_{ij}) \psi_\tau(z_{i'j})}{\phi(d - k)}, i \neq i' \qquad \textit{Unstructured}, \tag{12}$$

where $\phi = \frac{1}{n-k}\sum_{j=1}^{d}\sum_{i=1}^{n_j}\psi_\tau(z_{ij})^2$. In what follows, we report all the steps of the algorithm to estimate $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ and $\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau))$.

Algorithm The GMQEE algorithm

---

1 Let $\widehat{\boldsymbol{\beta}}^{(0)}(\tau) = \widehat{\boldsymbol{\beta}}_I(\tau)$ and $\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}^{(0)}(\tau)) = \frac{1}{d}\mathbf{I}_k$ denote the starting values for the algorithm.

2 Given $\widehat{\boldsymbol{\beta}}^{(b)}(\tau)$ at the b-th iteration, set $\widehat{\epsilon}_{ij}^{(b+1)} = \tilde{y}_{ij} - \mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}}^{(b)}(\tau)$ and compute σ_τ, z_{ij} and ϕ as:

$$\widehat{\sigma}_\tau^{(b+1)} = \frac{\text{Med}\{\mid \widehat{\epsilon}_{ij}^{(b+1)} - \text{Med}\{\widehat{\epsilon}_{ij}^{(b+1)}\} \mid\}}{0.6745},$$

$$\widehat{z}_{ij}^{(b+1)} = \frac{\tilde{y}_{ij} - \mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}}^{(b)}(\tau)}{\widehat{\sigma}_\tau^{(b+1)}},$$

$$\widehat{\phi}^{(b+1)} = \frac{1}{n-k}\sum_{j=1}^{d}\sum_{i=1}^{n_j}\psi_\tau(\widehat{z}_{ij}^{(b+1)})^2,$$

with $\tilde{y}_{ij}$ being the i-th element of the vector $\tilde{\mathbf{y}}_j$ defined in Section 3.

3 Depending upon the choice of $\mathbf{C}_j(\mathbf{r}_j)$, update the correlation parameters $\widehat{\mathbf{r}}_j^{(b+1)}$ using $\widehat{z}_{ij}^{(b+1)}$ and $\widehat{\phi}^{(b+1)}$.

4 Given $\widehat{\mathbf{r}}_j^{(b+1)}$, update $\widehat{\boldsymbol{\beta}}^{(b)}(\tau)$ and $\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}^{(b)}(\tau))$ by:

$$\widehat{\boldsymbol{\beta}}^{(b+1)}(\tau) = \widehat{\boldsymbol{\beta}}^{(b)}(\tau) + \left[-\frac{\partial\mathbf{U}(\boldsymbol{\beta}(\tau))}{\partial\boldsymbol{\beta}(\tau)}\right]^{-1}_{\widehat{\boldsymbol{\beta}}^{(b)}(\tau)}\left[\mathbf{U}(\boldsymbol{\beta}(\tau))\right]_{\widehat{\boldsymbol{\beta}}^{(b)}(\tau)},$$

$$\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}^{(b+1)}(\tau)) = \left[-\frac{\partial\mathbf{U}(\boldsymbol{\beta}(\tau))}{\partial\boldsymbol{\beta}(\tau)}\right]^{-1}_{\widehat{\boldsymbol{\beta}}^{(b)}(\tau)}\left[\text{Cov}(\mathbf{U}(\boldsymbol{\beta}(\tau)))\right]_{\widehat{\boldsymbol{\beta}}^{(b)}(\tau)}\left[-\frac{\partial\mathbf{U}(\boldsymbol{\beta}(\tau))}{\partial\boldsymbol{\beta}(\tau)}\right]^{-1}_{\widehat{\boldsymbol{\beta}}^{(b)}(\tau)},$$

where

$$\frac{\partial\mathbf{U}(\boldsymbol{\beta}(\tau))}{\partial\boldsymbol{\beta}(\tau)} = -\sum_{j=1}^{d}\mathbf{X}'_j\boldsymbol{\Sigma}_j^{-1}(\mathbf{r}_j)\mathbf{D}_j\mathbf{X}_j$$

and

$$\text{Cov}(\mathbf{U}(\boldsymbol{\beta}(\tau))) = \sum_{j=1}^{d}\mathbf{X}'_j\boldsymbol{\Sigma}_j^{-1}(\mathbf{r}_j)\mathbf{V}_j^{\frac{1}{2}}\psi_\tau(\mathbf{z}_j)\psi'_\tau(\mathbf{z}_j)\mathbf{V}_j^{\frac{1}{2}}\boldsymbol{\Sigma}_j^{-1}(\mathbf{r}_j)\mathbf{X}_j,$$

with $\mathbf{D}_j$ being the diagonal matrix with i-th element $[\mathbf{D}]_{ij} = \frac{\partial\psi_\tau(z_{ij})}{\partial z_{ij}}$.

5 Repeat 2-4, until convergence. In this work, convergence is achieved when the difference between the estimated model parameters obtained from two successive iterations is less than 10^{-8}.

At the end of the procedure, we compute the estimate $\widehat{\pi}_{\mathbf{u},\mathbf{x}}(\tau)$ and $\widehat{H}^+_{\mathbf{u},\mathbf{x}}(\tau)$ in (8) and (9). Keeping τ fixed, we repeat the algorithm by varying the direction $\mathbf{u}$ over a finite subset $\mathcal{S}_B^{p-1} \subset \mathcal{S}^{p-1}$ of all possible directions, $B \in \mathbb{N}$. For each $\mathbf{u} \in \mathcal{S}_B^{p-1}$, the model is re-estimated and the corresponding $\widehat{\pi}_{\mathbf{u},\mathbf{x}}(\tau)$ and $\widehat{H}^+_{\mathbf{u},\mathbf{x}}(\tau)$ are recorded. In this way, we obtain a sequence

$\{\widehat{H}^{+}_{\mathbf{u},\mathbf{x}}(\tau), \mathbf{u} \in \mathcal{S}^{p-1}_{B}\}$ which allows us to compute the estimate $\widehat{R}_{\mathbf{x}}(\tau)$ of $R_{\mathbf{x}}(\tau)$ as:

$$\widehat{R}_{\mathbf{x}}(\tau) = \bigcap_{\mathbf{u} \in \mathcal{S}^{p-1}_{B}} \{\widehat{H}^{+}_{\mathbf{u},\mathbf{x}}(\tau)\}. \tag{13}$$

We then estimate the contour $\partial\widehat{R}_{\mathbf{x}}(\tau)$ from (13).

To analyse the shape of $\partial\widehat{R}_{\mathbf{x}}(\tau)$ and provide a simple representation of its variability, we construct confidence regions for $\partial\widehat{R}_{\mathbf{x}}(\tau)$. Following Molchanov (2005) and Molchanov & Molinari (2018) let us denote by $\text{Haus}(A, B)$ the Hausdorff distance between two sets, say A and B. Our objective is to construct an asymptotically valid confidence set, $\mathcal{C}_{d,1-\alpha}$, such that:

$$\Pr(\partial R_{\mathbf{x}}(\tau) \subset \mathcal{C}_{d,1-\alpha}) = 1 - \alpha, \tag{14}$$

as $d \to \infty$. Let $W = \text{Haus}(\partial\widehat{R}_{\mathbf{x}}(\tau), \partial R_{\mathbf{x}}(\tau))$ and define:

$$w_{1-\alpha} = F^{-1}_{W}(1 - \alpha). \tag{15}$$

Then, it is easy to see that:

$$\Pr(\partial R_{\mathbf{x}}(\tau) \subset \partial\widehat{R}_{\mathbf{x}}(\tau) \oplus w_{1-\alpha}) \geq 1 - \alpha. \tag{16}$$

To approximate the distribution of W following Chen et al. (2017) and Molchanov & Molinari (2018), we adopt a nonparametric block bootstrap approach which preserves the group dependencies. Let $((\mathbf{Y}^{\star}_{1}, \mathbf{X}^{\star}_{1}), ..., (\mathbf{Y}^{\star}_{d^{\star}}, \mathbf{X}^{\star}_{d^{\star}}))$ be a bootstrap sample and let $\partial\widehat{R}^{\star}_{\mathbf{x}}(\tau)$ denote the corresponding estimate of the order-τ M-quantile regression contour. We define $W^{\star} = \text{Haus}(\partial\widehat{R}^{\star}_{\mathbf{x}}(\tau), \partial\widehat{R}_{\mathbf{x}}(\tau))$ and define the bootstrap estimate of $w_{1-\alpha}$ as:

$$\widehat{w}_{1-\alpha} = F^{-1}_{W^{\star}}(1 - \alpha). \tag{17}$$

Then the bootstrap confidence set for $\partial R_{\mathbf{x}}(\tau)$ is $\partial\widehat{R}_{\mathbf{x}}(\tau) \oplus \widehat{w}_{1-\alpha}$. In particular, this procedure allows us to construct asymptotically valid confidence envelopes for $\partial R_{\mathbf{x}}(\tau)$ (see Chen et al. 2017 and Molchanov & Molinari 2018) and identify potential influential observations depending on whether they fall inside or outside the estimated envelope.

4.1 Asymptotic properties

This section presents the asymptotic properties of the GMQEE estimator. Specifically, we derive the Bahadur-type (Bahadur 1966) representation, consistency and asymptotic normality for $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ for fixed τ and $\mathbf{u}$. Throughout this section, let $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_j(\mathbf{r}_j)$, $j = 1, ..., d$.

Consider the following assumptions:

(i) The distribution of the random vector $\mathbf{Y}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^p$, with density that has connected support, and admits finite first-order moments.

(ii) $(\mathbf{Y}_j, \mathbf{X}_j)$, $j = 1, ..., d$ is an i.i.d. sample from $(\mathbf{Y}, \mathbf{X})$.

(iii) The function $\rho(\cdot)$ related to (1) is continuous and strictly monotonic.

(iv) The function $\psi(\cdot)$ in (3) is bounded, non-decreasing and is twice differentiable at $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$, with the convention $\psi(0) = 0$.

(v) $\mathbb{E}[|| \mathbf{U}(\boldsymbol{\beta}(\tau)) ||^2] < \infty, \forall \boldsymbol{\beta}(\tau) \in \mathbb{R}^k$.

(vi) Let $\mathbf{H}$ denote the $k \times k$ matrix:

$$\mathbf{H} = \frac{1}{d} \sum_{j=1}^{d} \mathbf{X}_j' \boldsymbol{\Sigma}_j^{-1} \mathbb{E}[\mathbf{D}_j] \mathbf{X}_j, \tag{18}$$

with $\mathbf{H}$ being positive definite.

