TI 2022-094/III
Tinbergen Institute Discussion Paper

Multivariate quantile regression using superlevel sets of conditional densities

Annika Camehl[1]
Dennis Fok[1]
Kathrin Gruber[1]

1 Erasmus University Rotterdam

Multivariate quantile regression using superlevel sets of conditional densities

Annika Camehl, Dennis Fok, and Kathrin Gruber*

Econometric Institute, Erasmus University Rotterdam

December 9, 2022

Abstract

In multiple-output quantile regression the simultaneous study of multiple response variables requires multivariate quantiles. Current definitions of such quantiles often lack a clear probability interpretation, as the defined quantiles can cover large parts of the distribution where little probability mass is located or their enclosed area does not equal the quantile level. We suggest superlevel-sets of conditional multivariate density functions as an alternative multivariate quantile definition. Such a quantile set contains all points in the domain for which the density exceeds a certain level. By applying this to a conditional density, the quantile becomes a function of the conditioning variables. We show that such a quantile has favorable mathematical and intuitive features. For implementation, we, first, use an overfitted Gaussian mixture model to fit the multivariate density and, next, calculate the multivariate quantile for a conditional or marginal density of interest. Operating on the same estimated multivariate density guarantees logically consistent quantiles. In particular, the quantiles at multiple percentiles are non-crossing. We use simulation to demonstrate that we recover the true quantiles for distributions with correlation, heteroskedasticity, or asymmetry in the disturbances and we apply our method to study heterogeneity in household expenditures.

Keywords: Multiple Response, Bayesian Quantile Regression, Gaussian Mixture Model

*camehl@ese.eur.nl (A. Camehl), dfok@ese.eur.nl (D. Fok), gruber@ese.eur.nl (K. Gruber)

1 Introduction

Regression quantiles (Koenker and Bassett, 1978) allow researchers to investigate relationships between variables not only at the center, but over the entire conditional distribution of the response. This advantage over conventional mean regression led to a rapidly expanding literature on quantile regression with countless applications in virtually all scientific disciplines (see, e.g., Koenker, 2017, for a comprehensive survey).

Now, an observation rarely comes as a single quantity and statistical analyses often require to explore the joint quantiles of multiple response variables. The simplest approach is to decompose the multiple-output problem into individual sub-problems, each of which considers a different focal response variable as a function of the other variables (e.g., Ali et al., 2016). Although the information about the joint conditional dependence between the variables is retained, these quantiles are inherently one-dimensional and do not fully reflect the multivariate structure of the response. Substantial extensions to multiple-output regression quantiles aim to order the observed points in the multivariate data space, typically in relation to a global center (see, Chaudhuri, 1996; Chakraborty, 2003; Hallin et al., 2010; Kong and Mizera, 2012). The literature offers numerous such multivariate quantile proposals and the exact approaches differ widely in their properties and computational ease (an overview is provided in the classic paper of Serfling, 2002, a more recent review can be found in Hallin and Šiman, 2018). However, the demarked quantile regions often very much depend on the underlying empirical distribution and thus, are not guaranteed to cover a certain preset probability (Hallin et al., 2021; del Barrio et al., 2022). In case the response variables feature non-linear dependencies or multi-modalities, the determined center may also lie well afield of the bulk of the data and the constructed quantile region can cover large parts of the support of a distribution with little or no probability mass (Zuo and Serfling, 2000; Carlier et al., 2016). This turns the resulting statistical objects to be difficult to interpret in practice.

In this paper, we introduce an intuitive definition of a multivariate quantile that avoids both of the above mentioned problems. Specifically, we define the quantile region of a multivariate response variable, seen as a function of a set of other variables, as the superlevel-set of a conditional multivariate density, that is, the set of points for which the conditional density function equals or exceeds a specific threshold. This threshold is set in relation to the chosen α-level of the quantile such that a probabilistic interpretation is automatically supported. Since we directly build on the multivariate distribution of the variables, instead of their multivariate center, the quantile contours adapt to the underlying data, even for highly asymmetric and multi-modal shapes.

We embed this definition of a multivariate superlevel-set quantile in a practical three step estimation procedure: (i) fitting a general multivariate distribution to the observed data, (ii) deriving a conditional distribution of interest, and (iii) applying the quantile definition to this estimated conditional distribution. This way, we can construct multivariate as well as univariate quantiles for variables, or a variable, conditional on certain values for a set of other variables. Since all quantiles are retrieved from a single, global distribution, inconsistencies, such as decreasing conditional quantile functions at increasing α-level values, are automatically avoided.

Various methods can be used to estimate the multivariate density (see, e.g., Hartigan, 1987; Müller and Sawitzki, 1991; Polonik, 1995). However, we have two requirements: the calculation of marginal and conditional distributions given the estimation result, and of the (full, marginal and conditional) cumulative distribution function must be supported. We approach the density estimation task from a Bayesian perspective with a flexible prior for the unknown distribution (Escobar and West, 1995; Müller et al., 1996). Specifically, we opt for an overfitted finite mixture model with a hierarchical Dirichlet prior on the mixing weights (Nobile and Fearnside, 2007; Rousseau and Mengersen, 2011). An efficient Markov Chain Monte Carlo algorithm makes it easy to sample the unknown parameters together with the number of components from a joint posterior (Malsiner-Walli et al., 2016), and in the end to quantify the parameter uncertainty in the quantile estimates. Thus, we can retrieve multivariate and univariate conditional as well as marginal quantiles in a coherent, computational attractive framework and also improve upon Bayesian considerations of multiple-output quantile regression (see, Cai, 2010; Taddy and Kottas, 2010; Bhattacharya and Ghosal, 2021; Guggisberg, 2022).

The rest of the paper is organized as follows. In Section 2, we discuss our proposal of a conditional multivariate superlevel-set quantile in relation to prototypical one- and multidimensional quantile definitions, and establish its theoretical properties with regard to equivariance, nestedness and uniqueness. In Section 3, we detail our setup of an overfitted finite mixture model for the joint distribution of the data. The construction principle of the quantiles is discussed in Section 4. In Section 5, we evaluate the finite-sample performance of our overfitted mixture approach in relation to some selected benchmarks in a simulation study. Thereby, we focus on the univariate conditional quantile concept, due to the lack of competing estimation methods for the same multidimensional estimand. Section 6 illustrates univariate and multivariate conditional quantile inferences on a typical application from the economic literature. We conclude with a discussion in Section 7. Details on the sampling algorithm and additional simulation results are collected in the appendices. The computer codes are written in MATLAB (The Math Works Inc., 2020) and Julia (Bezanson et al., 2017), and will be made public available upon publication of the manuscript.

2 Multivariate Quantiles

We begin with a discussion on different notions of multivariate quantiles. Thereby, we start with one-dimensional (i.e., scalar-valued) considerations and build up towards the multidimensional (i.e., vector-valued) case. Finally, we introduce our proposal of a multivariate superlevel-set quantile and establish its theoretical properties.

2.1 Scalar-valued Approaches

The simplest notion of a multivariate quantile is that of a vector with the classical univariate quantiles as its components. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_K)'$ denote a random K-dimensional vector of response variables in $\mathbb{R}^K$. The α-quantile of the k-th random variable Y_k is the number $Q_{Y_k}(\alpha)$ such that $\Pr[Y_k \leq Q_{Y_k}(\alpha)]$ is at least α:

$$Q_{Y_k}(\alpha) = \inf\left\{y \in \mathbb{R} : F_{Y_k}\big(y\big) \geq \alpha\right\},\ \alpha \in (0,1)\ , \tag{1}$$

with $F_{Y_k}(\cdot)$ the right-continuously defined cumulative distribution function of Y_k.

A straightforward extension, to account for potential dependencies between the response components, is to define the quantiles of a response variable conditional on the remaining variables. Accordingly, let $\boldsymbol{Y}_{(-k)}$ denote the vector without the k-th response variable. The conditional quantile of Y_k is a function of the α-level and a (specified) value $\boldsymbol{y}_{(-k)}$ of $\boldsymbol{Y}_{(-k)}$:

$$\begin{aligned} Q_{Y_k|\boldsymbol{Y}_{(-k)}=\boldsymbol{y}_{(-k)}}(\alpha) &= \inf\left\{y \in \mathbb{R} : F_{Y_k|\boldsymbol{Y}_{(-k)}=\boldsymbol{y}_{(-k)}}\big(y\big) \geq \alpha\right\} \\ &= \inf\left\{y \in \mathbb{R} : \Pr[Y_k \leq y|\boldsymbol{Y}_{(-k)} = \boldsymbol{y}_{(-k)}] \geq \alpha\right\} \\ &= \inf\left\{y \in \mathbb{R} : \mathrm{E}[\rho_\alpha(Y_k - y)|\boldsymbol{Y}_{(-k)} = \boldsymbol{y}_{(-k)}]\right\}, \end{aligned} \tag{2}$$

where $F_{Y_k|\boldsymbol{Y}_{(-k)}=\boldsymbol{y}_{(-k)}}$ is the cumulative distribution function of Y_k conditional on the event that variable $\boldsymbol{Y}_{(-k)}$ takes on value $\boldsymbol{y}_{(-k)}$, and $\rho_\alpha(z) = (1-\alpha)|z|I_{[z<0]} + \alpha z I_{[z\geq 0]}$ is the standard check function (see, Koenker and Bassett, 1978).

In the standard (i.e. Koenker-Basset) quantile regression approach the conditional quantile $Q_{Y_k|\boldsymbol{Y}_{(-k)}=\boldsymbol{y}_{(-k)}}(\alpha)$ is a linear function of $\boldsymbol{y}_{(-k)}$, that is, $a + \boldsymbol{b}'\boldsymbol{y}_{(-k)}$. This is a global approach and a and $\boldsymbol{b}$ are obtained by solving

$$\inf\left\{a\in\mathbb{R},\boldsymbol{b}\in\mathbb{R}^{K-1}:\mathrm{E}[\rho_\alpha(Y_k-a-\boldsymbol{b}'\boldsymbol{Y}_{(-k)})]\right\}, \tag{3}$$

with the expectation taken over Y_k and $\boldsymbol{Y}_{(-k)}$. If $\boldsymbol{Y}$ follows a multivariate Gaussian, the solution to the problem in Eq. (3) coincides with that of Eq. (2), for every value of $\boldsymbol{y}_{(-k)}$. However, if the distribution is not multivariate Gaussian, the probability statement in Eq. (2) will not hold, since the expectation is taken over $\boldsymbol{Y}$, whereas the expectation in Eq. (2) is only taken over Y_k. Though, in the standard quantile regression setting the "regressors" are assumed fixed, in the context of multiple-output problems this assumption does not hold. Thus, the simultaneous study of multiple-outcome variables requires a vector-valued take on quantiles.

2.2 Vector-valued Approaches

The directional adaptation of the univariate Koenker-Bassett approach defines the multivariate quantile as the intersection of a collection of sets (Hallin et al., 2010; Kong and Mizera, 2012). Included in this collection are not only the half-spaces for each of the K-variables, but also the half-spaces for all of their linear combinations. These half-spaces are constructed as follows. Define $\mathcal{S}^{K-1}$ as the $K-1$ dimensional unit sphere, that is, the set of vectors in $\mathbb{R}^K$ with unit length. Choose a direction $\boldsymbol{u}\in\mathcal{S}^{K-1}$ and define $\boldsymbol{\Gamma}_{\boldsymbol{u}}$ as a $K\times(K-1)$ matrix of unit vectors such that $(\boldsymbol{u},\boldsymbol{\Gamma}_{\boldsymbol{u}})$ is an orthonormal basis of $\mathbb{R}^K$. Next, we obtain the α-level regression quantile of the univariate random variable $\boldsymbol{u}'\boldsymbol{Y}$, in the (standard) Koenker-Basset sense, using the $K-1$ dimensional random vector $\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y}$, which is orthogonal to $\boldsymbol{u}'\boldsymbol{Y}$. That is, we solve

$$(a_{\boldsymbol{u}},\boldsymbol{b}_{\boldsymbol{u}})=\arg\inf\left\{a\in\mathbb{R},\boldsymbol{b}\in\mathbb{R}^{K-1}:\mathrm{E}[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y}-a-\boldsymbol{b}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y}]\right\}.$$

This generates a hyperplane $\boldsymbol{u}'\boldsymbol{y} = a_{\boldsymbol{u}} + \boldsymbol{b}_{\boldsymbol{u}}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{y}$ for every $\boldsymbol{u} \in \mathcal{S}^{K-1}$ that splits $\mathbb{R}^K$ in two half-spaces with α of the total probability mass on one side and $(1-\alpha)$ probability mass on the other side. The hyperplane for the combination $(\alpha, \boldsymbol{u})$ equals the one for $(1-\alpha, -\boldsymbol{u})$. The final multivariate quantile is the intersection of all half-spaces for $\boldsymbol{u} \in \mathcal{S}^{K-1}$. The separated region is bounded by the so-called directional quantile contour.

