Università degli Studi di Cagliari

Dipartimento di Matematica e Informatica
Dottorato di Ricerca in Matematica e Calcolo Scientifico
Ciclo XXVII

Ph.D. Thesis

Parametric modeling of dependence of bivariate quantile regression residuals' signs

S.S.D. SECS-S/01 STATISTICA

Candidate
Silvia Columbu

Supervisor
Prof. Monica Musio

PhD Coordinator
Prof. Giuseppe Rodriguez

Final examination academic year 2013/2014

Abstract

In this thesis, we propose a non-parametric method to study the dependence of the quantiles of a multivariate response conditional on a set of covariates. We define a statistic that measures the conditional probability of concordance of the signs of the residuals of the conditional quantiles of each univariate response. The probability of concordance is bounded from below by the value of largest possible negative dependence and from above by that of largest possible positive dependence. The value corresponding to the case of independence is contained in the interior of that interval. We recommend two distinct regression methods to model the conditional probability of concordance. The first is a logistic regression with a logit link modified. The second one is a nonlinear regression method, where the outcome is modeled as a polynomial function of the linear predictor. Both are conceived to constrain the predicted probabilities to lie within the feasible range. The estimated probabilities can be tested against the values of largest possible dependence and independence. The method permits to capture important aspects of the dependence of multivariate responses and assess possible effects of covariates on such dependence. We use data on pulmonary disfunctions to illustrate the potential of the proposed method. We suggest also graphical tools for a correct interpretation of results.

Declaration

I declare that to the best of my knowledge the contents of this thesis are original and my work except where indicated otherwise.

Aknowledgments

In first instance, I would like to thank my supervisor Monica Musio for supporting and encouraging me during all these years. She gave me the opportunity to start my career in the research.

I want to express my best gratitude to Matteo Bottai, who offered me the opportunity to collaborate with him and his research team at Karolinska Institutet in Stockholm. His advices and research suggestions were really precious.

I am also thankful to all my colleagues and friends that shared with me the experience of the PhD thesis, with constant encouragement and for helping to reduce the daily stress. Among them I want to mention Marianna, Rossella and Valentina, that were always ready for consultations and to cheer me up in the moments of discomfort.

A special mention goes to Paolo, Celia, Cecilia and all my colleagues in Stockholm. They helped me to get adapted in a completely new environement. They offered me their support at work sharing their knowledge with me and, not less important, they are good friends in life.

I would also like to thank all members of the Department of Mathematics of University of Cagliari.

I especially want to thank my boyfriend Giancarlo for having been patiently at my side during this experience and for encouraging me to continue my work when it was too hard to do so.

Finally, but most importantly, I would like to thank my parents and my sisters for their constant support over all my studies and for believing in me in this project.

Contents

List of Figures

List of Tables

Introduction

In univariate data analysis, quantile regression has become a popular and useful technique to characterize the impact of explicative variables on the whole distribution of variables of interest [22]. Quantile regression has many advantages compared to the standard linear regression. It permits a complete description of the conditional distribution, it is robust to the presence of outlyers values, it is invariant to non decreasing monotonic transformations of the outcome. Furthermore it does not require any parametrical assumptions on the distribution of the error terms.

When analysing the effect of a set of variables on more than one outcome simultaneously univariate regression methods are not completely satisfactory. Indeed, correlations and associations connecting dependent variables might be determinant to the correct detection of covariates effects and of their interactions. The techniques used must then guarantee the contemplation of existing dependence structures in the data. The lack of a univoquous definition of multidimensional quantiles poses an obstacle to the extension to multiple-outputs in quantile regression. Despite the huge literature devoted to the topic (see for instance the review in [33] and other related papers [5], [6], [11], [17], [27], [28] and [39]), to the best of our knowledge, none of the methods proposed up to now has shown clear advantages over the others.

In this dissertation we present an alternative approach to study the multivariate associations in quantile frameworks. Our proposal does not require any definition of multivariate quantile. We suggest the use of a method based on the simultaneous analysis and comparison of univariate quantile residuals.

The method proposed is innovative, appealing from an applied perspective and offers an extension to the consideration of other multivariate structures.

The first part of this thesis is a literature review; it does not contain any original material. Chapter 1 is an introduction to quantile regression with focus on the aspects useful for the developement of the method proposed. Chapter 2 reviews the existing methods to extend quantile regression to multivariate settings. Chapter 3 is a compendium of association structures and indexes to their measure.

The original contribution of the thesis is contained entirely in the second part. In Chapter 4 we introduce the method of bivariate quantile residuals signs dependence. We define an index of concordance of quantile regressions residuals analyzing its properties. In Chapter 5 we describe two regression methods for modelong the conditional probability. Some mathematical asymptotic properties are proved and simulation studies to validate finite sample properties are reported. In Chapter 6 we illustrate the insights of the method developed through the application to a real dataset on pulmonary disfunctions. We suggest the use of some graphical tools, and a statistical test to assess independence among quantile residuals. Finally, Chapter 7 summarizes the findings of our work and suggests further extensions of the method.

Part I

Background and literature review

Chapter 1

Introduction to Quantile Regression

In this chapter we present the method of linear quantile regression, paying particular attention to the aspects of it that will be used throguhout the thesis.

Quantile regression is a statistical tool developed to study the distribution of a dependent variable conditionally to a set of covariates of interest. In this sense, it can be viewed as complement and extension of least squares methods, used to estimate and make inference on the centrality of the distribution through the conditional mean. It is a semiparametric method since it enables to make inference on conditional quantiles without relying on any parametric distributional form of the error terms.

Quantile regression, in its main idea, dates back to early 1757, when it was introduced by Boscovich [3]. He was the first to prove that the median of a variable can be derived as the solution to the minimization of a sum of absolute deviations.

Because of the difficulties related to the computation of absolute values the method remained mostly unused until the introduction of linear programming in the 1950's. Indeed the simplicity of least squares, introduced later by Legendre in 1805, is the reason why, up to the present, this technique is the most known and used in statistical inference problems.

1.1 Quantiles and Quantile function

Given a random variable Y, its probability distribution is usually described by considering the probability density function (*pdf*) and the cumulative distribution function (*cdf*). The *cdf* of Y is defined as

$$F_Y(y) = P(Y \leq y).$$

Another quantity that can be helpful to explore the probability Y is the quantile function.

Definition 1. *Let us consider the left limit for the cdf* $F_Y(y^-) = \lim_{t\uparrow y} F_Y(t)$. *Given* $\tau \in (0,1)$, *a value of* y *such that*

$$F_Y(y^-) = P(Y < y) \leq \tau \quad \textit{and} \quad F_Y(y) = P(Y \leq y) \geq \tau$$

is called quantile of order τ *of the distribution. A quantile is a value that splits the distribution in two complementary regions.*

There is a sort of inverse relation between quantiles and the values of the *cdf.* This relation cannot be defined in general as an ordinary inverse function since the distribution function is not always a bijection. It can then be possible to have more than one value τ associated to a single y value. This is always the case with discrete random variables, an example of that situation is illustrated in figure 1.1 Left. For this reason it is necessary to define a generalized inverse of the distribution function.

We therefore define the quantile function as

$$Q_Y(\tau) = F^{-1}(\tau) = \inf\{y \in \mathbb{R} | F_Y(y) \geq \tau\}.$$

For $\tau = 0.5$, $Q_Y(0.5)$ is the median.

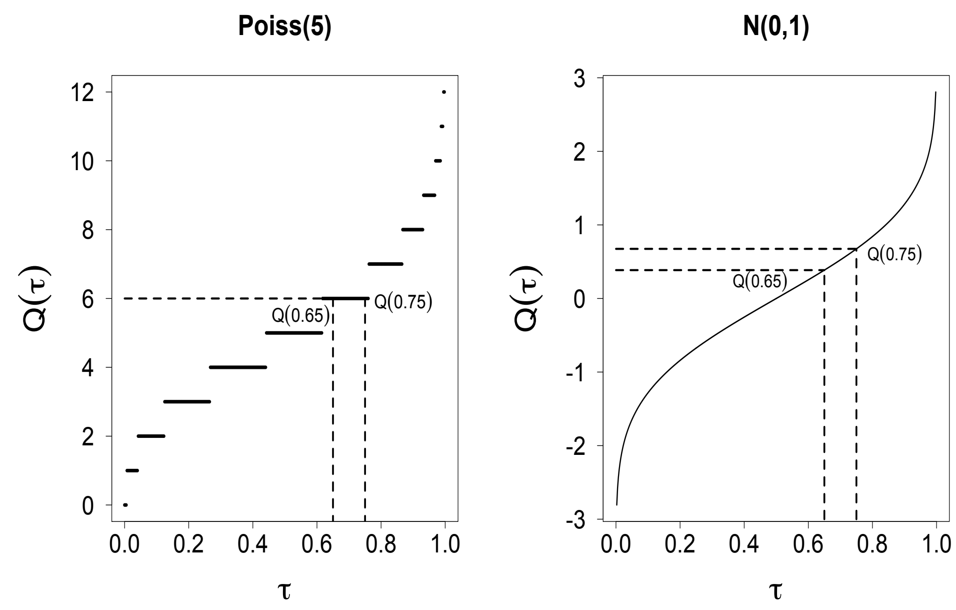

Figure 1.1: Left: quantile function of a discrete random variable (Poisson with parameter $\lambda = 5$). Right: quantile function for a continuous random variable (standard Normal).

A τth quantile of a random variable Y can be also obtained as the minimizer of an expected loss function.

Let us define the following asymmetric piecewise loss linear function

$$\begin{aligned}\rho_\tau(u) &= \begin{cases} \tau u & \text{for } u > 0 \\ (\tau - 1)u & \text{for } u \leq 0 \end{cases} \\ &= u(\tau - \mathbb{I}(u \leq 0))\end{aligned} \tag{1.1}$$

which is represented in figure 1.2 for different values of τ.

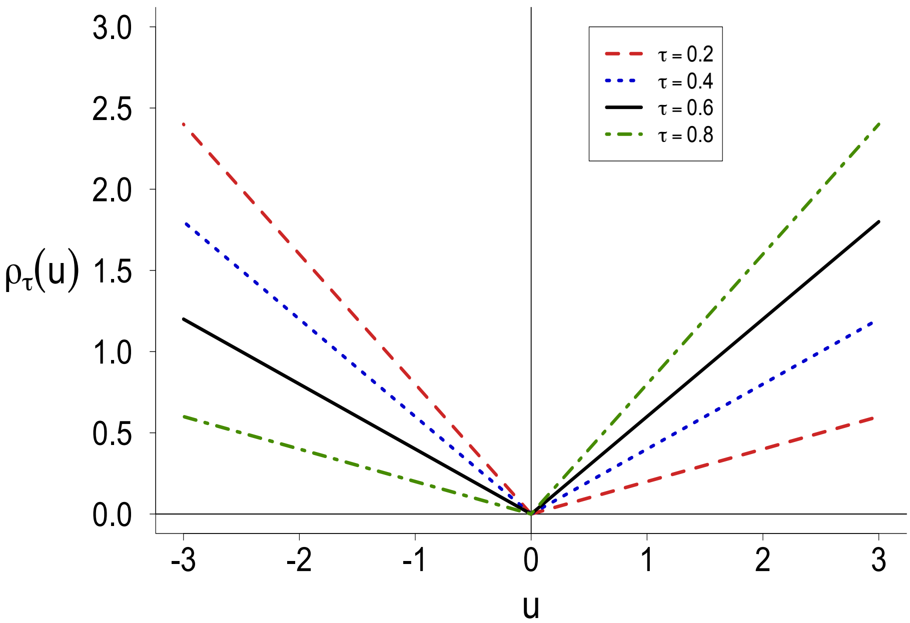

Figure 1.2: Piecewise loss function.

The defined loss function, for $\tau \neq 0.5$, assigns different weights to positive and negative values of u. It reduces to the simple absolute value when $\tau = 0.5$. It has also the property of being convex, this last result arising from the triangle inequality applied to absolute values as the norm on the real line.

We would like to find a value ξ which minimizes the expected loss:

$$\begin{aligned}E(\rho_\tau(Y - \xi)) &= \int_{-\infty}^{\infty} \rho_\tau(y - \xi) d\xi \\ &= \tau \int_{y>\xi} (y - \xi) f_Y(y) dy + (\tau - 1) \int_{y<\xi} (y - \xi) f_Y(y) dy \\ &= (\tau - 1) \int_{-\infty}^{\xi} y f_Y(y) dy - \xi(\tau - 1) F_Y(\xi) + \tau \int_{\xi}^{\infty} y f_Y(y) dy - \xi\tau[1 - F_Y(\xi)]\end{aligned}$$

$$= \tau\mathbb{E}(Y) - \int_{-\infty}^{\xi} y f_Y(y)dy - \xi[\tau - F_Y(\xi)].$$

Taking the derivative with respect to ξ and equating to zero we get

$$0 = \frac{d}{d\xi}\left[\tau E[Y] - \int_{-\infty}^{\xi} y f_Y(y)dy - \xi[\tau - F_Y(\xi)]\right] = -\xi f_Y(\xi) - [\tau - F_Y(\xi)] + \xi f_Y(\xi) = -\tau + F_Y(\xi).$$

$E(\rho_\tau(Y-\xi))$ being a convex function, we have that its minimum is attained at any ξ such that $\tau = F_Y(\xi)$. When the solution is unique $\xi = F_Y^{-1}(\tau)$, otherwise we have an interval of τ-th quantiles from which, by convention, the smallest must be chosen.

Let us consider n independent and identical distributed (i.i.d.) variables $Y_1, \dots, Y_n$ with *cdf* $F_Y(y)$ and let consider $(y_1, \dots, y_n)$ one possible realization of it. If we replace the *cdf* F_Y by the empirical distribution function

$$F_n(y) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}(Y_j \leq y_j)$$

we can derive the sample τ-th quantile ξ of Y, in an analogous way, by minimizing the sample loss expected function

$$\begin{aligned} \min_{\xi\in\mathbb{R}} \int \rho_\tau(y-\xi)dF_n(y) &= \min_{\xi\in\mathbb{R}} \frac{1}{n}\sum_{j=1}^{n} \rho_\tau(y-\xi) \qquad (1.2) \\ &= \min_{\xi\in\mathbb{R}} \frac{1}{n}\left[\sum_{j:y_j>\xi} \tau(y_j-\xi) + \sum_{j:y_j\leq\xi} (\tau-1)(y_j-\xi)\right] \end{aligned}$$

The sample median is simply the minimizer of the sum of absolute deviations

$$\min_{\xi\in\mathbb{R}} \sum_{j=1}^{n} |y_j - \xi|.$$

The sample piecewise asymmetric loss is convex since is the sum of convex functions.

1.2 Linear quantile regression

Let us suppose to have a sample of $Y_1, \dots, Y_n$ i.i.d. distributed according to a *cdf* F whose shape is not precisely known.

Let $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_q\}$ be an $n \times q$ matrix of observed variables. We would like to establish a linear relationship between the τth quantile of $\{y_1, \dots, y_n\}$, observations from the random sample $Y_1, \dots, Y_n$, and the matrix of covariates $\mathbf{X}$, by the estimation of an unknown q-parameter vector $\boldsymbol{\beta}_\tau = \{\beta_{1,\tau}, \ldots, \beta_{q,\tau}\}$. We assume the relationship to be described by the following linear model

$$\begin{aligned} y_j = \mathbf{x}_j^T \boldsymbol{\beta}_\tau + \varepsilon_j, \quad j = 1, \ldots, n \\ \varepsilon_j \text{ i.i.d } u_j \sim F \\ Q_{y_j}(\tau|\mathbf{x}_j) = \mathbf{x}_j^T \boldsymbol{\beta}_\tau \text{ for any } \tau \in (0,1). \end{aligned} \tag{1.3}$$

where $\mathbf{x}_j$ is a row of the design matrix $\mathbf{X}$.

We make the additional assumption that $Q_{\varepsilon_j}(\tau|\mathbf{x}_j) = 0$, where ε_i are random variables representing the errors of the model.

The regression parameter vector $\boldsymbol{\beta}_\tau$ can be estimated through the following minimization problem

$$\hat{\boldsymbol{\beta}}_\tau = \min_{\boldsymbol{\beta}_\tau \in \mathbb{R}^q} \sum_{j=1}^{n} \rho_\tau(y_j - \mathbf{x}_j^T \boldsymbol{\beta}_\tau). \tag{1.4}$$

where ρ_τ is the loss piecewise function defined in 1.1. When $q = 1$ and $\mathbf{x}_j = 1$ for all j, the quantile regression minimization problem coincides to the location model defined in the previous section.

Note that the objective function

$$R(\boldsymbol{\beta}_\tau) = \sum_{j=1}^{n} \rho_\tau(y_j - \mathbf{x}_i^T \boldsymbol{\beta}_\tau) \tag{1.5}$$

is continuous and differentiable except for the points where one or more residuals are zero. This partial nondifferentiability, implies that the usual first order condition is not sufficient for the determination of its minimum.

At points of nonderivability, $R(\boldsymbol{\beta}_\tau)$ has partial derivatives in all directions [22], depending, however, on the directions of evaluation.

The directional derivative at direction $\mathbf{w}$ is

$$\begin{aligned} \nabla R(\boldsymbol{\beta}_\tau, \mathbf{w}) &= \frac{d}{dt} R(\boldsymbol{\beta} + t\mathbf{w})|_{\mathbf{t=0}} \\ &= \sum_{j=1}^{n} (y_j - \mathbf{x}_j^T \boldsymbol{\beta} - x_j^T tw)[t - (\mathbb{I}(y_j - x_j^T \boldsymbol{\beta} - \mathbf{x}_j^T t\mathbf{w}) < \mathbf{0})]|_{\mathbf{t=0}} \\ &= -\sum \Psi_\tau^*(y_j - \mathbf{x}_j^T \boldsymbol{\beta}, -\mathbf{x}_j^T \mathbf{w}) \mathbf{x_j^T} \mathbf{w}, \end{aligned}$$

where

$$\Psi_\tau^* = \begin{cases} \tau - \mathbb{I}(\mathbf{u} < \mathbf{0}) & \text{if } \mathbf{u} \neq \mathbf{0} \\ \tau - \mathbb{I}(\mathbf{v} < \mathbf{0}) & \text{for } \mathbf{u} = \mathbf{0} \end{cases}$$

If at a point $\hat{\boldsymbol{\beta}}_\tau$ the directional derivatives are all nonnegative

$$\nabla R(\hat{\boldsymbol{\beta}}_\tau, \mathbf{w}) \geq 0 \quad \text{for all } \mathbf{w} \in \mathbb{R}^q \quad \text{with } ||\mathbf{w}|| = 1$$

then $\hat{\boldsymbol{\beta}}_\tau$ minimizes $R(\boldsymbol{\beta}_\tau)$. This is a natural generalization of simply setting $\nabla R(\boldsymbol{\beta}_\tau) = 0$ when R is smooth. It simply requires that the function is increasing as we move away from the point $\hat{\boldsymbol{\beta}}_\tau$ regardless of the direction in which we move.

