SAPIENZA
UNIVERSITÀ DI ROMA

On Quantile Regression Models for Multivariate Data

Scuola di Scienze Statistiche

Dottorato di Ricerca in Statistica Metodologica – XXXIV Ciclo

Candidate

Luca Merlo

ID number 1597379

Thesis Advisor

Prof. Lea Petrella

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistical Sciences

January 2022

On Quantile Regression Models for Multivariate Data
Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Version: February 22, 2022

Author's email: luca.merlo@uniroma1.it

Thesis defended on 22 February 2022
in front of a Board of Examiners composed by:

Prof. Salvatore Ingrassia (chairman)
Prof. Maura Mezzetti
Prof. Domenico Vistocco

Acknowledgments

This thesis is the result of four research works I carried forwards during my Ph.D. Throughout this journey, I had the opportunity to work with many people who have contributed to this dissertation and deserve credit for their support and efforts.

First, I primarily wish to express my profound gratitude and sincere thanks to my Advisor Prof. Lea Petrella who has always encouraged and supported me in the course of my studies. Not only she assisted me during all the stages of my thesis, but she has been a constant for when I sought guidance both on a professional and personal front.

Second, my particular thanks go to Prof. Valentina Raponi, Prof. Nikos Tzavidis and Prof. Nicola Salvati for their unconditional help and support. I would like to express my gratitude for the dedication, passion, commitment and patience they showed me. I am also deeply indebted to Prof. Nikos Tzavidis for his hospitality during my research visit at the University of Southampton.

I express my personal utmost gratitude to Prof. Marco Alfò and Prof. Antonello Maruotti for having shared their experience and valuable suggestions with me over the past three years. They both played a constructive part in the work and their support was vital for the accomplishment of this dissertation.

I am grateful to my friends, colleagues and professors at the Department of Statistical Sciences for all that they have taught me and the help that I have received.

Lastly, I thank the reviewers, Prof. Maria Giovanna Ranalli and Prof. Maria Francesca Marino, for their valuable comments and careful review of this thesis.

Abstract

The goal of this thesis is to bridge the gap between univariate and multivariate quantiles by extending the study of univariate quantile regression and its generalizations to multivariate responses. The statistical analysis focuses on a multivariate framework where we consider vector-valued quantile functions associated with multivariate distributions, providing inferential procedures and establishing the asymptotic properties of the proposed estimators. We illustrate their applicability in a wide variety of scientific settings, including time series, longitudinal and clustered data. The dissertation is divided into four chapters, each of them focusing on various aspects of multivariate analysis and different data types and structures. The methodologies we propose are supported by theoretical results and illustrated using simulation studies and real-world data.

Contents

List of Figures

List of Tables

Introduction and Overview

Since the seminal work of Koenker & Bassett (1978), quantile regression has become a widely used technique in many empirical applications. This method allows us to model the whole conditional distribution of a response variable in terms of a set of covariates, offering a more detailed picture of the relationship between the dependent and explanatory variables compared to traditional mean regression. Univariate quantile regression methods are now well established in the statistical literature, either in the frequentist or Bayesian settings, and have been extended to different data structures, including, among others, cross-sectional, time series, multilevel data, survival analysis and high-dimensional data; see, e.g., Yu & Moyeed (2001), Kozumi & Kobayashi (2011), Koenker (2005), Lum et al. (2012), Davino et al. (2013), Bernardi et al. (2015), Marino & Farcomeni (2015), Koenker et al. (2017), Furno & Vistocco (2018), Taylor (2019), Maruotti et al. (2021) and the references therein.

Several generalizations related to the notion of quantiles have also been introduced over the years. One extension is provided by the expectile regression (Newey & Powell 1987), which can be thought of as a "quantile-like" generalization of the classical mean regression. A second important extension is represented by the M-quantile regression of Breckling & Chambers (1988). This method extends the ideas of M-estimation of Huber (1964) based on asymmetric influence functions, combining the robustness and efficiency properties of quantiles and expectiles in a common framework. M-quantile regression models have attracted much research interest and have received increasing attention in both the parametric and non-parametric framework; see, for instance, Chambers & Tzavidis (2006), Tzavidis et al. (2008), Pratesi et al. (2009), Alfò et al. (2017), Bianchi et al. (2018), Alfò et al. (2021).

When multivariate response variables are concerned, the existing literature on quantile methods is less extensive since there is no "natural" ordering in a p-dimensional space, $p > 1$ (Serfling 2002, Chakraborty 2003, Hallin et al. 2010, Kong & Mizera 2012, Paindaveine & Šiman 2012, Koenker et al. 2017, Chavas 2018, Petrella & Raponi 2019). As a consequence, univariate quantile, M-quantile, and neither expectile regression models do straightforwardly extend to higher dimensions. Nevertheless, in most situations of practical interest, the purpose of the matter being investigated lies in describing the distribution of a multivariate response variable. In these cases, a multivariate approach is more appropriate than a univariate one because it takes into account dependence among responses. An extension to the multidimensional setting of these concepts is thus extremely desirable and sought after by researchers and data analysts.

Motivated by the necessity to provide useful tools for characterizing multivariate distributions, the goal of this thesis is to bridge the gap between univariate and

multivariate quantiles by extending the study of univariate quantile regression and its generalizations to multivariate responses. The statistical analysis focuses on a multivariate framework where we consider vector-valued quantile functions associated with multivariate distributions, providing both descriptive and inferential procedures.

When analyzing real-world data, observations are often interconnected with each other across time, space, or other dimensions, like groups, and their analysis demands specific analysis tools. In this context, the research interest may focus not only on investigating the association among responses, but also on accounting for the specific dependence structure embedded in the data. Dependency of observations arises in a wide variety of scientific settings, including time series, longitudinal and clustered data (Diggle et al. 2002, Bergsma et al. 2009, Goldstein 2011, Zucchini et al. 2016, Hamilton 2020). Particularly, in time series where data points are recorded over time, modeling the temporal dependence and serial correlation across observations are a crucial aspect of the analysis. On the other hand, when dealing with longitudinal data, because measurements recorded on the same individuals are likely correlated, the potential association between dependent observations should be taken into account. Either ignoring these factors or relying on untenable modeling assumptions may produce biased and inconsistent parameter estimates, which in turn leads to wrong interpretations and conclusions of the phenomenon under study. In this dissertation, we thus extend univariate quantile methods for modeling the entire conditional distribution of multivariate responses, accounting for the dependence among the outcomes and incorporating all characteristics of such complex data structures.

There are, however, numerous applications where features of the conditional distribution are not objects of direct interest, meanwhile, practitioners wish to determine the effects of relevant predictors on the unconditional distribution of the response. Indeed, an important branch of the literature focuses on the estimation of the unconditional distribution of the dependent variable, spanning from social programs evaluation to the identification of distributional effects for particular drug treatments. When the unconditional distribution is the ultimate research objective, using the univariate conditional quantile regression of Koenker & Bassett (1978) would yield misleading inferences (Firpo et al. 2009, Borah & Basu 2013, Frölich & Melly 2013). In the literature, several proposals have been introduced to estimate these unconditional quantile effects; see Gosling et al. (2000), Machado & Mata (2005), Melly (2005), Firpo et al. (2009). Those studies, however, focus on the univariate regression framework. When the problem under investigation involves multivariate dependent variables, extending such univariate procedures to multivariate techniques is a challenging task, given that there is no natural ordering for multivariate observations. This clearly represents an important gap to be filled in the literature on unconditional effects estimation, highlighting the need for new statistical tools and techniques to treat this problem. In this thesis, we contribute to the current literature extending the univariate unconditional approach of Firpo et al. (2009) to a more general multivariate setting. More specifically, we propose a unified unconditional regression approach that encompasses multivariate quantiles, M-quantiles and expectiles, which allows us to evaluate the impact of changes in the distribution of each explanatory variable across the entire unconditional distribution of the responses.

The dissertation is divided into four chapters, each of them focusing on various aspects of multivariate analysis and different data types and structures. The methodologies we propose are supported by theoretical results and illustrated using simulation studies and real-world data.

In Chapter 1, we introduce a Quantile Mixed Hidden Markov Model for joint estimation of multiple quantiles in multivariate longitudinal data. The approach is based on the Multivariate Asymmetric Laplace (MAL) distribution, which allows to simultaneously model the quantiles of the univariate conditional distributions of a multivariate response in a linear regression framework, accounting for possible correlation between the outcomes. Unobserved heterogeneity sources and serial dependence due to repeated measures are modeled through the introduction of individual-specific, time-constant random coefficients and time-varying parameters evolving over time with a Markovian structure, respectively. The inferential procedure is carried out through a suitable Expectation-Maximization algorithm without parametric assumptions on the random effects distribution and closed form M-step update expressions are derived for all model parameters. The validity of the approach is analyzed both by a simulation study and through the empirical analysis of the UK Millennium Cohort Study data. The content of this chapter is based upon a joint work with Prof. Petrella L. and Prof. Tzavidis N., which has been recently published in Merlo et al. (2022).

Chapter 2 generalizes the MAL quantile regression approach considered in Chapter 1 to a time-varying setting for the analysis of multivariate financial time series. Specifically, extending the univariate work of Taylor (2019), we propose a multiple linear quantile regression for jointly predicting tail risk measures, namely Value at Risk (VaR) and Expected Shortfall (ES). The proposed methodology permits simultaneous modelling of multiple conditional quantiles of a multivariate response variable and accounts for the dependence structure among financial assets. By exploiting the properties of the MAL distribution, we propose a new portfolio optimization method that minimizes portfolio risk and controls for well-known characteristics of financial data. We evaluate the advantages of the proposed approach on both simulated and real data, using weekly returns on three major stock market indices. We show that our method outperforms other existing models and provides more accurate risk measure forecasts than univariate methods. The methods and findings of this analysis are contained and published in Merlo, Petrella & Raponi (2021).

In Chapter 3, we develop 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 the 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. The considered modeling framework is the result of a collaborative work with Prof. Petrella L., Prof. Tzavidis N. and Prof. Salvati N., and it is currently under revision in a high-quality journal.

Finally, in Chapter 4 we build a unified regression approach to model unconditional quantiles, M-quantiles and expectiles of multivariate dependent variables exploiting the multidimensional Huber's function. To assess the impact of changes in the covariates across the entire unconditional distribution of the responses, we extend the work of Firpo et al. (2009) by running a mean regression of the recentered influence function on the explanatory variables. We discuss the estimation procedure and establish the asymptotic properties of the derived estimators. A data-driven procedure is also presented to select the optimal tuning constant of the Huber's function. The validity of the proposed methodology is explored with simulation studies and through an application using the Survey of Household Income and Wealth 2016 conducted by the Bank of Italy. The content of this chapter was written together with Prof. Petrella L., Prof. Tzavidis N. and Prof. Salvati N., and it has been submitted to an international journal for publication.

Chapter 1

Quantile Mixed Hidden Markov Models for multivariate longitudinal data

1.1 Introduction

Ever since quantile regression was first introduced in the seminal work of Koenker & Bassett (1978), it has attracted researches' and practitioners' attention. It provides a way to model the conditional quantiles of a response variable with respect to a set of covariates in order to have a more complete picture of the entire conditional distribution compared to the classical mean regression. In a univariate quantile regression framework, both the classical and Bayesian inferential approaches have been proposed in the literature to estimate the model parameters. In the frequentist setting, the inferential approach used to estimate the parameters relies on the minimization of the asymmetric loss function (see Koenker & Bassett 1978) while, in the Bayesian setting, and in a likelihood inferential approach, the Asymmetric Laplace (AL) distribution has been introduced as a likelihood inferential tool. The two approaches are well-justified by the relationship between the quantile loss function and the AL density. Indeed, Yu & Moyeed (2001) showed that the minimization of the quantile loss function is equivalent, in terms of parameter estimates, to the maximization of the likelihood associated with the AL density. Therefore, the AL distribution may offer a convenient approach to make inference in a quantile regression analysis. Quantile regression methods have become widely used in literature because they are suitable in those situations where skewness, fat-tails, outliers, truncation, censoring and heteroscedasticity arise. They have been implemented in a wide range of different fields, both in a frequentist paradigm and in a Bayesian setting, spanning from medicine (see Cole & Green 1992, Royston & Altman 1994, Reich et al. 2011 and Waldmann 2018), financial and economic research (see Bassett & Chen 2002, Kozumi & Kobayashi 2011, Bernardi et al. 2015, Petrella et al. 2018, Laporta et al. 2018, Tian et al. 2018, Bernardi, Bottone & Petrella 2018 and Petrella & Raponi 2019), and environmental modeling (Hendricks & Koenker 1992, Pandey & Nguyen 1999 and Reich et al. 2011). For a detailed review and list of references, Koenker (2005) and Koenker et al. (2017) provide an overview of the most used quantile

regression techniques in a classical setting. In longitudinal studies, quantile methods with random effects have been positively considered in order to account for the dependence between serial observations on the same subject (see Koenker 2004, Geraci & Bottai 2006, Farcomeni 2012, Luo et al. 2012, Marino & Farcomeni 2015, Alfò et al. 2017, Marino et al. 2018, Kulkarni et al. 2019, Merlo, Maruotti & Petrella 2021 and Alfò et al. 2021).

When multivariate response variables are concerned, the existing literature on quantile regression is less extensive due to the fact that there is not a unique definition of quantile for a multivariate random variable since there is no "natural" ordering in a p-dimensional space, for $p > 1$. As a consequence, the univariate quantile regression method does not straightforwardly extend to higher dimensions. Nevertheless, in most situations of practical interest, the purpose of the matter being investigated lies in describing the distribution of a multivariate response variable. In these cases, a multivariate approach is more appropriate than a univariate one because it takes into account dependence among marginals. An extension to the multidimensional setting of the definition of quantile above is thus desirable. For this reason, the search for a satisfactory notion of multivariate quantile has led to a flourishing literature on this topic despite its definition is still a debatable issue (see Chakraborty 2003, Hallin et al. 2010, Kong & Mizera 2012, Koenker et al. 2017, Stolfi et al. 2018, Chavas 2018, Charlier et al. 2020 and the references therein for relevant studies).

Recently, Petrella & Raponi (2019) generalized the AL distribution inferential approach of the univariate quantile regression to a multivariate framework by using the Multivariate Asymmetric Laplace (MAL) distribution defined in Kotz et al. (2012). By using the MAL distribution as a likelihood based inferential tool, the authors sidestep the problem of defining the quantiles of a multivariate distribution, and instead implement joint estimation for the univariate quantiles of the conditional distribution of a multivariate response variable given covariates, accounting for possible correlation among the responses.

When dealing with longitudinal data, because measurements recorded on the same individuals are likely correlated, the potential association between dependent observations should be taken into account in order to provide correct inferences. In such cases, random effect models have been advocated to accommodate for time-constant, within-subject correlation and between subject heterogeneity (see Liu & Bottai 2009 and Geraci & Bottai 2014): time-constant individual-specific random coefficients are added in the regression model to capture this unobserved heterogeneity. However, when the assumption that heterogeneity is constant over time does not hold, adopting such model specification may lead to biased parameter estimates (see Bartolucci & Farcomeni 2009). To account for serial heterogeneity, Farcomeni (2012) suggested the use of Hidden Markov Models (HMM). In such a context, a latent homogeneous Markov chain is defined in order to capture the temporal evolution of unobserved heterogeneity and state-dependent parameters are introduced to account for response variability due to time-varying omitted covariates. The application of HMMs is well justified by their versatility and mathematical tractability in longitudinal studies where the evolution of a latent individual characteristic is of interest (see Cappé et al. 2006, Maruotti 2011, Maruotti & Rocci 2012 and Zucchini et al. 2016).

In real data applications, unobserved heterogeneity may both evolve and/or stay constant over time. In addition, the available covariates may not be able to capture all the individual heterogeneity sources. In this case, a set of random effects should be included in the model specification to capture such unobserved differences between individuals (Maruotti & Rocci 2012). In order to handle such a complex data structure, it is possible to consider the well-known Mixed Hidden Markov Models (MHMMs, see Altman 2007). The MHMM, obtained by combining the features of hidden Markov and Mixed Effects Models, encompasses Generalized Linear Mixed Models and HMMs as it accommodates time-constant and time-varying sources of random variation jointly. In the application of quantile regression to longitudinal data, Marino et al. (2018) introduced a Mixed Hidden Markov quantile regression model for longitudinal continuous responses, extending the Linear Quantile Mixed Model of Geraci & Bottai (2014) and the Linear Quantile Hidden Markov Model of Farcomeni (2012). These proposals are, however, designed for univariate dependent variables and consequently, they neglect the dependence structure between multiple outcomes of interest measured over each unit.

The purpose of this article is to extend the work of Petrella & Raponi (2019) by introducing a Mixed Hidden Markov Model to the longitudinal data setting. We develop a Quantile Mixed Hidden Markov Model (QMHMM) to jointly estimate the quantiles of the univariate conditional distributions of a multivariate response, accounting for the dependence structure between the outcomes. In particular, time-constant unobserved heterogeneity is described via individual-specific random coefficients while temporal effects are captured through state-specific parameters that evolve over time depending on a hidden Markov chain. In order to prevent inconsistent parameter estimates due to misspecification of the random effects distribution, we adopt the Non-Parametric Maximum Likelihood (NPML) approach of Lindsay et al. (1983) where the distribution is left unspecified and approximated by a multivariate discrete finite mixture distribution estimated from the data. Within this scheme, our modeling framework reduces to a multivariate finite mixture of HMM quantile regressions. We propose to estimate the model parameters through Maximum Likelihood (ML) by using the MAL distribution as working likelihood in a regression framework. Specifically, as in Petrella & Raponi (2019) we consider a reparameterization of the MAL distribution, subject to some specific constraints, which allows us to estimate the regression coefficients via ML. In particular, we build an Expectation-Maximization (EM) algorithm which exploits the Gaussian location-scale mixture representation of the MAL distribution where both the hidden Markov chain and the random effects parameters are treated as missing data. From a computational perspective, we provide an efficient version of the EM algorithm with M-step updates in closed form for all model parameters.

Using simulation experiments we assess the validity of our approach by considering different data generating processes. Moreover, we apply our methodology to real data by analysing the Millennium Cohort Study (MCS). The MCS is a longitudinal birth cohort study following children born in the UK, providing multiple measures of the cohort members' physical, socio-emotional, cognitive and behavioural development over time. We develop a QMHMM for children's emotional and behavioral disorders as a function of demographic and socio-economics risk factors taking into account both the potential dependence between children's disorders, the time-constant and time-

varying unobserved heterogeneity. The proposed approach for modeling conditional quantiles simultaneously, can offer a considerably insight to child psychologists on the effect of selected risk factors on children's behavioural problems.

The rest of the paper is organized as follows. In Section 1.2, we introduce the proposed QMHMM regression framework. Section 1.3 illustrates the EM-based ML approach to estimate model parameters together with M-step updates in closed form. In Section 1.4 we present the simulation results while Section 1.5 discusses the empirical application. Section 1.6 summarizes our conclusions. All the proofs are provided in Appendix.

1.2 Methodology

Let $\mathbf{Y}_{it} = (Y_{it}^{(1)}, \ldots, Y_{it}^{(p)})$ be a continuous p-variate response variable vector and $\mathbf{X}_{it} = (X_{it}^{(1)}, \ldots, X_{it}^{(k)})$ be a k-dimensional vector of explanatory variables for subject $i = 1, \ldots, N$ and time occasion $t = 1, \ldots, T_i$. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$ denote p quantile indexes with $\tau_j \in (0, 1)$, for $j = 1, \ldots, p$. In particular, the p quantile indexes do not need to be the same for all of the elements of $\mathbf{Y}_{it}$. Further, let $S_{it}(\boldsymbol{\tau})$, $i = 1, \ldots, N$, $t = 1, \ldots, T_i$ be a homogeneous, first-order, aperiodic and irreducible hidden Markov chain defined over a discrete states space $\mathcal{S} = \{1, \ldots, M\}$ with initial and transition probabilities denoted by $\mathbf{q} = (q_1, \ldots, q_M)$ and $\mathbf{Q} = \{q_{jk}\}$ over $\mathcal{S} \times \mathcal{S}$ common to all subjects, respectively. Finally, let $\mathbf{b}_i(\boldsymbol{\tau})$ be a time-constant, subject-specific, random effects matrix having distribution $f_{\mathbf{b}}(\cdot \mid \mathbf{X}_{it}, \boldsymbol{\tau})$ with support $\mathcal{B}$, where $\mathbb{E}(\mathbf{b}_i(\boldsymbol{\tau})) = 0$ is used for parameter identifiability. We assume that the τ_j-th quantile of each of the j-th components of $\mathbf{Y}_{it}$ can be modeled as a function of explanatory variables. Let $\boldsymbol{\beta}(\boldsymbol{\tau}) = (\boldsymbol{\beta}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{\beta}_p(\boldsymbol{\tau}))$ be the $k \times p$ matrix of unknown quantile regression coefficients. Then, the QMHMM is defined as follows:

$$\mathbf{Y}_{it} = \mathbf{X}_{it}\boldsymbol{\beta}(\boldsymbol{\tau}) + \mathbf{Z}_{it}\mathbf{b}_i(\boldsymbol{\tau}) + \mathbf{W}_{it}\boldsymbol{\alpha}_{S_{it}}(\boldsymbol{\tau}) + \epsilon_{it}(\boldsymbol{\tau}) \tag{1.1}$$

where $\mathbf{Z}_{it}$ is a subset of $\mathbf{X}_{it}$, $\mathbf{W}_{it}$ is a further subset of $\mathbf{X}_{it}$ whose effects are assumed to vary over time, $\epsilon_{it}(\boldsymbol{\tau})$ denotes a p-dimensional vector of error terms with univariate component-wise quantiles (at fixed levels $\tau_1, \ldots, \tau_p$, respectively) equal to zero and where the coefficients matrix $\boldsymbol{\alpha}_{S_{it}}(\boldsymbol{\tau})$ evolves over time according to the hidden Markov chain, $S_{it}(\boldsymbol{\tau})$, and takes one of the values in the set $\{\boldsymbol{\alpha}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{\alpha}_M(\boldsymbol{\tau})\}$. In particular, the parameters $\mathbf{b}_i(\boldsymbol{\tau}), i = 1, \ldots, N$, and $\{\boldsymbol{\alpha}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{\alpha}_M(\boldsymbol{\tau})\}$ are designed to account for within-individual dependence by considering unobserved time-constant and time-varying sources of unobserved heterogeneity, respectively.

Our objective is to provide joint estimation of the p quantiles of the univariate conditional distributions of $\mathbf{Y}_{it}$ taking into account for potential correlation among the dependent variables. The QMHMM framework is based on the following central assumptions, which are standard in mixed effects models. The random effects $\mathbf{b}_i(\boldsymbol{\tau})$ are independent of the hidden Markov chain, $S_{it}(\boldsymbol{\tau})$, as they are meant to capture different unobserved characteristics, and furthermore, it is assumed that the covariates $\mathbf{X}_{it}$ are uncorrelated with $\mathbf{b}_i(\boldsymbol{\tau})$, that is $f_{\mathbf{b}}(\cdot \mid \mathbf{X}_{it}, \boldsymbol{\tau}) = f_{\mathbf{b}}(\cdot \mid \boldsymbol{\tau})$. Regarding the longitudinal responses, they must satisfy the contemporary dependence and conditional independence conditions. The former states that for the i-th subject at time t, the distribution of $\mathbf{Y}_{it}$, given the state variables $(S_{i1}(\boldsymbol{\tau}), ..., S_{iT_i}(\boldsymbol{\tau}))$ and the

time-constant individual-specific random effects $\mathbf{b}_i(\boldsymbol{\tau})$, depends only on the current state $S_{it}(\boldsymbol{\tau})$; the latter entails that the responses $(\mathbf{Y}_{i1}, ..., \mathbf{Y}_{iT_i})$ are conditionally independent, given the hidden state occupied at time t by $S_{it}(\boldsymbol{\tau})$ and the individual-specific random coefficients $\mathbf{b}_i(\boldsymbol{\tau})$. These assumptions imply that the following equality holds:

$$f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{y}_{i1:t-1}, \mathbf{x}_{i1:t-1}, s_{i1:t}, \mathbf{b}_i, \boldsymbol{\tau}) = f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_i, \boldsymbol{\tau}) \tag{1.2}$$

where $\mathbf{y}_{i1:t-1}$ and $\mathbf{x}_{i1:t-1}$ represent the history of the responses and the observed covariates for the i-th subject up to time $t-1$, respectively, and $s_{i1:t}$ is the individual sequence of states up to time t.

Generalizing the approach of Petrella & Raponi (2019), for the model in (1.1) we consider the MAL distribution, $\mathcal{MAL}(\boldsymbol{\mu}, \mathbf{D}\tilde{\xi}, \mathbf{D\Sigma D})$, (see Kotz et al. 2012) having density function:

$$f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_i, \boldsymbol{\tau}) = \frac{2\exp\left\{(\mathbf{y}_{it} - \boldsymbol{\mu}_{it})'\mathbf{D}^{-1}\boldsymbol{\Sigma}^{-1}\tilde{\xi}\right\}}{(2\pi)^{p/2}|\mathbf{D\Sigma D}|^{1/2}} \left(\frac{\tilde{m}_{it}}{2+\tilde{d}}\right)^{\nu/2} K_{\nu}\left(\sqrt{(2+\tilde{d})\tilde{m}_{it}}\right), \tag{1.3}$$

where the location parameter $\boldsymbol{\mu}_{it}$ is defined by the Mixed Hidden Markov Model:

$$\boldsymbol{\mu}_{it} = \boldsymbol{\mu}(s_{it}, \mathbf{b}_i, \boldsymbol{\tau}) = \mathbf{X}_{it}\boldsymbol{\beta}(\boldsymbol{\tau}) + \mathbf{Z}_{it}\mathbf{b}_i(\boldsymbol{\tau}) + \mathbf{W}_{it}\boldsymbol{\alpha}_{s_{it}}(\boldsymbol{\tau}), \tag{1.4}$$

$\mathbf{D}\tilde{\xi}$ is the skew parameter with $\mathbf{D} = \text{diag}[d_1, \ldots, d_p]$, $d_j > 0$ and $\tilde{\xi} = [\tilde{\xi}_1, \ldots, \tilde{\xi}_p]'$ having generic element $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$, $j = 1, \ldots, p$. $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix such that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Psi\Lambda}$, with $\boldsymbol{\Psi}$ being an unstructured correlation matrix of dimension p and $\boldsymbol{\Lambda} = \text{diag}[\sigma_1, \ldots, \sigma_p]$, with $\sigma_j^2 = \frac{2}{\tau_j(1-\tau_j)}$, $j = 1, \ldots, p$. Moreover, $\tilde{m}_{it} = (\mathbf{y}_{it} - \boldsymbol{\mu}_{it})'(\mathbf{D\Sigma D})^{-1}(\mathbf{y}_{it} - \boldsymbol{\mu}_{it})$, $\tilde{d} = \tilde{\xi}'\boldsymbol{\Sigma}^{-1}\tilde{\xi}$, and $K_{\nu}(\cdot)$ denotes the modified Bessel function of the third kind with index parameter $\nu = (2-p)/2$.

One of the key benefits of the MAL distribution is that, using (1.1) and (1.3), and following Kotz et al. (2012), $\mathbf{Y} \sim \mathcal{MAL}(\boldsymbol{\mu}, \mathbf{D}\tilde{\xi}, \mathbf{D\Sigma D})$ can be written as a location-scale mixture, having the following representation:

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{D}\tilde{\xi}\tilde{C} + \sqrt{\tilde{C}}\mathbf{D}\boldsymbol{\Sigma}^{1/2}\mathbf{Z} \tag{1.5}$$

where $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ denotes a p-variate standard Normal distribution and $\tilde{C} \sim \text{Exp}(1)$ has a standard exponential distribution, with $\mathbf{Z}$ being independent of $\tilde{C}$. In particular, the constraints imposed on $\tilde{\xi}$ and $\boldsymbol{\Lambda}$ represent necessary conditions for model identifiability for any fixed quantile level $\tau_1, \ldots, \tau_p$ and guarantee that $\mu_{it}^{(j)}$ is the τ_j-th conditional quantile function of $Y_{it}^{(j)}$ given $S_{it}(\boldsymbol{\tau})$ and $\mathbf{b}_i$, for $j = 1, \ldots, p$. This allows researchers to jointly study the occurrence of extreme values of several dependent variables, by considering either low or high quantile levels for all outcomes or to focus on different quantile levels when the interpretation of the responses runs in opposite directions. That is, high values of some variables are associated with at-risk situations while high values of the others correspond to low-risk situations, making it reasonable to set different quantile levels, $\tau_1, \ldots, \tau_p$, for each marginal $Y_{it}^{(j)}, j = 1, \ldots, p$.

As shown in Petrella & Raponi (2019), using this approach we are able to conduct inference on the quantiles of the univariate conditional distributions of $\mathbf{Y}_{it}$

simultaneously, taking into account the possible correlation between the outcomes. For a given quantile level $\boldsymbol{\tau}$, following Kotz et al. (2012) and by simple calculations it is possible to show that the covariance matrix of $\mathbf{Y}$ can be written as:

$$\mathbf{S} = \mathbf{D}(\tilde{\xi}\tilde{\xi}' + \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda})\mathbf{D}, \tag{1.6}$$

where the off-diagonal elements of $\mathbf{S}$ provide an indirect measure of association between the outcomes.

Two remarks are also noteworthy regarding the methodology introduced above. First, our model can be thought of as an extension to multivariate longitudinal data of: (i) the Linear Quantile Hidden Markov Model by Farcomeni (2012) when $\mathbf{W}_{it} = \mathbf{1}$ and $\mathbf{b}_i(\boldsymbol{\tau}) = \mathbf{0}$ for all $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$; (ii) the Linear Quantile Mixed Model (LQMM) proposed in Geraci & Bottai (2014) when there is only one state of the hidden Markov chain, i.e. $M = 1$. Also, our proposal reduces to the linear quantile mixed hidden Markov model of Marino et al. (2018) in the case of a single response variable when $p = 1$. Second, the proposed approach differs substantially from the ones by Kulkarni et al. (2019) and Alfò et al. (2021). In the former, the authors consider univariate quantile regression models where the dependence across time and responses is captured by time-constant outcome-specific normally distributed random coefficients. In the latter, the proposed method targets a different set of location parameters, i.e. the M-quantiles (Breckling & Chambers 1988) of the distribution of the dependent variables, which are more difficult to interpret than quantiles. The authors then define univariate M-quantile regression models with outcome-specific random effects, where dependence between outcomes for each unit is introduced by assuming correlated, subject-specific random effects in the univariate models.

Estimation of model parameters can be pursued using a ML approach. To ease the notation, unless specified otherwise, hereinafter we omit the quantile levels vector $\boldsymbol{\tau}$, yet all model parameters are allowed to depend on the p quantile indexes. Thus, let us denote by $\boldsymbol{\Phi}_{\boldsymbol{\tau}} = (\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\Psi}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M, \mathbf{q}, \mathbf{Q})$ the set of model parameters. Given the modeling assumptions introduced so far, the observed data likelihood is defined by:

$$L(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) = \prod_{i=1}^{N} \int_{\mathcal{B}} \left\{ \sum_{\mathcal{S}^{T_i}} \left[\prod_{t=1}^{T_i} f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_i) \right] q_{s_{i1}} \prod_{t=2}^{T_i} q_{s_{it-1}s_{it}} \right\} f_{\mathbf{b}}(\mathbf{b}_i) \mathrm{d}\mathbf{b}_i. \tag{1.7}$$

The maximization of the likelihood in (1.7) generally may prove to be excessively cumbersome because it involves a multidimensional integral over the random coefficients distribution $f_{\mathbf{b}}(\cdot)$ and a summation over M^{T_i} terms for each unit. In addition, the choice of an appropriate distribution for the random effects is not straightforward. Ideally $f_{\mathbf{b}}(\cdot)$ should be data driven and resistant to misspecification (Marino & Farcomeni 2015), otherwise an incorrect distributional assumption for the random effects has unfavorable influence on statistical inferences (see Agresti et al. 2004, Maruotti 2011 and Neuhaus et al. 2013). In the next section, we discuss how we specify the random effects distribution and how we may avoid evaluating the integral in (1.7) for ML estimation.

1.2.1 Specification of the random coefficients distribution

In the literature, typically the Gaussian distribution is a convenient choice for $f_{\mathbf{b}}(\cdot)$ from a computational point of view. In this case, we may approximate the integral in (1.7) using Gaussian quadrature or adaptive Gaussian quadrature schemes (see Rabe-Hesketh et al. 2005, Pinheiro & Chao 2006 and Crowther et al. 2014). A disadvantage of such approaches lies in the required computational effort, which is exponentially increasing with the number of the random parameters. For this reason, potential alternatives proposed the use of simulation methods such as Monte Carlo and simulated ML approaches (McCulloch 1997). However, for samples of finite size and short individual sequences, these methods may not provide a good approximation of the true mixing distribution (Alfò et al. 2017). As a robust alternative to the Gaussian choice, the multivariate Symmetric Laplace or multivariate Student t distributions have been considered by Geraci & Bottai (2014) and Farcomeni & Viviani (2015). However, a parametric assumption on the distribution of the random coefficients could be rather restrictive and misspecification of the mixing distribution can lead to biased parameter estimates (see Alfò & Maruotti 2010). Following Marino et al. (2018), in this work we exploit the approach based on the Non-Parametric Maximum Likelihood (NPML) estimation of Laird (1978) and extend it to the multivariate context. In particular, we do not parametrically specify $f_{\mathbf{b}}(\cdot)$ but we approximate it by using a discrete distribution defined on $G < N$ multivariate locations, $\mathbf{b}_g(\boldsymbol{\tau})$, with associated probabilities defined by:

$$\pi_g(\boldsymbol{\tau}) = \Pr(\mathbf{b}_i(\boldsymbol{\tau}) = \mathbf{b}_g(\boldsymbol{\tau})), \tag{1.8}$$

with $\pi_g \geq 0$, $\forall\ g = 1, \ldots, G$ and $\sum_{g=1}^{G} \pi_g = 1$. More concisely, we can write:

$$\mathbf{b}_i(\boldsymbol{\tau}) \sim \sum_{g=1}^{G} \pi_g(\boldsymbol{\tau}) \delta_{\mathbf{b}_g}(\boldsymbol{\tau}), \tag{1.9}$$

where δ_θ is a one-point distribution putting a unit mass at θ. With this approach, the parametric problem is thus converted to a semiparametric one, where $\mathbf{b}_g(\boldsymbol{\tau})$ and $\pi_g(\boldsymbol{\tau})$ define the discrete probability distribution of the random effects defined on G distinct support points. In this context, time-constant unobserved heterogeneity in the data is represented by a finite mixture with unknown proportions $\pi_g(\boldsymbol{\tau})$ and locations $\mathbf{b}_g(\boldsymbol{\tau})$ common to all subjects in the g-th group. Since locations and masses are completely free to vary over the corresponding support, this is a flexible method that can readily accommodate a wide range of shapes, including fat-tailed and asymmetric distributions, and it is more robust against deviations from model assumptions (for a detailed survey about this method see also Aitkin & Alfó 1998, Alfò & Aitkin 2000, Aitkin & Alfò 2003, Alfò et al. 2017, 2021 and Merlo, Maruotti & Petrella 2021).

In this setting, the observed data likelihood in (1.7) reduces to:

$$L(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) = \prod_{i=1}^{N} \sum_{g=1}^{G} \left\{ \sum_{\mathcal{S}^{T_i}} \left[\prod_{t=1}^{T_i} f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_g) \right] q_{s_{i1}} \prod_{t=2}^{T_i} q_{s_{it-1}s_{it}} \right\} \pi_g, \tag{1.10}$$

where $\boldsymbol{\Phi}_{\boldsymbol{\tau}} = (\boldsymbol{\beta}, \mathbf{D}, \boldsymbol{\Psi}, \mathbf{b}_1, \ldots, \mathbf{b}_G, \pi_1, \ldots, \pi_G, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M, \mathbf{q}, \mathbf{Q})$ denotes the vector of model parameters and $f_{\mathbf{Y}}(\mathbf{y}_{it} \mid \mathbf{x}_{it}, s_{it}, \mathbf{b}_g)$ represents the response distribution

of unit i being in the state s_{it} at time t and belonging to the g-th component of the finite mixture, which is assumed to follow the MAL as in (1.3) with location parameter given by:

$$\boldsymbol{\mu}_{it} = \boldsymbol{\mu}(s_{it}, \mathbf{b}_g, \boldsymbol{\tau}) = \mathbf{X}_{it}\boldsymbol{\beta}(\boldsymbol{\tau}) + \mathbf{Z}_{it}\mathbf{b}_g(\boldsymbol{\tau}) + \mathbf{W}_{it}\boldsymbol{\alpha}_{s_{it}}(\boldsymbol{\tau}). \tag{1.11}$$

By looking at the likelihood in (1.10), one can see that it resembles the likelihood of a finite mixture of HMM models with G homogeneous sub-populations where the presence of latent time-constant heterogeneity is described by discrete multivariate random effects. From the estimation perspective, locations $\mathbf{b}_g$ and corresponding probabilities π_g are unknown parameters which need to be estimated along with other model parameters. The number of mixture components G is unknown, and it is usually treated as fixed and estimated via penalized likelihood criteria (see e.g. Böhning 1999). Furthermore, as an important by-product, the computational complexity of the likelihood evaluation in (1.10) is of linear order with respect to G, which greatly facilitates the implementation of EM-type algorithm, as is described in the following section.

1.3 Maximum Likelihood estimation and inference

As mentioned in the previous sections, the MAL density represents a convenient tool to jointly model the univariate quantiles of the conditional distribution of a multivariate response variable in a quantile regression framework. In this section we introduce a ML approach to estimate and make inference on model parameters and build a suitable EM algorithm (Dempster et al. 1977). We will show that the M-step update of all model parameters can be easily obtained in closed form, hence reducing the computational burden of the algorithm compared to direct maximization of the likelihood in (1.10). Specifically, we derive the EM algorithm by exploiting the Gaussian location-scale mixture representation in (1.5) of the MAL distribution under the constraints on $\tilde{\xi}$ and $\boldsymbol{\Lambda}$.

1.3.1 The EM algorithm

The EM algorithm alternates between an expectation (E) step, which defines the expectation of the complete log-likelihood evaluated at the current parameters estimates, and a maximization (M) step, which computes parameter estimates by maximizing the expected complete log-likelihood obtained in the E-step. The complete log-likelihood, the expected complete log-likelihood function and the optimal parameter estimators are given below in the following propositions.

Given the representation in (1.10), let us denote by w_{ig} the indicator variable that is equal to 1 if the i-th unit belongs to the g-th component of the finite mixture, and 0 otherwise. Similarly, let u_{itj} be equal to 1 if unit i is in state j at time t and 0 otherwise; let v_{itjk} be equal to 1 if unit i is in state j at time $t-1$ and in state k at time t, and 0 otherwise. Finally, we denote by z_{itjg} the indicator of the i-th individual being in state j at time t and coming from the g-th component of the mixture. The complete data log-likelihood is presented in the following proposition.

Proposition 1. *For any fixed* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$*, G mixture components and M hidden states, the complete data log-likelihood function is proportional to:*

$$\begin{aligned}\ell_c(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) \propto \sum_{i=1}^{N} \Bigg\{ & \sum_{g=1}^{G} w_{ig} \log \pi_g + \sum_{j=1}^{M} u_{i1j} \log q_j + \sum_{t=2}^{T_i} \sum_{j=1}^{M} \sum_{k=1}^{M} v_{itjk} \log q_{jk} \\ & - \frac{1}{2} T_i \log \mid \mathbf{D}\boldsymbol{\Sigma}\mathbf{D} \mid + \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} z_{itjg} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})' \mathbf{D}^{-1} \boldsymbol{\Sigma}^{-1} \tilde{\xi} \\ & - \frac{1}{2} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} z_{itjg} \frac{1}{\tilde{C}_{itjg}} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})' (\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it}) \\ & - \frac{1}{2} \tilde{\xi}' \boldsymbol{\Sigma}^{-1} \tilde{\xi} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} z_{itjg} \tilde{C}_{itjg} \Bigg\},\end{aligned} \tag{1.12}$$

where $\tilde{C}_{itjg}$ *is a latent variable that follows an exponential distribution with parameter* 1.

In the E-step of the algorithm, the presence of the unobserved indicator variables $w_{ig}, u_{itj}, v_{itjk}$ and z_{itjg} is handled by taking their conditional expectation given the observed data and the current parameter estimates. Calculation of such quantities may be addressed via an adaptation of the forward and backward variables; see Welch (2003). Similarly, the conditional expectations of $\frac{1}{\tilde{C}_{itjg}}$ and $\tilde{C}_{itjg}$ are considered.

For the implementation of the algorithm, forward and backward variables are defined for the longitudinal measures. We define the probability of observing the partial sequence ending up in state j at time t, given the g-th component, as:

$$a_{it}(j,g) = f(\mathbf{y}_{i1:t}, S_{it} = j \mid \mathbf{b}_g) \qquad \text{and} \qquad a_{i1}(j,g) = q_j f(\mathbf{y}_{i1} \mid S_{i1} = j, \mathbf{b}_g). \tag{1.13}$$

The quantity $a_{it}(j,g)$ can be rewritten using the following recurrence relationship:

$$a_{it}(j,g) = \sum_{h=1}^{M} a_{it-1}(h,g) q_{hj} f(\mathbf{y}_{it} \mid S_{it} = j, \mathbf{b}_g). \tag{1.14}$$

Backward variables are defined as the probability of the longitudinal sequence from time $t+1$ to the last available observation T_i, conditional on being in state j at time t, given the g-th component:

$$b_{it}(j,g) = f(\mathbf{y}_{it+1:T_i} \mid S_{it} = j, \mathbf{b}_g) \qquad \text{and} \qquad b_{iT_i}(j,g) = 1. \tag{1.15}$$

Accordingly, the backward variable $b_{it}(j,g)$ can be rewritten as:

$$b_{it}(j,g) = \sum_{k=1}^{M} b_{it+1}(k,g) q_{jk} f(\mathbf{y}_{it+1} \mid S_{it+1} = k, \mathbf{b}_g). \tag{1.16}$$

Finally, the expected values of $w_{ig}, u_{itj}, v_{itjk}$ and z_{itjg} can be computed as:

$$\begin{aligned}
\hat{w}_{ig} &= \frac{\pi_g \sum_{j=1}^{M} a_{iT_i}(j,g)}{\sum_{g=1}^{G} \pi_g \sum_{j=1}^{M} a_{iT_i}(j,g)}, \\
\hat{z}_{itjg} &= \frac{a_{it}(j,g) b_{it}(j,g) \pi_g}{\sum_{g=1}^{G} \sum_{j=1}^{M} a_{it}(j,g) b_{it}(j,g) \pi_g}, \\
\hat{u}_{itj} &= \sum_{g=1}^{G} \hat{z}_{itjg}, \\
\hat{v}_{itjk} &= \frac{\sum_{g=1}^{G} a_{it-1}(j,g) q_{jk} f(\mathbf{y}_{it} \mid S_{it}=k, \mathbf{b}_g) b_{it}(k,g) \pi_g}{\sum_{g=1}^{G} \sum_{j=1}^{M} \sum_{k=1}^{M} a_{it-1}(j,g) q_{jk} f(\mathbf{y}_{it} \mid S_{it}=k, \mathbf{b}_g) b_{it}(k,g) \pi_g}.
\end{aligned} \tag{1.17}$$

By substituting the corresponding posterior expectations in (1.17) into the complete data likelihood in (1.12), the expected complete data log-likelihood function is provided in the following proposition.

Proposition 2. *For any fixed* $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)$*, G mixture components and M hidden states, the expected complete data log-likelihood function is proportional to:*

$$\begin{aligned}
\mathcal{O}(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) \propto \sum_{i=1}^{N} \Bigg\{ & \sum_{g=1}^{G} \hat{w}_{ig} \log \pi_g + \sum_{j=1}^{M} \hat{u}_{i1j} \log q_j + \sum_{t=2}^{T_i} \sum_{j=1}^{M} \sum_{k=1}^{M} \hat{v}_{itjk} \log q_{jk} \\
& - \frac{1}{2} T_i \log \mid \mathbf{D}\boldsymbol{\Sigma}\mathbf{D} \mid + \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} \hat{z}_{itjg} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})' \mathbf{D}^{-1} \boldsymbol{\Sigma}^{-1} \tilde{\xi} \\
& - \frac{1}{2} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} \hat{z}_{itjg} \hat{\tilde{z}}_{itjg} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})' (\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it}) \\
& - \frac{1}{2} \tilde{\xi}' \boldsymbol{\Sigma}^{-1} \tilde{\xi} \sum_{t=1}^{T_i} \sum_{j=1}^{M} \sum_{g=1}^{G} \hat{z}_{itjg} \hat{\tilde{c}}_{itjg} \Bigg\},
\end{aligned} \tag{1.18}$$

where

$$\hat{\tilde{c}}_{itjg} = \left(\frac{\tilde{m}_{itjg}}{2+\tilde{d}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)}{K_{\nu}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)}, \qquad \hat{\tilde{z}}_{itjg} = \left(\frac{2+\tilde{d}}{\tilde{m}_{itjg}} \right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)}{K_{\nu}\left(\sqrt{(2+\tilde{d})\tilde{m}_{itjg}} \right)} - \frac{2\nu}{\tilde{m}_{itjg}}, \tag{1.19}$$

with

$$\tilde{m}_{itjg} = (\mathbf{y}_{it} - \boldsymbol{\mu}_{it})' (\mathbf{D}\boldsymbol{\Sigma}\mathbf{D})^{-1} (\mathbf{y}_{it} - \boldsymbol{\mu}_{it}), \qquad \tilde{d} = \tilde{\xi}' \boldsymbol{\Sigma}^{-1} \tilde{\xi}. \tag{1.20}$$

Therefore, the EM algorithm can be implemented as follows:

E-step: At r-th iteration of the algorithm, let $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}$ denote the current parameter estimates. Then, conditionally on the observed data and $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}$, calculate the conditional expectations in (1.17) and (1.19). We denote such quantities $\hat{w}_{ig}^{(r)}, \hat{z}_{itjg}^{(r)}, \hat{u}_{itj}^{(r)}, \hat{v}_{itjk}^{(r)}$, and $\hat{\tilde{c}}_{itjg}^{(r)}, \hat{\tilde{z}}_{itjg}^{(r)}$.

M-step: Use $\hat{w}_{ig}^{(r)}, \hat{z}_{itjg}^{(r)}, \hat{u}_{itj}^{(r)}, \hat{v}_{itjk}^{(r)}$, and $\hat{\tilde{c}}_{itjg}^{(r)}, \hat{\tilde{z}}_{itjg}^{(r)}$ to maximize $\mathcal{O}(\boldsymbol{\Phi}_{\boldsymbol{\tau}} \mid \hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)})$ with respect to $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$, and obtain the update parameter estimates. Based on the introduced modeling assumptions, the maximization can be partitioned into orthogonal

subproblems, i.e. the maximization with respect to the fixed, hidden Markov chain and discrete mixing distribution parameters can be performed separately. The initial probabilities q_j, transition probabilities q_{jk} and mixing proportions π_g are estimated by:

$$\hat{q}_j^{(r)} = \frac{\sum_{i=1}^N \hat{u}_{i1j}^{(r)}}{N}, \qquad \hat{q}_{jk}^{(r)} = \frac{\sum_{i=1}^N \sum_{t=1}^{T_i} \hat{v}_{itjk}^{(r)}}{\sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{k=1}^M \hat{v}_{itjk}^{(r)}}, \qquad \hat{\pi}_g^{(r)} = \frac{\sum_{i=1}^N \hat{w}_{ig}^{(r)}}{N}. \quad (1.21)$$

If we let $S_{it} = j$, this implies that $\boldsymbol{\alpha}_{s_{it}} = \boldsymbol{\alpha}_j$ and the M-step updates of model parameters $\boldsymbol{\beta}, \mathbf{b}_g, \boldsymbol{\alpha}_j, \boldsymbol{\Sigma}$ and $\mathbf{D}$, are given in the following proposition.

Proposition 3. *The values of* $\boldsymbol{\beta}, \mathbf{b}_g, \boldsymbol{\alpha}_j, \boldsymbol{\Sigma}$ *and* $\mathbf{D}$ *maximizing* (1.18) *are:*

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{(r)} = &(\sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{j=1}^M \hat{\tilde{z}}_{itjg}^{(r)} \hat{\tilde{z}}_{itjg}^{(r)} \mathbf{X}_{it}' \mathbf{X}_{it})^{-1} (\sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{j=1}^M \hat{\tilde{z}}_{itjg}^{(r)} \hat{\tilde{z}}_{itjg}^{(r)} \mathbf{X}_{it}' \tilde{\mathbf{Y}}_{it} \\ &- \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{j=1}^M z_{itjg}^{(r)} \mathbf{X}_{it}' \tilde{\xi}' \mathbf{D}^{(r-1)}),\end{aligned} \quad (1.22)$$

where $\tilde{\mathbf{Y}}_{it} = \mathbf{Y}_{it} - \mathbf{Z}_{it} \hat{\mathbf{b}}_g^{(r-1)} - \mathbf{W}_{it} \hat{\boldsymbol{\alpha}}_j^{(r-1)}$.

$$\hat{\mathbf{b}}_g^{(r)} = (\sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{j=1}^M \hat{\tilde{z}}_{itjg}^{(r)} \hat{\tilde{z}}_{itjg}^{(r)} \mathbf{Z}_{it}' \mathbf{Z}_{it})^{-1} (\sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{j=1}^M \hat{\tilde{z}}_{itjg}^{(r)} \hat{\tilde{z}}_{itjg}^{(r)} \mathbf{Z}_{it}' \tilde{\mathbf{Y}}_{it} - \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{j=1}^M z_{itjg}^{(r)} \mathbf{Z}_{it}' \tilde{\xi}' \mathbf{D}^{(r-1)}), \quad (1.23)$$

where $\tilde{\mathbf{Y}}_{it} = \mathbf{Y}_{it} - \mathbf{X}_{it} \hat{\boldsymbol{\beta}}^{(r)} - \mathbf{W}_{it} \hat{\boldsymbol{\alpha}}_j^{(r-1)}$.

$$\hat{\boldsymbol{\alpha}}_j^{(r)} = (\sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \hat{z}_{itjg}^{(r)} \hat{\tilde{z}}_{itjg}^{(r)} \mathbf{W}_{it}' \mathbf{W}_{it})^{-1} (\sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \hat{z}_{itjg}^{(r)} \hat{\tilde{z}}_{itjg}^{(r)} \mathbf{W}_{it}' \tilde{\mathbf{Y}}_{it} - \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G z_{itjg}^{(r)} \mathbf{W}_{it}' \tilde{\xi}' \mathbf{D}^{(r-1)}), \quad (1.24)$$

where $\tilde{\mathbf{Y}}_{it} = \mathbf{Y}_{it} - \mathbf{X}_{it} \hat{\boldsymbol{\beta}}^{(r)} - \mathbf{Z}_{it} \hat{\mathbf{b}}_g^{(r)}$.

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}^{(r)} = &\frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{j=1}^M \hat{\tilde{z}}_{itjg}^{(r)} \hat{\tilde{z}}_{itjg}^{(r)} \hat{\mathbf{D}}^{-1(r-1)} (\mathbf{Y}_{it} - \hat{\boldsymbol{\mu}}_{it}^{(r)})' (\mathbf{Y}_{it} - \hat{\boldsymbol{\mu}}_{it}^{(r)}) \hat{\mathbf{D}}^{-1(r-1)} \\ &+ \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{j=1}^M \hat{z}_{itjg}^{(r)} \hat{\tilde{c}}_{itjg}^{(r)} \tilde{\xi} \tilde{\xi}' - \frac{2}{\sum_{i=1}^N T_i} \hat{\mathbf{D}}^{-1(r-1)} \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{j=1}^M \hat{z}_{itjg}^{(r)} (\mathbf{Y}_{it} - \hat{\boldsymbol{\mu}}_{it}^{(r)})' \tilde{\xi}',\end{aligned} \quad (1.25)$$

where $\hat{\boldsymbol{\mu}}_{it}^{(r)} = \mathbf{X}_{it} \hat{\boldsymbol{\beta}}^{(r)} + \mathbf{Z}_{it} \hat{\mathbf{b}}_g^{(r)} + \mathbf{W}_{it} \hat{\boldsymbol{\alpha}}_j^{(r)}$.

Finally, the elements $d_j, j = 1, \ldots, p$ *of the diagonal scale matrix* $\mathbf{D}$ *are estimated by:*

$$\hat{d}_j^{(r)} = \frac{1}{\sum_{i=1}^N T_i} \sum_{i=1}^N \sum_{t=1}^{T_i} \sum_{g=1}^G \sum_{k=1}^M \hat{z}_{itkg}^{(r)} \rho_{\tau_j} (Y_{it}^{(j)} - \hat{\mu}_{it}^{(j)(r)}), \quad (1.26)$$

where $\rho_\tau(\cdot)$ *is the quantile check function of Koenker & Bassett (1978):*

$$\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0)) \tag{1.27}$$

and $\hat{\mu}_{it}^{(j)(r)}$ *is the* j*-th element of the vector* $\hat{\boldsymbol{\mu}}_{it}^{(r)}$.

The E- and M-steps are alternated until convergence, that is when $|\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r)} - \hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}|$ is smaller than a predetermined threshold. In this paper, we set this convergence criterion equal to 10^{-6}.

Because both the number of components of the finite mixture and hidden states of the Markov chain are unknown a-priori, we select the optimal value of G and M using the BIC (Schwarz et al. 1978):

$$BIC_{(G,M)} = -2\ell(\boldsymbol{\Phi}_{\boldsymbol{\tau}}) + \log(N)\nu_f, \tag{1.28}$$

where $\ell(\boldsymbol{\Phi}_{\boldsymbol{\tau}})$ is the observed data log-likelihood in (1.10), N is the number of observed individuals and ν_f denotes the number of free model parameters in $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$. Following Marino et al. (2018), to avoid convergence to local maxima and better explore the parameter space, for fixed $\boldsymbol{\tau}, G$ and M, we fit the QMHMM model using a multiple random starts strategy with 50 different starting points and retain the solution corresponding to the maximum likelihood value. We then repeat this procedure for a grid of values of G and M, and select the best combination of the pair (G, M) corresponding to the lowest BIC value. The validity of the proposed EM algorithm and model selection procedure have been assessed using also a simulation exercise (see Section 1.4).