A more direct solution is the elliptical quantile (Hlubinka and Šiman, 2013; Hallin and Šiman, 2016). Here, we search for an ellipsoid such that α% of the probability mass is within the ellipsoid. Clearly, there exists an infinite number of such ellipsoids. Hallin and Šiman (2016) make the ellipsoid unique by suggesting to solve

$$\inf \left\{ \boldsymbol{A} \in \mathbb{R}^{K \times K}, \boldsymbol{b} \in \mathbb{R}^K, c \in \mathbb{R} : \mathrm{E}\left[\rho_\alpha(\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{b} - c)\right] \right\}, \tag{4}$$

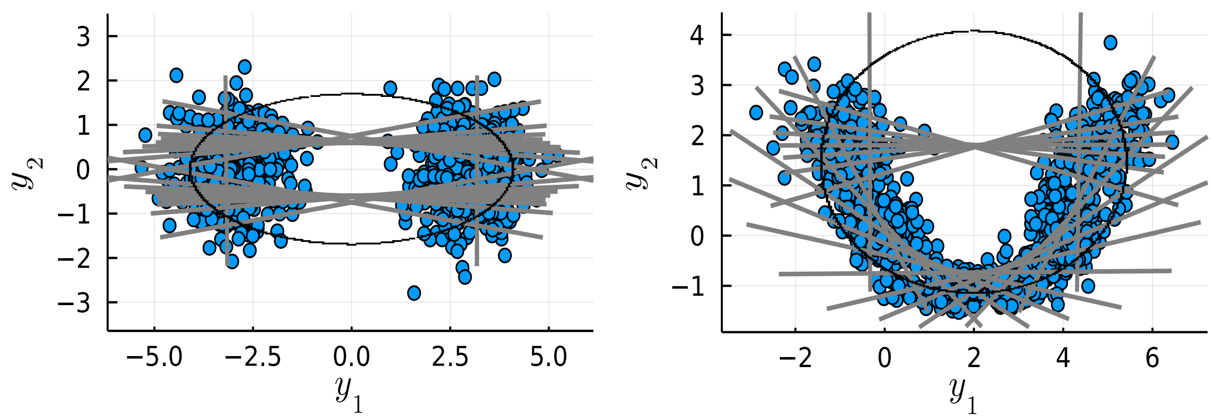
subject to $\boldsymbol{A}$ being symmetric, positive semidefinite and having determinant equal to 1. The optimal c is the α-quantile of the random variable $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{b}$. This guarantees that at the optimum $c^* = c^*(\boldsymbol{A}, \boldsymbol{b})$ we have $\Pr[\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{b} - c^* \leq 0] = \alpha$. In words, α probability mass is inside the ellipsoid defined by $\boldsymbol{A}$ and $\boldsymbol{b}$. Intuitively, the optimal values $\boldsymbol{A}^*$ and $\boldsymbol{b}^*$ are set to minimize the spread of the quadratic form $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{b}$ above and below c^*. The final quantile $\boldsymbol{Q}_{\boldsymbol{Y}}^{\text{elliptical}}(\alpha)$ is then defined as[1]

$$\boldsymbol{Q}_{\boldsymbol{Y}}^{\text{elliptical}}(\alpha) = \{\boldsymbol{y} \in \mathbb{R}^K : \boldsymbol{y}'\boldsymbol{A}^*\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{b}^* - c^* \leq 0\}. \tag{5}$$

Figure 1 compares the directional Koenker-Bassett quantiles (areas within the gray straight lines indicating 20 different directions) and elliptical quantiles (areas within the black lines) on two different bivariate Gaussian mixture distributions for $\alpha = .8$. Given that α probability is in each half-space, the probability within the enclosed region will be much smaller than α. For comparison, the ellipsoids are guaranteed - by construction - to

[1] Substituting the inequality in (5) for an equality defines the quantile as the contour of the ellipsoid.

Figure 1: Directional Koenker-Bassett and elliptical quantile contours for two examples of non-Gaussian data. Scatter plots include 1,000 simulated observations from two bivariate Gaussian distributions. Areas enclosed within the gray lines give the 80% directional quantile (20 different directions); areas enclosed within the thin black line give the 80% elliptical quantile.



cover the specified probability and thus, the 80% elliptical quantile regions are much larger. However, both the directional and the elliptical quantile definitions, cover large parts of $\mathbb{R}^2$ with very low probability mass, which turns the multivariate statistical objects difficult to interpret in practice.

Alternatively, Wei (2008) proposes the median of the distribution to construct the directional quantiles. The boundary of the quantile region gives the α-coverage interval in a particular direction from the median and is parameterized with a smooth spline as a function of the angle of the direction. Though, the resulting quantile covers α probability, the dependence on the median as a global center, limits its applicability to unimodal distributions. Likewise, for the examples in Figure 1, this quantile definition will yield a region containing large areas with low probability mass, since the density at the multivariate median is close to zero.

In sum, the directional quantile approach yields statistical objects that are difficult to interpret: the probability coverage is not known beforehand and the enclosed areas may not accurately reflect the concentration of probability mass in the data. The coverage probability problem can be resolved with the elliptical and Wei (2008) quantile approaches. Still, the quantile regions remain misleading in terms of the location of the probability mass. On top of these two limitations, the above mentioned multivariate quantile definitions do not directly allow one to obtain multivariate quantiles conditional on a set of variables. One therefore cannot easily study how the multivariate quantile depends on other (control) variables. Such dependence is at the heart of the popularity of univariate quantile regression.

2.3 The Superlevel-Set Approach

We take a different approach and directly define the multivariate quantile as a property of the density function of a subset of $\boldsymbol{Y}$ conditional on another, non-overlapping, subset of $\boldsymbol{Y}$. This allows us to easily investigate quantile dependencies. Let $\mathcal{K} \subseteq \{1, 2, \ldots, K\}$ be the set of indices of focal response variables, and $\mathcal{C} \subset \{1, 2, \ldots, K\}$ the indices of the variables collected in the conditioning set, where $\mathcal{K} \neq \emptyset$ and $\mathcal{K} \cap \mathcal{C} = \emptyset$.[2] If $|\mathcal{K}| > 1$, we obtain a multivariate distribution. $\mathcal{C}$ might be equal to $\{1, \ldots, K\} \setminus \mathcal{K}$, $\emptyset$, or something in between.

For notation, $\mathcal{K}$ or $\mathcal{C}$ in a subscript selects the corresponding elements from a vector, matrix or function. Accordingly, let $\mathcal{Y}_{\mathcal{K}}(t|\boldsymbol{y}_{\mathcal{C}})$ be the set of values $\boldsymbol{y}_{\mathcal{K}} \in \mathbb{R}^{|\mathcal{K}|}$ such that the conditional distribution of $\boldsymbol{Y}_{\mathcal{K}}$ given concrete values for the variables in the output-vector $\boldsymbol{y}_{\mathcal{C}}$, and (potential) covariates in an input-vector $\boldsymbol{x}$, equals at least t:

$$\mathcal{Y}_{\mathcal{K}}(t|\boldsymbol{y}_{\mathcal{C}}) = \{\boldsymbol{y}_{\mathcal{K}} : f_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\boldsymbol{y}_{\mathcal{K}}|\boldsymbol{x}, \boldsymbol{y}_{\mathcal{C}}) \geq t\}. \tag{6}$$

Eq. (6) gives a so-called superlevel-set of the conditional multivariate density. The corresponding superlevel-set quantile is

$$\begin{aligned} &Q_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha) = \mathcal{Y}_{\mathcal{K}}(t^*_\alpha|\boldsymbol{y}_{\mathcal{C}}), \text{ where} \\ &t^*_\alpha = \sup\{t : \Pr[\boldsymbol{Y}_{\mathcal{K}} \in \mathcal{Y}_{\mathcal{K}}(t|\boldsymbol{y}_{\mathcal{C}})|\boldsymbol{Y}_{\mathcal{C}} = \boldsymbol{y}_{\mathcal{C}}] \geq \alpha\}. \end{aligned} \tag{7}$$

[2] Note that we do not require $\mathcal{K} \cup \mathcal{C} = \{1, 2, \ldots, K\}$.

The quantile $Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$ aims to cover α probability mass with the smallest possible set. Whether this goal is exactly attainable depends on the smoothness of $\Pr[\boldsymbol{Y}_\mathcal{K} \in \mathcal{Y}_\mathcal{K}(t|\boldsymbol{y}_\mathcal{C})|\boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}]$ as a function of t. This function will be non-smooth in case the conditional density of $\boldsymbol{Y}_\mathcal{K}$ contains areas where the probability density is exactly constant. A small increase in t, may then lead to a jump in $\Pr[\boldsymbol{Y}_\mathcal{K} \in \mathcal{Y}_\mathcal{K}(t|\boldsymbol{y}_\mathcal{C})|\boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}]$. If this jump is located such that $\Pr[\boldsymbol{Y}_\mathcal{K} \in \mathcal{Y}_\mathcal{K}(t|\boldsymbol{y}_\mathcal{C})|\boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}] = \alpha$ has no solution, $Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$ will contain more than α probability. In case the conditional density is uniform on $[0,1]^{|\mathcal{K}|}$, we have $Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha) = [0,1]^{|\mathcal{K}|}$ for all $\alpha \in [0,1]$. However, this rather extreme situation will not often occur in practice. In most cases, we have $\Pr[\boldsymbol{Y}_\mathcal{K} \in Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)|\boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}] = \alpha$. The exact probability coverage of $Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$ is easy to obtain, in fact it is a by-product of the algorithm presented in Section 4.1.

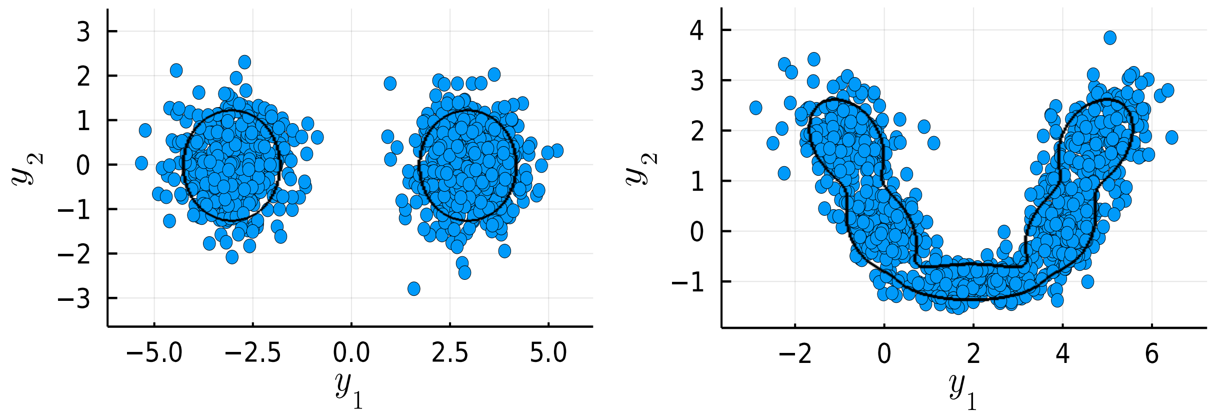
We do not restrict the shape of the multivariate quantile. If the density function is univariate, our definition gives a union of bounded intervals. If the distribution is multimodal, $Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$ is a union of sets. If the distribution features nonlinear dependencies, the shape of $Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$ can be far from elliptical. In Figure 2 we show the 80% multivariate quantile for the same distributions as used in Figure 1. We obtain intuitive multivariate quantiles for the multi-modal distribution as well as the distribution that features non-linear dependence. Finally, the explicit conditioning on $\boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}$ allows us to study how the multivariate quantile of $\boldsymbol{Y}_\mathcal{K}$ depends on the value $\boldsymbol{y}_\mathcal{C}$.