Note that the objective function in (1.5) can also be interpreted as a strictly proper scoring rule for quantiles (see [14] for an exhaustive description of the topic).

1.3 Quantile regression as a linear program

The quantile regression problem in (1.4) can be reformulated and solved as a linear programming problem. Wagner, in 1959 [41], was the first to formalize this result.

In this section we will introduce some notions of linear programming to interpret the quantile regression in these terms. For further details on linear programming see Hillier and Lieberman (2005) [20], whereas for the connection to quantile regression we refer to Koenker (2004) [22].

In general notation, a standard linear program is an optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{b}^T\mathbf{x} \tag{1.6}$$

subject to linear constraints

$$\mathbf{A}\mathbf{x} = \mathbf{c}$$

$$\mathbf{x} \geq 0$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^n$, and $\mathbf{b}^T\mathbf{x}$ is the objective function. The matrix $\mathbf{A}$ of constraints is such that $rank(\mathbf{A}) = rank(\mathbf{A}|\mathbf{c}) = m$.

Definition 2. *A vector $\mathbf{x}^* \in \mathbb{R}^n$ is said a feasible solution of* (1.6) *if it satisfies all the constraints. The set of all such vectors constitutes the set of all feasible solutions.*

Definition 3. *A vector $\mathbf{x}^{**} \in \mathbb{R}^n$ is said an optimal solution to* (1.6) *if it is a feasible solution of it and $\mathbf{b}^T\mathbf{x}^{**} \leq \mathbf{b}^T\mathbf{x}$.*

Let $\mathbf{A}(h) \in \mathbb{R}^{m\times m}$ be a submatrix formed using m linearly independent columns of $\mathbf{A}$, where $h \in \mathcal{H}$ indicates the subset of m-columns selected. We will indicate with $\bar{h}$ the complementary set associated to the remaining $n-m$ columns. The system of equality constraints $\mathbf{A}\mathbf{x} = \mathbf{c}$ can be then rewritten as

$$\begin{pmatrix} \mathbf{A}(h) & \mathbf{A}(\bar{h}) \end{pmatrix} \begin{pmatrix} \mathbf{x}(h) \\ \mathbf{x}(\bar{h}) \end{pmatrix} = \mathbf{c}. \tag{1.7}$$

For $\mathbf{x}(\bar{h}) = 0$ we have

$$\mathbf{x}(h) = \mathbf{A}(h)^{-1}\mathbf{c}.$$

Definition 4. *The vector*

$$\begin{pmatrix} \mathbf{x}(h) \\ \mathbf{x}(\bar{h}) \end{pmatrix} = \begin{pmatrix} \mathbf{A}(h)^{-1}\mathbf{c} \\ 0 \end{pmatrix}$$

is a solution of the system of constraints (1.7) *and is called basic solution. If* $\mathbf{x} \geq 0$*, then the basic solution is feasible.*

Each basic solution is associated to a base of the matrix $\mathbf{A}$.

Theorem 1. *For each* $\mathbf{b} \in \mathbb{R}^n$ *such that the linear program 1.6 has an optimum, there exists an optimal solution* $\mathbf{x}^{**}$ *which is a feasible basic solution.*

Note that the number of feasible basic solutions is bounded from above by $\binom{n}{m}$, which represents all possible combinations of m columns among n. Indeed, not all such combinations correspond to basic solutions; the columns could be linearly dependent or the corresponding basic solution could be not feasible.

Let us rewrite the quantile regression model in (1.3) as

$$y_j = \mathbf{x}_j^T \boldsymbol{\beta}_\tau + \varepsilon_j = \mathbf{x}_j^T \boldsymbol{\beta}_\tau + (u_j - v_j)$$

where $u_j = \varepsilon_j \mathbb{I}(\varepsilon_j > 0)$ and $v_j = |\varepsilon_j| \mathbb{I}(\varepsilon_j < 0)$.

The quantile minimization problem in (1.4)

$$\min_{\boldsymbol{\beta}_\tau \in \mathbb{R}^q} \sum_{j=1}^{n} \rho_\tau(y_j - \mathbf{x}_j^T \boldsymbol{\beta}_\tau)$$

can be reformulated in terms of a standard linear program as

$$\min_{\boldsymbol{\beta}_\tau, \mathbf{u}, \mathbf{v}} [\tau \imath^T \mathbf{u} + (1-\tau) \boldsymbol{\imath}^T \mathbf{v}] \tag{1.8}$$

subject to

$$\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_\tau = \mathbf{u} - \mathbf{v}$$

and with nonnegativity constraints

$$\mathbf{u} \geq 0 \text{ and } \mathbf{v} \geq 0.$$

Here $\boldsymbol{\beta}_\tau \in \mathbb{R}^q$ and $\boldsymbol{\imath}^T = (1, \ldots, 1)$ is an n-vector of ones.

To convert the linear quantile minimization problem in a linear programming problem we have made the following identifications: $\mathbf{b} = (O_q^T, \tau \boldsymbol{\imath}^T, (1-\tau)^T \boldsymbol{\imath}^T)$, $\mathbf{x} = (\mathbf{c}^T, bfu^T, bfv^T)^T$, $\mathbf{A} = [\mathbf{X}|\mathbf{I}| - \mathbf{I}]$, $\mathbf{c} = \mathbf{y}$.

A feasible basic solution for the linear program in (1.8), corresponding to the quantile regression problem, takes the form

$$\boldsymbol{\beta}_\tau(h) = \mathbf{X}^{-1}(h)\mathbf{y}(h)$$

$$\mathbf{u}(h) = \mathbf{v}(h) = 0.$$

This last result will be discussed in more detail in section 1.4.

In its reformulation, the quantile regression problem can be efficiently solved by the application of linear programming algorithms. Among them we recall the simplex method, for which we refer to Barrodale and Roberts (1974) [1], and the class of interior point methods based on a modification of the first one (see Koenker and D'Orey (1987) [25] and Koenker and Portnoy (1997) [24]).

1.4 Characterization of solutions: finite sample and asymptotic properties of quantile regression

The solutions of the regression quantile minimization problem have some nice finite sample and asymptotic properties. We remind some of them, the complete theory with proofs of cited theorems can be found in Koenker and Bassett (1978) [23] and Koenker (2004) [22].

It is important to observe that most of the following properties come from the fact that the quantile regression problem can be rewritten and solved as a linear programming problem as shown in section 1.3.

Let us remember that regressing a response variable on a set of q covariates $q \geq 2$, is geometrically equivalent to the determination of the parameters of a hyperplane.

Suppose

- $\mathbf{B}_\tau$ is the solution sets containing the elements $\hat{\boldsymbol{\beta}}_\tau$;
- $\mathcal{L} = \{1, \ldots, n\}$ is the set containing the indexes of the n observations;
- $\mathcal{H}$ is the set of all $\binom{n}{q}$ q-element subsets of $\mathcal{L}$, obtained by selecting q different elements each time.

Every $h \in \mathcal{H}$ has a relative complement $\bar{h} = \mathcal{L} \setminus h$ which is the set containing the $n - q$ elements of $\mathcal{L}$ not in h. Together h and $\bar{h}$ serve to partition the design matrix $\mathbf{X}$. Let $\mathbf{y}(h)$, for example, be a k-vector with coordinates $\{y_i | i \in h\}$ and $\mathbf{X}(h)$ the square matrix with rows $\{\mathbf{x}_i | i \in h\}$.

Finally, let

$$H = \{h \in \mathcal{H} | \text{ rank } \mathbf{X}(h) = q\}$$

be the subset of $\mathcal{H}$ whose elements h are such that the matrix $\mathbf{X}(h)$ is non singular.

The set $\mathbf{B}_\tau$ is never empty but the solution to the minimization problem in (1.4) is not unique in general.

Theorem 2 (Koenker and Bassett, 1978 [23]). *If $\mathbf{X}$ has full column rank q then the set of regression quantiles, $\mathbf{B}_\tau$, has at least one element of the form,*

$$\hat{\boldsymbol{\beta}}_\tau = \mathbf{X}(h)^{-1}\mathbf{y}(h)$$

for some $h \in H$. Moreover, $\mathbf{B}_\tau$, is the convex hull of all solutions of this form.

The theorem states that a regression hyperplane passes through at least q points between the n observed. This follows directly from the linear programming formulation of the minimization problem (section 1.3). This result gives us the opportunity to introduce an important property related to the sign of residuals.

Another important finite sample consequence of solving the quantile regression problem as a linear programming one, is that, provided that the design matrix contains a column of ones, there will be approximately $n(1-\tau)$ positive residuals and $n\tau$ negative ones. This is formally stated in the two following results.

Given $\hat{\boldsymbol{\beta}}_\tau \in \mathbf{B}_\tau$, a solution to the minimization problem let us denote with:

- $P(\mathbf{r\hat{e}s}_\tau)$ the number of positive residuals, $\mathbf{r\hat{e}s}_\tau = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\tau > 0$;
- $N(\mathbf{r\hat{e}s}_\tau)$ the number of negative residuals, $\mathbf{r\hat{e}s}_\tau = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\tau < 0$;
- $Z(\mathbf{r\hat{e}s}_\tau)$ the number of zero residuals, $\mathbf{r\hat{e}s}_\tau = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_\tau = 0.$;

Theorem 3. *Given $P(\mathbf{r\hat{e}s}_\tau)$, $N(\mathbf{r\hat{e}s}_\tau)$, and $Z(\mathbf{r\hat{e}s}_\tau)$, if $\mathbf{X}$ contains a column of ones, then for any $\hat{\boldsymbol{\beta}}_\tau \in \mathbf{B}_\tau$ we have*

$$N(\mathbf{r\hat{e}s}_\tau) \le n\tau \le n - P(\mathbf{r\hat{e}s}_\tau) = N(\mathbf{r\hat{e}s}_\tau) + Z(\mathbf{r\hat{e}s}_\tau) \tag{1.9}$$

and

$$P(\mathbf{r\hat{e}s}_\tau) \le n(1-\tau) \le P(\mathbf{r\hat{e}s}_\tau) + Z(\mathbf{r\hat{e}s}_\tau). \tag{1.10}$$

If $\hat{\boldsymbol{\beta}}_\tau$ is unique then the inequalities hold strictly.

To better understand the statement of the last theorem let us divide every member of the inequalities by $n = N(\mathbf{r\hat{e}s}_\tau) + P(\mathbf{r\hat{e}s}_\tau) + Z(\mathbf{r\hat{e}s}_\tau)$. This leads to the result expressed by the following corollary.

Corollary 1. *If $Z(\mathbf{r\hat{e}s}_\tau) = q$, then the proportion of negative residuals is approximately τ*

$$\frac{N(\mathbf{r\hat{e}s}_\tau)}{n} \le \tau \le \frac{N(\mathbf{r\hat{e}s}_\tau)}{n} + \frac{Z(\mathbf{r\hat{e}s}_\tau)}{n}$$

and the number of positive residuals is approximately $(1-\tau)$

$$\frac{P(\mathbf{r\hat{e}s}_\tau)}{n} \le 1-\tau \le \frac{P(\mathbf{r\hat{e}s}_\tau)}{n} + \frac{Z(\mathbf{r\hat{e}s}_\tau)}{n}.$$

It is worth remarking that if F is continuous then $Z(\mathbf{r\hat{e}s}_\tau) = q$ with probability one.

The quantile regression estimator defined in theorem 2 satisfies asymptotic distributional properties. These can be derived from the results concerning the large sample distribution of quantiles.

Theorem 4. *Let $\{Y_1, \ldots, Y_n\}$ be i.i.d. random variables from the distribution F and for some quantile $\xi_\tau = F^{-1}(\tau)$. Let us consider the sample quantile*

$$\hat{\xi}_\tau = \min_{\xi \in \mathbb{R}^q} \sum_{j=1}^{n} \rho_\tau(y_j - \xi). \tag{1.11}$$

If F is continuous and has continuous and positive density f at $\xi(\tau)$, then

$$\sqrt{n}(\hat{\xi}_\tau - \xi_\tau) \sim N\left(0, \frac{\tau(1-\tau)}{f^2(\xi_\tau)}\right).$$

The extension to the estimator of a quantile regression model is discussed in the following theorem.

Theorem 5. *Let $\hat{\beta}_{(n,\tau_1)}$ denotes an estimated regression quantile from model (1.3). Let $\xi_\tau = F^{-1}(\tau)$, $\boldsymbol{\xi_\tau} = (\xi_\tau, 0, \ldots, 0) \in \mathbb{R}^q$ and $\hat{\boldsymbol{\xi}}_{(n,\tau)} = \hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}$. Assume:*

(i) F is continuous and has continuous and positive density f at $\xi(\tau)$

(ii) $\mathbf{x}_{i1} = 1$, $i = 1, \ldots, n$ and $\lim_{n\to\infty} n^{-1}\mathbf{X}\mathbf{X}^T = \boldsymbol{Q}$, a positive definite matrix,

then

$$\sqrt{n}[\hat{\boldsymbol{\xi}}_{\boldsymbol{(n,\tau)}} - \boldsymbol{\xi_\tau}] \sim N_q\left(0, \frac{\tau(1-\tau)}{f(\xi_\tau)^2}\boldsymbol{Q}^{-1}\right). \tag{1.12}$$

1.5 Comparison between quantile regression and linear regression

We will compare quantile regression with the standard univariate linear regression trying to synthesize the advantages of the first one over the second.

We first recall that given a random sample $\{Y_1, \ldots, Y_n\}$ from a distribution F, an observed matrix of covariates $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$ and a set of q unknown parameters $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_q\}$, a linear regression model is written as

$$\begin{aligned} y_j = \mathbf{x}_j^T\boldsymbol{\beta} + \varepsilon_j, &\quad j = 1, \ldots, n \\ \varepsilon_j \text{ i.i.d } u_j &\sim N(0, \sigma^2) \\ \mathbb{E}(y_j|\mathbf{x}_j) &= \mathbf{x}_j^T\boldsymbol{\beta}. \end{aligned} \tag{1.13}$$

The estimates of the q-vector $\boldsymbol{\beta}$ is the result of the minimization of the sum of regression residuals

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta} \in \mathbb{R}^q} \sum_{j=1}^{n} (y_j - \mathbf{x}_j^T\boldsymbol{\beta})^2. \tag{1.14}$$

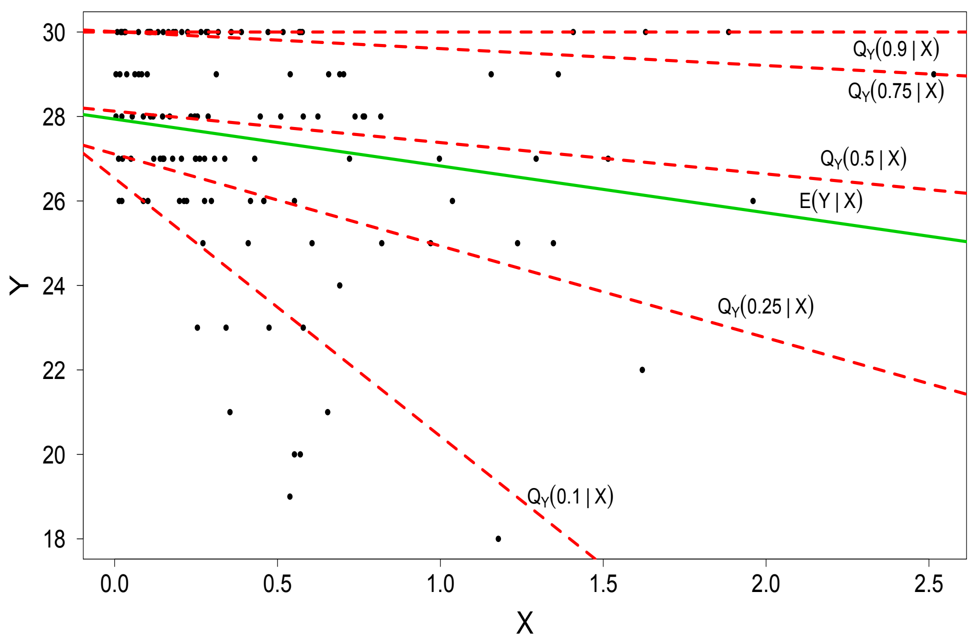

Figure 1.3: Example of linear quantile regression estimates compaired with OLS on the mean for a dataset with continuous response Y and covariate X. The distribution of Y is clearly asymmetric. The red dashed lines correspond to estimates for $\tau = \{0.10, 0.25, 0.50, 0.75, 0.90\}$, while the green solid line corresponds to the classical OLS estimate.

The problem in (1.14) can be easily solved getting a closed form of the parameter estimates

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

Then $\hat{\boldsymbol{\beta}}$ is a good estimator, in the sense that it satisfies the statistical properties of unbiasedness, consistency, efficiency and asymptotic normality.

The strength of OLS regression is related to its simplicity and completness, however there are some weaknesses in this method that can be overcome by the extension to the regression on quantiles.

The first limitation of linear regression, which we have seen when modeling the conditional mean of the dependent variable, is that it provides only a partial view of the relationship with a set of covariates. This might be not just incomplete, but also misleading if the research interest is more on the tails of the distribution. Quantile regression, on the other hand, permits inference on the entire shape of the conditional distribution, by modeling the covariates' effects on each quantile function separately, as summarized in figure 1.4.

Another important feature of quantile regression is that no distributional assumptions about the regression error terms are required. This implies that the method could be applied to any dataset regardless of the functional form, furthermore there are no difficulties in the presence of eteroskedasticity.

The other advantages of the method are strictly related to the definition of quantile itself. It is well known that in general the mean is sensible to the presence of outlyers, while quantiles are a robust measure to summarize the data in such conditions. This is perfectly reflected in the inference problem.

Finally we recall that quantiles and, as a consequence, inference on quantiles is invariant to increasing monotonic transformations h, while this is not true for the mean. This last property is really important, and together with other equivariance features will be analyzed more in detail in section 1.6.

1.6 Equivariance

Regression quantiles satisfy interesting properties of equivariance that can be really helpful for a coherent interpretation of regression results. Some of them are in common with least-squares regression, the last one that we will state is directly related to the definition of quantile itself.

Let us start by denoting by $\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X})$ the estimate of the τ-th regression quantile based on observations $(\mathbf{y}, \mathbf{X})$, with $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$ a $n \times q$ matrix. The following are four basic properties of the solution of a quantile regression problem.