Standard errors of model parameters are computed using a non-parametric block bootstrap. That is, by re-sampling individuals with replacement and retaining the corresponding sequence of measurements to preserve the within individual dependence structure (see Geraci & Bottai 2014, Marino & Farcomeni 2015 and Marino et al. 2018 for example). We refit the model to H bootstrap samples and approximate the standard error of each model parameter with the square root of the variance of the matrix:

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}) = \sqrt{\frac{1}{H-1}\sum_{h=1}^{H}(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(h)} - \bar{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(h)} - \bar{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})'}, \tag{1.29}$$

where $\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}^{(h)}$ is the set of parameter estimates for the h-th bootstrap sample and $\bar{\boldsymbol{\Phi}}_{\boldsymbol{\tau}}$ denotes the mean of the model parameters over bootstrap iterations. The standard errors are given by the diagonal elements of $\widehat{\text{Cov}}(\hat{\boldsymbol{\Phi}}_{\boldsymbol{\tau}})$.

1.4 Simulation study

In this section we conduct a simulation study to evaluate the finite sample properties of the proposed method and show that the introduced methodology represents a valid procedure to estimate the quantile regression coefficients. This simulation exercise addresses the following questions. First, we consider different distributional choices for the error term to study the performance of the model in the presence

of non-Gaussian errors. Second, we evaluate the robustness of the non-parametric approach to non-Gaussian distributions for the subject-specific, random coefficients. Finally, we analyze the performance of penalized likelihood criteria in selecting the optimal number of mixture components G and hidden states M.

We consider three sample sizes $N = (100, 200, 300)$ and three longitudinal lengths $T_i = T = (5, 10, 15)$, $i = 1, \ldots, N$, for a continuous response variable of dimension $p = 2$ and two explanatory variables $X_{it}^{(1)} \sim \mathcal{N}(0, 1)$ and $X_{it}^{(2)} \sim \text{Ber}(0.5)$. The observations are generated from a two state homogeneous Markov chain, i.e. $M = 2$, using the following data generating process:

$$\mathbf{Y}_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_{it}\mathbf{b}_i + \mathbf{W}_{it}\boldsymbol{\alpha}_{S_{it}} + \epsilon_{it}. \tag{1.30}$$

Regarding the hidden Markov chain, the simulation scheme is similar to the one adopted by Marino et al. (2018). The true values of the fixed, $\boldsymbol{\beta}$, state dependent parameters, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M)$, and the initial probabilities, $\mathbf{q}$, and transition probabilities, $\mathbf{Q}$, are given by, respectively:

$$\boldsymbol{\beta} = \begin{pmatrix} 2 & -0.8 \\ -1.4 & 3 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} 5 & -2 \\ -5 & 2 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}, \qquad \mathbf{q} = \begin{pmatrix} 0.7 & 0.3 \end{pmatrix}. \tag{1.31}$$

We consider a time-varying random intercept by setting $\mathbf{W}_{it} = \mathbf{1}$ and a random slope $\mathbf{Z}_{it} = X_{it}^{(1)}$. Hence, $\mathbf{b}_i$ are time-constant random slopes that capture individual departures from the marginal effect $\boldsymbol{\beta}$. For each sample size, two different simulation scenarios for the error distributions and for the random coefficients distributions are considered:

> $(\mathcal{N} - \mathcal{N})$: $\mathbf{b}_i$ represent i.i.d. draws from a standard bivariate Gaussian with variance-covariance matrix, $\boldsymbol{\Omega} = \left(\begin{smallmatrix} 1 & 0.25 \\ 0.25 & 1 \end{smallmatrix}\right)$ and the error terms, ϵ_{it}, are generated from a bivariate Normal random variable with zero mean vector and variance-covariance matrix equal to $\tilde{\boldsymbol{\Omega}}$;
>
> $(\mathcal{T} - \mathcal{T})$: $\mathbf{b}_i$ are sampled from a bivariate Student t with 3 degrees of freedom, centered around zero and scale matrix $\boldsymbol{\Omega} = \left(\begin{smallmatrix} 1 & 0.25 \\ 0.25 & 1 \end{smallmatrix}\right)$ while, ϵ_{it} are generated from a bivariate Student t distribution with 3 degrees of freedom, zero mean and scale matrix $\tilde{\boldsymbol{\Omega}}$.

Each simulation scenario is repeated twice by generating the errors ϵ_{it} with low ($\tilde{\boldsymbol{\Omega}} = \left(\begin{smallmatrix} 1 & 0.3 \\ 0.3 & 1 \end{smallmatrix}\right)$) and high ($\tilde{\boldsymbol{\Omega}} = \left(\begin{smallmatrix} 1 & 0.8 \\ 0.8 & 1 \end{smallmatrix}\right)$) correlation between the responses of unit i at a given time t. Figure 1.1 shows the scatter plot matrices of the simulated datasets for all scenarios considered with $N = 200$ and $T = 10$.

To fit the proposed model, we consider a varying number of mixture components $G = (2, \ldots, 10)$ and retained the model with the lowest BIC value. We analyze four different quantile levels: in the first case, we assume $\boldsymbol{\tau} = (0.50.0.50)$; in the second one, we set $\boldsymbol{\tau} = (0.25, 0.25)$; in the third, we set $\boldsymbol{\tau} = (0.75, 0.75)$ and finally we consider a more extreme case $\boldsymbol{\tau} = (0.25, 0.75)$. For each model, we carry out $B = 250$ Monte Carlo replications and report the following indicators. The Average Relative Bias (ARB) defined as:

$$ARB(\hat{\theta}_{\boldsymbol{\tau}}) = \frac{1}{B}\sum_{b=1}^{B}\frac{(\hat{\theta}_{\boldsymbol{\tau}}^{(b)} - \theta_{\boldsymbol{\tau}})}{\theta_{\boldsymbol{\tau}}} \times 100, \tag{1.32}$$

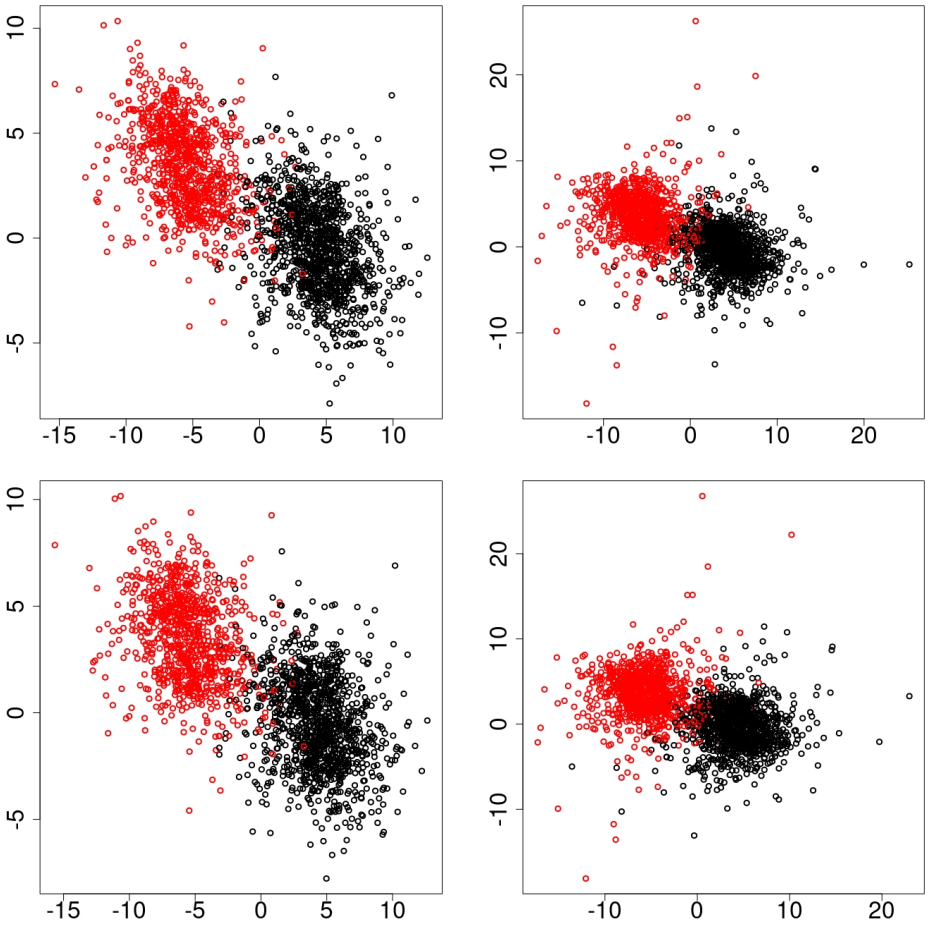

Figure 1.1. Scatter plots of simulated datasets using $\rho_{12} = 0.3$ (first row) and $\rho_{12} = 0.8$ (second row) for the $\mathcal{N} - \mathcal{N}$ (first column) and $\mathcal{T} - \mathcal{T}$ (second column) scenarios with $N = 200$ and $T = 10$. Red and black data points distinguish the two latent states.

where $\hat{\theta}_{\boldsymbol{\tau}}^{(b)}$ is the estimated parameter at quantile level $\boldsymbol{\tau}$ for the b-th replication and $\theta_{\boldsymbol{\tau}}$ is the corresponding "true" value of the parameter. Secondly, the Root Mean Square Error (RMSE) of model parameters averaged across the B simulations:

$$RMSE(\hat{\theta}_{\boldsymbol{\tau}}) = \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}_{\boldsymbol{\tau}}^{(b)} - \theta_{\boldsymbol{\tau}})^2}. \tag{1.33}$$

Tables 1.1, 1.2 and 1.3 report the results for the fixed parameters $\boldsymbol{\beta}$ and state-specific coefficients $\boldsymbol{\alpha}$.

As can be noted, the proposed model under the Normal and the Student t error distributions is able to recover the true fixed parameters and state-dependent intercept values for both low (Panels A) and high (Panels B) degree of dependence. Not surprisingly, the bias effect is quite small when we analyze the median levels (see columns 1 and 5). As the quantile levels become more extreme (see columns 2-4 and 6-8), the ARB slightly increases but it still remains reasonably small. Such differences are due to the reduced amount of information in the tails of the

distribution. Moreover, we can see a slightly higher bias for the parameters β_{21} and β_{22} associated to the binary covariate than those of the continuous covariate, which can possibly be due to some aliasing between the effect of the binary covariate $X_{it}^{(2)}$ and the state-dependent intercepts $\boldsymbol{\alpha}$. This situation might be exacerbated by the degree of overlap among the latent states of the hidden Markov chain, as shown in Figure 1.1. However, both the ARB and the RMSE tend to decrease with increasing sample sizes and number of measurement occasions (see Table 1.3). Also, under the scenario where $\mathbf{b}_i \sim \mathcal{T}_2(\mathbf{0}, \boldsymbol{\Omega})$ and $\epsilon_{it} \sim \mathcal{T}_2(\mathbf{0}, \tilde{\boldsymbol{\Omega}})$, the heavier tails of the Student t contribute to higher ARB and RMSE especially at the 25-th and 75-th percentiles. Concerning the hidden process, it is worth noting that we observe sensible differences in terms of efficiency for the state dependent parameters $\boldsymbol{\alpha}$. Given the true values of $\mathbf{Q}$ and $\mathbf{q}$, most of the units are in the first state of the latent Markov chain, sharing the common intercept value $\boldsymbol{\alpha}_1$. Hence, the intercept corresponding to the second state $\boldsymbol{\alpha}_2$ is estimated with lower precision due to lack of transitions from one state to the other. However, when the number of repeated measurements increases, we observe more frequent transitions towards the second state with the effect of reducing the RMSE. Again, such difference is more evident in the tails of the distribution. These findings are generally consistent with the ones in Marino et al. (2018).

To evaluate the performance of the model selection procedure described in Section 1.3, we considered the same simulation experiment with $N = 200$, $T = 10$, $M = 2$, $\boldsymbol{\tau} = (0.50, 0.50)$ and $B = 100$. Following Marino et al. (2018), for each of the simulated datasets we fit the QMHMM for $G = (2, \ldots, 8)$ and $M = (2, \ldots, 4)$, and select the optimal value of the pair (G, M) by using the AIC (Akaike 1998) and BIC in (1.28). Because the time-constant random slopes $\mathbf{b}_i$ in (1.30) are generated from continuous distributions, we only report the distribution of absolute frequencies of the hidden states M selected by the two penalized likelihood criteria. Table 1.4 summarizes the results.

As one can see, the BIC works well and outperform the AIC, with an average of correctly identified number of hidden states of more than the 80% across all simulation scenarios and levels of correlation. Furthermore, regardless of the distributional assumptions on the random slopes or on the error terms, the BIC captures the serial heterogeneity in the data in a more parsimonious manner compared to the AIC, hence offering easier interpretation about unobserved heterogeneity. In concluding, it is worth noting that the ability of both criteria in recovering the true number of hidden states tends to improve as the correlation between the outcomes increases. This is possibly due to the gain in efficiency of the proposed multivariate method w.r.t. running univariate quantile regressions on each outcome separately. That is, by estimating multiple conditional quantiles in one step and accounting for the association between the elements of $\mathbf{Y}_{it}$, the methodology introduced allows to borrow strength among the responses to improve inference and the precision of the estimates (Petrella & Raponi 2019), producing the largest efficiency gains when the variables are highly correlated.

	$(\mathcal{N}-\mathcal{N})$				$(\mathcal{T}-\mathcal{T})$			
$\boldsymbol{\tau}$	$(0.50, 0.50)$	$(0.25, 0.25)$	$(0.75, 0.75)$	$(0.25, 0.75)$	$(0.50, 0.50)$	$(0.25, 0.25)$	$(0.75, 0.75)$	$(0.25, 0.75)$
Panel A: $\rho_{12} = 0.3$								
$\beta_{11} = 2$	0.386 (0.114)	0.316 (0.110)	0.470 (0.116)	0.514 (0.116)	−0.310 (0.168)	−0.325 (0.175)	−0.214 (0.174)	−0.888 (0.189)
$\beta_{12} = -0.8$	0.716 (0.122)	1.317 (0.123)	0.912 (0.125)	0.843 (0.125)	0.951 (0.158)	0.313 (0.166)	0.871 (0.173)	0.670 (0.169)
$\beta_{21} = -1.4$	−0.281 (0.071)	0.942 (0.081)	−2.603 (0.091)	0.221 (0.092)	0.546 (0.075)	2.137 (0.102)	−1.752 (0.100)	−0.961 (0.106)
$\beta_{22} = 3$	−0.019 (0.073)	−0.652 (0.091)	0.838 (0.089)	0.331 (0.089)	0.044 (0.083)	−1.354 (0.111)	1.163 (0.103)	0.366 (0.108)
$\alpha_{11} = 5$	−0.028 (0.057)	−0.252 (0.063)	0.017 (0.070)	1.352 (0.071)	0.033 (0.058)	−0.201 (0.078)	0.535 (0.091)	1.265 (0.088)
$\alpha_{12} = -2$	0.397 (0.052)	0.387 (0.069)	0.042 (0.062)	0.365 (0.068)	−0.180 (0.068)	1.215 (0.082)	−1.459 (0.093)	1.022 (0.093)
$\alpha_{21} = -5$	−0.165 (0.074)	−0.034 (0.082)	−0.096 (0.080)	−0.397 (0.087)	0.179 (0.081)	0.763 (0.113)	0.066 (0.097)	−0.192 (0.100)
$\alpha_{22} = 2$	−0.120 (0.075)	−0.057 (0.083)	0.293 (0.084)	−0.712 (0.090)	0.011 (0.078)	−1.862 (0.108)	0.795 (0.095)	0.768 (0.100)
Panel B: $\rho_{12} = 0.8$								
$\beta_{11} = 2$	0.281 (0.113)	0.178 (0.111)	0.294 (0.116)	0.390 (0.111)	−0.339 (0.179)	−0.668 (0.185)	−0.655 (0.188)	−0.697 (0.179)
$\beta_{12} = -0.8$	0.951 (0.126)	1.170 (0.127)	0.700 (0.128)	0.213 (0.123)	1.413 (0.165)	1.876 (0.180)	1.946 (0.173)	0.489 (0.168)
$\beta_{21} = -1.4$	−0.522 (0.073)	2.041 (0.087)	−3.086 (0.091)	−2.622 (0.101)	0.282 (0.075)	3.756 (0.108)	−3.297 (0.113)	−1.614 (0.107)
$\beta_{22} = 3$	0.175 (0.074)	−1.205 (0.094)	1.502 (0.094)	−0.424 (0.092)	−0.106 (0.081)	−2.019 (0.119)	1.717 (0.119)	−0.532 (0.116)
$\alpha_{11} = 5$	−0.054 (0.055)	−0.438 (0.073)	0.168 (0.068)	3.139 (0.090)	0.012 (0.063)	−0.695 (0.094)	0.813 (0.097)	1.917 (0.099)
$\alpha_{12} = -2$	0.302 (0.052)	1.255 (0.078)	−0.434 (0.067)	0.861 (0.082)	0.050 (0.068)	1.997 (0.093)	−2.190 (0.099)	−0.149 (0.103)
$\alpha_{21} = -5$	−0.031 (0.071)	0.170 (0.083)	−0.383 (0.077)	−1.050 (0.105)	0.028 (0.086)	1.068 (0.121)	−0.582 (0.118)	0.403 (0.106)
$\alpha_{22} = 2$	−0.211 (0.076)	−0.457 (0.081)	1.068 (0.081)	−2.188 (0.096)	0.012 (0.082)	−2.349 (0.115)	1.873 (0.112)	1.124 (0.106)

Table 1.1. ARB and RMSE (in brackets) for longitudinal and state-parameter estimates with a sample size $N = 100$ and length of longitudinal sequences $T = 5$.

	$(\mathcal{N}-\mathcal{N})$				$(\mathcal{T}-\mathcal{T})$			
$\boldsymbol{\tau}$	$(0.50, 0.50)$	$(0.25, 0.25)$	$(0.75, 0.75)$	$(0.25, 0.75)$	$(0.50, 0.50)$	$(0.25, 0.25)$	$(0.75, 0.75)$	$(0.25, 0.75)$
Panel A: $\rho_{12} = 0.3$								
$\beta_{11} = 2$	0.173 (0.077)	0.008 (0.075)	−0.004 (0.075)	−0.027 (0.076)	0.098 (0.118)	−0.076 (0.120)	−0.010 (0.120)	0.197 (0.115)
$\beta_{12} = -0.8$	−0.432 (0.076)	−0.273 (0.076)	0.031 (0.076)	−0.081 (0.078)	−1.486 (0.107)	−0.503 (0.109)	−1.429 (0.106)	−1.273 (0.111)
$\beta_{21} = -1.4$	−0.291 (0.042)	1.947 (0.056)	−2.187 (0.055)	0.995 (0.054)	−0.108 (0.044)	1.628 (0.060)	−2.230 (0.066)	−0.561 (0.060)
$\beta_{22} = 3$	0.066 (0.036)	−1.013 (0.056)	1.052 (0.055)	0.676 (0.055)	−0.098 (0.043)	−1.165 (0.069)	1.280 (0.070)	0.438 (0.058)
$\alpha_{11} = 5$	−0.096 (0.032)	−0.197 (0.041)	0.241 (0.041)	0.589 (0.042)	0.030 (0.035)	−0.278 (0.047)	0.551 (0.057)	−0.034 (0.062)
$\alpha_{12} = -2$	0.116 (0.032)	0.409 (0.041)	−0.657 (0.039)	0.112 (0.046)	0.062 (0.036)	1.425 (0.059)	−1.398 (0.063)	0.702 (0.064)
$\alpha_{21} = -5$	−0.038 (0.037)	0.262 (0.052)	−0.296 (0.047)	−0.142 (0.052)	0.061 (0.039)	0.666 (0.067)	−0.349 (0.058)	1.438 (0.060)
$\alpha_{22} = 2$	−0.040 (0.038)	−0.862 (0.047)	0.372 (0.042)	−0.470 (0.047)	−0.021 (0.038)	−1.695 (0.067)	1.568 (0.070)	1.076 (0.069)
Panel B: $\rho_{12} = 0.8$								
$\beta_{11} = 2$	0.040 (0.075)	−0.028 (0.078)	−0.028 (0.079)	0.024 (0.081)	0.280 (0.123)	−0.113 (0.125)	−0.055 (0.119)	0.113 (0.120)
$\beta_{12} = -0.8$	−0.117 (0.075)	0.151 (0.080)	0.156 (0.077)	−0.062 (0.079)	−1.511 (0.115)	−0.685 (0.112)	−0.718 (0.116)	−0.963 (0.115)
$\beta_{21} = -1.4$	−0.385 (0.044)	3.246 (0.070)	−3.284 (0.070)	−1.472 (0.060)	−0.048 (0.051)	4.132 (0.087)	−4.411 (0.085)	−1.962 (0.074)
$\beta_{22} = 3$	0.080 (0.045)	−1.532 (0.072)	1.500 (0.070)	−0.528 (0.053)	−0.036 (0.051)	−1.968 (0.088)	2.047 (0.089)	−0.168 (0.066)
$\alpha_{11} = 5$	−0.040 (0.035)	−0.703 (0.058)	0.594 (0.052)	2.685 (0.068)	0.031 (0.037)	−0.993 (0.075)	0.942 (0.077)	−0.179 (0.069)
$\alpha_{12} = -2$	0.159 (0.036)	1.878 (0.059)	−1.637 (0.055)	−2.424 (0.068)	−0.006 (0.036)	2.839 (0.080)	−2.511 (0.080)	0.190 (0.066)
$\alpha_{21} = -5$	−0.028 (0.039)	0.666 (0.058)	−0.702 (0.062)	−1.048 (0.071)	0.054 (0.042)	1.031 (0.084)	−0.950 (0.077)	2.647 (0.076)
$\alpha_{22} = 2$	0.045 (0.037)	−1.822 (0.057)	1.933 (0.061)	1.125 (0.069)	−0.074 (0.044)	−2.590 (0.084)	2.615 (0.081)	1.447 (0.089)

Table 1.2. ARB and RMSE (in brackets) for longitudinal and state-parameter estimates with a sample size $N = 200$ and length of longitudinal sequences $T = 10$.

	$(\mathcal{N}-\mathcal{N})$				$(\mathcal{T}-\mathcal{T})$			
$\boldsymbol{\tau}$	$(0.50, 0.50)$	$(0.25, 0.25)$	$(0.75, 0.75)$	$(0.25, 0.75)$	$(0.50, 0.50)$	$(0.25, 0.25)$	$(0.75, 0.75)$	$(0.25, 0.75)$
Panel A: $\rho_{12} = 0.3$								
$\beta_{11} = 2$	0.059 (0.061)	0.025 (0.063)	0.375 (0.064)	0.264 (0.062)	−0.183 (0.099)	−0.001 (0.099)	−0.092 (0.098)	0.062 (0.097)
$\beta_{12} = -0.8$	−0.115 (0.063)	−0.021 (0.062)	0.100 (0.064)	0.163 (0.063)	−0.227 (0.091)	−0.972 (0.093)	−0.539 (0.092)	−0.580 (0.087)
$\beta_{21} = -1.4$	−0.103 (0.034)	0.159 (0.048)	−0.589 (0.053)	0.435 (0.048)	−0.219 (0.037)	1.323 (0.061)	−0.683 (0.060)	−0.741 (0.048)
$\beta_{22} = 3$	0.100 (0.032)	−0.305 (0.049)	0.058 (0.049)	0.297 (0.049)	−0.050 (0.033)	−1.299 (0.059)	1.302 (0.062)	0.501 (0.048)
$\alpha_{11} = 5$	−0.049 (0.025)	−0.336 (0.041)	0.447 (0.039)	−0.065 (0.035)	−0.048 (0.026)	−0.596 (0.050)	0.700 (0.054)	1.155 (0.068)
$\alpha_{12} = -2$	0.097 (0.023)	0.312 (0.039)	−0.177 (0.038)	−0.091 (0.033)	0.051 (0.026)	0.251 (0.062)	−0.279 (0.058)	1.568 (0.050)
$\alpha_{21} = -5$	−0.008 (0.027)	0.406 (0.041)	−0.559 (0.045)	0.056 (0.034)	−0.015 (0.029)	0.651 (0.055)	−0.596 (0.051)	−0.030 (0.045)
$\alpha_{22} = 2$	−0.089 (0.028)	−0.284 (0.042)	0.245 (0.041)	0.358 (0.031)	0.035 (0.031)	−1.049 (0.057)	1.247 (0.065)	1.082 (0.049)
Panel B: $\rho_{12} = 0.8$								
$\beta_{11} = 2$	0.083 (0.064)	0.146 (0.061)	0.114 (0.060)	0.133 (0.065)	−0.122 (0.096)	−0.192 (0.097)	−0.193 (0.100)	−0.043 (0.092)
$\beta_{12} = -0.8$	0.104 (0.065)	0.175 (0.064)	0.074 (0.064)	−0.120 (0.063)	−0.048 (0.092)	−0.931 (0.093)	−0.390 (0.091)	−0.837 (0.088)
$\beta_{21} = -1.4$	−0.133 (0.037)	0.240 (0.060)	−0.185 (0.060)	−0.200 (0.044)	−0.152 (0.038)	0.769 (0.075)	−0.485 (0.080)	−1.229 (0.056)
$\beta_{22} = 3$	0.090 (0.037)	−0.128 (0.061)	0.397 (0.057)	−0.404 (0.042)	0.013 (0.034)	−1.534 (0.079)	1.519 (0.075)	−0.166 (0.054)
$\alpha_{11} = 5$	−0.028 (0.025)	−0.013 (0.060)	0.960 (0.058)	0.238 (0.069)	−0.085 (0.026)	−1.453 (0.088)	1.414 (0.083)	1.676 (0.094)
$\alpha_{12} = -2$	−0.020 (0.025)	0.301 (0.062)	−0.539 (0.061)	0.940 (0.065)	0.259 (0.026)	0.786 (0.089)	−0.681 (0.087)	0.660 (0.071)
$\alpha_{21} = -5$	−0.006 (0.029)	0.459 (0.060)	−0.552 (0.063)	−0.335 (0.066)	−0.038 (0.031)	1.328 (0.084)	−0.496 (0.088)	−0.356 (0.047)
$\alpha_{22} = 2$	−0.019 (0.030)	−0.219 (0.064)	0.627 (0.063)	−0.527 (0.059)	0.185 (0.032)	−1.051 (0.086)	1.025 (0.095)	0.572 (0.050)

Table 1.3. ARB and RMSE (in brackets) for longitudinal and state-parameter estimates with a sample size $N = 300$ and length of longitudinal sequences $T = 15$.

Correlation	$\rho_{12} = 0.3$				$\rho_{12} = 0.8$			
Scenario	$(\mathcal{N} - \mathcal{N})$		$(\mathcal{T} - \mathcal{T})$		$(\mathcal{N} - \mathcal{N})$		$(\mathcal{T} - \mathcal{T})$	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
# of hidden states M								
2	20	67	79	92	33	68	92	96
3	34	22	19	8	36	25	6	4
4	46	11	2	0	31	7	2	0

Table 1.4. Absolute frequency distributions of the selected hidden states M via AIC and BIC with a sample size $N = 200$ and length of longitudinal sequences $T = 10$, over $B = 100$ Monte Carlo replications.

1.5 Application

In this section we present the application of the proposed methodology to the MCS dataset. The main aim of this empirical analysis is to assess how selected risk factors affect children's emotional (internalizing) and behavioural (externalizing) problems simultaneously, at different quantile levels of interest, and take into account the potential tail dependence between internalizing and externalizing scores. Indeed, it is possible that certain risk factors have a more pronounced effect at the top end where children display a high, perhaps abnormal, level of development problems than at the bottom end of the distribution of the internalizing and behavioural problems. In this case, jointly modeling the quantiles of the conditional distribution of the responses may be more appropriate than the conditional mean and can provide a more complete picture of the determinants of children's problems.

1.5.1 Data description

The MCS study (http://www.cls.ioe.ac.uk) is a longitudinal study that follows the lives of around 19,000 young people born across UK in 2000-02. It is designed to over-represent children from deprived backgrounds and aims at better addressing the effects of social disadvantage on children's outcomes. The information collected includes child development, social stratification and family life providing important evidence on how economic circumstances, parenting and relationships in the very first stages of life can influence later health and development. This study is widely regarded as the basis of the most reliable estimates of cognitive development problems in young people and it has been deeply investigated in the fields of child psychology and pedagogy; see, among others, the works of Griffiths et al. (2011), Goodman & Goodman (2011), Tzavidis et al. (2016), Bell et al. (2019), Ahn et al. (2018) and Alfò et al. (2021). In particular, Tzavidis et al. (2016) developed a univariate M-quantile regression with time-constant random effects for emotional and behavioral disorders, meanwhile Alfò et al. (2021) built a joint M-quantile model for both disorders, allowing for potential endogeneity issues through an auxiliary regression approach.

One of the most widely and internationally used measure of child mental health is provided by the Strengths and Difficulties Questionnaire (SDQ, see Goodman 1997 and Goodman & Goodman 2009). It represents a balanced coverage of children and young people's behaviours, emotions and relationships and it has been designed

to measure children's emotional and behavioural problems in psychological research. On one hand, internalizing behaviours are typified by inward symptoms such as being withdrawn, fearful or anxious; on the other hand, externalizing behaviours are outward and may be described as aggressive, non-compliant, impulsive or fidgety. The SDQ score is the sum of the main caregiver's responses to a series of items that describe children's internalizing and externalizing problems. This covers five different domains: emotional symptoms, peer problems, conduct problems, hyperactivity, and pro-social behavior. Each domain is measured by five items, for a total of 25 items. For each item, a score equal to 0 is given if the response is not true, 1 if it is somewhat true and 2 if it is certainly true. The internalizing SDQ score is the sum of the scores for responses to the five items in the domains of emotional and peer problems while the externalizing SDQ score is the sum of responses to the five conduct problems and hyperactivity items. Therefore, both SDQ scores range from 0 to 20 and were collected at ages 3, 5 and 7 years. In this context, the availability of longitudinal data can shed light on the evolution of SDQ scores over time and on how they are affected by risk factors and other family and child characteristics.

An extensive literature has examined and documented the effect of mother's characteristics (McMunn et al. 2012), neighbourhood context (Flouri & Sarmadi 2016) and family risk factors (see Tzavidis et al. 2016 and Wickham et al. 2017) on MCS children's trajectories of SDQ scores. Consequently, to analyse the effect of demographic and socio-economic factors on children disorders, we include the following set of predictors. ALE 11 measures the number, out of 11 events, of potentially stressful life events experienced by the family between two consecutive sweeps. The events, classified on the basis of the scale proposed by Tiet et al. (1998), are family member died, negative change in financial situation, new step-parent, sibling left home, child got seriously sick or injured, divorce or separation, family moved, parent lost job, new natural sibling, new stepsibling and mother diagnosed with or treated for depression. SED 4 measures the household's socio-economic disadvantage condition by combining information on overcrowding (more than 1.5 people per room excluding the bathroom and kitchen), not owning a home, receipt of means-tested income support and income poverty. Mother's personal characteristics and distress psychological indicators are included such as maternal education (no qualification (baseline), university degree or General Certificate of Secondary Education (GCSE)) and maternal depression, Kessm, measured by the Kessler score. Furthermore, child's age in years, centered around the mean, age year scal, the quadratic effect of child's age, age2 year scal, ethnicity (non-white (baseline) or white) and gender (female (baseline) or male) were included in the model. Finally, three explanatory variables evaluate the area characteristics. Imdscore is a time varying variable which measures neighbourhood deprivation by the index of multiple-deprivation score. A design variable which allows for the stratification of the MCS sampling design: the stratification variable of the MCS consists of three categories, namely the advantaged stratum (baseline category), the ethnic stratum, Eng eth stratum, and the disadvantaged stratum, Eng dis stratum. The considered covariates are the same as in Tzavidis et al. (2016). For a more detailed description of the covariates see also Tzavidis et al. (2016) and Alfò et al. (2021).

The data that we use in this paper are SDQ internalizing, SDQ_{Int}, and SDQ

externalizing, SDQ_{Ext}, scores recorded on children who were observed at all measurement occasions, i.e. the considered sample consists of $N = 5342$ units and $T_i = T = 3$, for all $i = 1, \ldots, N$. As is customary in child psychology, SDQ scores are treated as though they are continuous variables. Table 1.5 presents the main descriptive statistics of continuous and categorical variables considered in the sample. The average values of ALE 11, SED 4 and Kessm are 1.405, 0.555 and 2.597 but, there are cases with much higher scores as demonstrated by their maximum values. 43% of children have mothers who hold a degree and 48% have mothers with GCSE or other qualification. Around 50% of children are males and, in relation to ethnicity, 89% of members are white. The sample also includes 8.2% and 38% of families from the ethnic and disadvantaged strata respectively. Furthermore, as expected, the empirical correlation between SDQ_{Int} and SDQ_{Ext} equals 0.369, confirming that internalizing and externalizing problems are positively associated.

Variable	Minimum	1-st quartile	Median	Mean†	3-rd quartile	Maximum
SDQ_{Int}	0	1	2	2.493	4	19
SDQ_{Ext}	0	2	5	5.144	7	20
ALE 11	0	1	1	1.405	2	7
SED 4	0	0	0	0.555	1	4
Kessm	0	0	2	2.597	4	24
Degree				43.429		
GCSE				47.997		
White				89.012		
Male				50.337		
IMD	1	3	6	5.619	8	10
Eth stratum				8.199		
Dis stratum				38.132		

Table 1.5. Summary statistics for the MCS data. † means for dummy variables are reported in %.

As a preliminary step, we study the conditional distributions of SDQ scores given a set of covariates by fitting two univariate linear mixed models separately. In particular, we use a two-level random-intercepts model for SDQ outcomes with random effects specified at the level of the child. The model includes the following predictors, namely ALE11, Kessm, SED 4, Eng dis stratum, Eth dis stratum, child's age, gender and ethnicity. Figure 1.2 presents normal probability plots of level 1 and level 2 residuals. These reveal the presence of potentially influential observations in the data, indicate severe departures from the Gaussian assumption of the random-intercepts model for both SDQ outcomes and show that data are severely skewed. To further justify our approach, Figure 1.3 shows the individual trajectories of SDQ scores for a random subset of children in the sample. The overall trend, as estimated by a local polynomial regression, is shown in red along with the 95% confidence bands highlighted in grey. Two things stand out. Firstly, while the general trend is relatively constant over time, individual children may vary in that trajectory. Secondly, for most children, temporal trajectories display rapid changes and highlight a "U"-shaped curve for externalizing score measurement occasions. These indicate that time-constant random effects may not be suitable to describe high individual-

specific heterogeneity in the individual trajectories thus, their temporal evolution could be better captured by introducing time-varying intercepts. For these reasons, a QMHMM for SDQ scores seems to be appropriate and can offer useful information for clinicians and educationalists.

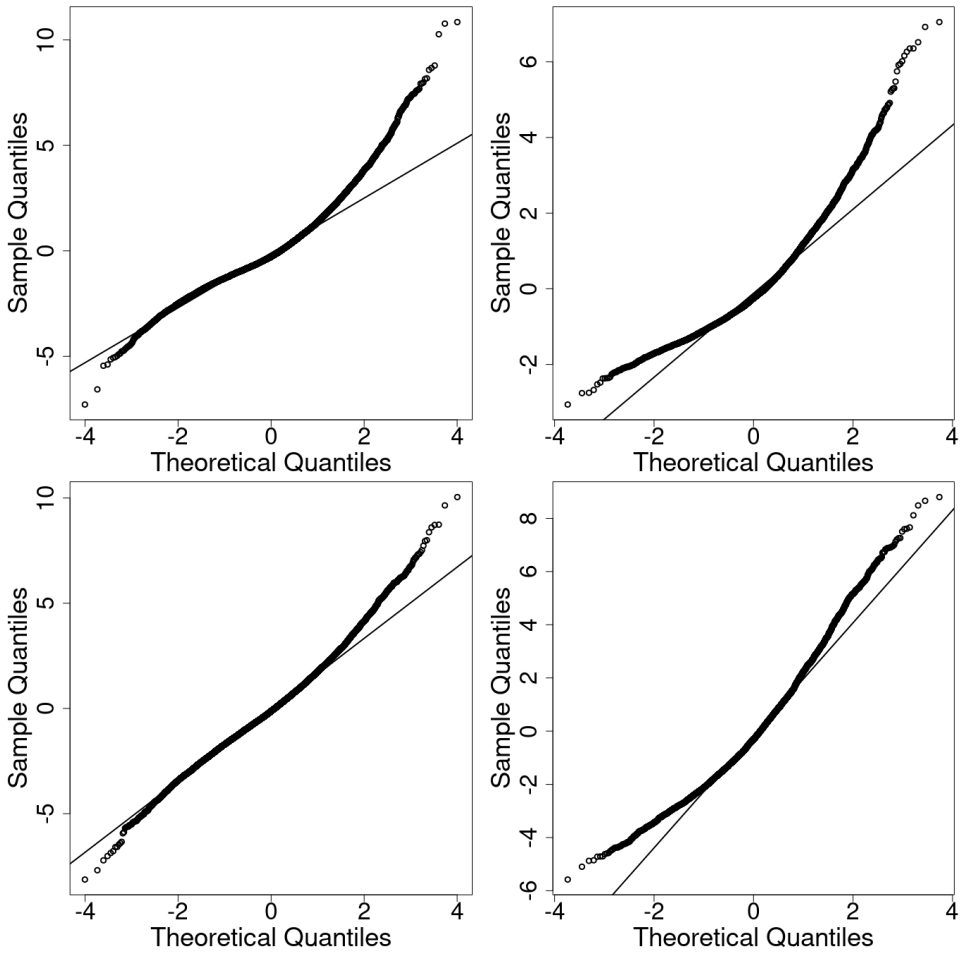

Figure 1.2. Normal probability plots of level 1 (first column) and level 2 (second column) residuals from a linear mixed model for SDQ internalizing (first row) and externalizing (second row) problems.

1.5.2 Results for SDQ scores

In this section, we analyze internalizing and externalizing data on disorders of children collected in the MCS dataset. We are interested in investigating the impact of environmental, parental, and child factors across both the distributions of SDQ scores. In order to account for all the data features described in the previous section, we consider a bivariate QMHMM with state-dependent random intercepts and constant random slopes specified for age to jointly model internalizing and externalizing disorders. In this case, random slopes allow the correlation structure to depend on age and may offer a more realistic structure for repeated measures data with respect to random intercepts which imply a simpler, uniform exchangeable

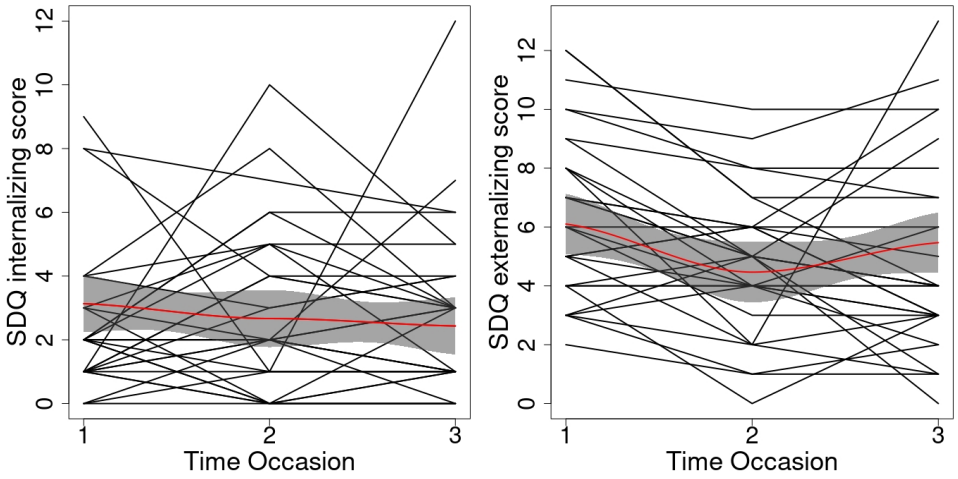

Figure 1.3. Individual trajectories for a random subsample of 30 children for SDQ internalizing (left) and externalizing (right) problems.

correlation structure (Tzavidis et al. 2016). We fitted the proposed model at quantile levels $\boldsymbol{\tau} = (0.25, 0.25)$, $\boldsymbol{\tau} = (0.50, 0.50)$ and $\boldsymbol{\tau} = (0.75, 0.75)$. Considering the 75-th percentile puts more emphasis on children with more severe problems generally associated with higher levels of SDQ scores. Even if we used the same quantile level for both SDQ scores, i.e., $\tau_1 = \tau_2$, our methodology allows researchers to focus on different parts of the SDQ score distributions simultaneously by considering different quantile levels for each of the two domains.

We estimated the QMHMM for a varying number of hidden states ($M = 2, \ldots, 7$) and mixture components ($G = 2, \ldots, 7$) employing the multi-start strategy described in Section 1.3, and then selected the optimal value of the pair (G, M) corresponding to the lowest BIC value. To enhance model interpretability and produce meaningful results, we retain only those solutions ensuring $\pi_g > 0.05$ for $g = 1, \ldots, G$ and $q_j > 0.05$ for $j = 1, \ldots, M$.

In addition to the proposed model, we compare our methodology with two well-known univariate alternatives for modeling longitudinal data: (i) the Linear Random Effects Model (LREM) with both random intercepts and slopes specified for age; (ii) the LQMM of Geraci & Bottai (2014) with both random intercepts and slopes specified for age at quantile levels τ, 0.25, 0.50 and 0.75. Specifically, the two models are estimated on SDQ_{Int} and SDQ_{Ext} scores independently. The reason why we consider the LREM is because it is a popular model for targeting the conditional expectation of the response given the explanatory variables. Whilst it produces efficient results when the normality assumptions hold, the LREM could potentially miss out important information related to other parts of the distribution of the outcome. In this case, the conditional mean may not offer the best summary; by contrast, the LQMM allows for modeling the entire conditional distribution of the outcome.

We start by commenting on the QMHMM results. Firstly, Table 1.6 reports the BIC values for the fitted models at the investigated quantile levels. As one can see,

the model selection procedure described in Section 1.3 leads to an increasing number of mixture components G equal to 3, 5 and 5, and a decreasing number of hidden states M equal to 5, 4 and 3 at quantile levels (0.25, 0.25), (0.50, 0.50) and (0.75, 0.75), respectively. The chosen values for G and M confirm the presence of constant and serial latent heterogeneity in the data, and support the exploratory analysis of individual SDQ trajectories in Figure 1.3.

M	2	3	4	5	6	7
			Panel A: $\boldsymbol{\tau} = (0.25, 0.25)$			
G						
2	146503.0	145105.3	144408.3	144830.9	146901.8	147143.7
3	146563.8	145128.4	144211.1	**143924.2**	145989.6	-
4	146569.7	145025.1	144116.5	144646.4	146198.6	-
5	146505.6	145027.0	144149.8	-	-	-
6	146504.8	-	-	-	-	-
7	-	-	-	-	-	-
			Panel B: $\boldsymbol{\tau} = (0.50, 0.50)$			
G						
2	150189.7	148382.3	148139.2	147586.2	147751.5	148352.8
3	150077.1	148241.8	148168.0	147522.3	148353.3	149095.7
4	150144.2	148350.3	148066.7	147685.4	148619.4	149037.1
5	150276.3	148423.3	**147215.9**	147578.3	148390.6	
6	150099.4	148072.9	147515.9	-	-	-
7	150149.6	148042.5	147883.2	-	-	-
			Panel C: $\boldsymbol{\tau} = (0.75, 0.75)$			
G						
2	158438.8	155016.5	155035.3	156646.2	-	-
3	158206.9	154993.0	155048.5	-	-	-
4	158470.3	154990.6	155870.4	-	-	-
5	158370.4	**154919.3**	155305.7	-	-	-
6	158386.2	156523.8	156528.4	-	-	-
7	158419.3	158532.2	-	-	-	-

Table 1.6. BIC values for a varying number of mixture components G and hidden states M. Bold font highlights the best values for the considered criterion (lower-is-better) while "−" denotes that the solution has been discarded because some $\pi_g, g = 1, \ldots, G$ or $q_j, j = 1, \ldots, M$ are less than 0.05.

Tables 1.7 and 1.9 report point estimates of the parameters and standard errors (in parentheses) based on $B = 1000$ bootstrap re-samples for the selected models at the investigated quantile levels. Parameter estimates are displayed in boldface when significant at the standard 5% level.

The second crucial finding is that the coefficient estimates vary with the quantile level $\boldsymbol{\tau}$. In particular, increasing adverse life events, socio-economic disadvantage, maternal depression and low maternal education are associated with higher SDQ

scores. The effect of these covariates appears to be more pronounced when looking at the upper tail compared with the lower tail of the distribution. Regarding income, there is evidence that poorer children are more likely to suffer from both physical and mental health problems (Currie 2009), hence the role of family income is likely to be concentrated at low incomes (see Fitzsimons et al. 2017).

Moreover, maternal depression has a more pronounced effect at the top end where children display critical levels of adjustment problems than at the bottom end of the distribution (see Kiernan & Huerta 2008). These considerations suggest that low socioeconomic status creates stress within the household, causing poor child health. On the other hand, gender and neighborhood deprivation are significantly associated with internalizing scores only up to the 25-th and 50-th percentiles, respectively; meanwhile ethnicity and ethnic stratification variables do not appear to be associated with the response at the investigated quantiles, except at the median level.

By looking at the fixed parameter estimates for the SDQ_{Ext}, it is possible to observe that, in contrast with internalizing scores, boys presents more externalizing problems than girls (Flouri & Sarmadi 2016) and the effect is more exacerbated in the right tail of the distribution. In general, girls are at lower risk of behavioral problems than boys which experience an increased risk for conduct and hyperactivity problems (Carona et al. 2014). Stressful life events, socio-economic disadvantage, maternal depression and maternal education are all significantly associated with internalizing scores. The effect of the included covariates is not uniform across quantiles but it is more apparent as the quantile level increases. This highlights the importance of considering a quantile regression approach revealing some possible underlying truth that can not be detected by classical mean regression. Moreover, the impact of such variables is more pronounced, across the distribution, on externalizing than on internalizing scores. This is consistent with other studies on behaviour disorders in child psychopathology claiming that poverty and material deprivation and education are more strongly related with children's externalizing problems compared with internalizing problems (see Costello et al. 2003 and Dearing et al. 2006). Overall, point estimates of regression coefficients are consistent with child development theory, as well as with the results discussed in Tzavidis et al. (2016) and in Alfò et al. (2021).

In order to highlight the practical relevance of the proposed methodology, we compare our findings with the parameter estimates of the univariate LREM and LQMM reported in Table 1.8. At first, we observe that the LQMM results are generally in line with our findings, except for $\tau = 0.25$. However, in this case we experienced slow convergence of the algorithm when fitting the model (see also Tzavidis et al. 2016). Overall, we notice that the estimated parameters for the SDQ_{Ext} are greater than those for the SDQ_{Int} scores and the effect of the considered covariates increases moving towards the right tail of the conditional distribution of each outcome. Consistently with the QMHMM, mother's education and family environmental risk factors are important for predicting emotional and behavioural disorders in early childhood. Further, males typically present higher SDQ_{Int} and SDQ_{Ext} problems than females whereas, contrary to the results obtained from the QMHMM, lower SDQ_{Int} scores are generally associated with white children.
In addition to that, both the LQMM and LREM analyze children's disorders by fitting two univariate models separately and hence, they disregard the possible association between the SDQ scores. In contrast, one of the main benefits of

the proposed multivariate approach is the possibility to study the magnitude and direction of the dependence structure between the responses at different quantile levels of interest. Following Kotz et al. (2012), we can compute the correlation between SDQ scores using (1.6) and understand whether their association structure becomes stronger for children with more pronounced problems. In particular, the estimated correlation coefficient, r_{12}, reported in Table 1.7 gives a measure of tail correlation and, consistently with the recent work of Alfò et al. (2021), it indicates that internalizing and externalizing disorders are positively associated and this association increases with the quantile level $\boldsymbol{\tau}$. Hence, children may present a constellation of symptoms comprised of both disorders which is aggravated in disadvantaged ones by the accumulation of risk factors. This finding is also in line with the existence of positive covariation among psychiatric diagnoses (see Lilienfeld 2003, Liu 2004 and Cicchetti & Toth 2014).

We conclude the analysis by reporting selected diagnostics for the fitted QMHMM. Table 1.9 summarizes the estimates for the initial and transition probabilities of the hidden Markov chain. The transition matrices describe how, and how frequently, children move from low to high level of disorders. For all examined quantiles, they do not tend to move through states over time. At first, it is worth noting that the estimated state-dependent intercepts $\boldsymbol{\alpha}$, consistently with the quantile regression framework, tend to increase when moving from lower to upper quantiles resulting in higher levels of children's disorders. For $\boldsymbol{\tau} = (0.25, 0.25)$ (Panel A), the initial probability distribution defined on $\mathcal{S}$ is relatively uniformly distributed and the probability of not moving from states 1 and 4 is also very high, i.e. $\widehat{q}_{11} = 0.991$ and $\widehat{q}_{44} = 0.993$, respectively. This implies that almost every child in the lower tails of the outcomes distributions, starts and maintains low-level disorders over time. If any transition is observed, units tend to move towards states 1 and 4 with lower intercepts and a reduction in juvenile developmental disorders with temporary jumps to moderate values of disorders. On the other hand, for $\boldsymbol{\tau} = (0.50, 0.50)$ (Panel B), by looking at the initial probabilities one can see that half of the units ($\widehat{q}_3 = 0.415$) starts the study with low values of emotional and behavioral disorders and transitions between states are unlikely. When $\boldsymbol{\tau} = (0.75, 0.75)$ (Panel C), the majority of children starts with moderate values of developmental difficulties ($\widehat{q}_1 + \widehat{q}_2 > 0.80$) and transitions to more severe disorders are more frequent.

It is also noteworthy that suitable constraints can be applied on the parameters space of the model to ease the interpretation of results on the hidden process or test hypotheses on the corresponding parameters. Specifically, these constraints may be posed on the random intercepts or on the initial and transition probabilities (Bartolucci et al. 2012). In these cases, the ML estimates of the parameters can be obtained using the EM algorithm in Section 1.3, in which only the M-step has to be modified according to the constraints of interest.

Finally, Figure 1.4 shows the estimated marginal cumulative density functions of the discrete random slopes for both SDQ outcomes. In both plots, it is clear that the estimated distribution functions depart substantially from the Gaussian distribution, having pronounced asymmetries. Hence, the underlying assumption of normally distributed random intercepts in the LQMM and LREM specifications is inappropriate. In contrast, the discrete mixture performs is more flexible and is able to accommodate possible departures from the Gaussianity assumption (Alfò et al.

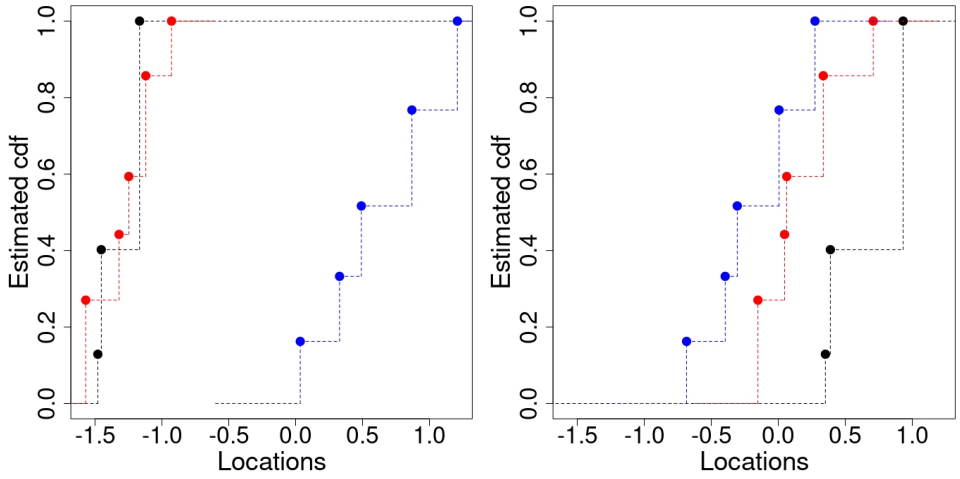

Figure 1.4. Estimated cumulative density function of the discrete random slopes for SDQ internalizing (left) and externalizing (right) problems scores at the 0.25 (black), 0.50 (red) and 0.75 (blue) quantile levels.

2017).

τ-th quantile	(0.25, 0.25) [G = 3, M = 5]		(0.50, 0.50) [G = 5, M = 4]		(0.75, 0.75) [G = 5, M = 3]	
Variable	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}
Age year scal	**−0.053** (0.010)	**−0.289** (0.016)	**−0.051** (0.009)	**−0.440** (0.011)	−0.040 (0.021)	**−0.450** (0.031)
Age2 year scal	**0.045** (0.005)	**0.193** (0.009)	**0.077** (0.011)	**0.208** (0.013)	**0.104** (0.018)	**0.287** (0.024)
ALE 11	**0.022** (0.008)	**0.036** (0.016)	**0.086** (0.018)	**0.113** (0.019)	**0.116** (0.039)	**0.205** (0.055)
SED 4	**0.070** (0.023)	**0.105** (0.045)	**0.175** (0.030)	**0.221** (0.030)	**0.201** (0.051)	**0.398** (0.076)
Kessm	**0.090** (0.009)	**0.143** (0.012)	**0.167** (0.009)	**0.189** (0.012)	**0.208** (0.018)	**0.299** (0.025)
Degree	**−0.350** (0.109)	**−0.894** (0.149)	**−0.526** (0.114)	**−1.482** (0.171)	**−0.703** (0.160)	**−1.267** (0.232)
GCSE	**−0.217** (0.110)	**−0.582** (0.150)	**−0.352** (0.113)	**−0.430** (0.166)	**−0.413** (0.149)	**−0.427** (0.213)
White	−0.090 (0.061)	−0.143 (0.097)	−0.075 (0.149)	0.059 (0.160)	−0.216 (0.203)	0.320 (0.303)
Male	**−0.062** (0.016)	**0.793** (0.030)	0.027 (0.037)	**0.950** (0.045)	0.082 (0.080)	**0.944** (0.119)
IMD	**−0.022** (0.005)	**−0.036** (0.009)	**−0.025** (0.008)	**−0.027** (0.010)	−0.027 (0.020)	−0.045 (0.029)
Eng eth stratum	−0.043 (0.159)	−0.168 (0.174)	0.122 (0.193)	0.144 (0.220)	0.174 (0.241)	−0.025 (0.374)
Eng dis stratum	−0.003 (0.031)	0.003 (0.072)	**0.085** (0.042)	**0.106** (0.050)	0.156 (0.111)	0.337 (0.175)
α_1	**0.437** (0.001)	**1.040** (0.114)	**0.983** (0.013)	**4.769** (0.049)	**2.082** (0.016)	**2.954** (0.054)
α_2	**3.930** (0.051)	**4.140** (0.089)	**2.480** (0.007)	**8.329** (0.077)	**2.735** (0.119)	**7.769** (0.014)
α_3	**1.008** (0.005)	**6.362** (0.000)	**1.164** (0.016)	**1.954** (0.044)	**6.957** (0.091)	**6.347** (0.201)
α_4	**1.523** (0.130)	**1.281** (0.099)	**4.829** (0.008)	**4.670** (0.128)		
α_5	**0.477** (0.043)	**3.115** (0.013)				
r_{12}	**0.356** (0.016)		**0.211** (0.023)		**0.602** (0.021)	
$\ell(\boldsymbol{\Phi}_\tau)$	-71665.9		-73333.3		-77223.6	
ν_f	69		64		55	
AIC	143470.0		146794.6		154557.2	
BIC	143924.2		147215.9		154919.3	

Table 1.7. Point estimates with standard errors in parentheses for different quantile levels. Parameter estimates are displayed in boldface when significant at the standard 5% level.