The so-defined multivariate superlevel-set quantile enjoys desirable theoretical properties of a well-behaved quantile concept, that is, (affine) equivariance, nestedness, and uniqueness (see Hallin and Šiman, 2018). We prove these properties next.

Property 1 (equivariance) *For a vector* $\boldsymbol{a}$ *and a given invertible matrix* $\boldsymbol{A}$*, both of appropriate size:* $\boldsymbol{y}_\mathcal{K} \in Q_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha) \iff \boldsymbol{a} + \boldsymbol{A}\boldsymbol{y}_\mathcal{K} \in Q_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$ *for all* $\boldsymbol{y}_\mathcal{K} \in \mathbb{R}^{|\mathcal{K}|}$*,* $\boldsymbol{y}_\mathcal{C} \in \mathbb{R}^{|\mathcal{C}|}$ *and* $\alpha \in (0,1)$

Proof. Let $f_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\cdot)$ denote the conditional density of $\boldsymbol{Y}_\mathcal{K}$ and $f_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\cdot)$ denote

Figure 2: Multivariate superlevel-set quantile contours for two examples of non-Gaussian data. Scatter plots include 1,000 simulated observations from two bivariate Gaussian distributions. Areas enclosed within the lines (i.e., quantile contours) correspond to 80% probability mass.



the conditional density of the transformed random vector $\boldsymbol{a} + \boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}}$. Next, let t^*_α be the α-quantile threshold of $\boldsymbol{Y}_{\mathcal{K}}$ and define $\tilde{t} = |\boldsymbol{A}|^{-1} t^*_\alpha$. Given $f_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{y}_{\mathcal{K}}) = |\boldsymbol{A}|^{-1} f_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\boldsymbol{y}_{\mathcal{K}})$ for any $\boldsymbol{y}_{\mathcal{K}} \in \mathbb{R}^{\mathcal{K}}$ and $\boldsymbol{y}_{\mathcal{C}} \in \mathbb{R}^{\mathcal{C}}$

$$\begin{aligned}
\boldsymbol{y}_{\mathcal{K}} \in \mathcal{Y}_{\boldsymbol{Y}_{\mathcal{K}}}(t^*_\alpha|\boldsymbol{y}_{\mathcal{C}}) \quad &\iff \quad f_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\boldsymbol{y}_{\mathcal{K}}) \geq t^*_\alpha \\
&\iff \quad f_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{y}_{\mathcal{K}})|\boldsymbol{A}| \geq t^*_\alpha = |\boldsymbol{A}|\tilde{t} \\
&\iff \quad f_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{y}_{\mathcal{K}}) \geq \tilde{t} \\
&\iff \quad \boldsymbol{a} + \boldsymbol{A}\boldsymbol{y}_{\boldsymbol{\mathcal{K}}} \in \mathcal{Y}_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}}}(\tilde{t}|\boldsymbol{y}_{\mathcal{C}}).
\end{aligned}$$

It follows $\Pr[\boldsymbol{a} + \boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}} \in \mathcal{Y}_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}}}(\tilde{t}|\boldsymbol{y}_{\mathcal{C}})|\boldsymbol{Y}_{\mathcal{C}} = \boldsymbol{y}_{\mathcal{C}}] = \Pr[\boldsymbol{Y}_{\mathcal{K}} \in \mathcal{Y}_{\boldsymbol{Y}_{\mathcal{K}}}(t^*_\alpha|\boldsymbol{y}_{\mathcal{C}})|\boldsymbol{Y}_{\mathcal{C}} = \boldsymbol{y}_{\mathcal{C}}]$ such that $Q_{\boldsymbol{a}+\boldsymbol{A}\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha) = \boldsymbol{a} + \boldsymbol{A}Q_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha)$. Consequently, the quantile is equivariant under invertible linear transformations. □

Property 2 (nestedness) *The quantile regions form a sequence of nested sets such that* $Q_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha_1) \subseteq Q_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha_2)$ *for* $0 < \alpha_1 \leq \alpha_2 < 1$ *and all* $\boldsymbol{y}_{\mathcal{C}} \in \mathbb{R}^{|\mathcal{C}|}$.

Proof. The nestedness property follows directly from the construction of the quantiles.

The set $\mathcal{Y}_{\mathcal{K}}(t|\boldsymbol{y}_{\mathcal{C}})$ increases with decreasing t and therefore, the probability $\Pr[\boldsymbol{Y}_{\mathcal{K}} \in \mathcal{Y}_{\mathcal{K}}(t|\boldsymbol{y}_{\mathcal{C}})|\boldsymbol{Y}_{\mathcal{C}} = \boldsymbol{y}_{\mathcal{C}}]$ does not decrease as t decreases. Thus, it follows $t^*_{\alpha_2} \leq t^*_{\alpha_1}$ and we obtain $Q_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha_1) = \mathcal{Y}_{\mathcal{K}}(t^*_{\alpha_1}|\boldsymbol{y}_{\mathcal{C}}) \subseteq \mathcal{Y}_{\mathcal{K}}(t^*_{\alpha_2}|\boldsymbol{y}_{\mathcal{C}}) = Q_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha_2)$. □

Property 3 (uniqueness) *Only one subset of* $\mathbb{R}^{|\mathcal{K}|}$ *satisfies the definition of the multivariate quantile* $Q_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha)$ *for a given* α *and* $\boldsymbol{y}_{\mathcal{C}}$.

Proof. Assume there are two distinct sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ both satisfying Eq. (7). If both corresponding threshold values are identical, the sets are identical. If the threshold values are not identical, it holds $t^*_{\alpha_1} > t^*_{\alpha_2}$ or $t^*_{\alpha_1} < t^*_{\alpha_2}$, and the sets $\mathcal{Y}_1$ and $\mathcal{Y}_2$ cannot both satisfy the definition of the multivariate quantile. □

3 Finite Mixture Regression

We proceed with the estimation of the multivariate density for observed data, that is, we aim to describe the joint distribution of K response variables given a common set of G covariates for N observations with a flexible multivariate model. Henceforth, let $\boldsymbol{y}_n = (y_{n1}, \ldots, y_{nK})$ be the K-dimensional output-vector of response variables and $\boldsymbol{x}_n = (x_{n1}, \ldots, x_{nG})$ be the G-dimensional input-vector of covariates for the n-th observation unit. We consider the distribution of $\boldsymbol{y}_n$ to be an M-component mixture of multivariate Gaussian densities where the component means can be a function of $\boldsymbol{x}_n$:

$$f\big(\boldsymbol{y}_n|\boldsymbol{x}_n\big) = \sum_{m=1}^{M} \kappa_m \phi\big(\boldsymbol{y}_n; \boldsymbol{g}_m(\boldsymbol{x}_n), \boldsymbol{\Sigma}_m\big), \tag{8}$$

with scalar-valued component mixing weights $\{\kappa_m\}$, subject to constraints: $\kappa_m \geq 0$ and $\sum_{m=1}^{M} \kappa_m = 1$; component-specific regression functions:

$$\boldsymbol{g}_m(\boldsymbol{x}_n) = \boldsymbol{\mu}_m + \boldsymbol{B}_m \boldsymbol{x}_n, \tag{9}$$

with location vectors $\{\boldsymbol{\mu}_m\}$ of dimension $K \times 1$, coefficient matrices $\{\boldsymbol{B}_m\}$ of dimension $K \times G$, component-specific error covariance matrices $\{\boldsymbol{\Sigma}_m\}$ of dimension $K \times K$; and density function $\phi(\cdot)$.

The Gaussian mixture formulation is quite general and, given the information encoded in the (exogenous) explanatory variables, covers a wide range of multivariate models. For $M > 1$, we obtain a flexible parameterization and for $M = 1$, a multivariate linear formulation is recovered. The number of components is treated as unknown and estimated jointly with the component-specific parameters from the data. Note that other component distribution choices, e.g., the multivariate skew-Gaussian or the multivariate skew-t (e.g., Lin et al., 2007; Frühwirth-Schnatter and Pyne, 2010), are possible as well. However, given enough Gaussian components, large classes of conditional multivariate densities can be approximated, with sufficient accuracy (Roeder and Wasserman, 1997). Though the assumption of a Gaussian base-distribution restricts the specified model to continuous response variables, this limitation does not apply to the definition of the multivariate superlevel-set quantile.

3.1 Prior Distributions

We assume an intentionally overfitted model with a finite but comparatively large number of components. The mixture is reduced to the final model through a sparse hierarchical prior on the collection of possible density functions. Accordingly, we specify a hierarchical Dirichlet prior for the component mixing weights:

$$\boldsymbol{\kappa}|\{\bar{\rho}_m\} \ \sim \mathcal{D}\big(\bar{\rho}_1, \cdots, \bar{\rho}_M\big), \tag{10}$$

where the component-specific concentration parameters $\{\bar{\rho}_m\}$ follow a Gamma distribution:

$$\bar{\rho}_m \sim \mathcal{G}\big(\underline{a}_1, 1/(\underline{a}_2 M)\big), \tag{11}$$

with hyperparameters $\underline{a}_1, \underline{a}_2 > 0$. The component locations are given a multivariate Gaussian prior:

$$\boldsymbol{\mu}_m|\bar{\boldsymbol{v}}_0, \overline{\boldsymbol{V}}_0 \sim \mathcal{N}\big(\bar{\boldsymbol{v}}_0, \overline{\boldsymbol{V}}_0\big), \tag{12}$$

with an improper Gaussian prior placed on the component prior mean $\bar{\boldsymbol{v}}_0$ and a hierarchical prior on the component prior variance $\overline{\boldsymbol{V}}_0$:

$$\bar{\boldsymbol{v}}_0 \sim \mathcal{N}\big(\underline{\boldsymbol{v}}, \underline{\boldsymbol{V}}\big), \tag{13}$$

$$\overline{\boldsymbol{V}}_0 = \text{diag}(R_1^2\lambda_1, \ldots, R_K^2\lambda_K), \tag{14}$$

$$\lambda_k \sim \mathcal{G}\big(\underline{b}_1, 1/\underline{b}_2\big), \tag{15}$$

where $\underline{\boldsymbol{v}} = median(\boldsymbol{y}_n)$ and $\underline{\boldsymbol{V}}^{-1} = \mathbf{0}$. $\{R_k\}$ correspond to the response variable-specific value ranges and $\{\lambda_k\}$ are local shrinkage factors following a Gamma distribution with hyperparameters $\underline{b}_1, \underline{b}_2 > 0$ (see, Brown and Griffin, 2010). We complete our model with a multivariate Gaussian prior for the vector of component-specific regression coefficients: $\text{vec}(\boldsymbol{B}_m) \sim \mathcal{N}\big(\boldsymbol{c}_0, \boldsymbol{C}_0\big)$ with prior mean $\boldsymbol{c}_0$ and prior variance $\boldsymbol{C}_0$; and a conjugate inverse-Wishart prior for the component-specific error covariance matrices: $\boldsymbol{\Sigma}_m \sim \mathcal{IW}\big(\boldsymbol{S}_0, s_0\big)$ with scale matrix $\boldsymbol{S}_0 = \boldsymbol{I}$ and degrees of freedom $s_0 > 2 + K$, to rule out variances equal or close to zero.

Asymptotically, the hierarchical Dirichlet prior on the mixing weights (Eq. (10)–(11)) matches a Dirichlet process prior with expectation $E(\bar{\rho}_m) = 1/M$ and variance $\text{Var}(\bar{\rho}_m) = 1/(\underline{a}_2 M^2)$ (see, Ishwaran et al., 2001). Thus, as the number of mixture distributions M increases, the relative size of the component mixing weights $\{\bar{\rho}_m\}$ decreases, and the sequence of Gaussian mixtures becomes increasingly sparse. The hyperparameters $\underline{a}_1$ and $\underline{a}_2$ have a regularizing effect such that the posterior distribution of the weights is encouraged to concentrate at zero for superfluous components. The multivariate Normal-Gamma prior (Eq. (13)–(15)) then puts strong shrinkage on the component means via the prior variance, particularly for heavily overlapping densities, to reduce the bias (caused by spurious components) for components with a small number of observations (see also, Richardson and Green, 1997; Stephens, 2000; Yau and Holmes, 2011).