Theorem 6. *(Koenker and Bassett, 1978 [23]) Let* $\mathbf{A}$ *be any* $q \times q$ *nonsingular matrix,* $\boldsymbol{\gamma} \in \mathbb{R}^q$*, and* $a > 0$*. Then, for any* $\tau \in [0, 1]$*,*

(i) $\hat{\boldsymbol{\beta}}(\tau; a\mathbf{y}, \mathbf{X}) = a\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X})$

(ii) $\hat{\boldsymbol{\beta}}(\tau; -a\mathbf{y}, \mathbf{X}) = -a\hat{\boldsymbol{\beta}}(1-\tau; \mathbf{y}, \mathbf{X})$

(iii) $\hat{\boldsymbol{\beta}}(\tau; \mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X}) = \hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) + \boldsymbol{\gamma}$

(iv) $\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}\mathbf{A}) = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X})$.

Properties (i) and (ii) state that if all values of the outcome $\mathbf{y}$ are multiplied by a quantity a, the solution will be modified at the same way. We talk of scale equivariance.

Property (iii) is usually called shift or regression equivariance. This property says that, when adding a linear combination of explicative variables to the dependent variable, the solution corresponds to the sum of the solution of the linear model with $\mathbf{y}$ as outcome plus the vector $\boldsymbol{\gamma}$ in the linear combination. In particular we observe that regression residuals for the two models coincide:

$$[\mathbf{y} + \mathbf{X}\boldsymbol{\gamma}] - \mathbf{X}\hat{\boldsymbol{\beta}}(\tau; \mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X}) \quad = \quad \mathbf{y} + \mathbf{X}\gamma - \mathbf{X}\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) + \boldsymbol{\gamma}$$

$$\begin{aligned} &= \mathbf{y} + \mathbf{X}\boldsymbol{\gamma} - \mathbf{X}\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}) - \mathbf{X}\boldsymbol{\gamma} \qquad (1.15) \\ &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\tau; \mathbf{y}, \mathbf{X}). \end{aligned}$$

Thus regression residuals are equivariant under transformations of this kind.

The last property, (iv), is called equivariance to reparametrization of design.

Another property, which is much stronger than those shown above, is the one known as equivariance to monotone transformations. Let h be a nondecreasing function on $\mathbb{R}$. For any random variable Y

$$Q_{h(y)}(\tau) = h\{Q_y(\tau)\},$$

meaning that the quantiles of the transformed outcome Y are the transformed quantiles of the original Y.

This follows from the simple probabilistic equality

$$P(Y \leq y) = P\{h(Y) \leq h(y)\}.$$

When convenient, one can then model directly the transformed $h(y)$ on the covariates via quantile regression. Hence, having estimated a linear model, $\mathbf{X}\hat{\boldsymbol{\beta}}_\tau$, for the conditional τth quantile of $h(y)$ given $\mathbf{X}$, we can interpret $h^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}}_\tau)$ as an appropriate estimate of the conditional τth quantile of $\mathbf{y}$ given $\mathbf{X}$. This property does not hold for the expectation operator because, in general, $E[h(y)] \neq h(E(y))$, except when h is linear.

This last property can be used to handle in a simple way different modelization problems. Among others we recall the method of Logistic Quantile Regression introduced by Bottai et al. in 2010 [4] for the modelization of bounded outcomes. This method is based on the use of a logistic transformation of the dependent variable, constructed in a way that predictions are constrained to lie within the feasible range of the bounded variable. For an application of the method see Columbu and Bottai (2015)[7].

Chapter 2

Multiple responses

Univariate regression methods can be extended to consider the effect of a set of covariates on more than one response variable simultaneously. Appropriate statistical techniques must be introduced for this purpose. In fact, ignoring the presence of a multivariate structure in the data, and then conducting separate analysis on each of the responses, would be exhaustive only if there is independence between them. Generally, the candidate response variables in a dataset are correlated in some way, and the global understanding of the relationships between explanatory variables and responses might be lost, if the dependence among the last ones is not taken into account.

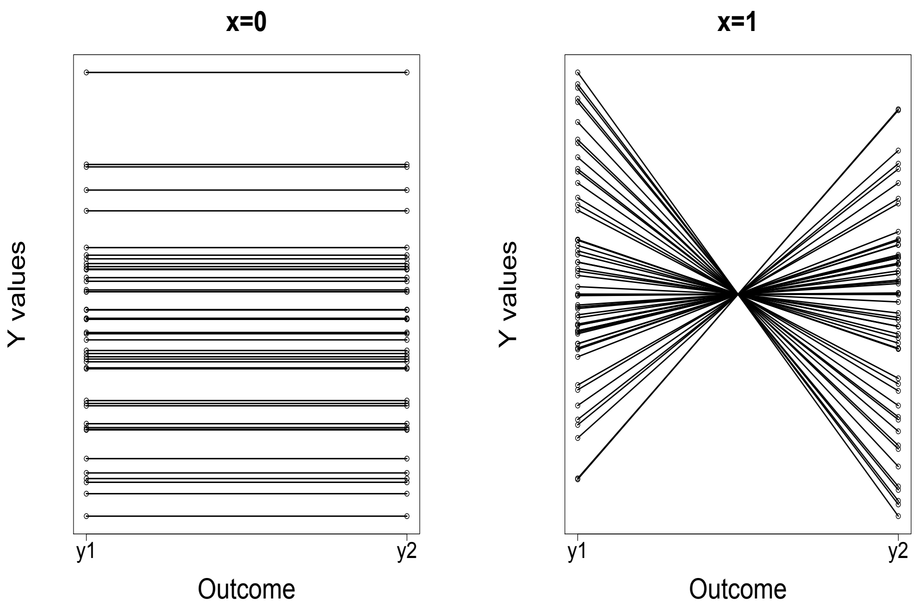

Figure 2.1: Bivariate dependence structure.

To better illustrate and understand the importance of the study of multivariate dependencies in multiple response problems we will consider the following example.

We refer for simplicity only to the bivariate case, however the considerations below can be easily extended to higher dimensions. Let's suppose to study the effect of a dichotomous predictor $X = \{0, 1\}$ on a bivariate dependent variable (Y^1, Y^2), from which we extract the i.i.d. samples $\mathbf{y}^1 = \{y_1^1, \ldots, y_n^1\}$ and $\mathbf{y}^2 = \{y_1^2, \ldots, y_n^2\}$ of size n. We can analyze the relationship between $\mathbf{y}^1$ and $\mathbf{y}^2$ in the two groups determined by $X = 0$ and $X = 1$. The effect of the covariates can induce completely different association structures between the responses. Indeed, let us consider, for example, the situation depicted in figure 4.1. When $X = 0$ there is a situation of maximal positive dependence between $\mathbf{y}^1$ and $\mathbf{y}^2$, the same subject which shows high values for the first response has high values also for the second one. The dependency is inverted when $X = 1$, in this case subjects with high values of $\mathbf{y}^1$ have a low value for $\mathbf{y}^2$ and viceversa. It is clear that all this information would have not been captured without introducing the multivariate framework.

In this chapter we first recall the formalization of multivariate regression models on the mean, which can be done with very little complication considering the properties of multivariate normal distribution. We will then review the concepts of multivariate quantiles and multivariate quantile regression with reference to the various existing methods in the literature.

Note that the extension to the multivariate case is not straightforward when considering quantiles. In fact, due to the absence of a numerical order in dimensions higher or equal than 2, there is not a natural extension of the definition of quantile, and therefore of related regression methods, in multidimensional spaces. As we will see in section 2.2, there exist many different definitions of multivariate quantiles, but, unfortunally, all of them sacrifice some of the properties of univariate quantiles.

2.1 Multivariate linear regression

We will refer to [32] for the description of multivariate regression.

Let us consider a set of m random variable $(Y^1, \ldots, Y^m)$ on which we would like to assess the effect of a set of q predictors $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$. A multivariate regression model, for a sample of size n, can be written in matricial notation as

$$\mathbf{Y}_{(n \times m)} = \mathbf{X}_{(n \times q)} \mathbf{B}_{(q \times m)} + \mathbf{E}_{(n \times m)},$$

where $\mathbf{Y}$ is the matrix stacking together the m observable vectors $\{\mathbf{y}^1, \ldots, \mathbf{y}^m\}$, $\mathbf{X}$ is the usual design matrix, and $\mathbf{E}$ is the matrix with columns the error vectors $\{\boldsymbol{\varepsilon}^1, \ldots, \boldsymbol{\varepsilon}^m\}$.

In analogy with the univariate linear regression we assume a normal distribution for the errors. So the hypotheses of the model are:

1. $E(\mathbf{Y}) = \mathbf{XB}$ or $E(\mathbf{E}) = 0$

2. $cov(\mathbf{y}^i, \mathbf{y}^j) = \mathbf{\Sigma}$ for all $i, j = 1, \ldots, m$

3. $cov(\mathbf{y}_h, \mathbf{y}_k) = 0$ for all $h, k = 1, \ldots, n$ and $h \neq k$

The covariance matrix $\mathbf{\Sigma}$ in assumption 2 contains the variances and covariances $y_h^1, \ldots, y_h^m$ in any unit $\mathbf{y}_h$

$$cov(\mathbf{y}^i, \mathbf{y}^j) = \mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_{2,2} & \cdots & \sigma_{2m} \\ \vdots & & \ddots & \vdots \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma_{mm} \end{pmatrix}$$

In conclusion $\mathbf{Y} \sim N_m(\mathbf{X}\boldsymbol{\beta}, \mathbf{\Sigma})$, or equivalently $\mathbf{E} \sim N_m(0, \mathbf{\Sigma})$.

By analogy to the univariate OLS, the matrix $\mathbf{B}$ of parameters can be estimated as

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

obtained by minimizing the least square error matrix $(\mathbf{Y} - \mathbf{XB})^T(\mathbf{Y} - \mathbf{XB})$.

The estimator fulfills the properties of a good estimator. Additionally to the univariate least squares it has the property that all $\hat{\beta}_h^i$ in $\hat{\mathbf{B}}$ are correlated with each other. The β's within a given column of $\hat{\mathbf{B}}$ are correlated because $\{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$ are correlated. Thus the relationship of the x's to each other affects the relationship of the β's within each column to each other. On the other hand, the β's in each column are correlated with β's in other columns because $\mathbf{y}^1, \ldots, \mathbf{y}^m$ are correlated.

The correlations between the parameters of regression entail the need of adequate multivariate hypotheses tests.

2.2 Multivariate quantiles and multivariate quantile regression

Unlike the mean, it is not possible to naturally extend the definition of quantile to the multivariate case. While the multivariate mean of a set of variables is simply defined as the vector of the means of each variable, the same cannot be done for quantiles. Quantiles are defined by considering the concept of order in the real line $\mathbb{R}$. Unfortunately it is not possible to generalize the natural order for $\mathbb{R}^d$ when $d \geq 2$. To prove it let us consider the definition of total order, as in [42].

Definition 5. *The set C is totally ordered under the relation $\leq$ if and only if for every $a_1, a_2, a_3 \in C$ the following conditions are satisfied*

1. *if $a_1 \leq a_2$ and $a_2 \leq a_1 \Rightarrow a_1 = a_2$ (antisymmetry);*

2. *if $a_1 \leq a_2$ and $a_2 \leq a_3 \Rightarrow a_1 \leq a_3$ (transitivity);*

3. either $a_1 \leq a_2$ or $a_2 \leq a_1$ (dichotomy).

It is common to consider the Euclidean norm to define an order on $\mathbb{R}^2$

$$(y_1^1, y_1^2) \leq (y_2^1, y_2^2) \iff \sqrt{(y_1^1)^2 + (y_1^2)^2} \leq \sqrt{(y_2^1)^2 + (y_2^2)^2}.$$

Unfortunately, the order induced by the Euclidean norm is not a total order. In fact $(1,2) \leq (2,1)$ and $(2,1) \leq (1,2)$ but $(1,2) = (2,1)$. This contradicts the antisymmetry property. It is however possible to specify total orders for vectors in $\mathbb{R}^2$. Consider for example the lexicographical order

$$(a,b) < (c,d) \text{ if and only if } a < c \text{ or } (a = c \wedge b < d);$$

and the antilexicographical order

$$(a,b) < (c,d) \text{ if and only if } b < d \text{ or } (b = d \wedge a < c)$$

However such orders are not satisfactory because they do not extend the intuitive notion of numerical order, for example we are not able to decide if $(5,7) \geq (1,30)$. Hence, this is the reason why the definition of univariate quantile cannot be naturally extended to the multidimensional case in a way that is possible to specify unique quantiles and define orders in statistical terms.

Let us consider again the hypothetical situation described in figure 4.1. In this example, restraining the analysis to the mean would probably be misleading. In fact, the inversion of association caused by the effect of the binary variable seems actually to be more in the tails of the distribution for the two outcomes. Thus it would be helpful to have methods to handle the problem of multivariate regression for quantiles.

In the literature there are many approaches to solving multivariate regression problems for quantiles. Each of them is based on a specific geometrical definition of multidimensional quantile. They are mostly based on vector valued concepts of ranks. These definitions make use of the orientation information in a way that becomes possible to define low and high points in a multidimensional data cloud. To give a picture of the variety and the complexity of the situation we briefly report the definitions for some of the existing approaches. We then evaluate if those definitions satisfy some properties required to extend the concept of quantiles in a multivariate setting. We refer to Serfling (2002) [33].

2.2.1 Geometrical median-oriented quantiles

To better understand the definitions of multivariate quantiles that we will introduce, it is necessary to define a framework of geometry of multivariate data clouds. First of all, we redefine univariate quantiles introducing the concept of center oriented order where a natural linear order is lacking.

For a $\tau \in (0,1)$, let us consider $u = 2\alpha - 1$ a linear one-to-one-way transform. This function maps the unit interval $(0,1)$ into the open interval $(-1,1)$ in such a way that τ's corresponding to extreme quantiles are mapped to values close to $+1$ or -1, whereas those corresponding to central quantiles are mapped to values close to zero. If F is a univariate *cdf* the definition of quantile function with reindexed τ values is (see Serfling (2010) in [34])

$$Q(u,F) = F^{-1}\left(\frac{1+u}{2}\right) \qquad -1 < u < 1.$$

For a given u each point $y \in \mathbb{R}$ can be represented by its quantile $y = Q(u,F)$. For $u = 0$ we have the median $Q(0,F)$. When $u \neq 0$ quantiles are indexed relatively to the median: its sign indicates the direction from the median whereas its magnitude $|u|$ measures the outlyingness from the median. For $|u| = c \in (0,1)$ the values $F^{-1}\left(\frac{1-c}{2}\right)$ and $F^{-1}\left(\frac{1-c}{2}\right)$ represent the boundary points (or equivalently the contour) demarking the upper and lower tails of equal probability weight $\frac{1-c}{2}$. The closed interval between the two values is a τth quantile median-oriented inner or central region.

In the multivariate case, where $F \in \mathbb{R}^d$ with $d \geq 2$, a quantile function can be indexed by elements of the unit open ball $\mathbb{B}^{(d-1)} = \{\boldsymbol{u}|\boldsymbol{u} \in \mathbb{R}^d, ||\boldsymbol{u}|| < 1\}$. Such indexes associate to each point $\mathbf{y} \in \mathbb{R}^d$ a quantile representation $\boldsymbol{Q}(\boldsymbol{u}, F)$ and generate nested contours $\{\boldsymbol{Q}(\boldsymbol{u}, F : ||\boldsymbol{u}|| = c)\}$, $0 \leq c < 1$. For $\boldsymbol{u} = 0$ the most central point is interpreted as the d-median $\boldsymbol{M} = \boldsymbol{Q}(0, F)$. In analogy to the univariate case, for $\boldsymbol{u} \neq 0$ the index $\boldsymbol{u}$ represents a direction from the median, and its magnitude $||\boldsymbol{u}||$ is an outlyingness parameter with higher values corresponding to more extreme points. The contours for $||\boldsymbol{u}|| = c$ thus represent equivalent classes of points of equal oulyingness, these contours can be used to define multivariate quantiles.

2.2.2 Multivariate quantile functions based on depth functions

Statistical *depth functions* are used in multivariate analyses to provide orderings of points in $\mathbb{R}^d$. They give a measure of the centrality of a point in a cloud of data. We will define these functions restraining only to nonnegative bounded depth as in [43]. A depth function with good ordering data properties should satisfy the four requirements below.

1. The depth of a point $\mathbf{y} \in \mathbb{R}^d$ should not depend on the underlying coordinate system or, in particular, on the scales of the underlying measurements. This property is called affine equivariance.

2. For a distribution having a uniquely defined center (or, more generally, a point of symmetry with respect to a symmetry measure), the depth function should attain its maximal value at this center.

3. As a point $\mathbf{y} \in \mathbb{R}^d$ moves away from the deepest point (the point at which the depth function attains its maximal value; for a symmetric distribution this point coincides with the center) along any fixed trajectory through the center, the depth at $\mathbf{y}$ should decrease monotonically. That is, we require monotonicity relative to the deepest point.

4. The depth of a point $\mathbf{y}$ should approach zero as $\|\mathbf{y}\|$ approaches infinity.

With these properties we can give a formal definition of a depth function. In the following definition we will denote by $\mathfrak{F}$ the class of *cdf* in $\mathbb{R}^d$.

Definition 6. *Let us consider the mapping* $D(\cdot;\cdot) : \mathbb{R}^d \times \mathfrak{F} \to \mathbb{R}$ *nonnegative, bounded and satisfying the above properties. That is, we assume:*

(i) $D(A\mathbf{y} + b; F_{AY+b}) = D(\mathbf{y}; F_Y)$ *holds for any random vector* $Y \in \mathbb{R}^d$, *any* $d \times d$ *nonsingular matrix* A, *and any* d*-vector* b;

(ii) $D(\theta; F) = \sup_{\mathbf{y}\in\mathbb{R}^d} D(\mathbf{y}; F)$ *holds for any* $F \in \mathfrak{F}$ *having center in* θ;

(iii) for any $F \in \mathfrak{F}$ *having deepest point* θ, $D(\mathbf{y}; F) \leq D(\theta + \tau(\mathbf{y} - \theta); F)$ *holds for any* $\tau \in [0, 1]$;

(iv) $D(\mathbf{y}; F) \to 0$ *as* $\|\mathbf{y}\| \to \infty$, *for each* $F \in \mathfrak{F}$. *The* $D(\cdot; F)$ *is called a statistical depth function.*

The point of maximal depth if unique, otherwise the average of such points, is used as a notion of multivariate median.

The most commonly used depth function is the half space depth (also known as Tukey, named after Tukey depth who introduced it in 1925), which is defined as the minimal probability attached to any closed halfspace containing $\mathbf{y}$

$$HD(\mathbf{y}, F) = \inf\{F(H)|H \text{ is a closed halfspace}, \mathbf{y} \in H\}.$$

In particular, the sample halfspace depth of $\mathbf{y}$ is the minimal number of the data points lying in a closed halfspace containing $\mathbf{y}$. This depth function satisfies all of the properties stated above (see [43]) and has ellipsoidal contours, as can be seen in the right panel of figure 2.2.2.