Theorem 1. *Let assumptions (i)-(vi) hold. Then,*

$$\sqrt{d}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau) - \boldsymbol{\beta}(\tau)) = \frac{1}{\sqrt{d}} \mathbf{H}^{-1} \sum_{j=1}^{d} \mathbf{U}_j(\boldsymbol{\beta}(\tau)) + o(1) \tag{19}$$

and

$$\sqrt{d}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau) - \boldsymbol{\beta}(\tau)) \xrightarrow{p} \mathcal{N}(\mathbf{0}, \mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1}) \quad \text{as} \quad d \to \infty, \tag{20}$$

with $\mathbf{B}$ *being*

$$\mathbf{B} = \frac{1}{d} \sum_{j=1}^{d} \mathbf{X}_j' \boldsymbol{\Sigma}_j^{-1} \mathbf{V}_j^{\frac{1}{2}} \mathbb{E}[\psi_\tau(\mathbf{z}_j)\psi_\tau'(\mathbf{z}_j)] \mathbf{V}_j^{\frac{1}{2}} \boldsymbol{\Sigma}_j^{-1} \mathbf{X}_j. \tag{21}$$

Proof. By assumptions (iii)-(iv), the Huber loss function $\rho(\cdot)$ with constant c bounded away from zero is continuous, differentiable and convex, thus the estimating equation $\mathbf{U}(\boldsymbol{\beta}(\tau))$ in (5) is continuous in $\boldsymbol{\beta}(\tau)$. Furthermore, $\mathbf{H}$ is positive definite by assumption (v) (see Bianchi & Salvati 2015). Then, Theorem 4 of Niemiro et al. (1992) applies which establishes the Bahadur representation in (19). Subsequently, (20) follows from (19) by the multivariate Central Limit Theorem and the Slutsky's Theorem. □

It is worth noting that assumptions (i)-(vi) are quite mild and standard in robust estimation theory. For example, assumption (i) holds when $\mathbf{Y}$ is multivariate Gaussian or multivariate Student t distribution with $\nu > 2$ degrees of freedom; assumption (iii) is a technical moment condition required for the asymptotic representation of $\boldsymbol{\beta}(\tau)$ while assumption (iv) is an identifiability condition. In assumption (iv) instead, the existence and positive-definiteness ensure the invertibility of $\mathbf{H}$ needed for the Bahadur representation.

In order to use Theorem 1 to build confidence intervals and hypothesis tests for $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$, a consistent estimator of the asymptotic covariance matrix $\mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1}$ in (20) is needed. We estimate $\mathbf{H}$ and $\mathbf{B}$ using a generalization of the robust estimator in White (1980) based on the well known sandwich approach, i.e.:

$$\widehat{\mathbf{H}} = \frac{1}{d} \sum_{j=1}^{d} \mathbf{X}_j' \widehat{\boldsymbol{\Sigma}}_j^{-1} \widehat{\mathbf{D}}_j \mathbf{X}_j, \tag{22}$$

$$\widehat{\mathbf{B}} = \frac{1}{d} \sum_{j=1}^{d} \mathbf{X}_j' \widehat{\boldsymbol{\Sigma}}_j^{-1} \mathbf{V}_j^{\frac{1}{2}} \psi_\tau(\widehat{\mathbf{z}}_j)\psi_\tau'(\widehat{\mathbf{z}}_j) \mathbf{V}_j^{\frac{1}{2}} \widehat{\boldsymbol{\Sigma}}_j^{-1} \mathbf{X}_j, \tag{23}$$

with $\widehat{\mathbf{D}}_j = \frac{\partial \psi_\tau(\widehat{z}_{ij})}{\partial z_{ij}}$ and $\widehat{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}_j(\widehat{\mathbf{r}}_j)$.

We now show that the covariance matrix estimator $\widehat{\mathbf{H}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{H}}^{-1}$ is consistent.

Theorem 2. *Let assumptions (i)-(vi) hold. Then,*

$$\widehat{\mathbf{H}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{H}}^{-1} - \mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1} \xrightarrow{p} 0, \tag{24}$$

where the notation is understood to indicate convergence of the matrices element by element.

Proof. To prove consistency of $\widehat{\mathbf{H}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{H}}^{-1}$, it suffices to apply Theorem 5 in Bianchi & Salvati (2015). □

Finally, following Prentice & Zhao (1991) and Yan & Fine (2004), the robust covariance estimator for the correlation parameter $\mathbf{r}_j$ is:

$$\boldsymbol{\Omega}(\widehat{\mathbf{r}}_j) = \Big(\sum_{j=1}^{d} \mathbf{K}_j'\mathbf{K}_j\Big)^{-1}\Big(\sum_{j=1}^{d} \mathbf{K}_j'\mathrm{Cov}(\widehat{\mathbf{s}}_j)\mathbf{K}_j\Big)\Big(\sum_{j=1}^{d} \mathbf{K}_j'\mathbf{K}_j\Big)^{-1}, \tag{25}$$

where $\mathbf{K}_j = \partial\boldsymbol{\alpha}_j/\partial\mathbf{r}_j$, $\boldsymbol{\alpha}_j$ and $\widehat{\mathbf{s}}_j$ are the $n_j(n_j-1)/2$ vectors of pairwise correlations in $\mathbf{C}_j(\mathbf{r}_j)$ and of upper triangular elements of the matrix $\psi_\tau(\widehat{\mathbf{z}}_j)\psi_\tau'(\widehat{\mathbf{z}}_j)$ in vector form, respectively.

5 Simulation study

In this section we conduct a simulation study to evaluate the finite sample properties of the proposed method. We address the following issues. First, we consider a subset of directions in $\mathcal{S}^{p-1}$ to study: (i) the efficiency of the MMQ model with respect to the independence assumption case and (ii) its robustness to outlying values and misspecification of the true correlation structure for different distributional choices of the error term and degrees of dependence among clustered units. Second, we provide a visual representation of the dependence between the $\mathbf{Y}$'s, and location and shape of M-quantile contours conditional on the covariates under different data generating mechanisms.

The observations are generated from the following bivariate, $p = 2$, regression model:

$$\mathbf{Y}_{ij} = \mathbf{X}_{ij}'\mathbf{B} + \boldsymbol{\epsilon}_{ij}, \qquad i = 1, ..., n_j \text{ and } j = 1, ..., d, \tag{26}$$

where $n_j = 7$ for $j = 1, ..., d$ with $d = 120$ and $\mathbf{X}_{ij} = (1, X_{ij}^{(1)})'$. The explanatory variable is generated from a standard Normal distribution and $\mathbf{B} = \left(\begin{smallmatrix}100 & 110\\ 2 & 1\end{smallmatrix}\right)$. Following Cho (2016), two error distributions are considered for $\boldsymbol{\epsilon}_j = (\boldsymbol{\epsilon}_{1j}, ..., \boldsymbol{\epsilon}_{n_j j})'$:

($\mathcal{N}$): multivariate Normal distribution with mean 0, marginal variance 1 and an exchangeable correlation structure with a correlation coefficient r;

($\mathcal{T}$): multivariate Student t distribution with 3 degrees of freedom, non centrality parameter equal to 0 and an exchangeable correlation structure with a correlation coefficient r.

This enables us to set both the correlation coefficient within the observations in the j-th group and the one between different response variables $k = 1, ..., p$ over the same i-th unit to be r, i.e. $\mathrm{Cor}(Y_{ijk}, Y_{i'jk}) = \mathrm{Cor}(Y_{ijk}, Y_{ijk'}) = r$ with $i \neq i'$ and $k \neq k'$. Similarly to Lu & Fan (2015), Fu & Wang (2012) and Lin et al. (2020), we consider errors with low ($r = 0.3$) and high ($r = 0.8$) correlation. To investigate the robustness of the proposed method to the presence of outliers, in the $\mathcal{N}$-scenario we contaminate the responses by using $\mathbf{Y}_{ij} + \delta_{ij}\mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma})$, where $\delta_{ij} \sim \text{Ber}(\alpha)$, with $\alpha = \Pr(\delta_{ij} = 1)$, and where $\mathbf{\Sigma}$ is a $p \times p$ diagonal variance-covariance matrix with marginal variances equal to 100 and 150. The proportion of contaminated observations α is chosen to be 10%. Naturally, when $\alpha = 0$ there is no contamination and errors follow a Normal distribution, whereas the other setting corresponds to clear deviations from normality to more heavy-tailed distributions.

For each simulation configuration, we select three quantile levels $\tau = (0.1, 0.5, 0.9)$ and three directions, namely $\mathbf{u}_1 = (1, 0), \mathbf{u}_2 = (\frac{1}{3}, \frac{2}{3})$ and $\mathbf{u}_3 = (0, 1)$, where the first and last vectors points vertically in the $Y^{(1)}$ and $Y^{(2)}$ direction, respectively. For $\mathbf{u}_1$ and $\mathbf{u}_3$, our MMQ model reduces to two component-wise univariate regressions where each marginal of $\mathbf{Y}$ is regressed onto the covariates $\mathbf{X}$, while the second direction weights equally $Y^{(1)}$ and $Y^{(2)}$. We first project $\mathbf{Y}$ onto each direction $\mathbf{u}$ and then regress $\mathbf{u}'\mathbf{Y}$ on the explanatory variables $\mathbf{X}$ using the MMQ model. For a given τ and $\mathbf{u}$, the true vector of the MMQ model parameters $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \beta_1(\tau))$ can be computed as $\boldsymbol{\beta}(\tau) = \mathbf{B}\mathbf{u}$, where the intercept $\beta_0(\tau)$ has been corrected to ensure that the conditional τ-th M-quantile of $\mathbf{u}'\mathbf{Y}$ is equal to $\mathbf{X}'\boldsymbol{\beta}(\tau)$. To evaluate the impact of misspecifying the working correlation structure on inference, we fit the MMQ model using the Exchangeable (E), Autoregressive of order one (AR1) and Unstructured (U) correlation matrices, and compare the results with the simplifying Independence (I) hypothesis which explicitly disregards the dependency between clustered observations. The tuning constant c in (3) has been set to 1.345 which gives reasonably efficiency under normality and protects against outliers (Huber & Ronchetti 2009).