τ-th quantile	(0.25, 0.25)		(0.50, 0.50)		LREM - Mean		(0.75, 0.75)	
Variable	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}	SDQ_{Int}	SDQ_{Ext}
Intercept	**2.025** (0.173)	**3.308** (0.215)	**2.325** (0.176)	**3.821** (0.219)	**2.629** (0.148)	**4.364** (0.221)	**2.742** (0.180)	**4.615** (0.235)
Age year scal	−0.000 (0.015)	**−0.408** (0.017)	**−0.037** (0.010)	**−0.453** (0.014)	−0.003 (0.009)	**−0.451** (0.012)	0.002 (0.012)	**−0.478** (0.018)
Age2 year scal	0.000 (0.009)	**0.205** (0.015)	**0.069** (0.010)	**0.246** (0.010)	**0.083** (0.007)	**0.252** (0.008)	**0.102** (0.013)	**0.267** (0.016)
ALE 11	0.000 (0.011)	**0.106** (0.025)	**0.062** (0.018)	**0.100** (0.030)	**0.085** (0.016)	**0.108** (0.021)	**0.097** (0.028)	**0.153** (0.038)
SED 4	0.000 (0.028)	**0.191** (0.058)	**0.132** (0.027)	**0.204** (0.041)	**0.124** (0.024)	**0.186** (0.032)	**0.209** (0.046)	**0.287** (0.091)
Kessm	0.000 (0.032)	**0.160** (0.018)	**0.142** (0.012)	**0.201** (0.013)	**0.156** (0.007)	**0.189** (0.009)	**0.201** (0.014)	**0.217** (0.023)
Degree	**−0.673** (0.110)	**−1.357** (0.163)	**−0.549** (0.113)	**−1.323** (0.159)	**−0.715** (0.097)	**−1.602** (0.149)	**−0.475** (0.126)	**−1.287** (0.177)
GCSE	**−0.571** (0.111)	**−0.690** (0.164)	**−0.374** (0.108)	**−0.502** (0.168)	**−0.376** (0.092)	**−0.692** (0.141)	−0.170 (0.111)	**−0.408** (0.200)
White	**−0.455** (0.119)	−0.020 (0.135)	**−0.232** (0.105)	0.300 (0.155)	**−0.371** (0.090)	0.115 (0.138)	−0.147 (0.110)	0.261 (0.169)
Male	0.012 (0.025)	**0.656** (0.088)	**0.080** (0.039)	**0.936** (0.089)	**0.150** (0.049)	**0.940** (0.075)	**0.193** (0.078)	**1.227** (0.144)
IMD	−0.009 (0.012)	−0.025 (0.017)	**−0.023** (0.010)	**−0.039** (0.018)	**−0.044** (0.010)	**−0.049** (0.015)	−0.014 (0.016)	−0.018 (0.031)
Eng eth stratum	0.050 (0.127)	−0.194 (0.186)	0.114 (0.112)	−0.017 (0.191)	0.178 (0.111)	−0.011 (0.169)	0.224 (0.120)	0.319 (0.197)
Eng dis stratum	0.001 (0.001)	0.162 (0.099)	0.054 (0.050)	**0.208** (0.079)	0.068 (0.061)	**0.230** (0.092)	0.099 (0.075)	**0.453** (0.160)
σ^2	0.750	4.508	2.065	5.297	2.428	6.282	2.976	5.973
σ^2_{age}	0.000	0.178	0.060	0.182	0.151	0.267	0.224	0.339

Table 1.8. Univariate LREM for the mean and LQMM results for internalizing and externalizing scores at the investigated quantile levels. Standard errors (in parentheses) are based on 200 bootstrap re-samples. Parameter estimates are displayed in boldface when significant at the standard 5% level.

	1	2	3	4	5
			Panel A: $\boldsymbol{\tau} = (0.25, 0.25)$		
q	0.201 (0.017)	0.111 (0.013)	0.269 (0.017)	0.165 (0.017)	0.254 (0.020)
1	0.991 (0.016)	0.000 (0.000)	0.000 (0.000)	0.009 (0.004)	0.000 (0.029)
2	0.000 (0.002)	0.883 (0.044)	0.007 (0.019)	0.110 (0.008)	0.000 (0.012)
3	0.000 (0.000)	0.089 (0.012)	0.723 (0.032)	0.000 (0.000)	0.187 (0.014)
4	0.003 (0.015)	0.004 (0.042)	0.000 (0.001)	0.993 (0.009)	0.000 (0.013)
5	0.096 (0.003)	0.016 (0.009)	0.027 (0.029)	0.029 (0.000)	0.832 (0.034)
			Panel B: $\boldsymbol{\tau} = (0.50, 0.50)$		
q	0.287 (0.019)	0.160 (0.016)	0.415 (0.018)	0.138 (0.014)	
1	0.863 (0.036)	0.034 (0.034)	0.071 (0.007)	0.032 (0.024)	
2	0.053 (0.017)	0.850 (0.038)	0.000 (0.000)	0.096 (0.013)	
3	0.006 (0.031)	0.000 (0.000)	0.967 (0.013)	0.027 (0.034)	
4	0.049 (0.015)	0.015 (0.024)	0.067 (0.011)	0.868 (0.036)	
			Panel C: $\boldsymbol{\tau} = (0.75, 0.75)$		
q	0.541 (0.021)	0.293 (0.017)	0.165 (0.015)		
1	0.942 (0.011)	0.006 (0.020)	0.052 (0.029)		
2	0.072 (0.004)	0.825 (0.024)	0.103 (0.020)		
3	0.185 (0.010)	0.109 (0.015)	0.707 (0.031)		

Table 1.9. Initial probabilities, **q**, and transition probabilities, **Q**, estimates for different quantiles.

1.6 Conclusions

Longitudinal data allows us to understand the evolution of a certain phenomenon over time. In this context, it becomes of crucial importance to determine an appropriate modeling framework to assess the effects of unobserved factors and hidden heterogeneity which can be either time-invariant or time-varying; ignoring these factors may induce bias and lead to invalid conclusions. Moreover, the literature on this topic which is traditionally focused on the conditional mean, might not provide a good summary of the response distribution. To account for the complex data structure, this work generalizes the multivariate quantile approach of Petrella & Raponi (2019) for the analysis of multivariate longitudinal data by combining the features of quantile regression and MHMMs (Altman 2007). The proposed model allows for the quantile-specific effects to be quantified and jointly modeling of several outcomes. The model further allows for different sources of heterogeneity to be distinguished, i.e. between individual heterogeneity and time heterogeneity are modeled through the state-specific effects. In order to avoid possibly misleading inferences caused by erroneous assumption on the random effects distribution, we

rely on the NPML estimation theory and we approximate this distribution by a multivariate discrete latent variable.

As illustrated in the real data application, the proposed method models simultaneously the quantiles of children's emotional and behavioral disorders as a function of demographic and socio-economics risk factors. The results show that behavioral and emotional difficulties are mainly affected by the family poverty conditions and mother's characteristics. Such effects are much stronger in the upper tail of the response distribution, i.e. for those children experiencing more severe internalizing and externalizing problems. In addition, the analysis reveals moderate levels of codependency between internalizing and externalizing disorders, that cannot be detected by univariate models.

The methodology can be further extended to allow for a non-homogeneous hidden Markov process where transition probabilities are allowed to depend on covariates. Finally, the hidden Markov chain implicitly assumes that the sojourn time, i.e. the number of consecutive time points that the process spends in a given state, is geometrically distributed. As a further generalization of this work, one may consider a semi-Markov process which is designed to relax this condition by allowing the sojourn time to be modeled by more flexible distributions.

1.7 Appendix

Proof of Proposition 1

Under the constraints imposed on $\tilde{\xi}$ and $\mathbf{\Lambda}$, the representation in (1.5) implies that:

$$\mathbf{Y} \mid \tilde{C} = \tilde{c} \sim \mathcal{N}_p(\boldsymbol{\mu} + \mathbf{D}\tilde{\xi}\tilde{c},\, \tilde{c}\mathbf{D}\mathbf{\Sigma}\mathbf{D}), \qquad \tilde{C} \sim \text{Exp}(1). \tag{1.34}$$

This implies that the joint density function of $\mathbf{Y}$ and $\tilde{C}$ is:

$$f_{\mathbf{Y},\tilde{C}}(\mathbf{y},\tilde{c}) = \frac{\exp\left\{(\mathbf{y}-\boldsymbol{\mu})'\mathbf{D}^{-1}\mathbf{\Sigma}^{-1}\tilde{\xi}\right\}}{(2\pi)^{p/2} \mid \mathbf{D}\mathbf{\Sigma}\mathbf{D} \mid^{1/2}} \left(\tilde{c}^{-p/2} \exp\left\{-\frac{1}{2}\frac{\tilde{m}}{\tilde{c}} - \frac{1}{2}\tilde{c}(\tilde{d}+2)\right\}\right). \tag{1.35}$$

Then, the complete log-likelihood function (up to additive constant terms) can be written as follows:

$$\begin{aligned}\log \ell_c(\mathbf{\Phi}_{\boldsymbol{\tau}} \mid \mathbf{y}, \mathbf{x}, \tilde{\mathbf{c}}, \mathbf{s}, \mathbf{b}) = \sum_{i=1}^{N}\Bigg\{ &\sum_{g=1}^{G} w_{ig} \log \pi_g + \sum_{j=1}^{M} u_{i1j} \log q_j + \sum_{t=2}^{T_i}\sum_{j=1}^{M}\sum_{k=1}^{M} v_{itjk} \log q_{jk} \\ &+ \sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg} \log f_{\mathbf{Y},\tilde{C}}(\mathbf{y}_{it}, \tilde{c}_{it} \mid \mathbf{x}_{it}, S_{it}=j, \mathbf{b}_g)\Bigg\}.\end{aligned} \tag{1.36}$$

By substituting (1.35) in (1.36), we obtain:

$$\begin{aligned}\ell_c(\mathbf{\Phi}_{\boldsymbol{\tau}}) = \sum_{i=1}^{N}\Bigg\{ &\sum_{g=1}^{G} w_{ig} \log \pi_g + \sum_{j=1}^{M} u_{i1j} \log q_j + \sum_{t=2}^{T_i}\sum_{j=1}^{M}\sum_{k=1}^{M} v_{itjk} \log q_{jk} \\ &- \frac{1}{2} T_i \log \mid \mathbf{D}\mathbf{\Sigma}\mathbf{D} \mid + \sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})'\mathbf{D}^{-1}\mathbf{\Sigma}^{-1}\tilde{\xi} \\ &- \frac{1}{2}\sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg} \frac{1}{\tilde{C}_{itjg}} (\mathbf{Y}_{it} - \boldsymbol{\mu}_{it})'(\mathbf{D}\mathbf{\Sigma}\mathbf{D})^{-1}(\mathbf{Y}_{it} - \boldsymbol{\mu}_{it}) \\ &- \frac{1}{2}\tilde{\xi}'\mathbf{\Sigma}^{-1}\tilde{\xi}\sum_{t=1}^{T_i}\sum_{j=1}^{M}\sum_{g=1}^{G} z_{itjg}\tilde{C}_{itjg}\Bigg\}.\end{aligned} \tag{1.37}$$

Proof of Proposition 2

The E-step of the EM algorithm considers the conditional expectation of the complete log-likelihood function given the observed data and the current parameter estimates $\hat{\mathbf{\Phi}}_{\boldsymbol{\tau}}^{(r-1)}$. The conditional expectations of $w_{ig}, u_{itj}, v_{itjk}$ and z_{itjg} can be computed using standard arguments in the HMM literature as shown in (1.17). To compute the conditional expectation of $\tilde{C}$ and $\tilde{C}^{-1}$, in the E-step of the EM algorithm, $\tilde{C}$ is treated as an additional latent variable and, hence, not observable. Using the joint distribution of $\mathbf{Y}$ and $\tilde{C}$ derived in (1.35) and the MAL density of $\mathbf{Y}$ given in (1.3), we have that:

$$f_{\tilde{C}}(\tilde{C} \mid \mathbf{Y} = \mathbf{y}) = \frac{f_{\tilde{C},\mathbf{Y}}(\tilde{c},\mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} = \frac{\tilde{c}^{-p/2}\left(\frac{2+\tilde{d}}{\tilde{m}}\right)^{\nu/2} \exp\left\{-\frac{\tilde{m}}{2\tilde{c}} - \frac{\tilde{c}(2+\tilde{d})}{2}\right\}}{2K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}}\right)}, \tag{1.38}$$

which corresponds to a Generalized Inverse Gaussian (GIG) distribution with parameters $\nu, 2+\tilde{d}, \tilde{m}_i$, i.e.[1]

$$f_{\tilde{C}}(\tilde{C} \mid \mathbf{Y}=\mathbf{y}) \sim \text{GIG}\left(\nu, \tilde{d}+2, \tilde{m}\right). \tag{1.39}$$

It follows that

$$\mathbb{E}[\tilde{C} \mid \cdot] = \left(\frac{\hat{\tilde{m}}}{2+\hat{\tilde{d}}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)} \tag{1.40}$$

and

$$\mathbb{E}[\tilde{C}^{-1} \mid \cdot] = \left(\frac{2+\hat{\tilde{d}}}{\hat{\tilde{m}}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}}\right)} - \frac{2\nu}{\hat{\tilde{m}}}. \tag{1.41}$$

Denoting the two conditional expectations in (1.40) and (1.41) by $\hat{\tilde{c}}$ and $\hat{\tilde{z}}$ respectively, concludes the proof.

Proof of Proposition 3

Imposing the first order conditions on (1.18) with respect to each component of the set $\boldsymbol{\Phi}_{\boldsymbol{\tau}}$, gives the parameter estimates in (1.21), (1.22) and (1.25). However, there is not closed formula solution to update the elements of the scale matrix $\mathbf{D}$; hence, the M-step update requires using numerical optimization techniques to maximize (1.18). A considerable disadvantage of this procedure is the necessary high computational effort which could be very time-consuming. For this reason, we utilize a simpler estimator for the scale parameters $d_j, j=1,\ldots,p$ which follows directly from the fact that all marginals of the MAL distribution are univariate AL distributions (see Yu & Zhang 2005 and Marino et al. 2018):

$$\hat{d}_j = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{g=1}^{G} \sum_{k=1}^{M} \hat{z}_{itkg} \rho_{\tau_j}(Y_{it}^{(j)} - \hat{\mu}_{it}^{(j)}). \tag{1.42}$$

[1] The pdf of a GIG(p,a,b) distribution is defined as $f_{GIG}(x;p,a,b) = \frac{\left(\frac{a}{b}\right)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} e^{-\frac{1}{2}\left(ax+bx^{-1}\right)}$, with $a>0$, $b>0$ and $p \in \mathcal{R}$.

Chapter 2

Forecasting VaR and ES using a joint quantile regression and its implications in portfolio allocation

2.1 Introduction

The events of the ongoing credit crisis and past financial crises have emphasized the necessity for appropriate risk measures. The use of quantitative risk measures has become an essential management tool providing advice, analysis and support for financial and asset management decisions to market participants and regulators. The most widely used risk measure is Value at Risk (VaR). VaR measures the maximum loss that a financial operator can incur over a defined time horizon and for a given confidence level. Its clear meaning and computational ease made it very popular among practitioners, so much so that it has widely infiltrated the banking regulatory framework. However, VaR has a number of drawbacks (Artzner et al. 1997, 1999). First, VaR does not account for tail risk; i.e., it does not warn us about the size of the losses that occur with a probability lower than the predetermined confidence level. Second, VaR is not a "coherent" risk measure (Artzner et al. 1999) since it does not satisfy the sub-additivity property, and hence, it does not take into consideration the benefits of diversification. As a result, investors and risk managers are likely to construct positions with unintended weaknesses that result in greater losses under conditions beyond the VaR level (Yamai & Yoshiba 2005). Market participants could solve such problems by adopting the Expected Shortfall (ES) risk measure, which is defined as the conditional expectation of exceedances beyond VaR (see Acerbi & Tasche 2002 and Rockafellar & Uryasev 2000). Unlike VaR, ES is a coherent risk measure and provides more information on the shape and the heaviness of the tails of the loss distribution. Therefore, ES has gained increasing attention from risk managers, banking regulators and investors as an alternative measure of risk, complementing the VaR measure.

However, despite its interesting properties, and in contrast with VaR, little work exists on modeling ES. This is in part because ES is not an "elicitable" measure, in

the sense that there does not exist a loss function such that ES is the solution that minimizes the expected loss. Several works have been proposed in the literature to overcome the problem of elicitability (see, e.g., Engle & Manganelli 2004, Cai & Wang 2008, Taylor 2008, Zhu & Galbraith 2011, Du & Escanciano 2017, Patton et al. 2019 and Bu et al. 2019). Recently, using the results of Fissler & Ziegel (2016), who show that ES is jointly elicitable with VaR, Taylor (2019) uses the Asymmetric Laplace (AL) distribution to jointly estimate dynamic models for both VaR and ES. In particular, Taylor (2019) shows that the negative of the log-likelihood associated with the AL distribution belongs to the class of loss functions presented in Fissler & Ziegel (2016), and hence it can be used to estimate and forecast the VaR and ES measures in one step. In his paper, the joint estimation of VaR and ES is obtained in a univariate quantile regression framework, exploiting the interesting result that ES can be expressed in terms of the scale parameter of the AL density.

The literature mentioned above, however, has mainly focused on univariate time series, which completely disregards the strong interrelation among assets in financial markets. To capture the degree of tail interdependence between assets, several quantile-based methods have also been proposed to estimate VaR, but they do not specify a model for the ES component; see, for example, the relevant works of Baur (2013), Bernardi et al. (2015), White et al. (2015), Kraus & Czado (2017) and Bonaccolto et al. (2019).

In this paper, we extend the univariate approach of Taylor (2019) to a multivariate framework, with the objective of obtaining joint estimates of both VaR and ES for multiple financial assets simultaneously, accounting for their dependence structure. To this end, we generalize the Multivariate Asymmetric Laplace (MAL) quantile regression approach of Petrella & Raponi (2019) to a time-varying setting by allowing the parameters of the MAL distribution to vary over time. For each asset, we model the evolution of VaR and ES as functions of the location and scale parameters of the distribution. In particular, for the VaR component, we adopt a Conditional Autoregressive Value at Risk (CAViaR) specification (Engle & Manganelli 2004).

The advantages of our methodology are manifold. First, our approach is a joint modeling framework where both the model parameters and the pair (VaR, ES) of multiple returns are estimated simultaneously, generalizing the univariate results of Bassett et al. (2004) and Taylor (2019). Second, our theory captures empirical characteristics of financial data such as peakedness, skewness, and heavy tails (see e.g., Kraus & Litzenberger 1976, Friend & Westerfield 1980 and Barone-Adesi 1985) without relying on the limitation of normally distributed returns.

The inferential problem is solved by developing a suitable Expectation-Maximization (EM) algorithm, which exploits the mixture representation of the MAL distribution (see Petrella & Raponi 2019) properly generalized to the case of time-varying parameters. The finite sample properties of the proposed estimation method are also evaluated using a simulation exercise, where we show the validity and the robustness of our procedure under different data generating processes.

A further contribution of the paper concerns the evaluation of VaR and ES in the context of portfolio optimization (see, e.g., Yiu 2004 and Alexander & Baptista 2008). In recent years, the MAL density has attracted wide attention in the literature because of its flexibility in modeling financial data (Mittnik & Rachev 1991, Kotz et al. 2012 and Paolella 2015) and for its interesting properties, which can be exploited to

derive optimal portfolio allocations (see Zhao et al. 2015 and Shi et al. 2018). In the classic Mean-Variance (MV) methodology of Markowitz (1952), portfolio risk is measured using the standard deviation of the portfolio. However, the MV approach is reasonably applicable only in cases where either the returns follow a Gaussian distribution or the investor utility function is quadratic. Given the empirical evidence showing that market participants have a preference for positive skewness and they are more concerned about the downside risk (see Arditti 1971 and Konno & Suzuki 1995, among others), the MAL distribution could represent a more effective tool for selecting optimal portfolio allocations in the case of risk-averse agents. Therefore, in this paper, we exploit the MAL properties to incorporate skewness directly into the portfolio optimization method and to identify the optimal allocation weights. We then compute the corresponding portfolio VaR and ES as a function of the multivariate structure of the data. We prove that this result follows directly from the fact that any linear combination of the MAL components is still AL distributed, with location, skew and scale parameters that are functions of the MAL parameters and the portfolio weights. Therefore, once we obtain the Maximum Likelihood (ML) estimates of the MAL parameters from the proposed dynamic quantile regression model, we fix a desired level of risk for any target portfolio and search for the optimal allocation weights according to the adopted strategy.

Specifically, we consider the Skewness Mean-Variance (SMV) strategy of Zhao et al. (2015), where the optimal allocation is obtained by minimizing the portfolio variance, while controlling for the skewness of the asset returns. However, Zhao et al. (2015) employed the method of moments to estimate the portfolio variance; in contrast, we estimate the MAL parameters in a ML framework by using an EM algorithm.

Empirically, we analyze the weekly returns of the FTSE 100, NIKKEI 225 and Standard & Poor's 500 (S&P 500) market indices from April 1985 to February 2021. In a first out-of-sample exercise, we jointly estimate the VaR and ES of the three stock market indices using the proposed dynamic joint quantile regression model, hence taking into account the correlation among the three indices. To evaluate VaR and ES forecasts and to show the main advantages of the proposed method, we use different backtesting procedures, where we compare the out-of-sample VaR and ES predictions with the ones obtained by applying the univariate method of Taylor (2019). In particular, to perform a joint evaluation of VaR and ES, we follow Fissler et al. (2015), Nolde et al. (2017), Patton et al. (2019) and Taylor (2019) and extend their approach by introducing a new scoring function based on the MAL distribution. We find that our multivariate method always provides more accurate VaR and ES predictions than other well-known approaches, such as the Quantile AutoRegression of Koenker & Xiao (2006) and the dynamic quantile regression of Taylor (2019). Moreover, in line with Taylor (2019), our results show that the Asymmetric Slope CAViaR specification of Engle & Manganelli (2004) yields the best VaR and ES forecasts for all three indices at different quantile levels, confirming the existence of relevant asymmetries in the impact of positive and negative returns.

In a second empirical analysis, we aggregate the stock market indices to form a financial portfolio with a predetermined level of risk, and we estimate its optimal allocation weights by implementing our new optimization procedure. We then compute the out-of-sample portfolio's VaR and ES and evaluate the predictions

using the univariate backtesting procedures of Taylor (2019), Nolde et al. (2017) and Patton et al. (2019). The empirical analysis reveals that our multivariate method produces the lowest average losses compared to other existing strategies based on the multivariate Normal and Student-t distributions, regardless of the scoring function used. In addition, we find that the proposed methodology yields the highest Sharpe Ratio overall as well as the least concentrated portfolio allocations.

The rest of this paper is organized as follows. In Section 2.2, we introduce the dynamic multiple quantile regression and propose a joint model for VaR and ES. We then illustrate the EM-based ML approach for the simultaneous estimation of VaR and ES. Section 2.3 develops the portfolio allocation problem. Section 2.4 introduces a new scoring function for the joint evaluation of VaR and ES forecasts. In Section 2.5 we discuss the main empirical results, while Section 2.6 concludes the paper. All the proofs are provided in Appendix A and the simulation study is presented in Appendix B.

2.2 Multivariate framework

In this paper, we generalize the univariate regression approach of Taylor (2019). Specifically, by extending the MAL density of Petrella & Raponi (2019) – allowing the location and scale parameters of the MAL distribution to vary over time – we estimate the pair of VaR and ES associated with each asset using a joint quantile regression framework. In this way, we are able to calculate the time-varying VaR and ES simultaneously for all marginal response variables, accounting for possible correlation among the considered assets. For the VaR components, we assume a CAViaR specification (see Engle & Manganelli 2004). Parameter estimation is carried out using a suitable EM algorithm as in Petrella & Raponi (2019), properly extended to deal with the time-varying setting. In this way, the estimated parameters account for tail interdependence among multiple returns, and they convey this information to the VaR and ES estimates.

We start by introducing the time-varying joint quantile regression model in Section 2.2.1, where we consider a dynamic generalization of the MAL density proposed in Petrella & Raponi (2019). We then show in Section 2.2.2 how the resulting time-varying scale parameter of the MAL distribution can be used to model the ES vector and derive a parsimonious approach for the simultaneous estimation of VaR and ES in a multidimensional setting. The parameter estimation and the EM algorithm are described in Section 2.2.3.

2.2.1 Dynamic joint quantile regression

Let $\mathbf{Y}_t = [Y_{t1}, ..., Y_{tp}]'$ be a p-variate response variable and denote by $\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1})$ the τ_j-quantile function of the j-th component of $\mathbf{Y}_t$ conditional on the information set $\mathcal{F}_{t-1}$ available at time $t-1$, for $j = 1, ..., p$ and $t = 1, ..., T$. Then, for a given τ_j, we consider the following autoregressive dynamic:

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \omega_j + \eta_j \mathcal{Q}_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \ell(\boldsymbol{\beta}_j, Y_{t-1j}), \tag{2.1}$$

where $\omega_j = \omega_j(\tau_j)$, $\eta_j = \eta_j(\tau_j)$ and $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j(\tau_j) = [\beta_{1j}, ..., \beta_{Kj}]'$ are model parameters that depend on the chosen level τ_j and where we suppress the index τ_j for

simplicity of notation. The dynamic specification in (2.1) is well-known in the literature as the CAViaR model of Engle & Manganelli (2004), which aims to compute the τ-th level VaR by estimating the τ-th level quantile of the asset returns through a conditional autoregressive structure. The function $\ell(\cdot)$ represents the so-called News Impact Curve (NIC), originally introduced by Engle & Ng (1993). For each j-th component, the NIC function essentially feeds back the last available observation (Y_{t-1j}) into the present value of the conditional quantile through the $K \times 1$ parameter vector $\boldsymbol{\beta}_j$. Following the CAViaR literature, we will consider different specifications for $\ell(\cdot)$ to model the marginal quantiles, which will be described in the next section.

Using matrix notation, the representation in (2.1) can be embedded in the following multivariate linear regression model:

$$\mathbf{Y}_t = \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \qquad t = 1, ..., T \tag{2.2}$$

where $\boldsymbol{\epsilon}_t$ denotes a $p \times 1$ vector of error terms, with each marginal quantile (at fixed levels $\tau_1, ..., \tau_p$) equal to zero, to ensure that $\boldsymbol{\mu}_t = \mathcal{Q}_{\mathbf{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$ at $\boldsymbol{\tau} = [\tau_1, ..., \tau_p]$. As in Chapter 1, the p quantile indices, $\tau_1, ..., \tau_p$, do not need to be the same for all the components of $\mathbf{Y}_t$.

To estimate the regression model in (2.2), we consider a dynamic generalization of the MAL distribution introduced in Petrella & Raponi (2019) and Kotz et al. (2012); i.e., we consider the time-varying distribution $\text{MAL}_p\left(\boldsymbol{\mu}_t, \mathbf{D}_t\tilde{\xi}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t\right)$, with density function:

$$f_{\mathbf{Y}_t}(\mathbf{y}_t|\boldsymbol{\mu}_t, \mathbf{D}_t\tilde{\xi}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t, \mathcal{F}_{t-1}) = \frac{2\exp\left\{(\mathbf{y}_t - \boldsymbol{\mu}_t)'\mathbf{D}_t^{-1}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\xi}\right\}}{(2\pi)^{p/2}|\mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t|^{1/2}} \left(\frac{\tilde{m}_t}{2+\tilde{d}}\right)^{\nu/2} K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}_t}\right). \tag{2.3}$$

In (2.3), $\boldsymbol{\mu}_t$ represents the location parameter vector, $\mathbf{D}_t\tilde{\xi} \in \mathcal{R}^p$ is the scale (or skew) parameter, with $\mathbf{D}_t = \text{diag}[\delta_{t1}, ..., \delta_{tp}]$, $\delta_{tj} > 0$ and $\tilde{\xi} = [\tilde{\xi}_1, ..., \tilde{\xi}_p]'$, having generic element $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$. $\tilde{\boldsymbol{\Sigma}}$ is a $p \times p$ positive definite matrix such that $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}$, with $\boldsymbol{\Psi}$ having the structure of a correlation matrix[1] and $\tilde{\boldsymbol{\Lambda}} = \text{diag}[\tilde{\sigma}_1, ..., \tilde{\sigma}_p]$, with $\tilde{\sigma}_j^2 = \frac{2}{\tau_j(1-\tau_j)}$, $j = 1, ..., p$. Moreover, $\tilde{m}_t = (\mathbf{y}_t - \boldsymbol{\mu}_t)'(\mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_t)$, $\tilde{d} = \tilde{\xi}'\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\xi}$, and $K_\nu(\cdot)$ denotes the modified Bessel function of the third kind with index parameter $\nu = (2-p)/2$.

Notice that, as stressed in Petrella & Raponi (2019), the specification in (2.3) should not be viewed as a parametric assumption in model (2.2) but rather as a convenient tool to jointly estimate the marginal dynamic quantiles of a multivariate response variable in a quantile regression framework. Moreover, as clarified in their paper, the constraints $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$ and $\tilde{\sigma}_j^2 = \frac{2}{\tau_j(1-\tau_j)}$ must be imposed to guarantee model identifiability (see Petrella & Raponi 2019, Proposition 2) and to ensure that the dynamic quantile specification in (2.1) holds, i.e., that $\mathbb{P}(Y_{tj} < \mu_{tj}) = \tau_j$ holds for each $j = 1, ..., p$.

[1] In greater detail, $\boldsymbol{\Psi}$ represents the correlation matrix of the (latent) Gaussian process that defines the mixture representation of the MAL distribution (see Equation (9) in Petrella & Raponi 2019). Moreover, by simple calculations, it is possible to show that the covariance matrix of $\mathbf{Y}$ depends on $\boldsymbol{\Psi}$ through the following relationship: $\text{cov}(\mathbf{Y}) = \mathbf{D}(\tilde{\xi}\tilde{\xi}' + \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}})\mathbf{D}$. In other words, $\boldsymbol{\Psi}$ represents a shifted and scaled version of the sample correlation matrix of $\mathbf{Y}$ through the vector $\tilde{\xi}$ and the matrix $\mathbf{D}$, respectively.

In addition, when these constraints are satisfied, each marginal component of the MAL distribution in (2.3) follows a univariate AL distribution, that is, $Y_{tj} \sim \text{AL}(\mu_{tj}, \tau_j, \delta_{tj})$, where δ_{tj} represents the time-varying scale parameter of Y_{tj}. This allows us to exploit the result of Taylor (2019), who showed the link between the scale parameter of the AL distribution and the ES risk measure in a univariate framework. By extending these results, we provide new insights on how to estimate the conditional VaR and ES jointly in a multidimensional setting, which accounts for correlations between marginals. This is explained in detail in the next section.

2.2.2 Modeling VaR and ES jointly

Following Engle & Manganelli (2004), the CAViaR specification in (2.1) allows us to derive the VaR of an asset at level τ_j by estimating the corresponding quantile at the τ_j-th level, through a conditional autoregressive structure. In what follows, we consider several CAViaR formulations, depending on the choice of the NIC function $\ell(\cdot)$. We then extend the idea of Taylor (2019) to a multivariate setting in order to model and estimate the ES component dynamically.

The CAViaR specifications that we consider are the following:

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \omega_j + \eta_j \mathcal{Q}_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \beta_{1j}|Y_{t-1j}|, \qquad \textit{Symmetric Absolute Value (SAV)} \tag{2.4}$$

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \omega_j + \eta_j \mathcal{Q}_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \beta_{1j}Y^{+}_{t-1j} + \beta_{2j}Y^{-}_{t-1j}, \qquad \textit{Asymmetric Slope (AS)} \tag{2.5}$$

$$\mathcal{Q}_{Y_{tj}}(\tau_j|\mathcal{F}_{t-1}) = \left(\omega_j + \eta_j \mathcal{Q}^2_{Y_{t-1j}}(\tau_j|\mathcal{F}_{t-2}) + \beta_{2j}Y^2_{t-1j}\right)^{1/2} \qquad \textit{Indirect GARCH(1,1) (IG)} \tag{2.6}$$

where $\boldsymbol{\omega} = [\omega_1, ..., \omega_p]', \boldsymbol{\eta} = [\eta_1, ..., \eta_p]'$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_p]'$, with $\boldsymbol{\beta}_j = [\beta_{1j}, \beta_{2j}]'$, are unknown parameters to be estimated, and where $y^+ = \max(y, 0)$ and $y^- = -\min(y, 0)$, denote the positive and negative parts of y, respectively.

For the ES component, we exploit the interesting link provided in Bassett et al. (2004), which relates univariate quantile regression to conditional ES through the following relation:

$$ES_{tj} = \mathbb{E}[Y_{tj}] - \frac{\mathbb{E}[(Y_{tj} - \mathcal{Q}_{Y_{tj}})(\tau_j - \mathbf{1}_{(Y_{tj} < \mathcal{Q}_{Y_{tj}})})]}{\tau_j} \tag{2.7}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Following Taylor (2019), the expression in (2.7) can be rearranged so that the conditional ES can be expressed in terms of the conditional AL scale parameter δ_{tj}. Specifically, recalling that each marginal of the MAL distribution has a univariate AL density with a conditional scale parameter equal to $\delta_{tj} = \mathbb{E}[(Y_{tj} - \mathcal{Q}_{Y_{tj}})(\tau_j - \mathbf{1}_{(Y_{tj} < \mathcal{Q}_{Y_{tj}})})]$, (2.7) reduces to:

$$ES_{tj} = \mathbb{E}[Y_{tj}] - \frac{\delta_{tj}}{\tau_j}, \qquad t = 1, ..., T, \quad j = 1, ..., p \tag{2.8}$$

implying that

$$\delta_{tj} = \tau_j \left(\mathbb{E}[Y_{tj}] - ES_{tj}\right), \qquad t = 1, ..., T, \quad j = 1, ..., p \tag{2.9}$$

To ensure that each estimated ES does not cross the corresponding estimated quantile, we model the ES in (2.9) as the product of the quantile and a constant factor (see, e.g., Gourieroux et al. 2012 and Taylor 2019) as follows:

$$ES_{tj} = (1 + e^{\gamma_{0j}})\mathcal{Q}_{Y_{tj}}(\tau_j), \qquad t = 1, ..., T, \quad j = 1, ..., p, \tag{2.10}$$

where γ_{0j} is an unconstrained parameter to be estimated such that $1 + e^{\gamma_{0j}}$ is greater than 1 and $\mathbf{Y}_t$ is assumed to have zero mean. We collect the unknown parameters in the vector $\boldsymbol{\gamma}_0 = [\gamma_{01}, ..., \gamma_{0p}]'$. As explained in Taylor (2019), this formulation correctly describes the relationship between ES and VaR for different data generating processes, such as a GARCH process with a Student-t distribution. Therefore, the representation in (2.10) provides a simple and parsimonious approach to estimating VaR and ES simultaneously in a dynamic framework.

In (2.10), however, only the quantile is dynamic, while the factor $1 + e^{\gamma_{0j}}$ remains constant over time. Therefore, to generalize this approach, we also consider the alternative formulation for the ES presented in Taylor (2019), where the difference between the ES and the VaR is modeled using an AutoRegressive (AR) specification as follows:

$$ES_{tj} = \mathcal{Q}_{Y_{tj}}(\tau_j) - x_{tj}, \qquad t = 1, ..., T, \quad j = 1, ..., p, \tag{2.11}$$

$$x_{tj} = (\gamma_{1j} + \gamma_{2j}(\mathcal{Q}_{Y_{t-1j}}(\tau_j) - Y_{t-1j}) + \gamma_{3j}x_{t-1j})\mathbf{1}_{(Y_{tj} \leq \mathcal{Q}_{Y_{tj}})} + x_{t-1j}\mathbf{1}_{(Y_{tj} > \mathcal{Q}_{Y_{tj}})}, \tag{2.12}$$

where we define the nonnegative parameter $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_p]'$, with $\boldsymbol{\gamma}_j = [\gamma_{1j}, \gamma_{2j}, \gamma_{3j}]'$, to ensure that the VaR and ES estimates do not cross.
In the next section, we show how to estimate the model parameters using a ML approach based on a dynamic modification of the EM algorithm proposed by Petrella & Raponi (2019).

2.2.3 Parameter estimation using the EM algorithm

Before describing the main steps of the EM algorithm, we introduce the notation $\mathcal{Q}_t = \mathcal{Q}_{\mathbf{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$, $\mathbf{D}_t(\boldsymbol{\gamma})$ and $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})$ to clarify that the vector $\mathcal{Q}_{\mathbf{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$ and the matrices $\mathbf{D}_t$ and $\tilde{\boldsymbol{\Sigma}}$ depend on the unknown parameters $\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\Psi}$. The derivation of the EM algorithm is based on Proposition 3 of Petrella & Raponi (2019), properly extended to deal with the autoregressive structure of the quantile function $\mathcal{Q}_t$ and the time dependency of the scale matrix $\mathbf{D}_t(\boldsymbol{\gamma})$.

Let $\boldsymbol{\Phi} = \{\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Psi}\}$ denote the global set of parameters, and define $\hat{\boldsymbol{\Phi}} = \{\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Psi}}\}$ as the corresponding set of parameter estimates. For a given vector $\boldsymbol{\tau}$, the expected complete log-likelihood function (up to additive constants), given the observed data $\mathbf{Y}_t$ and the parameter estimates $\hat{\boldsymbol{\Phi}}$, is:

$$E\left[l_c(\boldsymbol{\Phi}|\mathbf{Y}_t, \hat{\boldsymbol{\Phi}})\right] = -\frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{D}_t(\boldsymbol{\gamma})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})\mathbf{D}_t(\boldsymbol{\gamma})| + \sum_{t=1}^{T}(\mathbf{Y}_t - \mathcal{Q}_t)'\mathbf{D}_t(\boldsymbol{\gamma})^{-1}\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})^{-1}\tilde{\xi} \tag{2.13}$$

$$-\frac{1}{2}\sum_{t=1}^{T} z_t(\mathbf{Y}_t-\mathcal{Q}_t)'(\mathbf{D}_t(\boldsymbol{\gamma})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})\mathbf{D}_t(\boldsymbol{\gamma}))^{-1}(\mathbf{Y}_t-\mathcal{Q}_t) \tag{2.14}$$

$$-\frac{1}{2}\tilde{\xi}'\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\Psi})^{-1}\tilde{\xi}\sum_{t=1}^{T} u_t, \tag{2.15}$$

where

$$u_t = E[\tilde{C}_t|\mathbf{Y}_t,\hat{\boldsymbol{\Phi}}] = \left(\frac{\hat{\tilde{m}}_t}{2+\hat{\tilde{d}}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_t}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_t}\right)} \tag{2.16}$$

$$z_t = E[\tilde{C}_t^{-1}|\mathbf{Y}_t,\hat{\boldsymbol{\Phi}}] = \left(\frac{2+\hat{\tilde{d}}}{\hat{\tilde{m}}_t}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_t}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\tilde{d}})\hat{\tilde{m}}_t}\right)} - \frac{2\nu}{\hat{\tilde{m}}_t}, \tag{2.17}$$

with

$$\hat{\tilde{m}}_t = (\mathbf{y}_t-\mathcal{Q}_t)'(\mathbf{D}_t(\hat{\boldsymbol{\gamma}})\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}})\mathbf{D}_t(\hat{\boldsymbol{\gamma}}))^{-1}(\mathbf{y}_t-\mathcal{Q}_t), \qquad \hat{\tilde{d}} = \tilde{\xi}'\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}})^{-1}\tilde{\xi}, \tag{2.18}$$

and where $\tilde{C}_t$ follows a standard exponential distribution.
For a given vector $\boldsymbol{\tau}$, the expected complete log-likelihood in (2.13)-(2.15) is then maximized with respect to the parameter set $\boldsymbol{\Phi}$, yielding the M-step updates $\hat{\boldsymbol{\Phi}}$. Notice that, unlike Petrella & Raponi (2019), closed-form solutions for $\hat{\boldsymbol{\omega}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ do not exist, due to the autoregressive structure of the data and, therefore, numerical optimization is required. Updated estimates of $\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}})$ can instead be derived using the following expression:

$$\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}}) = \frac{1}{T}\sum_{t=1}^{T} z_t\mathbf{D}_t(\hat{\boldsymbol{\gamma}})^{-1}(\mathbf{Y}_t-\mathcal{Q}_t)(\mathbf{Y}_t-\mathcal{Q}_t)'\mathbf{D}_t(\hat{\boldsymbol{\gamma}})^{-1} \tag{2.19}$$

$$+\frac{1}{T}\sum_{t=1}^{T} u_t\tilde{\xi}\tilde{\xi}' - \frac{2}{T}\sum_{t=1}^{T}\mathbf{D}_t(\hat{\boldsymbol{\gamma}})^{-1}(\mathbf{Y}_t-\mathcal{Q}_t)\tilde{\xi}'. \tag{2.20}$$

Therefore, the EM algorithm can be implemented as follows:
E-step: Set the iteration number $h=1$. Fix the vector $\boldsymbol{\tau}$ at the chosen quantile levels $\tau_1,...,\tau_p$ of interest, and initialize the parameter set $\boldsymbol{\Phi} = \{\boldsymbol{\omega},\boldsymbol{\eta},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\Psi}\}$. Then, given $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{\Phi}}^{(h)} = \{\hat{\boldsymbol{\omega}}^{(h)},\hat{\boldsymbol{\eta}}^{(h)},\hat{\boldsymbol{\beta}}^{(h)},\hat{\boldsymbol{\gamma}}^{(h)},\hat{\boldsymbol{\Psi}}^{(h)}\}$, at each iteration h, calculate the weights:

$$\hat{u}_t^{(h)} = \left(\frac{\hat{\tilde{m}}_t^{(h)}}{2+\hat{\tilde{d}}^{(h)}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}}\right)} \tag{2.21}$$

$$\hat{z}_t^{(h)} = \left(\frac{2+\hat{\tilde{d}}^{(h)}}{\hat{\tilde{m}}_t^{(h)}}\right)^{\frac{1}{2}} \frac{K_{\nu+1}\left(\sqrt{(2+\hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}}\right)}{K_{\nu}\left(\sqrt{(2+\hat{\tilde{d}}^{(h)})\hat{\tilde{m}}_t^{(h)}}\right)} - \frac{2\nu}{\hat{\tilde{m}}_t^{(h)}} \tag{2.22}$$

where

$$\hat{\tilde{m}}_t^{(h)} = (\mathbf{y}_t - \mathcal{Q}_t^{(h)})'(\mathbf{D}_t(\hat{\boldsymbol{\gamma}}^{(h)})\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}}^{(h)})\mathbf{D}_t(\hat{\boldsymbol{\gamma}}^{(h)}))^{-1}(\mathbf{y}_t - \mathcal{Q}_t^{(h)}), \tag{2.23}$$

$$\hat{\tilde{d}}^{(h)} = \tilde{\xi}'\tilde{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\Psi}}^{(h)})^{-1}\tilde{\xi}. \tag{2.24}$$

M-step: Use the estimates $\hat{u}_t^{(h)}$ and $\hat{z}_t^{(h)}$ to maximize $E[l_c(\boldsymbol{\Phi}|\hat{\boldsymbol{\Phi}}^{(h)})]$ with respect to $\boldsymbol{\Phi}$, and obtain the updated set of parameter estimates $\hat{\boldsymbol{\Phi}}^{(h+1)}$.

The optimization procedure is iterated until convergence, that is, when the difference between the likelihood function evaluated at two consecutive iterations is smaller than 10^{-5}. We initialize the EM algorithm by providing the univariate parameter estimates of Taylor (2019) for each asset, while the initial value for the correlation matrix $\boldsymbol{\Psi}$ in (2.3) is calibrated using the empirical correlation matrix of the data. We fit the univariate models following the estimation procedure in Engle & Manganelli (2004) and Taylor (2019). In addition, we consider a strategy of multiple random starts with 100 different starting points to better explore the parameter space, and we retain the solution corresponding to the maximum likelihood value. This strategy prevents convergence issues and prevents the algorithm from being trapped in local maxima. From an algorithmic point of view, the EM method exploits the Nelder-Mead and Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization routines to obtain the updated estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and it uses (2.19) to compute the updated estimate of $\tilde{\boldsymbol{\Sigma}}$. The computational analysis was conducted using the `R` (version 4.0.2) software, where the functions to update $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\Sigma}}$ were coded with efficient `C++` object-oriented programming.
The validity and performance of the proposed EM algorithm were also assessed using a simulation exercise (see Appendix B).

2.3 Portfolio construction

In this section, we approach the problem of portfolio allocation. Specifically, we construct the Skewness Mean-Variance (SMV) portfolio of Zhao et al. (2015), taking into account both the multivariate structure and the skewness of asset returns. Following Stolfi et al. (2018) and Zhao et al. (2015), we exploit an interesting property characterizing the MAL distribution in (2.3). We show that any linear combination of its marginal components follows a univariate AL distribution whose parameters are a function of the MAL parameters in (2.3). Note that while the MAL density has thus far been regarded as a convenient tool for estimating the marginal quantiles, in this section, the MAL distribution is used as a data-driven assumption to describe the empirical characteristics of asset returns. As already stated in the Introduction, this choice has been positively accepted in the recent financial literature to detect the peakedness, fat-tails, and skewness of financial assets, overcoming the possible deficiencies of standard approaches relying, for example, on the Gaussian distribution assumption. We then evaluate the riskiness of the selected portfolio by calculating its corresponding VaR and ES using the results of Section 2.2.2.

2.3.1 Linear combinations of MAL components

Let us assume that $\mathbf{Y}_t$ is a p-dimensional random variable describing the joint dynamics of p variables at time t. Let us consider a linear combination (with weights to be determined) of each component of $\mathbf{Y}_t$. Then, the following proposition holds.

Proposition 4. *Let* $\mathbf{Y}_t \sim MAL_p(\boldsymbol{\mu}_t, \mathbf{D}_t\tilde{\xi}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)$*, with density function defined in (2.3). Let* $\mathbf{b}_t = (b_{t1}, ..., b_{tp})' \in \mathcal{R}^p$ *be a vector of weights such that* $\mathbf{b}_t \neq \mathbf{0}_p$*, with* $\mathbf{0}_p$ *denoting a p-vector of zeros. Define the random variable* $Y_t^{\mathbf{b}} = \sum_{j=1}^{p} b_{tj} Y_{tj}$*. Then,*

$$Y_t^{\mathbf{b}} \sim AL(\mu_t^\star, \tau_t^\star, \delta_t^\star) \tag{2.25}$$

where

$$\begin{aligned}
\mu_t^\star &= \mathbf{b}_t'\boldsymbol{\mu}_t, \\
\tau_t^\star &= \frac{1}{2}\left(1 - \frac{\mathbf{b_t'D_t}\tilde{\xi}}{\sqrt{2(\mathbf{b_t'D_t}\tilde{\boldsymbol{\Sigma}}\mathbf{D_tb_t}) + (\mathbf{b_t'D_t}\tilde{\xi})^2}}\right), \\
\delta_t^\star &= \frac{(\mathbf{b_t'D_t}\tilde{\boldsymbol{\Sigma}}\mathbf{D_tb_t})}{2\sqrt{2(\mathbf{b_t'D_t}\tilde{\boldsymbol{\Sigma}}\mathbf{D_tb_t}) + (\mathbf{b_t'D_t}\tilde{\xi})^2}}.
\end{aligned} \tag{2.26}$$

Proposition 4 brings out two main considerations. First, the distribution of $Y_t^{\mathbf{b}}$ is still AL, which greatly facilitates the computation of the VaR and ES in our context. Second, the parameters of $Y_t^{\mathbf{b}}$ are expressed as a function of the multivariate parameters $\boldsymbol{\mu}_t, \mathbf{D}_t$ and $\tilde{\boldsymbol{\Sigma}}$ of the MAL distribution in (2.3). This allows us to take into account the possible association among the marginal components of $\mathbf{Y}_t$ when choosing the allocation weights $\mathbf{b}_t$. In the next section, we exploit this property to retrieve the distribution of returns of a financial portfolio, whose optimal weights can be derived by solving a simple constrained optimization problem. Given the resulting optimal portfolio weights, we can then use the results of Section 2.2.2 to derive appropriate measures of the portfolio's VaR and ES.

2.3.2 The portfolio optimization problem

Assume that $\mathbf{Y}_t$ follows the distribution in (2.3). Given this specification, at each time t, investors may be interested in deriving a portfolio $Y_t^b = \sum_{j=1}^{p} b_{tj} Y_{tj}$ by investing a portion b_{tj} of their capital on the asset Y_{tj} so that $\sum_{j=1}^{p} b_{tj} = 1$. Then, in this setting, the result of Proposition 4 can be applied easily, yielding a portfolio with location, skewness and scale parameters equal to, respectively, $\mu_t^\star$, $\tau_t^\star$ and $\delta_t^\star$ as in (2.26).

Typically, in risk management applications, the skewness parameter is fixed a priori by the researcher at a certain level (constant over time) $\tau_t^\star = \tilde{\tau}$, as it essentially measures the overall riskiness of a financial product (a portfolio, in our case). Therefore, once we estimate the time-varying MAL parameters from the quantile regression model in (2.2), for a fixed level of risk $\tilde{\tau}$, the investor's portfolio decision is based on the solution of the selected portfolio strategy. As stated above, to obtain the optimal portfolio allocation, we adopt the SMV strategy of Zhao et al. (2015), which seeks to minimize the portfolio variance and at the same time control

for the skewness of the asset returns. This approach can be seen as an extension of the classical MV approach of Markowitz 1952 where typical empirical features of financial data are incorporated into the optimization problem. According to Proposition 4, the objective function of the considered optimization problem, i.e., the portfolio variance, is given by expression $\mathbf{b}_t'\mathbf{D}_t\tilde{\mathbf{\Sigma}}\mathbf{D}_t\mathbf{b}_t$. Then, for a given level of risk $\tilde{\tau}$, the SMV portfolio solves the following constrained optimization problem:

$$\begin{aligned} \underset{\mathbf{b}_t \in \mathcal{R}^p}{\arg\min} \quad & \mathbf{b}_t'\mathbf{D}_t\tilde{\mathbf{\Sigma}}\mathbf{D}_t\mathbf{b}_t && \text{(2.27a)} \\ \text{s.t.} \quad & \tau_t^\star = \tilde{\tau}, \quad \forall t && \text{(2.27b)} \\ & \mathbf{b}_t'\mathbf{1}_p = 1 && \text{(2.27c)} \end{aligned}$$

where $\tilde{\mathbf{\Sigma}}$ was introduced in (2.3) and accounts for the covariance matrix of the returns, while $\mathbf{b}_t$ denotes the portfolio's weights at time t held by the investor over the period $[t, t+1)$.

From an empirical point of view, the constraint in (2.27b) implies that the portfolio weights must be adjusted at each holding period to guarantee that the VaR of the portfolio has a constant level $\tilde{\tau}$, namely, $\mathbb{P}(Y_t^{\mathbf{b}} < \mu_t^\star | \mathcal{F}_{t-1}) = \tilde{\tau}$. Once we obtain the optimal portfolio weights for the period $[t, t+1)$, we can compute the conditional portfolio's VaR and ES at level $\tilde{\tau}$ by simply applying the result in (2.8) to the univariate case. In addition, from a computational standpoint, optimizing (2.27a) is computationally advantageous because $\mathbf{b}_t'\mathbf{D}_t\tilde{\mathbf{\Sigma}}\mathbf{D}_t\mathbf{b}_t$ is a quadratic objective function which can be solved more efficiently than using directly the scale parameter of the portfolio distribution $\delta_t^\star$.
As explained above, since the parameters $\mu_t^\star, \tau_t^\star$ and $\delta_t^\star$ depend on the parameter estimates of the MAL distribution, information on the dependence structure and on the empirical characteristics embedded in the data is channeled through these estimates into the portfolio's VaR and ES forecasts. This motivates our approach even further, since it can offer an operative and useful tool to help investors and asset managers in deriving optimal portfolio allocations and, at the same time, monitoring multiple VaR and ES jointly.

2.4 Assessment of VaR and ES forecasts

To assess the performance of VaR and ES predictions jointly, we introduce a new backtesting procedure, based on the multivariate approach discussed in Section 2.2.

Backtesting techniques are based on quantitative tests that scrutinize model performance in terms of accuracy and precision with respect to a defined criterion. Existing approaches, however, rely on tests that analyze VaR and ES predictions separately; i.e., they focus only on the individual evaluation of one risk measure or the other. VaR evaluation is typically based on coverage tests, which measure the percentage of times that the returns exceed the estimated VaR at a chosen probability level τ (see, e.g., the unconditional coverage (LR_{uc}) test of Kupiec 1995, the conditional coverage (LR_{cc}) test of Christoffersen 1998 and the Dynamic Quantile (DQ) test of Engle & Manganelli 2004).

To evaluate ES forecasts, the backtesting analysis becomes more complicated since ES is not an elicitable measure (Gneiting 2011) and therefore suitable scoring

functions cannot be determined (Taylor 2019). The test of McNeil & Frey (2000) is commonly used in this context, which is based on the discrepancy between the observed return and the ES forecast for the periods in which the return exceeds the VaR forecast. Another suitable option is the backtesting procedure of Du & Escanciano (2017), which is based on the Unconditional ES (U_{ES}) and Conditional ES (C_{ES}) tests.