For a depth function $D(\mathbf{y}, F)$, let us consider its associated nested contours which enclose the median M and bound inner regions of the form $\{\mathbf{y} : D(\mathbf{y}, F) \geq \alpha\}$, for $\alpha > 0$.

Definition 7. *The depth contour induces the quantile function* $\boldsymbol{Q}(\boldsymbol{u}, F)$, *for* $\boldsymbol{u} \in \mathbb{B}^{(d)}$, *with each* $\mathbf{y} \in \mathbb{R}^d$ *as follow. For* $\mathbf{y} = \boldsymbol{M}$, *denote it by* $\boldsymbol{Q}(0, F)$. *For* $\mathbf{y} \neq \boldsymbol{M}$ *denote it by* $\boldsymbol{Q}(\boldsymbol{u}, F)$, *with* $\boldsymbol{u} = p\boldsymbol{v}$, *where* p *is the probability weight of the central region with* $\mathbf{y}$ *on its boundary and* $\boldsymbol{v}$ *is the unit vector from* $\mathbf{M}$ *toward* $\mathbf{y}$.

Different versions of quantile functions are associated to different depth functions.

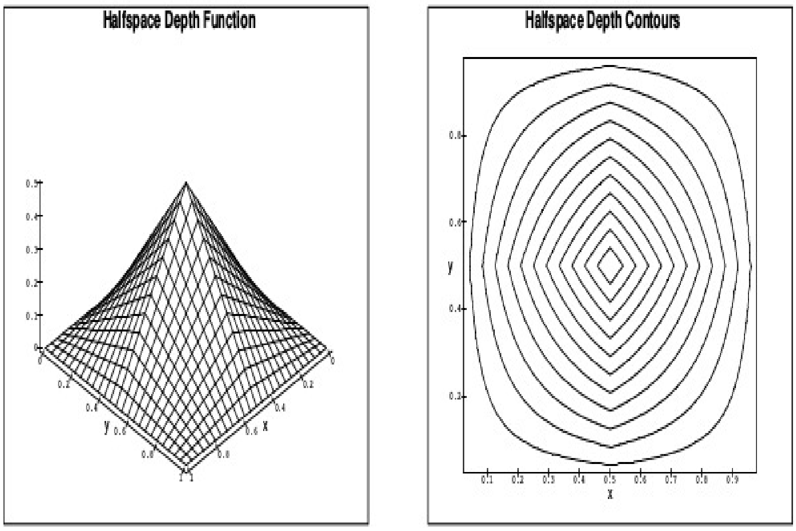

Figure 2.2: Halfspace Depth, F Uniform on Unit Square

2.2.3 Multivariate quantile based on norm minimization: the geometric quantile

For a univariate random variable Y, we have already shown as a quantile arises from a minimization problem of the form

$$\min_{\xi \in \mathbb{R}} \mathbb{E}[(Y-\xi)(\tau - \mathbb{I}(Y \leq \xi)]$$

this can be also rewritten in a useful way as

$$\min_{\xi \in \mathbb{R}} \mathbb{E}[|Y-\xi| + (2\tau - 1)(Y-\xi)]. \tag{2.1}$$

Generalizations of this last formulation of the minimization problem have been considered to define multidimensional notions of quantiles. Among the many we recall one of the most succesful definitions in these terms, given by Chauduri in 1996 [6]. For the extension to $\mathbb{R}^d$ let us rewrite the objective function in (2.1) as

$$\mathbb{E}[|Y-\xi| + u(Y-\xi)]$$

where $u = 2\tau - 1$ is defined in the open interval $(-1, 1)$. The extension to the multivariate case can be done by considering the multivariate geometrical framework introduced in 2.2.1. Therefore, let us consider the index $\mathbf{u} \in \mathbb{B}^{(d)}$, where $\mathbb{B}^{(d)}$ indicates the open unit ball.

Definition 8. *Letting $Y \in \mathbb{R}^d$ be a random multidimensional variable with $d \geq 2$, the $\mathbf{u}$-th geometric quantile of Y is*

$$\tilde{Q}(\mathbf{u}) = \min_{\xi \in \mathbb{R}^d} E[\Phi(\mathbf{u}, X - \xi) - \Phi(\mathbf{u}, X)]$$

where $\Phi(\mathbf{u}, \mathbf{t}) = \|\mathbf{t}\| + <\mathbf{u}, \mathbf{t}>$, with $\|\cdot\|$ the Euclidian norm and $<\cdot,\cdot>$ the inner product.

2.2.4 Multivariate quantiles as inversions of mappings

Univariate quantiles are usually defined by considering the inversion function of the cdf $F^{-1}(\tau)$ for $\tau \in (0, 1)$. Since it is not possible to consider the inverse of a multivariate cdf, we could then think of substituting the *cdf* with an alternative invertible function G_F, able to characterize the probability distribution. Dudley and Koltchinskii, in an unpublished manuscript of 1992 [11], defined such a function as a one-to-one map from $\mathbb{R}^d$ into the unit open ball $\mathbb{B}^{(d)}$ such that

$$G_F(\mathbf{t}) = -\mathbb{E}\left(\frac{Y - \mathbf{t}}{\|Y - \mathbf{t}\|}\right) \quad \mathbf{t} \in \mathbb{R}^d. \tag{2.2}$$

Multivariate quantiles can be then defined as

$$\tilde{Q}(\mathbf{u}) = G_F^{-1}(\mathbf{u})$$

with $\mathbf{u} \in \mathbb{B}^{d-1}(0)$. For the complete theory, see Koltchinskii (1997) [26].

The geometric quantile introduced in Definition 8, can also be viewed as the inversion of the map 2.2 if the random variable Y is absolutely continuous in $\mathbb{R}^d$.

2.2.5 Comparison of multivariate quantile definitions and their application to multiple response regression

A quantile function $Q(\cdot, \cdot)$, either in the univariate or in the multivariate case, should always have some properties of interpretability. These properties can be summarized in three points.

- The probabilistic interpretability of empirical and distribution quantiles. For each fixed $\tau \in [0, 1)$ the set $\{Q(u, t) | 0 \leq t \leq \tau, \text{all u}\}$ should include a τth quantile inner region with boundary points $Q(u, \tau)$ and with median $Q(u, 0) = M$.

- Directional monotonicity to permit the description of location. For each fixed direction u from M, the distance $\|Q(u, \tau) - M\|$ should increase with τ for $\tau \in [0, 1]$.

- Suitable set theoretic interpretation. The τth quantile inner regions $\{Q(u,t)|0 \leq t \leq \tau, \text{all u}\}$ are required to be affine equivariant, nested, compact and connected.

The three definitions of multivariate quantiles discussed are quite satisfactory in terms of quantile interpretability and satisfy most of the requirements. However, each of them presents some limitation.

The depth approach verifies all the properties of interpretability; unfortunately, the implementation of efficient algorithms to their calculation is really challenging. It involves the calculation of the infimum over an infinite number of direction vectors. Furthermore, even if possible, the extension to regression settings is not straightforward. The halfspace depth has been considered for this extension. Depth regression was first introduced by Struyf and Rousseeuw in 1999 [39]. In a paper from 2012 [27], Kong e Mizera developed a slightly modified quantile regression, based on projectional quantiles to estimate depth contours.

Liu and Zuo (2014) [28] proposed two exact algorithms for the exact calculation of halfspace depth and regression depth.

The definition based on norm minimization does not have the property of affine equivariance. Multidimensional quantiles defined in this way are equivariant under orthogonal transformations or rotations of the response vectors, but they are not under arbitrary nonsingular transformations, including also coordinatewise scale transformations. The extension to the regression is quite natural and was also discussed in Chauduri's paper [6].

In regression problems the lack of equivariance for multidimensional quantiles, defined as minimizers of a norm, was overcome by Chakraborty [5] who considered a transformation retransformation procedure based on a data driven coordinate system. The method first linearly transforms the data to a new invariant coordinate system, in which the regression parameters are estimated on the transformed coordinates. Then the estimates are retransformed to the original coordinate system.

In a paper from 2010 Hallin et al. [17] give a definition of multivariate quantiles by making the connection between two of the enhanced classical definitions above. They defined multivariate quantiles as a norm minimization problem in a way that makes possible to define a strict interconnection with depth contours. This method also leads to a concept of multiple-output regression halfspace depth.

Chapter 3

Association and dependence structures

In this chapter we will review the existing tools to detect and measure associations between variables. In what follows we will mainly refer to Embrechts et al. (1998) [13] and Trivedi and Zimmer (2005) [40].

Given a set of random variables there are many different ways to measure the relationships between them. These are mostly described by considering measures of dependence or association such as concordance, correlation, tail dependence. All these indexes evaluate if such measured variables vary according to some pattern, and summarize the information in a single number.

The term correlation indicates a measure of linear dependence between random variables and is often erroneously considered as synonymous to the more general association.

3.1 Desired properties of dependence measures

In probabilistic terms two random variables (Y^1, Y^2) are said to be *independent* if their joint distribution function is the same as the product of the marginal *cdf*s, in other terms $F_{Y^1,Y^2}(y^1, y^2) = F_{Y^1}(y^1)F_{Y^2}(y^2)$.

Let $\delta(Y^1, Y^2)$ be a scalar measure of dependence, according to [13] and [40], a good measure should satisfy the following properties:

1. symmetry: $\delta(Y^1, Y^2) = \delta(Y^2, Y^1)$;
2. normalization: $-1 \leq \delta \leq 1$;
3. comonotonicity: $\delta(Y^1, Y^2) = 1 \iff (Y^1, Y^2)$ comonotonic;
4. countermonotonicity: $\delta(Y^1, Y^2) = -1 \iff (Y^1, Y^2)$ countermonotonic;

5. for a strictly monotonic transformation $T : \mathbb{R} \to \mathbb{R}$ of Y^1

$$\delta(T(Y^1), Y^2) = \begin{cases} \delta(Y^1, Y^2) & \text{if T is increasing} \\ -\delta(Y^1, Y^2) & \text{if T is decreasing;} \end{cases}$$

another property which is often required to be satistied is

6. $\delta(Y^1, Y^2) = 0 \iff Y^1, Y^2$ are independent.

The terms comonotonicity and countermonotonicity refer to the perfect positive and negative dependence between two random variables. The concept of comonotonicity is easily extendable to dimensions higher than 2 while this is not true for countermonotonicity. A formal definition can be given.

Definition 9. *Two components* (y_j^1, y_j^2) *and* (y_k^1, y_k^2) *of a pair of random vectors* (Y^1, Y^2) *are said to be comonotonic if* $\{y_j^1 \leq y_j^2, y_k^1 \leq y_k^2\}$ *or* $\{y_j^1 \geq y_j^2, y_k^1 \geq y_k^2\}$*. They are called countermonotonic if* $\{y_j^1 \leq y_j^2, y_k^1 \geq y_k^2\}$ *or* $\{y_j^1 \geq y_j^2, y_k^1 \leq y_k^2\}$*.*

The list of properties can be extended in different ways. Unfortunately, as it is proved in the following proposition, there does not exist a dependency measure satisfying all the requirements above.

Proposition 1. *(Embrechts 2002 [13]) There is no dependency measure satisfyng* 5 *and* 6 *simultaneously.*

Some of the properties (2, 3, 4 and 5 to be precise) could be modified to get measures that satisfy also property 6. The alternative requirements are

2b. $0 \leq \delta \leq 1$

3-4b. $\delta(Y^1, Y^2) = 1 \iff Y^1, Y^2$ comonotonic or countermonotonic

5b. for $T : \mathbb{R} \to \mathbb{R}$ strictly monotonic $\delta(T(Y^1), Y^2) = \delta(Y^1, Y^2)$.

3.2 Linear correlation

The correlation coefficient measures the linear dependence between two random variables and is one of the most common dependency indexes. It can be extended easily to the multidimensional case.

Definition 10. *Given two random variables* Y^1 *and* Y^2 *we define their linear correlation as*

$$\rho(Y^1, Y^2) = \frac{cov[Y^1, Y^2]}{\sigma_{Y^1} \sigma_{Y^2}}$$

where $cov[Y^1, Y^2] = E[Y^1 Y^2] - E[Y^1]E[Y^2]$*, and* $(\sigma_{Y^1}, \sigma_{Y^2}) > 0$ *denoting the standard deviations of* Y^1 *and* Y^2*, respectively.*

When $m \geq 3$ covariance and correlation are symmetric positive matrices with elements all the pairwise covariances and correlations.

The correlation has many of the properties required from a dependency measure. It is a symmetric measure. It is normalized, that is $-1 \leq \rho(Y^1, Y^2) \leq 1$ where the lower and the upper bounds correspond to perfect negative and positive dependence. Though, there is a limitation when considering the range of possible values for the correlation coefficient since the actual range of attainable values for each pair of variables is defined from the marginal distributions. It is invariant for strictly increasing linear transformations of the variables, but not under strictly increasing nonlinear transformations. That is $\rho(T(Y^1), T(Y^2)) \neq \rho(Y^1, Y^2)$ for $T : \mathbb{R} \to \mathbb{R}$. In case of independent random variables the correlation is null, $\rho(Y^1, Y^2) = 0$, since $cov(Y^1, Y^2) = 0$. However, the contrary is valid only if the pair (Y^1, Y^2) follows a bivariate normal distribution.

Linear correlation is in general easy to calculate, unfortunately, it cannot be defined if the variances of Y^1 and Y^2 are not finite.

Thanks to its straightforward interpretation and calculation, linear correlation is a reasonably good and popular dependence measure; however, the drawbacks mentioned above must be taken into consideration and thus motivate the introduction of alternative measures.

3.3 Concordance

The concordance measures the agreement between two random variables. In fact, a pair of random variables is said *concordant* if large values of one correspond to large values of the other and equivalently for small values. On the contrary if small values are associated to high values of the other variable the pair is said *discordant*.

Definition 11. *Given (Y^1, Y^2) a pair of continuous random variables from which we select two observations (y_i^1, y_i^2) and (y_j^1, y_j^2) we say that*

- *(y_i^1, y_i^2) and (y_j^1, y_j^2) are concordant if for $y_i^1 < y_j^2$ we have $y_i^2 < y_j^2$, or if when $y_i^1 > y_j^1$ then $y_i^2 > y_j^2$, this can be said equivalently as $(y_i^1 - y_j^1)(y_i^2 - y_j^2) > 0$*
- *(y_i^1, y_i^2) and (y_j^1, y_j^2) are discordant if $y_i^1 < y_j^2$ and $y_i^2 > y_j^2$, or if $y_i^1 > y_j^1$ and $y_i^2 < y_j^2$, we can alternatively write it as $(y_i^1 - y_j^1)(y_i^2 - y_j^2) < 0$.*

3.3.1 Spearman's rho and Kendall's tau coefficients

Spearmann's rho is a dependence coefficient built on the concept of concordance.

Definition 12. *Let (Y_1^1, Y_1^2), (Y_2^1, Y_2^2), and (Y_3^1, Y_3^2) be three independent random vectors. The Spearman rho coefficient is defined as*

$$\rho_S(Y^1, Y^2) = 3(P[(Y_1^1 - Y_2^2)(Y_1^2 - Y_3^2) > 0] - P[(Y_1^1 - Y_2^1)(Y_1^2 - Y_3^2) < 0])$$

$$\rho_S(Y^1, Y^2) = 3(P[concordance] - P[discordance])$$

where the pairs have a common distribution function $H_{Y^1Y^2}(y^1, y^2)$, and marginal distributions $F_{Y^1}(y^1)$, $G_{Y^2}(y^2)$.

The Spearman rho coefficient is thus expressed in terms of concordance and discordance since it is proportional to the probability of concordance minus the probability of discordance for the two pairs (Y_1^1, Y_1^2) and (Y_2^1, Y_3^2). Note that while the joint distribution function of (Y_1^1, Y_1^2) is $H_{Y^1Y^2}(y^1, y^2)$, the joint distribution function of (Y_2^1, Y_3^2) is $F_{Y^1}(y^1)G_{Y^2}(y^2)$ (because Y_2^1 and Y_3^2 are independent).

Spearman's rho can be also thought of as a rank correlation measure. In fact the values of variables compared are considered by evaluating their order or better their ranked values; concordance associates high ranked values between them and low ranked values for each pair. The following definition recalls more intuitively the association based on ranks.

Definition 13. *Spearman's rank correlation is the linear correlation between $F_1(Y^1)$ and $F_2(Y^2)$*

$$\rho_S(Y^1, Y^2) = \rho(F_1(Y^1), F_2(Y^2)) = \frac{cov(F_1(Y^1), F_2(Y^2))}{\sqrt{var(F_1(Y^1))var(F_2(Y^2))}} \tag{3.1}$$

where $F_1(Y^1) = U_1$ and $F_2(Y^2) = U_2$ are integral transforms of Y^1 and Y^2.

By integral transform we refer to the standard result of probability in the following proposition.

Proposition 2. *(Probability integral transform) Let X be a random variable with distribution function F. If F is continuous then the random variable $F(X)$ is standard-uniformly distributed, that is $F(X) \sim U(0, 1)$.*

Hence the Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables.

Another common measure of concordance is the Kendall tau coefficient defined as the relative difference between probability of concordance and discordance.

Definition 14. *Let (Y_1^1, Y_1^2) and (Y_2^1, Y_2^2) two independent random vectors, Kendall's tau coefficient is defined as*

$$\rho_\tau(Y^1, Y^2) = P[(Y_1^1 - Y_2^1)(Y_1^2 - Y_2^2) > 0] - P[(Y_1^1 - Y_2^1)(Y_1^2 - Y_2^2) < 0] \tag{3.2}$$

$$\rho_\tau(Y^1, Y^2) = P[concordance] - P[discordance].$$

As the Spearman rho coefficient, Kendall's tau measures also the rank correlation between two random variables.

Both ρ_S and ρ_τ satisfy the following properties of an independence coefficient. They are symmetric and normalized with $-1 \leq \rho_S, \rho_\tau \leq 1$. They have the property of comonotonicity and countermonotonocity with perfect positive dependence attained at 1 and perfect negative dependence reached at -1. Furthermore, if Y^1 and Y^2 are independent then $\rho_S(Y^1, Y^2) = \rho_\tau(Y^1, Y^2) = 0$.

In addition these rank correlation measures satisfy the property of being invariant under monotonic linear and nonlinear transformations and can capture perfect dependence. However, they are not always easy to calculate.

Given a pair of random variables the values corresponding to Kendall's tau and Spearman's rho can be quite different even if they both measure the probability of concordance. There are some inequalities that can explain the relationship between the two measures. We will report them; their proofs can be found in [30].