We carry out $H = 1000$ Monte Carlo replications and we calculate the Average Relative Bias (ARB) defined as:

$$ARB(\hat{\theta}_\tau) = \frac{1}{H}\sum_{h=1}^{H}\frac{(\hat{\theta}_\tau^{(h)} - \theta_\tau)}{\theta_\tau} \times 100, \tag{27}$$

where $\hat{\theta}_\tau^{(h)}$ is the estimated parameter at quantile level τ for the h-th replication and θ_τ is the corresponding "true" value of the parameter. To evaluate the efficiency of $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ w.r.t. $\widehat{\boldsymbol{\beta}}_I(\tau)$, we compute the Relative Efficiency (REF) measure defined as:

$$\text{REF}(\widehat{\boldsymbol{\beta}}(\tau)) = \frac{S^2(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau))}{S^2(\widehat{\boldsymbol{\beta}}_I(\tau))}, \tag{28}$$

where $S^2(\widehat{\boldsymbol{\beta}}(\tau)) = \frac{1}{H}\sum_{h=1}^{H}(\widehat{\boldsymbol{\beta}}(\tau)^{(h)} - \bar{\boldsymbol{\beta}}(\tau))^2$ and $\bar{\boldsymbol{\beta}}(\tau) = \frac{1}{H}\sum_{h=1}^{H}\widehat{\boldsymbol{\beta}}(\tau)^{(h)}$. The REF defined in (28) measures the efficiency gain of the estimates of $\boldsymbol{\beta}(\tau)$ using the proposed directional M-quantile regression method, $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$, over the independence assumption, $\widehat{\boldsymbol{\beta}}_I(\tau)$. When $\text{REF}(\widehat{\boldsymbol{\beta}}(\tau))$ is less than one, this indicates that $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ is preferable. Tables 1-2 show the ARB and REF measures of the proposed estimators $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ for each component of the parameter vector $\boldsymbol{\beta}(\tau)$ under the considered working correlation structures. As can be noted, when there are no outliers in the data, the proposed model under the Gaussian and the Student t error distributions is able to recover the regression coefficients for both low (Table 1) and high (Table 2) degree of dependence. Not surprisingly, the bias effect is quite small when we analyze the median levels. As the τ levels become more extreme, the ARB increases because of the reduced amount of information in the tails of the distribution but it remains reasonably small. In the presence of outliers, the proposed method still provides uniformly good results even when the working correlation matrix is incorrectly specified, as large residuals are down-weighted by the constant c of the Huber functions and do not produce much larger biases. The results with $\alpha = 5\%$ and $\alpha = 20\%$ confirm these findings and are available from the authors.

Furthermore, the estimator of the proposed model is more efficient than the corresponding estimator from classical M-quantile regression under the independence assumption. Examination of Table 1 shows that with a moderate correlation ($r = 0.3$), the relative efficiencies of the regression estimators $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ perform slightly better when the errors follow a multivariate Normal and Student t distributions. When the correlation increases ($r = 0.8$), the proposed estimator become much more efficient than the working independence estimator. This pattern is consistent across all three examined quantile levels even under the misspecified AR1 correlation structure, indicating the robustness of the proposed method. In the case of contamination ($\alpha = 10\%$), our estimator still outperforms the naive one $\widehat{\boldsymbol{\beta}}_I(\tau)$ with low and large r. This demonstrates that the $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ estimator yields positive results in settings with clear departures from normality as the MMQ model protects against outlying values and accounts for the specific dependence structure embedded in the data. In addition, by focusing on linear combinations of the responses (see Panels B in Tables 1-2), there is an even greater improvement in the estimation efficiency compared to the independence assumption because the working correlation matrix also accounts for the correlation between the outcomes within each cluster.

To evaluate the performance of the estimated variances as described in (20), we report the Coverage Probability (CP) of nominal 95% confidence intervals for $\beta_0(\tau)$ and $\beta_1(\tau)$ defined by the number of times the interval $\theta_\tau \pm 2\sqrt{\mathrm{Var}(\hat{\theta}_\tau)}$ contains the "true" population parameter divided by the number of Monte Carlo replicates H. The results presented in Tables 3 and 4 indicate that under the Gaussian, contaminated Gaussian scheme and Student t scenarios, our variance estimator leads to confidence intervals with coverage close to the theoretical value of 0.95 for all τ levels. This therefore suggests that the proposed large sample approximation is suitable for approximate inference of the MMQ regression parameters for moderate contamination levels and misspecification of the working correlation.

Finally, to get a graphical representation of how M-quantile contours behave empirically, we consider 50 equally spaced directions on the unit circle and plot $\partial\widehat{R}_{\mathbf{x}}(\tau)$ for $\tau = (0.05, 0.1, 0.25, 0.4)$ with $c = 1.345$. Figure 1 shows the estimated contours under the four data generating processes with a correlation coefficient of $r = 0.3$, conditional on the 0.05-th (violet), 0.5-th (orange) and 0.95-th (green) empirical quantiles of $X_{ij}^{(1)}$. In particular, one can see that $\partial\widehat{R}_{\mathbf{x}}(\tau)$ slowly ascend upward along the data cloud, demonstrating the positive dependence with increasing values of the covariate. The most obvious features of all plots are the fact that the enclosed area is decreasing with increasing τ, thus the contours are neatly nested, and their behaviour under different levels of contamination by outliers. These figures also suggest that, as the contour lines approach the convex hull of the sample data for small values of τ, they can be employed to detect possible outliers, corresponding to extreme points falling outside the estimated boundary.

6 Application

In this section we apply the proposed methodology to the Tennessee's Student/Teacher Achievement Ratio (STAR) dataset (http://fmwww.bc.edu/ec-p/data). The STAR experiment (see Word et al. 1990 and Finn & Achilles 1990) is a four-year longitudinal class-size

Model	Coef	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
		$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$		
Panel A: $\mathbf{u}_1$										
I	β_0	0.002 (1.000)	−0.011 (1.000)	−0.019 (1.000)	0.005 (1.000)	−0.010 (1.000)	0.006 (1.000)	−0.024 (1.000)	−0.023 (1.000)	−0.037 (1.000)
	β_1	−0.009 (1.000)	0.052 (1.000)	−0.064 (1.000)	−0.116 (1.000)	−0.171 (1.000)	−0.160 (1.000)	−0.025 (1.000)	−0.035 (1.000)	−0.111 (1.000)
E	β_0	0.002 (0.985)	−0.010 (1.013)	−0.020 (1.002)	0.005 (1.004)	−0.010 (0.981)	0.007 (0.997)	−0.025 (1.002)	−0.023 (1.002)	−0.038 (0.982)
	β_1	0.000 (0.828)	0.099 (0.786)	−0.090 (0.879)	−0.138 (0.950)	−0.064 (0.862)	−0.110 (1.008)	−0.124 (0.850)	−0.060 (0.825)	−0.049 (0.838)
AR1	β_0	0.002 (0.984)	−0.011 (1.019)	−0.020 (0.999)	0.005 (1.001)	−0.011 (0.978)	0.007 (0.999)	−0.024 (1.011)	−0.023 (1.018)	−0.039 (0.995)
	β_1	−0.037 (0.946)	0.091 (0.940)	−0.097 (0.972)	−0.156 (0.981)	−0.081 (0.968)	−0.131 (1.014)	−0.031 (1.003)	−0.028 (0.958)	−0.119 (1.002)
U	β_0	0.005 (1.027)	−0.011 (1.045)	−0.025 (1.057)	−0.000 (1.067)	−0.012 (1.017)	0.011 (1.051)	−0.002 (1.019)	−0.024 (1.009)	−0.060 (0.948)
	β_1	0.004 (0.874)	0.063 (0.812)	−0.087 (0.939)	−0.135 (1.036)	−0.113 (0.912)	0.020 (1.038)	0.048 (0.830)	−0.070 (0.885)	0.061 (0.880)
Panel B: $\mathbf{u}_2$										
I	β_0	0.006 (1.000)	−0.007 (1.000)	−0.018 (1.000)	0.008 (1.000)	−0.005 (1.000)	−0.005 (1.000)	−0.013 (1.000)	−0.018 (1.000)	−0.038 (1.000)
	β_1	0.003 (1.000)	0.016 (1.000)	−0.196 (1.000)	−0.099 (1.000)	−0.075 (1.000)	−0.105 (1.000)	−0.153 (1.000)	−0.009 (1.000)	−0.026 (1.000)
E	β_0	0.005 (1.009)	−0.007 (1.023)	−0.017 (1.036)	0.008 (0.994)	−0.004 (0.991)	−0.005 (1.006)	−0.014 (1.007)	−0.018 (0.988)	−0.036 (1.013)
	β_1	0.040 (0.718)	0.017 (0.627)	−0.060 (0.817)	−0.107 (0.910)	0.044 (0.730)	−0.073 (0.907)	−0.147 (0.714)	−0.007 (0.646)	−0.095 (0.648)
AR1	β_0	0.005 (1.014)	−0.007 (1.036)	−0.017 (1.044)	0.008 (0.993)	−0.004 (1.002)	−0.005 (1.005)	−0.013 (1.028)	−0.018 (1.010)	−0.036 (1.033)
	β_1	0.043 (0.877)	0.017 (0.877)	−0.095 (0.956)	−0.102 (0.967)	0.028 (0.868)	−0.111 (0.939)	−0.003 (0.883)	0.037 (0.859)	−0.130 (0.822)
U	β_0	0.011 (1.036)	−0.007 (1.043)	−0.022 (1.092)	0.003 (1.080)	−0.005 (1.040)	0.002 (1.078)	−0.004 (0.990)	0.004 (0.996)	−0.017 (0.973)
	β_1	0.004 (0.812)	0.014 (0.712)	−0.109 (0.910)	−0.009 (1.014)	0.026 (0.777)	−0.111 (1.034)	−0.158 (0.717)	−0.009 (0.674)	−0.022 (0.648)
Panel C: $\mathbf{u}_3$										
I	β_0	−0.000 (1.000)	−0.002 (1.000)	−0.011 (1.000)	0.002 (1.000)	0.003 (1.000)	0.006 (1.000)	−0.012 (1.000)	−0.017 (1.000)	−0.040 (1.000)
	β_1	−0.115 (1.000)	−0.183 (1.000)	−0.199 (1.000)	−0.200 (1.000)	0.101 (1.000)	0.148 (1.000)	0.067 (1.000)	−0.076 (1.000)	−0.154 (1.000)
E	β_0	−0.000 (1.005)	−0.003 (1.027)	−0.011 (1.015)	0.002 (0.990)	0.003 (1.006)	0.005 (0.969)	−0.016 (1.018)	−0.019 (1.000)	−0.039 (1.026)
	β_1	−0.068 (0.852)	−0.109 (0.825)	−0.112 (0.935)	−0.213 (0.986)	0.128 (0.885)	0.149 (0.961)	−0.078 (0.750)	−0.034 (0.775)	−0.219 (0.761)
AR1	β_0	0.000 (1.007)	−0.002 (1.039)	−0.011 (1.022)	0.002 (0.992)	0.003 (1.018)	0.005 (0.970)	−0.016 (1.023)	−0.019 (1.003)	−0.040 (1.044)
	β_1	−0.070 (0.951)	−0.096 (0.971)	−0.175 (1.000)	−0.227 (1.004)	0.111 (1.010)	0.135 (0.979)	0.047 (0.906)	0.023 (0.912)	−0.186 (0.917)
U	β_0	0.005 (1.085)	−0.003 (1.058)	−0.015 (1.085)	−0.001 (1.050)	0.002 (1.067)	0.008 (1.028)	0.004 (1.031)	−0.017 (1.036)	−0.060 (1.036)
	β_1	−0.167 (0.921)	−0.124 (0.869)	−0.092 (0.999)	−0.208 (1.061)	0.181 (0.921)	0.199 (1.068)	−0.062 (0.797)	−0.097 (0.814)	−0.132 (0.755)

Table 1: Values of ARB (in percentage) and REF (in brackets) of $\beta_0(\tau)$ and $\beta_1(\tau)$ over 1000 Monte Carlo simulations under the three data generating scenarios with low correlation ($\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = 0.3$).