3.2 Posterior Inference

Given data on N observations, we obtain posterior results via a data-augmented MCMC algorithm (Diebolt and Robert, 1994). Thus, we associate each observation $\boldsymbol{y}_n$ to a latent allocation variable $z_n \in \{1, \dots, M\}$ such that:

$$f\big(\boldsymbol{y}_n | z_n = m, \boldsymbol{x}_n\big) = \phi\big(\boldsymbol{y}_n; \boldsymbol{g}_m(\boldsymbol{x}_n), \boldsymbol{\Sigma}_m\big), \tag{16}$$

$$\Pr[z_n = m] = \kappa_m. \tag{17}$$

Exact details on the full conditionals and the sampling steps are provided in Appendix A. After each MCMC iteration a random permutation step is added to ensure that the sampler explores all modes of the full posterior distribution (Frühwirth-Schnatter, 2001). Most parts of this algorithm are quite standard with two exceptions: the full conditional distribution for the shrinkage factors (Malsiner-Walli et al., 2016) and the multivariate random-walk Metropolis-Hastings step for the component-specific concentration parameters.

In case component-specific posterior means or variances are needed, one needs to solve the label-switching problem present in most mixture models. We identify the component labels in a separate post-processing procedure. We cluster the component-specific posterior MCMC samples using a K-centroids Mahalanobis distance-based algorithm to reduce the dimensionality of the relabeling problem and to capture elliptical shapes and different volumes of the posteriors (a detailed description is given in Malsiner-Walli et al., 2016). This identification scheme can easily be applied to non-Gaussian mixture settings as well.

4 Implementation

Given samples from the posterior of the Gaussian mixture model, we can obtain posterior quantile estimates of $\boldsymbol{Y}_{\mathcal{K}}$ given $\boldsymbol{Y}_{\mathcal{K}} = \boldsymbol{y}_{\mathcal{C}}$, either at the posterior means or for each posterior draw. The latter allows a straightforward evaluation of the posterior uncertainty in the quantile estimates, while the former involves lower computational costs. Note that using posterior means requires the label-switching problem to be solved, as previously discussed.

Specifically, we aim to construct multivariate and univariate quantiles given some value of the other response variables, and some value for the covariates. Accordingly, the conditional distribution of $\boldsymbol{Y}_{\mathcal{K}}$ given the estimated model parameters collected in Θ, concrete values for the variables in the output-vector $\boldsymbol{y}_{\mathcal{C}}$, and the covariates in the input-vector $\boldsymbol{x}$, is needed. It is easy to show that this conditional distribution is again a M-component mixture of Gaussians:

$$f_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}\big(\boldsymbol{y}_{\mathcal{K}}|\boldsymbol{x}, \boldsymbol{y}_{\mathcal{C}}, \Theta\big) = \sum_{m=1}^{M} \omega_m^{\mathcal{C}}(\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x})\phi\big(\boldsymbol{y}_{\mathcal{K}}; \boldsymbol{\mu}_m^{\mathcal{K}|\mathcal{C}}(\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x}), \boldsymbol{\Sigma}_m^{\mathcal{K}|\mathcal{C}}\big), \tag{18}$$

with component-specific conditional mean $\boldsymbol{\mu}_m^{\mathcal{K}|\mathcal{C}}(\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x})$ and component-specific conditional variance $\boldsymbol{\Sigma}_m^{\mathcal{K}|\mathcal{C}}$:

$$\boldsymbol{\mu}_m^{\mathcal{K}|\mathcal{C}}(\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x}) = \boldsymbol{g}_{m,\mathcal{K}}(\boldsymbol{x}) + \boldsymbol{\Sigma}_{m,\mathcal{K},\mathcal{C}}\boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{C}}^{-1}\big(\boldsymbol{y}_{\mathcal{C}} - \boldsymbol{g}_{m,\mathcal{C}}(\boldsymbol{x})\big), \tag{19}$$

$$\boldsymbol{\Sigma}_m^{\mathcal{K}|\mathcal{C}} = \boldsymbol{\Sigma}_{m,\mathcal{K},\mathcal{K}} - \boldsymbol{\Sigma}_{m,\mathcal{K},\mathcal{C}}\boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{C}}^{-1}\boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{K}} \ , \tag{20}$$

and (updated) conditional mixing weight $\omega_m^{\mathcal{C}}$ defined in terms of the Gaussian components in the conditioning set:

$$\omega_m^{\mathcal{C}}(\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x}) = \frac{\kappa_m \phi\big(\boldsymbol{y}_{\mathcal{C}}; \boldsymbol{g}_{m,\mathcal{C}}(\boldsymbol{x}), \boldsymbol{\Sigma}_{m,\mathcal{C},\mathcal{C}}\big)}{\sum_{l=1}^{M} \kappa_l \phi\big(\boldsymbol{y}_{\mathcal{C}}; \boldsymbol{g}_{l,\mathcal{C}}(\boldsymbol{x}), \boldsymbol{\Sigma}_{l,\mathcal{C},\mathcal{C}}\big)} \ . \tag{21}$$

The constructed conditional distribution serves as input for the respective quantile formulation in the following subsections.

4.1 Multivariate Quantiles

The conditional Gaussian mixture distribution at given parameter values is used to numerically solve the optimization problem that defines the multivariate superlevel-set quantile (see, Eq. (7)). The procedure is summarized in Algorithm 1.

First, we select a subset of $\mathbb{R}^{|\mathcal{K}|}$ with a sufficiently large density, that is, we take the Cartesian product of the intervals $[F^{-1}_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\epsilon), F^{-1}_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(1-\epsilon)]$ for each $k \in \mathcal{K}$, where ϵ is a small value close to zero and $F^{-1}_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\cdot)$ is the quantile function for variable Y_k given $\boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}$ (see also Eq. (18)). We then place n_{grid} grid points on each interval, running from the corresponding ϵ-marginal to the $(1-\epsilon)$-marginal quantile, to split the subset into $(n_{\text{grid}}-1)^{|\mathcal{K}|}$ hypercubes. Next, we calculate the enclosed probability content for each hypercube based on the estimated cumulative density functions. Finally, we add hypercubes

Algorithm 1: Calculating the multivariate level-set quantile

Input : chosen coverage probability α
conditional distribution function $F_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\boldsymbol{y})$
grid boundary probability ϵ
dimension-specific grid point number n_{grid}

Output: actual coverage probability p
numerical quantile $\tilde{Q} = \tilde{Q}_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$ of size $n_{\text{grid}}^{|\mathcal{K}|}$

for $k \in \mathcal{K}$ **do**
  grid_k = equally spaced n_{grid} vector with values
    from $F^{-1}_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\epsilon)$ to $F^{-1}_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(1-\epsilon)$;
  $\tilde{Q}_{\boldsymbol{Y}_\mathcal{K}|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha) = |\mathcal{K}|$-dimensional array of zeros
P = empty $|\mathcal{K}|$-dimensional array to hold probabilities per hypercube
for $(i_1 \in 2 : n_{grid}),\ (i_2 \in 2 : n_{grid}),\ \ldots,\ (i_{|\mathcal{K}|} \in 2 : n_{grid})$ **do**
  $P_{i_1,i_2,\ldots,i_{|\mathcal{K}|}} = \Pr[Y_k \in [\text{grid}_{k,i_k-1}, \text{grid}_{k,i_k}]\ \forall k \in \mathcal{K} | \boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}]$
$p = 0$
while $p < \alpha$ **do**
  $\mathcal{I}$ = set of indices for which P equals $\max\{P\}$
  $p = p + \sum_{i \in \mathcal{I}} P_i$
  for $i \in \mathcal{I}$ **do**
    $\tilde{Q}_i = \alpha$
    $P_i = 0$

in the order of the highest probability to $\tilde{Q}_{\boldsymbol{Y}_{\mathcal{K}}|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(\alpha)$ until the chosen α-coverage probability is exceeded or (exactly) reached. Hypercubes with the same probability are added simultaneously.

The discretization of the probability space results in a numerical approximation of the multivariate quantile. This approximation is accurate as long as the number of grid points is sufficiently large. Otherwise, the enclosed probability content may be too large and the resulting quantile boundaries will be non-smooth. To avoid hypercubes at the border of the grid, ϵ should be set to a small value. In Figure 2, for instance, we set $\epsilon = 0.001$ and $n_{\text{grid}} = 250$. Note, that a too small ϵ value and/ or a too high grid point number will increase the necessary computation time.

4.2 Conditional Univariate Quantiles

Likewise, we can construct conditional univariate quantiles, in the traditional sense, directly from the (estimated) conditional distributions. For the conditional Gaussian mixture distribution, we obtain the univariate quantiles numerically as the solution to a simple root-finding problem. Algorithm 2 provides a stylized description of the procedure on the example of a variable k conditional on the values of all variables in $\mathcal{C}$, where $\{k\} \cap \mathcal{C} = \emptyset$.

First, we partition the output-vector and the corresponding mean and error variance estimates to calculate the conditional component-specific moments $\mu_m^{k|\mathcal{C}}(\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x})$ and $\Sigma_m^{k|\mathcal{C}}$ (see, Eq. (19) and (20)), and conditional mixing weights $\omega_m^{\mathcal{C}}(\boldsymbol{y}_{\mathcal{C}}, \boldsymbol{x})$ (see, Eq. (21)). Next, we derive the quantile directly from the distribution function of the corresponding conditional Gaussian mixture, that is, we use a simple bisection method to find the value q for which $F_{Y_k|\boldsymbol{Y}_{\mathcal{C}}=\boldsymbol{y}_{\mathcal{C}}}(q) = \alpha$. Thereby, the conditional component quantiles specify the interval for the root-finder, with the smallest and largest component α-quantiles providing the lower and upper bounds, respectively.

Algorithm 2: Calculating the conditional univariate quantiles

Input : chosen coverage probability α
index of focal response variable k
value of response variables $\boldsymbol{y}_\mathcal{C}$
estimated component-specific parameters $\{\boldsymbol{g}_m(\boldsymbol{x}_n)\}$, $\{\boldsymbol{\Sigma}_m\}$, $\{\boldsymbol{\kappa}_m\}$

Output : conditional univariate quantile $Q_{k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}(\alpha)$

for $m \in \{1, \dots, M\}$ **do**
    Calculate conditional moments $\bar{x}_m = \mu_m^{k|\mathcal{C}}(\boldsymbol{y}_\mathcal{C}, \boldsymbol{x})$ and $s_m^2 = \Sigma_m^{k|\mathcal{C}}$
    Calculate component-specific quantiles $q_m = \mu_m^{k|\mathcal{C}}(\boldsymbol{y}_\mathcal{C}, \boldsymbol{x}) + \Sigma_m^{k|\mathcal{C}} \Phi^{-1}(\alpha)$
    Calculate conditional weight $w_m = \omega_m^\mathcal{C}(\boldsymbol{y}_\mathcal{C}, \boldsymbol{x})$

for *all* $k \in \mathcal{K}$ **do**
    Define $h(q) = \sum_{m=1}^M w_m \Phi\big((q - \bar{x}_m)/\sqrt{s_m^2}\big) - \alpha$
    Use bisection to find root of $h(q)$ starting with interval $[\min_m q_m, \max_m q_m]$

$\Phi^{-1}(\alpha)$ is the (inverse) of the standard Gaussian distribution function.

4.3 Quantile-Specific Measures

Since inference for any set of quantile levels is based on a global posterior, quantile crossing, that is, the problem of non-monotonicity of the quantile contours, is automatically avoided. Hence, we can infer a valid posterior distribution for any linear combination of response variables. Likewise, we can create a simple model of treatment and control to calculate marginal quantile effects.

We define the local marginal effect in the α-level quantile of response variable Y_k conditional on $\boldsymbol{Y}_\mathcal{C} = \boldsymbol{y}_\mathcal{C}$ for a change from $\boldsymbol{x}$ to $\boldsymbol{x} + \Delta_g$:

$$\beta_{k|\mathcal{C}}^g(\alpha|\boldsymbol{y}_\mathcal{C}, \boldsymbol{x}) = Q_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}\big(\alpha|\boldsymbol{x} + \boldsymbol{\Delta}_g\big) - Q_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}\big(\alpha|\boldsymbol{x}\big), \tag{22}$$

where $\boldsymbol{\Delta}_g$ is a vector with a small value δ_g at position g and zeros elsewhere (Doksum, 1974). Alternatively, the marginal effect can be considered a derivative with δ_g dividing the result of Eq. (22). Similar marginal effects can be defined for changes in the conditioning variables $\boldsymbol{y}_\mathcal{C}$ in a straightforward way.