Theorem 7. *(Daniels 1950 [8]) Let Y^1 and Y^2 be continuous random variables. Then*

$$1 \leq 3\rho_\tau - 2\rho_S \leq 1.$$

Theorem 8. *Let Y^1, Y^2, ρ_τ, and ρ_S be as in Theorem 6. Then*

$$\frac{1+\rho_S}{2} \geq \left(\frac{1+\rho_\tau}{2}\right)^2 \tag{3.3}$$

and

$$\frac{1-\rho_S}{2} \geq \left(\frac{1-\rho_\tau}{2}\right)^2 \tag{3.4}$$

The inequalities in the preceding two theorems can be combined to get the final inequality in the corollary.

Corollary 2. *Let Y^1, Y^2, ρ_τ, and ρ_S be as in Theorem 6. Then*

$$\frac{3\rho_\tau - 1}{2} \leq \rho_S \leq \frac{1+\rho_\tau-\rho_\tau^2}{2}, \quad \rho_\tau \geq 0$$

and

$$\frac{\rho_\tau^2 + 2\rho_\tau - 1}{2} \leq \rho_S \leq \frac{1+3\rho_\tau}{2}, \quad \rho_\tau \leq 0.$$

The two rank correlations (3.1) and (3.2) above are valid for continuous variables and do not directly apply to discrete data. In these cases the coefficients must be adjusted to handle the discreteness. When considering discrete data, for example, there is the possibility of having ties, that is $P[\text{tie}] = P[Y_1^1 = Y_2^1 \text{ or } Y_1^2 = Y_2^2]$. A modified version of Kendall's tau allows one to take into consideration the probabilty of ties.

Definition 15. *Let (Y_1^1, Y_1^2) and (Y_2^1, Y_2^2) be two two pairs of nonnegative discrete random variables. Then Kendall's tau coefficient is defined as*

$$\begin{aligned}\rho_\tau(Y^1, Y^2) &= 4P[Y_2^1 < Y_1^1, Y_2^2 < Y_1^2] - 1 + P[Y_1^1 = Y_2^1 \ or \ Y_1^2 = Y_2^2] \\ &= 2P[concordance] - 1 + P[tie]\end{aligned}$$

This definition can be derived by considering that

$$P[\text{concordance}] - P[\text{discordance}] + P[\text{tie}] = 1.$$

With this new definition we can see how in the discrete case the Kendall's tau coefficient is dependent upon the margins of the compared variables. In addition in case of small sample size, it is likely to have a high proportion of ties and this reduces the range of feasible values for $\rho_\tau(Y^1, Y^2)$. There is a loss in efficiency going from the continuous to the discrete.

3.3.2 Positively quadrant dependent

There are situations in which it is not necessary to calculate the strength of association between two variables but is enough to know if they are concordant (positive associated) or discordant (negative associated). We recall two definitions of association concepts based on this philosophy.

Definition 16. *Two random variables Y^1 and Y^2 are positively quadrant dependent (PQD) if*

$$P(Y^1 > y^1, Y^2 > y^2) \geq P(Y^1 > y^1)P(Y^2 > y^2) \ for \ all \ x, y \in \mathbb{R}$$

or equivalently

$$P(Y^1 \leq y^1, Y^2 \leq y^2) \geq P(Y^1 \leq y^1)P(Y^2 \leq y^2) \ for \ all \ x, y \in \mathbb{R}$$

given that $P(Y^1 > y^1, Y^2 > y^2) = 1 - P(Y^1 \leq y^1) + P(Y^2 \leq y^2) - P(Y^1 \leq y^1, Y^2 \leq y^2)$. Analogously by reversing the sense of inequalities we can define the negative quadrant dependence (NQD).

Intuitively, Y^1 and Y^2 are PQD if the probability that they are simultaneously small (or simultaneously large) is at least as large as it would be if they were independent.

Definition 17. *Two random variables Y^1 and Y^2 are positively associated (PA) if*

$$E[g_1(Y^1)g_2(Y^2)] \geq E[g_1(Y^1)]E[g_2(Y^2)]$$

for all real-valued, measurable functions g_1 and g_2, which are increasing in both components and for which the expectations above are defined.

We remark that PQD and PA are invariant under increasing transformations. It can be proved (see [13]) that

$$\text{comonotonicity} \Rightarrow PA \Rightarrow PQD \Rightarrow \rho(Y^1, Y^2) \geq 0, \rho_S(Y^1, Y^2) \geq 0, \rho_\tau(Y^1, Y^2) \geq 0$$

and so comonotonicity is the strongest type of concordance of positive dependence.

3.4 Tail dependence

There are cases in which one is interested in detecting the association or concordance between extreme values or tails of two random variables. This dependence describes essentially the limiting proportion that one margin exceeds a certain threshold given that another margin has already exceeded that threshold.

Definition 18. *Let Y^1 and Y^2 be two random variables with distribution functions F_1 and F_2. The coefficient of upper tail dependence of Y^1 and Y^2 is*

$$\lim_{\alpha \to 1^-} P[Y^2 > F_2^{-1}(\alpha) | Y^1 > F_1^{-1}(\alpha)] = \lambda_u$$

while the coefficient of lower tail dependence is

$$\lim_{\alpha \to 0^+} P[Y^2 \leq F_2^{-1}(\alpha) | Y^1 \leq F_1^{-1}(\alpha)] = \lambda_l$$

provided that a limit $\lambda_u \in [0, 1]$ and a limit $\lambda_l \in [0, 1]$ exist. If $\lambda_u \in (0, 1]$, Y^1 and Y^2 are said to be asymptotically dependent in the upper tail; if $\lambda_u \in [0, 1)$, Y^1 and Y^2 are said to be asymptotically dependent in the lower tail; to conclude if $\lambda_u = 0$ they are said upper tail independent and if $\lambda_l = 0$ they are said lower tail independent.

3.5 Copula functions

When considering multivariate structures and dependencies between random variables copula functions are among one the most exhaustive statistical tools.

Let recall that, given a set of random variables $X_1, \ldots, X_n$, the dependency between them is completely described by their joint *cdf*

$$F_X(x_1, \ldots, x_n) = P[X_1 \leq x_1, \ldots, X_n \leq x_n].$$

The concept of copula arises from the idea of separating the multivariate *cdf* F into the marginal *cdf*s of the single variables plus the description of the dependence structure.

Let us consider again the random variables $X_1, \ldots, X_n$ with continuous marginal distributions $F_i(x) = P(X_i \leq x)$. Applying the probability integral transform in proposition 2 to each component of the vector, we obtain

$$(U_1, \ldots, U_n) = (F_1(X_1), \ldots, F_n(X_n))$$

where each marginal has uniform distribution. The joint distribution of the n random variables is called *copula* of $X_1, \ldots, X_n$. This can be written as

$$\begin{aligned} C(u_1, \ldots, u_n) &= C(F_1(x_1), \ldots, F_n(x_n)) = P[F_1(X_1) \leq F_1(x_1), \ldots, F_n(X_n) \leq F_n(x_n)] \\ &= F(x_1, \ldots, x_n). \end{aligned}$$

Definition 19. *A copula is the distribution function of a random variable in $\mathbb{R}^n$ with uniform-$(0,1)$ marginals. Alternatively a copula is any function $C : [0,1]^n \to [0,1]$ with the following properties:*

1. *$C(x_1, \ldots, x_n)$ is increasing in each component x_i;*
2. *$C(1, \ldots, 1, x_i, 1, \ldots, 1) = x_i$ for all $i \in \{1, \ldots, n\}$, $x_i \in [0,1]$;*
3. *for all $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in [0,1]^n$ with $a_i \leq b_i$ we have*

$$\sum_{i_1=1}^{2} \sum_{i_n=1}^{2} (-1)^{i_1+\ldots+i_n} C(x_{1i_1}, \ldots, x_{ni_n}) \geq 0 \tag{3.5}$$

where $x_{j1} = a_j$ and $x_{j2} = b_j$ for all $j \in \{1, \ldots, n\}$.

The two alternative definitions are equivalent. In fact, given that we consider the multivariate distribution having standard uniform marginals, the three properties can be easily derived. Property 1 is clear from the definition of distribution function; property 2 follows from the fact that the marginals are uniform-$(0,1)$; property 3 is true because the sum (3.5) can be interpreted as $P[a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n]$, which is non-negative.

One of the most famous results relative to the theory of copulas is the following due to Sklar. It dates back to 1959 [35].

Theorem 9. *(Sklar 1959 [35]) Let $F(x_1, \ldots, x_n)$ be a joint cumulative distribution function with marginals $F_i(x_i)$. Then there exists a copula C such that, for all real values $(x_1, \ldots, x_n)$*

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

If the marginals $F_i(x_i)$ are all continuous the copula is unique; otherwise it is uniquely determined on $range(F_1) \times \ldots \times range(F_n)$ which is the cartesian product of the ranges of the marginals cdf's. Conversely, if C is a copula and $F_i(x_i)$ are univariate cdf's then $F(x_1, \ldots, x_n)$ is a joint cdf with margins $F_i(x_i)$.

It is then sufficient to know marginal distributions to completely reconstruct the joint one. Any copula can be constructed by considering a set of marginal distributions and a parameter θ, called dependence parameter, which measures the structure of association between the marginals. In this sense a copula of $X_1, \ldots, X_n$ can be written also as $C(F_1(x_1), \ldots, F_n(x_n); \theta)$.

The following theorem, which defines bounds of a copula, is an important result on copula theory and is useful to define dependencies properties of such functions.

Theorem 10. *Let C be a copula and Y^1 and Y^2 two random variables. For every observed value y^1 and y^2*

$$\max\{y^1 + y^2 + 1 - 2, 0\} \leq C(y^1, y^2) \leq \min\{y^1, y^2\}$$

with

$$\max\{y^1 + y^2 + 1 - 2, 0\} = C_\ell(y^1, y^2) \text{ and } \min\{y^1, y^2\} = C_u(y^1, y^2).$$

These bounds are known as Fréchet-Hoeffding bounds.

Since copulas are seen as dependence functions it is interesting to show that they share some of the features of dependencies coefficients. We state them only for the bivariate case.

A copula function satisfy the following properties:

1 (Y^1, Y^2) are independent $\iff C(y^1, y^2) = F_1(y^1)F_2(y^2)$;

2 (Y^1, Y^2) are comonotonic $\iff C(y^1, y^2) = C_u(y^1, y^2)$;

3 (Y^1, Y^2) are countermonotonic $\iff C(y^1, y^2) = C_\ell(y^1, y^2)$;

4 if (T_1, T_2) are increasing continuous functions of $(Y^1, Y^2) \Rightarrow C(y^1, y^2) = C(T_1(y^1), T_2(y^2))$ (invariance).

Most of the dependence functions described in the previous paragraph can be directly related to copulas, and this can be used to describe the association structure carried by them in an easier way.

Definition 20. *Let (Y^1, Y^2) be a pair of continuous random variables with copula C, ρ_S and ρ_τ Spearman's rho and Kendall's tau coefficients, respectively. Then*

$$\rho_S(Y^1, Y^2) = 12 \int_0^1 \int_0^1 [C(u, v) - uv] du dv$$

and

$$\rho_\tau(Y^1, Y^2) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

where $u = F_1^{-1}(y^1)$ and $v = F_2^{-1}(y^2)$.

Definition 21. *Let (Y^1, Y^2) be a pair of continuous random variables with copula C. Then the coefficients of upper and lower tail dependence are given by the following expressions*

$$\lambda_u = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$

and

$$\lambda_\ell = \lim_{u \to 0^+} \frac{C(u, u)}{u}$$

where $u = F_1^{-1}(y^1)$.

Given a set of completely known marginal distributions it is an important task to choose the appropriate copula function to piece them together. There are in literature a large number of families of copulas, each of them imposes a different dependence structure. At the occurence it is important to select the one which better describes the data.

Part II

Modeling of dependence of quantile regression residuals' signs for bivariate responses

Chapter 4

Concordance of quantile regression residuals signs' as dependence measure

We propose a new method which allows us to model multivariate structures of quantiles in multiple response problems without relying on any definition of multivariate quantile. The method is non parametric, based on the joint analysis of residuals' signs of univariate quantile regression models. We define an index of dependence similar to the coefficients introduced in chapter 3. Such index measures the degree of concordance of residuals' signs. We then model the probability of concordance of the residuals' signs as a function of a set of covariates.

We will discuss the case of bivariate response variables. In the case of higher dimension the method could be generalized by taking all the possible combinations of pairs.

In this chapter we will define a coefficient of concordant residuals' signs and show its properties.

Let us consider n observations from a bivariate random variable (Y^1, Y^2) and a set of q covariates $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_q\}$. Let us indicate with $\mathbf{y}^i = \{y_1^i, \ldots, y_n^i\}$ the vectors of observations from (Y^1, Y^2), where $i = 1, 2,$.

Chosen a $\tau \in (0, 1)$, for each of the two outcomes we build a univariate τ-th quantile regression model as described in section 1.2

$$y_j^i = \mathbf{x}_j^T \boldsymbol{\beta}_\tau^i + \boldsymbol{\varepsilon}^i \quad i = 1, 2 \quad j = 1, \ldots, n$$

and estimate its unknown regression parameters $\boldsymbol{\beta}_\tau^i = \{\beta_{\tau,0}^i, \beta_{\tau,1}^i, \ldots, \beta_{\tau,q-1}^i\}$.

For any subject j we observe the estimated residuals $\hat{\varepsilon}_j^i = y_j^i - \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_\tau^i$. The concordance in their signs gives informations on the degree of dependence between them. A discordance in the residuals signs indicates a situation of negative dependence, while concordance can be interpreted as positive dependence. This is well depicted in figure 4 where the dashed red lines represent two different situations of positive and negative concordance, whereas the blue solid line is an example of discordance.

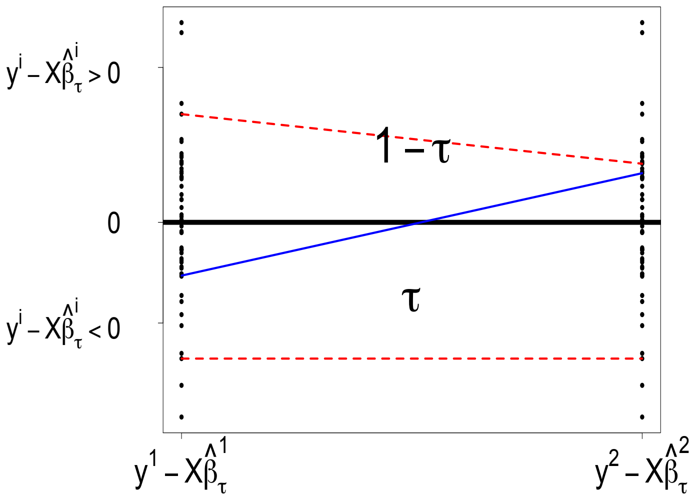

Figure 4.1: Bivariate dependence of residuals of quantile regression. The red dashed lines correspond to concordant residuals represent positive dependence. The blue solid line represent negative dependence, that is discordant residuals.

We then generate an indicator variable, called Z, whose values can be stacked in an n-dimensional vector $\mathbf{z} = (z_1, \ldots, z_n)$ with components

$$z_j = \begin{cases} 1 & \text{if } (y_j^1 \leq \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_\tau^1 \wedge y_j^2 \leq \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_\tau^2) \vee (y_j^1 > \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_\tau^1 \wedge y_j^2 > \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_\tau^2) \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

which describes the concordance of residuals for $j = 1, \ldots, n$.

We use the information contained in the vector $\mathbf{z}$ to construct a statistic which represents the proportion of observations with concordant quantile regression residuals' signs

$$\hat{\sigma} = \frac{1}{n} \sum_{j=1}^{n} z_j. \tag{4.2}$$

The $\hat{\sigma}$ statistic can be viewed as a coefficient of concordance satisfying many of the properties of a good measure of dependence introduced in chapter 3.

In order to measure the strength of dependence we need to calculate the values of $\hat{\sigma}$ corresponding to the three extreme hypotheses of association: perfect positive dependence, perfect negative dependence and independence.

To do so, we consider the marginal distribution of quantile regression residuals. From Theorem 3 in section 1.3 we know that there is approximately a proportion of τ negative and $1 - \tau$ positive quantile regression residuals.

To calculate limiting values for the $\hat{\sigma}$ statistic, we looked at the marginal quantile distribution of residuals by constructing 2×2 contingency tables that contain the proportion of subjects with positive or negative residuals in the two regression models.

The general table is of the form from which the $\hat{\sigma}$ statistic can be obtained

		$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$		
		+	-	
$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$	+	a	b	$1-\tau$
	-	c	d	τ
		$1-\tau$	τ	

Table 4.1: General contingency table of the distribution of quantile residuals regressions.

summing the proportion of subjects with concordant residuals along the diagonal as

$$\hat{\sigma} = a + d. \tag{4.3}$$

The two formulations of $\hat{\sigma}$ in 4.2 and 4.3 are obviously equivalent. Contingency tables are good tools to calculate the three limit values for $\hat{\sigma}$.

Let us start by considering a situation of independence. In this case the joint distribution of residuals corresponds to the product of the marginals. From table 4.2 we calculate the value of $\hat{\sigma}$ as

$$\sigma_{indep} = 1 + 2\tau^2 - 2\tau.$$

		$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$		
		+	-	
$sign(\mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2_\tau)$	+	$(1-\tau)^2$	$\tau-\tau^2$	$1-\tau$
	-	$\tau-\tau^2$	τ^2	τ
		$1-\tau$	τ	

Table 4.2: Distribution of residuals of quantile regression for independence.

Under the hypothesis of perfect positive dependence we assume that there are no observations that generate discordant residuals. Under this assumption, from table 4.3 we can then calculate the corresponding value of positive dependence for the coefficient of concordance as

$$\sigma_{dep+} = 1.$$

		$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$		
		+	-	
$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$	+	$1-\tau$	0	$1-\tau$
	-	0	τ	τ
		$1-\tau$	τ	

Table 4.3: Distribution of residuals of quantile regression for perfect positive dependence.

The perfect negative dependence is a little more complicated to assess. Under such a hypothesis we would not expect to have any concordant residual. However, as a consequence of the asymmetric structure of quantiles, the cells on the diagonal can never be simultaneously empty if $\tau \neq 0.5$. We must separate the cases with $\tau > 0.50$ and $\tau < 0.50$ and then construct two different tables (table 4.4 and table 4.5), each allowing a small amount of proportion of observations in a different cell of concordance. Note that the case corresponding to the median can be indistinctly considered in both tables.