However, since ES relies on observations exceeding the VaR, it is clear that assessment of ES forecasts cannot be independent of the predicted VaR values. This, together with the fact that ES is not elicitable, motivates the introduction of a scoring function for jointly evaluating VaR and ES forecasts. Based on the characterization of consistent scoring functions introduced by Fissler & Ziegel (2016) and Nolde et al. (2017), several scoring rules have been proposed in the literature for the univariate setting (see, e.g., Patton et al. 2019, Fissler et al. 2015 and Taylor 2019).

In what follows, we provide a new scoring rule that can be used in a multivariate setting to jointly evaluate VaR and ES forecasts of multiple (and possibly correlated) financial assets. To provide support for our proposal of estimating multiple VaR and ES by maximizing the MAL likelihood, we define a new scoring function (S_{MAL}) using the negative of the MAL log score:

$$
\begin{aligned}
S_{MAL}\left(\mathcal{Q}_t, \mathbf{ES}_t, \mathbf{y}_t; \tilde{\boldsymbol{\Sigma}}, \boldsymbol{\tau}\right) = & \\
& \frac{1}{2}\log\left(|\tilde{\boldsymbol{\Sigma}}|\right) + \log\left(|(\boldsymbol{\tau}\mathbf{ES}_t') \circ \mathbf{I}_p|\right) - \frac{\nu}{2}\log\left(\frac{\tilde{m}_t}{2+\tilde{d}}\right) \\
& + (\mathbf{y}_t - \mathcal{Q}_t)'\left((\boldsymbol{\tau}\mathbf{ES}_t') \circ \mathbf{I}_p\right)^{-1}\tilde{\boldsymbol{\Sigma}}^{-1}\tilde{\xi} - \log\left(K_\nu\left(\sqrt{(2+\tilde{d})\tilde{m}_t}\right)\right)
\end{aligned}
\tag{2.28}
$$

where $\circ$ denotes the Hadamard product and $\mathbf{I}_p$ represents the identity matrix of order p.

Notice that when p is equal to 1, the S_{MAL} in (2.28) reduces to the AL log score of Taylor (2019). When $p > 1$, the loss function S_{MAL} allows us to (i) perform a joint assessment of the pairs (VaR, ES) specific to each asset and, at the same time, (ii) control for the existing correlation among returns.

2.5 Empirical study

In this section, we apply the methodology presented in Sections 2.2 and 2.3 to real data in order to evaluate and compare the empirical implications with those obtained by using a univariate framework. Specifically, we follow Taylor (2019) and use the weekly returns of the FTSE 100, NIKKEI 225, and S&P 500 stock market indices from April 26, 1985, to February 01, 2021. Using a rolling window exercise, we estimate the one-week-ahead VaR and ES forecasts implied by the CAViaR specifications described in Section 2.2.2, and we select the most desirable model using the Diebold & Mariano (2002) test. In a second empirical exercise, we aggregate the market indices to form a financial portfolio and determine its optimal allocation weights by solving the optimization problem described in Section 2.3.2. We finally compute and assess the resulting portfolio's conditional VaR and ES for the out-of-sample period, which consists of the last 368 observations of the sample.

2.5.1 Data description

Our sample is collected from Bloomberg, and it consists of 1868 weekly returns for each of the three stock indices. The main summary statistics are displayed in Table 2.1 below, providing evidence of the well-known stylized facts on fat tails, high kurtosis and serial and cross-sectional correlation that typically characterize financial assets. Moreover, all series exhibit a negative skewness, the Jarque-Bera test significantly rejects the normality assumption, the Ljung-Box test indicates the presence of serial correlation and the Augmented Dickey-Fuller test supports the hypothesis of the absence of unit roots. These results clearly motivate us to consider a quantile regression approach as an investigative tool.

Index	Mean	Median	SD	Skewness	Kurtosis	J-B	L-B	ADF
FTSE 100	0.086	0.234	2.396	-1.456	14.717	**17517.036**	**62.674**	**-19.729**
NIKKEI 225	0.046	0.237	2.946	-0.748	6.421	**3383.250**	**175.024**	**-19.159**
S&P 500	0.163	0.320	2.338	-0.947	7.367	**4503.301**	**403.851**	**-20.325**
Correlation matrix								
	FTSE 100	NIKKEI 225	S&P 500					
FTSE 100	1							
NIKKEI 225	0.510	1						
S&P 500	0.709	0.501	1					

Table 2.1. Summary statistics of the weekly returns of the three indices for the entire sample from April 26, 1985, to February 01, 2021. The test statistics are displayed in boldface when the null hypothesis is rejected at the 1% significance level. J-B, L-B and ADF denote the Jarque-Bera test, the Ljung-Box test on squared returns with 4 lags and the Augmented Dickey-Fuller unit root test with 4 lags, respectively.

2.5.2 Out-of-sample VaR and ES forecasting

Using the approach introduced in Section 2.2, in this section, we derive a joint estimation of VaR and ES for the three stock market indices described above. Specifically, we estimate the out-of-sample series of VaR and ES by considering the three different specifications in (2.4), (2.5) and (2.6), with both the multiplicative factor in (2.10) and the AR formulation in (2.11)-(2.12) for the ES component. Moreover, since we are concerned with the downside risk, we evaluate the out-of-sample forecasts at three different probability levels, namely, $\boldsymbol{\tau} = [0.1, 0.1, 0.1], \boldsymbol{\tau} = [0.05, 0.05, 0.05]$ and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$.

The first objective is to assess the performance of the CAViaR specifications using the proposed multivariate framework. We start by evaluating the VaR forecasts using the conventional LR_{uc}, LR_{cc} and DQ tests, while we perform the U_{ES} and C_{ES} tests of Du & Escanciano (2017) to evaluate the ES predictions. The results are shown in Table 2.2, where Panel A refers to the case of the ES vector modeled as in (2.10) and Panel B refers to the AR specification in (2.11)-(2.12). Looking at the VaR forecasts, in both panels, for all three indices and for all three quantile levels, we find that the CAViaR-AS specification is always successfully backtested at the 5% significance level. The results are less clear for the other two CAViaR specifications. The same results are confirmed when evaluating the ES predictions, as the CAViaR-AS specification again yields outstanding performances for all three

indices and for all three quantile levels.
To jointly evaluate the VaR and ES forecasts associated with each stock market index, in addition to the results of the coverage tests, Table 2.3 reports the values of the loss functions S_{FZN} of Nolde et al. (2017) and S_{FZ0} of Patton et al. (2019) averaged over the out-of-sample period, where:

$$\begin{aligned} S_{FZN}(\mathcal{Q}_t, ES_t, y_t) = (\mathbf{1}_{(y_t<\mathcal{Q}_t)} - \tau)\frac{\mathcal{Q}_t}{2\tau\sqrt{-ES_t}} \\ - \frac{1}{2\sqrt{-ES_t}}(\mathbf{1}_{(y_t<\mathcal{Q}_t)}\frac{y_t}{\tau} - ES_t) + \sqrt{-ES_t} \end{aligned} \tag{2.29}$$

and

$$S_{FZ0}(\mathcal{Q}_t, ES_t, y_t) = \frac{1}{\tau ES_t}\mathbf{1}_{(y_t<\mathcal{Q}_t)}(y_t - \mathcal{Q}_t) + \frac{\mathcal{Q}_t}{ES_t} + \log(-ES_t) - 1. \tag{2.30}$$

The losses in (2.29) and (2.30) belong to the class of scoring rules proposed in Nolde et al. (2017) and Patton et al. (2019) and have the additional advantage of generating loss differences (between competing forecasts) that are homogeneous of degree 1/2 and zero, respectively.
Overall, the results show that both the CAViaR-AS and CAViaR-IG dynamics are associated with smaller losses compared to the CAViaR-SAV model, except for the case of $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$. This finding suggests that there are relevant asymmetries and leptokurtosis in the behavior of the return series of the three indices, which must be taken into due account to yield better out-of-sample forecasts. Moreover, in line with Taylor (2019), we find evidence of a better forecasting performance when using the constant multiplicative factor $(1 + e^{\gamma_{0j}})$ to model the ES parameter (Panel A) compared to the AR dynamics (Panel B).
Finally, to reinforce our analysis, we evaluate the forecasting performance of the three competing CAViaR models using the scoring function in (2.28). Specifically, at each time t, and for the specified level $\boldsymbol{\tau}$, we define by $S^{(j)}_{MAL_t}(\boldsymbol{\tau})$ the scoring function associated with model j, and we denote the difference between the scoring functions of model i and model j by $\Delta^{(i,j)}_{MAL,t} = S^{(i)}_{MAL_t}(\boldsymbol{\tau}) - S^{(j)}_{MAL_t}(\boldsymbol{\tau})$, where $i, j = 1, 2, 3$. We then test for the null hypothesis that $\mathbb{E}[\Delta^{(i,j)}_{MAL,t}] = 0$ against $\mathbb{E}[\Delta^{(i,j)}_{MAL,t}] < 0$ using the Diebold & Mariano (2002) test for all the pairs of models i and j. If the null hypothesis is rejected, then the forecasts delivered by model i are more accurate than those of model j, and therefore model i is preferable to model j. The results of the test, together with the corresponding p-values, are reported in Table 2.4. The table clearly shows that the CAViaR-AS specification outperforms both the CAViaR-IG and CAViaR-SAV models at all the three quantile levels and for both the adopted ES formulations of constant multiplicative factor (Panel A) and AR dynamics (Panel B). Therefore, to select the best performing model between the two CAViaR-AS in Panels A and B, we again apply the Diebold & Mariano (2002) test to the two competing CAViaR-AS specifications. The results are reported in Table 2.5, and they suggest that the CAViaR-AS model with the ES specified as a constant multiple of the VaR provides the most accurate predictions at all three quantile levels. This is in line with Taylor (2019), who also found that the same specification not only produces the smallest losses but also delivers the most accurate predictions compared

with all the other competing CAViaR dynamics[2]. These results corroborate the fact that accounting for asymmetries in the autoregressive process of a given quantile improves the model's forecasting ability (see, e.g., Engle & Manganelli 2004, Xiliang & Xi 2009, Taylor 2005 and Laporta et al. 2018).

To show the advantages and the different implications of our approach, we compare our results with those obtained by considering each asset separately, as if we ignored their possible dependence structure. Specifically, the three CAViaR models are estimated individually for each stock market index using the univariate approach of Taylor (2019). To assess the performance of the three models and to combine the individual forecasts of the three indices in a single value, we use the sum of the three corresponding AL log scores (see Taylor 2019) as a consistent scoring rule. That is, at each time t, and for each model j, we define the following scoring function:

$$S^{(j)}_{AL_t}(\boldsymbol{\tau}) = \sum_{p=1}^{3} S^{(j)}_{AL_{p,t}}(\tau_p) \tag{2.31}$$

where $S^{(j)}_{AL_{p,t}}(\tau_p)$ denotes the AL log-score of Taylor (2019), corresponding to model j and asset p:

$$S^{(j)}_{AL_{p,t}}(\tau_p) = -\log\left(\frac{\tau_p - 1}{ES^{(j)}_{p,t}}\right) - \frac{\left(y_{p,t} - \mathcal{Q}^{(j)}_{p,t}\right)\left(\tau_p - \mathbf{1}_{(y_{p,t} < \mathcal{Q}^{(j)}_{p,t})}\right)}{\tau_p ES^{(j)}_{p,t}}. \tag{2.32}$$

As explained in Frongillo & Kash (2015), summing the three AL scoring functions would produce a consistent scoring rule in this case, since each function $S^{(j)}_{AL_{p,t}}(\tau_p)$ elicits the pair (VaR, ES) for the corresponding p-th asset (see Fissler & Ziegel 2016 and Taylor 2019).
Then, as before, we define the difference between the scoring functions of model i and model j by $\Delta^{(i,j)}_{AL,t} = S^{(i)}_{AL_t}(\boldsymbol{\tau}) - S^{(j)}_{AL_t}(\boldsymbol{\tau})$ and apply the Diebold & Mariano (2002) test to look for the best model in terms of forecasting accuracy. The results are reported in Table 2.6 (Panels A and B). As shown in the table, the conclusion of the test is now less clear and does not provide any significant evidence in favor of a particular model. This is one of the primary advantages of our approach, as we are able to identify a clear hierarchy among competing models.

A second question of interest concerns the "efficiency gain" of the multiple approach compared to the univariate one. In this sense, we would like to test whether taking into account the association structure among the market indices would provide us with better predictions in terms of VaR and ES. To do this, we use the backtesting procedure to identify the most "efficient" model, that is, the model producing the best forecasts according to the Diebold & Mariano (2002) test.

[2]To further justify this choice, we also compare the CAViaR-AS model with the Quantile AutoRegression of Koenker & Xiao (2006). Specifically, we estimate the regression model of Petrella & Raponi (2019) using the lagged returns (at lag 1) of each asset as the covariates. Comparing these two models would allow one to evaluate the potential contribution of assuming a CAViaR specification in the quantile dynamics. According to the coverage tests and the scoring functions defined in (2.29) and (2.30), we still find that the performance of the CAViaR-AS specifications is better.

To measure the efficiency gain, we analyze the difference, if any, in the predictive accuracy between the forecast (VaR, ES) produced by our multivariate approach and the univariate ones. Therefore, for a given CAViaR specification, we test for the difference between the scores obtained with the scoring rule in (2.31) and those obtained with (2.28). The null hypothesis is that, on average, the difference is not statistically different from zero, i.e., that the two approaches have the same forecasting performance. The alternative hypothesis is that the difference is smaller than zero, i.e., that the multivariate approach delivers significantly better predictions (smaller losses).

Table 2.7 shows the resulting test statistics and the corresponding p-values for each of the possible pairs of competing models. Interestingly, for all the considered risk levels and for all three CAViaR specifications, we are always able to reject the null hypothesis at the 1% level, except for the comparison between the CAViaR-IG and CAViaR-AS specifications at $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ for which the null is rejected at the 5% level, providing evidence of the efficiency gain of our proposed joint approach[3].

To offer a graphical intuition of supporting the results, Figure 2.1 shows the time series of the difference between the scoring function in (2.31) that is consistent with the univariate approach and the scoring function proposed in (2.28) over the whole out-of-sample period and for the three considered CAViaR specifications, with the ES modeled as a multiple of the VaR. The left graph in Figure 2.1 refers to the case of $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, the center graph displays the case of $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$, and the right plot displays $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$. The black line represents the difference of the two scoring functions obtained by using the CAViaR-SAV specification in (2.4), while the red and blue lines refer to the CAViaR-AS and CAViaR-IG dynamics, respectively. The efficiency gain of the multivariate approach clearly emerges from these pictures. Indeed, for all the considered risk levels, and regardless of the dynamic specification of the quantiles, the difference between the two approaches is almost always positive. This confirms the idea that the losses associated with the univariate model can be very large if the dependence structure of the data is not accounted for.

In Figure 2.2, we display the series of out-of-sample forecasts of the VaR and ES for each of the three stock indices, which are estimated by assuming the selected CAViaR-AS specification in both the univariate and joint approaches. The VaR predictions obtained with the univariate approach of Taylor (2019) are represented by a dotted blue line, while the VaR estimates produced by our joint approach are depicted by the solid red line. The estimated ES is represented by the dotted green line (using the univariate approach) and the solid orange line (using the joint approach). The left panels of Figure 2.2 refer to $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, the center panels refer to $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ and the right ones refer to $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$. The gray dots denote the original return series of each stock index. In all the cases, the estimates of VaR and ES produced by our joint approach lie below the corresponding values obtained with the univariate setting, suggesting that our proposed method can lead to more conservative results.

[3] As a further robustness check, we also considered the best CAViaR specification for each asset and then applied the Diebold & Mariano (2002) test. The proposed joint method still appears to be more efficient than the univariate method of Taylor (2019), so our findings are unchanged.

Finally, to obtain a more intuitive representation of the relationship between the estimated VaR and ES over time and across quantile levels, Figure 2.3 displays the absolute difference between the out-of-sample VaR and ES forecasts for each of the three stock indices. The plots in the first row are obtained by assuming the CAViaR-AS specification with the ES modeled as in (2.10), while the plots in the second row refer to the case of the CAViaR-AS specification with the ES following the dynamic in (2.11)-(2.12) at the $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$ (left column), $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ (center column) and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ (right column) quantile levels. The blue, red and orange lines refer to the FTSE 100, NIKKEI 225 and S&P 500 stock market indices, respectively, while the gray bands correspond to the main recession periods and to various economic and financial crises that have occurred since 2014. As one can reasonably expect, the difference follows the overall market volatility. Indeed, we find that the difference between the estimated risk measures is typically smaller in calm periods and larger in periods of turbulent markets, with more pronounced upward spikes when the AR dynamics for the ES are used. High volatility is also clearly evident in correspondence and in the aftermath of major economic and financial crises, such as, for example, the Chinese stock market crash at the start of 2016, the Brexit in 2018 and the outbreak of the COVID-19 pandemic in 2020.

Based on these considerations, in the next section, we consider the CAViaR-AS specification in (2.5) with the ES expressed in (2.10) to implement the portfolio optimization procedure.

2.5.3 Out-of-sample portfolio VaR and ES forecasting

In this section, we use the three stock market indices FTSE 100, NIKKEI 225 and S&P 500 to build a SMV portfolio that delivers a certain fixed level of risk $\tilde{\tau}$. The optimal allocation weights are determined by solving the optimization problem described in Section 2.3.2 using the parameter estimates provided by the CAViaR-AS specification in Section 2.5.2.

We evaluate the benefits of our approach by considering alternative strategies. First, we use the estimation method of Zhao et al. (2015), where the covariance matrix $\tilde{\boldsymbol{\Sigma}}$ in (2.27a) is estimated using the sample variance and the sample mean of the return series. We call this strategy Moment-SMV. Second, we evaluate the classic MV of Markowitz (1952). In this case, we model the conditional covariance of the asset returns using several well-known autoregressive dynamics, i.e., the multivariate GARCH Dynamic Conditional Correlation model (Engle 2002) under both the multivariate Normal (MV-G-DCC-N) and Student-t (MV-G-DCC-t) error distributions and the asymmetric Dynamic Conditional Correlation model with multivariate Normal (MV-G-aDCC-N) and Student-t (MV-G-aDCC-t) errors. Moreover, since the MV strategy can often be inadequate in controlling for asymmetric risk-averse agents, we also consider the above strategies under the multivariate skew Normal (SN) and multivariate skew Student-t (St) distributions of Bauwens & Laurent (2005) as further competing strategies, which we denote by MV-G-DCC-SN, MV-G-DCC-St, MV-G-aDCC-SN and MV-G-aDCC-St, respectively. Then, for each model, we forecast the one-week-ahead conditional covariance matrix and plug it into the portfolio optimization problem.

We jointly estimate the VaR and ES of the resulting portfolios and analyze their

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$					$[0.05, 0.05, 0.05]$					$[0.01, 0.01, 0.01]$				
	LR_{uc}	LR_{cc}	DQ	U_{ES}	C_{ES}	LR_{uc}	LR_{cc}	DQ	U_{ES}	C_{ES}	LR_{uc}	LR_{cc}	DQ	U_{ES}	C_{ES}
Panel A: Multiplicative factor for the ES															
SAV															
FTSE 100	**0.043**	**0.455**	**7.194**	**1.467**	**7.735**	**2.658**	**3.469**	**8.288**	**0.128**	**6.711**	**0.756**	**1.111**	**6.350**	5.063	9.991
NIKKEI 225	5.571	**5.838**	10.463	**0.022**	**6.553**	**1.851**	**2.806**	**7.047**	**0.543**	11.086	**2.857**	**4.510**	11.047	**−1.280**	**0.913**
S&P 500	**0.450**	**0.804**	**9.431**	**1.604**	12.776	**1.203**	**3.890**	**8.917**	**0.931**	12.748	4.890	**5.815**	30.388	**−1.229**	9.568
AS															
FTSE 100	**1.067**	**1.210**	**6.641**	**0.663**	**8.773**	**1.203**	**2.314**	**6.266**	**0.434**	**5.243**	**1.096**	**4.184**	**3.351**	**−1.562**	**3.428**
NIKKEI 225	**3.571**	**4.932**	**1.064**	**0.261**	**8.043**	**2.658**	**3.469**	**8.094**	**0.579**	**9.193**	**1.747**	**2.019**	**8.332**	**−1.490**	**2.241**
S&P 500	**0.724**	**0.961**	**7.585**	**1.428**	**8.660**	**0.704**	**2.942**	**7.101**	**1.091**	**8.038**	**0.414**	**2.022**	**6.519**	**1.439**	**0.013**
IG															
FTSE 100	**0.450**	**0.456**	**5.705**	**1.011**	9.926	**2.658**	**3.469**	**8.381**	**0.263**	**6.421**	4.261	**4.399**	9.853	**1.670**	**6.078**
NIKKEI 225	6.547	6.717	11.814	**−0.295**	**8.787**	**1.851**	**2.806**	**6.954**	**0.597**	**9.478**	**1.747**	**2.019**	**5.738**	**−1.195**	**6.915**
S&P 500	**0.724**	**0.961**	10.492	**1.298**	12.164	**1.851**	**5.047**	11.778	**0.579**	**7.730**	**0.952**	**1.309**	**6.009**	**1.699**	**3.150**
Panel B: AR formulation for the ES															
SAV															
FTSE 100	4.589	**4.714**	9.787	**0.184**	**4.671**	**2.587**	**3.490**	11.336	**−1.699**	**2.256**	**1.055**	**2.922**	**9.309**	−3.699	**2.256**
NIKKEI 225	5.329	**5.431**	10.613	**−1.570**	**7.718**	**1.547**	**2.297**	**7.295**	**−1.330**	**2.575**	**2.070**	**4.164**	12.106	4.330	11.575
S&P 500	6.840	8.057	12.670	**0.152**	12.066	**1.476**	**3.516**	**2.591**	1.969	**0.315**	4.020	**5.144**	15.297	5.869	**0.315**
AS															
FTSE 100	**3.441**	**3.337**	**5.771**	**−1.594**	**1.128**	**2.249**	**5.271**	**3.573**	**1.470**	**7.960**	**3.182**	**2.913**	**3.248**	**−1.470**	**7.960**
NIKKEI 225	**3.651**	**5.168**	**6.650**	**−1.646**	**1.631**	**2.251**	**4.451**	**7.473**	**−1.350**	**2.542**	**0.177**	**4.879**	**3.313**	**1.350**	**2.542**
S&P 500	**2.250**	**3.371**	**1.745**	**0.523**	**0.008**	**1.535**	**3.321**	**6.457**	**1.103**	**6.304**	**2.922**	**3.042**	**6.036**	**−1.103**	**6.304**
IG															
FTSE 100	5.214	**1.282**	**9.456**	**0.251**	**7.350**	**2.284**	**3.146**	**8.951**	**1.553**	**5.164**	**3.036**	**3.935**	**9.082**	**1.840**	**4.991**
NIKKEI 225	12.095	6.845	**6.904**	**1.285**	9.997	**1.228**	**3.180**	**6.852**	**1.151**	**4.656**	**2.135**	**3.106**	**6.277**	**1.704**	**2.610**
S&P 500	9.859	**1.564**	**6.456**	**1.732**	**4.639**	**1.731**	**4.584**	10.327	**−1.703**	**1.925**	**1.041**	**5.188**	15.358	**−1.898**	**0.013**

Table 2.2. Marginal out-of-sample VaR and ES forecast evaluation using the joint approach with the multiplicative factor in (2.10) (Panel A) and the AR formulation in (2.11)-(2.12) (Panel B) for the ES. At the 5% significance level, the critical values of LR_{uc} and LR_{cc} are 3.84 and 5.99, respectively. The U_{ES} is rejected if the test statistic is greater (in absolute value) than 1.96. Finally, the DQ test uses lagged violations at lag 4 while the C_{ES} test considers the first 4 lagged autocorrelations, and the critical value for both is 9.49. The test statistics are displayed in boldface when the null hypotheses are not rejected at the 5% significance level.

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$		$[0.05, 0.05, 0.05]$		$[0.01, 0.01, 0.01]$	
	S_{FZ0}	S_{FZN}	S_{FZ0}	S_{FZN}	S_{FZ0}	S_{FZN}
Panel A: Multiplicative factor for the ES						
SAV						
FTSE 100	1.398	2.661	1.795	3.221	2.109	3.399
NIKKEI 225	1.545	3.091	2.051	3.666	2.231	3.678
S&P 500	1.368	2.552	1.753	3.126	1.907	3.215
AS						
FTSE 100	1.371	2.578	1.750	2.833	1.954	3.254
NIKKEI 225	1.585	2.890	1.949	3.155	2.111	3.636
S&P 500	1.346	2.431	1.706	2.690	1.867	3.142
IG						
FTSE 100	1.383	2.636	1.765	3.000	1.946	3.320
NIKKEI 225	1.539	3.025	2.023	3.387	2.196	3.689
S&P 500	1.357	2.535	1.715	2.943	1.897	3.257
Panel B: AR formulation for the ES						
SAV						
FTSE 100	1.414	2.878	2.116	2.362	4.866	5.740
NIKKEI 225	1.766	2.082	2.241	3.055	5.220	4.779
S&P 500	1.406	2.767	1.940	3.177	4.929	5.241
AS						
FTSE 100	1.496	1.845	1.985	2.325	6.998	5.009
NIKKEI 225	1.714	2.059	2.315	3.731	7.233	4.120
S&P 500	1.478	2.756	2.048	2.944	6.929	4.934
IG						
FTSE 100	1.411	1.849	2.112	2.800	3.201	5.025
NIKKEI 225	1.671	1.993	2.305	3.076	3.432	5.629
S&P 500	1.370	2.764	2.031	3.007	3.164	4.983

Table 2.3. Marginal out-of-sample VaR and ES forecast evaluation based on the average losses using the scoring functions in (2.29) and (2.30) for the joint approach with the multiplicative factor in (2.10) (Panel A) and the AR formulation in (2.11)-(2.12) (Panel B) for the ES.

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$			$[0.05, 0.05, 0.05]$			$[0.01, 0.01, 0.01]$		
	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG
Panel A: Multiplicative factor for the ES									
CAViaR-SAV	-	3.169	2.020	-	4.090	4.630	-	3.832	3.777
	-	(0.999)	(0.978)	-	(1.000)	(1.000)	-	(1.000)	(1.000)
CAViaR-AS	-3.169	-	-2.575	-4.090	-	-2.394	-3.832	-	-1.931
	(0.001)	-	(0.005)	(0.000)	-	(0.009)	(0.000)	-	(0.027)
CAViaR-IG	-2.020	2.575	-	-4.630	2.394	-	-3.777	1.931	-
	(0.022)	(0.995)	-	(0.000)	(0.991)	-	(0.000)	(0.973)	-
Panel B: AR formulation for the ES									
CAViaR-SAV	-	2.635	1.919	-	3.688	4.114	-	3.927	4.728
	-	(0.996)	(0.972)	-	(1.000)	(1.000)	-	(1.000)	(1.000)
CAViaR-AS	-2.635	-	-2.320	-3.688	-	6.090	-3.927	-	6.273
	(0.004)	-	(0.010)	(0.000)	-	(1.000)	(0.000)	-	(1.000)
CAViaR-IG	-1.919	2.320	-	-4.114	-6.090	-	-4.728	-6.273	-
	(0.028)	(0.990)	-	(0.000)	(0.000)	-	(0.000)	(0.000)	-

Table 2.4. Test statistics and p-values (in parentheses) of the Diebold & Mariano (2002) pairwise test between competing CAViaR models in predicting one-week-ahead returns using the joint approach with the multiplicative factor in (2.10) (Panel A) and the AR formulation in (2.11)-(2.12) (Panel B) for the ES. In each panel, the null hypothesis is that on average, the forecasts obtained with model i are not statistically different from those obtained with model j using the multivariate scoring rule $S^{(j)}_{MAL_t}(\boldsymbol{\tau})$ in (2.28).

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$
	AR formulation		
	CAViaR-AS	CAViaR-AS	CAViaR-AS
Multiplicative factor			
CAViaR-AS	-1.997	-6.273	-5.029
	(0.023)	(0.000)	(0.000)

Table 2.5. Test statistics and p-values (in parentheses) of the Diebold & Mariano (2002) pairwise test between the CAViaR-AS specifications using the joint approach with the constant multiplicative factor in (2.10) and the AR formulation in (2.11)-(2.12) for the ES component in predicting one-week-ahead returns. The null hypothesis is that the two approaches have the same forecasting performance.

out-of-sample performance using the last 368 returns of the sample. The backtesting results for the considered strategies are shown in Table 2.8, where we report the AL log-score of Taylor (2019) together with the S_{FZN} and S_{FZ0} loss functions in (2.29) and (2.30) for the joint evaluation of the pair (VaR, ES). The results clearly show that our approach stands out compared to the other strategies. Indeed, the strategies based on the multivariate Normal and t- distributions and their skewed counterparts produce highly volatile VaR forecasts and suffer from larger average losses over the out-of-sample period. On the other hand, the SMV and Moment-SMV models deliver better performance gains over the MV portfolios, with the SMV being preferred at the three VaR levels, especially in the most extreme case of $\tau = 0.01$. This gain may be traced back to the higher efficiency in the estimation procedure based on the ML approach proposed in Section 2.2.3. It is worth noting that these conclusions remain valid even when we use the S_{FZ0} and S_{FZN} scoring rules, which do not directly depend on the AL likelihood function.

From a financial viewpoint, in Table 2.8, we also evaluate the risk-adjusted returns of the competing portfolios, measured by the Sharpe Ratio (SR) and the Herfindahl-Hirchman Index of weights concentration (HHI). We find that the SMV strategy delivers the portfolio with the highest SR and the least concentrated portfolios at both $\tau = 0.1$ and $\tau = 0.05$, with average HHIs of 0.558 and 0.557, respectively. On the other hand, when $\tau = 0.01$, the MV strategy seems to yield the portfolios with the highest SRs, while the Mom-SMV strategy produces the lowest degree of weights concentration.

A graphical representation of the SMV portfolio weights and their evolution over time is provided in Figure 2.4. In each plot of the figure, the blue line denotes the allocation weights assigned to the FTSE 100, the red line refers to the NIKKEI 225, and the allocation weights of the S&P 500 are displayed in orange. The left panel concerns $\tau = 0.1$, the center one concerns $\tau = 0.05$, and on the right-hand side, we plot the results for $\tau = 0.01$. The SMV strategy tends to invest mainly in the FTSE 100 and the S&P 500, while it tends to hold only a small short position on the NIKKEI 225. It is also interesting that the portfolio weights exhibit the highest

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$			$[0.05, 0.05, 0.05]$			$[0.01, 0.01, 0.01]$		
	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG
Panel A: Multiplicative factor for the ES									
CAViaR-SAV	-	1.383	0.971	-	1.905	0.938	-	0.920	0.274
	-	(0.916)	(0.834)	-	(0.971)	(0.826)	-	(0.821)	(0.608)
CAViaR-AS	-1.383	-	-1.030	-1.905	-	-1.936	-0.920	-	-0.975
	(0.084)	-	(0.152)	(0.029)	-	(0.027)	(0.179)	-	(0.165)
CAViaR-IG	-0.971	1.030	-	-0.938	1.936	-	-0.274	0.975	-
	(0.166)	(0.848)	-	(0.174)	(0.973)	-	(0.392)	(0.835)	-
Panel B: AR formulation for the ES									
CAViaR-SAV	-	0.179	0.273	-	0.473	0.111	-	0.188	-0.740
	-	(0.571)	(0.608)	-	(0.682)	(0.544)	-	(0.575)	(0.230)
CAViaR-AS	-0.179	-	-0.003	-0.473	-	-0.537	-0.188	-	-0.520
	(0.429)	-	(0.499)	(0.318)	-	(0.296)	(0.425)	-	(0.302)
CAViaR-IG	-0.273	0.003	-	-0.111	0.537	-	0.740	0.520	-
	(0.392)	(0.501)	-	(0.456)	(0.704)	-	(0.770)	(0.698)	-

Table 2.6. Test statistics and p-values (in parentheses) of the Diebold & Mariano (2002) pairwise test between competing CAViaR models in predicting one-week-ahead returns using the univariate approach of Taylor (2019) with the multiplicative factor in (2.10) (Panel A) and the AR formulation in (2.11)-(2.12) (Panel B) for the ES. In each panel, the null hypothesis is that, on average the forecasts obtained with model i are not statistically different from those obtained with model j using the $S^{(j)}_{AL_t}(\boldsymbol{\tau})$ scoring rule in (2.31).

$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$			$[0.05, 0.05, 0.05]$			$[0.01, 0.01, 0.01]$		
	Univariate approach								
	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG	CAViaR-SAV	CAViaR-AS	CAViaR-IG
Joint approach									
Panel A: Multiplicative factor for the ES									
CAViaR-SAV	-4.368	-4.432	-4.488	-2.966	-2.748	-2.890	-3.166	-3.348	-3.290
	(0.000)	(0.000)	(0.000)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
CAViaR-AS	-4.386	-4.561	-4.601	-3.108	-2.973	-3.127	-3.066	-2.748	-3.190
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.001)
CAViaR-IG	-4.389	-4.523	-4.573	-3.063	-2.880	-3.029	-2.656	-2.048	-2.690
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.001)	(0.004)	(0.020)	(0.005)
Panel B: AR formulation for the ES									
CAViaR-SAV	-8.913	-8.622	-8.764	-6.290	-6.104	-6.118	-4.467	-4.859	-4.433
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAViaR-AS	-8.536	-8.278	-8.417	-6.413	-6.472	-6.306	-4.856	-5.353	-4.856
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CAViaR-IG	-8.737	-8.440	-8.613	-6.530	-6.467	-6.394	-4.894	-5.336	-4.852
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 2.7. Test statistics and p-values (in parentheses) of the Diebold & Mariano (2002) pairwise test between the competing joint and univariate approaches in predicting one-week-ahead returns with the multiplicative factor in (2.10) (Panel A) and the AR formulation in (2.11)-(2.12) (Panel B) for the ES. The null hypothesis is that the two approaches have the same forecasting performance.

volatility in periods of high uncertainty in the market, such as, for example at the end of 2016 and during the ongoing COVID-19 pandemic. Finally, we evaluate the evolution of the wealth generated by the portfolios at the three risk levels during the out-of-sample period. Figure 2.5 highlights a positive trend for all quantile levels $\tau = 0.1$ (violet line), $\tau = 0.05$ (green line) and $\tau = 0.01$ (yellow line) from 2014 to 2015 and from 2016 until the outbreak of COVID-19 at the beginning of 2020.

2.6 Discussion and conclusions

This paper proposes a dynamic joint quantile regression model for estimating the VaR and ES of multiple financial assets in one step, extending the univariate framework of Taylor (2019). To implement the methodology, we suggest a likelihood-based approach based on the MAL density proposed in Petrella & Raponi (2019), generalized to the case of time-varying parameters. This offers a powerful tool to model the dynamics of multiple VaR and ES jointly. Indeed, the location parameter of the MAL density represents the vector of the VaRs, while the ES can be expressed as a simple function of the density scale parameter.

We show that our approach can offer several important advantages, both theoretical and practical. First, it provides a significant gain in terms of estimation efficiency, as it allows us to estimate multiple VaR and ES in just one step. Second, it can significantly improve the forecast accuracy, since it accounts for the dependence structure among financial assets, which cannot be detected by univariate methods. These results are also confirmed empirically. Indeed, using three stock market indices as in Taylor (2019), we estimate the pairs (VaR, ES) for each of the three assets and evaluate the forecasts using a new scoring function based on the MAL density, which allows us to account for the dependence structure among the considered assets at each point in time. The forecasts of VaR and ES are compared with those obtained by the univariate approach of Taylor (2019), i.e., by considering the three stock market indices separately, as they are independent of each other. We find a significant gain in terms of the forecasting accuracy using the proposed multivariate framework, leading to more reliable risk measure estimates.

Following Zhao et al. (2015), we also exploit the properties of the time-varying MAL distribution to derive a new portfolio optimization method, where the optimal allocation weights are adjusted at each holding period to ensure that the portfolio meets a predetermined level of risk. Empirically, we find that our optimization method produces a portfolio with less concentrated allocation weights and a higher Sharpe Ratio than other existing strategies.

Several extensions and generalizations could be analyzed, leaving space for future research. Although we focused only on CAViaR models, one could consider other VaR-based models in the quantile regression framework or specify different ES dynamics where the factor $(1 + e^{\gamma_{tj}})$ varies over time according to an autoregressive process for γ_{tj}. Another interesting research problem concerns the evaluation of the portfolio performance when a larger set of indices is considered, which may help us in providing an empirical ranking based on the VaR and ES forecasts. In this case, a penalized approach, as used, for instance, by Petrella & Raponi (2019), could be adopted to deal with the curse of dimensionality, improve estimation,

Portfolio	Mean	SD	S_{FZ0}	S_{FZN}	S_{AL}	SR	HHI
Panel A: $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$							
SMV	-2.716	0.883	1.317	1.549	2.473	0.225	0.558
Mom-SMV	-2.785	0.859	1.405	1.592	2.486	0.221	0.563
MV-G-DCC-N	-2.519	1.175	1.399	1.621	2.516	0.192	0.680
MV-G-DCC-t	-2.391	1.107	1.380	1.911	2.491	0.198	0.662
MV-G-aDCC-N	-2.520	1.174	1.399	1.622	2.491	0.195	0.697
MV-G-aDCC-t	-2.392	1.107	1.379	1.912	2.487	0.193	0.673
MV-G-DCC-SN	-3.198	1.483	1.419	1.876	2.511	0.187	0.693
MV-G-DCC-St	-3.582	1.674	1.448	2.216	2.544	0.197	0.676
MV-G-aDCC-SN	-3.199	1.482	1.418	1.876	2.511	0.188	0.691
MV-G-aDCC-St	-3.582	1.633	1.459	2.208	2.544	0.200	0.678
Panel B: $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$							
SMV	-4.016	1.240	1.563	1.612	2.529	0.204	0.557
Mom-SMV	-3.759	1.244	1.592	1.615	2.615	0.189	0.562
MV-G-DCC-N	-3.232	1.508	1.696	1.624	2.735	0.192	0.680
MV-G-DCC-t	-3.174	1.476	1.655	2.024	2.696	0.198	0.662
MV-G-aDCC-N	-3.234	1.507	1.695	1.625	2.735	0.195	0.697
MV-G-aDCC-t	-3.176	1.476	1.655	2.025	2.696	0.193	0.673
MV-G-DCC-SN	-3.812	1.770	1.666	1.868	2.706	0.187	0.693
MV-G-DCC-St	-4.429	2.082	1.668	2.336	2.711	0.197	0.676
MV-G-aDCC-SN	-3.809	1.765	1.664	1.871	2.705	0.188	0.691
MV-G-aDCC-St	-4.430	2.027	1.679	2.327	2.722	0.200	0.678
Panel C: $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$							
SMV	-5.227	1.936	1.989	1.656	3.031	0.154	0.533
Mom-SMV	-5.346	2.036	2.018	1.756	3.043	0.141	0.527
MV-G-DCC-N	-4.570	2.132	2.523	2.201	3.524	0.192	0.680
MV-G-DCC-t	-4.913	2.312	2.263	2.073	3.266	0.198	0.662
MV-G-aDCC-N	-4.572	2.132	2.523	2.202	3.522	0.195	0.697
MV-G-aDCC-t	-4.914	2.311	2.263	2.074	3.266	0.193	0.673
MV-G-DCC-SN	-5.004	2.325	2.406	1.865	3.407	0.187	0.693
MV-G-DCC-St	-6.405	3.054	2.125	2.534	3.129	0.197	0.676
MV-G-aDCC-SN	-5.009	2.336	2.395	1.877	3.396	0.188	0.691
MV-G-aDCC-St	-6.395	2.948	2.145	2.503	3.149	0.200	0.678

Table 2.8. Evaluation of the out-of-sample forecasts of the portfolios VaR and ES. Mean and SD report the average and standard deviation of the portfolio VaR. S_{FZ0}, S_{FZN} and S_{AL} show the average losses using the scoring functions of Patton et al. (2019), Nolde et al. (2017) and Taylor (2019) in (2.30), (2.29) and (2.32), respectively. SR and HHI denote the portfolio Sharpe Ratio and the averaged Herfindahl-Hirschman Index.

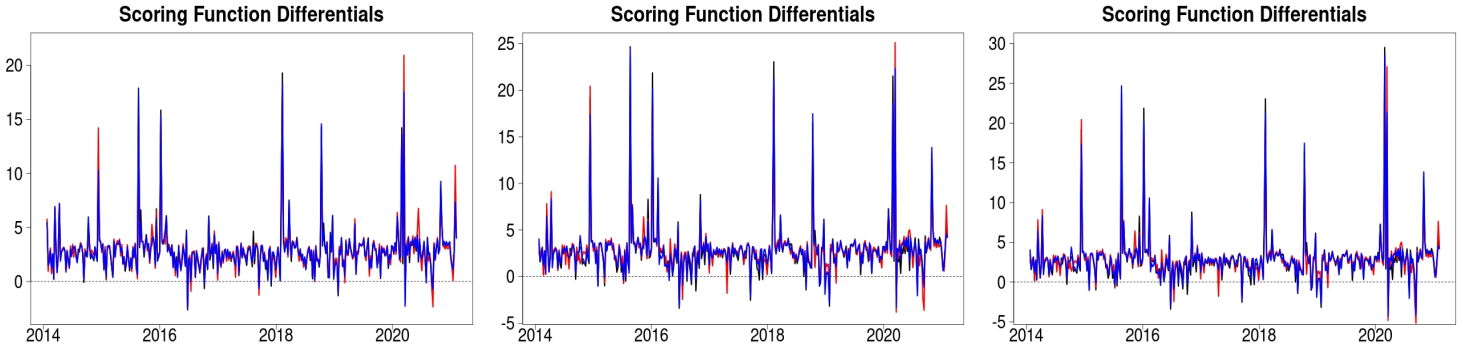

Figure 2.1. Scoring function differentials, $S_{AL_t}(\boldsymbol{\tau}) - S_{MAL_t}(\boldsymbol{\tau})$, between the $S_{AL_t}(\boldsymbol{\tau})$ loss in (2.31) of the univariate approach of Taylor (2019) and the $S_{MAL_t}(\boldsymbol{\tau})$ loss in (2.28) for the joint method, over the out-of-sample period at $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$ (left plot), $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ (center plot) and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ (right plot) for the CAViaR-SAV (black), CAViaR-AS (red) and CAViaR-IG (blue) specifications, with the ES modeled as in (2.10).

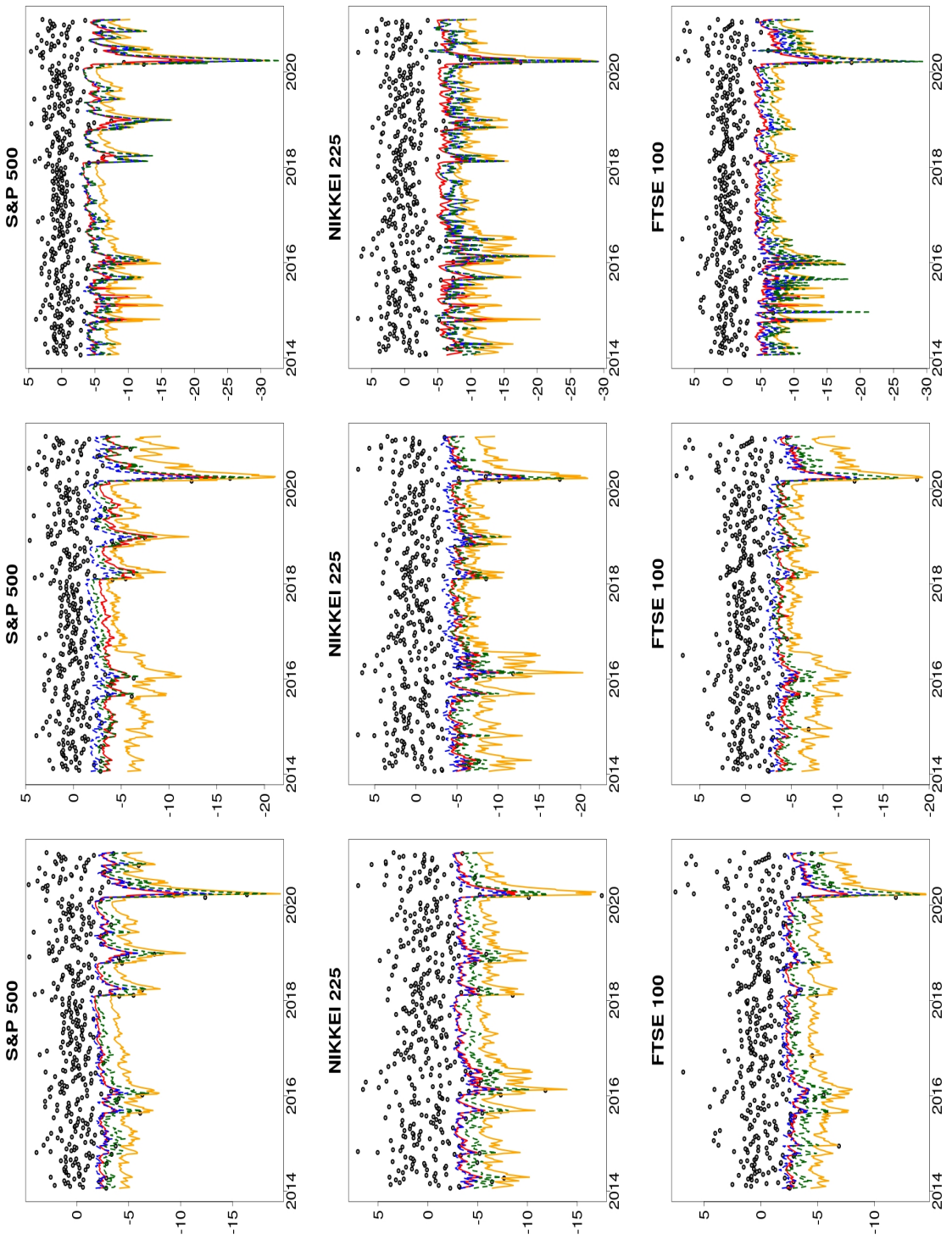

Figure 2.2. Out-of-sample forecasts of VaR and ES for the three stock indices, estimated with the CAViaR-AS specification using both the univariate and joint approaches, with the ES modeled as in (2.10). The dotted blue and the solid red lines refer to the VaR predictions, estimated with the univariate and the multiple approach, respectively. The estimated ES is represented by the dotted green line (for the univariate method of Taylor (2019)) and the solid orange line (for the multivariate approach). The left panels refer to $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, the center panels refer to $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ and the case of $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ is displayed in the right panels. The gray dots represent the observed weekly returns for the considered stock index.

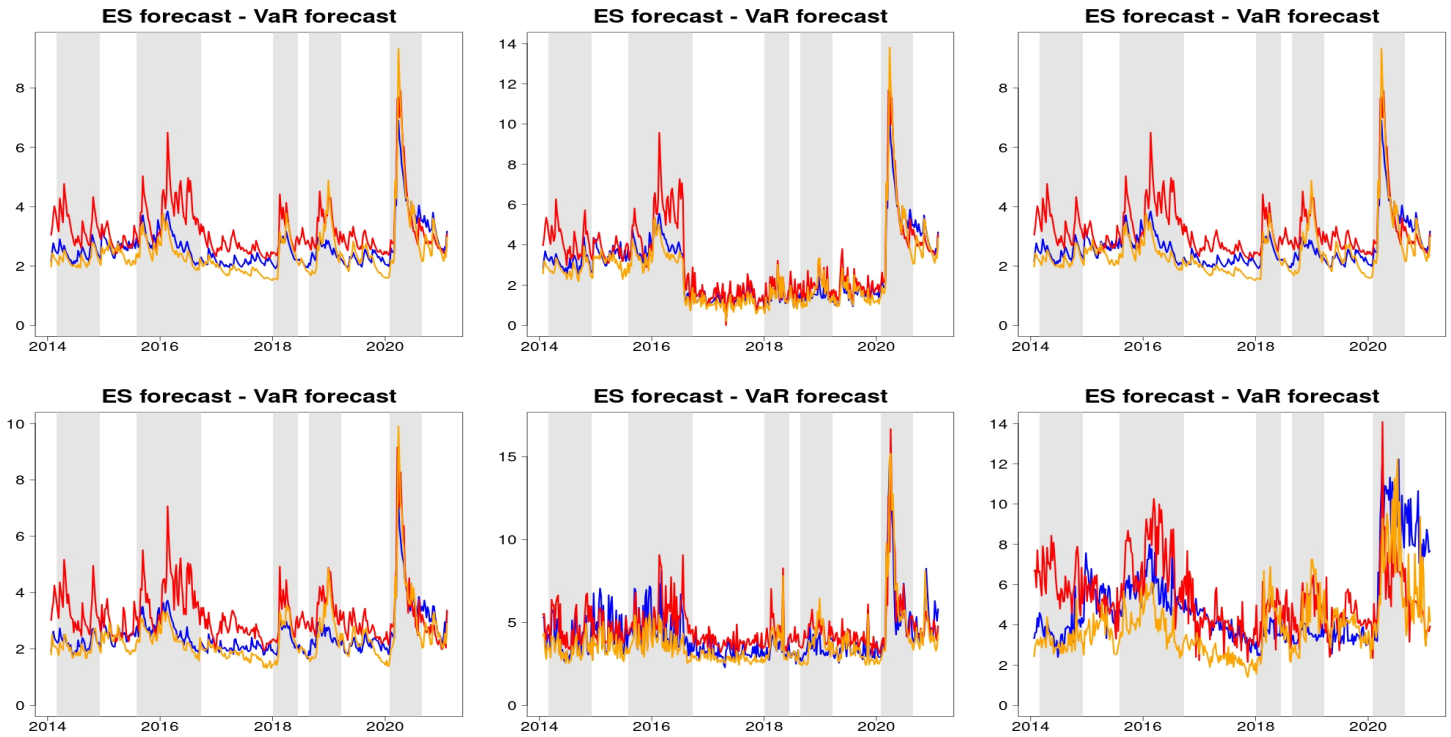

Figure 2.3. Absolute difference between the out-of-sample VaR and ES forecasts for the three stock indices, estimated with the CAViaR-AS specification and using the ES modeled both as in (2.10) (first row) and with the AR specification of (2.11)-(2.12) (second row), at the $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$ (left column), $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ (center column) and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$ (right column) quantile levels. The blue, red and orange lines refer to the FTSE 100, NIKKEI 225 and S&P 500 stock market indices, respectively. The gray bands correspond to the recession dates and to various economic and financial crises that occurred in 2014,03-2015,02; 2015,07-2016,09; 2018,01-2018,06; 2018,08-2019,03; and 2020,02-2020,03.

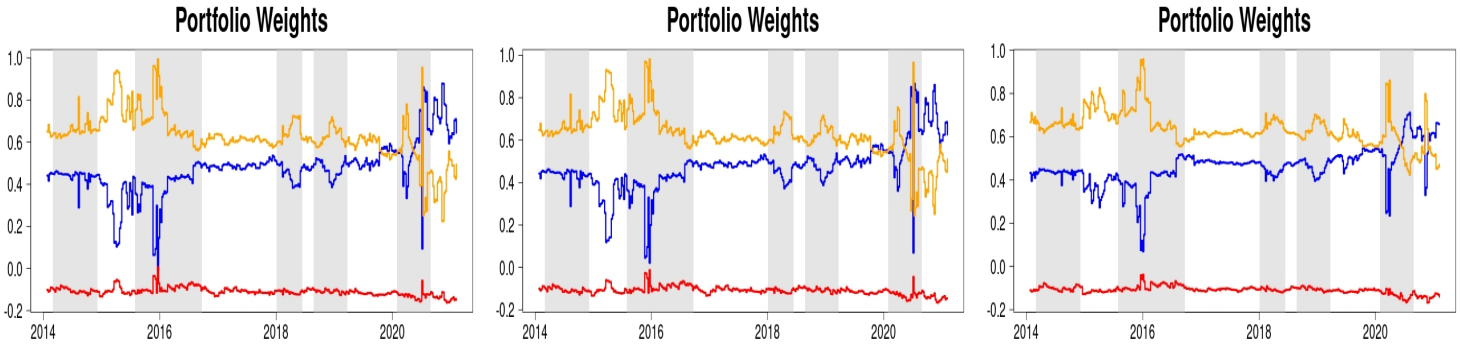

Figure 2.4. Optimal portfolio weights path over the out-of-sample period computed using the selected CAViaR-AS model at $\tau = 0.1$ (left panel), $\tau = 0.05$ (central panel) and $\tau = 0.01$ (right panel). The optimal portfolio weights comprise the FTSE 100 (blue), NIKKEI 225 (red) and S&P 500 (orange) stock market indices. The gray bands correspond to the recession dates and to various economic and financial crises that occurred in 2014,03-2015,02; 2015,07-2016,09; 2018,01-2018,06; 2018,08-2019,03; and 2020,02-2020,03.

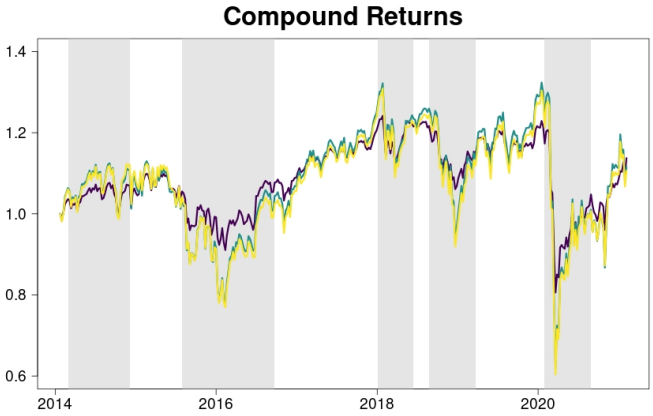

Figure 2.5. Compound returns over the out-of-sample period computed using the selected CAViaR-AS model at $\tau = 0.1$ (violet), $\tau = 0.05$ (green) and $\tau = 0.01$ (yellow). The gray bands correspond to the recession dates and to various economic and financial crises that occurred in 2014,03-2015,02; 2015,07-2016,09; 2018,01-2018,06; 2018,08-2019,03; and 2020,02-2020,03.

achieve greater parsimony and conduct a variable selection procedure. Finally, other portfolio strategies can be implemented, where the choice of the weights may be motivated by other practical considerations or regulatory restrictions.