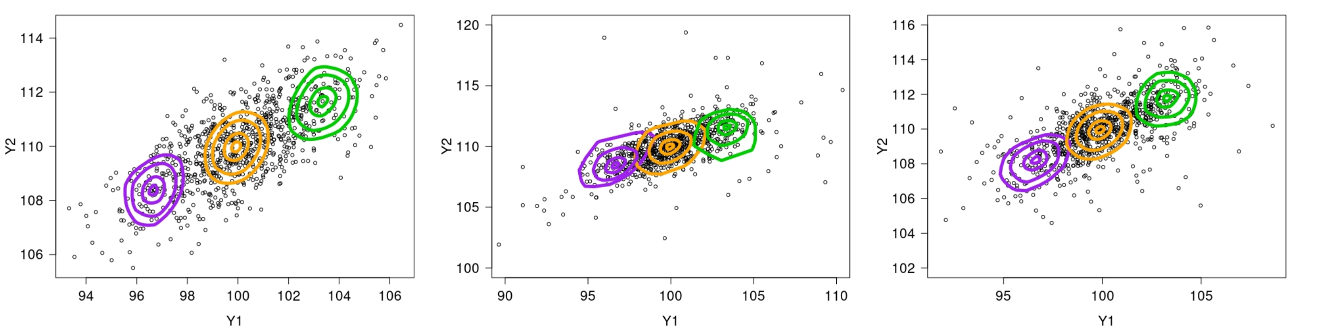

Figure 1: From left to right, estimated M-quantile contours under the $\mathcal{N}$, $\mathcal{N}-10\%$ and $\mathcal{T}$ simulation scenarios at level $\tau = (0.05, 0.1, 0.25, 0.4)$ (from the outside inwards), conditional on the 0.05-th (violet), 0.5-th (orange) and 0.95-th (green) empirical quantiles of $X_{ij}^{(1)}$.

Model	Coef	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
		$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$		
					Panel A: $\mathbf{u}_1$					
I	β_0	0.012 (1.000)	−0.012 (1.000)	−0.028 (1.000)	0.015 (1.000)	−0.005 (1.000)	0.000 (1.000)	−0.045 (1.000)	−0.030 (1.000)	−0.047 (1.000)
	β_1	−0.050 (1.000)	−0.073 (1.000)	−0.118 (1.000)	−0.057 (1.000)	−0.058 (1.000)	−0.077 (1.000)	−0.046 (1.000)	−0.060 (1.000)	0.001 (1.000)
E	β_0	0.008 (0.998)	−0.013 (0.989)	−0.025 (1.007)	0.015 (1.003)	−0.005 (1.006)	−0.002 (0.985)	−0.045 (1.032)	−0.030 (0.992)	−0.045 (0.986)
	β_1	−0.023 (0.347)	−0.014 (0.271)	−0.025 (0.382)	0.020 (0.857)	−0.005 (0.508)	0.019 (0.894)	−0.053 (0.344)	−0.064 (0.301)	0.044 (0.340)
AR1	β_0	0.009 (1.029)	−0.013 (1.007)	−0.026 (1.013)	0.016 (0.999)	−0.005 (1.007)	−0.002 (0.983)	−0.044 (1.066)	−0.031 (1.035)	−0.047 (1.018)
	β_1	−0.033 (0.534)	−0.012 (0.409)	−0.041 (0.558)	−0.028 (0.980)	−0.037 (0.674)	−0.044 (0.966)	−0.030 (0.532)	−0.053 (0.477)	0.007 (0.534)
U	β_0	0.014 (1.013)	−0.013 (0.989)	−0.032 (1.102)	−0.010 (1.105)	−0.006 (1.044)	0.024 (1.103)	−0.030 (1.051)	−0.031 (0.979)	−0.057 (1.007)
	β_1	−0.026 (0.376)	−0.024 (0.284)	−0.022 (0.430)	0.077 (0.961)	0.004 (0.508)	−0.011 (0.976)	0.010 (0.330)	−0.071 (0.307)	0.072 (0.354)
					Panel B: $\mathbf{u}_2$					
I	β_0	0.014 (1.000)	−0.010 (1.000)	−0.027 (1.000)	0.019 (1.000)	−0.006 (1.000)	−0.018 (1.000)	−0.034 (1.000)	−0.027 (1.000)	−0.052 (1.000)
	β_1	−0.064 (1.000)	−0.105 (1.000)	−0.116 (1.000)	−0.076 (1.000)	−0.099 (1.000)	−0.096 (1.000)	−0.083 (1.000)	−0.067 (1.000)	−0.053 (1.000)
E	β_0	0.012 (0.969)	−0.011 (1.026)	−0.023 (1.015)	0.020 (0.978)	−0.005 (0.979)	−0.017 (1.011)	−0.038 (1.046)	−0.028 (0.992)	−0.046 (1.037)
	β_1	0.008 (0.213)	−0.021 (0.166)	−0.004 (0.249)	0.036 (0.813)	0.054 (0.422)	−0.030 (0.817)	−0.060 (0.199)	−0.029 (0.167)	−0.011 (0.197)
AR1	β_0	0.013 (1.002)	−0.010 (1.037)	−0.024 (1.018)	0.021 (0.982)	−0.005 (0.998)	−0.017 (1.011)	−0.037 (1.085)	−0.029 (1.026)	−0.047 (1.087)
	β_1	−0.043 (0.348)	−0.020 (0.267)	−0.016 (0.387)	0.002 (0.945)	0.008 (0.587)	−0.111 (0.949)	−0.024 (0.325)	−0.014 (0.290)	−0.082 (0.318)
U	β_0	0.015 (1.032)	−0.012 (1.015)	−0.026 (1.063)	−0.015 (1.101)	−0.005 (1.045)	0.020 (1.176)	−0.028 (1.090)	−0.027 (0.989)	−0.054 (1.050)
	β_1	0.031 (0.233)	−0.008 (0.178)	−0.033 (0.275)	0.047 (0.914)	0.058 (0.429)	0.002 (0.935)	−0.021 (0.198)	−0.029 (0.174)	−0.043 (0.199)
					Panel C: $\mathbf{u}_3$					
I	β_0	0.013 (1.000)	−0.008 (1.000)	−0.028 (1.000)	0.014 (1.000)	−0.004 (1.000)	−0.017 (1.000)	−0.027 (1.000)	−0.026 (1.000)	−0.054 (1.000)
	β_1	−0.111 (1.000)	−0.243 (1.000)	−0.329 (1.000)	−0.189 (1.000)	−0.016 (1.000)	−0.086 (1.000)	−0.125 (1.000)	0.019 (1.000)	−0.122 (1.000)
E	β_0	0.012 (0.999)	−0.008 (1.005)	−0.024 (1.034)	0.015 (0.977)	−0.003 (1.005)	−0.015 (0.997)	−0.033 (1.047)	−0.028 (0.993)	−0.050 (1.008)
	β_1	0.008 (0.368)	−0.062 (0.305)	−0.022 (0.421)	−0.109 (0.853)	−0.015 (0.496)	−0.036 (0.902)	−0.064 (0.318)	−0.006 (0.246)	−0.170 (0.333)
AR1	β_0	0.013 (1.018)	−0.008 (1.035)	−0.024 (1.059)	0.016 (0.978)	−0.003 (1.039)	−0.016 (1.004)	−0.031 (1.084)	−0.028 (1.016)	−0.050 (1.045)
	β_1	−0.097 (0.546)	−0.081 (0.473)	0.022 (0.606)	−0.137 (0.977)	−0.099 (0.683)	−0.107 (1.051)	−0.008 (0.486)	0.071 (0.404)	−0.252 (0.527)
U	β_0	0.019 (1.037)	−0.007 (1.018)	−0.029 (1.066)	−0.011 (1.063)	−0.003 (1.056)	0.011 (1.114)	−0.018 (1.065)	−0.026 (0.998)	−0.063 (1.050)
	β_1	−0.014 (0.413)	−0.082 (0.315)	−0.050 (0.478)	−0.353 (0.999)	−0.077 (0.481)	−0.126 (1.010)	−0.068 (0.348)	0.043 (0.279)	−0.115 (0.320)

Table 2: Values of ARB (in percentage) and REF (in brackets) of $\beta_0(\tau)$ and $\beta_1(\tau)$ over 1000 Monte Carlo simulations under the three data generating scenarios with high correlation ($\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = 0.8$).

study funded by the Tennessee General Assembly and conducted by the State Department of Education. Over 7,000 students in 79 schools were randomly assigned into one of three interventions: small class (13 to 17 students per teacher), regular class (22 to 25 students per teacher), and regular-with-aide class (22 to 25 students with a full-time teacher's aide). Classroom teachers were also randomly assigned to the classes they would teach. The interventions were initiated as the students entered school in kindergarten and continued through third grade. The outcome variables of interest are the scores of mathematics and reading tests of the Stanford Achievement Test (SAT-9) which are representative of educational attainment in young students.