If $\tau \leq 0.50$

		$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$		
		+	-	
$sign(\mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2_\tau)$	+	$1-2\tau$	τ	$1-\tau$
	-	τ	0	τ
		$1-\tau$	τ	

Table 4.4: Distribution of residuals of quantile regression for perfect negative dependence when $\tau \leq 0.5$.

If $\tau \geq 0.50$

		$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$		
		+	-	
$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$	+	0	$1-\tau$	$1-\tau$
	-	$1-\tau$	$2\tau-1$	τ
		$1-\tau$	τ	

Table 4.5: Distribution of residuals of quantile regression for perfect negative dependence when $\tau \geq 0.5$.

Although we have distinguished the cases $\tau > 0.5$ and $\tau < 0.5$, the final limiting value of perfect negative dependence can be defined in a unique way by taking the absolute value of the sum along the diagonal, so that

$$\sigma_{dep-} = |2\tau - 1|.$$

It is is easy to see that the calculated limit values of $\hat{\sigma}$ are ordered as

$$0 \leq \sigma_{dep-} < \sigma_{indep} \leq \sigma_{dep+} = 1.$$

In the definition of the indicator of concordance, in equation (4.1), we have considered the null residuals in the same group of the negative ones.

An alternative could have been to isolate the proportion of null residuals since, from theorem 2 in section 1.3 we know that for every quantile regression model there are at least q residuals equal to zero. We then construct limiting values for the coefficient of concordance in a second way by considering residuals divided into the three categories: negative, null and positive.

Let $k \geq q$ be the number of null residuals, resulting in a proportion of $\frac{k}{n}$, where n is the number of observed subjects. It follows that there will be $\tau - \frac{k}{2n} = \frac{2n\tau - k}{2n}$ negative residuals and $(1-\tau) - \frac{k}{2n} = \frac{2n(1-\tau)-k}{2n}$ positive ones. We construct a corresponding contingency table, of dimension 3×3, and take the sum of the proportions on the diagonal as an estimate of the statistic $\hat{\sigma}$, in the same way as we have done in the first categorization of residuals (table 4.1). The new classification of residuals is reported in table 4.6 from which we calculate

$$\hat{\sigma} = a + e + l.$$

		$sign(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$			
		+	0	-	
$sign(\mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2_\tau)$	+	a	b	c	$\frac{2n(1-\tau)-k}{2n}$
	0	d	e	f	k/n
	-	h	i	l	$\frac{2n\tau-k}{2n}$
		$\frac{2n(1-\tau)-k}{2n}$	$\frac{k}{n}$	$\frac{2n\tau-k}{2n}$	

Table 4.6: Distribution of residuals of quantile regression if zero residuals apart.

The resulting three limiting values for the statistic remained unchanged except for the limit corresponding to independence. The latter becomes

$$\sigma_{indep} = \frac{[2n(1-\tau)-k]^2 + 4k^2 + (2n\tau - k)^2}{(2n)^2}.$$

The dependence on sample size for this threshold value introduces complications in the interpretation of the strength of association.

In fact, with finite sample sizes it may occur that the order $0 \leq \sigma_{dep-} < \sigma_{indep} \leq \sigma_{dep+} = 1$ is not valid anymore.

We preferred to keep the easy interpretation of the results coming from the first formulation. When the number of null residuals is too large to not be taken into consideration, one could always evenly split them between the negative and the positive ones.

The limiting values chosen for the $\hat{\sigma}$ statistic depend only on the quantile analyzed. As a reference we report in table 4.7 these values calculated for a selection of quantiles.

τ	0.1/0.9	0.2/0.8	0.3/0.7	0.4/0.6	0.5
σ_{dep-}	0.8	0.6	0.4	0.2	0.0
σ_{indep}	0.82	0.68	0.58	0.52	0.50
σ_{dep+}	1	1	1	1	1

Table 4.7: $\hat{\sigma}$ statistic values under hypothesis of perfect negative dependence, perfect independence and perfect positive dependence for 9 quantiles

In order to preserve the interpretation in terms of probability of concordance, we did not normalize our coefficient of dependence .

As to the other properties required from a good measure of dependence, we have that:

1 $\hat{\sigma}$ is symmetric: $\hat{\sigma}(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau, \mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2_\tau) = \hat{\sigma}(\mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2_\tau, \mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau)$;

2 $\hat{\sigma}(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau, \mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2_\tau) = max(\hat{\sigma}) \iff$ the signs of residuals are comonotonic;

3 $\hat{\sigma}(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1_\tau, \mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2_\tau) = min(\hat{\sigma}) \iff$ the signs of residuals are countermonotonic;

4 there is a unique value of the statistic $\hat{\sigma}$ at which independence is attained.

For what concerns invariance, we can prove that the $\hat{\sigma}$ coefficient is invariant to shift transforms of the outcome $\mathbf{y}^i$. That is, given $\boldsymbol{\gamma}_\tau \in \mathbb{R}^q$

$$\begin{aligned} &\hat{\sigma}(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1(\tau : \mathbf{y}^1, \mathbf{X}), \mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2(\tau; \mathbf{y}^2, \mathbf{X})) = \\ &= \hat{\sigma}([\mathbf{y}^1 + \mathbf{X}\boldsymbol{\gamma}_\tau] - \mathbf{X}\hat{\boldsymbol{\beta}}^1(\tau : \mathbf{y}^1 + \mathbf{X}\boldsymbol{\gamma}_\tau, \mathbf{X}), \mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2(\tau; \mathbf{y}^2, \mathbf{X})) \end{aligned}$$

where, to simplify the notation, we have indicated by $\hat{\boldsymbol{\beta}}^i(\tau; l, m)$ the estimated regression parameter based on the observations (l, m). This result is a direct consequence of the fact that, as shown with equation (1.15) in section 1.5, quantile residuals are invariant under this kind of transformations.

Another special property of equivariance for the $\hat{\sigma}$ coefficient is related to the equivariance property to monotonic transformations, discussed in section 1.5.

For a nondecreasing function h of the response $\mathbf{y}^i$ we have that

$$\hat{\sigma}(\mathbf{y}^1 - \mathbf{X}\hat{\boldsymbol{\beta}}^1(\tau : \mathbf{y}^1, \mathbf{X}), \mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2(\tau; \mathbf{y}^2, \mathbf{X})) = \hat{\sigma}(\mathbf{y}^1 - h^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}}^1(\tau : h(\mathbf{y}^1), \mathbf{X})), \mathbf{y}^2 - \mathbf{X}\hat{\boldsymbol{\beta}}^2(\tau; \mathbf{y}^2, \mathbf{X})).$$

This last result follows from the equality $\mathbf{X}\hat{\boldsymbol{\beta}}^1(\tau : \mathbf{y}^1, \mathbf{X}) = h^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}}^1(\tau : h(\mathbf{y}^1), \mathbf{X}))$ between the linear predictor of the quantile regression model on the simple outcome $\mathbf{y}^1$ and the inverse transform of the linear predictor of $h(\mathbf{y}^1)$.

Chapter 5

Modeling the concordance of probability

We recall that, under a specified quantile regression model, there is independence between regression residuals and covariates

$$P(\varepsilon^i \leq 0 \mid \mathbf{X}) = \tau \quad i = 1, 2.$$

The joint distribution of residuals, however, can still depend on such covariates. In fact, it may happen that

$$P(\varepsilon^1 \leq 0 \mid \varepsilon^2 \leq 0 \wedge \mathbf{X}) \neq P(\varepsilon^1 \leq 0 \mid \varepsilon^2 \leq 0).$$

For this reason the probability of concordant regression residuals, denoted by $\sigma_{\mathbf{X}} = P(Z = 1|\mathbf{X})$, can be modeled conditionally on covariates. Its modelization would allow us to detect if, once performed univariate analysis, there is still a residual conditional association. The set of covariates used in the modelization could be the same as the univariate quantile regressions, or a different one. For simplicity, in the following notation we still denote by $\mathbf{X}$ the matrix of independent variables, being aware that it could be different from the matrix used in the previous steps of the model.

The probability of concordance could be analyzed by considering different regression models. Our first instinct would be to use standard models for probabilities, such as logit and probit regressions. However, these could generate predictions outside the specific limits calculated from the concordance coefficient. We therefore prefer to define ad hoc regression models, flexible enough to predict probabilities in the suitable range.

In the following chapter we will propose two alternative models constructed as to constrain the predicted probabilities in the feasible range of the residuals' signs statistic. For each of them we will present statistical properties and insights.

5.1 Constrained logistic regression model

The first model proposed is a variant of the classical logistic regression. Logistic regression is a generalized linear model (*glm*) developed to analyze relationships between binary dependent variables and covariates. For the complete theory of glm methods we refer to McCullagh and Nelder (1989) [29]. Let denote with $\pi(\mathbf{X}) = P(Y = 1|\mathbf{X})$ the conditional probability associated to a random variable Y. Logistic regression models this probability by taking a continuous and increasing transform $g(\pi) = \log(\frac{\pi}{1-\pi})$ that maps the probabilities π, belonging to the unit interval $(0, 1)$, into the real line $(-\infty, \infty)$. The use of this transform enables the correct prediction of probabilities in their range of definition. The modelization is based on the assumption that the random response Y follows a Bernoulli distribution of probability π.

In our specific situation we want to model the probabiility $\sigma_{\mathbf{X}} = P(Z = 1|\mathbf{X})$ for the indicator of concordance Z. The limits imposed by the definition of $\hat{\sigma}$ reduce the range of allowed predicted probabilities when $\tau \neq 0.5$. The framework of logistic regression, whose predicted probabilities are defined in the entire interval $(0, 1)$, is not completely exhaustive. We therefore decided to modify the logit link of classical logistic regression in order to constrain predictions to the interval $(\sigma_{dep-}, \sigma_{dep+})$.

Let us rescale the probability $\sigma_{\mathbf{X}} \in (\sigma_{dep-}, \sigma_{dep+})$, extending its interval of definition to the entire unit interval $(0, 1)$

$$p = \frac{\sigma_{\mathbf{X}} - \sigma_{dep-}}{\sigma_{dep+} - \sigma_{dep-}}$$

p being defined in $(0, 1)$, in accordance with the *glm* theory, we need to select an appropriate link function $g : (0, 1) \to \mathbb{R}$ such that

$$g(p) = \mathbf{X}\boldsymbol{\gamma}_\tau$$

where the linear predictor $\mathbf{X}\boldsymbol{\gamma}_\tau \in \mathbb{R}$ and $\boldsymbol{\gamma}_\tau$ is the vector of unknown parameters of the regression. The inverse of any *cdf* $F : \mathbb{R} \to (0, 1)$ would always be a correct choice for this kind of link function. Following the literature on probability modelization, and for easy of interpretation, we decided to consider a logit link such that the final model of probability is defined as

$$\begin{aligned} \text{logit}^*(p) &= \log\left(\frac{p}{1-p}\right) = \log\left(\frac{\frac{\sigma_{\mathbf{X}}-\sigma_{dep-}}{\sigma_{dep+}-\sigma_{dep-}}}{1-\frac{\sigma_{\mathbf{X}}-\sigma_{dep-}}{\sigma_{dep+}-\sigma_{dep-}}}\right) \\ &= \log\left(\frac{\sigma_{\mathbf{X}} - \sigma_{dep-}}{\sigma_{dep+} - \sigma_{\mathbf{X}}}\right) = \mathbf{X}\boldsymbol{\gamma}_\tau + \eta \end{aligned} \tag{5.1}$$

where the subscript τ in the vector of parameters is referred to as the selected quantile used to define the vector of concordance $\mathbf{z}$, and $\boldsymbol{\eta}$ is the error term of the model.

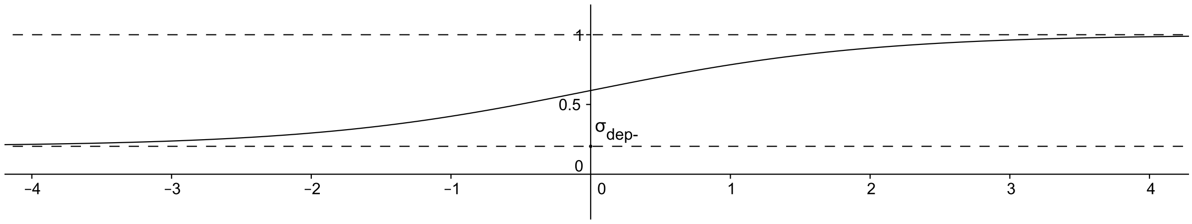

Figure 5.1: Logistic constrained function σ_t. Note that $\sigma_t \in (\sigma_{dep-}, \sigma_{dep+} = 1)$, boundaries are attained only asymptotically. The value of σ_{dep-} is 0.2 and corresponds to the choiche of a quantile $\tau = 0.6$ or $\tau = 0.4$.

The model in (5.1) assures predictions of probabilities inside the suitable range. In fact, the inverse transform of the logit-modified link, see figure 5.1, is

$$\sigma_{\mathbf{X}} = \frac{\sigma_{dep-} + \exp(\mathbf{X}\boldsymbol{\gamma}_\tau)}{\sigma_{dep+} + \exp(\mathbf{X}\boldsymbol{\gamma}_\tau)} \in (\sigma_{dep-}, \sigma_{dep+}).$$

This model, similarly to the classical logistic, can be expressed as a *glm* where the response $Z \sim \text{Bernoulli}(\sigma_{\mathbf{X}})$. Regression parameters can thus be estimated by considering the method of maximum likelihood.

Given a sample of n independent observations from the random variable Z, and a probability of concordance $\sigma_{\mathbf{X}}$ defined in $(\sigma_{dep-}, \sigma_{dep+})$, the likelihood function is expressed as

$$L(\boldsymbol{\gamma}_\tau | z_j, \mathbf{x}_j) = \prod_{j=1}^{n} \sigma_{\mathbf{x}_j}^{z_j} (1 - \sigma_{\mathbf{x}_j})^{1-z_j}$$

where $\boldsymbol{\gamma}_\tau = \{\gamma_{\tau,0}, \ldots, \gamma_{\tau,q-1}\}$ is the unknown vector of parameters to be estimated.

The expression for the log-likelihood is

$$\begin{aligned}
\ell(\boldsymbol{\gamma}_\tau | z_j, \mathbf{x}_j) &= \sum_{j=1}^{n} [z_j \log(\sigma_{\mathbf{x}_j}) + (1 - z_j) \log(1 - \sigma_{\mathbf{x}_j})] \\
&= \sum_{j=1}^{n} [\log(1 - \sigma_{\mathbf{x}_j}) + z_j \log(\sigma_{\mathbf{x}_j}) - z_j \log(1 - \sigma_{\mathbf{x}_j})] \\
&= \sum_{j=1}^{n} \left[log(1 - \sigma_{\mathbf{x}_j}) + z_j \log\left(\frac{\sigma_{\mathbf{x}_j}}{1 - \sigma_{\mathbf{x}_j}} \right) \right] \\
&= \sum_{j=1}^{n} \left[\log\left(1 - \frac{\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{1 + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \right) \right. \\
&+ \left. z_j \log\left(\frac{\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{1 + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau) - \sigma_{dep-} - \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \right) \right]
\end{aligned}$$

$$= \sum_{j=1}^{n} \left[\log \left(\frac{1-\sigma_{dep-}}{1+\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \right) + z_j \log \left(\frac{\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{1-\sigma_{dep-}} \right) \right]$$

where we recall that we always have $\sigma_{dep+} = 1$.

The first derivative of the loglikelihood with respect to $\boldsymbol{\gamma}_\tau$ is

$$\begin{aligned} S(\boldsymbol{\gamma}_\tau) &= \frac{\partial \ell(\boldsymbol{\gamma}_\tau | z_j, \mathbf{x}_j)}{\partial \boldsymbol{\gamma}_\tau} \\ &= \sum_{j=1}^{n} \left[-\frac{\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{1+\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \mathbf{x}_j + z_j \mathbf{x}_j \frac{\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \right] \qquad (5.2) \\ &= \sum_{j=1}^{n} \left[\mathbf{x}_j \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau) \left(-\frac{1}{1+\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} + \frac{z_j}{\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \right) \right]. \end{aligned}$$

Similar calculations yeld the second derivative with respect to $\boldsymbol{\gamma}_\tau$

$$\frac{\partial S(\boldsymbol{\gamma}_\tau)}{\partial \boldsymbol{\gamma}_\tau} = \sum_{j=1}^{n} \mathbf{x}_j \mathbf{x}_j^T \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau) \left[-\frac{1}{(1+\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau))^2} + \frac{z_j \sigma_{dep-}}{(\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau))^2} \right].$$

The maximum likelihood estimate for the vector $\boldsymbol{\gamma}_\tau$ of model (5.1) is calculated by setting (5.2) to zero.

Using a logistic regression model has the advantage of easy interpretation of parameters in terms of odds ratios. However, there are some drawbacks connected to the definition of the logit* function. Since the defined loglikelihood is strictly concave if its equation admits a solution, this solution is unique and coincides with the maximum likelihood estimator. Unfortunately, the likelihood equation does not always have a solution. Problems arises as the parameters tend to $\pm\infty$, i.e. if the observed probability is such that $\sigma_{\mathbf{X}} \leq \sigma_{dep-}$ or $\sigma_{\mathbf{X}} \geq \sigma_{dep+}$.

Such convergence problems are not so rare when analizing concordance for extreme quantile residuals. In such cases alternative methods should be considered.

Many authors have addressed this issue for the classical logistic regression. The solution proposed by Heinze and Schemper (2002) [19] could also be applied to our model.

5.2 Constrained polynomial regression model

The second model proposed is a nonlinear regression model, in which the relationship between the outcome and the covariates is studied through a nonlinear functional. This kind of model offers the opportunity to analyze a wide range of functional forms of the regression relationship.

A general nonlinear model is expressed as

$$z_j = h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau) + \eta_j, \quad j = 1, \ldots, n \qquad (5.3)$$

where the error terms are assumed to be i.i.d with conditional mean given by $\mathbb{E}[\eta_j | h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau)] = 0$ and homoskedastic. The function $h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau)$ is continuously differentiable with respect to the components of $\boldsymbol{\gamma}_\tau$.

Among the many different possible choices we opted for a transform able to constrain predicted probabilities in the range $[\sigma_{dep-}, \sigma_{dep+}]$. We recall that $\sigma_{dep+} = 1$.

The function that we consider is the one defined as

$$h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau) = \begin{cases} 1 & \text{if } \mathbf{x}_j^T \boldsymbol{\gamma}_\tau \geq 1 \\ (1 - \sigma_{dep-})(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)^2 (3 - 2\mathbf{x}_j^T \boldsymbol{\gamma}_\tau) + \sigma_{dep-} & \text{if } \sigma_{dep-} < \mathbf{x}_j^T \boldsymbol{\gamma}_\tau < 1 \\ \sigma_{dep-} & \text{if } \mathbf{x}_j^T \boldsymbol{\gamma}_\tau \leq 0 \end{cases} \tag{5.4}$$

which is a polynomial transform of the predictor $\mathbf{x}_j^T \boldsymbol{\gamma}_\tau$.