The derived marginal quantile effects permit a direct local interpretation at a certain value of $\boldsymbol{x}$ and $\boldsymbol{y}_{\mathcal{C}}$, or a global interpretation through a (weighted) average across all observations, $\frac{1}{N}\sum_{n=1}^{N}\beta^{g}_{k|\mathcal{C}}(\alpha|\boldsymbol{y}_{n,\mathcal{C}},\boldsymbol{x}_{n,\mathcal{C}})$. The computation of the corresponding interval estimates for posterior uncertainty quantification is straightforward: Eq. (22) is evaluated for each posterior draw and the results are summarized either in terms of a posterior variance or the highest posterior density region.

5 Simulation Study

We compare the performance of our Gaussian mixture-based quantiles to existing approaches in a Monte Carlo simulation study. The lack of a generally accepted multivariate quantile definition restricts this simulation exercise to univariate conditional quantiles. Still, a good fit of the conditional Gaussian mixture distributions readily translates into good estimation performance for the multivariate quantiles. We consider five data generating processes with varying degrees of asymmetry and non-convexity in the multivariate response distributions:

1. Multivariate Gaussian:

   $\boldsymbol{y}_n \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = [.2, .2, .2]'$ and $\Sigma_{jj} = .4$, $\Sigma_{jk} = .25$ $(\forall j \neq k)$.

2. Multivariate Student-t:

   $\boldsymbol{y}_n \sim t_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $r = 5$, $\boldsymbol{\mu} = [.2, .2, .2]'$ and $\Sigma_{jj} = .4$, $\Sigma_{jk} = .25$ $(\forall j \neq k)$.

3. Multivariate log-Gaussian:

   $\boldsymbol{y}_n \sim \log\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = [.2, .2, .2]'$ and $\Sigma_{jj} = .4$, $\Sigma_{jk} = .25$ $(\forall\ j \neq k)$.

4. Conditional heteroskedasticity:

   $\boldsymbol{y}_n \sim \mathcal{N}(\boldsymbol{\mu}, \Omega_n)$ with $\boldsymbol{\mu} = [.2, .2, .2]'$ and $\Omega_n = \exp(z_n)\boldsymbol{\Sigma}$ where $z_n \sim \mathcal{N}(0, 1)$ and $\Sigma_{jj} = .4$, $\Sigma_{jk} = .25$, $\forall j \neq k$.

5. Multivariate Gaussian mixture:

$\boldsymbol{y}_n \sim \mathcal{N}(\boldsymbol{\mu}_{z_n}, \boldsymbol{\Sigma}_{z_n})$ with $\Pr[z_n = m] = .33$, for $m = 1, 2, 3$, $\boldsymbol{\mu}_1 = [2, 2, 2]$, $\boldsymbol{\mu}_2 = [0, 0, 0]$, $\boldsymbol{\mu}_3 = [-2, .5, 1]$ and $\Sigma_{1,jj} = .4$, $\Sigma_{1,jk} = .25$, $\boldsymbol{\Sigma}_2 = \boldsymbol{I}$, $\Sigma_{3,jj} = .7$, $\Sigma_{3,jk} = .5$.

For each data generating process, we generate $1,000$ data sets with a sample size of $10,000$.

We analyze each simulated data set with four methods. (1) The conditional univariate quantiles retrieved from fitting our multivariate Gaussian mixture model to the data. We assume an overfitted mixture distribution of order $M = 5$ and set the hyperparameters for the hierarchical Dirichlet prior on the mixing weights to $\underline{a}_1 = 10$ and $\underline{a}_2 = 40$ to keep the conditional allocation probabilities to spurious components small. The hyperparameters for the Gamma prior on the component prior variance are set to $\underline{b}_1 = .5$ and $\underline{b}_2 = .5$. We keep every 10-th of $50,000$ MCMC draws after an initial burn-in phase of $10,000$. (2) A standard univariate quantile regression model with the linear regression quantiles estimated independently for each response variable. Here, the regression quantiles essentially correspond to a constant. (3) A univariate linear quantile regression model with the other response variables of the output-vector in the conditioning set, and (4) a univariate non-linear quantile regression model with the level and squared response variables in the conditioning set.[3]

All models are evaluated on the basis of the average squared deviation of the estimated regression quantiles from the true conditional α-level quantiles for $\alpha \in \{.2, .4, .6, .8\}$. The exact definitions of the respective population quantile functions are provided as a supplementary appendix. We use posterior means as point estimates for the Bayesian models. The corresponding simulated mean squared errors are summarized in Table 1. In general, the expansion of the model input-space, that is, using the other variables in the output-vector, greatly increases the predictive performance. The multivariate Gaussian mixture model recovers the conditional quantiles of the symmetric uni-/ and the multi-modal settings well as compared to the univariate linear and non-linear specifications. Likewise, the

[3] All univariate quantile regression models are implemented with the `rq_fnm` function (Xu, 2020). We use the `bisection` implementation of Sartorius (2020) for deriving the Gaussian mixture-based quantiles.

Table 1: Mean squared errors for five simulated multivariate response distributions

		Gaussian	Student-t	log-Gaussian	conditional heteroskedast.	Gaussian mixture
$\alpha = .2$	QReg	.216	.367	.214	.353	1.960
	QReg with $\boldsymbol{y}_{(-k)}$	.000	.013	.004	.024	.539
	QReg with $\boldsymbol{y}_{(-k)}$, $(\boldsymbol{y}_{(-k)})^2$	.000	.008	.002	.012	.269
	MQReg	.000	.004	.001	.003	.004
$\alpha = .4$	QReg	.196	.332	.328	.315	1.517
	QReg with $\boldsymbol{y}_{(-k)}$	.000	.001	.007	.002	.487
	QReg with $\boldsymbol{y}_{(-k)}$, $(\boldsymbol{y}_{(-k)})^2$	.000	.001	.004	.001	.253
	MQReg	.000	.003	.026	.002	.003
$\alpha = .6$	QReg	.197	.330	.492	.315	1.244
	QReg with $\boldsymbol{y}_{(-k)}$	.000	.001	.011	.002	.404
	QReg with $\boldsymbol{y}_{(-k)}$, $(\boldsymbol{y}_{(-k)})^2$	.000	.001	.005	.002	.247
	MQReg	.000	.003	.048	.001	.003
$\alpha = .8$	QReg	.216	.361	.850	.350	1.084
	QReg with $\boldsymbol{y}_{(-k)}$	.000	.013	.002	.024	.197
	QReg with $\boldsymbol{y}_{(-k)}$, $(\boldsymbol{y}_{(-k)})^2$	.000	.009	.001	.012	.124
	MQReg	.000	.004	.081	.002	.002

[a] QReg: univariate linear quantile regression; QReg with $\boldsymbol{y}_{(-k)}$: univariate linear quantile regression with the other response variables in the conditioning set; QReg with $\boldsymbol{y}_{(-k)}$, $(\boldsymbol{y}_{(-k)})^2$: univariate non-linear quantile regression with the other level and squared response variables in the conditioning set; MQReg: multivariate Gaussian mixture quantile regression.

[b] The reported mean squared errors are averaged across response variables and simulation draws.

asymmetries in the conditional heteroskedasticity design are accurately recovered, particularly in the lowest and highest quantiles of the conditional response distributions. Only for the extremely skewed multivariate log-Gaussian distribution our Gaussian mixture model's fit decreases in the tails.

For DGP 5, the Multivariate Gaussian mixture, Figure 3 compares the estimated quantiles of response variable y_1 against the true conditional Gaussian mixture α-level quantiles for different values of y_2 (on the horizontal axis) and y_3 (across the different subplots). The estimated conditional univariate Gaussian mixture quantiles perfectly match the true conditional quantiles while the conditional univariate linear and non-linear quantiles only give global approximations. Likewise, as illustrated in Figure 4, the asymmetries in the conditional heteroskedasticity process are perfectly retained, particularly for the highest and lowest quantiles of the conditional response distributions. For the simulated uni-modal and symmetric distributions all three models provide similar estimates (the figures of the remaining simulation designs are collected in a supplementary appendix).

Figure 3: Estimated (dashed lines) vs. true (solid lines) conditional quantiles of DGP 5, a multivariate Gaussian mixture response distribution, for $\alpha \in \{.2, .4, .6, .8\}$

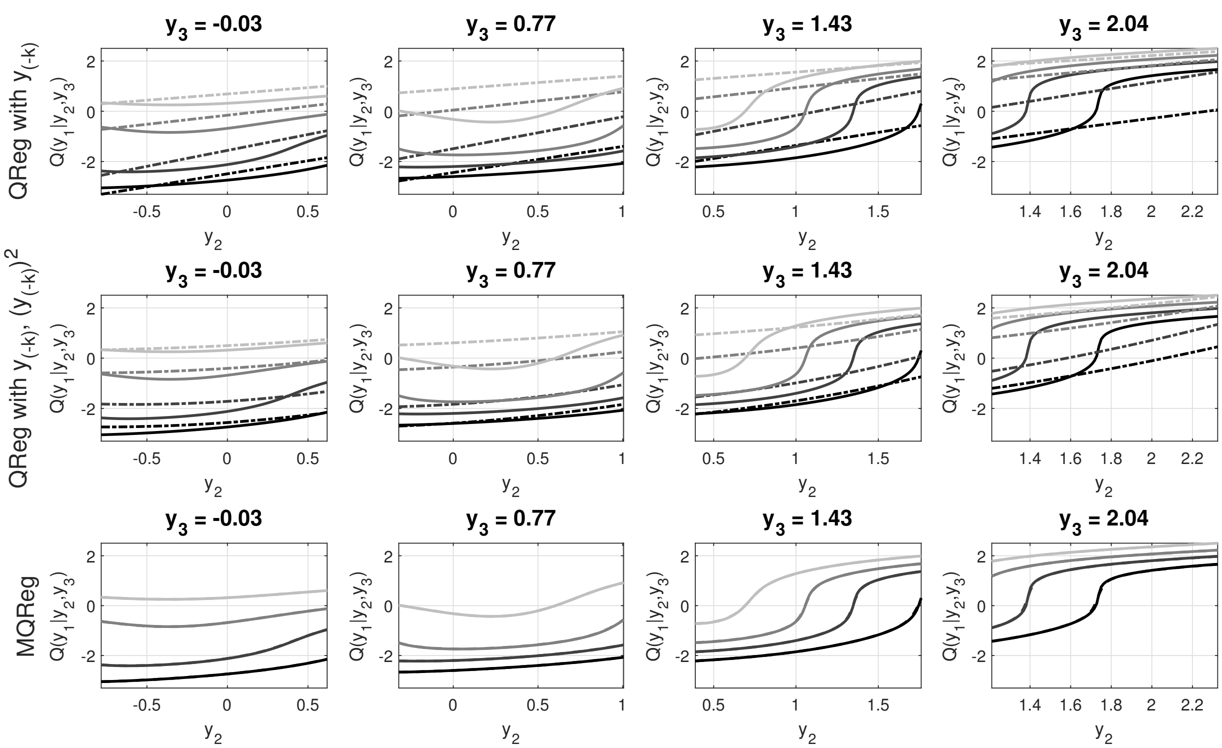
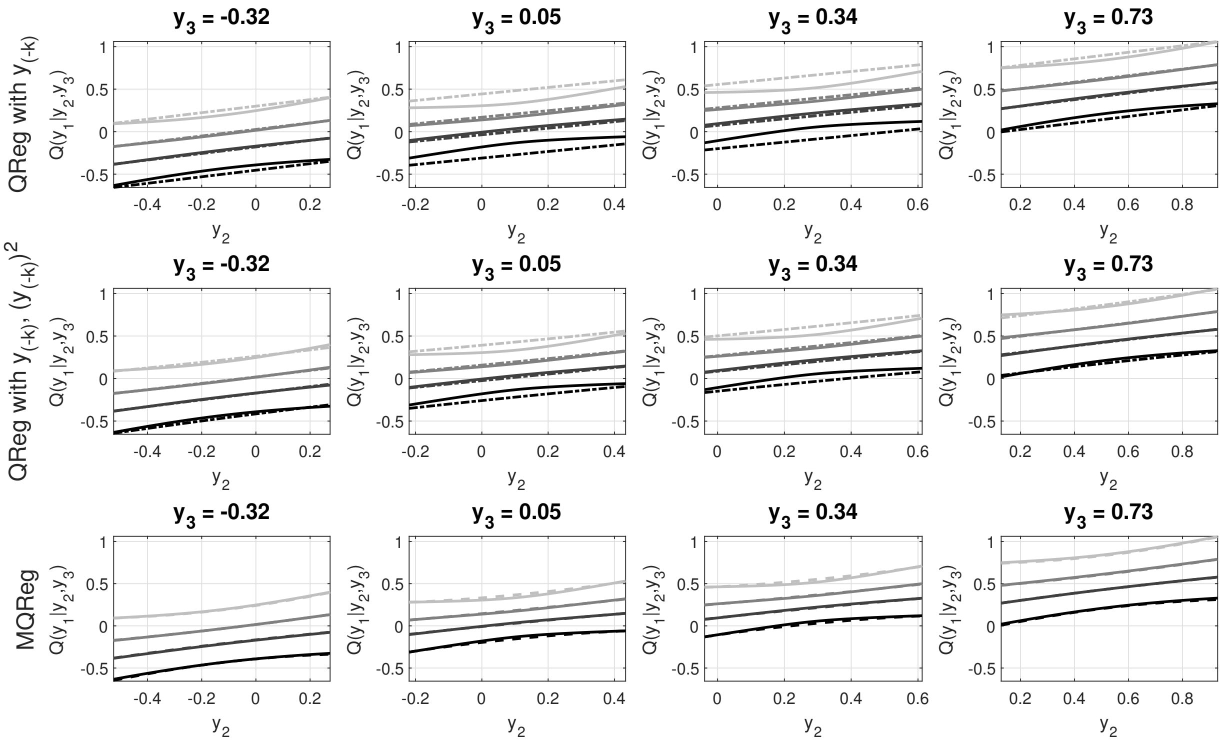