Schooling systems present an obvious example of dependency between observations, with pupils clustered within schools, which the analysis needs to take into due account in order to avoid misleading inferences. Previous studies examine mathematics and reading test scores independently using univariate statistical methods neglecting possible information about the relationship between the grades of the two subjects. In addition, linear models focused on how educational attainment is determined, on average, by various explanatory variables despite prior research suggests that the magnitude and direction of relationships may differ across the

Model	Coef	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
			$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$	
						Panel A: $\mathbf{u}_1$				
E	β_0	0.948	0.942	0.931	0.949	0.936	0.941	0.965	0.939	0.935
	β_1	0.941	0.957	0.933	0.960	0.938	0.957	0.942	0.957	0.950
AR1	β_0	0.949	0.940	0.932	0.948	0.938	0.944	0.961	0.937	0.931
	β_1	0.948	0.961	0.943	0.957	0.941	0.956	0.956	0.945	0.949
U	β_0	0.933	0.937	0.919	0.942	0.930	0.935	0.938	0.935	0.921
	β_1	0.929	0.948	0.917	0.941	0.927	0.940	0.915	0.941	0.931
						Panel B: $\mathbf{u}_2$				
E	β_0	0.943	0.948	0.933	0.949	0.953	0.942	0.958	0.952	0.913
	β_1	0.945	0.948	0.939	0.961	0.942	0.949	0.954	0.964	0.961
AR1	β_0	0.941	0.948	0.934	0.948	0.949	0.941	0.955	0.951	0.913
	β_1	0.940	0.948	0.941	0.960	0.950	0.943	0.953	0.963	0.959
U	β_0	0.933	0.948	0.905	0.938	0.947	0.944	0.936	0.944	0.937
	β_1	0.921	0.935	0.921	0.940	0.926	0.935	0.925	0.952	0.941
						Panel C: $\mathbf{u}_3$				
E	β_0	0.951	0.957	0.948	0.949	0.953	0.953	0.958	0.945	0.916
	β_1	0.949	0.945	0.951	0.960	0.952	0.947	0.949	0.964	0.957
AR1	β_0	0.947	0.957	0.944	0.949	0.953	0.953	0.961	0.942	0.919
	β_1	0.956	0.956	0.951	0.953	0.950	0.945	0.957	0.958	0.957
U	β_0	0.934	0.955	0.928	0.946	0.948	0.948	0.930	0.941	0.933
	β_1	0.919	0.937	0.931	0.947	0.932	0.929	0.925	0.954	0.937

Table 3: CP of $\beta_0(\tau)$ and $\beta_1(\tau)$ over 1000 Monte Carlo simulations under the three data generating scenarios with low correlation ($\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = 0.3$).

distribution of achievement gains (Haile & Nguyen 2008, Kelcey et al. 2019). For example, small classes are likely to be beneficial to students at risk for school failure than highly skilled pupils hence, the effect of class size on students' performance might be thought of as quantile-specific. Only recently, Guggisberg (2019) jointly analyzed math and reading scores within a Bayesian framework for estimation of directional quantiles.

The aim of this analysis is to investigate how the effect of classroom size and teacher's experience affect differently the achievement of proficient students (high quantiles) and less proficient students (low quantiles). We considered the subset of students in kindergarten for a sample size of $n = 3743$ divided in $d = 79$ schools, after removing missing data and, as in Guggisberg (2019), omitting large classrooms that had a teaching assistant. Since a pupil's

Model	Coef	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
		$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$		
					Panel A: $\mathbf{u}_1$					
E	β_0	0.946	0.948	0.935	0.935	0.953	0.932	0.960	0.945	0.918
	β_1	0.941	0.952	0.943	0.974	0.948	0.961	0.956	0.951	0.954
AR1	β_0	0.944	0.954	0.933	0.932	0.955	0.930	0.963	0.940	0.920
	β_1	0.937	0.949	0.937	0.970	0.958	0.958	0.954	0.945	0.954
U	β_0	0.929	0.951	0.912	0.938	0.950	0.928	0.949	0.945	0.891
	β_1	0.909	0.936	0.910	0.964	0.935	0.950	0.908	0.936	0.915
					Panel B: $\mathbf{u}_2$					
E	β_0	0.946	0.951	0.926	0.937	0.953	0.933	0.954	0.943	0.920
	β_1	0.942	0.944	0.945	0.965	0.953	0.967	0.964	0.967	0.966
AR1	β_0	0.945	0.953	0.922	0.938	0.952	0.935	0.949	0.943	0.912
	β_1	0.933	0.944	0.935	0.966	0.954	0.964	0.969	0.958	0.962
U	β_0	0.928	0.952	0.915	0.947	0.949	0.939	0.949	0.943	0.908
	β_1	0.927	0.935	0.924	0.965	0.941	0.953	0.905	0.944	0.913
					Panel C: $\mathbf{u}_3$					
E	β_0	0.941	0.954	0.918	0.936	0.952	0.935	0.959	0.951	0.918
	β_1	0.942	0.946	0.930	0.964	0.962	0.962	0.958	0.957	0.956
AR1	β_0	0.947	0.952	0.927	0.938	0.951	0.934	0.954	0.944	0.911
	β_1	0.932	0.930	0.934	0.963	0.958	0.957	0.952	0.958	0.957
U	β_0	0.927	0.953	0.910	0.943	0.946	0.938	0.949	0.949	0.935
	β_1	0.916	0.932	0.919	0.950	0.949	0.945	0.929	0.943	0.916

Table 4: CP of $\beta_0(\tau)$ and $\beta_1(\tau)$ over 1000 Monte Carlo simulations under the three data generating scenarios with high correlation ($\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = 0.8$).

performance is likely to depend not only on its abilities, but also on the characteristics of the school, to handle dependence between pupils within the same school and avoid convergence difficulties due to large sized clusters, we assume a parsimonious parametrization of the correlation matrix, namely an exchangeable correlation structure. Following established custom, the tuning constant c in (3) has been set to 1.345 which gives reasonably efficiency under normality and protects against outliers (Huber & Ronchetti 2009).

As a preliminary step, to support the choice of using a robust approach we study the conditional distributions of mathematics and reading scores by fitting separately two univariate Marginal Mean (MM) models under an exchangeable correlation structure. The model includes the following two predictors, namely classroom size and teacher's experience. Figure

2 shows the normal probability plot of the residuals for mathematics (left) and reading (right) test scores. These reveal the presence of potentially influential observations in the data, indicate severe departures from the Gaussian assumption for both outcomes and show that data are severely skewed. For these reasons, a robust approach based on M-quantile seems to be appropriate for these data.

Therefore, we estimate the MMQ model for mathematics and reading scores for specific directions of interest. Then, we inspect τ-th M-quantile contours when a subset of directions in $\mathcal{S}^{p-1}$ is considered simultaneously and directions are aggregated together as shown in (13). Sections 6.1 and 6.2 report the results, respectively.

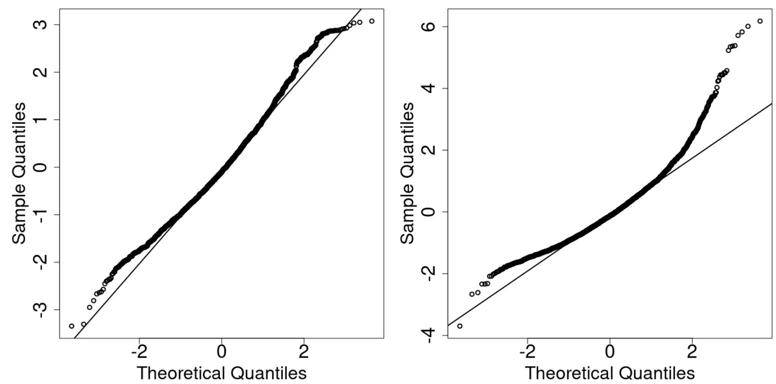

Figure 2: Normal probability plots residuals from a Marginal Mean model under an exchangeable correlation structure for mathematics (left) and reading (right) scores.

6.1 Fixed-u analysis

We fit the MMQ model for $\tau = (0.1, 0.25, 0.5, 0.75, 0.9)$ and select three different directions, i.e., $\mathbf{u}_1 = (1, 0)$, $\mathbf{u}_2 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $\mathbf{u}_3 = (0, 1)$, using the same covariates as above. The considered directions have a natural interpretation and allow us to construct linear combinations of mathematics and reading scores depending on how much importance the researcher wants to give to each subject. In the educational context, a weighted average mark is relevant because it represents multiple cognitive domains that has improved power compared with the most sensitive single test items (see Israel et al. 2001 and Kolen et al. 2012). Thus, $\mathbf{u}_1$ and $\mathbf{u}_3$ reduce the multidimensional problem to two MMQ regressions on each component of the bivariate response. On the other hand, $\mathbf{u}_2$ is equivalent to choosing the arithmetic mean of the two scores.

Table 5 shows point estimates of the regression coefficients and of the correlation parameter for the MMQ model at the investigated quantile levels. Statistical significance of regression coefficients is assessed by computing asymptotic standard errors as described in Section 4.1. Because we can investigate both mathematics and reading scores among their linear combinations, we compare the proposed MMQ model with existing univariate approaches in the literature. For each direction $\mathbf{u}$, relative to the standard GEE approach we report the results of the MM model with an exchangeable correlation matrix. In addition, we also consider the two-level M-quantile Random Effects (MQRE) model of Tzavidis et al. (2016) and the Linear Quantile Mixed Model (LQMM) of Geraci & Bottai (2014) with random intercepts, which is equivalent to assuming an exchangeable correlation structure. These allow us to evaluate the sensitivity of our methodology when random intercepts specified at the school level are included to account for a two level hierarchical structure in the data. Table 6 reports the estimated parameters of the MQRE model, as well as the estimated Intraclass Correlation Coefficient (ICC), defined as the ratio of the variance of the random intercepts to the total variance, which measures the proportion of variance explained by clustering. Table 7 shows the estimates of the regression coefficients and variance component (σ^2_{school}) for the two-level LQMM. Parameter estimates are displayed in boldface when significant at the 5% level.