The polynomial function, illustrated in figure 5.2, satisfies regularity conditions of continuity and differentiability and is nondecreasing on the real line.

A nonlinear regression model can be fitted using the method of moments. In particular, when the error terms are i.i.d. as in this case, nonlinear least squares is proved to be the most efficient method to be used from the class of moment estimators (see [9]).

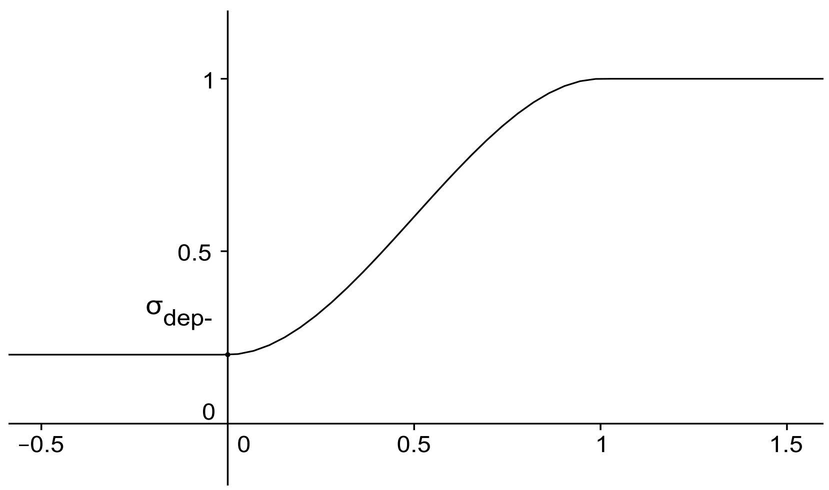

Figure 5.2: Polynomial constrained function $h(t) \in [\sigma_{dep-}, \sigma_{dep+} = 1]$. The function is well defined at the boundaries of the space. The value of extreme negative dependence used in the figure is $\sigma_{dep-} = 0.2$, it comes from the choice of $\tau = 0.4$ or $\tau = 0.6$.

The nonlinear square estimator for the model in (5.3) with functional $h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau)$ defined in (5.4) is the minimizer of the following sum of squares

$$S(\boldsymbol{\gamma}_\tau) = \sum_{j=1}^{n} (z_j - h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau))^2 \tag{5.5}$$

$$= \begin{cases} \sum_{j=1}^{n}(z_j - 1)^2 & \text{if } \mathbf{x}_j^T\boldsymbol{\gamma}_\tau \geq 1 \\ \sum_{j=1}^{n}[z_j - (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) + \sigma_{dep-}]^2 & \text{if } \sigma_{dep-} < \mathbf{x}_j^T\boldsymbol{\gamma}_\tau < 1 \\ \sum_{j=1}^{n}(z_j - \sigma_{dep-})^2 & \text{if } \mathbf{x}_j^T\boldsymbol{\gamma}_\tau \leq 0 \end{cases}$$

with first order derivatives given by

$$\frac{\partial S(\boldsymbol{\gamma}_\tau)}{\partial \boldsymbol{\gamma}_\tau} = 2\sum_{j=1}^{n}[z_j - h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau)]\frac{\partial h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau)}{\partial \boldsymbol{\gamma}_\tau} = \tag{5.6}$$

$$\begin{cases} 12\sum_{j=1}^{n}[z_j - (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) + \sigma_{dep-}]\mathbf{x_j}^T(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)(1-\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) \\ \qquad\qquad \text{if } \sigma_{dep-} < \mathbf{x}_j^T\boldsymbol{\gamma}_\tau < 1 \\ 0 \qquad\qquad \text{otherwise .} \end{cases}$$

The interpretation of the estimated parameters $\hat{\boldsymbol{\gamma}}_\tau$ in nonlinear regression models is not always straightforward. However, the interpretability of the $\boldsymbol{\gamma}_\tau$ parameters is beyond our scope. We rather seek to have a good prediction of probabilities. The flexibility offered by the introduction of nonlinear transformations of the predictor is a powerful tool. The polynomial $h(\mathbf{x}_j, \boldsymbol{\gamma}_\tau)$ in 5.4 presents the advantage of being always defined in the entire range; no problems arise when the estimates reach the boundaries of the parameter space. This is a major advantage compared to the constrained logistic alternative proposed in first instance.

5.3 Asymptotic properties of estimators of concordance probability

The vector $\mathbf{z}$ of observations of the dependent variable Z, used to model the probability of concordance, derives from the estimated residuals in univariate quantile regression models. The variance of quantile regression coefficients, $\boldsymbol{\beta}_\tau^1$ and $\boldsymbol{\beta}_\tau^2$, needs to be considered when estimating the variance of the coefficient vector $\boldsymbol{\gamma}_\tau$. This kind of estimation structure is known in the literature as two-step estimation problem. M-estimators, introduced by Huber in 1967 [21], can be used to provide large sample approximations of the variance. Stefanski and Boos (2002) [38] and Hardin (2002) [18] provide an exhaustive description of this kind of methods and referred to the particular case of two-stage estimates in which moment estimating equations are stacked together in a single partitioned estimating equation as "partial M-estimation".

Our goal is to derive a corrected estimator of the variance of $\boldsymbol{\gamma}_\tau$ in a form commonly known as sandwich variance estimator. The problem reduces to the resolution of a system of three estimating equations. Following Hardin (2002) [18], we write a unique 3×1 dimensional vector of estimating functions

$$\Psi(\mathbf{T}, \boldsymbol{\theta}) = \begin{bmatrix} \Psi_1(Y^1, \boldsymbol{\beta}_\tau^1) \\ \Psi_2(Y^2, \boldsymbol{\beta}_\tau^2) \\ \Psi_3(Z, \boldsymbol{\gamma}_\tau | \boldsymbol{\beta}_\tau^1, \boldsymbol{\beta}_\tau^2) \end{bmatrix} = 0 \tag{5.7}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \boldsymbol{\gamma}_\tau)$ is the complete vector of parameters to be estimated and $\mathbf{T} = (Y^1, Y^2, Z)$ is the vector of response variables.

M-estimators satisfy asymptotic properties of consistency under the following mild regularity conditions (see Boos and Stefanski (2013) [2]):

1. $T_j = (Y_j^1, Y_j^2, Z_j)$ with $j = 1, \ldots, n$ is a triplet of i.i.d. samples, each with a known *cdf* F_h, $h = 1, 2, 3$;

2. $\boldsymbol{\theta}$ is a compact parameter space in $\mathbb{R}^{q\times 3}$;

3. $\Psi(\mathbf{T}, \boldsymbol{\theta})$ is a vector of continuous and twice differentiable functions, where each of its components (Ψ_1, Ψ_2 and Ψ_3) is bounded by an integrable function of its response variable not dependent on the elements of $\boldsymbol{\theta}$;

4. $\mathbb{E}[\Psi(\mathbf{T}, \boldsymbol{\theta}_0)] = 0$ where $\boldsymbol{\theta}_0$ is the true vector of values.

Some additional conditions are needed to establish asymptotic normality.

Theorem 11. *(Asymptotic normality of $\hat{\gamma}_\tau$). Under the above assumptions 1-4, suppose that $\Psi^n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{j=1}^n \Psi(t_j, \hat{\boldsymbol{\theta}}) = o_p(n^{-\frac{1}{2}})$ and that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$. In addition, assume that*

1a. for each $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, there exist a vector of functions $g(t) = (g_1(\mathbf{y}^1), g_1(\mathbf{y}^2), g_1(\mathbf{z}))$ (possible depending on $\boldsymbol{\theta}_0$) such that for all $\ell, k, h \in \{1, 2, 3\}$

$$\left| \frac{\partial^2}{\partial\boldsymbol{\theta}_\ell \partial\boldsymbol{\theta}_k} \Psi_h(t_h, \boldsymbol{\theta}) \right| \leq g_h(t_h)$$

with $t_1 = (\mathbf{y}^1)$, $t_2 = (\mathbf{y}^2)$, $t_3 = (\mathbf{z})$;

2a. $\mathbf{A}(\boldsymbol{\theta}_0) = \mathbb{E}[-D\Psi(\mathbf{T}, \boldsymbol{\theta}_0)]$ exists and is non singular

3a. $\mathbf{B}(\boldsymbol{\theta}_0) = \mathbb{E}[\Psi(\mathbf{T}, \boldsymbol{\theta}_0)\Psi(\mathbf{T}, \boldsymbol{\theta}_0)^T]$ exists and is finite

4a. the joint pdf *$F_{Y^1,Y^2}(\mathbf{X}\boldsymbol{\beta}_\tau^1, \mathbf{X}\boldsymbol{\beta}_\tau^2)$ and the joint* cdf *$f_{Y^1,Y^2}(\mathbf{X}\boldsymbol{\beta}_\tau^1, \mathbf{X}\boldsymbol{\beta}_\tau^2)$ exist.*

Then

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} MVN(0, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^T) \quad \text{as } n \to \infty \tag{5.8}$$

and in particular

$$n^{1/2}(\hat{\boldsymbol{\gamma}}_\tau - \boldsymbol{\gamma}_{0,\tau}) \xrightarrow{d} N(0, \mathbf{V}_S(\boldsymbol{\gamma}_\tau)) \quad \text{as } n \to \infty$$

where

$$\mathbf{V}_S(\boldsymbol{\gamma}_\tau) = [A^{-1}(\boldsymbol{\theta})B(\boldsymbol{\theta})(A^{-1}(\boldsymbol{\theta}))^T]_{(3,3)}.$$

Proof. For the first part of the theorem we have followed the proof of Theorem 7.2 by Boos and Stefanski (pag. 327) [2]. The proof is based on a component-wise expansion of $\Psi^n(\hat{\boldsymbol{\theta}})$.

We know by assumption that $\Psi^n(\hat{\boldsymbol{\theta}}) = o_p(n^{-\frac{1}{2}})$ and thus a Taylor series expansion of the h-th component of $\Psi^n(\hat{\boldsymbol{\theta}})$ results in

$$\begin{aligned} o_p(n^{-\frac{1}{2}}) &= \Psi^n_h(\hat{\boldsymbol{\theta}}) \\ &= \Psi^n_h(\boldsymbol{\theta}_0) + \Psi^{n'}_h(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \Psi^{n''}_h(\tilde{\boldsymbol{\theta}}^*_h)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \Psi^n_h(\boldsymbol{\theta}_0) + \left\{ \Psi^{n'}_h(\boldsymbol{\theta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \Psi^{n''}_h(\tilde{\boldsymbol{\theta}}^*_h) \right\}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}^*_h$ is on the line segment joining $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, $h = 1, 2, 3$. Writing these 3 equations in matrix notation we have

$$o_p(n^{-\frac{1}{2}}) = \Psi^n(\boldsymbol{\theta}_0) + \left\{ \Psi^{n'}(\boldsymbol{\theta}_0) + \frac{1}{2}\tilde{Q}^* \right\}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\tilde{Q}^*$ is the 3×3 matrix with h-th row given by $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \Psi^{n''}_h(\tilde{\boldsymbol{\theta}}^*_h)$. Note that under condition $1a$, each entry in $\tilde{Q}^*$ is bounded by $||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|| n^{-1} \sum g_h(t_h) = o_p(1)$, and thus $\tilde{Q}^* = o_p(1)$. By the weak law of large numbers, $\Psi^{n'}(\boldsymbol{\theta}_0) \xrightarrow{p} -A(\boldsymbol{\theta}_0)$ which is nonsingular under condition $2a$. Thus for n sufficiently large, the matrix in brackets above is nonsingular with probability approaching 1. By the central limit theorem, $\sqrt{n}\Psi^n(\boldsymbol{\theta}_0) \xrightarrow{d} MVN(0, B(\boldsymbol{\theta}_0))$. On the set where the matrix in brackets is nonsingular (lets that set S_N), we have

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left\{ \Psi^{n'}(\boldsymbol{\theta}_0) + \frac{1}{2}\tilde{Q}^* \right\}^{-1} \{\sqrt{n}\Psi^n(\boldsymbol{\theta}_0) + o_p(1)\}.$$

Slutsky's Theorem [36] in combination with the previous results then gives

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} MVN(0, A^{-1}B(A^{-1})^T)$$

where we note that $P(S_N) = 1$. The estimated sandwich variance matrix can be written in a compact form as

$$\hat{V} = \begin{bmatrix} \{\hat{V}_S(\boldsymbol{\beta}^1)\}_{q\times q} & \{\hat{Cov}_S(\boldsymbol{\beta}^1, \boldsymbol{\beta}^2)\}_{q\times q} & \{\hat{Cov}_S(\boldsymbol{\beta}^1, \boldsymbol{\gamma}_\tau)\}_{q\times q} \\ \{\hat{Cov}_S(\boldsymbol{\beta}^2, \boldsymbol{\beta}^1)\}_{q\times q} & \{\hat{V}_S(\boldsymbol{\beta}^2)\}_{q\times q} & \{\hat{Cov}_S(\boldsymbol{\beta}^2, \boldsymbol{\gamma}_\tau)\}_{q\times q} \\ \{\hat{Cov}_S(\boldsymbol{\gamma}_\tau, \boldsymbol{\beta}^1)\}_{q\times q} & \{\hat{Cov}_S(\boldsymbol{\gamma}_\tau, \boldsymbol{\beta}^2)\}_{q\times q} & \{\hat{V}_S(\boldsymbol{\gamma}_\tau)\}_{q\times q} \end{bmatrix}$$

and we are interested in the lower right partition of the matrix.

Let us consider the two matrices defined in conditions $2a$ and $3a$

$$A(\boldsymbol{\theta}_0) = \mathbb{E} \begin{bmatrix} \frac{\partial \Psi_1}{\partial (\boldsymbol{\beta}^1_\tau)^T} & 0 & 0 \\ 0 & \frac{\partial \Psi_2}{\partial (\boldsymbol{\beta}^2_\tau)^T} & 0 \\ \frac{\partial \Psi_3}{\partial (\boldsymbol{\beta}^1_\tau)^T} & \frac{\partial \Psi_3}{\partial (\boldsymbol{\beta}^2_\tau)^T} & \frac{\partial \Psi_3}{\partial \boldsymbol{\gamma}^T_\tau} \end{bmatrix}, B(\boldsymbol{\theta}_0) = \mathbb{E} \begin{bmatrix} \Psi_1\Psi_1^T & \Psi_1\Psi_2^T & \Psi_1\Psi_3^T \\ \Psi_2\Psi_1^T & \Psi_2\Psi_2^T & \Psi_2\Psi_3^T \\ \Psi_3\Psi_1^T & \Psi_3\Psi_2^T & \Psi_3\Psi_3^T \end{bmatrix}$$

where we have used the fact that the following derivatives are null $\frac{\partial \Psi_1}{\partial (\boldsymbol{\beta}_\tau^2)^T} = \frac{\partial \Psi_1}{\partial \boldsymbol{\gamma}_\tau^T} = \frac{\partial \Psi_2}{\partial (\boldsymbol{\beta}_\tau^1)^T} = \frac{\partial \Psi_2}{\partial \boldsymbol{\gamma}_\tau^T} = 0$. This result follows directly from the definition of the vector of estimating functions in (5.7).

For the sake of notation, let us denote the matrices components as

$$-V_1^{-1} = \mathbb{E}\left(\frac{\partial \Psi_1}{\partial (\boldsymbol{\beta}_\tau^1)^T}\right), \quad -V_2^{-1} = \mathbb{E}\left(\frac{\partial \Psi_2}{\partial (\boldsymbol{\beta}_\tau^2)^T}\right), \quad -V_3^{-1} = \mathbb{E}\left(\frac{\partial \Psi_3}{\partial (\boldsymbol{\gamma}_\tau)^T}\right)$$

$$C_1^* = \mathbb{E}\left(\frac{\partial \Psi_3}{\partial (\boldsymbol{\beta}_\tau^1)^T}\right), \quad C_2^* = \mathbb{E}\left(\frac{\partial \Psi_3}{\partial (\boldsymbol{\beta}_\tau^2)^T}\right) \tag{5.9}$$

$$V_1^{*-1} = \mathbb{E}\left(\Psi_1 \Psi_1^T\right), \quad V_2^{*-1} = \mathbb{E}\left(\Psi_2 \Psi_2^T\right), \quad V_3^{*-1} = \mathbb{E}\left(\Psi_3 \Psi_3^T\right)$$

$$R_{12} = \mathbb{E}\left(\Psi_1 \Psi_2^T\right), \quad R_{13} = \mathbb{E}\left(\Psi_1 \Psi_3^T\right), \quad R_{23} = \mathbb{E}\left(\Psi_2 \Psi_3^T\right).$$

Then the matrices of the sandwich can be rewritten in the form

$$A = \begin{bmatrix} -V_1^{-1} & 0 & 0 \\ 0 & -V_2^{-1} & 0 \\ -C_1^* & -C_2^* & -V_3^{-1} \end{bmatrix}, B = \begin{bmatrix} V_1^{*-1} & R_{12} & R_{13} \\ R_{12}^T & V_2^{*-1} & R_{23} \\ R_{13}^T & R_{23}^T & V_3^{*-1} \end{bmatrix},$$

where A is invertible (cf. condition $2a$) with inverse

$$A^{-1} = \begin{bmatrix} -V_1 & 0 & 0 \\ 0 & -V_2 & 0 \\ V_3 C_1^* V_1 & V_3 C_2^* V_2 & -V_3 \end{bmatrix}.$$

After some algebra we get that the sandwich variance of the $\boldsymbol{\gamma}_\tau$ estimator is

$$\begin{aligned} V_S(\boldsymbol{\gamma}_\tau) &= (V_3 C_1^* V_1 V_1^{*-1} + V_3 C_2^* V_2 R_{12}^T - V_3 R_{13}^T)(V_1 C_1^{*T} V_3) + (V_3 C_1^* V_1 R_{12} + V_3 C_2^* V_2 V_2^{*-1} \\ &- V_3 R_{23}^T)(V_2 C_2^* V_3) - (V_3 C_1^* V_1 R_{13} + V_3 C_2^* V_2 R_{23} - V_3 V_3^{*-1}) V_3 \\ &= V_3 V_3^{*-1} V_3 + V_3 (C_1^* V_1 V_1^{*-1} V_1 C_1^{*T} + C_2^* V_2 R_{12}^T V_1 C_1^{*T} - R_{13}^T V_1 C_1^{*T} + C_1^* V_1 R_{12} V_2 C_2^{*T} \\ &+ C_2^* V_2 V_2^{*-1} V_2 C_2^{*T} - R_{23}^T V_2 C_2^{*T} - C_1^* V_1 R_{13} + C_2^* V_2 R_{23}) V_3 \\ &= V_{S3} + V_3 (C_1^* V_{S1} C_1^{*T} + C_2^* V_2 R_{12}^T V_1 C_1^{*T} - R_{13}^T V_1 C_1^{*T} + C_1^* V_1 R_{12} V_2 C_2^{*T} + C_2^* V_{S2} C_2^{*T} \\ &- R_{23}^T V_2 C_2^{*T} - C_1^* V_1 R_{13} + C_2^* V_2 R_{23}) V_3 \end{aligned}$$

where we indicate

$$V_1 V_1^{*-1} V_1 = V_{S1} = V_S(\boldsymbol{\beta}_\tau^1),$$
$$V_2 V_2^{*-1} V_2 = V_{S2} = V_S(\boldsymbol{\beta}_\tau^2),$$
$$V_3 V_3^{*-1} V_3 = V_{S3}.$$

□

After substituting the values of a selected sample of size n, the vector of equations in (5.7) can be more explicitly written as

$$\begin{cases} \Psi_1(\mathbf{y}^1, \boldsymbol{\beta}^1_\tau) = \frac{\partial}{\partial \boldsymbol{\beta}^1_\tau}(\boldsymbol{\omega}^1 - \tau)(\mathbf{y}^1 - \mathbf{X}\boldsymbol{\beta}^1_\tau) = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^1_\tau}(\omega^1_j - \tau)(y^1_j - \mathbf{x}^T_j \boldsymbol{\beta}^1_\tau) \\ \Psi_2(\mathbf{y}^2, \boldsymbol{\beta}^2_\tau) = \frac{\partial}{\partial \boldsymbol{\beta}^2_\tau}(\boldsymbol{\omega}^2 - \tau)(\mathbf{y}^2 - \mathbf{X}\boldsymbol{\beta}^2_\tau) = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^2_\tau}(\omega^2_j - \tau)(y^2_j - \mathbf{x}^T_j \boldsymbol{\beta}^2_\tau) \\ \Psi_3(\mathbf{z}, \boldsymbol{\gamma}_\tau | \boldsymbol{\beta}^1_\tau, \boldsymbol{\beta}^2_\tau) = \frac{\partial}{\partial \boldsymbol{\gamma}_\tau}(m(\boldsymbol{\gamma}_\tau | \mathbf{z}, \mathbf{X})) = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\gamma}_\tau}(m(\boldsymbol{\gamma}_\tau | z_j, \mathbf{x}_j)) \end{cases}$$

where $\omega^i = \mathbb{I}(\mathbf{y}^i \leq \mathbf{X}\boldsymbol{\beta}^i)$ and with m we denote the general equation representing the model selected to study the conditional probability of concordance. The first two equations, which are common to the two proposed models, are those of univariate quantile regression.