Figure 4: TEstimated (dashed lines) vs. true (solid lines) conditional quantiles of DGP 4, a conditional heteroskedasticity response distribution, for $\alpha \in \{.2, .4, .6, .8\}$

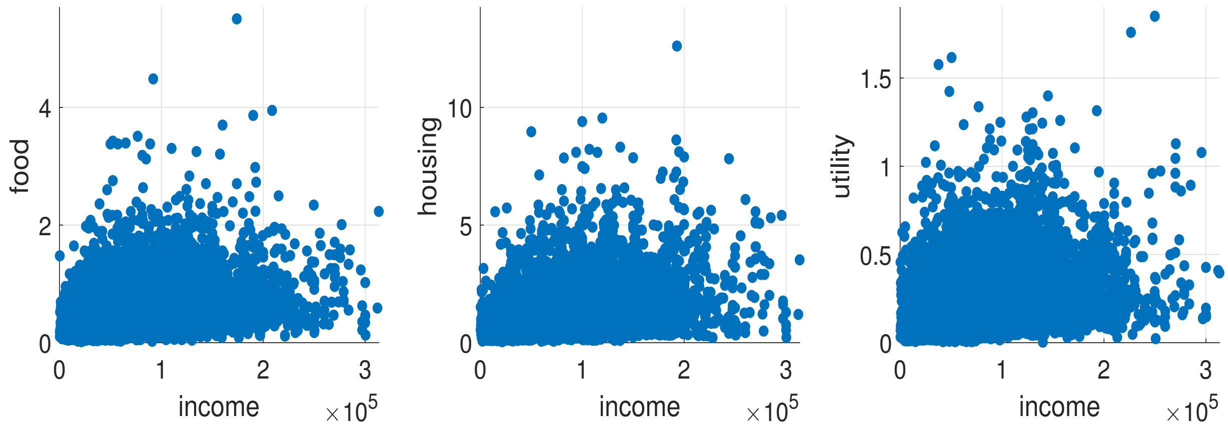


6 Empirical Application

The study of heterogeneity in household consumption patterns is a classical application of quantile regression in economics. In the following, we analyze the variation in annual real expenditures on food, housing and utilities (i.e., fuels, gas, electricity, telephone and public services) jointly with household income. The data are a subset of households from the U.S. Consumer Expenditure Survey collected by the Bureau of Labor Statistics in 2015.[4] We only consider households consisting of one or two earners, living in an owned or rented property, and having two or less children. This leaves us with N=29,988 observations for $K = 4$ response variables. The household income is the amount the household earned in

[4] The corresponding survey interview files are public available via https://www.bls.gov/cex/pumd_data.htm. The analysis is based on the variable labels: `houspq` (housing), `foodpq` (food), `utilpq` (utilities) and `fincbtxm` (income before taxes). The replication codes will be made available after publication.

Figure 5: Empirical distribution of the three expenditures categories food, housing, and utilities conditional on income before taxes

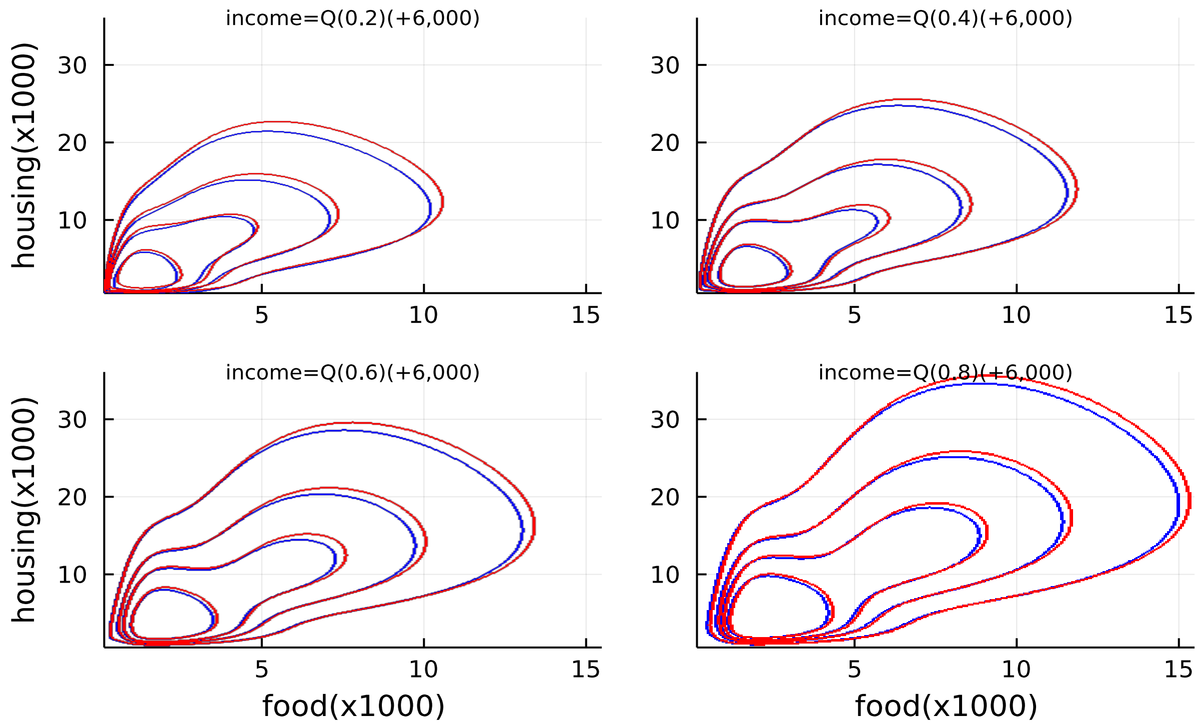


the chosen year before taxes. Figure 5 illustrates the empirical distribution of the three expenditure categories conditional on income. Note that the sum of the expenditures does not amount to the overall consumption of the household and thus, the considered categories are not complete.

Again, we assume an overfitted mixture distribution of order $M = 5$. We a-priori favor allocations to several components, to address the asymmetries in the empirical distribution, and set the hyperparameter for the hierarchical Dirichlet prior on the mixing weights to $\underline{a}_1 = 10$ and $\underline{a}_2 = 40$. The data is additionally log-transformed in order to fit the income as well as expenditure distributions with fewer mixture components. The hyperparameters for the Gamma prior on the component prior variance are set to $\underline{b}_1 = .5$ and $\underline{b}_2 = .5$. We keep every 40-th of $200,000$ MCMC draws after an initial burn-in phase of $400,000$. The quantile estimates are presented on the original scales. For the multivariate (conditional) quantiles, we obtain the distribution for the original data from the distribution function of the log-transformed data. For univariate (conditional) quantiles, the inverse transformation can be directly applied to the estimated quantile.

Figure 6 shows the bivariate quantiles for expenditures on food and housing ($\mathcal{K} = \{\text{food}, \text{housing}\}$) conditional on marginal α-level values of income ($\mathcal{C} = \{\text{income}\}$).

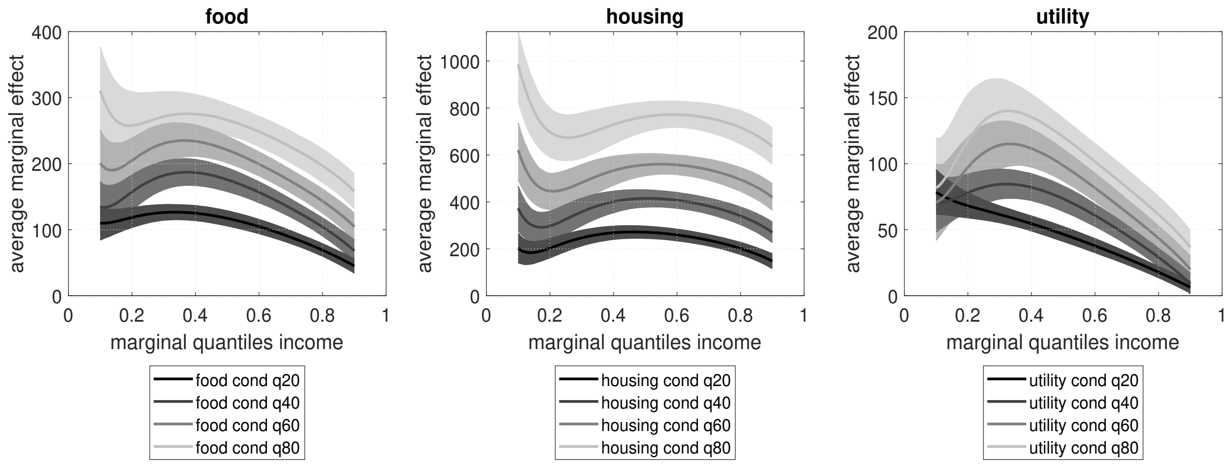
Figure 6: Bivariate quantiles for the two expenditure categories food and housing conditional on different levels of income. Blue lines correspond to $\alpha \in \{.2, .4, .6, .8\}$ conditional on the indicated quantiles of income, red lines correspond to an income increase of $6,000



We implicitly marginalize over the utilities expenditures. The blue lines correspond to the multivariate quantiles at $\alpha \in \{.2, .4, .6, .8\}$, the red lines give the multivariate quantiles conditional on an income increase of $6,000. Conditional on the four income levels, those households with the least food and housing expenditures (.2-quantile, smallest enclosed area) do not react considerably to the income increase for all four income levels. In contrast, those in the highest expenditures for housing and food (.8-quantile, largest enclosed area) do increase their spending substantially after a raise in income. This effect is strongest for the lowest income quantile. Moreover, the mulitvariate quantiles do not indicate clear substitution effects between the two categories.

Next, we investigate the quantiles from a univariate perspective. Figure 7 depicts the quantile-varying marginal effect for a raise in income of $6,000. The associated consumption changes greatly vary along the income distribution: Low-income households dedicate

Figure 7: Quantile-varying marginal effects for the three expenditures categories food, housing, and utilities conditional on different levels of income. Shaded areas indicate the corresponding 90% confidence intervals for four α-level values of income



most of the additional income to necessities that relate to spending on food and shelter. In contrast, households in the middle of the income distribution enjoy greater discretionary power and thus, seem to adjust their food and utilities expenditures the most. Households in the highest quantiles of the income distribution hardly increase their spending on food and utilities at all.