2.7 Appendix A. Proof of Proposition 4

As stated in Petrella & Raponi (2019), the $\text{MAL}_p(\boldsymbol{\mu}, \mathbf{D}\tilde{\xi}, \mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D})$ in (2.3) can be written as a location-scale mixture, having the following representation:

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{D}\tilde{\xi}\tilde{C} + \sqrt{\tilde{C}}\mathbf{D}\tilde{\boldsymbol{\Sigma}}^{1/2}\mathbf{Z} \tag{2.33}$$

where $\tilde{\xi} = [\tilde{\xi}_1, ..., \tilde{\xi}_p]'$, having generic element $\tilde{\xi}_j = \frac{1-2\tau_j}{\tau_j(1-\tau_j)}$, $j = 1, ..., p$. $\tilde{\boldsymbol{\Sigma}}$ is a $p \times p$ positive definite matrix such that $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Lambda}}\boldsymbol{\Psi}\tilde{\boldsymbol{\Lambda}}$, with $\boldsymbol{\Psi}$ being a correlation matrix and $\tilde{\boldsymbol{\Lambda}} = \text{diag}[\tilde{\sigma}_1, ..., \tilde{\sigma}_p]$, with $\tilde{\sigma}_j^2 = \frac{2}{\tau_j(1-\tau_j)}$, $j = 1, ..., p$. Finally, $\mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ denotes a p-variate standard Normal distribution and $\tilde{C} \sim \text{Exp}(1)$ has a standard Exponential distribution, with $\mathbf{Z}$ being independent of $\tilde{C}$. Notice that, under the constraints imposed on $\tilde{\xi}$ and $\tilde{\boldsymbol{\Lambda}}$, the representation in (2.33) implies that:

$$\mathbf{Y} \mid \tilde{C} = \tilde{c} \sim \mathcal{N}_p(\boldsymbol{\mu} + \mathbf{D}\tilde{\xi}\tilde{c}, \tilde{c}\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}). \tag{2.34}$$

Let $\phi_{\mathbf{Y}}(\mathbf{t})$ denote the characteristic function of $\mathbf{Y}$, with $\mathbf{t} \in \mathcal{R}^p$. Using the result in (2.33), it follows that:

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}_{\tilde{C}}\left[\mathbb{E}_Y[e^{i\mathbf{t}'\mathbf{Y}} \mid \tilde{C} = \tilde{c}]\right] = \int_0^\infty \mathbb{E}_Z[e^{i\mathbf{t}'(\boldsymbol{\mu}+\mathbf{D}\tilde{\xi}\tilde{c}+\sqrt{\tilde{c}}\mathbf{D}\tilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}\mathbf{Z})} \mid \tilde{C} = \tilde{c}]e^{-\tilde{c}}d\tilde{c}. \tag{2.35}$$

Now, using the conditional distribution of $\mathbf{Y}$ given $\tilde{C}$ in (2.34), we have that:

$$\mathbb{E}_Z[e^{i\mathbf{t}'(\boldsymbol{\mu}+\mathbf{D}\tilde{\xi}\tilde{c}+\sqrt{\tilde{c}}\mathbf{D}\tilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}\mathbf{Z})} \mid \tilde{C} = \tilde{c}] = e^{i\mathbf{t}'\boldsymbol{\mu}+i\mathbf{t}'\mathbf{D}\tilde{\xi}\tilde{c}-\frac{\tilde{c}}{2}\mathbf{t}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{t}}.$$

Substituting this result into (2.35) yields:

$$\phi_{\mathbf{Y}}(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu}} \int_0^\infty e^{-\tilde{c}(1+\frac{1}{2}\mathbf{t}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{t}-i\mathbf{t}'\mathbf{D}\tilde{\xi})}d\tilde{c}. \tag{2.36}$$

Finally, integrating over $\tilde{C}$, we obtain:

$$\phi_{\mathbf{Y}}(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu}}\left(1 + \frac{1}{2}\mathbf{t}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{t} - i\mathbf{t}'\mathbf{D}\tilde{\xi}\right)^{-1}. \tag{2.37}$$

Now, let $\mathbf{b} = (b_1, ..., b_p)' \in \mathcal{R}^p$ be a $p \times 1$ vector such that $\mathbf{b} \neq \mathbf{0}_p$ and consider a new random variable $Y^{\mathbf{b}} = \sum_{j=1}^p b_j Y_j = \mathbf{b}'\mathbf{Y}$, having characteristic function $\phi_{Y^{\mathbf{b}}}(z)$, with $z \in \mathcal{R}$. Notice that $Y^{\mathbf{b}}$ is a linear transformation of the marginals $Y_1, \ldots, Y_p$. Therefore, the relation $\phi_{Y^{\mathbf{b}}}(z) = \phi_{\mathbf{b}'\mathbf{Y}}(z) = \phi_{\mathbf{Y}}(\mathbf{b}z)$ holds, since:

$$\phi_{Y^{\mathbf{b}}}(z) = e^{iz\mathbf{b}'\boldsymbol{\mu}}\left(1 + \frac{1}{2}z^2\mathbf{b}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{b} - iz\mathbf{b}'\mathbf{D}\tilde{\xi}\right)^{-1}. \tag{2.38}$$

The characteristic function in (2.38) resembles the characteristic function of the AL univariate distribution discussed in Yu & Moyeed (2001) and Kozumi & Kobayashi (2011) where $\mu^\star, \tau^\star, \delta^\star$ are the scale, skewness and scale parameters, respectively. Therefore, the characteristic function of $Y^{\mathbf{b}}$ in (2.38) can be rewritten as the characteristic function of a univariate AL distribution with parameters:

$$\mu^\star = \mathbf{b}'\boldsymbol{\mu}, \quad \tau^\star = \frac{1}{2}\left(1 - \frac{\mathbf{b}'\mathbf{D}\tilde{\xi}}{\sqrt{2(\mathbf{b}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{b}) + (\mathbf{b}'\mathbf{D}\tilde{\xi})^2}}\right), \quad \delta^\star = \frac{(\mathbf{b}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{b})}{2\sqrt{2(\mathbf{b}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{b}) + (\mathbf{b}'\mathbf{D}\tilde{\xi})^2}}. \tag{2.39}$$

□

In conclusion, we obtain that $Y^{\mathbf{b}} \sim \text{AL}(\mu^\star, \tau^\star, \delta^\star)$ and $\mathbb{P}(Y^{\mathbf{b}} < \mu^\star) = \tau^\star$. Consequently, the parameter $\tau^\star$ controls the probability assigned to each side of $Y^{\mathbf{b}}$ and $\mu^\star$ is the corresponding quantile at level $\tau^\star$. Notice that the denominator $2(\mathbf{b}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{b}) + (\mathbf{b}'\mathbf{D}\tilde{\xi})^2$ in (2.39) is well defined on the positive real line since $(\mathbf{b}'\mathbf{D}\tilde{\xi})^2 \geq 0$ and $2(\mathbf{b}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{b}) > 0$ because $\tilde{\boldsymbol{\Sigma}}$ is a positive definite matrix. Furthermore, when $\boldsymbol{\tau} = [0.5, ..., 0.5]$, we have that $\mathbf{b}'\mathbf{D}\tilde{\xi} = 0$ and (2.39) simplifies to $\tau^\star = 0.5$ and $\delta^\star = \frac{\sqrt{\mathbf{b}'\mathbf{D}\tilde{\boldsymbol{\Sigma}}\mathbf{D}\mathbf{b}}}{2\sqrt{2}}$, which implies that the distribution of $Y^{\mathbf{b}}$ is symmetric around $\mu^\star$.

2.8 Appendix B. Simulation study

In this appendix we conduct a simulation study to evaluate the finite sample properties of the proposed method and its ability to jointly estimate the pair (VaR, ES) for multiple correlated assets. This simulation exercise addresses the following issues. First, we consider different distributional choices for the error term to investigate the behavior of the model in the presence of non-Gaussian errors. Second, we evaluate the bias and accuracy of the ML estimators when the interest of the research is focused upon the lower tails of the distributions.

We consider a sample size of $T = 1500$ and set $p = 3$. The observations are simulated using the following data generating process:

$$\mathbf{Y}_t = \mathcal{Q}_{\mathbf{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1}) + \boldsymbol{\epsilon}_t, \qquad t = 1, 2, ..., T, \tag{2.40}$$

where $\mathcal{Q}_{\mathbf{Y}_t}(\boldsymbol{\tau}|\mathcal{F}_{t-1})$ is generated according to the three different CAViaR specifications described in (2.4)-(2.6). For the ES component, we adopted both the multiplicative factor specification in (2.10) and the AR formulation in (2.11)-(2.12). Following Petrella & Raponi (2019), two different simulation scenarios are considered for the error terms $\boldsymbol{\epsilon}_t$ in (2.40):

(i) multivariate Normal distribution ($\mathcal{N}_3$) with zero mean and variance-covariance matrix equal to $\mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t$, that is $\boldsymbol{\epsilon}_t \sim \mathcal{N}_3(\mathbf{0}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)$;

(ii) multivariate Student-t distribution ($\mathcal{T}_3$) with 5 degrees of freedom, scale parameter $\mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t$ and non centrality parameter $\mathbf{D}_t\tilde{\xi}$, that is, $\boldsymbol{\epsilon}_t \sim \mathcal{T}_3(5, \mathbf{D}_t\tilde{\xi}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)$.

The true values of the CAViaR model and the ES dynamics are calibrated using the real data in the empirical application. Specifically, we set

$$\boldsymbol{\omega} = [-0.20, -0.12, -0.24]', \quad \boldsymbol{\eta} = [0.85, 0.70, 0.60]',$$
$$\boldsymbol{\beta}_1 = [-0.10, -0.05, -0.20]', \quad \boldsymbol{\beta}_2 = [0.05, 0.10, 0.20]'$$

and

$$\boldsymbol{\gamma}_0 = [-1.1, -1.5, -1.3]', \quad \boldsymbol{\gamma}_1 = [0.05, 0.10, 0.02]',$$
$$\boldsymbol{\gamma}_2 = [0.12, 0.05, 0.20]' \quad \text{and} \quad \boldsymbol{\gamma}_3 = [0.80, 0.70, 0.60]'.$$

For the CAViaR-IG dynamic, each element of the vector $\boldsymbol{\omega}$ is considered in terms of absolute value to guarantee that the autoregressive process in (2.6) is well-defined. Finally, we set $\boldsymbol{\Psi} = \begin{pmatrix} 1 & 0.3 & 0.7 \\ 0.3 & 1 & 0.5 \\ 0.7 & 0.5 & 1 \end{pmatrix}$.
Since we are interested in evaluating the downside risk, we analyze three different quantile vectors, namely, $\boldsymbol{\tau} = [0.1, 0.1, 0.1]$, $\boldsymbol{\tau} = [0.05, 0.05, 0.05]$ and $\boldsymbol{\tau} = [0.01, 0.01, 0.01]$. For each model, we carry out $B = 250$ Monte Carlo replications and report the percentage relative bias (Bias%) and the Root Mean Square Error (RMSE) averaged across the B simulations. Tables 2.9 and 2.10 report the results for the parameters $\boldsymbol{\omega}, \boldsymbol{\eta}$ and $\boldsymbol{\beta}$ of the three CAViaR specifications. As can be noted, our estimation method is able to recover the true CAViaR specifications under both the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios and for both the considered ES dynamics. Indeed, both the Bias% and the RMSE remain reasonably small under all the different scenarios even though, as expected, their values tend to increase slightly as the quantile level becomes more extreme (due to the reduced information available at the tails of the distribution) and when we consider a heavy-tailed distribution ($\mathcal{T}_3$ scenario). To computationally evaluate the speed of convergence of the EM algorithm, in the last row of each panel, we also report the median number of iterations and CPU Time (in seconds) required by the implemented `R` code using an Intel Xeon E5-2609 2.40GHz processor. The running times range from 9.613 seconds for the simplest SAV specification with a constant multiplicative factor specified for the ES component to 47.242 seconds for the most complex AS model with an autoregressive ES component, confirming the practical feasibility of our optimization algorithm.

CAViaR	SAV			AS			IG		
$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$
Panel A: $\boldsymbol{\epsilon}_t \sim \mathcal{N}_3(\mathbf{0}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)$									
ω_1	−0.120 (0.027)	0.758 (0.047)	3.874 (0.149)	−1.621 (0.036)	−2.048 (0.050)	−3.472 (0.151)	−1.494 (0.031)	2.520 (0.044)	3.749 (0.156)
ω_2	−1.020 (0.037)	−3.119 (0.052)	2.057 (0.145)	−1.265 (0.030)	−3.738 (0.054)	−4.086 (0.169)	−1.334 (0.036)	1.558 (0.053)	3.622 (0.142)
ω_3	−1.371 (0.028)	−2.245 (0.053)	3.900 (0.162)	−1.454 (0.035)	−2.617 (0.063)	−3.238 (0.147)	−1.195 (0.033)	2.751 (0.048)	3.964 (0.148)
η_1	1.184 (0.025)	1.337 (0.057)	−1.528 (0.153)	2.241 (0.030)	2.379 (0.049)	−3.185 (0.155)	0.548 (0.048)	−2.690 (0.059)	−2.941 (0.159)
η_2	1.228 (0.035)	1.753 (0.041)	−2.537 (0.143)	−0.179 (0.034)	−1.855 (0.062)	1.992 (0.199)	1.794 (0.039)	−2.314 (0.041)	−3.570 (0.144)
η_3	1.560 (0.027)	2.472 (0.036)	−1.132 (0.174)	1.830 (0.041)	2.991 (0.062)	4.230 (0.178)	1.981 (0.031)	2.220 (0.054)	2.781 (0.153)
β_{11}	−1.573 (0.031)	−2.688 (0.038)	3.971 (0.155)	−1.409 (0.035)	−3.308 (0.056)	−3.366 (0.168)	−0.141 (0.040)	2.704 (0.041)	3.497 (0.162)
β_{12}	−0.841 (0.033)	−1.521 (0.060)	2.459 (0.140)	−0.517 (0.031)	−0.797 (0.045)	1.871 (0.159)	1.427 (0.045)	1.981 (0.062)	2.082 (0.185)
β_{13}	−1.799 (0.045)	−2.066 (0.057)	3.211 (0.165)	−1.439 (0.039)	−1.937 (0.055)	3.540 (0.167)	1.899 (0.041)	2.853 (0.073)	3.198 (0.159)
β_{21}				1.862 (0.033)	2.008 (0.046)	5.335 (0.147)			
β_{22}				−0.187 (0.039)	1.705 (0.054)	4.059 (0.161)			
β_{23}				0.634 (0.039)	1.089 (0.049)	3.899 (0.158)			
Iterations	6	8	12	8	14	27	6	9	14
CPU Time	9.613	11.824	17.090	11.498	22.830	33.933	9.669	12.816	18.159
Panel B: $\boldsymbol{\epsilon}_t \sim \mathcal{T}_3(5, \mathbf{D}_t\tilde{\boldsymbol{\xi}}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)$									
ω_1	−1.196 (0.039)	−2.393 (0.063)	2.996 (0.156)	−0.196 (0.049)	−2.898 (0.051)	−2.393 (0.168)	−1.291 (0.029)	1.996 (0.056)	3.560 (0.163)
ω_2	−1.833 (0.045)	−1.572 (0.059)	1.385 (0.162)	−2.833 (0.035)	−2.380 (0.052)	−4.572 (0.179)	−1.533 (0.034)	1.385 (0.052)	2.608 (0.160)
ω_3	−1.397 (0.036)	−1.297 (0.041)	3.736 (0.168)	−2.397 (0.046)	−2.120 (0.061)	−4.297 (0.151)	−1.558 (0.032)	1.736 (0.068)	1.825 (0.169)
η_1	2.699 (0.051)	0.292 (0.067)	−2.097 (0.170)	2.699 (0.041)	2.266 (0.057)	2.292 (0.167)	1.132 (0.027)	−2.097 (0.050)	−1.940 (0.164)
η_2	1.050 (0.048)	1.849 (0.068)	−4.367 (0.177)	0.050 (0.038)	1.421 (0.047)	1.849 (0.198)	1.805 (0.037)	−2.367 (0.047)	−3.034 (0.153)
η_3	1.508 (0.042)	2.455 (0.061)	3.581 (0.167)	1.508 (0.032)	1.783 (0.049)	3.455 (0.151)	1.335 (0.036)	1.581 (0.057)	2.243 (0.140)
β_{11}	−1.718 (0.030)	−0.402 (0.050)	2.688 (0.154)	−2.718 (0.049)	−3.440 (0.068)	−3.702 (0.160)	−1.908 (0.036)	2.688 (0.054)	1.954 (0.162)
β_{12}	−1.615 (0.048)	−0.402 (0.052)	3.189 (0.163)	−1.615 (0.038)	−2.691 (0.064)	−2.402 (0.147)	−0.940 (0.034)	1.189 (0.053)	2.514 (0.189)
β_{13}	−2.557 (0.037)	1.942 (0.049)	3.087 (0.180)	−2.157 (0.047)	−2.542 (0.060)	1.942 (0.199)	1.769 (0.041)	2.087 (0.080)	3.941 (0.170)
β_{21}				1.371 (0.059)	1.057 (0.056)	3.782 (0.170)			
β_{22}				−0.803 (0.040)	−1.690 (0.052)	1.166 (0.166)			
β_{23}				0.489 (0.039)	1.607 (0.065)	2.210 (0.163)			
Iterations	7	10	15	11	16	28	7	11	16
CPU Time	9.617	12.799	18.141	11.691	22.983	34.631	9.628	13.812	18.385

Table 2.9. Bias% and RMSE (in brackets) of the point estimates for $\boldsymbol{\omega}, \boldsymbol{\eta}$ and $\boldsymbol{\beta}$ of the three CAViaR specifications in (2.4)-(2.6) with the ES modeled as a multiple of VaR as in (2.10), under the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios. The last two rows of each panel show the median number of iterations and CPU Time (in seconds) required to fit the model using a single run of the EM algorithm.

CAViaR	SAV			AS			IG		
$\boldsymbol{\tau}$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$	$[0.1, 0.1, 0.1]$	$[0.05, 0.05, 0.05]$	$[0.01, 0.01, 0.01]$
Panel A: $\boldsymbol{\epsilon}_t \sim \mathcal{N}_3(\mathbf{0}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)$									
ω_1	−0.688 (0.037)	−2.193 (0.094)	−3.853 (0.155)	−2.266 (0.060)	2.609 (0.156)	2.245 (0.196)	0.505 (0.059)	−2.726 (0.075)	2.522 (0.150)
ω_2	−0.413 (0.037)	−2.936 (0.053)	−3.462 (0.165)	−1.189 (0.065)	2.139 (0.055)	3.807 (0.191)	−1.986 (0.077)	−2.309 (0.084)	3.130 (0.162)
ω_3	−1.796 (0.055)	−1.046 (0.063)	−3.785 (0.145)	−1.949 (0.057)	1.749 (0.073)	2.424 (0.165)	−1.162 (0.064)	−4.810 (0.121)	3.644 (0.153)
η_1	1.691 (0.037)	2.940 (0.061)	2.999 (0.140)	0.642 (0.061)	−3.533 (0.067)	−2.804 (0.199)	1.430 (0.040)	−2.116 (0.074)	−3.672 (0.162)
η_2	1.884 (0.043)	−2.023 (0.050)	−2.133 (0.177)	−1.217 (0.052)	−2.763 (0.091)	−4.777 (0.140)	−1.813 (0.054)	−1.854 (0.119)	−3.088 (0.155)
η_3	1.018 (0.065)	−1.174 (0.057)	3.017 (0.162)	−3.500 (0.065)	−2.376 (0.045)	−2.197 (0.150)	1.158 (0.081)	−2.133 (0.136)	−2.316 (0.167)
β_{11}	−1.850 (0.048)	−3.514 (0.045)	3.345 (0.133)	1.850 (0.064)	1.395 (0.053)	4.856 (0.176)	−2.688 (0.060)	1.699 (0.102)	2.803 (0.189)
β_{12}	−1.609 (0.041)	−2.701 (0.057)	2.465 (0.186)	2.602 (0.066)	3.041 (0.091)	3.230 (0.141)	−2.357 (0.074)	−2.640 (0.128)	2.921 (0.182)
β_{13}	−1.224 (0.038)	−1.291 (0.045)	2.808 (0.164)	1.659 (0.063)	2.407 (0.067)	4.289 (0.119)	−1.928 (0.065)	−2.191 (0.117)	4.053 (0.143)
β_{21}				1.320 (0.058)	−2.763 (0.141)	−2.013 (0.193)			
β_{22}				1.599 (0.069)	−2.484 (0.121)	−3.276 (0.162)			
β_{23}				1.082 (0.060)	−2.179 (0.118)	3.936 (0.174)			
Iterations	12	14	19	15	23	35	10	15	18
CPU Time	13.791	20.739	25.738	16.112	22.956	45.946	13.796	20.781	25.784
Panel B: $\boldsymbol{\epsilon}_t \sim \mathcal{T}_3(5, \mathbf{D}_t\tilde{\xi}, \mathbf{D}_t\tilde{\boldsymbol{\Sigma}}\mathbf{D}_t)$									
ω_1	−1.687 (0.049)	−3.265 (0.121)	−3.792 (0.151)	−2.793 (0.037)	2.309 (0.126)	3.497 (0.215)	−1.006 (0.038)	−2.278 (0.117)	3.580 (0.142)
ω_2	−2.891 (0.078)	−2.992 (0.084)	−4.143 (0.189)	−1.244 (0.048)	4.041 (0.121)	4.667 (0.223)	−2.308 (0.069)	−3.344 (0.102)	2.875 (0.127)
ω_3	−1.169 (0.087)	−2.719 (0.128)	−4.852 (0.156)	−1.529 (0.070)	1.858 (0.093)	2.045 (0.216)	−1.845 (0.044)	−2.442 (0.125)	4.644 (0.194)
η_1	2.834 (0.038)	3.721 (0.096)	−3.089 (0.157)	1.081 (0.057)	2.490 (0.171)	−3.874 (0.190)	1.595 (0.051)	2.387 (0.093)	−4.524 (0.146)
η_2	−1.345 (0.062)	−2.788 (0.166)	−4.423 (0.154)	−1.449 (0.063)	−2.036 (0.176)	−3.810 (0.215)	0.404 (0.059)	−2.002 (0.086)	−4.258 (0.189)
η_3	1.642 (0.067)	−2.089 (0.094)	−2.026 (0.144)	−1.104 (0.073)	2.716 (0.096)	−2.981 (0.221)	2.186 (0.057)	3.929 (0.106)	2.565 (0.192)
β_{11}	−2.171 (0.046)	−3.313 (0.091)	2.090 (0.132)	0.920 (0.066)	3.032 (0.133)	3.917 (0.186)	−1.893 (0.040)	−3.994 (0.138)	3.613 (0.161)
β_{12}	−2.212 (0.050)	−3.605 (0.112)	1.394 (0.146)	1.261 (0.082)	2.959 (0.090)	3.883 (0.197)	−1.525 (0.047)	−2.143 (0.104)	3.346 (0.181)
β_{13}	−1.536 (0.042)	−2.111 (0.085)	3.773 (0.129)	1.971 (0.082)	3.189 (0.145)	2.862 (0.184)	−2.807 (0.054)	−3.592 (0.092)	4.842 (0.164)
β_{21}				1.171 (0.051)	1.847 (0.137)	3.866 (0.177)			
β_{22}				1.229 (0.067)	2.621 (0.154)	3.195 (0.196)			
β_{23}				1.456 (0.063)	2.158 (0.133)	3.535 (0.187)			
Iterations	13	14	20	16	24	37	11	16	20
CPU Time	13.797	20.760	25.790	16.783	22.176	47.242	13.794	20.752	26.766

Table 2.10. Bias% and RMSE (in brackets) of the point estimates for $\boldsymbol{\omega}, \boldsymbol{\eta}$ and $\boldsymbol{\beta}$ of the three CAViaR specifications in (2.4)-(2.6) with the AR process for the ES as in (2.11)-(2.12) under the $\mathcal{N}_3$ and $\mathcal{T}_3$ scenarios. The last two rows of each panel show the median number of iterations and CPU Time (in seconds) required to fit the model using a single run of the EM algorithm.

Chapter 3

Marginal M-quantile regression for multivariate dependent data

3.1 Introduction

Quantile regression has attracted considerable interest in many empirical studies since its introduction in the seminal paper of Koenker & Bassett (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 can be often 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, Bernardi, Durante, Jaworski, Petrella & Salvadori 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 2009). 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, conditional 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 3.2 we introduce the main notation and briefly review the M-quantile regression model. Section 3.3 describes the proposed methodology and Section 3.4 discusses the estimation procedure and develops the asymptotic theory. Section 3.5 presents the simulation study and the results. The application is presented in Section 3.6 and Section 3.7 presents some concluding remarks.

3.2 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:

$$\int \psi_\tau(y - \theta_{\mathbf{x}}(\tau)) dF_{Y|\mathbf{X}}(y \mid \mathbf{x}) = 0, \tag{3.1}$$

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:

$$\theta_{\mathbf{x}}(\tau) = \mathbf{x}'\boldsymbol{\beta}(\tau), \tag{3.2}$$

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

In this work, the influence function $\psi(\cdot)$ in (3.1) is chosen to be the well-known Huber influence function (Huber 1964):

$$\psi(u) = u\mathbf{1}_{(|u|\leq c)} + c\,\text{sign}(u)\mathbf{1}_{(|u|>c)}, \tag{3.3}$$

where c denotes a tuning constant bounded away from zero. The function in (3.3) 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 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 (3.2) 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.

3.3 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 $\mathbf{Y}_{ij} = (Y_{ij}^{(1)}, \ldots, Y_{ij}^{(p)})'$ with $\mathbf{y}_{ij}$ being the corresponding observed value and let $\mathbf{X}_{ij} = (X_{ij}^{(1)}, \ldots, X_{ij}^{(k)})$ 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 $n = \sum_{j=1}^{d} n_j$. Also, let $\mathbf{u}$ denote a unit norm direction vector ranging over the $(p-1)$-dimensional unit sphere $\mathcal{S}^{p-1} = \{\mathbf{z} \in \mathbb{R}^p : ||\mathbf{z}|| = 1\}$. To simplify the notation, we stack up the projected responses on $\mathbf{u}$ to the n_j dimensional vector $\tilde{\mathbf{Y}}_j = (\mathbf{u}'\mathbf{Y}_{1j}, \ldots, \mathbf{u}'\mathbf{Y}_{n_j j})'$, while $\mathbf{X}_j = (\mathbf{X}_{ij}, \ldots, \mathbf{X}_{n_j j})$ 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* (3.3)*. 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 3.2. Specifically, by varying the tuning constant c in (3.3), 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:

$$\theta_{\mathbf{u},\mathbf{x}}(\tau) = \mathbf{x}'_{ij}\boldsymbol{\beta}(\tau, \mathbf{u}), \qquad i = 1, ..., n_j \text{ and } j = 1, ..., d, \tag{3.4}$$

where $\mathbf{x}_{ij}$ is the covariates vector for the i-th subject in the j-th group and $\boldsymbol{\beta}(\tau, \mathbf{u})$ is the k-dimensional vector of regression coefficients. The linear model in (3.4) assumes that $\beta(\tau, \mathbf{u})$ represents the effect of the covariates on the conditional τ-th directional M-quantile of $\mathbf{Y}_{ij}$ given $\mathbf{X}_{ij} = \mathbf{x}_{ij}$ in the direction $\mathbf{u}$.
The considered directional approach is a viable way to take into account not only the dependence among variables, circumventing the problem of defining a multivariate M-quantile, but also to get useful insights on the dependence of $\mathbf{Y}$ on the set of covariates $\mathbf{X}$, fully characterizing the conditional distribution of the response.

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 $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ as the solution of the following Generalized M-quantile Estimating Equations (GMQEE):

$$\mathbf{U}(\boldsymbol{\beta}(\tau, \mathbf{u})) = \sum_{j=1}^{d} \mathbf{U}_j(\boldsymbol{\beta}(\tau, \mathbf{u})) = \sum_{j=1}^{d} \mathbf{X}_j' \boldsymbol{\Sigma}_j^{-1}(\mathbf{r}_j) \mathbf{V}_j^{\frac{1}{2}} \psi_\tau(\mathbf{z}_j) = \mathbf{0}, \tag{3.5}$$

where $\mathbf{z}_j = \mathbf{V}_j^{-\frac{1}{2}}(\tilde{\mathbf{Y}}_j - \mathbf{X}_j \boldsymbol{\beta}(\tau, \mathbf{u}))$ 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 $\tilde{\mathbf{Y}}_j - \mathbf{X}_j \boldsymbol{\beta}(\tau, \mathbf{u})$, $\psi_\tau(\cdot)$ is the influence function in (3.3) and

$$\boldsymbol{\Sigma}_j(\mathbf{r}_j) = \mathbf{V}_j^{\frac{1}{2}} \mathbf{C}_j(\mathbf{r}_j) \mathbf{V}_j^{\frac{1}{2}}, \tag{3.6}$$

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, (3.5) reduces to:

$$\mathbf{U}_I(\boldsymbol{\beta}(\tau, \mathbf{u})) = \sum_{j=1}^{d} \mathbf{X}_j' \mathbf{V}_j^{-\frac{1}{2}} \psi_\tau(\mathbf{z}_j) = \mathbf{0}, \tag{3.7}$$

where independence between clustered observations is assumed. In this case, we denote $\widehat{\boldsymbol{\beta}}_I(\tau, \mathbf{u})$ the estimator of $\boldsymbol{\beta}(\tau, \mathbf{u})$ as the solution of (3.7). Second, contrary to the well known GEE estimator of Liang & Zeger (1986), using the Huber's function ensures that $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ behaves robustly against outliers for finite values of c. Furthermore, by focusing on linear combinations of $\mathbf{Y}$, inference on $\boldsymbol{\beta}_{MMQ}(\tau, \mathbf{u})$ accounts for the possible correlation between the outcomes through the working correlation structure in (3.6). 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), (3.5) 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 $\widehat{\boldsymbol{\beta}}_I(\tau, \mathbf{u})$ (Liang & Zeger 1986, Crowder 1995, Wang & Carey 2003, 2004 and Hin & Wang 2009). We present the asymptotic properties of the proposed estimator $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ in Section 3.4.1.

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.

3.3.1 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:

$$\pi_{\mathbf{u},\mathbf{x}}(\tau) = \{\mathbf{y} \in \mathbb{R}^p : \mathbf{u}'\mathbf{y} = \theta_{\mathbf{u},\mathbf{x}}(\tau)\}, \tag{3.8}$$

where $\theta_{\mathbf{u},\mathbf{x}}(\tau)$ is defined in (3.4). For example, when $p = 2$, the hyperplanes in (3.8) 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 (3.8). 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:

$$R_{\mathbf{x}}(\tau) = \bigcap_{\mathbf{u} \in \mathcal{S}^{p-1}} H^{+}_{\mathbf{u},\mathbf{x}}(\tau). \tag{3.9}$$

The region defined in (3.9) 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 (3.3) 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.

3.4 Estimation and inference

In this section we provide the algorithm to compute an estimate of the robust estimator $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ in (3.5), 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 $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$.

In order to estimate $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ and the corresponding covariance matrix $\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u}))$ we propose to use the iterative Newton-Raphson algorithm to solve the GMQEE in (3.5). 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:

$$r = \frac{\sum_{j=1}^{d} \sum_{i<i'}^{n_j} \psi_\tau(z_{ij}) \psi_\tau(z_{i'j})}{\phi(\sum_{j=1}^{d} \frac{1}{2} n_j(n_j - 1) - k)} \qquad \textit{Exchangeable}, \tag{3.10}$$

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

$$r = \frac{\sum_{j=1}^{d} \sum_{i=1}^{n_j - 1} \psi_\tau(z_{ij}) \psi_\tau(z_{i+1j})}{\phi(\sum_{j=1}^{d} n_j(n_j - 1) - k)} \qquad \textit{Autoregressive}. \tag{3.11}$$

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

$$r_{ii'} = \frac{\sum_{j=1}^{d} \psi_\tau(z_{ij}) \psi_\tau(z_{i'j})}{\phi(d - k)}, i \neq i' \qquad \textit{Unstructured}, \tag{3.12}$$

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 $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ and $\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u}))$.

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 (3.8) and (3.9). 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 $\{\widehat{H}^{+}_{\mathbf{u},\mathbf{x}}(\tau), \mathbf{u} \in \mathcal{S}_B^{p-1}\}$ which allows us to compute the estimate $\widehat{R}_{\mathbf{x}}(\tau)$ of $R_{\mathbf{x}}(\tau)$ as:

$$\widehat{R}_{\mathbf{x}}(\tau) = \bigcap_{\mathbf{u} \in \mathcal{S}_B^{p-1}} \{\widehat{H}^{+}_{\mathbf{u},\mathbf{x}}(\tau)\}. \tag{3.13}$$

We then estimate the contour $\partial \widehat{R}_{\mathbf{x}}(\tau)$ from (3.13).

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:

$$\Pr(\partial R_{\mathbf{x}}(\tau) \subset \mathcal{C}_{d,1-\alpha}) = 1 - \alpha, \tag{3.14}$$

Algorithm The GMQEE algorithm

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

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

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

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

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

with $\tilde{y}_{ij}$ being the i-th element of the vector $\tilde{\mathbf{y}}_j$ defined in Section 3.3.

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 $\widehat{\boldsymbol{\beta}}^{(b)}(\tau,\mathbf{u})$ and $\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}}^{(b)}(\tau,\mathbf{u}))$ by:

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

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

where

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

and

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

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}.

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

$$w_{1-\alpha} = F_W^{-1}(1-\alpha). \tag{3.15}$$

Then, it is easy to see that:

$$\Pr(\partial R_{\mathbf{x}}(\tau) \subset \partial\widehat{R}_{\mathbf{x}}(\tau) \oplus w_{1-\alpha}) \geq 1-\alpha. \tag{3.16}$$

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 $((\mathbf{Y}_1^\star, \mathbf{X}_1^\star), ..., (\mathbf{Y}_{d^\star}^\star, \mathbf{X}_{d^\star}^\star))$ be a bootstrap sample and let $\partial\widehat{R}_{\mathbf{x}}^\star(\tau)$ denote the corresponding estimate of the order-τ M-quantile regression contour. We define $W^\star = \text{Haus}(\partial\widehat{R}_{\mathbf{x}}^\star(\tau), \partial\widehat{R}_{\mathbf{x}}(\tau))$ and define the bootstrap estimate of $w_{1-\alpha}$ as:

$$\widehat{w}_{1-\alpha} = F_{W^\star}^{-1}(1-\alpha). \tag{3.17}$$

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.

3.4.1 Asymptotic properties

This section presents the asymptotic properties of the GMQEE estimator. First, we derive the Bahadur-type (Bahadur 1966) representation, consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ for fixed τ and $\mathbf{u}$. Subsequently, we establish the asymptotic properties of the GMQEE estimator when several directions in $\mathcal{S}^{p-1}$ are considered simultaneously. 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 (3.1) is continuous and strictly monotonic.

(iv) The function $\psi(\cdot)$ in (3.3) is bounded, non-decreasing and is twice differentiable at $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$, with the convention $\psi(0) = 0$.

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

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

$$\mathbf{H} = \frac{1}{d}\sum_{j=1}^{d} \mathbf{X}_j' \boldsymbol{\Sigma}_j^{-1} \mathbb{E}[\mathbf{D}_j] \mathbf{X}_j, \tag{3.18}$$

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

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

$$\sqrt{d}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau,\mathbf{u})-\boldsymbol{\beta}(\tau,\mathbf{u}))=\frac{1}{\sqrt{d}}\mathbf{H}^{-1}\sum_{j=1}^{d}\mathbf{U}_j(\boldsymbol{\beta}(\tau,\mathbf{u}))+o(1) \tag{3.19}$$

and

$$\sqrt{d}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau,\mathbf{u})-\boldsymbol{\beta}(\tau,\mathbf{u}))\xrightarrow{p}\mathcal{N}(\mathbf{0},\mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1}) \quad \text{as} \quad d\to\infty, \tag{3.20}$$

with $\mathbf{B}$ *being*

$$\mathbf{B}=\frac{1}{d}\sum_{j=1}^{d}\mathbf{X}_j'\boldsymbol{\Sigma}_j^{-1}\mathbf{V}_j^{\frac{1}{2}}\mathbb{E}[\psi_\tau(\mathbf{z}_j)\psi_\tau'(\mathbf{z}_j)]\mathbf{V}_j^{\frac{1}{2}}\boldsymbol{\Sigma}_j^{-1}\mathbf{X}_j. \tag{3.21}$$

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 $\mathbf{U}(\boldsymbol{\beta}(\tau,\mathbf{u}))$ in (3.5) is continuous in $\boldsymbol{\beta}(\tau,\mathbf{u})$. 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 (3.19). Subsequently, (3.20) follows from (3.19) 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 $\boldsymbol{\beta}(\tau,\mathbf{u})$ 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 addition, to gain better estimation accuracy, we may estimate multiple MMQ models at different directions simultaneously by incorporating the associations among the considered directions. Hence, by proceeding as in the proof of Theorem 1 and applying the multivariate Central Limit Theorem to the Bahadur representation in (3.19), we derive the asymptotic distribution of the GMQEE estimator when multiple directions are considered jointly, as shown in the following remark.

Remark 1. *Let* $\{\mathbf{u}_1,\ldots,\mathbf{u}_J\}\subset\mathcal{S}^{p-1}$*, with* J *being a fixed positive integer. Suppose assumptions (i)-(vi) hold, then the joint asymptotic distribution of* $\sqrt{d}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau,\mathbf{u}_1)-\boldsymbol{\beta}(\tau,\mathbf{u}_1),\ldots,\widehat{\boldsymbol{\beta}}_{MMQ}(\tau,\mathbf{u}_J)-\boldsymbol{\beta}(\tau,\mathbf{u}_J))$ *is Gaussian with zero mean and the asymptotic covariance matrix between* $\sqrt{d}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau,\mathbf{u}_r)-\boldsymbol{\beta}(\tau,\mathbf{u}_r))$ *and* $\sqrt{d}(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau,\mathbf{u}_s)-\boldsymbol{\beta}(\tau,\mathbf{u}_s))$*, where* $1\leq r,s\leq J$*, is given by:*

$$\mathbf{H}(\mathbf{u}_r)^{-1}\mathbf{B}(\mathbf{u}_r,\mathbf{u}_s)\mathbf{H}(\mathbf{u}_s)^{-1}, \tag{3.22}$$

with

$$\mathbf{H}(\mathbf{u}_r)=\frac{1}{d}\sum_{j=1}^{d}\mathbf{X}_j'\boldsymbol{\Sigma}_j^{-1}(\mathbf{u}_r)\mathbb{E}[\mathbf{D}_j(\mathbf{u}_r)]\mathbf{X}_j, \tag{3.23}$$

$$\mathbf{B}(\mathbf{u}_r, \mathbf{u}_s) = \frac{1}{d}\sum_{j=1}^{d} \mathbf{X}_j' \boldsymbol{\Sigma}_j^{-1}(\mathbf{u}_r)\mathbf{V}_j^{\frac{1}{2}}(\mathbf{u}_r)\mathbb{E}[\psi_\tau(\mathbf{z}_j(\mathbf{u}_r))\psi_\tau'(\mathbf{z}_j(\mathbf{u}_s))]\mathbf{V}_j^{\frac{1}{2}}(\mathbf{u}_s)\boldsymbol{\Sigma}_j^{-1}(\mathbf{u}_s)\mathbf{X}_j \tag{3.24}$$

and where the i-th element of $\mathbf{z}_j(\mathbf{u}_r)$ *is* $z_{ij}(\mathbf{u}_r) = (\mathbf{u}_r'\mathbf{y}_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}(\tau, \mathbf{u}_r))/\sigma_{\tau,\mathbf{u}_r}$.

In order to use Theorem 1 to build confidence intervals and hypothesis tests for $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$, a consistent estimator of the asymptotic covariance matrix $\mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1}$ in (3.20) 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.:

$$\widehat{\mathbf{H}} = \frac{1}{d}\sum_{j=1}^{d} \mathbf{X}_j'\widehat{\boldsymbol{\Sigma}}_j^{-1}\widehat{\mathbf{D}}_j\mathbf{X}_j, \tag{3.25}$$

$$\widehat{\mathbf{B}} = \frac{1}{d}\sum_{j=1}^{d} \mathbf{X}_j'\widehat{\boldsymbol{\Sigma}}_j^{-1}\mathbf{V}_j^{\frac{1}{2}}\psi_\tau(\widehat{\mathbf{z}}_j)\psi_\tau'(\widehat{\mathbf{z}}_j)\mathbf{V}_j^{\frac{1}{2}}\widehat{\boldsymbol{\Sigma}}_j^{-1}\mathbf{X}_j, \tag{3.26}$$

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,*

$$\widehat{\mathbf{H}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{H}}^{-1} - \mathbf{H}^{-1}\mathbf{B}\mathbf{H}^{-1} \xrightarrow{p} 0, \tag{3.27}$$

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:

$$\boldsymbol{\Omega}(\widehat{\mathbf{r}}_j) = \Big(\sum_{j=1}^{d}\mathbf{K}_j'\mathbf{K}_j\Big)^{-1}\Big(\sum_{j=1}^{d}\mathbf{K}_j'\mathrm{Cov}(\widehat{\mathbf{s}}_j)\mathbf{K}_j\Big)\Big(\sum_{j=1}^{d}\mathbf{K}_j'\mathbf{K}_j\Big)^{-1}, \tag{3.28}$$

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.

3.4.2 Selection of working correlation structure

In the inferential procedure discussed above, the working correlation structure $\mathbf{C}(\mathbf{r})$ is generally unknown a-priori. From an applied perspective, the choice of an appropriate correlation matrix should be data driven and take into account the structure of the data to achieve greater asymptotic efficiency, using either statistical tools or the analyst's prior knowledge. In the literature, $\mathbf{C}(\mathbf{r})$ is typically estimated via selection model techniques, such as working correlation selection criteria (see e.g., Pan 2001, Hin & Wang 2009, Chen & Lazar 2012, Gosho et al. 2014, Fu et al.

2018). In this work, we select the optimal structure for $\mathbf{C}(\mathbf{r})$ using the Correlation Information Criterion (CIC) of Hin & Wang (2009). More formally, for a given τ and $\mathbf{u}$, the CIC is defined as:

$$CIC(\widehat{\mathbf{C}}) = \text{Tr}(\widehat{\mathbf{H}}_I \widehat{\boldsymbol{\Omega}}), \tag{3.29}$$

where $\widehat{\mathbf{H}}_I$ is the matrix in (3.25) evaluated at $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ under the working independence model, $\widehat{\boldsymbol{\Omega}}$ is the variance-covariance of $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau)$ estimated from (3.25) and (3.26), and $\text{Tr}(\mathbf{A})$ denotes the trace of a square matrix $\mathbf{A}$. Following Pan (2001) and Hin & Wang (2009), in order to estimate the optimal correlation structure, at first we fit the proposed MMQ model using different candidate working correlation structures (e.g., independence, exchangeable, autoregressive, unstructured) and then select the best one corresponding to the lowest value of the criteria in (3.29).

The validity of the proposed estimation algorithm and correlation structure selection procedure are assessed in the next section.

3.5 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; (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; (iii) the performance of the considered information criterion in selecting the optimal correlation structure. 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:

$$\mathbf{Y}_{ij} = \mathbf{X}'_{ij}\mathbf{B} + \boldsymbol{\epsilon}_{ij}, \qquad i = 1, ..., n_j \text{ and } j = 1, ..., d, \tag{3.30}$$

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. $\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = r$ 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 $\mathbf{Y}_{ij} + \delta_{ij}\mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$, where $\delta_{ij} \sim \text{Ber}(\alpha)$, with $\alpha = \Pr(\delta_{ij} = 1)$, and where $\boldsymbol{\Sigma}$ 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 settings correspond 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 $\boldsymbol{\beta}(\tau, \mathbf{u}) = (\beta_0(\tau, \mathbf{u}), \beta_1(\tau, \mathbf{u}))$ can be computed as $\boldsymbol{\beta}(\tau, \mathbf{u}) = \mathbf{B}\mathbf{u}$, where the intercept $\beta_0(\tau, \mathbf{u})$ has been corrected to ensure that the conditional τ-th M-quantile of $\mathbf{u}'\mathbf{Y}$ is equal to $\mathbf{X}'\boldsymbol{\beta}(\tau, \mathbf{u})$. 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 (3.3) 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:

$$ARB(\hat{\theta}_\tau) = \frac{1}{H}\sum_{h=1}^{H}\frac{(\hat{\theta}_\tau^{(h)} - \theta_\tau)}{\theta_\tau} \times 100, \tag{3.31}$$

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 $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ w.r.t. $\widehat{\boldsymbol{\beta}}_I(\tau, \mathbf{u})$, we compute the Relative Efficiency (REF) measure defined as:

$$\text{REF}(\widehat{\boldsymbol{\beta}}(\tau, \mathbf{u})) = \frac{S^2(\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u}))}{S^2(\widehat{\boldsymbol{\beta}}_I(\tau, \mathbf{u}))}, \tag{3.32}$$

where $S^2(\widehat{\boldsymbol{\beta}}(\tau, \mathbf{u})) = \frac{1}{H}\sum_{h=1}^{H}(\widehat{\boldsymbol{\beta}}(\tau, \mathbf{u})^{(h)} - \bar{\boldsymbol{\beta}}(\tau, \mathbf{u}))^2$ and $\bar{\boldsymbol{\beta}}(\tau, \mathbf{u}) = \frac{1}{H}\sum_{h=1}^{H}\widehat{\boldsymbol{\beta}}(\tau, \mathbf{u})^{(h)}$. The REF defined in (3.32) measures the efficiency gain of the estimates of $\boldsymbol{\beta}(\tau, \mathbf{u})$ using the proposed directional M-quantile regression method, $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$, over the independence assumption, $\widehat{\boldsymbol{\beta}}_I(\tau, \mathbf{u})$. When $\text{REF}(\widehat{\boldsymbol{\beta}}(\tau, \mathbf{u}))$ is less than one, this indicates that $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ is preferable. Tables 3.1-3.2 show the ARB and REF measures of the proposed estimators $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ for each component of the parameter vector $\boldsymbol{\beta}(\tau, \mathbf{u})$ 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 3.1) and high (Table 3.2) 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. 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 3.1 shows that with a moderate correlation ($r = 0.3$), the relative efficiencies of the regression estimators $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ 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 $\widehat{\boldsymbol{\beta}}_{I}(\tau, \mathbf{u})$ with low and large r. This demonstrates that the $\widehat{\boldsymbol{\beta}}_{MMQ}(\tau, \mathbf{u})$ 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 3.1-3.2), 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. The results with $\alpha = 5\%$ confirm these findings and are available from the authors. To evaluate how the MMQ estimator is affected under a more extreme level of contamination, we also consider $\alpha = 20\%$ and report the results in Table 3.3. As one can see, the effect of contamination is visible on both the ARB and REF compared to those in Tables 3.1-3.2, however, the increase in the ARB is only a small effect in relation to the increase in the proportion of contaminated data points from $\alpha = 10\%$ to 20%.

To assess the performance of the estimated variances as described in (3.20), we report the Coverage Probability (CP) of nominal 95% confidence intervals for $\beta_0(\tau, \mathbf{u})$ and $\beta_1(\tau, \mathbf{u})$ defined by the number of times the interval $\theta_\tau \pm 2\sqrt{\text{Var}(\hat{\theta}_\tau)}$ contains the "true" population parameter divided by the number of Monte Carlo replicates H. The results presented in Tables 3.4 and 3.5 indicate that under the Gaussian, contaminated Gaussian scheme with $\alpha = 10\%$ 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. When $\alpha = 20\%$, the obtained CPs come close to the 0.95 nominal level even tough there is some degree of over- and under- coverage. This should not surprise us given that in all simulation scenarios, the tuning constant c of the Huber function has been set to 1.345, which gives 95% efficiency in the normal case. However, when large values of α are used, the extent of contamination which can be tolerated by the proposed model can be increased by selecting smaller values of c to further reduce the effects of contamination on inference.

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

In order to determine the performance of the correlation structure selection procedure described in Section 3.4.2, we adopt the same simulation experiment and consider the exchangeable and the AR1 structures with low ($r = 0.3$) and high ($r = 0.8$) correlation for the error term $\boldsymbol{\epsilon}_j = (\boldsymbol{\epsilon}_{1j}, \ldots, \boldsymbol{\epsilon}_{n_j j})'$ related to (3.30). For each of the $H = 1000$ simulated datasets and the scenarios considered, we fit the MMQ model using three candidate working correlation structures (the I, E and AR1) and then select the best one associated to the lowest CIC value in (3.29). Table 3.6 reports the number of times the criterion correctly identifies the true structure that generated the data in all settings. As one can see, the CIC detects the correct working correlation matrix especially at $\tau = 0.50$, and the percentage of correct selection rises when the intra-cluster correlation increases. As we move towards the tails of the distribution of the responses (see $\tau = 0.1$ and $\tau = 0.9$), the performance of the CIC slightly decreases, nonetheless, it is always able to select the correct structure with a detection rate ranging between 42.3% and 90%.

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 3.1 shows the estimated contours under the four data generating processes with a correlation coefficient of

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)	−0.153 (0.999)	−0.077 (0.481)	−0.126 (1.010)	−0.068 (0.348)	0.043 (0.279)	−0.115 (0.320)

Table 3.2. Values of ARB (in percentage) and REF (in brackets) of $\beta_0(\tau, \mathbf{u})$ and $\beta_1(\tau, \mathbf{u})$ over 1000 Monte Carlo simulations under the three data generating scenarios with high correlation ($\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = 0.8$).

$r = 0.3$, 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.

3.6 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 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

Model	Coef	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
		$\mathbf{u}_1$			$\mathbf{u}_2$			$\mathbf{u}_3$		
		Panel A: $r = 0.3$								
I	β_0	−0.078 (1.000)	−0.056 (1.000)	−0.004 (1.000)	−0.084 (1.000)	−0.047 (1.000)	−0.016 (1.000)	−0.086 (1.000)	−0.021 (1.000)	0.073 (1.000)
	β_1	−0.099 (1.000)	−0.029 (1.000)	0.070 (1.000)	−0.155 (1.000)	−0.107 (1.000)	−0.109 (1.000)	−0.134 (1.000)	0.137 (1.000)	−0.142 (1.000)
E	β_0	−0.076 (0.978)	−0.056 (1.000)	−0.004 (0.997)	−0.083 (0.997)	−0.047 (1.006)	−0.016 (1.002)	−0.086 (0.999)	−0.022 (0.988)	0.074 (1.012)
	β_1	−0.028 (0.982)	−0.062 (0.909)	−0.063 (1.006)	−0.104 (0.976)	−0.054 (0.843)	−0.092 (0.981)	0.206 (1.015)	0.101 (0.913)	−0.120 (0.996)
AR1	β_0	−0.076 (0.978)	−0.056 (0.992)	−0.004 (0.996)	−0.083 (0.999)	−0.047 (1.011)	−0.016 (1.003)	−0.086 (0.996)	−0.022 (0.995)	0.074 (1.017)
	β_1	−0.069 (0.988)	−0.074 (0.964)	−0.094 (1.009)	−0.142 (0.976)	−0.091 (0.959)	−0.129 (0.976)	−0.174 (1.014)	0.101 (0.984)	−0.134 (0.998)
U	β_0	−0.084 (1.025)	−0.056 (1.006)	−0.004 (1.028)	−0.089 (1.077)	−0.047 (1.035)	−0.014 (0.990)	−0.088 (1.011)	−0.022 (1.044)	0.081 (1.073)
	β_1	0.117 (1.012)	−0.116 (0.943)	0.042 (1.023)	−0.299 (1.046)	0.022 (0.876)	−0.125 (1.005)	0.181 (1.068)	0.102 (1.031)	−0.136 (1.079)
		Panel B: $r = 0.8$								
I	β_0	−0.105 (1.000)	−0.048 (1.000)	−0.024 (1.000)	−0.091 (1.000)	−0.057 (1.000)	−0.060 (1.000)	−0.080 (1.000)	−0.041 (1.000)	−0.024 (1.000)
	β_1	−0.218 (1.000)	−0.090 (1.000)	−0.098 (1.000)	−0.166 (1.000)	−0.044 (1.000)	−0.057 (1.000)	0.199 (1.000)	−0.164 (1.000)	−0.200 (1.000)
E	β_0	−0.104 (0.990)	−0.048 (1.001)	−0.024 (0.991)	−0.091 (0.997)	−0.056 (0.977)	−0.059 (1.000)	−0.080 (0.997)	−0.040 (0.992)	−0.023 (1.003)
	β_1	−0.163 (0.994)	−0.012 (0.694)	0.108 (1.001)	−0.059 (0.992)	−0.060 (0.603)	−0.058 (0.989)	0.132 (1.006)	−0.031 (0.645)	−0.142 (0.996)
AR1	β_0	−0.105 (0.991)	−0.048 (0.996)	−0.024 (0.988)	−0.091 (0.997)	−0.056 (0.988)	−0.059 (1.003)	−0.079 (0.996)	−0.040 (1.016)	−0.023 (1.007)
	β_1	−0.117 (1.003)	−0.047 (0.826)	−0.093 (1.006)	−0.194 (1.006)	−0.134 (0.764)	−0.112 (0.998)	0.187 (1.013)	−0.106 (0.779)	−0.171 (1.006)
U	β_0	−0.127 (1.094)	−0.048 (1.054)	−0.012 (1.049)	−0.108 (1.079)	−0.056 (1.031)	−0.042 (1.031)	−0.093 (1.025)	−0.040 (1.039)	−0.007 (1.058)
	β_1	0.166 (1.075)	−0.002 (0.689)	−0.065 (1.056)	−0.135 (1.079)	−0.040 (0.637)	−0.156 (1.082)	−0.177 (1.111)	−0.081 (0.714)	−0.187 (1.116)

Table 3.3. Values of ARB (in percentage) and REF (in brackets) of $\beta_0(\tau, \mathbf{u})$ and $\beta_1(\tau, \mathbf{u})$ for $\alpha = 20\%$ of contamination over 1000 Monte Carlo simulations with low (Panel A) and high (Panel B) correlation.