Firstly, we observe that the MM and MMQ models produce comparable estimates at the center of the distribution ($\tau = 0.5$) however the MM regression model cannot be used to estimate the covariates' effects in the tails of the distribution. Secondly, consistently with the quantile regression framework, the estimated intercepts increase when moving from lower to upper quantiles. The results show that there is evidence of a negative association between the increase in classroom size and school performance across the examined quantiles. In particular, the size of the estimated effect is more pronounced at the upper tail of the distribution of each score than at the lower tail. Therefore, our results indicate that high-performing students are more affected than low-performing ones by a larger number of students in the classroom. On the other hand, teacher's experience is positively associated with the responses from the lower quantiles up to the 75-th percentile at the 10% significance level. Such effect is more evident for $\mathbf{u}_1$ and $\mathbf{u}_2$, suggesting that teachers preparation possibly influences students' cognitive abilities and skills in various school subjects. As it is evident, the estimates of the MQRE and the MMQ are similar among the three directions both in the center and in the tails of the distributions of the responses. The major difference is in the magnitude of the estimated regression coefficients with respect to the LQMM as the estimate of the teacher's experience effect is only statistically significant at the tails of the conditional distribution of the outcome. Finally, by looking at the within-group correlations, the estimated correlation parameters (r) of the MM and MMQ models with $\tau = 0.5$ are similar among the three directions. Similarly, the correlation coefficient estimates of the MMQ and the ICC values of the MQRE models are very close at the five investigated τ levels. The sign of the estimates indicates that pupils in the same school are more alike than students in different schools, highlighting the importance of clustering. It is also worth noticing that the intra-school correlation shows an inverted U-shape effect as the quantile level increases, i.e., the correlation between observations that belong to the same cluster is high at the center and low at the tail of the distribution of the outcomes. Differences between schools, therefore, seem to play a less prominent role in explaining mathematics and reading scores below- and above-the-average students' performance (Geraci & Bottai 2014).

6.2 Fixed-τ analysis

To provide a graphical representation of the effects of classroom size and teachers experience at the tails of the distribution of test scores, we fit the MMQ model at $\tau = (0.005, 0.1)$ for 100 equispaced directions and construct M-quantile regression contours using (13). Figure 3 illustrates the estimated $\partial \widehat{R}_{\mathbf{x}}(\tau)$ conditional on small (red curves) and large (blue curves) classes at the 0.01-th (top-left), 0.25-th (top-right), 0.75-th (bottom-left) and 0.99-th (bottom-right) empirical quantiles of teacher's experience, which correspond to 0, 4, 13 and 27 years of experience. The shaded areas represent 95% confidence envelopes obtained through the non-parametric bootstrap method of Section 3.1 using 1000 re-samples. For comparison purposes, we also consider the directional quantile contours of Kong & Mizera (2012) by fitting the proposed MMQ model with $c = 0.01$ under the working independence correlation structure

u	Variable	MM	MMQ				
			0.1	0.25	0.5	0.75	0.9
$\mathbf{u}_1$	Intercept	**486.541** (3.915)	**443.343** (3.086)	**462.121** (3.370)	**484.136** (3.921)	**509.513** (4.732)	**537.082** (6.078)
	Class size	$-$**9.306** (2.805)	$-$**6.773** (2.505)	$-$**7.350** (2.425)	$-$**8.499** (2.716)	$-$**10.922** (3.392)	$-$**14.556** (4.442)
	Teacher Experience	**0.585** (0.280)	**0.576** (0.207)	**0.588** (0.237)	**0.559** (0.280)	0.552 (0.338)	0.667 (0.483)
	r	**0.192** (0.028)	**0.135** (0.035)	**0.184** (0.040)	**0.200** (0.034)	**0.163** (0.026)	**0.097** (0.020)
$\mathbf{u}_2$	Intercept	**653.411** (4.236)	**606.399** (3.179)	**626.367** (3.598)	**650.296** (4.348)	**677.574** (5.433)	**707.618** (6.785)
	Class size	$-$**11.233** (3.023)	$-$**7.808** (2.875)	$-$**8.678** (2.660)	$-$**10.641** (2.933)	$-$**13.711** (3.696)	$-$**16.221** (4.731)
	Teacher Experience	**0.662** (0.305)	**0.640** (0.244)	**0.677** (0.257)	**0.674** (0.307)	0.701 (0.391)	0.688 (0.547)
	r	**0.203** (0.028)	**0.155** (0.041)	**0.205** (0.043)	**0.222** (0.034)	**0.174** (0.025)	**0.098** (0.019)
$\mathbf{u}_3$	Intercept	**437.504** (2.516)	**411.791** (1.541)	**422.004** (1.806)	**434.494** (2.454)	**448.991** (3.489)	**466.199** (4.693)
	Class size	$-$**6.569** (1.735)	$-$**4.507** (1.467)	$-$**5.108** (1.442)	$-$**5.981** (1.651)	$-$**7.118** (2.244)	$-$**8.081** (3.081)
	Teacher Experience	0.353 (0.186)	**0.291** (0.135)	**0.339** (0.139)	**0.363** (0.185)	0.421 (0.272)	0.478 (0.363)
	r	**0.194** (0.026)	**0.145** (0.036)	**0.205** (0.038)	**0.236** (0.036)	**0.203** (0.043)	**0.111** (0.034)

Table 5: MM and MMQ model parameter estimates at the investigated quantile levels. Boldface denote statistical significance at the 5% level.

u	Variable	MQRE				
		0.1	0.25	0.5	0.75	0.9
$\mathbf{u}_1$	Intercept	**443.356** (3.087)	**462.128** (3.371)	**484.141** (3.925)	**509.519** (4.738)	**537.090** (6.084)
	Class size	$-$**6.776** (2.507)	$-$**7.356** (2.427)	$-$**8.510** (2.719)	$-$**10.935** (3.396)	$-$**14.566** (4.445)
	Teacher Experience	**0.575** (0.207)	**0.587** (0.238)	**0.559** (0.281)	0.552 (0.339)	0.667 (0.484)
	ICC	0.139	0.197	0.219	0.176	0.102
$\mathbf{u}_2$	Intercept	**606.406** (3.180)	**626.375** (3.600)	**650.308** (4.352)	**677.589** (5.437)	**707.654** (6.795)
	Class size	$-$**7.809** (2.876)	$-$**8.683** (2.661)	$-$**10.652** (2.935)	$-$**13.723** (3.697)	$-$**16.234** (4.735)
	Teacher Experience	**0.640** (0.244)	**0.676** (0.257)	**0.673** (0.307)	0.700 (0.392)	0.685 (0.549)
	ICC	0.158	0.218	0.238	0.185	0.104
$\mathbf{u}_3$	Intercept	**411.796** (1.542)	**422.011** (1.808)	**434.499** (2.456)	**448.999** (3.486)	**466.232** (4.702)
	Class size	$-$**4.507** (1.468)	$-$**5.110** (1.442)	$-$**5.984** (1.651)	$-$**7.122** (2.241)	$-$**8.091** (3.082)
	Teacher Experience	**0.291** (0.135)	**0.338** (0.139)	0.362 (0.185)	0.420 (0.271)	0.475 (0.364)
	ICC	0.149	0.214	0.246	0.210	0.118

Table 6: MQRE model parameter estimates and ICC values at the investigated quantile levels. Boldface denote statistical significance at the 5% level.

(see Figure 4).

There are several interesting findings. The contours for smaller τ capture the effects of students who perform exceptionally well on mathematics and reading or exceptionally poorly on mathematics and reading. Meanwhile, the contours for larger τ capture the effects for students at the center of the distribution i.e. those who do not stand out from their peers. The larger contours are affected by abnormal observations while the smaller ones are less sensitive to outliers. The elongated and positively oriented contours indicate that there is more variability in the mathematics scores and confirm the existence of positive covariation between reading and mathematics grades. It can also be easily seen that both M-contours and quantile contours shift up and to the right as years of teaching experience increase which

u	Variable	LQMM				
		0.1	0.25	0.5	0.75	0.9
$\mathbf{u}_1$	Intercept	**455.487** (6.998)	**476.333** (4.855)	**484.779** (3.848)	**492.306** (5.117)	**501.977** (5.568)
	Class size	**−10.799** (2.837)	**−9.739** (2.728)	**−8.500** (2.555)	**−6.847** (2.931)	**−7.416** (3.213)
	Teacher Experience	0.066 (0.341)	−0.000 (0.285)	0.594 (0.345)	**1.081** (0.366)	**1.215** (0.540)
	σ^2_{school}	23.989	18.661	17.135	20.025	29.015
$\mathbf{u}_2$	Intercept	**589.880** (7.261)	**640.025** (4.973)	**651.089** (4.067)	**677.652** (6.964)	**687.842** (8.783)
	Class size	**−6.204** (3.021)	**−10.089** (2.978)	**−10.114** (2.792)	**−12.199** (3.510)	**−11.411** (4.311)
	Teacher Experience	**0.791** (0.321)	0.296 (0.321)	0.572 (0.327)	**0.886** (0.425)	**0.984** (0.450)
	σ^2_{school}	16.763	20.810	20.484	25.208	37.518
$\mathbf{u}_3$	Intercept	**420.779** (4.365)	**431.182** (3.248)	**432.613** (2.810)	**442.018** (3.175)	**455.566** (4.721)
	Class size	**−6.854** (1.854)	**−6.865** (1.771)	**−6.338** (1.792)	**−5.072** (2.025)	−4.573 (2.652)
	Teacher Experience	−0.000 (0.235)	0.237 (0.180)	0.215 (0.217)	0.482 (0.248)	**0.692** (0.341)
	σ^2_{school}	14.834	11.526	12.451	13.452	20.529

Table 7: LQMM parameter estimates at the investigated quantile levels. Standard errors are computed via block bootstrap using 500 resamples. Boldface denote statistical significance at the 5% level.

highlights the centrality of teachers' skills and attitudes in students' achievement. Most importantly, the results obtained suggest that both high-performing students and those who are at risk of failure benefit from smaller classes, as this generates substantial gains in the two subjects (Finn & Achilles 1999, Biddle & Berliner 2002, Guggisberg 2019) at both $\tau = 0.005$ and $\tau = 0.1$ levels. Further, one observes that the quantile contours in Figure 4 are closer to the convex hull of the sample data and the enclosed areas are greater than those produced by the M-quantile contours. Finally, the presented confidence regions give an insight into the estimation uncertainty, which is higher in sparse regions of the data as the size of these envelopes is much larger at $\tau = 0.005$ than $\tau = 0.1$. Also, they are helpful to detect, possible, conditional outliers in the multivariate space, identified as those points that fall outside the estimated fence and are located far away from the bulk of the data.

7 Conclusions

In the univariate setting, M-quantiles (Breckling & Chambers 1988) allow to target different parts of the distribution of the response given the covariates instead of just the expected value of the conditional distribution of the outcome variable. The Huber M-quantiles (Huber 1964) are very versatile because they can trade robustness for efficiency in inference by selecting the tuning constant of the influence function and they offer computational stability because they are based on a continuous influence function (Tzavidis et al. 2016, Bianchi et al. 2018). Unfortunately, M-quantiles have remained relegated to univariate problems due to the lack of a natural ordering in a p-dimensional space, $p > 1$. Yet, an extension to higher dimensions could prove to be very useful role in many fields of applied statistics when the problem being studied involves the characterization of the distribution of a multivariate response.