7 Conclusion

We propose a new multivariate superlevel-set conditional quantile definition specified directly on the (estimated) multivariate density function. The superlevel-set is defined as the set of all outcomes in the sample space for which the conditional density equals, or exceeds a certain threshold. For a given α level, we search for the density-threshold such that the probability content of the superlevel-set is as close as possible to α, but not smaller than α. This multivariate quantile has several favorable properties: it has a clear coverage probability interpretation, works for various distributional forms, can have a non-convex shape, and is equivariant, nested, and unique.

To estimate the multivariate density, we use a finite overfitted Gaussian mixture model. This model adapts to heteroskedastic and non-convex disturbances. The mixture model can itself be used as a multiple-output quantile regression model when focusing on univariate conditional quantiles. Even though we only consider the multivariate Gaussian distribution as component distribution, the principle is also applicable to a model using a mixture of non-Gaussian distributions.

With the implementation via the Gaussian mixture model, we provide a widely applicable model for estimating multivariate and univariate (conditional) quantiles without the need to specify the exact multivariate density. The resulting multivariate quantiles are easy to interpret and apply, for example to study substitution effects. As such, we present a comprehensive and computationally attractive Bayesian framework that greatly simplifies the application of flexible multiple-output quantile regression models in practice.

8 Acknowledgments

We thank Andreas Alfons and Chen Zhou for their insightful comments. The paper was presented at ESOBE 2022 and COMPSTAT 2022. Comments by the participants are gratefully acknowledged. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

References

Ali, A., J. Z. Kolter, and R. J. Tibshirani (2016). The multiple quantile graphical model. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 29. Curran Associates, Inc.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah (2017). Julia: A fresh approach to numerical computing. *SIAM Review 59*(1), 65–98.

Bhattacharya, I. and S. Ghosal (2021). Bayesian multivariate quantile regression using dependent Dirichlet process prior. *Journal of Multivariate Analysis 185*, 104763.

Brown, P. J. and J. E. Griffin (2010). Inference with Normal-Gamma prior distributions in regression problems. *Bayesian Analysis 5*(1), 171–188.

Cai, Y. (2010). Multivariate quantile function models. *Statistica Sinica 20*, 481–496.

Carlier, G., V. Chernozhukov, A. Galichon, et al. (2016). Vector quantile regression: an optimal transport approach. *The Annals of Statistics 44*(3), 1165–1192.

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

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association 91*(434), 862–872.

del Barrio, E., A. G. Sanz, and M. Hallin (2022). Nonparametric multiple-output center-outward quantile regression.

Diebolt, J. and C. P. Robert (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 56*(2), 363–375.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The Annals of Statistics 2*(2), 267–277.

Escobar, M. D. and M. West (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association 90*(430), 577–588.

Frühwirth-Schnatter, S. and S. Pyne (2010, 01). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics 11*(2), 317–336.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association 96*(453), 194–209.

Guggisberg, M. (2022). A Bayesian approach to multiple-output quantile regression. *Journal of the American Statistical Association* (forthcoming), 1–27.

Hallin, M., E. Del Barrio, J. Cuesta-Albertos, and C. Matrán (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics 49*(2), 1139–1165.

Hallin, M., D. Paindaveine, and M. Šiman (2010). Multivariate quantiles and muliple-output regression quantiles: from L_1 optimization to halfspace depth. *The Annals of Statistics 38*(2), 635–703.

Hallin, M. and M. Šiman (2016). Elliptical multiple-output quantile regression and convex optimization. *Statistics & Probability Letters 109*, 232–237.

Hallin, M. and M. Šiman (2018). Multiple-output quantile regression. In R. Koenker, V. Chernozhukov, X. He, and L. Peng (Eds.), *Handbook of Quantile Regression*, pp. 185–207. Chapman and Hall/CRC Press.

Hartigan, J. A. (1987). Estimation of a convex density contour in two dimensions. *Journal of the American Statistical Association 82*(397), 267–270.

Hlubinka, D. and M. Šiman (2013). On elliptical quantiles in the quantile regression setup. *Journal of Multivariate Analysis 116*, 163–171.

Ishwaran, H., L. F. James, and J. Sun (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association 96*(456), 1316–1332.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics 9*, 155–176.

Koenker, R. and G. Bassett (1978). Regression Quantiles. *Econometrica 46*(1), 33–50.

Kong, L. and I. Mizera (2012). Quantile tomography: using quantiles with multivariate data. *Statistica Sinica 22*, 1589–1610.

Lin, T. I., J. C. Lee, and W. J. Hsieh (2007). Robust mixture modeling using the skew t distribution. *Statistics and Computing 17*(2), 81–92.

Malsiner-Walli, G., S. Frühwirth-Schnatter, and B. Grün (2016). Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing 26*, 303–324.

Müller, D. W. and G. Sawitzki (1991). Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association 86*(415), 738–746.

Müller, P., A. Erkanli, and M. West (1996). Bayesian curve fitting using multivariate Normal mixtures. *Biometrika 83*(1), 67–79.

Nobile, A. and A. T. Fearnside (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics and Computing 17*(2), 147–162.

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *The Annals of Statistics*, 855–881.

Richardson, S. and P. J. Green (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59*(4), 731–792.

Roeder, K. and L. Wasserman (1997). Practical Bayesian density estimation using mixtures of Normals. *Journal of the American Statistical Association 92*(439), 894–902.

Rousseau, J. and K. Mengersen (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(5), 689–710.

Sartorius, S. (2020). Bisection Method Root Finding. https://github.com/sky-s/bisection, last accessed on 2022-10-24.

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica 56*(2), 214–232.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of Statistics 28*(1), 40–74.

Taddy, M. A. and A. Kottas (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business & Economic Statistics 28*(3), 357–369.

The Math Works Inc. (2020). *MATLAB: Version 7.10.0 (R2020a)*. Natick, Massachusetts.

Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *Journal of the American Statistical Association 103*(481), 397–409.

Xu, P. (2020). Quantile Regression. https://github.com/zjph602xtc/Quantile_reg/releases/tag/1.0.3, last accessed on 2022-10-24.

Yau, C. and C. Holmes (2011). Hierarchical Bayesian nonparametric mixture models for clustering with variable relevance determination. *Bayesian Analysis 6*(2), 329.

Zuo, Y. and R. Serfling (2000). General notions of statistical depth function. *Annals of Statistics 28*(3), 461–482.

A Posterior Simulation Algorithm

1. Simulate mixture parameters conditional on $\boldsymbol{z}_n$ $(n = 1, \ldots, N,\ m = 1, \ldots, M)$:

   - Sample $\{\kappa_m\}$ from $\mathcal{D}(\bar{\rho}_1, \ldots, \bar{\rho}_M)$ where $\bar{\rho}_m = \rho_m + N_m$ and $N_m = \#\{n : z_n = m\}$, the number of observations allocated to the m-th component.

   - Sample $\{\boldsymbol{\mu}_m\}$ from $\mathcal{N}(\bar{\boldsymbol{v}}_m, \overline{\boldsymbol{V}}_m)$:
     - $\bar{\boldsymbol{v}}_m = \overline{\boldsymbol{V}}_m(\overline{\boldsymbol{V}}_0^{-1}\bar{\boldsymbol{v}}_0 + \boldsymbol{\Sigma}_m^{-1} N_m \bar{\boldsymbol{y}}_m)$ with $\bar{\boldsymbol{y}}_m = \frac{1}{N_m} \sum_{n:z_n=m} (y_n - \boldsymbol{B}_m \boldsymbol{x}_n)$.
     - $\overline{\boldsymbol{V}}_m = (\overline{\boldsymbol{V}}_0^{-1} + \boldsymbol{N}_m \Sigma_m^{-1})^{-1}$

   - Sample $\{\boldsymbol{B}_m\}$ from $\mathcal{N}(\boldsymbol{c}_m, \boldsymbol{C}_m)$:
     - $\boldsymbol{c}_m = \boldsymbol{C}_m\big(\boldsymbol{C}_0^{-1}\boldsymbol{c}_0 + \sum_{n:z_n=m} (\boldsymbol{x}_n \otimes \boldsymbol{I})\boldsymbol{\Sigma}_m^{-1}(\boldsymbol{y}_n - \boldsymbol{\mu}_m)\big)$
     - $\boldsymbol{C}_m = (\boldsymbol{C}_0^{-1} + \sum_{n:z_n=m} \boldsymbol{x}_n \boldsymbol{x}_n' \otimes \boldsymbol{\Sigma}_m^{-1})^{-1}$

   - Sample $\{\boldsymbol{\Sigma}_m\}$ from $\mathcal{IW}(\boldsymbol{S}_m, s_m)$:
     - $s_m = s_0 + \ N_m$
     - $\boldsymbol{S}_m = \boldsymbol{S}_0 + \sum_{n:z_n=m} (\boldsymbol{y}_n - \boldsymbol{\mu}_m - \boldsymbol{B}_m \boldsymbol{x}_n)(\boldsymbol{y}_n - \boldsymbol{\mu}_m - \boldsymbol{B}_m \boldsymbol{x}_n)'$

2. Sample z_n to classify observations conditional on mixture parameters $(n = 1, \ldots, N)$:

   - $\pi_m \equiv \Pr[z_n = m | \boldsymbol{y}_m, \boldsymbol{\kappa}, \boldsymbol{\mu}, \boldsymbol{B}, \boldsymbol{\Sigma}] \propto \kappa_m \phi\big(\boldsymbol{y}_n; g_m(\boldsymbol{x}_n), \boldsymbol{\Sigma}_m\big)$.

   - Sample $\{z_n\}$ from $\mathcal{M}(\pi_1, \ldots, \pi_M)$.

3. Sample hyperparameters:

   - Sample $\{\bar{\rho}_m\}$ simulatneously via a random walk Metropolis Hastings step with proposal density $\log(\rho_m) \sim \mathcal{N}(\log(\rho_m), s_{\rho_m}^2)$ from

$$p(\bar{\rho}_m|\boldsymbol{\kappa}) \propto p(\bar{\rho}_m)\frac{\Gamma(M\bar{\rho}_m)}{\Gamma(\bar{\rho}_m)^M}\left(\prod_{m=1}^{M}\kappa_m\right)^{\bar{\rho}_m-1}$$

- Sample $\{\lambda_k\}$ from a generalized inverted Gamma distribution $\mathcal{GIG}(\underline{b}_1 - M/2, 2\underline{b}_2, \delta_k)$ where $\delta_k = \sum_{m=1}^{M}(\mu_{m,k} - \bar{v}_{0,k})^2/R_k^2$ is the distance between the component locations and prior means.

- Sample $\bar{\boldsymbol{v}}_0$ from $\mathcal{N}\big(\sum_{m=1}^{M}\boldsymbol{\mu}_k/M, \overline{\boldsymbol{V}}_0/M\big)$ with $\overline{\boldsymbol{V}}_0 = \text{diag}(R_1^2\lambda_1, \ldots, R_K^2\lambda_K)$.

If no observation is allocated to component m during classification in step 2, all component-specific parameters of this empty component are sampled from their priors.

SUPPLEMENTARY APPENDIX
to Multivariate quantile regression using superlevel sets of conditional densities

Population Quantile Functions The conditional population quantiles from the simulation exercise are based on the distribution of $\boldsymbol{y}_n$ given the true parameters. We consider each time the quantile of a focal response variable k conditional on all other variables in the output-vector, that is, $\mathcal{K} = \{k\}$ and $\mathcal{C} = \{1, \ldots, K\} \setminus \{k\}$.