(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 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 performance is likely to depend not only

Model	Coef	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
		$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$			$\mathcal{N}-20\%$		
		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	0.956	0.935	0.936
	β_1	0.941	0.957	0.933	0.960	0.938	0.957	0.942	0.957	0.950	0.970	0.933	0.957
AR1	β_0	0.949	0.940	0.932	0.948	0.938	0.944	0.961	0.937	0.931	0.959	0.943	0.937
	β_1	0.948	0.961	0.943	0.957	0.941	0.956	0.956	0.945	0.949	0.968	0.940	0.954
U	β_0	0.933	0.937	0.919	0.942	0.930	0.935	0.938	0.935	0.921	0.955	0.943	0.938
	β_1	0.929	0.948	0.917	0.941	0.927	0.940	0.915	0.941	0.931	0.960	0.922	0.954
		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	0.966	0.941	0.921
	β_1	0.945	0.948	0.939	0.961	0.942	0.949	0.954	0.964	0.961	0.971	0.939	0.958
AR1	β_0	0.941	0.948	0.934	0.948	0.949	0.941	0.955	0.951	0.913	0.966	0.944	0.923
	β_1	0.940	0.948	0.941	0.960	0.950	0.943	0.953	0.963	0.959	0.968	0.955	0.962
U	β_0	0.933	0.948	0.905	0.938	0.947	0.944	0.936	0.944	0.937	0.955	0.940	0.925
	β_1	0.921	0.935	0.921	0.940	0.926	0.935	0.925	0.952	0.941	0.963	0.924	0.953
		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	0.969	0.942	0.966
	β_1	0.949	0.945	0.951	0.960	0.952	0.947	0.949	0.964	0.957	0.958	0.938	0.964
AR1	β_0	0.947	0.957	0.944	0.949	0.953	0.953	0.961	0.942	0.919	0.970	0.944	0.965
	β_1	0.956	0.956	0.951	0.953	0.950	0.945	0.957	0.958	0.957	0.959	0.937	0.965
U	β_0	0.934	0.955	0.928	0.946	0.948	0.948	0.930	0.941	0.933	0.965	0.952	0.957
	β_1	0.919	0.937	0.931	0.947	0.932	0.929	0.925	0.954	0.937	0.948	0.931	0.958

Table 3.4. CP of $\beta_0(\tau, \mathbf{u})$ and $\beta_1(\tau, \mathbf{u})$ over 1000 Monte Carlo simulations under the four data generating scenarios with low correlation ($\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = 0.3$).

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 (3.3) 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 3.2 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

Model	Coef	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
		$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$			$\mathcal{N}-20\%$		
		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	0.958	0.956	0.976
	β_1	0.941	0.952	0.943	0.974	0.948	0.961	0.956	0.951	0.954	0.970	0.940	0.969
AR1	β_0	0.944	0.954	0.933	0.932	0.955	0.930	0.963	0.940	0.920	0.956	0.952	0.977
	β_1	0.937	0.949	0.937	0.970	0.958	0.958	0.954	0.945	0.954	0.967	0.939	0.969
U	β_0	0.929	0.951	0.912	0.938	0.950	0.928	0.949	0.945	0.891	0.954	0.940	0.975
	β_1	0.909	0.936	0.910	0.964	0.935	0.950	0.908	0.936	0.915	0.958	0.946	0.964
		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	0.950	0.937	0.928
	β_1	0.942	0.944	0.945	0.965	0.953	0.967	0.964	0.967	0.966	0.966	0.955	0.954
AR1	β_0	0.945	0.953	0.922	0.938	0.952	0.935	0.949	0.943	0.912	0.948	0.954	0.929
	β_1	0.933	0.944	0.935	0.966	0.954	0.964	0.969	0.958	0.962	0.967	0.952	0.947
U	β_0	0.928	0.952	0.915	0.947	0.949	0.939	0.949	0.943	0.908	0.945	0.942	0.938
	β_1	0.927	0.935	0.924	0.965	0.941	0.953	0.905	0.944	0.913	0.952	0.945	0.944
		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	0.959	0.953	0.964
	β_1	0.942	0.946	0.930	0.964	0.962	0.962	0.958	0.957	0.956	0.966	0.954	0.967
AR1	β_0	0.947	0.952	0.927	0.938	0.951	0.934	0.954	0.944	0.911	0.959	0.952	0.966
	β_1	0.932	0.930	0.934	0.963	0.958	0.957	0.952	0.958	0.957	0.969	0.944	0.970
U	β_0	0.927	0.953	0.910	0.943	0.946	0.938	0.949	0.949	0.935	0.962	0.944	0.959
	β_1	0.916	0.932	0.919	0.950	0.949	0.945	0.929	0.943	0.916	0.949	0.947	0.955

Table 3.5. CP of $\beta_0(\tau, \mathbf{u})$ and $\beta_1(\tau, \mathbf{u})$ over 1000 Monte Carlo simulations under the four data generating scenarios with high correlation ($\text{Cor}(Y_{ijk}, Y_{i'jk}) = \text{Cor}(Y_{ijk}, Y_{ijk'}) = 0.8$).

True model	$r = 0.3$									$r = 0.8$								
	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
	$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$			$\mathcal{N}$			$\mathcal{N}-10\%$			$\mathcal{T}$		
	Panel A: $\mathbf{u}_1$																	
E	660	887	647	440	817	454	660	870	637	751	839	725	625	930	604	645	868	675
AR1	683	916	672	432	789	424	585	881	554	882	983	900	465	867	473	802	981	756
	Panel B: $\mathbf{u}_2$																	
E	720	933	747	532	901	520	673	918	670	695	777	639	666	906	679	620	798	597
AR1	717	923	667	456	819	440	577	904	532	890	980	883	488	891	491	790	981	750
	Panel C: $\mathbf{u}_3$																	
E	660	887	647	440	817	454	660	870	637	699	858	715	588	920	630	631	859	665
AR1	704	914	689	423	817	449	586	907	535	898	981	886	499	882	482	786	981	765

Table 3.6. Number of correctly identified working correlation structure using the CIC, over 1000 Monte Carlo simulations under the three data generating scenarios.

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

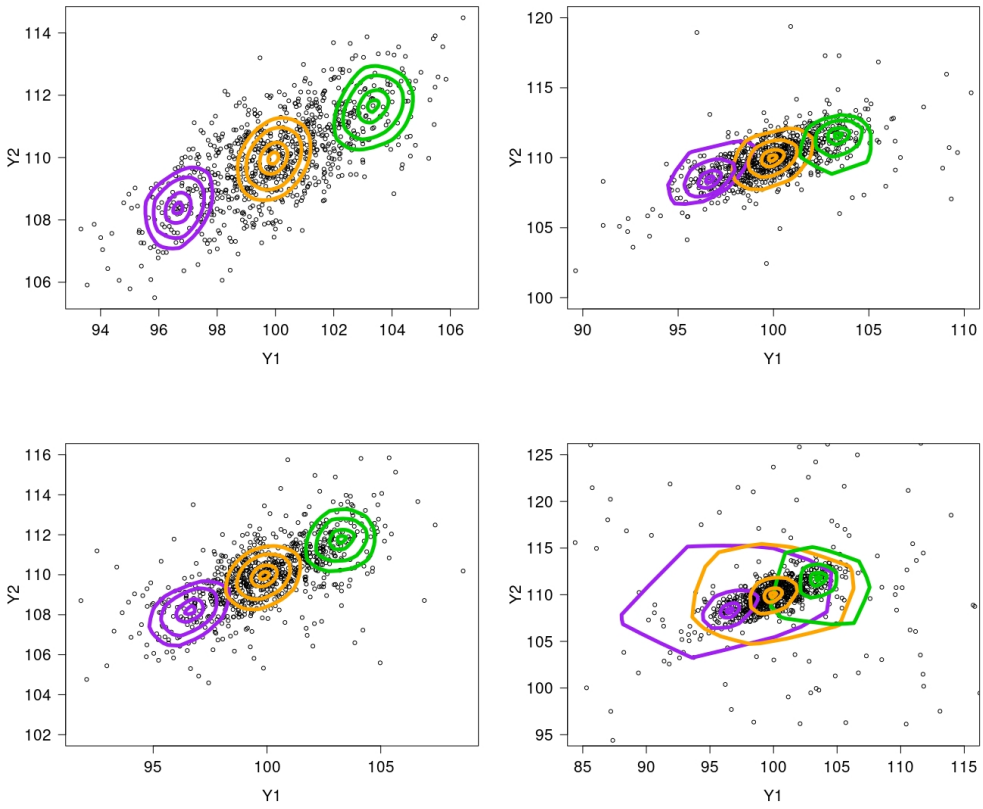

Figure 3.1. From left to right and top to bottom, estimated M-quantile contours under the $\mathcal{N}$, $\mathcal{N}-10\%$, $\mathcal{T}$ and $\mathcal{N}-20\%$ simulation scenarios at level $\tau = (0.05, 0.1, 0.25, 0.4)$ (from the outside inwards), conditional on the 0.05-th (violet), 0.5-th (orange) and 0.95-th (green) empirical quantiles of $X_{ij}^{(1)}$.

together as shown in (3.13). Sections 3.6.1 and 3.6.2 report the results, respectively.

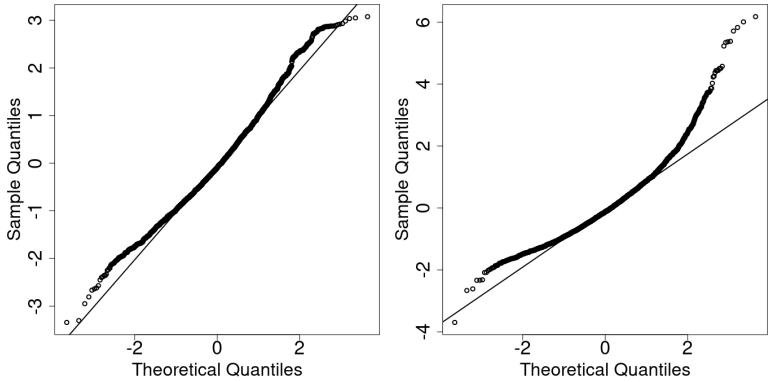

Figure 3.2. Normal probability plots residuals from a Marginal Mean model under an exchangeable correlation structure for mathematics (left) and reading (right) scores.

3.6.1 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. In this context, the $\mathbf{u} = (u_1, u_2)$ directions, where u_1 is the mathematics score dimension and u_2 is the reading score dimension, 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, allowing us to rank students on the basis of their scores on achievement tests in mathematics and reading, respectively. On the other hand, $\mathbf{u}_2$ is equivalent to choosing the arithmetic mean of the two scores. In this case, the response variable is given by the average test score obtained by each student.

Since a pupil's performance is likely to depend not only on its abilities, but also on the characteristics of the school, to handle dependence between pupils nested in the same school, we estimate the MMQ model using three candidate working correlation structures (I, E and AR1), and then select the optimal one corresponding to the lowest CIC value in (3.29).

Table 3.7 shows point estimates of the regression coefficients and of the correlation parameter and the CIC values associated to the selected MMQ models at the investigated quantile levels. Statistical significance of regression coefficients is assessed by computing asymptotic standard errors as described in Section 3.4.1. 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 using the working correlation structure selected by the CIC. 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 3.8 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 3.9 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.

We start by commenting on the MMQ results in Table 3.7. As one can see, the CIC presented in Section 3.4.2 leads us to select the exchangeable correlation structure over the independence and the AR1 structures. This result is consistent with the data structure as the academic performance of pupils enrolled in the same school will likely be influenced by the learning environment and characteristics of the school.

By looking at the parameter estimates, we firstly 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).

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)
	CIC	8.492	2.591	4.613	7.155	8.139	7.698
$\mathbf{u}_2$	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)
	CIC	8.501	3.001	4.830	7.368	8.602	7.934
$\mathbf{u}_3$	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)
	CIC	8.212	2.685	4.652	8.178	11.545	10.016

Table 3.7. MM and MMQ model parameter estimates at the investigated quantile levels. Boldface denote statistical significance at the 5% level.

$\mathbf{u}$	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 3.8. MQRE model parameter estimates and ICC values at the investigated quantile levels. Boldface denote statistical significance at the 5% level.

$\mathbf{u}$	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 3.9. 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.

3.6.2 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 (3.13). Figure 3.3 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 nonparametric bootstrap method of Section 3.3.1 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 (see Figure 3.4).

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 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 3.4 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.

3.7 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 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

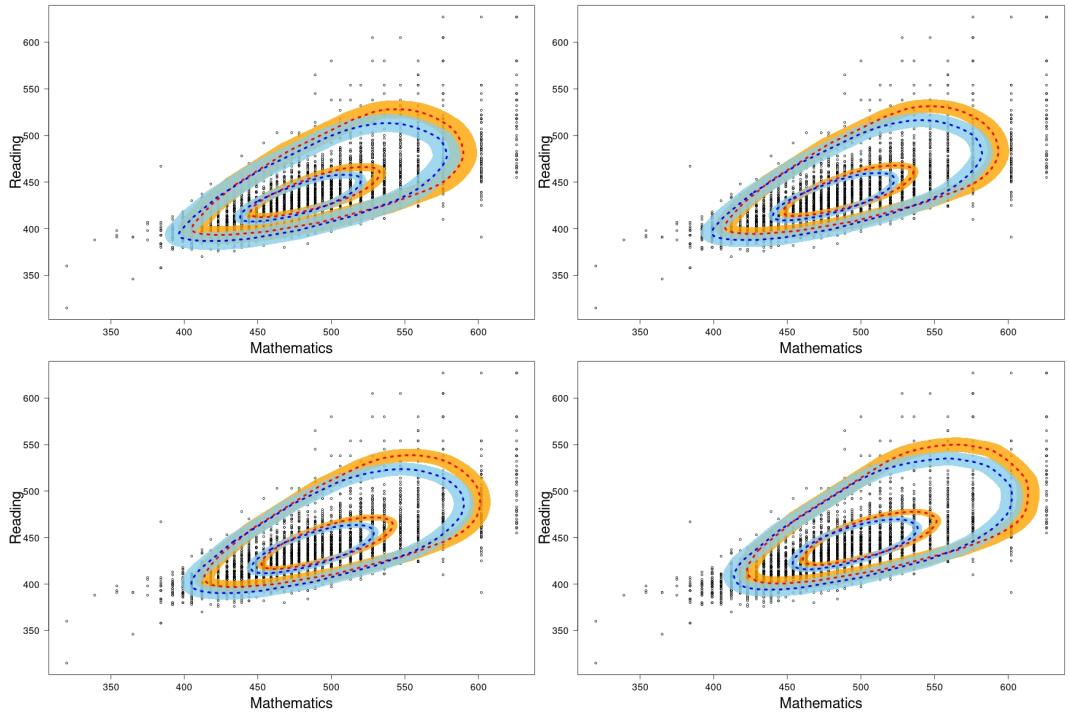

Figure 3.3. Estimated M-quantile contours at $\tau = (0.005, 0.1)$ for small (red) and large (blue) classes, conditional on the 0.01-th (top-left), 0.25-th (top-right), 0.75-th (bottom-left) and 0.99-th (bottom-right) empirical quantiles of years of teaching experience. The shaded surfaces represent 95% confidence envelopes for M-quantile contours obtained using nonparametric bootstrap.

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 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.

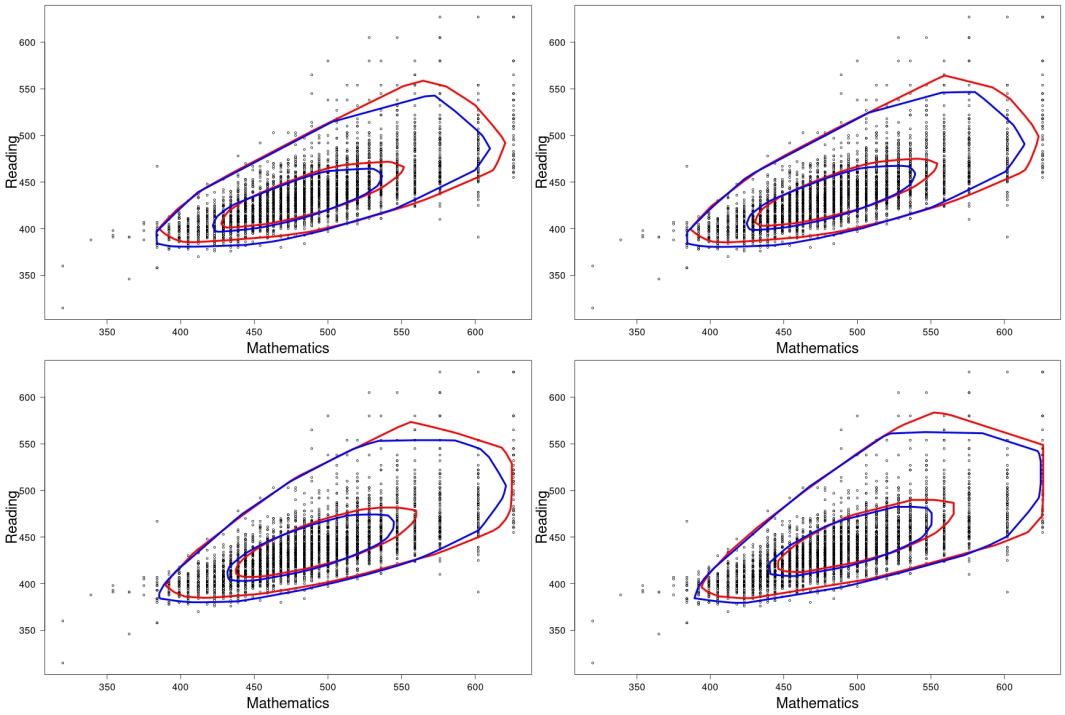

Figure 3.4. Estimated quantile contours at $\tau = (0.005, 0.1)$ for small (red) and large (blue) classes, conditional on the 0.01-th (top-left), 0.25-th (top-right), 0.75-th (bottom-left) and 0.99-th (bottom-right) empirical quantiles of years of teaching experience.

Chapter 4

Unified unconditional regression for multivariate quantiles, M-quantiles and expectiles

4.1 Introduction

When researchers wish to determine the effect of relevant predictors across the entire distribution of the dependent variable of interest, Quantile Regression (QR), as introduced by Koenker & Bassett (1978), plays a crucial role in providing a much more complete statistical analysis compared to the classical mean regression. Indeed, it allows to model conditional quantiles of a response as a function of explanatory variables and it has been greatly exploited for the study of non-Gaussian, heavy-tailed and highly skewed data. For a detailed survey and list of references of the most used QR techniques, please refer to Koenker (2005) and Koenker et al. (2017).

However, if one is interested in how the whole unconditional distribution of the outcome responds to changes in the covariates, QR methods would yield misleading inferences (see Firpo et al. 2009, Borah & Basu 2013, Maclean et al. 2014, Killewald & Bearak 2014). As explained by Frölich & Melly (2013) in a simple example relating wages to years of education, the unconditional 90-th quantile refers to the high wage workers, whereas the 90-th quantile conditional on education refers to the high wage workers within each education class, who however may not necessarily be high earners overall. Presuming a strong positive correlation between education and wages, it may well be that the 90-th quantile among high school dropouts is lower than, say, the median of all Ph.D. graduates. The interpretation of the 90-th quantile is thus different for conditional and unconditional quantiles. From a policy perspective, while the welfare of highly educated people with relatively low wages catches little interest, the welfare of the poor, i.e., those located in the lower end of the unconditional distribution of wages, attracts a lot of attention in the political debate. There are, indeed, numerous applications of practical relevance where the ultimate research objective is the unconditional distribution of the dependent variable, as in the context of the earnings disparities between different groups of workers, the effect of education on earnings, or the distributional impacts of a particular treatment in a given population (Borah & Basu 2013, Frölich & Melly 2013, Huffman et al. 2017,

Firpo et al. 2018). Therefore, a number of proposals has been introduced in the literature to estimate these unconditional effects (Gosling et al. 2000, Machado & Mata 2005, Melly 2005).

Motivated by this interest, Firpo et al. (2009) proposed the Unconditional Quantile Regression (UQR) approach for modeling unconditional quantiles of a dependent variable as a function of the explanatory variables. This method builds upon the concept of Recentered Influence Function (RIF) which originates from a widely used tool in robust statistics, namely the Influence Function (IF, see Hampel 1974, Huber & Ronchetti 2009). In particular, the UQR of Firpo et al. (2009) consists of regressing the RIF of the unconditional quantile of the outcome variable on the explanatory variables using either Ordinary Least Squares (OLS), logistic regression or the nonparametric regression of Newey (1994). Their approach represents an important contribution on quantile regression methods and its validity is also demonstrated by the growing scientific literature spanning from medicine (Borah & Basu 2013), economics (Murakami & Seya 2019, Dong et al. 2020), social inequalities (Rodriguez-Caro et al. 2016, Huffman et al. 2017) and agriculture (Mishra et al. 2015, Bonanno et al. 2018).

These studies, however, focus on a univariate regression framework. When the problem under investigation involves multivariate dependent variables, the method of Firpo et al. (2009) cannot be easily extended to higher dimensions due to the non existence of a natural ordering in a p-dimensional space, $p > 1$ (see Serfling 2002, Kong & Mizera 2012, Koenker et al. 2017, Merlo et al. 2022).

With this paper, we contribute to the current literature extending the univariate UQR approach of Firpo et al. (2009) to a more general multivariate setting. In particular, we propose to employ the multidimensional Huber's function defined in Hampel et al. (2011) to build a unified unconditional regression approach that encompasses multivariate quantiles, M-quantiles (Breckling & Chambers 1988) and expectiles (Newey & Powell 1987).

In the statistical literature, the Huber's function has been employed to define the M-quantile for robust modeling of the entire distribution of univariate response variables, extending the ideas of M-estimation of Huber (1964) and Huber & Ronchetti (2009). This method provides a "quantile-like" generalization of the mean based on influence functions that combines in a common framework the robustness and efficiency properties of quantiles and expectiles, depending on the choice of the Huber's tuning constant. In the multivariate framework, the multidimensional Huber's function (Hampel et al. 2011) has been exploited by Breckling & Chambers (1988) to define the multivariate M-quantile using a direction vector in the Euclidean p-dimensional space in order to establish a suitable ordering procedure for multivariate observations. Subsequently, Kokic et al. (2002) generalized their definition by introducing a class of multivariate M-quantiles based on weighted estimating equations, which includes multivariate quantiles and expectiles depending on the value of the tuning constant.

In our paper, we rely on the Kokic et al. (2002) approach using the multidimensional Huber's function to model unconditional quantiles, M-quantiles and expectiles of multivariate response variables in a unified regression framework, by choosing the tuning constant appropriately.

In order to analyze the impact of changes in the distribution of explanatory variables on the entire unconditional distribution of the responses, following Firpo et al. (2009),

we regress the RIF of the proposed model on the covariates, producing the Unconditional Quantile, M-Quantile and Expectile Partial Effect (UQPE, UMQPE, UEPE) according to the selected value of the tuning constant. From the theoretical point of view, we establish the asymptotic properties of the corresponding estimators using the Bahadur representation (Bahadur 1966). Furthermore, we propose a data-driven method based on cross-validation for selecting the optimal tuning constant that accounts for possible outliers in the data.

Using simulation studies, we illustrate the finite sample properties of the proposed methodology under different data generating processes. From an empirical standpoint, we demonstrate the usefulness of this method through the analysis of the Survey on Household Income and Wealth (SHIW) 2016 conducted by the Bank of Italy. In particular, we fit the proposed model to evaluate the effect of economic and socio-demographic characteristics of Italian households on the unconditional distributions of family wealth and consumption, accounting both for the correlation between the outcomes and influential observations in the sample. The proposed multivariate approach allows us to consider consumption and wealth as part of a collective framework and it can be of great interest to investigate whether the effect of the covariates is more pronounced on low-quantiles (low-consumption and low-wealth families) than on high-quantiles (high-level spending and wealthy households) of the responses' unconditional distribution.

The remainder of the paper is organized as follows. In Section 4.2, we revise the RIF and its properties. Section 4.3 introduces the proposed unconditional regression model for multivariate response variables and provides a detailed discussion of the asymptotic properties of the introduced estimators. Section 4.4 discusses the simulation study and the results. Finally, the empirical application is presented in Section 4.5 while Section 4.6 concludes.

4.2 Notation and preliminary results

In this section, we present the main notation and concepts which we use throughout the paper. Specifically, we review the notion of Recentered Influence Function (RIF) which originates from the Influence Function (IF) of Hampel (1974) and the Unconditional Partial Effect (UPE) introduced by Firpo et al. (2009) that leads us to analyze the impact of changes in the covariates on the unconditional distribution of the response variable.

Let $\mathbf{Y}$ denote a random variable belonging to an arbitrary sample space $\mathcal{Y}$ with absolutely continuous distribution function $F_{\mathbf{Y}}$ and consider a vector-valued functional $\nu(F_{\mathbf{Y}})$ where $\nu : \mathcal{F}_\nu \to \mathbb{R}^p$, such that $\mathcal{F}_\nu$ is the collection of all distributions on $\mathcal{Y}$ for which ν is defined.

The IF allows to study the effect of an infinitesimal contamination in the underlying distribution $F_{\mathbf{Y}}$ at a point $\mathbf{y}$ on the statistic $\nu(F_{\mathbf{Y}})$ we are interested in. Let us consider $\Delta_{\mathbf{y}}$ the probability measure that puts mass 1 at the value $\mathbf{y}$ and let $F_{\mathbf{Y},t\Delta_{\mathbf{y}}} = (1-t)F_{\mathbf{Y}} + t\Delta_{\mathbf{y}}$ represent the mixing distribution with $t \in [0,1]$.

Following Hampel et al. (2011), the p-dimensional IF of $\nu(F_{\mathbf{Y}})$ is defined as:

$$IF(\mathbf{y};\nu) = \lim_{t\to 0} \frac{\nu(F_{\mathbf{Y},t\Delta_{\mathbf{y}}}) - \nu(F_{\mathbf{Y}})}{t}. \tag{4.1}$$

Using the definition of IF in (4.1), Firpo et al. (2009) considered the RIF to analyze the statistic $\nu(F_{\mathbf{Y}})$ after a perturbation of $F_{\mathbf{Y}}$ in the direction of $\Delta_{\mathbf{y}}$. In particular, the RIF is defined as the first two terms of the von Mises linear approximation (Mises 1947) of the corresponding statistic $\nu(F_{\mathbf{Y},t\Delta_{\mathbf{y}}})$ with $t = 1$, namely:

$$RIF(\mathbf{y};\nu) = \nu(F_{\mathbf{Y}}) + \int IF(\mathbf{s};\nu)d\Delta_{\mathbf{y}}(\mathbf{s}) = \nu(F_{\mathbf{Y}}) + IF(\mathbf{y};\nu). \tag{4.2}$$

In the presence of a set of covariates $\mathbf{X} \in \mathcal{X}$, with $\mathcal{X} \subset \mathbb{R}^k$ being the support of $\mathbf{X}$, Firpo et al. (2009) suggested the use of the RIF in (4.2) for analyzing the impact on $\nu(F_{\mathbf{Y}})$ due to changes in the distribution of $\mathbf{X}$. In particular, in order to incorporate the effect of the explanatory variables, by the law of iterated expectations and the fact that the IF in (4.1) integrates up to zero, it follows from (4.2) that:

$$\nu(F_{\mathbf{Y}}) = \int RIF(\mathbf{y};\nu)dF_{\mathbf{Y}}(\mathbf{y}) = \int \mathbb{E}[RIF(\mathbf{Y};\nu) \mid \mathbf{X} = \mathbf{x}]dF_{\mathbf{X}}(\mathbf{x}), \tag{4.3}$$

where $F_{\mathbf{X}}$ is the distribution function of $\mathbf{X}$.
From (4.3) it can be seen that when one is interested in the impact of a change in the covariates $\mathbf{X}$ on a specific distributional statistic $\nu(F_{\mathbf{Y}})$, the $\mathbb{E}[RIF(\mathbf{Y};\nu) \mid \mathbf{X} = \mathbf{x}]$ can be modeled as a function of $\mathbf{X}$, which can be easily implemented using regression methods for the conditional mean (see Firpo et al. 2009, 2018 and Rios-Avila 2020). More formally, under the assumption that the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$, $F_{\mathbf{Y}|\mathbf{X}}$, is unaffected by changes in the law of $\mathbf{X}$, the partial effects of a small location shift in the distribution of $\mathbf{X}$ is given by:

$$\boldsymbol{\alpha}(\nu) = \int \frac{d\mathbb{E}[RIF(\mathbf{Y};\nu) \mid \mathbf{X} = \mathbf{x}]}{d\mathbf{x}} dF_{\mathbf{X}}(\mathbf{x}), \tag{4.4}$$

where $d\mathbb{E}[RIF(\mathbf{Y};\nu) \mid \mathbf{X} = \mathbf{x}]/d\mathbf{x}$ is understood to indicate the Jacobian matrix of all its first-order partial derivatives with respect to $[\mathbf{x}_j]_{j=1}^k$. Firpo et al. (2009) call the quantity $\boldsymbol{\alpha}(\nu)$ in (4.4) as the Unconditional Partial Effect (UPE).
In the case of a dummy variable $X \in \{0, 1\}$, the UPE represents the effect of a small increase in the probability that $X = 1$, namely:

$$\boldsymbol{\alpha}(\nu) = \mathbb{E}[RIF(\mathbf{Y};\nu) \mid X = 1] - \mathbb{E}[RIF(\mathbf{Y};\nu) \mid X = 0]. \tag{4.5}$$

As stated in Firpo et al. (2009), the approach based on the RIF is applicable to a wide range of distributional statistics to capture how they are affected by changes in the distribution of $\mathbf{X}$. In the following section we exploit these properties for the analysis of unconditional quantile, M-quantile and expectile regressions associated to multivariate response variables.

4.3 Methodology

In this section, we generalize the univariate approach of Firpo et al. (2009) by developing a unifying regression method to model unconditional quantiles, M-quantiles and expectiles of vector-valued responses using the RIF approach illustrated in Section 4.2. Firstly, we introduce the Huber's multidimensional M-function for estimating multivariate quantiles, M-quantiles and expectiles by varying the value of the tuning constant in an appropriate way. Then, in order to assess the effect of the covariates at different parts of the unconditional distribution of the dependent variable, we introduce the Unconditional Quantile, M-Quantile and Expectile Partial Effect (UQPE, UMQPE, UEPE) and establish the asymptotic properties of the corresponding estimators.

The Huber's function, in the univariate case, has been used by Breckling & Chambers (1988) to define the M-quantile, extending the concept of M-estimation of Huber (1964) to the quantile framework. Huber M-quantiles are a generalized form of M-estimators that include in a single modeling approach, quantiles and expectiles through a tuning constant to adjust the robustness of the estimator in the presence of outliers. In higher dimensions, extending these univariate notions to multivariate data is not a trivial task since there does not exist a natural ordering in p dimensions, $p > 1$. Already in their proposal, Breckling & Chambers (1988) considered the multivariate extension of Huber's function to estimate M-quantiles by considering a directional unit norm vector to set up a suitable ordering procedure for multidimensional data. Further, Kokic et al. (2002) generalized their approach by introducing a weighted estimating equation based on the multidimensional Huber's influence function that encompasses multivariate quantiles, M-quantiles and expectiles, depending on the value of the related tuning constant.
Here and in what follows, in order to present a unified unconditional regression approach to model multivariate quantiles, M-quantiles and expectiles, we adopt the approach of Kokic et al. (2002) based on the multidimensional Huber's function.

Formally, the multidimensional Huber's influence function in Hampel et al. (2011) is defined as:

$$\Psi(\mathbf{r}) = \begin{cases} \frac{\mathbf{r}}{c}, & \|\mathbf{r}\| < c \\ \frac{\mathbf{r}}{\|\mathbf{r}\|}, & \|\mathbf{r}\| \geq c \end{cases}, \qquad \mathbf{r} \in \mathbb{R}^p, \tag{4.6}$$

where $c \geq 0$ is the tuning constant that can be adjusted to trade robustness for efficiency, with increasing robustness when it is chosen to be close to 0 and increasing efficiency when it is chosen to be large. To show how one can use the $\Psi(\mathbf{r})$ function in (4.6) to estimate multivariate quantiles, M-quantiles and expectiles, we introduce the following additional notation. Let $\mathcal{Y} = \mathbb{R}^p$, consider a continuous p-dimensional random variable $\mathbf{Y}$ and let $\mathbf{u}$ denote a unit norm direction vector ranging over the p-dimensional unit sphere $\mathcal{S}^{p-1} = \{\mathbf{z} \in \mathbb{R}^p : ||\mathbf{z}|| = 1\}$, where $||\cdot||$ denotes the Euclidean norm. Following Kokic et al. (2002), for a general value of c, we obtain the τ-th multivariate M-quantile of $\mathbf{Y}$ in the direction of $\mathbf{u}$, $\boldsymbol{\theta}_{\tau,\mathbf{u}}$, with $\tau \in (0, \frac{1}{2}]$, by satisfying the equation:

$$\int \eta_\delta(\varphi)\Psi(\mathbf{y} - \boldsymbol{\theta}_{\tau,\mathbf{u}})dF_{\mathbf{Y}}(\mathbf{y}) = \mathbf{0}, \tag{4.7}$$

where

$$\eta_\delta(\varphi) = \begin{cases} (1-\cos\varphi)^\delta\zeta + 2\tau, & \varphi \in (-\frac{\pi}{2}, \frac{\pi}{2}) \\ -(1-\cos\varphi)^\delta\zeta + 2(1-\tau), & \varphi \in [-\pi, -\frac{\pi}{2}] \cup [\frac{\pi}{2}, \pi], \end{cases}$$

is a weighting function with $\zeta = 1-2\tau$, $\delta > 0$ and φ being the angle between $\mathbf{Y}-\boldsymbol{\theta}$ and $\mathbf{u}$, so $\cos\varphi = \frac{(\mathbf{Y}-\boldsymbol{\theta})'\mathbf{u}}{\|\mathbf{Y}-\boldsymbol{\theta}\|}$.

The function $\eta_\delta(\varphi)$ gives asymmetric weights to the residual $\mathbf{Y}-\boldsymbol{\theta}$ depending both on its length and the angle it forms with $\mathbf{u}$. Most importantly, the tuning constant c determines where the weighted scheme based on $\eta_\delta(\varphi)$ defines multivariate quantiles when $c = 0$ and yields multivariate expectiles as $c \to \infty$, which could be particularly fruitful when the use of outlier-robust estimation methods is not justified but there is still interest in modeling the entire distribution of $\mathbf{Y}$ (Tzavidis et al. 2010). Computationally, an estimate of $\boldsymbol{\theta}_{\tau,\mathbf{u}}$ in (4.7) can be efficiently obtained by using Iteratively Reweighted Least Squares (IRLS, Breckling et al. 2001). Clearly, the multivariate M-quantile in (4.7) includes the traditional notion of univariate M-quantile. Indeed, if $p = 1$ and $u = 1$, then (4.7) reduces to the estimating equation of the univariate M-quantile, θ_τ, because $\cos\varphi = \text{sgn}(Y-\theta_\tau)$, which implies that $\eta_\delta(\varphi) = 1-\zeta\text{sgn}(Y-\theta_\tau)$.

A particular issue in this context may be the choice of the direction $\mathbf{u}$, which is often selected on the basis of the empirical problem at hand to produce meaningful results (see Paindaveine & Šiman 2011, Kong & Mizera 2012, Geraci et al. 2020 and Farcomeni et al. 2020). One option is also to use the principal component of a Principal Component Analysis by maximizing the variance of the projected data $\mathbf{u}'\mathbf{Y}$ as proposed in Korhonen & Siljamäki (1998) and Geraci et al. (2020), or to consider the direction by minimizing a measure of skewness.

When the directional approach is adopted, considering theoretically all directions in $\mathcal{S}^{p-1}$ simultaneously yields multivariate M-quantiles centrality regions, which allow us to provide a visual description of the location, spread, shape and dependence between the responses distribution. These quantities are of crucial interest as they are able to adapt to the underlying shape of the distribution of $\mathbf{Y}$ without being constrained to particular shapes, such as convex bodies or ellipses (Breckling et al. 2001). Specifically, the τ-th M-quantile region, $R_\tau \subset \mathbb{R}^p$, is defined as the set whose vertices are:

$$R_\tau = \{\boldsymbol{\theta}_{\tau,\mathbf{u}} \,|\, \mathbf{u} \in \mathcal{S}^{p-1}\}. \tag{4.8}$$

The region defined in (4.8) is a closed surface and the corresponding M-quantile contour of order τ is defined as the boundary ∂R_τ of R_τ. For fixed τ, when $c = 0$ it defines quantile contours and it generates expectile contours when $c \to \infty$. Meanwhile, for any $c \geq 0$, the contours are nested as τ increases. As $\tau \to 0$, instead, the τ-th M-quantile contour approaches the convex hull of the sample data providing valuable information about the extent of extremality of points (see Serfling 2002 and Kokic et al. 2002).

To build our model, it follows from (4.7) and Hampel et al. (2011) that the IF for the unconditional multivariate M-quantile $\boldsymbol{\theta}_{\tau,\mathbf{u}}$ is defined as:

$$IF(\mathbf{y}; \boldsymbol{\theta}_{\tau,\mathbf{u}}) = \mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1}\eta_\delta(\varphi)\Psi(\mathbf{y}-\boldsymbol{\theta}_{\tau,\mathbf{u}}) \tag{4.9}$$

with $\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$ being the $p \times p$ matrix given by:

$$\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = -\int \nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}}(\eta_\delta(\varphi)\Psi(\mathbf{y}-\boldsymbol{\theta}))dF_{\mathbf{Y}}(\mathbf{y}), \tag{4.10}$$

where $\nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}}(\cdot)$ is the $p \times p$ matrix of first order derivatives of $\eta_\delta(\varphi)\Psi(\mathbf{y}-\boldsymbol{\theta})$ in (4.7) with respect to $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}_{\tau,\mathbf{u}}$. Then, following the idea in (4.2), the RIF is obtained from (4.9) by adding back the multivariate M-quantile $\boldsymbol{\theta}_{\tau,\mathbf{u}}$:

$$RIF(\mathbf{y};\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \boldsymbol{\theta}_{\tau,\mathbf{u}} + IF(\mathbf{y};\boldsymbol{\theta}_{\tau,\mathbf{u}}). \tag{4.11}$$

Two remarks are worth noticing. Firstly, (4.11) generalizes the RIF of the multivariate quantile when $c = 0$, where the matrix of first order derivatives of $\Psi(\mathbf{r})$ is equal to:

$$\frac{d\Psi(\mathbf{r})}{d\mathbf{r}} = \frac{1}{\|\mathbf{r}\|}\Big\{\mathbf{I}_p - \frac{\mathbf{r}\mathbf{r}'}{\|\mathbf{r}\|^2}\Big\}, \tag{4.12}$$

with $\mathbf{I}_p$ being the identity matrix of dimension p, and it coincides with the RIF of the multivariate expectile when $c \to \infty$, which implies $\frac{d\Psi(\mathbf{r})}{d\mathbf{r}} \propto \mathbf{I}_p$. Secondly, when $p = 1$ and $u = 1$, (4.11) reduces to the univariate RIF of standard M-quantiles which, in turn, includes the RIF of the quantile in Firpo et al. (2009) and the RIF of the expectile for c arbitrarily large. Further, using this approach we are able to investigate the correlation structure of multivariate responses at different values of τ. More in detail, to study the association between multiple outcomes we analyze the covariance matrix of the RIF in (4.11) which, by simple calculations, can be written as:

$$\boldsymbol{\Delta}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \mathbb{E}[IF(\mathbf{Y};\boldsymbol{\theta}_{\tau,\mathbf{u}})IF(\mathbf{Y};\boldsymbol{\theta}_{\tau,\mathbf{u}})']. \tag{4.13}$$

Given $\mathbf{u}$, τ and c, the off-diagonal elements of $\boldsymbol{\Delta}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$ provide a measure of tail correlation between the components of $\mathbf{Y}$.

In a regression framework where covariates $\mathbf{X}$ are available, from (4.11) we define the unified unconditional regression model as follows:

$$\mathbb{E}[RIF(\mathbf{Y};\boldsymbol{\theta}_{\tau,\mathbf{u}}) \mid \mathbf{X} = \mathbf{x}] = \boldsymbol{\theta}_{\tau,\mathbf{u}} + \mathbb{E}[IF(\mathbf{Y};\boldsymbol{\theta}_{\tau,\mathbf{u}}) \mid \mathbf{X} = \mathbf{x}]. \tag{4.14}$$

Our objective is to identify how changes in the distribution of $\mathbf{X}$ affect the multivariate quantile, M-quantile and expectile of the unconditional distribution of $\mathbf{Y}$. Following (4.4), for a given level τ, direction $\mathbf{u}$, and constant $c \geq 0$, the Unconditional M-Quantile Partial Effect (UMQPE), $\boldsymbol{\alpha}_{\tau,\mathbf{u}}$, is formally defined as:

$$\begin{aligned}\boldsymbol{\alpha}_{\tau,\mathbf{u}} &= \int \frac{d\mathbb{E}[RIF(\mathbf{Y};\boldsymbol{\theta}_{\tau,\mathbf{u}}) \mid \mathbf{X} = \mathbf{x}]}{d\mathbf{x}} dF_{\mathbf{X}}(\mathbf{x}) \\ &= \mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1} \int \frac{d\mathbb{E}[\eta_\delta(\varphi)\Psi(\mathbf{Y}-\boldsymbol{\theta}_{\tau,\mathbf{u}}) \mid \mathbf{X} = \mathbf{x}]}{d\mathbf{x}} dF_{\mathbf{X}}(\mathbf{x}).\end{aligned} \tag{4.15}$$

It is worth noting that the proposed approach has several appealing properties. Firstly, the UMQPE in (4.15) is easy to compute as it does not depend on the density of $\mathbf{Y}$ which would entail the use of nonparametric density estimation procedures (see Kokic et al. 2002 and Scott 2015). Secondly, this methodology allows us to

directly control the robustness to outliers and estimation efficiency by means of the tuning constant c. In particular, when $c = 0$ we have the UQPE and when $c \to \infty$ we have the UEPE.

Before concluding this section, it is relevant to compare the proposed unconditional regression model with the standard conditional regression approach. Suppose that $\mathbf{Y} = h(\mathbf{X}) + \boldsymbol{\epsilon}$, where $h : \mathcal{X} \to \mathbb{R}^p$ is an unknown function with bounded first partial derivatives and $\boldsymbol{\epsilon}$ is a p-dimensional random error independent of $\mathbf{X}$. For a given τ, $\mathbf{u}$ and $c \geq 0$, we denote the τ-th multivariate M-quantile of $\mathbf{Y}$ conditional on $\mathbf{X} = \mathbf{x}$, $\boldsymbol{\theta}_{\tau,\mathbf{u}}(\mathbf{x}) = h(\mathbf{x})$, as the solution of the following estimating equation:

$$\int \eta_\delta(\varphi)\Psi(\mathbf{y} - \boldsymbol{\theta}_{\tau,\mathbf{u}}(\mathbf{x}))dF_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) = \mathbf{0}. \tag{4.16}$$

Consequently, the effect of a small change in $\mathbf{X}$ on the conditional M-quantile of $\mathbf{Y}$, $\boldsymbol{\theta}_{\tau,\mathbf{u}}(\mathbf{x})$, which we denote as Conditional M-Quantile Partial Effect (CMQPE), is given by:

$$\boldsymbol{\alpha}_{\tau,\mathbf{u}}(\mathbf{x}) = \frac{dh(\mathbf{x})}{d\mathbf{x}}. \tag{4.17}$$

In order to clarify the interpretation of (4.15), following Firpo et al. (2009), we provide a useful representation of the UMQPE in terms of the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$ and show how it is related to the CMQPE in (4.17). Let $\mathbf{W}_{\tau,\mathbf{u}} : \mathcal{X} \to \mathbb{R}^{p\times p}$ define the weighting matrix function:

$$\mathbf{W}_{\tau,\mathbf{u}}(\mathbf{x}) = \mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1}\mathbb{E}[\nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}}(\eta_\delta(\varphi)\Psi(\mathbf{Y} - \boldsymbol{\theta}_{\tau,\mathbf{u}})) \mid \mathbf{X} = \mathbf{x}] \tag{4.18}$$

and let $s_{\tau,\mathbf{u}}$ be the function $s_{\tau,\mathbf{u}} : \mathcal{X} \to (0, \frac{1}{2}]$ which can be thought as a "matching" function that maps each conditional multivariate M-quantile onto the unconditional multivariate M-quantile of $\mathbf{Y}$, i.e.:

$$s_{\tau,\mathbf{u}}(\mathbf{x}) = \{\tilde{\tau} : \boldsymbol{\theta}_{\tilde{\tau},\mathbf{u}}(\mathbf{x}) = \boldsymbol{\theta}_{\tau,\mathbf{u}}\}. \tag{4.19}$$

The next Theorem establishes a link between the UMQPE and the CMQPE.

Theorem 3. *Assume that $\mathbf{Y} = h(\mathbf{X}) + \boldsymbol{\epsilon}$ where $h(\cdot)$ is an unknown function with bounded first partial derivatives and $\boldsymbol{\epsilon}$ is an error term independent of $\mathbf{X}$. For a given $\tau \in (0, \frac{1}{2}]$, $\mathbf{u} \in \mathcal{S}^{p-1}$ and $c \geq 0$, the UMQPE $\boldsymbol{\alpha}_{\tau,\mathbf{u}}$ can be written as:*

$$\boldsymbol{\alpha}_{\tau,\mathbf{u}} = \mathbb{E}[\mathbf{W}_{\tau,\mathbf{u}}(\mathbf{X})\, \boldsymbol{\alpha}_{s_{\tau,\mathbf{u}}(\mathbf{X}),\mathbf{u}}(\mathbf{X})], \tag{4.20}$$

where the expectation is taken over the distribution of $\mathbf{X}$.

Proof. See Proof of Theorem 3 in Appendix. □

From Theorem 3 there follow several interesting considerations. Firstly, it formally shows that, unlike conditional means which average up to the unconditional mean thanks to the law of iterated expectations, conditional multivariate M-quantiles do not average up to their unconditional counterparts. On the contrary, the UMQPE is equal to a weighted average, over the distribution of the covariates, of the CMQPE at

the $s_{\tau,\mathbf{u}}(\mathbf{X})$-th conditional M-quantile level corresponding to the τ-th unconditional M-quantile of $\mathbf{Y}$ in the direction $\mathbf{u}$. Secondly, it generalizes the result in Firpo et al. (2009) to the multivariate setting which also holds in the univariate case when $p = 1$ and $u = 1$. Furthermore, Theorem 3 is useful for interpreting the parameters of the proposed regression method. For instance, in a linear model where $h(\mathbf{X}) = \boldsymbol{\beta}'\mathbf{X}$, the UMQPE and CMQPE are both equal to the matrix of regression coefficients $\boldsymbol{\beta}$ for any choice of τ and $\mathbf{u}$. More generally, the UMQPE and CMQPE will be different depending on the structural form of $h(\cdot)$ and the distribution of $\mathbf{X}$ as described in Theorem 3.

4.3.1 Estimation

In this section, we discuss the estimation of the UQPE, UMQPE and UEPE using the RIF regression approach. Following Firpo et al. (2009), for a given τ, direction $\mathbf{u}$ and $c \geq 0$, we estimate the $\boldsymbol{\alpha}_{\tau,\mathbf{u}}$ in (4.15) via an OLS regression of the $RIF(\mathbf{Y}; \boldsymbol{\theta}_{\tau,\mathbf{u}})$ as a dependent variable onto the covariates $\mathbf{X}$ by using a two-step procedure. Specifically, an estimate $\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}$ of $\boldsymbol{\theta}_{\tau,\mathbf{u}}$ is obtained by solving (4.7) via IRLS, substitute $\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}$ in (4.11) and then estimate $\boldsymbol{\alpha}_{\tau,\mathbf{u}}$ by regressing the $RIF(\mathbf{Y}; \widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}})$ on $\mathbf{X}$. Let $(\mathbf{Y}_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, denote a random sample of size n, the estimator of the UMQPE in (4.15), $\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}$, is defined as follows:

$$\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}} = \widehat{\Omega}_{\mathbf{X}}^{-1} \frac{1}{n} \sum_{i=1}^{n} \{\mathbf{X}_i RIF'(\mathbf{Y}_i; \widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}})\}. \tag{4.21}$$

where $\widehat{\Omega}_{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i'$. Similarly, the estimators of the UQPE and UEPE related to (4.15) can be obtained following the same procedure by setting $c = 0$ and c large enough such that $||\mathbf{Y}_i - \boldsymbol{\theta}_{\tau,\mathbf{u}}|| < c$, $\forall i = 1, \ldots, n$, respectively. In all other cases, the UMQPE estimator in (4.21) depends on the unknown tuning constant c of the Huber's influence function.

The choice of an appropriate value for c is not straightforward. Ideally, it should be data-driven and account for possible outliers in the data. In the literature on univariate M-estimation, c can be either fixed a-priori or defined by the data analyst to achieve a specified asymptotic efficiency under normality (Huber & Ronchetti 2009), maximize the asymptotic efficiency (Wang et al. 2007) or it can be estimated in a likelihood framework as illustrated by Bianchi et al. (2018). In our multivariate context, we propose to select the optimal value of c, denoted with $c^\star$, via K-fold cross-validation which allows us to consider c as a data-driven parameter. In particular, for fixed τ and $\mathbf{u}$, we construct a uniform grid of values of c from $c_{min} = 0.1$ to $c_{max} = \max_{i=1,\ldots,n} || \mathbf{Y}_i ||$. Then, for each value of $c \in [c_{min}, \ldots, c_{max}]$, we fit the proposed model and determine the optimal value $c^\star$.

4.3.2 Asymptotic properties

This section presents the asymptotic properties of the estimator $\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}$ in (4.21). Specifically, we derive the Bahadur-type (Bahadur 1966) representation, consistency and asymptotic normality for fixed τ, direction $\mathbf{u}$ and c. To prove the following results, we follow Firpo et al. (2009) where they consider the IF and not its recentered

version. Either using the IF or the RIF, all regression coefficients are the same, the only exception being the intercept.

Consider the following assumptions:

(A1) The distribution of the random vector $\mathbf{Y}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^p$, with a density bounded on every compact subset of $\mathbb{R}^p$.

(A2) The observations $(\mathbf{Y}_i, \mathbf{X}_i), i = 1, \ldots, n$ are an i.i.d. sample from $(\mathbf{Y}, \mathbf{X})$.

(A3) The function $\Psi(\cdot)$ in (4.6) is bounded, non-decreasing in each argument and possesses bounded derivatives up to the second order with the convention $\Psi(\mathbf{0}) = \mathbf{0}$.

(A4) $\mathbb{E}[|| \eta_\delta(\varphi)\Psi(\mathbf{Y} - \boldsymbol{\theta}) ||^2] < \infty, \forall \boldsymbol{\theta} \in \mathbb{R}^p$.

(A5) The $p \times p$ matrix $\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$ in (4.10) is positive definite.

(A6) $\widehat{\Omega}_{\mathbf{X}}$ is nonsingular almost surely for n sufficiently large and converges to $\Omega_{\mathbf{X}} = \mathbb{E}[\mathbf{X}\mathbf{X}']$.

It is worth noticing that assumptions A1-A6 are quite mild and are standard in robust estimation theory. For instance, assumptions A1-A5 are needed for the Bahadur (Bahadur 1966) representation and ensure the invertibility of $\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$.

In order to present the asymptotic properties $\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}$, we first need to establish the Bahadur-type representation for $\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}$ and its limiting distribution.

Theorem 4. *Let assumptions A1-A5 hold. Then, for any* $\tau \in (0, \frac{1}{2}]$ *and* $\mathbf{u} \in \mathcal{S}^{p-1}$, *the following asymptotic linear representation holds:*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}} - \boldsymbol{\theta}_{\tau,\mathbf{u}}) = \mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_\delta(\varphi_i)\Psi(\mathbf{Y}_i - \boldsymbol{\theta}_{\tau,\mathbf{u}}) + o_p(1) \tag{4.22}$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}} - \boldsymbol{\theta}_{\tau,\mathbf{u}}) \xrightarrow{p} \mathcal{N}(\mathbf{0}, \mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1}\mathbf{D}(\boldsymbol{\theta}_{\tau,\mathbf{u}})\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1}) \qquad \text{as} \qquad n \to \infty, \tag{4.23}$$

where $\mathbf{D}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$ *defines a* $p \times p$ *matrix:*

$$\mathbf{D}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \mathbb{E}[\eta_\delta^2(\varphi)\Psi(\mathbf{Y} - \boldsymbol{\theta}_{\tau,\mathbf{u}})\Psi'(\mathbf{Y} - \boldsymbol{\theta}_{\tau,\mathbf{u}})]. \tag{4.24}$$

Proof. See Proof of Theorem 4 in Appendix. □

To prove consistency and asymptotic normality of $\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}$, we exploit Theorem 4 and define the $k \times p$ matrix of OLS regression coefficients of $IF(\mathbf{Y}; \boldsymbol{\theta}_{\tau,\mathbf{u}})$ on $\mathbf{X}$:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \widehat{\Omega}_{\mathbf{X}}^{-1} \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \, IF'(\mathbf{Y}_i; \boldsymbol{\theta}_{\tau,\mathbf{u}}) \tag{4.25}$$

whose population counterpart is:

$$\boldsymbol{\beta}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \Omega_{\mathbf{X}}^{-1}\mathbb{E}[\mathbf{X}\, IF'(\mathbf{Y};\boldsymbol{\theta}_{\tau,\mathbf{u}})]. \tag{4.26}$$

Also, let us denote for $i = 1, \ldots, n$,

$$\boldsymbol{\gamma}_i^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \text{vec}\Big(\Omega_{\mathbf{X}}^{-1}\mathbf{X}_i\mathbf{z}_i'(\boldsymbol{\theta}_{\tau,\mathbf{u}})\Big) \qquad \text{and} \qquad \mathbf{z}_i(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = IF(\mathbf{Y}_i;\boldsymbol{\theta}_{\tau,\mathbf{u}}) - \boldsymbol{\beta}'(\boldsymbol{\theta}_{\tau,\mathbf{u}})\mathbf{X}_i. \tag{4.27}$$

where the vec$(\cdot)$ operator converts a matrix into a column vector by stacking its columns on top of one another. Finally, we define:

$$\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}^{\star} = \text{vec}(\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}) \qquad \text{and} \qquad \boldsymbol{\alpha}_{\tau,\mathbf{u}}^{\star} = \text{vec}(\boldsymbol{\alpha}_{\tau,\mathbf{u}}). \tag{4.28}$$

Theorem 5. *Let assumptions A1-A6 hold. Then, for any* $\tau \in (0, \frac{1}{2}]$ *and* $\mathbf{u} \in \mathcal{S}^{p-1}$*, the following asymptotic linear representation holds:*

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}^{\star} - \boldsymbol{\alpha}_{\tau,\mathbf{u}}^{\star}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{S}_i(\boldsymbol{\theta}_{\tau,\mathbf{u}}) + o_p(1) \tag{4.29}$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}^{\star} - \boldsymbol{\alpha}_{\tau,\mathbf{u}}^{\star}) \xrightarrow{p} \mathcal{N}\Big(\mathbf{0}, \mathbb{E}\Big[\mathbf{S}(\boldsymbol{\theta}_{\tau,\mathbf{u}})\mathbf{S}'(\boldsymbol{\theta}_{\tau,\mathbf{u}})\Big]\Big) \qquad \text{as} \qquad n \to \infty, \tag{4.30}$$

where $\mathbf{S}_i(\boldsymbol{\theta}_{\tau,\mathbf{u}})$ *is a kp-dimensional vector:*

$$\mathbf{S}_i(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}}\boldsymbol{\beta}^{\star}(\boldsymbol{\theta})\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1}\eta_{\delta}(\varphi_i)\Psi(\mathbf{Y}_i - \boldsymbol{\theta}_{\tau,\mathbf{u}}) + \boldsymbol{\gamma}_i^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}), \tag{4.31}$$

and $\nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}}\boldsymbol{\beta}^{\star}(\boldsymbol{\theta})$ *is the derivative of* $\boldsymbol{\beta}^{\star}(\boldsymbol{\theta})$ *with respect to* $\boldsymbol{\theta}$ *evaluated at* $\boldsymbol{\theta}_{\tau,\mathbf{u}}$*.*

Proof. See Proof of Theorem 5 in Appendix. □

In order to use the previous Theorem to build confidence intervals and hypothesis tests, in what follows we provide a consistent estimator of the asymptotic covariance matrix of $\widehat{\boldsymbol{\alpha}}_{\tau,\mathbf{u}}^{\star}$ in (4.30). The analytical form of the asymptotic covariance matrix suggests the following estimator:

$$\widehat{\mathbf{V}}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{S}}_i(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}})\widehat{\mathbf{S}}_i'(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) \tag{4.32}$$

where $\widehat{\mathbf{S}}_i(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}})$ is the kp-dimensional vector:

$$\widehat{\mathbf{S}}_i(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) = \Big(\nabla_{\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}}\widehat{\boldsymbol{\beta}}^{\star}(\boldsymbol{\theta})\mathbf{M}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}})^{-1}\eta_{\delta}(\widehat{\varphi}_i)\Psi(\mathbf{Y}_i - \widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) + \widehat{\boldsymbol{\gamma}}_i^{\star}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}})\Big) \tag{4.33}$$

and $\nabla_{\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}}\widehat{\boldsymbol{\beta}}^{\star}(\boldsymbol{\theta})$ can be obtained via numerical differentiation.