In the present paper we generalize univariate M-quantile regression to the multivariate setting for the analysis of dependent data. Extending the notion of directional quantiles of

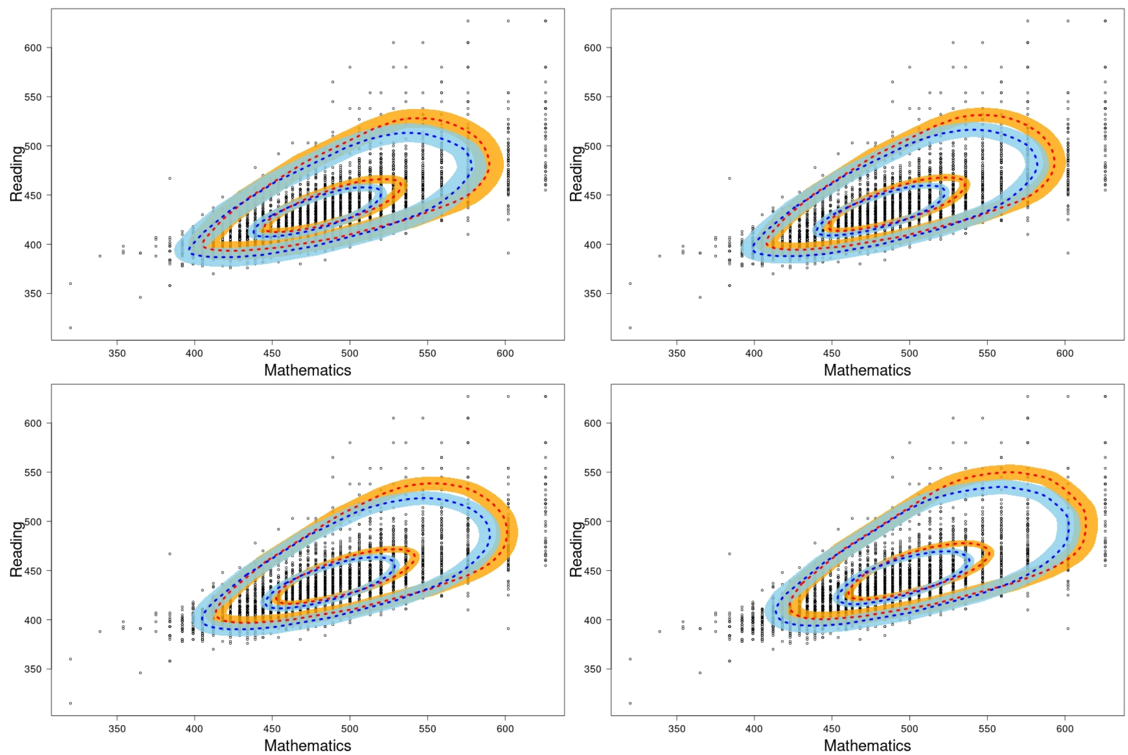

Figure 3: Estimated M-quantile contours at $\tau = (0.005, 0.1)$ for small (red) and large (blue) classes, conditional on the 0.01-th (top-left), 0.25-th (top-right), 0.75-th (bottom-left) and 0.99-th (bottom-right) empirical quantiles of years of teaching experience. The shaded surfaces represent 95% confidence envelopes for M-quantile contours obtained using nonparametric bootstrap.

Kong & Mizera (2012), we introduce directional M-quantiles which are obtained as projections of the original data on a specified unit norm direction. In order to take into consideration the possible within cluster correlation, we develop an M-Quantile Marginal (MMQ) regression model (Liang & Zeger 1986, Zeger & Liang 1986, Heagerty et al. 2000, Diggle et al. 2002). To estimate the model parameters, we extend the well-known GEE approach of Liang & Zeger (1986) and present the robust Generalized M-Quantile Estimating Equations (GMQEE). For a fixed direction, we derive asymptotic properties for the proposed estimator and establish consistency and asymptotic normality. When theoretically all directions are considered simultaneously, the proposed directional approach allows to determine M-quantile regions and contours for a given quantile level. We propose to use M-quantile contour lines to investigate the effect of covariates on the location, spread and shape of the distribution of the responses. To identify potential outliers and provide a simple visual representation of the variability of the M-quantile contours estimator, we construct confidence envelopes via nonparametric bootstrap. Using real data, we apply the MMQ regression model to study the impact of class size and teacher's experience on the joint distribution of the mathematics and reading scores. The obtained results from the fixed-$\mathbf{u}$ and fixed-τ analyses show that small classroom and teacher's experience help improve performance in both subjects.

The methodology can be further extended to take advantage of the longitudinal structure of the STAR study and allow for school effects over time, or cross-classified models to allow for the impact of local area. An interesting research problem would involve the estimation of

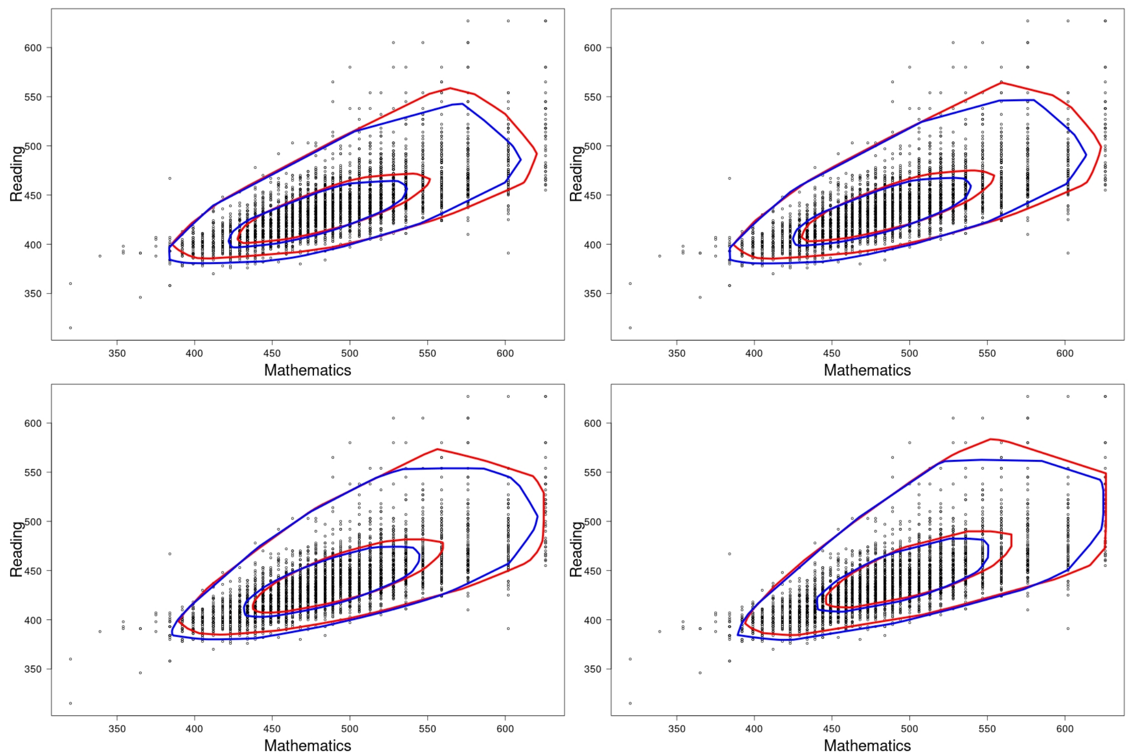

Figure 4: Estimated quantile contours at $\tau = (0.005, 0.1)$ for small (red) and large (blue) classes, conditional on the 0.01-th (top-left), 0.25-th (top-right), 0.75-th (bottom-left) and 0.99-th (bottom-right) empirical quantiles of years of teaching experience.

the proposed M-quantile contours in applications to dependent data, where the contour lines also might vary with time. Lastly, a conditional M-quantile model for robust clustering can be developed where M-quantile contours can help us identify the existence of group structures within the study population.

References

Alfò, M., Marino, M. F., Ranalli, M. G., Salvati, N. & Tzavidis, N. (2021), 'M-quantile regression for multivariate longitudinal data with an application to the Millennium Cohort Study', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **70**(1), 122–146.

Alfò, M., Salvati, N. & Ranallli, M. G. (2017), 'Finite mixtures of quantile and M-quantile regression models', *Statistics and Computing* **27**(2), 547–570.

Bahadur, R. R. (1966), 'A note on quantiles in large samples', *The Annals of Mathematical Statistics* **37**(3), 577–580.

Barry, A., Oualkacha, K. & Charpentier, A. (2018), 'Weighted asymmetric least squares regression for longitudinal data using GEE', *arXiv preprint arXiv:1810.09214* .

Bassett, G. W. & Chen, H.-L. (2002), *Portfolio style: Return-based attribution using quantile regression*, Physica-Verlag HD, Heidelberg, pp. 293–305.

Bergsma, W., Croon, M. A. & Hagenaars, J. A. (2009), *Marginal models: For dependent, clustered, and longitudinal categorical data*, Springer Science & Business Media.

Bernardi, M., Durante, F., Jaworski, P., Petrella, L. & Salvadori, G. (2018), 'Conditional risk based on multivariate hazard scenarios', *Stochastic Environmental Research and Risk Assessment* **32**(1), 203–211.

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

Bianchi, A., Fabrizi, E., Salvati, N. & Tzavidis, N. (2018), 'Estimation and testing in M-quantile regression with applications to small area estimation', *International Statistical Review* **86**(3), 541–570.

Bianchi, A. & Salvati, N. (2015), 'Asymptotic properties and variance estimators of the M-quantile regression coefficients estimators', *Communications in Statistics-Theory and Methods* **44**(11), 2416–2429.

Biddle, B. J. & Berliner, D. C. (2002), 'Small class size and its effects', *Educational Leadership* **59**(5), 12–23.

Borgoni, R., Del Bianco, P., Salvati, N., Schmid, T. & Tzavidis, N. (2018), 'Modelling the distribution of health-related quality of life of advanced melanoma patients in a longitudinal multi-centre clinical trial using M-quantile random effects regression', *Statistical Methods in Medical Research* **27**(2), 549–563.

Breckling, J. & Chambers, R. (1988), 'M-quantiles', *Biometrika* **75**(4), 761–771.

Cascos, I. & Ochoa, M. (2021), 'Expectile depth: Theory and computation for bivariate datasets', *Journal of Multivariate Analysis* **184**, 104757.

Chambers, R., Chandra, H., Salvati, N. & Tzavidis, N. (2014), 'Outlier robust small area estimation', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* pp. 47–69.