1. Multivariate Gaussian:

$$Q_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}\big(\alpha\big) = \mu^{k|\mathcal{C}}(\boldsymbol{y}_\mathcal{C}) + \Sigma^{k|\mathcal{C}}\Phi^{-1}(\alpha),$$

   with conditional mean $\mu^{k|\mathcal{C}}(\boldsymbol{y}_\mathcal{C})$ and conditional variance $\Sigma^{k|\mathcal{C}}$:

$$\mu^{k|\mathcal{C}}(\boldsymbol{y}_\mathcal{C}) = \mu_k + \Sigma_{k,\mathcal{C}}\Sigma^{-1}_{\mathcal{C},\mathcal{C}}\big(\boldsymbol{y}_\mathcal{C} - \mu_\mathcal{C}\big),$$
$$\Sigma^{k|\mathcal{C}} = \Sigma_{kk} - \Sigma_{k,\mathcal{C}}\Sigma^{-1}_{\mathcal{C},\mathcal{C}}\Sigma_{\mathcal{C},k}.$$

2. Multivariate Student-t:

$$Q_{Y_k|\boldsymbol{Y}_\mathcal{C}=\boldsymbol{y}_\mathcal{C}}\big(\alpha\big) = \mu^{k|\mathcal{C}}(\boldsymbol{y}_\mathcal{C}) + \Sigma^{k|\mathcal{C}}F^{-1}(\alpha; r^k),$$

   where $F^{-1}(\cdot)$ is the inverse of the cumulative distribution function of a Student-t with degrees of freedom $r^k = r + K - 1$, conditional mean $\mu^{k|\mathcal{C}}(\boldsymbol{y}_\mathcal{C})$ equivalent to the multivariate Gaussian, and conditional variance $\Sigma^{k|\mathcal{C}}$:

$$\Sigma^{k|\mathcal{C}} = \frac{r + \big(\boldsymbol{y}_\mathcal{C} - \mu_\mathcal{C}\big)\Sigma^{-1}_{\mathcal{C},\mathcal{C}}\big(\boldsymbol{y}_\mathcal{C} - \mu_\mathcal{C}\big)'}{r^k}\big(\Sigma_{kk} - \Sigma_{k,\mathcal{C}}\Sigma^{-1}_{\mathcal{C},\mathcal{C}}\Sigma_{\mathcal{C},k}\big),$$

   (see, e.g., Ding, 2016).

3. Multivariate log-Gaussian:

$$Q_{Y_k|\boldsymbol{Y}_\mathcal{C}}(\alpha) = \exp\{\mu^{k|\mathcal{C}}(\log \boldsymbol{y}_\mathcal{C}) + \Sigma^{k|\mathcal{C}}\Phi^{-1}(\alpha)\},$$

   with $\mu^{k|\mathcal{C}}(\cdot)$ and $\Sigma^{k|\mathcal{C}}$ equivalent to the multivariate Gaussian.

4. Conditional heteroskedasticity:

   We approximate the population quantile with the conditional quantile of a multivariate $M = 1,000$ component mixture of Gaussians, with components $\boldsymbol{\mu}_m = \boldsymbol{\mu}$, $\Omega_m = \exp(z_m)\boldsymbol{\Sigma}$ where z_m is simulated from $\mathcal{N}(0,1)$, and equal mixture weights $\kappa_m = 1/M$. The calculation of the quantiles follows Algorithm 2 and Eq. (18)–(21).

Further Simulation Results Figures 1 to 3 provide the conditional quantiles for the three remaining data generating processes in Section 5. Again, we compare the estimated versus true conditional quantiles of response variable y_1 given a range of values of y_2 and fixed values for y_3 for $\alpha \in \{.2, .4, .6, .8\}$. All settings are investigated with three methods:

- QReg with $\boldsymbol{y}_{(-k)}$: univariate linear quantile regression with the other response variables in the conditioning set,
- QReg with $\boldsymbol{y}_{(-k)}$, $(\boldsymbol{y}_{(-k)})^2$: univariate non-linear quantile regression with the other level and squared response variables in the conditioning set, and
- MQReg: multivariate Gaussian mixture quantile regression.

Figure 1: Estimated (dashed lines) vs. true (solid lines) conditional quantiles of DGP 1, a multivariate Gaussian response distribution, for $\alpha \in \{.2, .4, .6, .8\}$

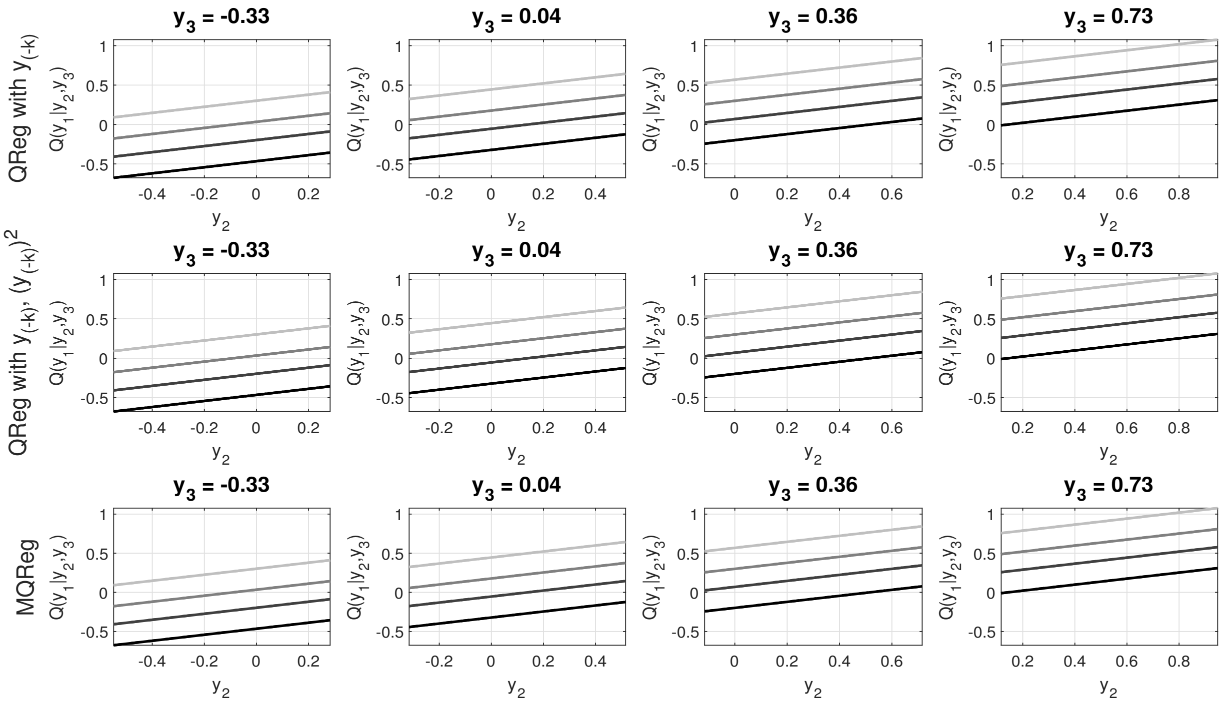


Figure 2: Estimated (dashed lines) vs. true (solid lines) conditional quantiles of DGP 2, a multivariate Student-t response distribution, for $\alpha \in \{.2, .4, .6, .8\}$

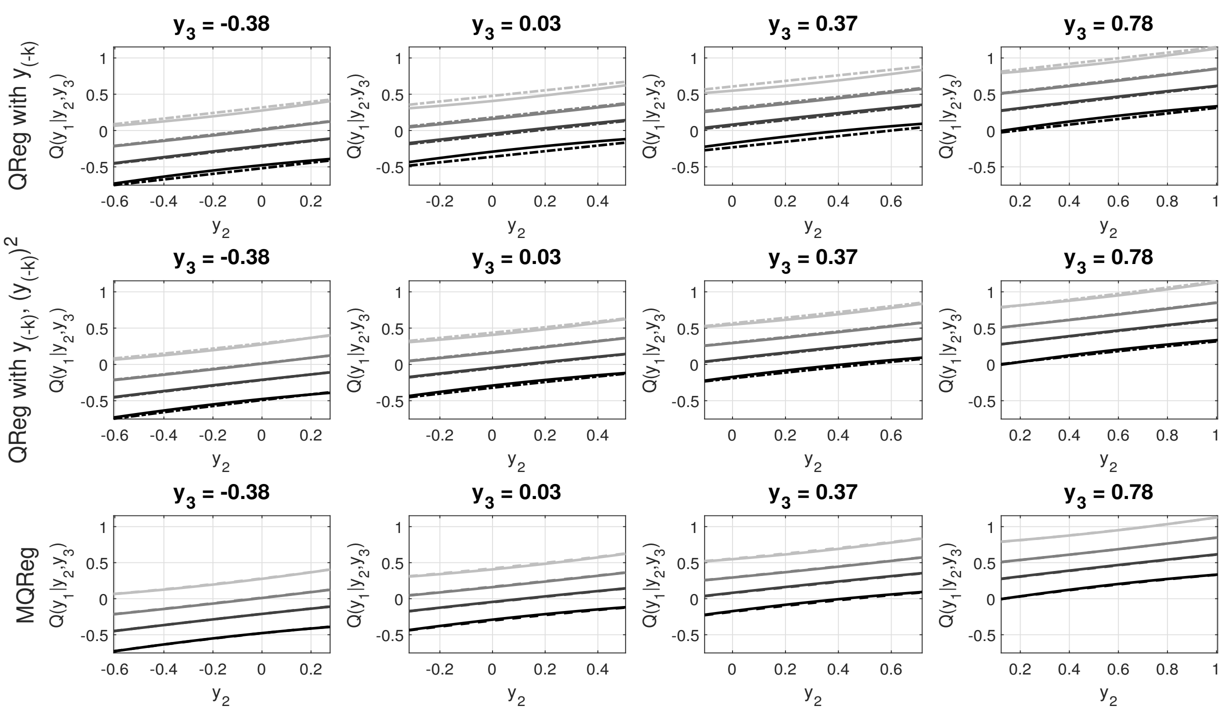
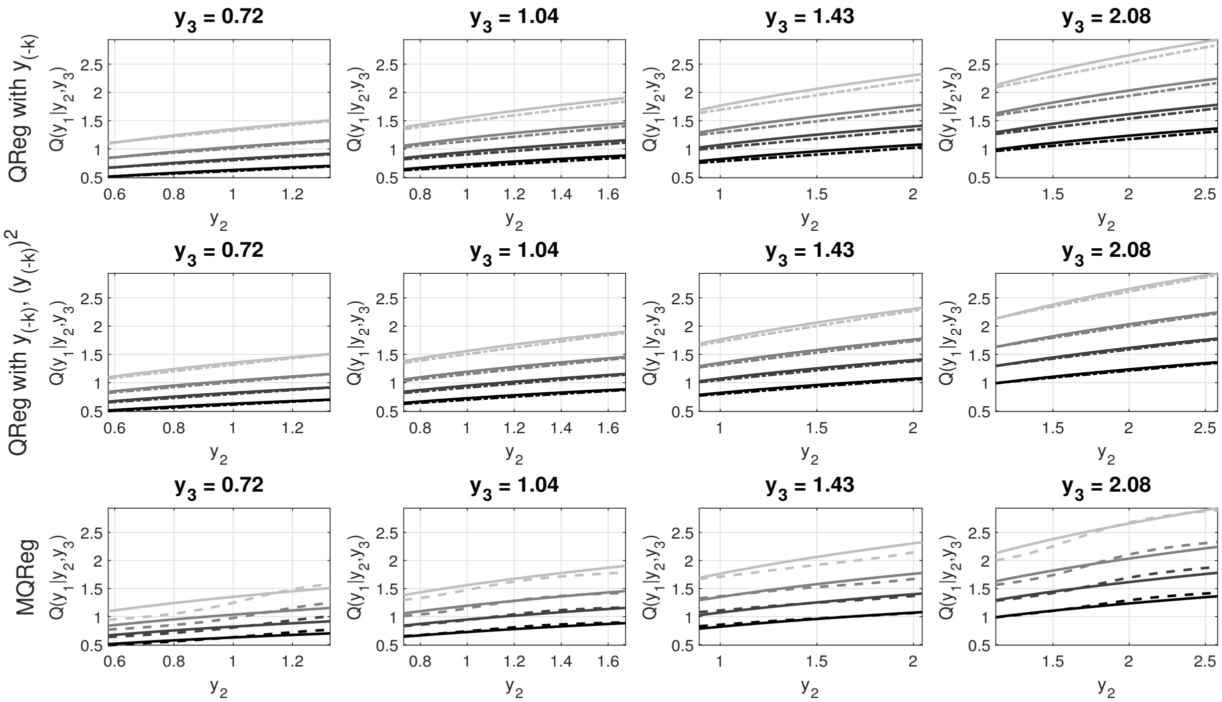


Figure 3: Estimated (dashed lines) vs. true (solid lines) conditional quantiles of DGP 3, a multivariate log-Gaussian response distribution, for $\alpha \in \{.2, .4, .6, .8\}$



References

Ding, P. (2016). On the Conditional Distribution of the Multivariate t Distribution. *The American Statistician 70*(3), 293–295.