The next Theorem proves consistency of $\widehat{\mathbf{V}}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}})$.

Theorem 6. *Let assumptions A1-A6 hold,*

$$\widehat{\mathbf{V}}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) - \mathbb{E}\Big[\mathbf{S}(\boldsymbol{\theta}_{\tau,\mathbf{u}})\mathbf{S}'(\boldsymbol{\theta}_{\tau,\mathbf{u}})\Big] \xrightarrow{p} 0, \tag{4.34}$$

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

Proof. See Proof of Theorem 6 in Appendix. □

4.4 Simulation study

In this section, we conduct a simulation study to evaluate the finite sample properties of the proposed method, showing that it represents a valid procedure to estimate the effect of the covariates on the quantiles, M-quantiles and expectiles of the unconditional distribution of multivariate responses. This simulation exercise addresses the following issues. First, we consider different distributional choices for the error term to study the performance of the model in the presence of non-Gaussian errors. Second, we evaluate the efficiency of the introduced estimators using different values of the tuning constant c of the multivariate Huber's influence function in (4.6). Finally, we assess the performance of the cross-validation procedure described in Section 4.3.1 to select the optimal c.

We set $p = 2$, $n = 1000$ and consider two explanatory variables, i.e., $X_i^{(1)} \sim \mathcal{N}(0, 0.5)$ and $X_i^{(2)} \sim \mathcal{N}(0, 0.5)$. The observations are generated according to the following bivariate regression model:

$$\mathbf{Y}_i = \mathbf{B}'\mathbf{X}_i + \boldsymbol{\epsilon}_i, \tag{4.35}$$

where $\mathbf{X}_i = (1, X_i^{(1)}, X_i^{(2)})'$ and $\mathbf{B} = \begin{pmatrix} 2 & 5 \\ -1.7 & 1.5 \\ 1 & -0.8 \end{pmatrix}$. It is worth recalling that under the data generating process in (4.35), it follows from Theorem 3 that the slope parameters in $\mathbf{B}$ coincides with the UMQPE of $\mathbf{X}$.
To address the first aim of this simulation study, following Dawber et al. (2020), the error terms $\boldsymbol{\epsilon}_i$ in (4.35) are sampled from three distributions, namely:

(N): a multivariate standard Normal distribution with zero mean and variance-covariance matrix equal to $\boldsymbol{\Sigma}_1 = \left(\begin{smallmatrix} 1 & 0.3 \\ 0.3 & 1 \end{smallmatrix}\right)$, that is $\boldsymbol{\epsilon}_i \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_1)$;

(CN): a contaminated multivariate Normal distribution $\boldsymbol{\epsilon}_i \sim (1-\lambda)\mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_1) + \lambda\mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}_2)$, with $\lambda = 5\%$ and $\boldsymbol{\Sigma}_2 = \left(\begin{smallmatrix} 100 & 85.732 \\ 85.732 & 150 \end{smallmatrix}\right)$, which corresponds to a correlation between errors of 0.7;

(T): a multivariate Student-t distribution with 3 degrees of freedom and variance-covariance matrix equal to $\boldsymbol{\Sigma}_1$, that is $\boldsymbol{\epsilon}_i \sim \mathcal{T}_2(3, \boldsymbol{\Sigma}_1)$.

To answer the second question, we compare the efficiency of the introduced estimators using three values for the tuning constant in (4.6), i.e., $c = 0$, $c = 1.5$ and an arbitrarily large value $c = 100$. In the first and third case our model reduces, respectively, to the unconditional quantile regression and unconditional expectile regression while the value $c = 1.5$ in between down-weights the influence of outlying observations offering moderate protection against outliers. We select three τ levels, $\tau = 0.10, \tau = 0.25$ and $\tau = 0.45$, three directions in the unit circle, $\mathbf{u}_1 = (1, 0), \mathbf{u}_2 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and $\mathbf{u}_3 = (0, 1)$, and report the following indicators over $H = 1000$ Monte Carlo replications. The Average Relative Bias (ARB) defined as:

$$ARB(\hat{\theta}_\tau) = \frac{1}{H}\sum_{h=1}^{H} \frac{(\hat{\theta}_\tau^{(h)} - \theta_\tau)}{\theta_\tau} \times 100, \tag{4.36}$$

where $\hat{\theta}_\tau^{(h)}$ is the estimated parameter at level τ for the h-th replication and θ_τ is the corresponding "true" value of the parameter. Secondly, the Root Mean Square Error (RMSE) of model parameters averaged across the H simulations:

$$RMSE(\hat{\theta}_\tau) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{\theta}_\tau^{(h)} - \theta_\tau)^2}. \tag{4.37}$$

Tables 4.1, 4.2 and 4.3 report the ARB (in percentage) and the RMSE (in brackets) of the slope parameters $\mathbf{B}_{11} = -1.7$, $\mathbf{B}_{12} = 1.5$, $\mathbf{B}_{21} = 1$ and $\mathbf{B}_{22} = -0.8$, for the three considered values of c.

Firstly, one can see that all models are able to recover the true parameter values for all considered directions and values of the tuning constant c. Indeed, both the ARB and the RMSE remain reasonably small at the center of the distribution ($\tau = 0.45$), but they tend to increase as the quantile level becomes more extreme ($\tau = 0.10$) when we consider contaminated or heavy-tailed distributions (CN and T scenarios). Secondly, to evaluate the efficiency of the considered estimators, we compare the RMSE for different tuning constants under the three scenarios for the error term. It is possible to observe that with normal errors our approach achieves the highest efficiency when $c = 100$ for all quantile levels. As expected, for normally distributed data, resistance against outliers is not required and a large value for the tuning constant is preferred. In contrast, for contaminated normal and Student-t errors, our estimators are more efficient when a smaller tuning constant is selected, which reflect the presence of potential influential observations that should be down-weighted (Dawber et al. 2020). These findings are consistent across the considered three directions and provide evidence that the proposed model offers great flexibility in controlling for the presence of outlying values.

From a computational point of view, to evaluate the execution times of the estimation procedure, we report the median CPU Time (in seconds) required by the implemented `R` code using an Intel Xeon E5-2609 2.40GHz processor. As can be seen, the fastest running times are associated to the unconditional expectile regression with a minimum of 0.575 seconds, while the slowest ones are associated to the unconditional quantile regression with a maximum of 0.981 seconds. These results reveal decent computational performance, confirming the practical feasibility of our estimation procedure.

Furthermore, we evaluate the proposed asymptotic standard errors of the UMQPE estimators described in Section 4.3.2. We start by comparing the empirical Monte Carlo standard errors, $S(\hat{\theta}_\tau) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{\theta}_\tau^{(h)} - \bar{\theta}_\tau)^2}$ with $\bar{\theta}_\tau = \frac{1}{H}\sum_{h=1}^{H}\hat{\theta}_\tau^{(h)}$, and the estimated standard errors, $S_A(\hat{\theta}_\tau)$, based on expression (4.32) averaged over $H = 1000$ replications. For all considered scenarios and values of c, the results are reported in Tables 4.4, 4.5 and 4.6. Firstly, as expected, we can see that both the empirical and asymptotic standard error estimates are larger when we are modeling extreme (M-)quantiles ($\tau = 0.10$) near the convex hull of the data than of (M-)quantiles close to the center of the distribution ($\tau = 0.45$). Secondly, it can be observed that the asymptotic standard errors offer a good approximation to the corresponding empirical ones. In some cases, however, we notice a slight overestimation of the empirical standard errors when $c = 0$ and when $\tau = 0.1$ and

$\tau = 0.25$, which can be ascribed to data sparsity in low density regions. In addition, we examine the confidence interval coverage of the proposed estimators constructed using (4.32). In Table 4.7 we report the Coverage Probability (CP) of nominal 95% confidence intervals of the UMQPEs defined as the ratio between the number of times the interval $\theta_\tau \pm 2\sqrt{\text{Var}(\hat{\theta}_\tau)}$ contains the "true" population parameter and the number of Monte Carlo replicates, H. One can observe that the estimated CPs are generally accurate, although few values are slightly away from the nominal 95% level when data are generated under the CN and T scenarios and at levels $\tau = 0.10$ and $\tau = 0.25$, due to the reduced amount of information as we approach the convex hull of the sample. We expect that the results can potentially be improved by either increasing the number of observations and/or the number of Monte Carlo replications.

$\mathbf{u}$	Error	τ	$\mathbf{B}_{11} = -1.7$	$\mathbf{B}_{12} = 1.5$	$\mathbf{B}_{21} = 1$	$\mathbf{B}_{22} = -0.8$	CPU Time
		0.10	0.315 (0.303)	0.482 (0.239)	0.699 (0.199)	0.173 (0.156)	0.860
	Gaussian	0.25	0.413 (0.205)	−0.378 (0.182)	0.677 (0.141)	−0.072 (0.127)	0.962
		0.45	−0.104 (0.178)	0.498 (0.160)	0.033 (0.121)	−0.083 (0.111)	0.929
		0.10	0.648 (0.289)	0.178 (0.249)	0.383 (0.206)	0.526 (0.166)	0.981
$\mathbf{u}_1$	Contamined	0.25	0.030 (0.217)	−0.035 (0.209)	0.141 (0.148)	−0.214 (0.136)	0.943
		0.45	0.168 (0.178)	0.029 (0.164)	0.393 (0.125)	−0.042 (0.114)	0.909
		0.10	−0.603 (0.324)	0.493 (0.259)	−0.865 (0.221)	0.477 (0.171)	0.941
	Student-t	0.25	−0.712 (0.216)	−0.712 (0.214)	0.361 (0.151)	0.613 (0.141)	0.906
		0.45	0.273 (0.199)	−0.290 (0.176)	0.052 (0.137)	−0.360 (0.122)	0.880
		0.10	−0.036 (0.251)	−0.696 (0.237)	0.009 (0.163)	−0.947 (0.160)	0.951
	Gaussian	0.25	0.525 (0.203)	0.507 (0.184)	0.502 (0.139)	0.665 (0.129)	0.933
		0.45	−0.033 (0.173)	0.335 (0.163)	−0.137 (0.119)	−0.023 (0.113)	0.906
		0.10	0.606 (0.252)	0.804 (0.239)	0.554 (0.172)	0.741 (0.161)	0.946
$\mathbf{u}_2$	Contamined	0.25	−0.092 (0.217)	−0.711 (0.186)	0.090 (0.147)	−0.821 (0.131)	0.918
		0.45	0.081 (0.179)	0.123 (0.174)	0.253 (0.123)	0.105 (0.121)	0.892
		0.10	−0.067 (0.274)	−0.648 (0.257)	−0.664 (0.189)	−0.075 (0.184)	0.919
	Student-t	0.25	0.092 (0.218)	0.507 (0.185)	−0.181 (0.149)	−0.490 (0.135)	0.885
		0.45	−0.218 (0.197)	−0.247 (0.174)	−0.497 (0.137)	0.362 (0.122)	0.866
		0.10	0.239 (0.240)	−0.467 (0.280)	−0.118 (0.160)	−0.547 (0.177)	0.974
	Gaussian	0.25	0.566 (0.188)	−0.094 (0.199)	−0.316 (0.129)	−0.160 (0.134)	0.953
		0.45	0.173 (0.178)	0.127 (0.171)	0.000 (0.125)	−0.213 (0.116)	0.921
		0.10	−0.090 (0.263)	0.938 (0.290)	−0.170 (0.175)	0.680 (0.193)	0.963
$\mathbf{u}_3$	Contamined	0.25	−0.078 (0.218)	0.687 (0.191)	−0.030 (0.141)	−0.733 (0.136)	0.935
		0.45	0.164 (0.191)	0.513 (0.166)	0.295 (0.128)	0.446 (0.118)	0.902
		0.10	0.219 (0.268)	0.265 (0.278)	−0.349 (0.181)	0.596 (0.192)	0.924
	Student-t	0.25	−0.350 (0.205)	−0.241 (0.203)	−0.171 (0.141)	−0.187 (0.141)	0.900
		0.45	0.361 (0.176)	−0.232 (0.181)	−0.004 (0.126)	−0.191 (0.126)	0.876

Table 4.1. Values of ARB (in percentage), RMSE (in brackets) and median CPU Time (in seconds) required to fit the model over 1000 Monte Carlo simulations, under the three data generating processes and using $c = 0$.

Finally, to evaluate the performance of the considered cross-validation procedure to select c, we consider the same simulation experiment where we include only the intercept with no covariates. The simulations are carried out on null models as the effect of different tuning constant will be most prominent on the intercept coefficient (Dawber et al. 2020). Specifically, we construct a uniform grid of 50 values for c from 0.1 to 5 as described in Section 4.3.1. Then, for fixed $\mathbf{u}$ and τ, we use a K-fold cross-validation to select the tuning constant. Table 4.8 reports the median value of

$\mathbf{u}$	Error	τ	$\mathbf{B}_{11} = -1.7$	$\mathbf{B}_{12} = 1.5$	$\mathbf{B}_{21} = 1$	$\mathbf{B}_{22} = -0.8$	CPU Time
		0.10	0.234 (0.135)	0.350 (0.107)	0.289 (0.113)	−0.354 (0.095)	0.716
	Gaussian	0.25	0.014 (0.098)	0.152 (0.087)	0.024 (0.083)	0.039 (0.081)	0.751
		0.45	−0.055 (0.084)	−0.004 (0.077)	−0.130 (0.074)	−0.192 (0.073)	0.728
		0.10	0.122 (0.140)	0.105 (0.117)	0.218 (0.133)	0.384 (0.104)	0.781
$\mathbf{u}_1$	Contamined	0.25	0.109 (0.102)	0.133 (0.094)	0.276 (0.095)	0.391 (0.086)	0.743
		0.45	0.068 (0.088)	0.228 (0.083)	0.257 (0.080)	0.393 (0.079)	0.719
		0.10	−0.059 (0.160)	0.242 (0.127)	−0.613 (0.140)	0.697 (0.112)	0.771
	Student-t	0.25	0.035 (0.114)	0.230 (0.099)	−0.378 (0.101)	−0.466 (0.092)	0.731
		0.45	0.015 (0.096)	0.136 (0.087)	−0.348 (0.089)	0.230 (0.083)	0.713
		0.10	−0.206 (0.108)	0.024 (0.100)	−0.263 (0.091)	0.379 (0.094)	0.755
	Gaussian	0.25	−0.106 (0.091)	−0.012 (0.084)	−0.159 (0.079)	0.042 (0.080)	0.729
		0.45	−0.051 (0.084)	−0.006 (0.077)	−0.143 (0.073)	−0.179 (0.074)	0.712
		0.10	0.264 (0.117)	0.439 (0.114)	0.284 (0.108)	−0.444 (0.105)	0.754
$\mathbf{u}_2$	Contamined	0.25	0.156 (0.095)	0.315 (0.093)	0.256 (0.087)	−0.397 (0.086)	0.723
		0.45	0.070 (0.088)	0.237 (0.084)	0.247 (0.079)	0.368 (0.079)	0.709
		0.10	−0.094 (0.137)	0.259 (0.129)	−0.814 (0.125)	0.817 (0.120)	0.760
	Student-t	0.25	−0.080 (0.107)	0.184 (0.097)	−0.555 (0.098)	0.442 (0.093)	0.719
		0.45	−0.010 (0.095)	0.121 (0.087)	−0.385 (0.089)	0.220 (0.083)	0.705
		0.10	−0.016 (0.112)	−0.066 (0.117)	−0.116 (0.092)	0.511 (0.105)	0.772
	Gaussian	0.25	−0.101 (0.093)	−0.064 (0.089)	−0.231 (0.079)	0.034 (0.083)	0.746
		0.45	−0.047 (0.083)	0.003 (0.078)	−0.160 (0.073)	−0.168 (0.074)	0.722
		0.10	−0.006 (0.118)	0.504 (0.138)	0.025 (0.101)	0.193 (0.123)	0.768
$\mathbf{u}_3$	Contamined	0.25	0.035 (0.097)	0.399 (0.100)	0.135 (0.086)	0.331 (0.091)	0.738
		0.45	0.064 (0.088)	0.253 (0.084)	0.238 (0.079)	0.361 (0.079)	0.714
		0.10	−0.089 (0.130)	0.424 (0.147)	−0.561 (0.118)	0.852 (0.132)	0.763
	Student-t	0.25	−0.122 (0.108)	0.144 (0.103)	−0.554 (0.098)	−0.345 (0.096)	0.728
		0.45	−0.037 (0.096)	0.092 (0.088)	−0.428 (0.089)	−0.178 (0.083)	0.712

Table 4.2. Values of ARB (in percentage), RMSE (in brackets) and median CPU Time (in seconds) required to fit the model over 1000 Monte Carlo simulations, under the three data generating processes and using $c = 1.5$.

c being selected, $\bar{c}^{\star}$, and the Proportion of Huberised Residuals (PHR) over 1000 Monte Carlo replications, using $K = 5$ and $K = 10$ folds. The results show that the considered approach reduces fewer observations as c increases and as τ is far from 0.5, for all three distributions. As one can see, the selected tuning constant down-weights influential observations in heavy-tailed distributions (see Panels B and C) when estimating the center of the distribution and, on the other hand, the PHR tends to be smaller when we are interested in extreme τ levels. Furthermore, both the median value of the selected tuning constant and the PHR are very similar when using $K = 10$ in place of $K = 5$ for all scenarios considered. This therefore suggests that the considered procedure is appropriate for determining a suitable tuning constant from the data.

4.5 Application

In this section, we consider data from the Survey on Household Income and Wealth (SHIW) 2016 conducted by the Bank of Italy to show the relevance of our methodology. We analyze the impact of economic and socio-demographic factors on households wealth and consumption levels, accounting both for the presence of outliers and the correlation structure between the two outcomes. We are interested in evaluating

$\mathbf{u}$	Error	τ	$\mathbf{B}_{11} = -1.7$	$\mathbf{B}_{12} = 1.5$	$\mathbf{B}_{21} = 1$	$\mathbf{B}_{22} = -0.8$	CPU Time
		0.10	−0.046 (0.106)	0.018 (0.085)	0.112 (0.092)	0.122 (0.081)	0.612
	Gaussian	0.25	−0.017 (0.077)	−0.006 (0.070)	0.056 (0.071)	0.045 (0.070)	0.626
		0.45	−0.012 (0.066)	−0.045 (0.064)	−0.025 (0.063)	−0.029 (0.065)	0.598
		0.10	−0.453 (0.317)	0.802 (0.309)	0.204 (0.312)	−0.564 (0.294)	0.656
$\mathbf{u}_1$	Contamined	0.25	−0.273 (0.195)	0.599 (0.222)	0.124 (0.193)	−0.129 (0.212)	0.616
		0.45	−0.177 (0.161)	0.466 (0.193)	0.114 (0.157)	0.082 (0.185)	0.589
		0.10	0.039 (0.191)	0.162 (0.148)	−1.151 (0.189)	1.090 (0.144)	0.634
	Student-t	0.25	0.028 (0.130)	0.164 (0.119)	−0.953 (0.130)	0.877 (0.118)	0.603
		0.45	0.049 (0.108)	0.101 (0.108)	−0.837 (0.110)	0.729 (0.108)	0.576
		0.10	−0.100 (0.086)	−0.060 (0.080)	−0.054 (0.077)	0.259 (0.080)	0.649
	Gaussian	0.25	−0.044 (0.072)	−0.071 (0.068)	−0.020 (0.067)	0.087 (0.070)	0.622
		0.45	−0.014 (0.066)	−0.053 (0.063)	−0.034 (0.063)	−0.021 (0.065)	0.597
		0.10	−0.788 (0.419)	1.194 (0.508)	0.997 (0.401)	−1.182 (0.484)	0.681
$\mathbf{u}_2$	Contamined	0.25	−0.336 (0.206)	0.787 (0.251)	0.308 (0.202)	−0.350 (0.243)	0.621
		0.45	−0.186 (0.161)	0.487 (0.194)	0.137 (0.157)	0.045 (0.186)	0.584
		0.10	−0.036 (0.172)	0.333 (0.174)	−1.014 (0.172)	1.117 (0.167)	0.639
	Student-t	0.25	−0.020 (0.125)	0.208 (0.125)	−0.950 (0.126)	0.810 (0.123)	0.604
		0.45	0.035 (0.108)	0.105 (0.108)	−0.855 (0.110)	0.705 (0.108)	0.575
		0.10	−0.053 (0.087)	−0.156 (0.093)	−0.070 (0.077)	0.247 (0.088)	0.651
	Gaussian	0.25	−0.041 (0.073)	−0.123 (0.071)	−0.076 (0.067)	0.060 (0.072)	0.624
		0.45	−0.012 (0.066)	−0.062 (0.063)	−0.046 (0.063)	−0.025 (0.065)	0.594
		0.10	−0.704 (0.284)	1.026 (0.444)	0.531 (0.272)	−0.977 (0.431)	0.664
$\mathbf{u}_3$	Contamined	0.25	−0.349 (0.189)	0.745 (0.246)	0.258 (0.183)	−0.300 (0.239)	0.616
		0.45	−0.181 (0.161)	0.481 (0.194)	0.144 (0.156)	0.051 (0.186)	0.583
		0.10	−0.067 (0.141)	0.458 (0.190)	−1.159 (0.142)	0.832 (0.176)	0.633
	Student-t	0.25	−0.045 (0.118)	0.192 (0.130)	−1.003 (0.119)	0.669 (0.125)	0.602
		0.45	0.032 (0.107)	0.094 (0.108)	−0.860 (0.109)	0.677 (0.108)	0.575

Table 4.3. Values of ARB (in percentage), RMSE (in brackets) and median CPU Time (in seconds) required to fit the model over 1000 Monte Carlo simulations, under the three data generating processes and using $c = 100$.

whether these effects are more pronounced on more disadvantaged families than on richer ones. In this setting, the limitation of using a conditional (M-)quantile model is that the effect of the covariates at different quantile levels may be masked by the set of conditioning variables, i.e., the characteristics of the family. Once we have conditioned on the explanatory variables, for instance, the 10-th (M-)quantile of the unconditional distribution of the responses may potentially be very different from the 10-th (M-)quantile of the conditional distribution, so the coefficients of conditional (M-)quantile regression cannot be interpreted as unconditional effects. On the other hand, by using our unconditional method, the UMQPE provides an estimate of the impact of covariates across the entire population and not merely among population subgroups, consisting of families who share the same values of the included covariates. In what follows, we fit the proposed regression method at different points of the unconditional distributions of family wealth and consumption, and illustrate the difference between the conditional and unconditional approaches.

4.5.1 Data description

The SHIW (https://www.bancaditalia.it) is an annual survey conducted by the Bank of Italy whose aims are to provide information on the economic and financial behaviours of Italian households and collect reliable, comparable and representative

$\mathbf{u}$	Error	τ	$S(\hat{\mathbf{B}}_{11})$	$S_A(\hat{\mathbf{B}}_{11})$	$S(\hat{\mathbf{B}}_{12})$	$S_A(\hat{\mathbf{B}}_{12})$	$S(\hat{\mathbf{B}}_{21})$	$S_A(\hat{\mathbf{B}}_{21})$	$S(\hat{\mathbf{B}}_{22})$	$S_A(\hat{\mathbf{B}}_{22})$
		0.10	0.3034	0.3603	0.2386	0.2492	0.1992	0.2313	0.1556	0.1681
	Gaussian	0.25	0.2049	0.2162	0.1818	0.1815	0.1407	0.1466	0.1267	0.1268
		0.45	0.1780	0.1643	0.1597	0.1570	0.1208	0.1174	0.1108	0.1127
		0.10	0.2881	0.3701	0.2495	0.2742	0.2061	0.2404	0.1663	0.1843
$\mathbf{u}_1$	Contamined	0.25	0.2171	0.2276	0.2088	0.2044	0.1476	0.1556	0.1356	0.1410
		0.45	0.1777	0.1792	0.1645	0.1701	0.1245	0.1278	0.1139	0.1208
		0.10	0.3242	0.4272	0.2588	0.3076	0.2212	0.2774	0.1707	0.2060
	Student-t	0.25	0.2159	0.2288	0.2141	0.1962	0.1508	0.1605	0.1410	0.1406
		0.45	0.1987	0.1813	0.1757	0.1755	0.1370	0.1317	0.1219	0.1269
		0.10	0.2508	0.2941	0.2366	0.2841	0.1628	0.1910	0.1595	0.1844
	Gaussian	0.25	0.2033	0.2052	0.1835	0.1947	0.1392	0.1404	0.1294	0.1335
		0.45	0.1735	0.1750	0.1633	0.1620	0.1194	0.1223	0.1130	0.1141
		0.10	0.2516	0.3269	0.2385	0.3202	0.1715	0.2120	0.1612	0.2104
$\mathbf{u}_2$	Contamined	0.25	0.2171	0.2156	0.1861	0.2089	0.1472	0.1485	0.1311	0.1442
		0.45	0.1794	0.1753	0.1744	0.1668	0.1232	0.1253	0.1213	0.1197
		0.10	0.2739	0.3384	0.2565	0.3321	0.1892	0.2279	0.1839	0.2225
	Student-t	0.25	0.2185	0.2193	0.1841	0.2071	0.1491	0.1542	0.1342	0.1467
		0.45	0.1967	0.1908	0.1736	0.1841	0.1371	0.1365	0.1225	0.1308
		0.10	0.2405	0.2576	0.2798	0.3213	0.1604	0.1724	0.1771	0.2071
	Gaussian	0.25	0.1883	0.2019	0.1993	0.2212	0.1294	0.1386	0.1344	0.1481
		0.45	0.1778	0.1693	0.1706	0.1658	0.1246	0.1199	0.1161	0.1160
		0.10	0.2634	0.2968	0.2901	0.3748	0.1749	0.1957	0.1927	0.2413
$\mathbf{u}_3$	Contamined	0.25	0.2181	0.2110	0.1906	0.2300	0.1410	0.1464	0.1360	0.1557
		0.45	0.1909	0.1717	0.1662	0.1650	0.1278	0.1234	0.1182	0.1188
		0.10	0.2682	0.2833	0.2776	0.3688	0.1807	0.1938	0.1925	0.2444
	Student-t	0.25	0.2047	0.2185	0.2033	0.2397	0.1415	0.1539	0.1412	0.1639
		0.45	0.1758	0.1893	0.1815	0.1852	0.1265	0.1351	0.1256	0.1310

Table 4.4. Empirical, $S(\cdot)$, and asymptotic, $S_A(\cdot)$, standard error estimates of the UMQPEs over 1000 Monte Carlo simulations, under the three data generating processes and using $c = 0$.

data of the population resident in Italy. This survey is widely regarded as the basis of the most reliable estimates for macroeconomics studies. The sample is drawn in two stages, the primary and secondary sampling units are municipalities and households, respectively. Before the primary units are selected, they are stratified by region and population size. Data are collected mainly via an electronic questionnaire using the Computer Assisted Personal Interviewing program while the remaining interviews are conducted using the Paper And Pencil Personal Interviewing program.

In this work, following established custom we transform the dependent variables, i.e., household consumption (LCON) and net wealth (LWEA) to natural logarithm. In particular, LCON is defined as the sum of household's expenditure on durables and non-durable goods while net wealth is obtained as the algebraic sum of real assets, financial assets and financial liabilities. The set of considered predictors includes the log of net disposable income (LINC), defined as the sum of payroll income, pensions, net transfers, net self-employment income and property income sources, and relevant information on the household's head such as age (Age) and age squared (Age2) measured in years. Also, gender (male (baseline)), marital (married (baseline), never married, separated, widowed), employment status (employee (baseline), self-employed, not-employed) and educational level (no education (baseline), elementary, middle, vocational, high school, university or higher) are included as dummy variables.

$\mathbf{u}$	Error	τ	$S(\hat{\mathbf{B}}_{11})$	$S_A(\hat{\mathbf{B}}_{11})$	$S(\hat{\mathbf{B}}_{12})$	$S_A(\hat{\mathbf{B}}_{12})$	$S(\hat{\mathbf{B}}_{21})$	$S_A(\hat{\mathbf{B}}_{21})$	$S(\hat{\mathbf{B}}_{22})$	$S_A(\hat{\mathbf{B}}_{22})$
		0.10	0.1354	0.1125	0.1068	0.0953	0.1127	0.1041	0.0954	0.0891
	Gaussian	0.25	0.0981	0.0860	0.0872	0.0813	0.0834	0.0879	0.0807	0.0837
		0.45	0.0843	0.0702	0.0772	0.0690	0.0736	0.0702	0.0735	0.0690
		0.10	0.1396	0.1261	0.1175	0.1056	0.1333	0.1172	0.1036	0.0989
$\mathbf{u}_1$	Contamined	0.25	0.1020	0.0887	0.0937	0.0838	0.0947	0.0866	0.0860	0.0821
		0.45	0.0879	0.0759	0.0834	0.0753	0.0799	0.0759	0.0787	0.0753
		0.10	0.1598	0.1465	0.1273	0.1195	0.1398	0.1367	0.1123	0.1126
	Student-t	0.25	0.1139	0.1002	0.0994	0.0934	0.1015	0.0981	0.0922	0.0917
		0.45	0.0958	0.0853	0.0872	0.0835	0.0887	0.0853	0.0833	0.0835
		0.10	0.1078	0.0982	0.0999	0.0983	0.0915	0.0904	0.0937	0.0897
	Gaussian	0.25	0.0909	0.0780	0.0839	0.0770	0.0787	0.0760	0.0804	0.0748
		0.45	0.0837	0.0701	0.0771	0.0691	0.0733	0.0701	0.0736	0.0690
		0.10	0.1167	0.1113	0.1142	0.1125	0.1078	0.1032	0.1054	0.1039
$\mathbf{u}_2$	Contamined	0.25	0.0955	0.3202	0.0927	0.1597	0.0872	0.2294	0.0861	0.1414
		0.45	0.0877	0.0758	0.0835	0.0753	0.0794	0.0758	0.0788	0.0753
		0.10	0.1367	0.1321	0.1285	0.1327	0.1244	0.1239	0.1201	0.1235
	Student-t	0.25	0.1071	0.0967	0.0974	0.0954	0.0983	0.0949	0.0931	0.0935
		0.45	0.0953	0.0852	0.0873	0.0835	0.0886	0.0852	0.0835	0.0835
		0.10	0.1116	0.0934	0.1168	0.1081	0.0924	0.0881	0.1045	0.0996
	Gaussian	0.25	0.0930	0.0771	0.0886	0.0796	0.0793	0.0756	0.0829	0.0773
		0.45	0.0834	0.0700	0.0776	0.0691	0.0734	0.0700	0.0738	0.0691
		0.10	0.1184	0.2425	0.1377	0.2141	0.1010	0.1678	0.1231	0.1689
$\mathbf{u}_3$	Contamined	0.25	0.0968	0.0835	0.1000	0.0880	0.0857	0.0820	0.0910	0.0857
		0.45	0.0882	0.0758	0.0841	0.0754	0.0794	0.0758	0.0793	0.0754
		0.10	0.1304	0.1168	0.1469	0.1442	0.1184	0.1110	0.1322	0.1336
	Student-t	0.25	0.1078	0.0940	0.1027	0.0985	0.0980	0.0926	0.0960	0.0960
		0.45	0.0965	0.0851	0.0877	0.0836	0.0886	0.0851	0.0833	0.0836

Table 4.5. Empirical, $S(\cdot)$, and asymptotic, $S_A(\cdot)$, standard error estimates of the UMQPEs over 1000 Monte Carlo simulations, under the three data generating processes and using $c = 1.5$.

Finally, a categorical variable is included to investigate the presence of regional divergences in wealth and consumption levels depending on the region of residence (north (baseline), centre, south and islands). Table 4.9 summarizes the descriptive statistics of the included variables. The considered sample contains 7027 households.

As a preliminary step, we study the unconditional distributions of households wealth and consumption. The histograms of LWEA and LCON unconditional distributions in Figure 4.1 reveal that, while normality seems tenable for LCON, there are potentially influential observations in the distribution of LWEA and indicate a departure from the Gaussian assumption, having fat tails and pronounced asymmetries. Furthermore, the empirical correlation between LCON and LWEA equals to 0.477. As expected, consumption and wealth are positively correlated, justifying the need for a multivariate approach that considers these two dimensions together. Consequently, the presented unconditional regression model is appropriate to account for outlying observations and investigate how the relationship between responses and explanatory variables can vary across the unconditional distribution of family wealth and consumption.

$\mathbf{u}$	Error	τ	$S(\hat{\mathbf{B}}_{11})$	$S_A(\hat{\mathbf{B}}_{11})$	$S(\hat{\mathbf{B}}_{12})$	$S_A(\hat{\mathbf{B}}_{12})$	$S(\hat{\mathbf{B}}_{21})$	$S_A(\hat{\mathbf{B}}_{21})$	$S(\hat{\mathbf{B}}_{22})$	$S_A(\hat{\mathbf{B}}_{22})$
		0.10	0.1057	0.0941	0.0846	0.0845	0.0923	0.0886	0.0808	0.0787
	Gaussian	0.25	0.0774	0.0716	0.0703	0.0697	0.0711	0.0702	0.0705	0.0681
		0.45	0.0658	0.0635	0.0636	0.0634	0.0629	0.0635	0.0650	0.0634
		0.10	0.3169	0.3421	0.3092	0.3096	0.3118	0.3135	0.2939	0.2954
$\mathbf{u}_1$	Contamined	0.25	0.1953	0.1932	0.2218	0.2145	0.1930	0.1893	0.2119	0.2113
		0.45	0.1610	0.1549	0.1928	0.1838	0.1568	0.1550	0.1848	0.1839
		0.10	0.1914	0.1846	0.1478	0.1536	0.1887	0.1763	0.1441	0.1464
	Student-t	0.25	0.1300	0.1263	0.1192	0.1206	0.1294	0.1246	0.1183	0.1189
		0.45	0.1082	0.1080	0.1076	0.1082	0.1095	0.1080	0.1081	0.1082
		0.10	0.0860	0.0858	0.0796	0.0874	0.0772	0.0793	0.0804	0.0798
	Gaussian	0.25	0.0724	0.0697	0.0681	0.0700	0.0672	0.0680	0.0704	0.0680
		0.45	0.0657	0.0634	0.0634	0.0634	0.0627	0.0634	0.0651	0.0634
		0.10	0.4190	0.4041	0.5079	0.4772	0.4010	0.3932	0.4843	0.4677
$\mathbf{u}_2$	Contamined	0.25	0.2062	0.1995	0.2510	0.2380	0.2020	0.1982	0.2429	0.2368
		0.45	0.1614	0.1551	0.1938	0.1846	0.1570	0.1552	0.1861	0.1847
		0.10	0.1725	0.1739	0.1739	0.1773	0.1719	0.1672	0.1670	0.1695
	Student-t	0.25	0.1253	0.1242	0.1252	0.1252	0.1256	0.1228	0.1226	0.1236
		0.45	0.1079	0.1080	0.1080	0.1083	0.1093	0.1080	0.1082	0.1084
		0.10	0.0868	0.0824	0.0927	0.0930	0.0772	0.0774	0.0881	0.0865
	Gaussian	0.25	0.0734	0.0692	0.0712	0.0714	0.0671	0.0678	0.0716	0.0696
		0.45	0.0658	0.0634	0.0634	0.0635	0.0626	0.0634	0.0650	0.0635
		0.10	0.2840	0.2788	0.4438	0.4472	0.2722	0.2683	0.4310	0.4256
$\mathbf{u}_3$	Contamined	0.25	0.1893	0.1834	0.2458	0.2358	0.1827	0.1810	0.2386	0.2331
		0.45	0.1611	0.1547	0.1937	0.1845	0.1563	0.1547	0.1861	0.1846
		0.10	0.1415	0.1484	0.1896	0.1864	0.1420	0.1423	0.1761	0.1771
	Student-t	0.25	0.1177	0.1193	0.1296	0.1270	0.1183	0.1178	0.1248	0.1252
		0.45	0.1073	0.1078	0.1083	0.1083	0.1088	0.1079	0.1082	0.1084

Table 4.6. Empirical, $S(\cdot)$, and asymptotic, $S_A(\cdot)$, standard error estimates of the UMQPEs over 1000 Monte Carlo simulations, under the three data generating processes and using $c = 100$.

4.5.2 Modeling household wealth and consumption

We analyze the SHIW 2016 data to jointly model households' log-wealth, LWEA, and log-consumption, LCON, as a function of the predictors in Table 4.9. As explained in Section 4.3, a meaningful direction $\mathbf{u}$ shall be determined for ordering multivariate observations. Specifically, in economics and finance, there has been by now a long tradition of studies investigating the relationship between consumption and wealth (see Campbell & Mankiw 1989, Deaton et al. 1992 and Campbell 1993 for example). In this literature, a central object is the wealth-consumption ratio (Li 2005, Lustig et al. 2013). In our specific application where $\mathbf{Y} = (Y_{(1)}, Y_{(2)})'$ denote the multivariate random vector collecting households' log-wealth and log-consumption, it is immediate to see that $Z = \log\left(\frac{Y_{(1)}^{u_1}}{Y_{(2)}^{-u_2}}\right)$ is the log wealth-consumption ratio if $\mathbf{u} = (1, -1)'$. Following the definition in (4.7), we can justify this choice from a practical point of view as each residual, $\mathbf{Y} - \boldsymbol{\theta}_{\tau,\mathbf{u}}$, is assigned a weight depending on the ratio between wealth and consumption. In order to explicitly account for the existing link between wealth and consumption in the choice of $\mathbf{u}$, we use this direction throughout the rest of the section. We fit the proposed approach at levels $\tau = 0.10$, $\tau = 0.50$ and $\tau = 0.90$, which can estimated by simply noting that for $\tau \in (0, \frac{1}{2}]$, $\boldsymbol{\theta}_{1-\tau,\mathbf{u}} = \boldsymbol{\theta}_{\tau,-\mathbf{u}}$ (see Kokic et al. 2002).

$\mathbf{u}$	Error	$\tau = 0.10$				$\tau = 0.25$				$\tau = 0.45$			
		$\mathbf{B}_{11}$	$\mathbf{B}_{12}$	$\mathbf{B}_{21}$	$\mathbf{B}_{22}$	$\mathbf{B}_{11}$	$\mathbf{B}_{12}$	$\mathbf{B}_{21}$	$\mathbf{B}_{22}$	$\mathbf{B}_{11}$	$\mathbf{B}_{12}$	$\mathbf{B}_{21}$	$\mathbf{B}_{22}$
Panel A: $c = 0$													
	Gaussian	0.93	0.91	0.95	0.94	0.92	0.92	0.94	0.95	0.92	0.95	0.94	0.96
$\mathbf{u}_1$	Contamined	0.94	0.92	0.95	0.94	0.92	0.92	0.95	0.96	0.92	0.92	0.94	0.96
	Student-t	0.94	0.91	0.97	0.96	0.93	0.94	0.96	0.96	0.91	0.94	0.95	0.96
	Gaussian	0.94	0.94	0.96	0.96	0.93	0.93	0.95	0.95	0.91	0.95	0.95	0.95
$\mathbf{u}_2$	Contamined	0.93	0.93	0.96	0.96	0.93	0.93	0.95	0.96	0.91	0.91	0.94	0.96
	Student-t	0.96	0.96	0.97	0.97	0.93	0.94	0.95	0.96	0.93	0.94	0.96	0.96
	Gaussian	0.93	0.93	0.95	0.96	0.91	0.93	0.94	0.95	0.95	0.91	0.94	0.95
$\mathbf{u}_3$	Contamined	0.92	0.93	0.96	0.97	0.92	0.93	0.96	0.95	0.90	0.93	0.95	0.95
	Student-t	0.94	0.95	0.96	0.97	0.91	0.94	0.95	0.96	0.93	0.94	0.95	0.96
Panel B: $c = 1.5$													
	Gaussian	0.91	0.93	0.93	0.94	0.90	0.93	0.94	0.94	0.95	0.93	0.95	0.94
$\mathbf{u}_1$	Contamined	0.93	0.93	0.92	0.94	0.93	0.94	0.94	0.94	0.92	0.93	0.94	0.95
	Student-t	0.93	0.94	0.95	0.96	0.92	0.95	0.95	0.96	0.92	0.95	0.94	0.95
	Gaussian	0.93	0.95	0.96	0.95	0.90	0.93	0.96	0.94	0.95	0.93	0.95	0.94
$\mathbf{u}_2$	Contamined	0.94	0.95	0.94	0.96	0.92	0.93	0.95	0.96	0.92	0.93	0.94	0.95
	Student-t	0.95	0.96	0.96	0.96	0.93	0.95	0.94	0.96	0.92	0.94	0.94	0.95
	Gaussian	0.90	0.94	0.95	0.94	0.91	0.92	0.96	0.94	0.95	0.93	0.95	0.94
$\mathbf{u}_3$	Contamined	0.92	0.93	0.94	0.95	0.92	0.92	0.94	0.94	0.91	0.92	0.94	0.95
	Student-t	0.92	0.95	0.94	0.95	0.91	0.95	0.94	0.95	0.92	0.94	0.94	0.95
Panel C: $c = 100$													
	Gaussian	0.93	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.94	0.96	0.95	0.95
$\mathbf{u}_1$	Contamined	0.97	0.95	0.97	0.95	0.96	0.95	0.96	0.95	0.94	0.95	0.96	0.95
	Student-t	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.95	0.96	0.95	0.96	0.95
	Gaussian	0.95	0.97	0.97	0.95	0.95	0.96	0.96	0.95	0.95	0.95	0.96	0.95
$\mathbf{u}_2$	Contamined	0.98	0.96	0.97	0.96	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.95
	Student-t	0.97	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.96	0.96	0.96	0.95
	Gaussian	0.94	0.96	0.96	0.95	0.94	0.96	0.96	0.95	0.94	0.96	0.96	0.95
$\mathbf{u}_3$	Contamined	0.96	0.96	0.96	0.97	0.95	0.95	0.96	0.96	0.95	0.95	0.96	0.95
	Student-t	0.97	0.96	0.96	0.95	0.96	0.96	0.96	0.95	0.95	0.95	0.96	0.95

Table 4.7. CP of the UMQPEs over 1000 Monte Carlo simulations, under the three data generating processes and three values of c.

The estimation of the optimal tuning constant $c^\star$ is obtained using a 5-fold cross-validation[1] over a sequence of 200 possible values as described in Section 4.3.1. The results are reported in Table 4.10 ($c = c^\star$) where we compare the proposed unconditional regression with the standard conditional regression approach. To display the sampling variation, asymptotic standard errors obtained using the results in Section 4.3.2 are presented in parentheses. Parameter estimates are displayed in boldface when significant at the standard 5% level.

The main findings can be summarised as follows. The selected values of c are 0.492, 7.050 and 9.889 at level 0.10, 0.50 and 0.90, respectively. This implies that the estimates are close to the quantile case at $\tau = 0.10$ as more than 90% of residuals are down-weighted (Huberised) meanwhile, at $\tau = 0.50$ and $\tau = 0.90$ correspond to

[1] As a robustness check, we have also considered a cross-validation with $K = 10$ folds. Doing so, leads to comparable results at the investigated levels τ, making our findings unchanged.

	$K=5$						$K=10$					
	$\tau=0.10$		$\tau=0.25$		$\tau=0.45$		$\tau=0.10$		$\tau=0.25$		$\tau=0.45$	
	$\bar{c}^\star$	PHR	$\bar{c}^\star$	PHR	$\bar{c}^\star$	PHR	$\bar{c}^\star$	PHR	$\bar{c}^\star$	PHR	$\bar{c}^\star$	PHR
Panel A: Gaussian												
$\mathbf{u}_1$	4.583	0.000	4.270	0.000	3.332	0.005	4.583	0.000	4.270	0.000	3.332	0.006
$\mathbf{u}_2$	4.791	0.000	4.479	0.000	3.540	0.003	4.791	0.000	4.479	0.000	3.540	0.003
$\mathbf{u}_3$	4.583	0.000	4.270	0.000	3.436	0.004	4.583	0.000	4.270	0.000	3.332	0.005
Panel B: Contamined												
$\mathbf{u}_1$	0.934	0.137	1.038	0.095	0.830	0.150	0.934	0.137	1.038	0.095	0.830	0.148
$\mathbf{u}_2$	0.934	0.186	0.934	0.113	0.830	0.156	0.934	0.186	0.934	0.114	0.830	0.163
$\mathbf{u}_3$	0.830	0.155	0.934	0.125	0.726	0.202	0.830	0.156	0.830	0.127	0.726	0.209
Panel C: Student-t												
$\mathbf{u}_1$	1.247	0.331	1.143	0.323	0.830	0.459	1.247	0.332	1.143	0.323	0.830	0.459
$\mathbf{u}_2$	1.247	0.337	1.143	0.313	0.934	0.405	1.247	0.337	1.143	0.313	0.934	0.415
$\mathbf{u}_3$	1.247	0.333	1.143	0.328	0.830	0.436	1.247	0.332	1.143	0.325	0.830	0.460

Table 4.8. Median selected tuning constant $\bar{c}^\star$ and PHR, under the three data generating processes and three directions.

expectile estimation as almost no observations are Huberised. The estimated values for c reflect the negatively skewed distribution of wealth and support the exploratory analysis in Figure 4.1.

Because the selected values of c lead towards expectile estimation, especially above $\tau = 0.50$, we also consider the case $c = 0$ (see Table 4.11) which allows us to estimate the impact of the covariates on the unconditional quantiles of the responses. Comparing Tables 4.10 and 4.11, slight differences in terms of estimation can be found. This is attributable to the choice of c as the two models allow to target different population parameters by selecting different values for the tuning constant of the Huber function. The above points demonstrate the flexibility of the methodology proposed to extend the classical OLS regression for assessing the effect of covariates, not only at the center, but also at different parts of the unconditional distribution of interest.

Nevertheless, there are still similar results between Tables 4.10 and 4.11. Point estimates generally increase in magnitude when moving outward from the bulk of the data. Income elasticity is always positively associated with both wealth and consumption for all investigated τ levels. One can see that there are small differences in consumption expenditure among males and females. Moreover, education, marital and employment status are important determinants of family's consumption and wealth levels, with an increasing trend as $\tau \to 0$. There also appears to be significant regional disparities across the distributions of the responses as southern regions and islands generally have lower levels of wealth and consumption. It is important to note that the effect of education, marital status and LINC between the conditional and unconditional models is very different, especially at $c = 0$. As shown in Theorem 3, this corresponds to the case where large differences exist between the UMQPE and the CMQPE. This may be due to the fact that the matching, $s_{\tau,\mathbf{u}}(\mathbf{X})$, and weighting, $\mathbf{W}_{\tau,\mathbf{u}}(\mathbf{X})$, functions vary across the values of $\mathbf{X}$, which means that the

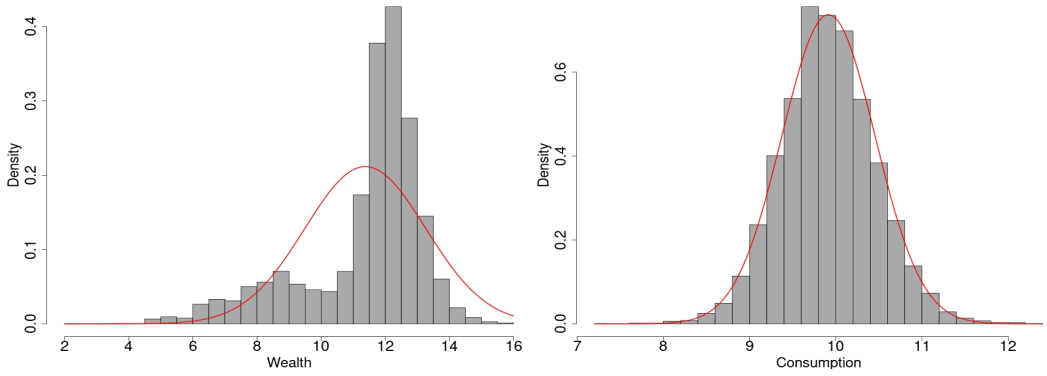

Figure 4.1. Histograms of LWEA (left) and LCON (right) unconditional distributions. The red curves denote the Gaussian densities with mean and standard deviation respectively given by the empirical mean and standard deviation of each outcome.

conditional effects do not average up to their respective unconditional effects. By contrast, the conditional and unconditional models provide similar estimates for age and employment status, for example, suggesting that both functions are relatively constant for all values of $\mathbf{X}$, and that the conditional and unconditional distributions might be similar in this case.

As a means of further comparison, we analyze the log-levels of wealth and consumption with the UQR of Firpo et al. (2009) by fitting two univariate models independently. Table 4.12 reports the corresponding parameter estimates and standard errors at the examined τ levels. We can observe that the results are generally in line with our findings. However, some differences can be identified due to the selected direction $\mathbf{u}$ and the fact that the univariate UQR completely disregards the dependence between wealth and consumption. By contrast, the proposed model allows to study the direction and magnitude of such correlation at different levels τ. In particular, using (4.13) we represent in Tables 4.10 and 4.11 the estimated correlation coefficient, r_{12}, which indicates that consumption and wealth are strongly correlated with each other and this association slightly decreases for households at the upper end of the responses distribution. At $\tau = 0.10$, the estimated coefficients (0.664 and 0.749) suggest that low-consumption households are highly likely to be accompanied by low wealth. From the median to the 90-th percentile instead, the estimates decrease but emphasise that high net worth families tend to report high spending patterns.

We conclude the analysis by providing a graphical representation of the quantile and expectile regions described in Section 4.3 and by evaluating their sensitivity to different choices of c. Formally, we consider 100 equispaced directions in the unit circle to construct unconditional quantile ($c = 0$) and expectile ($c = 100$) contours using (4.8) at $\tau = (0.01, 0.25, 0.40)$ (see Figure 4.2). In the left column, we report quantile (top) and expectile (bottom) contour cuts at the empirical quantile of LINC at level 0.10 (red), 0.50 (blue) and 0.90 (orange). In the right column, meanwhile, we report quantile (top) and expectile (bottom) contour cuts at three education levels (no education (red), high school (blue) and university or higher (orange)).

All plots adapt to the unconditional distribution of the outcomes reasonably well,

Variable	Min	25-th	Mean†	Median	75-th	Max
LWEA	2.303	10.953	11.377	11.939	12.520	15.955
LCON	7.378	9.561	9.915	9.903	10.261	12.234
LINC	0.684	9.775	10.148	10.160	10.594	13.203
Age	16	51	62.258	63	75	101
Sex			38.110			
Marital status						
married			52.099			
never married			18.045			
separated			8.638			
widowed			21.218			
Education level						
no education			3.202			
elementary school			22.414			
middle school			27.878			
vocational school			7.969			
high school			25.943			
university or higher			12.594			
Employment status						
employee			36.559			
self-employed			8.866			
not-employed			54.575			
Geographical area						
north			43.418			
centre			22.129			
south and islands			34.453			

Table 4.9. Descriptive statistics of the outcome variables and covariates. † Means for dummy variables are reported in %.

however, they differ in location, shape and size. The contours for smaller τ capture the effects for the most extreme families i.e., disadvantaged and wealthy families. By contrast, those associated to larger τ capture the effects of more central households. The former are affected by abnormal observations while the latter are less sensitive to outliers and present a smoother surface. One also observes that the contours shift up and to the right for increasing values of LINC, demonstrating the positive dependence with increasing values of the covariate. Moving on to the right-hand side, the plotted contours provide an estimate of returns to schooling on the distribution of wealth and consumption. The different locations and forms of the curves shown in Figure 4.2 indicate that increasing the educational level generates a positive wealth effect and contributes to lifting consumption, and such education premium is even larger for university attainment. These considerations highlight that the proposed contours play an important role in data visualization analysis and are able to detect covariate-dependent features of the responses, while ensuring robustness to outlying data.