Chambers, R. & Tzavidis, N. (2006), 'M-quantile models for small area estimation', *Biometrika* **93**(2), 255–268.

Chen, Y.-C., Genovese, C. R. & Wasserman, L. (2017), 'Density level sets: Asymptotics, inference, and visualization', *Journal of the American Statistical Association* **112**(520), 1684–1696.

Cho, H. (2016), 'The analysis of multivariate longitudinal data using multivariate marginal models', *Journal of Multivariate Analysis* **143**, 481–491.

Crowder, M. (1995), 'On the use of a working correlation matrix in using generalised linear models for repeated measures', *Biometrika* **82**(2), 407–410.

Diggle, P., Diggle, P. J., Heagerty, P., Liang, K.-Y., Heagerty, P. J., Zeger, S. et al. (2002), *Analysis of longitudinal data*, Oxford University Press.

Farcomeni, A. (2012), 'Quantile regression for longitudinal data based on latent Markov subject-specific parameters', *Statistics and Computing* **22**(1), 141–152.

Farcomeni, A., Geraci, M. & Viroli, C. (2020), 'Directional quantile classifiers', *arXiv preprint arXiv:2009.05007* .

Finn, J. D. & Achilles, C. M. (1990), 'Answers and questions about class size: A statewide experiment', *American Educational Research Journal* **27**(3), 557–577.

Finn, J. D. & Achilles, C. M. (1999), 'Tennessee's class size study: Findings, implications, misconceptions', *Educational Evaluation and Policy Analysis* **21**(2), 97–109.

Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. (2012), *Applied longitudinal analysis*, Vol. 998, John Wiley & Sons.

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

Geraci, M., Boghossian, N. S., Farcomeni, A. & Horbar, J. D. (2020), 'Quantile contours and allometric modelling for risk classification of abnormal ratios with an application to asymmetric growth-restriction in preterm infants', *Statistical Methods in Medical Research* **29**(7), 1769–1786.

Geraci, M. & Bottai, M. (2014), 'Linear quantile mixed models', *Statistics and Computing* **24**(3), 461–479.

Giovannetti, G., Sanfilippo, M. & Velucchi, M. (2018), 'Diverse twins: analysing China's impact on Italian and German exports using a multilevel quantile regressions approach', *Applied Economics* **50**(28), 3051–3065.

Goldstein, H. (2011), *Multilevel statistical models*, Vol. 922, John Wiley & Sons.

Guggisberg, M. (2019), A Bayesian approach to multiple-output quantile regression, Technical report, Institute for Defense Analyses.

Haile, G. A. & Nguyen, A. N. (2008), 'Determinants of academic attainment in the United States: A quantile regression analysis of test scores', *Education Economics* **16**(1), 29–57.

Hallin, M., Paindaveine, D., Šiman, M., Wei, Y., Serfling, R., Zuo, Y., Kong, L. & Mizera, I. (2010), 'Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth', *The Annals of Statistics* pp. 635–703.

Heagerty, P. J., Zeger, S. L. et al. (2000), 'Marginalized multilevel models and likelihood inference', *Statistical Science* **15**(1), 1–26.

Heise, L. L. & Kotsadam, A. (2015), 'Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys', *The Lancet Global Health* **3**(6), 332–340.

Hendricks, W. & Koenker, R. (1992), 'Hierarchical spline models for conditional quantiles and the demand for electricity', *Journal of the American statistical Association* **87**(417), 58–68.

Hin, L.-Y. & Wang, Y.-G. (2009), 'Working-correlation-structure identification in generalized estimating equations', *Statistics in Medicine* **28**(4), 642–658.

Huber, P. J. (1964), 'Robust estimation of a location parameter', *Annals of Mathematical Statistics* **35**(1), 73–101.

Huber, P. J. & Ronchetti, E. M. (1981), 'Robust statistics john wiley & sons', *New York* **1**(1).

Huber, P. & Ronchetti, E. (2009), *Robust Statistics*, Wiley.

Israel, G. D., Beaulieu, L. J. & Hartless, G. (2001), 'The influence of family and community social capital on educational achievement', *Rural Sociology* **66**(1), 43–68.

Jones, M. C. (1994), 'Expectiles and M-quantiles are quantiles', *Statistics & Probability Letters* **20**(2), 149–153.

Kelcey, B., Hill, H. C. & Chin, M. J. (2019), 'Teacher mathematical knowledge, instructional quality, and student outcomes: a multilevel quantile mediation analysis', *School Effectiveness and School Improvement* **30**(4), 398–431.

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

Koenker, R. & Bassett Jr, G. (1978), 'Regression quantiles', *Econometrica: Journal of the Econometric Society* pp. 33–50.

Koenker, R., Chernozhukov, V., He, X. & Peng, L. (2017), *Handbook of Quantile Regression*, CRC press.

Kokic, P., Breckling, J. & Lübke, O. (2002), A new definition of multivariate M-quantiles, *in* 'Statistical data analysis based on the L1-norm and related methods', Springer, pp. 15–24.

Kokic, P., Chambers, R., Breckling, J. & Beare, S. (1997), 'A measure of production performance', *Journal of Business & Economic Statistics* **15**(4), 445–451.

Kolen, M. J., Wang, T. & Lee, W.-C. (2012), 'Conditional standard errors of measurement for composite scores using IRT', *International Journal of Testing* **12**(1), 1–20.

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

Korhonen, P. & Siljamäki, A. (1998), 'Ordinal principal component analysis theory and an application', *Computational Statistics & Data Analysis* **26**(4), 411–424.

Kozumi, H. & Kobayashi, G. (2011), 'Gibbs sampling methods for bayesian quantile regression', *Journal of Statistical Computation and Simulation* **81**(11), 1565–1578.

Liang, K.-Y. & Zeger, S. L. (1986), 'Longitudinal data analysis using generalized linear models', *Biometrika* **73**(1), 13–22.

Lin, F., Tang, Y. & Zhu, Z. (2020), 'Weighted quantile regression in varying-coefficient model with longitudinal data', *Computational Statistics & Data Analysis* **145**, 106915.

Lindsey, J. K. (1999), *Models for Repeated Measurements*, Oxford University Press.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. & Zhao, L. P. (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**(4), 463–476.

Lu, X. & Fan, Z. (2015), 'Weighted quantile regression for longitudinal data', *Computational Statistics* **30**(2), 569–592.

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

Marino, M. F., Tzavidis, N. & 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.

Merlo, L., Maruotti, A. & Petrella, L. (2021), 'Two-part quantile regression models for semi-continuous longitudinal data: A finite mixture approach', *Statistical Modelling* p. 1471082X21993603.

Merlo, L., Petrella, L. & Raponi, V. (2021), 'Forecasting VaR and ES using a joint quantile regression and its implications in portfolio allocation', *Journal of Banking & Finance* p. 106248.

Molchanov, I. (2005), *Theory of random sets*, Vol. 19, Springer.

Molchanov, I. & Molinari, F. (2018), *Random sets in econometrics*, Vol. 60, Cambridge University Press.

Molenberghs, G. & Verbeke, G. (2006), *Models for discrete longitudinal data*, Springer Science & Business Media.

Newey, W. K. & Powell, J. L. (1987), 'Asymmetric least squares estimation and testing', *Econometrica: Journal of the Econometric Society* pp. 819–847.

Niemiro, W. et al. (1992), 'Asymptotics for M-estimators defined by convex minimization', *The Annals of Statistics* **20**(3), 1514–1533.

Paindaveine, D. & Šiman, M. (2011), 'On directional multiple-output quantile regression', *Journal of Multivariate Analysis* **102**(2), 193–212.

Pandey, G. R. & Nguyen, V.-T.-V. (1999), 'A comparative study of regression based methods in regional flood frequency analysis', *Journal of Hydrology* **225**(1-2), 92–101.

Prentice, R. L. & Zhao, L. P. (1991), 'Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses', *Biometrics* pp. 825–839.

Reich, B. J., Fuentes, M. & Dunson, D. B. (2011), 'Bayesian spatial quantile regression', *Journal of the American Statistical Association* **106**(493), 6–20.

Salvati, N., Fabrizi, E., Ranalli, M. & Chambers, R. (2021), 'Small area estimation with linked data', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **83**(1), 78–107.

Sinha, S. K. & Rao, J. (2009), 'Robust small area estimation', *Canadian Journal of Statistics* **37**(3), 381–399.

Smith, L. B., Reich, B. J., Herring, A. H., Langlois, P. H. & Fuentes, M. (2015), 'Multilevel quantile function modeling with application to birth outcomes', *Biometrics* **71**(2), 508–519.

Street, J. O., Carroll, R. J. & Ruppert, D. (1988), 'A note on computing robust regression estimates via iteratively reweighted least squares', *The American Statistician* **42**(2), 152–154.

Tzavidis, N., Salvati, N., Pratesi, M. & Chambers, R. (2008), 'M-quantile models with application to poverty mapping', *Statistical Methods and Applications* **17**(3), 393–411.

Tzavidis, N., Salvati, N., Schmid, T., Flouri, E. & Midouhas, E. (2016), 'Longitudinal analysis of the strengths and difficulties questionnaire scores of the Millennium Cohort Study children in England using M-quantile random-effects regression', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **179**(2), 427.

Wang, Y.-G. & Carey, V. (2003), 'Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance', *Biometrika* **90**(1), 29–41.

Wang, Y.-G. & Carey, V. J. (2004), 'Unbiased estimating equations from working correlation models for irregularly timed repeated measures', *Journal of the American Statistical Association* **99**(467), 845–853.

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica: Journal of the Econometric Society* pp. 817–838.

Word, E. R. et al. (1990), 'The State of Tennessee's student/teacher achievement ratio (STAR) Project: Technical Report (1985-1990)'.

Yan, J. & Fine, J. (2004), 'Estimating equations for association structures', *Statistics in Medicine* **23**(6), 859–874.

Yang, C.-C., Chen, Y.-H. & Chang, H.-Y. (2017), 'Joint regression analysis of marginal quantile and quantile association: application to longitudinal body mass index in adolescents', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **66**(5), 1075–1090.

Zeger, S. L. & Liang, K.-Y. (1986), 'Longitudinal data analysis for discrete and continuous outcomes', *Biometrics* pp. 121–130.

Zeger, S. L., Liang, K.-Y. & Albert, P. S. (1988), 'Models for longitudinal data: a generalized estimating equation approach', *Biometrics* pp. 1049–1060.

Zhao, W., Zhang, W. & Lian, H. (2020), 'Marginal quantile regression for varying coefficient models with longitudinal data', *Annals of the Institute of Statistical Mathematics* **72**(1), 213–234.