	UNCONDITIONAL REGRESSION						CONDITIONAL REGRESSION					
τ	0.10		0.50		0.90		0.10		0.50		0.90	
Variable	W	C	W	C	W	C	W	C	W	C	W	C
Intercept	**−31.954**	**2.907**	**−5.858**	**5.427**	−0.084	**5.931**	**−16.512**	**3.376**	**−6.215**	**5.280**	**−1.341**	**6.136**
	(2.758)	(0.270)	(0.391)	(0.095)	(0.258)	(0.114)	(0.582)	(0.118)	(0.389)	(0.093)	(0.293)	(0.114)
LINC	**3.007**	**0.612**	**1.300**	**0.431**	**0.935**	**0.390**	**2.161**	**0.641**	**1.337**	**0.446**	**1.006**	**0.363**
	(0.203)	(0.021)	(0.032)	(0.008)	(0.021)	(0.009)	(0.048)	(0.010)	(0.032)	(0.008)	(0.024)	(0.009)
Sex	0.066	−0.011	0.010	**−0.025**	−0.008	**−0.031**	0.004	−0.022	0.014	**−0.024**	0.020	**−0.026**
	(0.140)	(0.019)	(0.044)	(0.011)	(0.029)	(0.013)	(0.066)	(0.013)	(0.044)	(0.011)	(0.033)	(0.013)
Age	**0.297**	**0.019**	**0.082**	0.001	**0.044**	−0.001	**0.101**	0.000	**0.081**	0.001	**0.057**	−0.001
	(0.032)	(0.004)	(0.008)	(0.002)	(0.005)	(0.002)	(0.012)	(0.002)	(0.008)	(0.002)	(0.006)	(0.002)
Age2	**−0.002**	**−0.000**	**−0.001**	−0.000	**−0.000**	−0.000	**−0.001**	−0.000	**−0.001**	−0.000	**−0.000**	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Marital status												
never married	0.276	**−0.170**	0.089	**−0.155**	0.019	**−0.167**	**0.322**	**−0.080**	0.101	**−0.151**	0.041	**−0.171**
	(0.180)	(0.025)	(0.056)	(0.014)	(0.037)	(0.016)	(0.084)	(0.017)	(0.056)	(0.013)	(0.042)	(0.016)
separated	**−1.134**	**−0.194**	**−0.318**	**−0.111**	**−0.170**	**−0.099**	**−0.342**	−0.033	**−0.300**	**−0.103**	**−0.208**	**−0.115**
	(0.237)	(0.032)	(0.071)	(0.017)	(0.046)	(0.021)	(0.106)	(0.021)	(0.071)	(0.017)	(0.054)	(0.021)
widowed	0.067	**−0.105**	0.048	**−0.107**	−0.017	**−0.126**	**0.250**	**−0.048**	0.061	**−0.102**	−0.008	**−0.116**
	(0.197)	(0.027)	(0.062)	(0.015)	(0.040)	(0.018)	(0.093)	(0.019)	(0.062)	(0.015)	(0.047)	(0.018)
Education level												
elementary school	0.374	**0.197**	**0.300**	**0.086**	**0.333**	0.014	**0.434**	**0.078**	**0.297**	**0.085**	**0.362**	**0.084**
	(0.349)	(0.049)	(0.111)	(0.027)	(0.072)	(0.032)	(0.165)	(0.033)	(0.110)	(0.026)	(0.083)	(0.032)
middle school	**0.839**	**0.328**	**0.544**	**0.187**	**0.540**	**0.111**	**0.671**	**0.185**	**0.536**	**0.184**	**0.567**	**0.167**
	(0.367)	(0.051)	(0.116)	(0.028)	(0.076)	(0.034)	(0.174)	(0.035)	(0.116)	(0.028)	(0.087)	(0.034)
vocational school	0.762	**0.288**	**0.602**	**0.203**	**0.626**	**0.152**	**0.783**	**0.152**	**0.592**	**0.199**	**0.598**	**0.211**
	(0.409)	(0.057)	(0.130)	(0.031)	(0.085)	(0.038)	(0.194)	(0.039)	(0.129)	(0.031)	(0.098)	(0.038)
high school	**1.722**	**0.440**	**0.917**	**0.276**	**0.817**	**0.210**	**0.928**	**0.221**	**0.899**	**0.269**	**0.879**	**0.260**
	(0.379)	(0.052)	(0.119)	(0.029)	(0.078)	(0.034)	(0.178)	(0.036)	(0.119)	(0.028)	(0.090)	(0.035)
university	**1.347**	**0.445**	**0.969**	**0.364**	**0.961**	**0.338**	**0.811**	**0.257**	**0.937**	**0.351**	**0.963**	**0.349**
	(0.401)	(0.056)	(0.127)	(0.031)	(0.083)	(0.037)	(0.190)	(0.038)	(0.127)	(0.030)	(0.096)	(0.037)
Employment status												
self-employed	**1.915**	**0.102**	**0.946**	0.032	**0.789**	0.039	**0.864**	0.005	**0.942**	0.031	**0.830**	−0.013
	(0.253)	(0.033)	(0.068)	(0.017)	(0.045)	(0.020)	(0.102)	(0.021)	(0.068)	(0.016)	(0.051)	(0.020)
not-employed	**1.763**	**0.160**	**0.667**	**0.042**	**0.502**	**0.042**	**0.690**	**0.050**	**0.669**	**0.043**	**0.596**	0.030
	(0.213)	(0.028)	(0.060)	(0.014)	(0.039)	(0.017)	(0.089)	(0.018)	(0.059)	(0.014)	(0.045)	(0.017)
Geographical area												
centre	**0.663**	**0.078**	**0.210**	**0.023**	**0.112**	0.010	**0.286**	**0.029**	**0.213**	**0.024**	**0.145**	0.004
	(0.154)	(0.021)	(0.047)	(0.011)	(0.031)	(0.014)	(0.071)	(0.014)	(0.047)	(0.011)	(0.036)	(0.014)
south and islands	**0.845**	−0.020	**0.190**	**−0.104**	0.044	**−0.144**	**0.348**	**−0.051**	**0.202**	**−0.099**	**0.127**	**−0.143**
	(0.153)	(0.021)	(0.044)	(0.011)	(0.028)	(0.013)	(0.065)	(0.013)	(0.043)	(0.010)	(0.033)	(0.013)
r_{12}	0.664		0.476		0.416							

Table 4.10. Unconditional and conditional regression coefficient estimates at the investigated τ levels and direction $\mathbf{u} = (1, -1)'$, using the optimal tuning constant $c = c^\star$. Parameter estimates are displayed in boldface when significant at the standard 5% level.

4.6 Conclusions

Extending the univariate work of Firpo et al. (2009), this paper proposes a unified approach to model the entire unconditional distribution of a multivariate response variable in a regression setting. We make several contributions to the literature. First, by employing the multidimensional Huber's function in Hampel et al. (2011) we are able to build a comprehensive modeling framework to estimate multivariate unconditional quantiles, M-quantiles and expectiles, choosing the tuning constant in an appropriate manner. Second, in contrast to univariate methods, our multivariate model accounts for the, potentially asymmetric, association structure between the outcome variables. Third, the proposed methodology is easy to implement through an OLS regression of the RIF on the explanatory variables. From a theoretical standpoint, we show that the introduced estimators are consistent, asymptotically normal and can be written as a weighted average of conditional effects. In addition, we propose a data-driven procedure based on cross-validation to select the optimal tuning constant for estimating the UMQPE. Finally, we contribute to the empirical literature by analyzing log-levels of wealth and consumption of Italian households

	UNCONDITIONAL REGRESSION						CONDITIONAL REGRESSION					
τ	0.10		0.50		0.90		0.10		0.50		0.90	
Variable	W	C	W	C	W	C	W	C	W	C	W	C
Intercept	**−22.249**	**1.612**	**2.842**	**5.848**	**4.832**	**6.033**	**−16.625**	**3.342**	**−5.114**	**3.825**	**1.335**	**4.477**
	(7.463)	(0.821)	(0.255)	(0.132)	(0.262)	(0.163)	(0.524)	(0.133)	(0.263)	(0.094)	(0.212)	(0.128)
LINC	**2.336**	**0.705**	**0.678**	**0.384**	**0.596**	**0.372**	**2.165**	**0.647**	**1.276**	**0.597**	**0.891**	**0.522**
	(0.534)	(0.060)	(0.021)	(0.011)	(0.021)	(0.013)	(0.043)	(0.011)	(0.021)	(0.008)	(0.017)	(0.010)
Sex	0.040	−0.011	−0.010	**−0.042**	−0.034	−0.032	0.009	−0.005	−0.019	−0.019	0.003	−0.012
	(0.108)	(0.023)	(0.027)	(0.015)	(0.028)	(0.018)	(0.059)	(0.015)	(0.030)	(0.011)	(0.024)	(0.015)
Age	**0.228**	**0.028**	**0.040**	**0.006**	**0.021**	0.003	**0.100**	−0.000	**0.088**	−0.002	**0.030**	0.000
	(0.055)	(0.007)	(0.005)	(0.003)	(0.005)	(0.003)	(0.011)	(0.003)	(0.005)	(0.002)	(0.004)	(0.003)
Age2	**−0.002**	**−0.000**	**−0.000**	**−0.000**	**−0.000**	−0.000	**−0.001**	−0.000	**−0.001**	0.000	**−0.000**	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Marital status												
never married	0.191	**−0.165**	−0.052	**−0.182**	−0.064	**−0.186**	**0.323**	**−0.080**	**0.172**	**−0.116**	0.035	**−0.135**
	(0.146)	(0.031)	(0.035)	(0.019)	(0.035)	(0.023)	(0.076)	(0.019)	(0.038)	(0.014)	(0.031)	(0.018)
separated	**−0.873**	**−0.232**	**−0.183**	**−0.126**	**−0.095**	**−0.119**	**−0.370**	−0.046	**−0.113**	**−0.038**	−0.038	**−0.070**
	(0.271)	(0.042)	(0.044)	(0.024)	(0.045)	(0.029)	(0.096)	(0.024)	(0.048)	(0.017)	(0.039)	(0.023)
widowed	0.042	**−0.108**	**−0.107**	**−0.161**	**−0.130**	**−0.169**	**0.233**	**−0.068**	**0.190**	**−0.057**	0.019	**−0.090**
	(0.154)	(0.032)	(0.038)	(0.021)	(0.039)	(0.025)	(0.083)	(0.021)	(0.042)	(0.015)	(0.034)	(0.020)
Education level												
elementary school	0.274	**0.209**	**0.349**	0.007	**0.228**	−0.077	**0.512**	0.062	**0.322**	**0.068**	**0.363**	0.048
	(0.271)	(0.057)	(0.069)	(0.038)	(0.069)	(0.045)	(0.148)	(0.038)	(0.074)	(0.027)	(0.060)	(0.036)
middle school	**0.641**	**0.352**	**0.543**	**0.155**	**0.426**	0.035	**0.758**	**0.175**	**0.343**	**0.137**	**0.471**	**0.110**
	(0.299)	(0.061)	(0.074)	(0.040)	(0.073)	(0.048)	(0.156)	(0.040)	(0.078)	(0.028)	(0.063)	(0.038)
vocational school	0.578	**0.309**	**0.632**	**0.201**	**0.587**	**0.120**	**0.873**	**0.145**	**0.427**	**0.153**	**0.509**	**0.158**
	(0.326)	(0.068)	(0.081)	(0.044)	(0.082)	(0.053)	(0.174)	(0.044)	(0.087)	(0.031)	(0.070)	(0.043)
high school	**1.335**	**0.496**	**0.822**	**0.281**	**0.670**	**0.166**	**1.006**	**0.213**	**0.564**	**0.189**	**0.655**	**0.173**
	(0.384)	(0.066)	(0.075)	(0.041)	(0.075)	(0.049)	(0.160)	(0.041)	(0.080)	(0.029)	(0.065)	(0.039)
university	**1.051**	**0.486**	**0.847**	**0.315**	**0.843**	**0.277**	**0.897**	**0.238**	**0.575**	**0.225**	**0.737**	**0.213**
	(0.353)	(0.067)	(0.079)	(0.043)	(0.080)	(0.052)	(0.171)	(0.043)	(0.085)	(0.031)	(0.069)	(0.042)
Employment status												
self-employed	**1.452**	**0.159**	**0.495**	0.001	**0.621**	**0.059**	**0.861**	0.008	**0.610**	0.018	**0.625**	−0.010
	(0.373)	(0.053)	(0.042)	(0.023)	(0.045)	(0.028)	(0.092)	(0.023)	(0.046)	(0.017)	(0.037)	(0.022)
not-employed	**1.355**	**0.219**	**0.425**	0.036	**0.378**	**0.055**	**0.687**	**0.054**	**0.321**	0.025	**0.434**	**0.058**
	(0.336)	(0.044)	(0.037)	(0.020)	(0.038)	(0.024)	(0.080)	(0.020)	(0.040)	(0.014)	(0.032)	(0.020)
Geographical area												
centre	**0.505**	**0.096**	**0.120**	**0.034**	**0.074**	**0.039**	**0.291**	0.017	**0.154**	0.021	**0.056**	−0.004
	(0.156)	(0.028)	(0.029)	(0.016)	(0.030)	(0.019)	(0.064)	(0.016)	(0.032)	(0.011)	(0.026)	(0.016)
south and islands	**0.633**	0.004	−0.035	**−0.137**	**−0.111**	**−0.172**	**0.351**	**−0.053**	**0.186**	**−0.049**	**0.048**	**−0.087**
	(0.189)	(0.030)	(0.027)	(0.015)	(0.027)	(0.018)	(0.059)	(0.015)	(0.029)	(0.011)	(0.024)	(0.014)
r_{12}	0.749		0.546		0.499							

Table 4.11. Unconditional and conditional regression coefficient estimates at the investigated τ levels and direction $\mathbf{u} = (1, -1)'$, using $c = 0$. Parameter estimates are displayed in boldface when significant at the standard 5% level.

collected in the SHIW 2016 data.

This approach can be further extended to a time-varying setting. Specifically, one can exploit the panel dimension of the SHIW and build a panel dataset to model the unconditional distribution of macroeconomic variables as a function of socio-economic and demographic household characteristics over time.

Lastly, the proposed methodology can be further generalized to multivariate longitudinal or clustered observations. Particularly, in order to account for the within-cluster dependence and between subject heterogeneity, we may consider the proposed unconditional regression approach with random effects to model the conditional expectation of the RIF in (4.14) to provide correct inferences. In this case, our object of interest is to identify how changes in the distribution of the explanatory variables affect the unconditional distribution of the multivariate response, given some random group-effects capturing this unobserved heterogeneity. Hence, by proceeding as in Section 4.2 and from (4.15), one might estimate the conditional expectation of the RIF given the covariates and the random effects using mixed effect models targeting the conditional mean of the response. The estimation of the model parameters can be obtained by ML or Restricted ML methods with either a

τ	0.10		0.50		0.90	
Variable	W	C	W	C	W	C
Intercept	**−25.146**	**4.856**	**3.054**	**5.582**	**5.183**	**6.218**
	(2.370)	(0.159)	(0.195)	(0.100)	(0.269)	(0.165)
LINC	**2.575**	**0.440**	**0.668**	**0.401**	**0.655**	**0.426**
	(0.179)	(0.015)	(0.014)	(0.009)	(0.022)	(0.017)
Sex	−0.003	**−0.026**	**−0.025**	**−0.049**	**−0.049**	0.002
	(0.021)	(0.006)	(0.003)	(0.002)	(0.005)	(0.003)
Age	**0.154**	**−0.018**	**0.036**	**0.007**	**0.017**	**0.008**
	(0.014)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Age2	**−0.001**	**0.000**	**−0.000**	**−0.000**	**−0.000**	**−0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Marital status						
never married	**0.582**	**−0.137**	**−0.057**	**−0.163**	**−0.065**	**−0.208**
	(0.038)	(0.007)	(0.006)	(0.005)	(0.006)	(0.008)
separated	**−0.476**	**−0.063**	**−0.164**	**−0.095**	**−0.030**	**−0.148**
	(0.047)	(0.008)	(0.006)	(0.003)	(0.009)	(0.007)
widowed	**0.577**	**−0.083**	**−0.100**	**−0.121**	**−0.036**	**−0.072**
	(0.047)	(0.005)	(0.005)	(0.005)	(0.009)	(0.008)
Education level						
elementary school	**−0.191**	**0.448**	**0.356**	**0.055**	**0.019**	**−0.166**
	(0.060)	(0.041)	(0.014)	(0.007)	(0.004)	(0.007)
middle school	**0.188**	**0.600**	**0.577**	**0.175**	**0.214**	**−0.124**
	(0.062)	(0.052)	(0.016)	(0.006)	(0.013)	(0.007)
vocational school	**0.401**	**0.581**	**0.650**	**0.224**	**0.273**	**−0.131**
	(0.041)	(0.054)	(0.014)	(0.010)	(0.012)	(0.009)
high school	**0.876**	**0.594**	**0.863**	**0.299**	**0.454**	0.004
	(0.064)	(0.054)	(0.022)	(0.008)	(0.025)	(0.006)
university	**0.659**	**0.535**	**0.878**	**0.329**	**1.028**	**0.334**
	(0.068)	(0.051)	(0.018)	(0.008)	(0.034)	(0.014)
Employment status						
self-employed	**1.479**	**−0.060**	**0.490**	**0.011**	**0.948**	**0.143**
	(0.119)	(0.008)	(0.010)	(0.001)	(0.040)	(0.007)
not-employed	**0.929**	**0.052**	**0.404**	**0.035**	**0.407**	**0.088**
	(0.075)	(0.004)	(0.009)	(0.004)	(0.015)	(0.006)
Geographical area						
centre	**0.446**	**0.046**	**0.115**	**0.046**	**−0.083**	**−0.040**
	(0.045)	(0.002)	(0.004)	(0.003)	(0.008)	(0.003)
south and islands	**0.660**	**−0.076**	**−0.066**	**−0.104**	**−0.126**	**−0.115**
	(0.062)	(0.006)	(0.010)	(0.003)	(0.006)	(0.004)

Table 4.12. Univariate URQ coefficient estimates at the investigated τ levels. Standard errors are computed via nonparametric bootstrap using 1000 resamples and parameter estimates are displayed in boldface when significant at the standard 5% level.

parametric specification of the random effects distribution or, similarly to Chapter 1, a nonparametric specification based on NPML (Laird 1978) to further robustify the UMQPE to model misspecification.

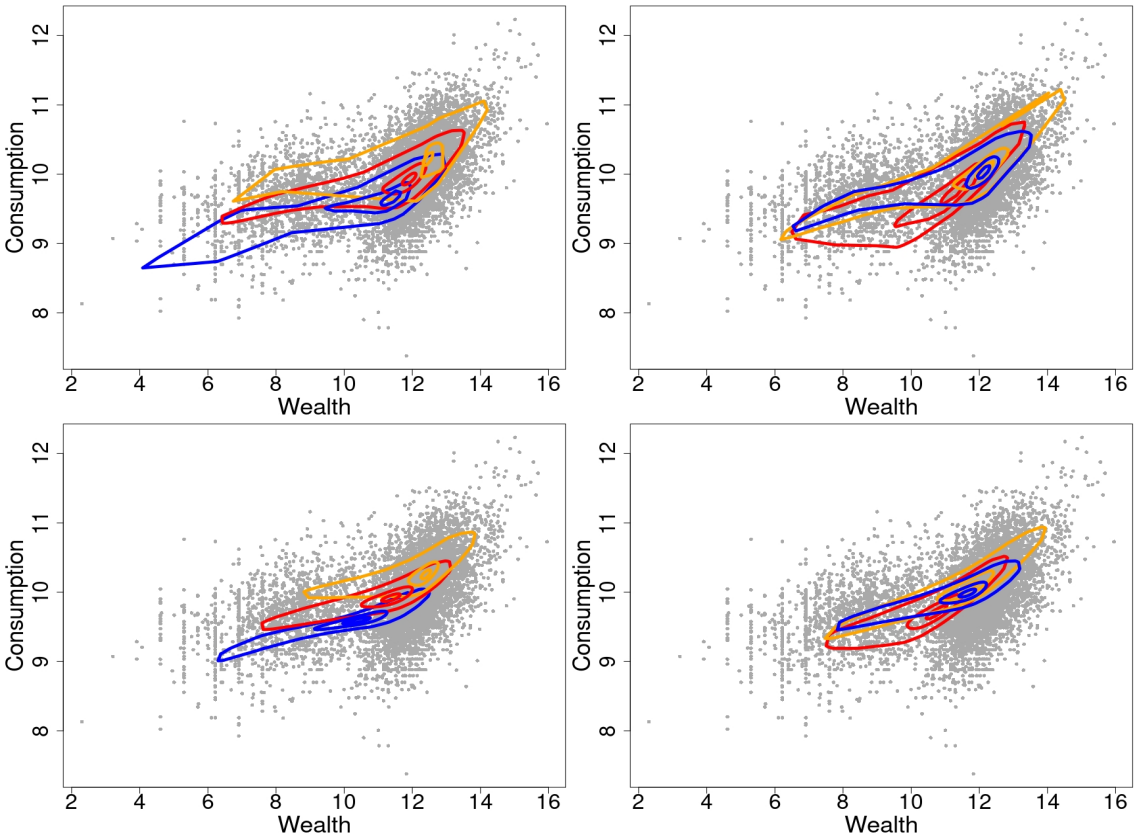

Figure 4.2. Unconditional contours at $\tau = (0.01, 0.25, 0.40)$ (from the outside inwards). In the left column, quantile (top) and expectile (bottom) contour cuts at the empirical quantile of LINC at level 0.10 (red), 0.50 (blue) and 0.90 (orange). In the right column, quantile (top) and expectile (bottom) contour cuts at three education levels: no education (red), high school (blue) and university or higher (orange).

4.7 Appendix

Proof of Theorem 3

Proof. For fixed level τ and direction $\mathbf{u}$, from (4.15) it follows that:

$$\boldsymbol{\alpha}_{\tau,\mathbf{u}} = \int \frac{d}{d\mathbf{x}} \Big(\mathbb{E}[\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1} \eta_\delta(\varphi) \Psi(\mathbf{Y} - \boldsymbol{\theta}_{\tau,\mathbf{u}}) \mid \mathbf{X} = \mathbf{x}] \Big) dF_{\mathbf{X}}(\mathbf{x}), \tag{4.38}$$

with $\mathbf{Y} = h(\mathbf{X}) + \epsilon$. Taking the derivative with respect to $\mathbf{x}$ and under the assumption of independence between ϵ and $\mathbf{X}$ we obtain:

$$\boldsymbol{\alpha}_{\tau,\mathbf{u}} = \int \mathbb{E}[\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1} \nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}} (\eta_\delta(\varphi) \Psi(\mathbf{Y} - \boldsymbol{\theta}_{\tau,\mathbf{u}})) \mid \mathbf{X} = \mathbf{x}] \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}} dF_{\mathbf{X}}(\mathbf{x}). \tag{4.39}$$

We denote with $\mathbf{W}_{\tau,\mathbf{u}}(\mathbf{X})$ the conditional expectation in (4.39). Now, let $\boldsymbol{\theta}_{\tilde{\tau},\mathbf{u}}(\mathbf{x})$ denote the multivariate M-quantile of the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$, at level $\tilde{\tau}$, with $\tilde{\tau} \in (0, \frac{1}{2}]$ in the direction $\mathbf{u}$. Then, by the independence assumption between ϵ and $\mathbf{X}$,

$$\boldsymbol{\alpha}_{\tilde{\tau},\mathbf{u}}(\mathbf{x}) = \frac{\partial h(\mathbf{X})}{\partial \mathbf{x}}. \tag{4.40}$$

Finally, by letting $s_\tau(\mathbf{x}) = \{\tilde{\tau} : \boldsymbol{\theta}_{\tilde{\tau},\mathbf{u}}(\mathbf{x}) = \boldsymbol{\theta}_{\tau,\mathbf{u}}\}$, we conclude that:

$$\boldsymbol{\alpha}_{\tau,\mathbf{u}} = \mathbb{E}[\mathbf{W}_{\tau,\mathbf{u}}(\mathbf{X})\, \boldsymbol{\alpha}_{s_\tau(\mathbf{X}),\mathbf{u}}(\mathbf{X})]. \tag{4.41}$$

□

Proof of Theorem 4

Proof. By assumptions (A1)-(A5), the expectation defining $\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$ exist finitely for $p \geq 2$ and $\mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$ is positive definite. Moreover, assumptions (A1)-(A5) guarantee that $\Psi(\cdot)$ possesses bounded derivatives up to the second order. Then, Theorem 4 of Niemiro et al. (1992) applies. Finally, the asymptotic normality of the multivariate M-quantile follows from the Slutsky's Theorem and the multivariate Central Limit Theorem, which completes the proof. □

Proof of Theorem 5

Proof. We express the difference $\widehat{\boldsymbol{\alpha}}^\star_{\tau,\mathbf{u}} - \boldsymbol{\alpha}^\star_{\tau,\mathbf{u}}$ in the following manner and study each term of the sum separately:

$$\begin{aligned}
&\widehat{\boldsymbol{\alpha}}^\star_{\tau,\mathbf{u}} - \boldsymbol{\alpha}^\star_{\tau,\mathbf{u}} && (4.42)\\
&= \widehat{\boldsymbol{\beta}}^\star(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) - \widehat{\boldsymbol{\beta}}^\star(\boldsymbol{\theta}_{\tau,\mathbf{u}}) && (4.43)\\
&+ \widehat{\boldsymbol{\beta}}^\star(\boldsymbol{\theta}_{\tau,\mathbf{u}}) - \boldsymbol{\beta}^\star(\boldsymbol{\theta}_{\tau,\mathbf{u}}) && (4.44)\\
&+ \boldsymbol{\beta}^\star(\boldsymbol{\theta}_{\tau,\mathbf{u}}) - \boldsymbol{\alpha}^\star_{\tau,\mathbf{u}}. && (4.45)
\end{aligned}$$

Consider the term (4.43). An expression for the first term $\widehat{\boldsymbol{\beta}}^\star(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) - \widehat{\boldsymbol{\beta}}^\star(\boldsymbol{\theta}_{\tau,\mathbf{u}})$, is derived in the following expression using a first-order Taylor expansion:

$$\widehat{\boldsymbol{\beta}}^\star(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) - \widehat{\boldsymbol{\beta}}^\star(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}} \boldsymbol{\beta}^\star(\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}} - \boldsymbol{\theta}_{\tau,\mathbf{u}}) + O_p(|| \widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}} - \boldsymbol{\theta}_{\tau,\mathbf{u}} ||^2) + R, \tag{4.46}$$

where $R = (\nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}} \widehat{\boldsymbol{\beta}}^{\star}(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}} \boldsymbol{\beta}^{\star}(\boldsymbol{\theta}))(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}} - \boldsymbol{\theta}_{\tau,\mathbf{u}})$. By Theorem 4, the multivariate M-quantile estimator $\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}$ is $\sqrt{n}$-consistent, hence $O_p(|| \widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}} - \boldsymbol{\theta}_{\tau,\mathbf{u}} ||^2) = O_p(\frac{1}{n})$ and $R = O_p(\frac{1}{n})$.

Now we turn our attention to the second term (4.44), $\widehat{\boldsymbol{\beta}}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) - \boldsymbol{\beta}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$, and consider the following linear asymptotic representation of the OLS estimator $\widehat{\boldsymbol{\beta}}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$:

$$\widehat{\boldsymbol{\beta}}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) - \boldsymbol{\beta}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\gamma}_i^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) + o_p(\frac{1}{\sqrt{n}}). \tag{4.47}$$

Finally, we consider the bias term (4.45), $\boldsymbol{\beta}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) - \boldsymbol{\alpha}^{\star}_{\tau,\mathbf{u}}$. Under the condition that $\mathbb{E}[\mathbf{z} \mid \mathbf{X}] = \mathbf{0}$, we have that $\mathbb{E}[IF(\mathbf{Y}; \boldsymbol{\theta}_{\tau,\mathbf{u}})] \mid \mathbf{X}] = \boldsymbol{\beta}'(\boldsymbol{\theta}_{\tau,\mathbf{u}})\mathbf{X}$. Then, from (4.15) it follows that $\frac{d\mathbb{E}[IF(\mathbf{Y};\boldsymbol{\theta}_{\tau,\mathbf{u}})|\mathbf{X}=\mathbf{x}]}{d\mathbf{x}} = \boldsymbol{\beta}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$, hence $\boldsymbol{\beta}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \boldsymbol{\alpha}_{\tau,\mathbf{u}}$.

A combination of previous results yields:

$$\widehat{\boldsymbol{\beta}}^{\star}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) - \boldsymbol{\beta}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{S}_i + o_p(\frac{1}{\sqrt{n}}), \tag{4.48}$$

where $\mathbf{S}_i = \nabla_{\boldsymbol{\theta}_{\tau,\mathbf{u}}} \boldsymbol{\beta}^{\star}(\boldsymbol{\theta}) \mathbf{M}(\boldsymbol{\theta}_{\tau,\mathbf{u}})^{-1} \eta_{\delta}(\varphi_i) \Psi(\mathbf{Y}_i - \boldsymbol{\theta}_{\tau,\mathbf{u}}) + \boldsymbol{\gamma}_i^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$.

Finally, asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\alpha}}^{\star}_{\tau,\mathbf{u}} - \boldsymbol{\alpha}^{\star}_{\tau,\mathbf{u}})$ follows from the Slutsky's Theorem and the multivariate Central Limit Theorem. □

Proof of Theorem 6

Proof. By assumptions (A1)-(A6) and Theorems 4 and 5, it follows that $\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}} \xrightarrow{p} \boldsymbol{\theta}_{\tau,\mathbf{u}}$ and $\widehat{\boldsymbol{\beta}}^{\star}(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) \xrightarrow{p} \boldsymbol{\beta}^{\star}(\boldsymbol{\theta}_{\tau,\mathbf{u}})$. Subsequently, from the continuous mapping theorem we have that $\frac{1}{n}\sum_{i=1}^{n} \widehat{\mathbf{S}}_i(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) \widehat{\mathbf{S}}_i'(\widehat{\boldsymbol{\theta}}_{\tau,\mathbf{u}}) \xrightarrow{p} \mathbb{E}\Big[\mathbf{S}(\boldsymbol{\theta}_{\tau,\mathbf{u}}) \mathbf{S}'(\boldsymbol{\theta}_{\tau,\mathbf{u}})\Big]$. □

Bibliography

Acerbi, C. & Tasche, D. (2002), 'On the coherence of Expected Shortfall', *Journal of Banking & Finance* **26**(7), 1487–1503.

Agresti, A., Caffo, B. & Ohman-Strickland, P. (2004), 'Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies', *Computational Statistics & Data Analysis* **47**(3), 639–653.

Ahn, J. V., Sera, F., Cummins, S. & Flouri, E. (2018), 'Associations between objectively measured physical activity and later mental health outcomes in children: findings from the UK Millennium Cohort Study', *Journal of Epidemiology and Community Health* **72**(2), 94–100.

Aitkin, M. & Alfó, M. (1998), 'Regression models for binary longitudinal responses', *Statistics and Computing* **8**(4), 289–307.

Aitkin, M. & Alfò, M. (2003), 'Longitudinal analysis of repeated binary data using autoregressive and random effect modelling', *Statistical Modelling* **3**(4), 291–303.

Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, *in* 'Selected papers of Hirotugu Akaike', Springer, pp. 199–213.

Alexander, G. J. & Baptista, A. M. (2008), 'Active portfolio management with benchmarking: Adding a Value-at-Risk constraint', *Journal of Economic Dynamics and Control* **32**(3), 779 – 820.

Alfò, M. & Aitkin, M. (2000), 'Random coefficient models for binary longitudinal responses with attrition', *Statistics and Computing* **10**(4), 279–287.

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. & Maruotti, A. (2010), 'Two-part regression models for longitudinal zero-inflated count data', *Canadian Journal of Statistics* **38**(2), 197–216.

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

Altman, R. M. (2007), 'Mixed Hidden Markov models: an extension of the Hidden Markov model to the longitudinal data setting', *Journal of the American Statistical Association* **102**(477), 201–210.

Arditti, F. D. (1971), 'Another look at mutual fund performance', *Journal of Financial and Quantitative Analysis* **6**(3), 909–912.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1997), 'Thinking coherently', *Risk* **10**(11), 68–72.

Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), 'Coherent measures of risk', *Mathematical Finance* **9**(3), 203–228.

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

Barone-Adesi, G. (1985), 'Arbitrage equilibrium with skewed asset returns', *Journal of Financial and Quantitative Analysis* **20**(3), 299–313.

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

Bartolucci, F. & Farcomeni, A. (2009), 'A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure', *Journal of the American Statistical Association* **104**(486), 816–831.

Bartolucci, F., Farcomeni, A. & Pennoni, F. (2012), *Latent Markov models for longitudinal data*, CRC Press.

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

Bassett, G. W., Koenker, R. & Kordas, G. (2004), 'Pessimistic portfolio allocation and choquet expected utility', *Journal of Financial Econometrics* **2**(4), 477–492.

Baur, D. G. (2013), 'The structure and degree of dependence: A quantile regression approach', *Journal of Banking & Finance* **37**(3), 786–798.

Bauwens, L. & Laurent, S. (2005), 'A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models', *Journal of Business & Economic Statistics* **23**(3), 346–354.

Bell, S. L., Audrey, S., Gunnell, D., Cooper, A. & Campbell, R. (2019), 'The relationship between physical activity, mental wellbeing and symptoms of mental health disorder in adolescents: a cohort study', *International Journal of Behavioral Nutrition and Physical Activity* **16**(1), 138.

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

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

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.

Böhning, D. (1999), *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others*, Vol. 81, CRC press.

Bonaccolto, G., Caporin, M. & Paterlini, S. (2019), 'Decomposing and backtesting a flexible specification for CoVaR', *Journal of Banking & Finance* **108**, 105659.

Bonanno, A., Bimbo, F., Cleary, R. & Castellari, E. (2018), 'Food labels and adult BMI in Italy–an unconditional quantile regression approach', *Food Policy* **74**, 199–211.

Borah, B. J. & Basu, A. (2013), 'Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence', *Health Economics* **22**(9), 1052–1070.

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.

Breckling, J., Kokic, P. & Lübke, O. (2001), 'A note on multivariate M-quantiles', *Statistics & Probability Letters* **55**(1), 39–44.

Bu, D., Liao, Y., Shi, J. & Peng, H. (2019), 'Dynamic Expected Shortfall: A spectral decomposition of tail risk across time horizons', *Journal of Economic Dynamics and Control* **108**, 103753.

Cai, Z. & Wang, X. (2008), 'Nonparametric estimation of conditional VaR and Expected Shortfall', *Journal of Econometrics* **147**(1), 120–130.

Campbell, J. Y. (1993), 'Intertemporal asset pricing without consumption data', *The American Economic Review* **83**(3), 487–512.

Campbell, J. Y. & Mankiw, N. G. (1989), 'Consumption, income, and interest rates: Reinterpreting the time series evidence', *NBER macroeconomics annual* **4**, 185–216.

Cappé, O., Moulines, E. & Rydén, T. (2006), *Inference in hidden Markov models*, Springer Science & Business Media.

Carona, C., Moreira, H., Silva, N., Crespo, C. & Canavarro, M. C. (2014), 'Social support and adaptation outcomes in children and adolescents with cerebral palsy', *Disability and Rehabilitation* **36**(7), 584–592.

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

Chakraborty, B. (2003), 'On multivariate quantile regression', *Journal of Statistical Planning and Inference* **110**(1-2), 109–132.

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.

Charlier, I., Paindaveine, D. & Saracco, J. (2020), 'Multiple-output quantile regression through optimal quantization', *Scandinavian Journal of Statistics* **47**(1), 250–278.

Chavas, J.-P. (2018), 'On multivariate quantile regression analysis', *Statistical Methods & Applications* **27**(3), 365–384.

Chen, J. & Lazar, N. A. (2012), 'Selection of working correlation structure in generalized estimating equations via empirical likelihood', *Journal of Computational and Graphical Statistics* **21**(1), 18–41.

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.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**(4), 841–862.

Cicchetti, D. & Toth, S. L. (2014), 'A developmental perspective on internalizing and externalizing disorders', *Internalizing and Externalizing Expression of Dysfunction* pp. 1–19.

Cole, T. J. & Green, P. J. (1992), 'Smoothing reference centile curves: the LMS method and penalized likelihood', *Statistics in Medicine* **11**(10), 1305–1319.

Costello, E. J., Compton, S. N., Keeler, G. & Angold, A. (2003), 'Relationships between poverty and psychopathology: A natural experiment', *Journal of the American Medical Association* **290**(15), 2023–2029.

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

Crowther, M. J., Look, M. P. & Riley, R. D. (2014), 'Multilevel mixed effects parametric survival models using adaptive Gauss–Hermite quadrature with application to recurrent events and individual participant data meta-analysis', *Statistics in Medicine* **33**(22), 3844–3858.

Currie, J. (2009), 'Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development', *Journal of Economic Literature* **47**(1), 87–122.

Davino, C., Furno, M. & Vistocco, D. (2013), *Quantile regression: theory and applications*, Vol. 988, John Wiley & Sons.

Dawber, J., Salvati, N., Schmid, T. & Tzavidis, N. (2020), 'Scale estimation and data-driven tuning constant selection for M-quantile regression', *arXiv preprint arXiv:2011.10522* .

Dearing, E., McCartney, K. & Taylor, B. A. (2006), 'Within-child associations between family income and externalizing and internalizing problems.', *Developmental Psychology* **42**(2), 237.

Deaton, A. et al. (1992), *Understanding consumption*, Oxford University Press.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–38.

Diebold, F. X. & Mariano, R. S. (2002), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **20**(1), 134–144.

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.

Dong, X., Li, C. & Yoon, S.-M. (2020), 'Asymmetric dependence structures for regional stock markets: An unconditional quantile regression approach', *The North American Journal of Economics and Finance* **52**, 101111.

Du, Z. & Escanciano, J. C. (2017), 'Backtesting Expected Shortfall: accounting for tail risk', *Management Science* **63**(4), 940–958.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business & Economic Statistics* **20**(3), 339–350.

Engle, R. F. & Manganelli, S. (2004), 'CAViaR: conditional autoregressive value at risk by regression quantiles', *Journal of Business & Economic Statistics* **22**(4), 367–381.

Engle, R. F. & Ng, V. K. (1993), 'Measuring and testing the impact of news on volatility', *The Journal of Finance* **48**(5), 1749–1778.

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* .

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

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.

Firpo, S., Fortin, N. M. & Lemieux, T. (2009), 'Unconditional quantile regressions', *Econometrica: Journal of the Econometric Society* **77**(3), 953–973.

Firpo, S. P., Fortin, N. M. & Lemieux, T. (2018), 'Decomposing wage distributions using recentered influence function regressions', *Econometrics* **6**(2), 28.

Fissler, T. & Ziegel, J. F. (2016), 'Higher order elicitability and Osband's principle', *The Annals of Statistics* **44**(4), 1680–1707.

Fissler, T., Ziegel, J. F. & Gneiting, T. (2015), 'Expected Shortfall is jointly elicitable with Value at Risk - implications for backtesting', *arXiv preprint arXiv:1507.00244* .

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

Fitzsimons, E., Goodman, A., Kelly, E. & Smith, J. P. (2017), 'Poverty dynamics and parental mental health: Determinants of childhood mental health in the UK', *Social Science & Medicine* **175**, 43–51.

Flouri, E. & Sarmadi, Z. (2016), 'Prosocial behavior and childhood trajectories of internalizing and externalizing problems: The role of neighborhood and school contexts', *Developmental Psychology* **52**(2), 253.

Friend, I. & Westerfield, R. (1980), 'Co-skewness and capital asset pricing', *The Journal of Finance* **35**(4), 897–913.

Frölich, M. & Melly, B. (2013), 'Unconditional quantile treatment effects under endogeneity', *Journal of Business & Economic Statistics* **31**(3), 346–357.

Frongillo, R. & Kash, I. A. (2015), Vector-Valued property elicitation, *in* 'Conference on Learning Theory', pp. 710–727.

Fu, L., Hao, Y. & Wang, Y.-G. (2018), 'Working correlation structure selection in generalized estimating equations', *Computational Statistics* **33**(2), 983–996.

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

Furno, M. & Vistocco, D. (2018), *Quantile regression: Estimation and simulation*, Vol. 216, John Wiley & Sons.

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

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.

Gneiting, T. (2011), 'Making and evaluating point forecasts', *Journal of the American Statistical Association* **106**(494), 746–762.

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

Goodman, A. & Goodman, R. (2009), 'Strengths and difficulties questionnaire as a dimensional measure of child mental health', *Journal of the American Academy of Child & Adolescent Psychiatry* **48**(4), 400–403.

Goodman, A. & Goodman, R. (2011), 'Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in britain', *Journal of Child Psychology and Psychiatry* **52**(1), 100–108.

Goodman, R. (1997), 'The Strengths and Difficulties Questionnaire: a research note', *Journal of Child Psychology and Psychiatry* **38**(5), 581–586.

Gosho, M., Hamada, C. & Yoshimura, I. (2014), 'Selection of working correlation structure in weighted generalized estimating equation method for incomplete longitudinal data', *Communications in Statistics-Simulation and Computation* **43**(1), 62–81.

Gosling, A., Machin, S. & Meghir, C. (2000), 'The changing distribution of male wages in the uk', *The Review of Economic Studies* **67**(4), 635–666.

Gourieroux, C., Liu, W. & Liu, G. (2012), 'Converting Tail-VaR to VaR: An econometric study', *Journal of Financial Econometrics* **10**(2), 233–264.

Griffiths, L. J., Dezateux, C. & Hill, A. (2011), 'Is obesity associated with emotional and behavioural problems in children? findings from the millennium cohort study', *International Journal of Pediatric Obesity* **6**(sup3), e423–432.

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.

Hamilton, J. D. (2020), *Time series analysis*, Princeton University Press.

Hampel, F. R. (1974), 'The influence curve and its role in robust estimation', *Journal of the American Statistical Association* **69**(346), 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (2011), *Robust statistics: the approach based on influence functions*, Vol. 196, John Wiley & Sons.

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

Huffman, M. L., King, J. & Reichelt, M. (2017), 'Equality for whom? Organizational policies and the gender gap across the German earnings distribution', *ILR Review* **70**(1), 16–41.

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.

Kiernan, K. E. & Huerta, M. C. (2008), 'Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood', *The British Journal of Sociology* **59**(4), 783–806.

Killewald, A. & Bearak, J. (2014), 'Is the motherhood penalty larger for low-wage women? a comment on quantile regression', *American Sociological Review* **79**(2), 350–357.

Koenker, R. (2004), 'Quantile regression for longitudinal data', *Journal of Multivariate Analysis* **91**(1), 74–89.

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

Koenker, R. & Bassett, G. (1978), 'Regression Quantiles', *Econometrica: Journal of the Econometric Society* **46**(1), 33–50.

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

Koenker, R. & Xiao, Z. (2006), 'Quantile autoregression', *Journal of the American Statistical Association* **101**(475), 980–990.

Kokic, P., Breckling, J. & Lübke, O. (2002), A new definition of multivariate M-quantiles, *in* 'Statistical data analysis based on the L_1-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.

Konno, H. & Suzuki, K.-i. (1995), 'A Mean-Variance-Skewness portfolio optimization model', *Journal of the Operations Research Society of Japan* **38**(2), 173–187.

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

Kotz, S., Kozubowski, T. & Podgorski, K. (2012), *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*, Springer Science & Business Media.

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

Kraus, A. & Litzenberger, R. H. (1976), 'Skewness preference and the valuation of risk assets', *The Journal of Finance* **31**(4), 1085–1100.

Kraus, D. & Czado, C. (2017), 'D-vine copula based quantile regression', *Computational Statistics & Data Analysis* **110**, 1–18.

Kulkarni, H., Biswas, J. & Das, K. (2019), 'A joint quantile regression model for multiple longitudinal outcomes', *AStA Advances in Statistical Analysis* **103**(4), 453–473.

Kupiec, P. H. (1995), 'Techniques for verifying the accuracy of risk measurement models', *The Journal of Derivatives* **3**(2), 73–84.

Laird, N. (1978), 'Nonparametric maximum likelihood estimation of a mixing distribution', *Journal of the American Statistical Association* **73**(364), 805–811.

Laporta, A. G., Merlo, L. & Petrella, L. (2018), 'Selection of Value at Risk models for energy commodities', *Energy Economics* **74**, 628–643.

Li, Y. (2005), 'The wealth-consumption ratio and the consumption-habit ratio', *Journal of Business & Economic Statistics* **23**(2), 226–241.

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

Lilienfeld, S. O. (2003), 'Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions', *Journal of Abnormal Child Psychology* **31**(3), 285–291.

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

Lindsay, B. G. et al. (1983), 'The geometry of mixture likelihoods: a general theory', *The Annals of Statistics* **11**(1), 86–94.

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.

Liu, J. (2004), 'Childhood externalizing behavior: Theory and implications', *Journal of Child and Adolescent Psychiatric Nursing* **17**(3), 93–103.

Liu, Y. & Bottai, M. (2009), 'Mixed-effects models for conditional quantiles with longitudinal data', *The International Journal of Biostatistics* **5**(1).

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

Lum, K., Gelfand, A. E. et al. (2012), 'Spatial quantile multiple regression using the asymmetric Laplace process', *Bayesian Analysis* **7**(2), 235–258.

Luo, Y., Lian, H. & Tian, M. (2012), 'Bayesian quantile regression for longitudinal data models', *Journal of Statistical Computation and Simulation* **82**(11), 1635–1649.

Lustig, H., Van Nieuwerburgh, S. & Verdelhan, A. (2013), 'The wealth-consumption ratio', *The Review of Asset Pricing Studies* **3**(1), 38–94.

Machado, J. A. & Mata, J. (2005), 'Counterfactual decomposition of changes in wage distributions using quantile regression', *Journal of Applied Econometrics* **20**(4), 445–465.

Maclean, J. C., Webber, D. A. & Marti, J. (2014), 'An application of unconditional quantile regression to cigarette taxes', *Journal of Policy Analysis and Management* **33**(1), 188–210.

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.

Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* **7**(1), 77–91.

Maruotti, A. (2011), 'Mixed Hidden Markov models for longitudinal data: an overview', *International Statistical Review* **79**(3), 427–454.

Maruotti, A., Petrella, L. & Sposito, L. (2021), 'Hidden semi-Markov-switching quantile regression for time series', *Computational Statistics & Data Analysis* **159**, 107208.

Maruotti, A. & Rocci, R. (2012), 'A mixed non-homogeneous hidden markov model for categorical data, with application to alcohol consumption', *Statistics in Medicine* **31**(9), 871–886.

McCulloch, C. E. (1997), 'Maximum likelihood algorithms for generalized linear mixed models', *Journal of the American Statistical Association* **92**(437), 162–170.

McMunn, A., Kelly, Y., Cable, N. & Bartley, M. (2012), 'Maternal employment and child socio-emotional behaviour in the UK: longitudinal evidence from the UK Millennium Cohort Study', *Journal of Epidemiology and Community Health* **66**(7), e19–e19.

McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach', *Journal of Empirical Finance* **7**(3-4), 271–300.

Melly, B. (2005), 'Decomposition of differences in distribution using quantile regression', *Labour Economics* **12**(4), 577–590.

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.

Merlo, L., Petrella, L. & Tzavidis, N. (2022), 'Quantile Mixed Hidden Markov Models for multivariate longitudinal data: an application to children's strengths and difficulties questionnaire scores', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* .

Mises, R. v. (1947), 'On the asymptotic distribution of differentiable statistical functions', *The Annals of Mathematical Statistics* **18**(3), 309–348.

Mishra, A. K., Mottaleb, K. A. & Mohanty, S. (2015), 'Impact of off-farm income on food expenditures in rural Bangladesh: an unconditional quantile regression approach', *Agricultural Economics* **46**(2), 139–148.

Mittnik, S. & Rachev, S. T. (1991), Alternative multivariate stable distributions and their applications to financial modeling, *in* 'Stable processes and related topics', Springer, pp. 107–119.

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.

Murakami, D. & Seya, H. (2019), 'Spatially filtered unconditional quantile regression: Application to a hedonic analysis', *Environmetrics* **30**(5), e2556.

Neuhaus, J. M., McCulloch, C. E. & Boylan, R. (2013), 'Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes', *Statistics in Medicine* **32**(14), 2419–2429.

Newey, W. K. (1994), 'The asymptotic variance of semiparametric estimators', *Econometrica: Journal of the Econometric Society* pp. 1349–1382.

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.

Nolde, N., Ziegel, J. F. et al. (2017), 'Elicitability and backtesting: Perspectives for banking regulation', *The Annals of Applied Statistics* **11**(4), 1833–1874.

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

Paindaveine, D. & Šiman, M. (2012), 'Computing multiple-output regression quantile regions from projection quantiles', *Computational statistics* **27**(1), 29–49.

Pan, W. (2001), 'Akaike's information criterion in generalized estimating equations', *Biometrics* **57**(1), 120–125.

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.

Paolella, M. S. (2015), 'Multivariate asset return prediction with mixture models', *The European Journal of Finance* **21**(13-14), 1214–1252.

Patton, A. J., Ziegel, J. F. & Chen, R. (2019), 'Dynamic semiparametric models for Expected Shortfall (and Value-at-Risk)', *Journal of Econometrics* **211**(2), 388–413.

Petrella, L., Laporta, A. G. & Merlo, L. (2018), 'Cross-Country Assessment of Systemic Risk in the European Stock Market: Evidence from a CoVaR Analysis', *Social Indicators Research* pp. 1–18.

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

Pinheiro, J. C. & Chao, E. C. (2006), 'Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models', *Journal of Computational and Graphical Statistics* **15**(1), 58–81.

Pratesi, M., Ranalli, M. G. & Salvati, N. (2009), 'Nonparametric M-quantile regression using penalised splines', *Journal of Nonparametric Statistics* **21**(3), 287–304.

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.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2005), 'Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects', *Journal of Econometrics* **128**(2), 301–323.

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

Rios-Avila, F. (2020), 'Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition', *The Stata Journal* **20**(1), 51–94.

Rockafellar, R. T. & Uryasev, S. (2000), 'Optimization of conditional Value-at-Risk', *Journal of Risk* **2**(3), 21–42.

Rodriguez-Caro, A., Vallejo-Torres, L. & Lopez-Valcarcel, B. (2016), 'Unconditional quantile regressions to determine the social gradient of obesity in Spain 1993–2014', *International Journal for Equity in Health* **15**(1), 1–13.

Royston, P. & Altman, D. G. (1994), 'Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **43**(3), 429–453.

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.

Schwarz, G. et al. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**(2), 461–464.

Scott, D. W. (2015), *Multivariate density estimation: theory, practice, and visualization*, John Wiley & Sons.

Serfling, R. (2002), 'Quantile functions for multivariate analysis: approaches and applications', *Statistica Neerlandica* **56**(2), 214–232.

Shi, Y., Ng, C. T. & Yiu, K.-F. C. (2018), 'Portfolio selection based on asymmetric Laplace distribution, coherent risk measure, and expectation-maximization estimation', *Quantitative Finance and Economics* **2**(4), 776–797.

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.

Stolfi, P., Bernardi, M. & Petrella, L. (2018), 'The sparse method of simulated quantiles: An application to portfolio optimization', *Statistica Neerlandica* .

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.

Taylor, J. W. (2005), 'Generating volatility forecasts from Value at Risk estimates', *Management Science* **51**(5), 712–725.

Taylor, J. W. (2008), 'Estimating Value at Risk and Expected Shortfall using expectiles', *Journal of Financial Econometrics* **6**(2), 231–252.

Taylor, J. W. (2019), 'Forecasting Value at Risk and Expected Shortfall using a semiparametric approach based on the asymmetric Laplace distribution', *Journal of Business & Economic Statistics* **37**(1), 121–133.

Tian, F., Gao, J. & Yang, K. (2018), 'A quantile regression approach to panel data analysis of health-care expenditure in organisation for economic co-operation and development countries', *Health Economics* **27**(12), 1921–1944.

Tiet, Q. Q., Bird, H. R., Davies, M., Hoven, C., Cohen, P., Jensen, P. S. & Goodman, S. (1998), 'Adverse life events and resilience', *Journal of the American Academy of Child & Adolescent Psychiatry* **37**(11), 1191–1200.

Tzavidis, N., Salvati, N., Geraci, M. & Bottai, M. (2010), M-quantile and expectile random effects regression for multilevel data, Technical report, Southampton Statistical Sciences Research Institute.

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**, 427–452.

Waldmann, E. (2018), 'Quantile regression: a short story on how and why', *Statistical Modelling* **18**(3-4), 203–218.

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.

Wang, Y.-G., Lin, X., Zhu, M. & Bai, Z. (2007), 'Robust estimation using the Huber function with a data-dependent tuning constant', *Journal of Computational and Graphical Statistics* **16**(2), 468–481.

Welch, L. R. (2003), 'Hidden Markov Models and the Baum-Welch algorithm', *IEEE Information Theory Society Newsletter* **53**(4), 10–13.

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

White, H., Kim, T.-H. & Manganelli, S. (2015), 'VAR for VaR: Measuring tail dependence using multivariate regression quantiles', *Journal of Econometrics* **187**(1), 169–188.

Wickham, S., Whitehead, M., Taylor-Robinson, D. & Barr, B. (2017), 'The effect of a transition into poverty on child and maternal mental health: a longitudinal analysis of the UK Millennium Cohort Study', *The Lancet Public Health* **2**(3), e141–e148.

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

Xiliang, Z. & Xi, Z. (2009), Estimation of Value-at-Risk for energy commodities via CAViaR model, *in* 'Cutting-Edge Research Topics on Multiple Criteria Decision Making', Springer, pp. 429–437.

Yamai, Y. & Yoshiba, T. (2005), 'Value-at-Risk versus Expected Shortfall: A practical perspective', *Journal of Banking & Finance* **29**(4), 997–1015.

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.

Yiu, K. (2004), 'Optimal portfolios under a Value-at-Risk constraint', *Journal of Economic Dynamics and Control* **28**(7), 1317 – 1334.

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

Yu, K. & Zhang, J. (2005), 'A three-parameter Asymmetric Laplace distribution and its extension', *Communications in Statistics—Theory and Methods* **34**(9-10), 1867–1879.

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, S., Lu, Q., Han, L., Liu, Y. & Hu, F. (2015), 'A Mean-CVaR-Skewness portfolio optimization model based on asymmetric Laplace distribution', *Annals of Operations Research* **226**(1), 727–739.

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.

Zhu, D. & Galbraith, J. W. (2011), 'Modeling and forecasting Expected Shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions', *Journal of Empirical Finance* **18**(4), 765–778.

Zucchini, W., MacDonald, I. L. & Langrock, R. (2016), *Hidden Markov models for time series: an introduction using R*, Chapman and Hall/CRC.