The indicator variable of concordance Z can be expressed in a useful way in terms of the indicator function

$$Z = \mathbb{I}(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau)\mathbb{I}(\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau) + (1 - \mathbb{I}(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau))(1 - \mathbb{I}(\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)).$$

From the properties of the indicator function we can rewrite Z as

$$Z = \mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cap (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)} + 1 - \mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cup (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)}$$

whose expected value is

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}(\mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cap (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)} + 1 - \mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cup (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)}) \\ &= P(\mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cap (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)}) + 1 - P(\mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cup (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)}) \\ &= 2F_{Y^1,Y^2}(\mathbf{X}\boldsymbol{\beta}^1_\tau, \mathbf{X}\boldsymbol{\beta}^2_\tau) + 1 - F_{Y^1}(\mathbf{X}\boldsymbol{\beta}^1_\tau) - F_{Y^2}(\mathbf{X}\boldsymbol{\beta}^2_\tau). \end{aligned}$$

In the computations we will need the expected value of Z^2, which is given by

$$\begin{aligned} \mathbb{E}(\mathbf{Z}^2) &= \mathbb{E}[(\mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cap (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)} + 1 - \mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cup (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)})^2] \\ &= \mathbb{E}(3\mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cap (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)} + 1 - \mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cup (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)}) \\ &= 3P(\mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cap (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)}) + 1 - P(\mathbb{I}_{(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}^1_\tau) \cup (\mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}^2_\tau)}) \\ &= 4F_{Y^1,Y^2}(\mathbf{X}\boldsymbol{\beta}^1_\tau, \mathbf{X}\boldsymbol{\beta}^2_\tau) + 1 - F_{Y^1}(\mathbf{X}\boldsymbol{\beta}^1_\tau) - F_{Y^2}(\mathbf{X}\boldsymbol{\beta}^2_\tau). \end{aligned}$$

Theorem 5 in section 1.3 gives the asymptotic expression of the variance for the vectors of quantile regression parameters $\boldsymbol{\beta}^1_\tau$ and $\boldsymbol{\beta}^2_\tau$. Theorem 11, allows us to calculate their asymptotic distribution also in sandwich form, by the application of (5.8), yelding

$$V_{S1} = \frac{\tau(1-\tau)}{\mathbf{X}^T\mathbf{X} f_{Y^1}(\mathbf{X}\boldsymbol{\beta}^1_\tau)^2} \quad V_{S2} = \frac{\tau(1-\tau)}{\mathbf{X}^T\mathbf{X} f_{Y^2}(\mathbf{X}\boldsymbol{\beta}^2_\tau)^2}. \tag{5.10}$$

The two estimating quantile regression functions $\Psi_1(\boldsymbol{\beta}^1_\tau)$ and $\Psi_2(\boldsymbol{\beta}^2_\tau)$ are not differentiable everywhere. We could then interchange the order of differentiation and expectation, at the true parameter value $\boldsymbol{\beta}^i_{0\tau}$, to have

$$\begin{aligned}
-V_1^{-1} &= \mathbb{E}\left(\frac{\partial \Psi_1}{\partial(\boldsymbol{\beta}_\tau^1)^T}\right) = \frac{\partial}{\partial(\boldsymbol{\beta}_\tau^1)^T}\mathbb{E}(\Psi_1)\Big|_{\boldsymbol{\beta}_\tau^1=\boldsymbol{\beta}_{0,\tau}^1} = \frac{\partial}{\partial \beta_\tau^1}\mathbb{E}(-\mathbf{X}(\omega^1-\tau)) \\
&= -\frac{\partial}{\partial \beta_\tau^1}\mathbf{X}(F_{Y^1}(\mathbf{X}\boldsymbol{\beta}_\tau^1)-\tau) = -\mathbf{X}^T\mathbf{X}f_{Y^1}(\mathbf{X}^T\beta_\tau^1).
\end{aligned}$$

Analogously

$$-V_2^{-1} = \mathbb{E}\left(\frac{\partial \Psi_2}{\partial(\boldsymbol{\beta}_\tau^2)^T}\right) = \frac{\partial}{\partial(\boldsymbol{\beta}_\tau^2)^T}\mathbb{E}(\Psi_2)\Big|_{\boldsymbol{\beta}_\tau^2=\boldsymbol{\beta}_{0,\tau}^2} = -\mathbf{X}\mathbf{X}^T f_{Y^2}(\mathbf{X}\beta_\tau^2).$$

We can then derive the central matrix of the sandwich variance as

$$\begin{aligned}
V_1^{*-1} = E(\Psi_1\Psi_1^T) &= \mathbb{E}[\mathbf{X}(\boldsymbol{\omega}^1-\tau)(\mathbf{X}(\boldsymbol{\omega}^1-\tau))^T] = \mathbb{E}[\mathbf{X}^T\mathbf{X}(\boldsymbol{\omega}^1-\tau)^2] \qquad (5.11)\\
&= \mathbb{E}[\mathbf{X}^T\mathbf{X}((\boldsymbol{\omega}^1)^2+\tau^2-\boldsymbol{\omega}^1\tau)] \\
&= \mathbf{X}^T\mathbf{X}\mathbb{E}((\boldsymbol{\omega}^1)^2+\tau^2-\boldsymbol{\omega}^1\tau) = \mathbf{X}^T\mathbf{X}[F_{Y^1}(\mathbf{X}\boldsymbol{\beta}^1\tau)+\tau^2-2\tau] \\
&= \mathbf{X}^T\mathbf{X}[\tau+\tau^2-2\tau] = \mathbf{X}^T\mathbf{X}(\tau-\tau^2) = \mathbf{X}^T\mathbf{X}(\tau(1-\tau))
\end{aligned}$$

and, in the same way, for the second quantile regression model

$$V_2^{*-1} = E(\Psi_2\Psi_2^T) = \mathbb{E}[\mathbf{X}(\boldsymbol{\omega}^2-\tau)(\mathbf{X}(\boldsymbol{\omega}^2-\tau))^T] = \mathbb{E}[\mathbf{X}^T\mathbf{X}(\boldsymbol{\omega}^2-\tau)^2] = \mathbf{X}^T\mathbf{X}(\tau(1-\tau))$$

Computing the products $V_1V_1^{*-1}V_1$ and $V_2V_2^{*-1}V_2$ we obtain the expressions for the variance in (5.10).

We remark that during these computations we employ the strict connection between quantile and cumulative distribution function

$$F_{Y^1}(\mathbf{X}\boldsymbol{\beta}_\tau^1) = F_{Y^2}(\mathbf{X}\boldsymbol{\beta}_\tau^2) = \tau$$

at the true value of the $\boldsymbol{\beta}_\tau^i$ parameter.

In the following paragraphs we will calculate the sandwich variance of $\boldsymbol{\gamma}_\tau$ for the two proposed regression models of probability of concordance.

We will simplify the notation in the following way:

$$F_{Y^1} = F_{Y^1}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1) \quad F_{Y^2} = F_{Y^2}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2) \quad F_{Y^1,Y^2} = F_{Y^1,Y^2}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1, \mathbf{x}_j^T\boldsymbol{\beta}_\tau^2)$$

$$f_{Y^1} = f_{Y^1}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1) \quad f_{Y^2} = f_{Y^2}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2) \quad f_{Y^1,Y^2} = f_{Y^1,Y^2}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1, \mathbf{x}_j^T\boldsymbol{\beta}_\tau^2)$$

$$f_{Y^1,Y^2}dt = f_{Y^1,Y^2}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1, t)dt \quad f_{Y^1,Y^2}ds = f_{Y^1,Y^2}(s, \mathbf{x}_j^T\boldsymbol{\beta}_\tau^2)ds.$$

5.3.1 Asymptotic variance for the constrained logistic regression parameters

Let us consider the vector of observed estimating equations corresponding to the complete model in which the last step involves logistic constrained regression:

$$\begin{cases} \Psi_1(\mathbf{y}^1,\boldsymbol{\beta}^1_\tau) = \frac{\partial}{\partial \boldsymbol{\beta}^1_\tau}(\boldsymbol{\omega}^1-\tau)(\mathbf{y}^1-\mathbf{X}\boldsymbol{\beta}^1_\tau) = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^1_\tau}(\omega^1_j-\tau)(y^1_j-\mathbf{x}_j^T\boldsymbol{\beta}^1_\tau) \\ \Psi_2(\mathbf{y}^2,\boldsymbol{\beta}^2_\tau) = \frac{\partial}{\partial \boldsymbol{\beta}^2_\tau}(\boldsymbol{\omega}^2-\tau)(\mathbf{y}^2-\mathbf{X}\boldsymbol{\beta}^2_\tau) = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^2_\tau}(\omega^2_j-\tau)(y^2_j-\mathbf{x}_j^T\boldsymbol{\beta}^2_\tau) \\ \Psi_3(\mathbf{z},\boldsymbol{\gamma}_\tau|\boldsymbol{\beta}^1_\tau,\boldsymbol{\beta}^2_\tau) = \frac{\partial}{\partial \boldsymbol{\gamma}_\tau}\left[\log\left(\frac{1-\sigma_{dep-}}{1+\exp(\mathbf{X}\boldsymbol{\gamma}_\tau)}\right)+\mathbf{z}\log\left(\frac{\sigma_{dep-}+\exp(\mathbf{X}\boldsymbol{\gamma}_\tau)}{1-\sigma_{dep-}}\right)\right] \\ \qquad = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\gamma}_\tau}\left[\log\left(\frac{1-\sigma_{dep-}}{1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)+z_j\log\left(\frac{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{1-\sigma_{dep-}}\right)\right] \end{cases}$$

where after differentiation the single j elements are

$$(\Psi_1)_j = -\mathbf{x}_j(\omega^1_j-\tau) \quad (\Psi_2)_j = -\mathbf{x}_j(\omega^2_j-\tau)$$

$$(\Psi_3)_j = \left[\mathbf{x}_j\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)\left(-\frac{1}{1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}+\frac{z_j}{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)\right].$$

Following the proof of theorem 11 on the asymptotic normality of $\hat{\boldsymbol{\gamma}}_\tau$ we can calculate the corrected variance step by step, yelding

$$\begin{aligned} (V_3^{-1})_j &= \mathbb{E}\left(-\frac{\partial\Psi_3}{\partial(\boldsymbol{\gamma}_\tau)^T}\right)_j = \mathbb{E}\left(\frac{\partial}{\partial(\boldsymbol{\gamma}_\tau)^T}\mathbf{x}_j^T\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)\left[\frac{1}{1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}-\frac{z_j}{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right]\right) \\ &= \mathbf{x}_j\mathbf{x}_j^T\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)} - \mathbf{x}_j\mathbf{x}_j^T\frac{\exp(\mathbf{x}_j^T\gamma)}{\sigma_{dep-}+\exp(\mathbf{x}_j\boldsymbol{\gamma}_\tau)}[2F_{Y^1,Y^2}+1-2\tau] \end{aligned}$$

$$\begin{aligned} V_{3j}^{*-1} &= \mathbb{E}(\Psi_3\Psi_3^T)_j \\ &= \mathbb{E}\Bigg[\mathbf{x}_j\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)\left(\frac{z_j}{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}-\frac{1}{1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)\mathbf{x}_j^T\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) \\ &\quad \left(\frac{z_j}{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}-\frac{1}{1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)\Bigg] \\ &= \mathbf{x}_j\mathbf{x}_j^T\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2\Bigg[\frac{4F_{Y^1,Y^2}+1-2\tau}{(\sigma_{dep-}+\exp(\mathbf{x}_i^T\boldsymbol{\gamma}_\tau))^2}-\frac{4F_{Y^1,Y^2}+2-4\tau}{(1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau))(\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau))} \\ &+ \frac{1}{(1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau))^2}\Bigg] \end{aligned}$$

$$(V_{S3})_j = (V_3V_3^{*-1}V_3)_j$$

$$= \frac{\left(\frac{4F_{Y^1,Y^2}+1-2\tau}{(\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau))^2} - \frac{4F_{Y^1,Y^2}+2-4\tau}{(1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau))(\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau))} + \frac{1}{(1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau))^2}\right)}{\mathbf{x}_j\mathbf{x}_j^T\left[\frac{1}{1+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)} - \frac{1}{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}(2F_{Y^1,Y^2}+1-2\tau)\right]^2}$$

Note that the variable Z is, by definition, not necessarily differentiable with respect of the parameters of quantile regression $\boldsymbol{\beta}_\tau^1$ and $\boldsymbol{\beta}_\tau^1$. Anyway, the factors C_{1j}^* (5.12) and C_{2j}^* (5.13) can be correctly calculated applying the property of exchangeability of differentiation and expectation operators.

$$\begin{aligned}(C_1^*)_j &= \mathbb{E}\left(\frac{\partial\Psi_3}{(\partial\boldsymbol{\beta}_\tau^1)^T}\right)_j = \frac{\partial}{\partial\boldsymbol{\beta}_\tau^1}\mathbb{E}\left[\mathbf{x}_j\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)\left(\frac{z_i}{\sigma_{dep-}+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)} - \frac{1}{1+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)\right] \\ &= \mathbf{x}_j\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)\frac{\partial}{\partial\boldsymbol{\beta}_\tau^1}\left(\frac{2F_{Y^1,Y^2}+1-F_{Y^1}-F_{Y^2}}{\sigma_{dep-}+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)} - \frac{1}{1+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right) \qquad (5.12) \\ &= \mathbf{x}_j\mathbf{x}_j^T\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - f_{Y^1}\right]\end{aligned}$$

$$\begin{aligned}(C_2^*)_j &= \mathbb{E}\left(\frac{\partial\Psi_3}{\partial\boldsymbol{\beta}_\tau^2}\right)_j = \frac{\partial}{\partial\boldsymbol{\beta}_\tau^2}\mathbb{E}\left[\mathbf{x}_j\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)\left(\frac{z_i}{\sigma_{dep-}+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)} - \frac{1}{1+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)\right] \\ &= \mathbf{x}_j\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)\frac{\partial}{\partial\boldsymbol{\beta}_\tau^2}\left(\frac{2F_{Y^1,Y^2}+1-F_{Y^1}-F_{Y^2}}{\sigma_{dep-}+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)} - \frac{1}{1+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right) \qquad (5.13) \\ &= \mathbf{x}_j\mathbf{x}_j^T\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-}+\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}\right]\end{aligned}$$

$$(C_1^*V_{S1}C_1^{*T})_j = \mathbf{x}_j\mathbf{x}_j^T\left(\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-}+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)^2\frac{\tau(1-\tau)}{f_{Y^1}^2}\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - f_{Y^1}\right]^2$$

$$(C_2^*V_{S2}C_2^{*T})_j = \mathbf{x}_j\mathbf{x}_j^T\left(\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-}+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)^2\frac{\tau(1-\tau)}{f_{Y^2}^2}\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}\right]^2$$

$$\begin{aligned}(R_{12})_j &= \mathbb{E}(\Psi_1\Psi_2^T) = \mathbb{E}[\mathbf{x}_j\mathbf{x}_j^T(\omega^1-\tau)(\omega^2-\tau)] = \mathbb{E}[\mathbf{x}_j\mathbf{x}_j^T(\mathbb{I}(\mathbf{y}^1 \le \mathbf{x}_j^T\boldsymbol{\beta}_\tau^1)-\tau)(\mathbb{I}(\mathbf{y}^2 \le \mathbf{x}_j^T\boldsymbol{\beta}_\tau^2)-\tau)] \\ &= \mathbf{x}_j\mathbf{x}_j^T[F_{Y^1,Y^2}-\tau^2]\end{aligned}$$

$$(R_{13})_j = \mathbb{E}(\Psi_1\Psi_3^T)_j = \mathbb{E}\left[-\mathbf{x}_j\mathbf{x}_j^T(\omega^1-\tau)\left(\frac{z_j\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-}+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)} - \frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{1+exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)\right]$$

$$= -\mathbf{x}_j\mathbf{x}_j^T\left[(F_{Y^1,Y^2} - 2F_{Y^1,Y^2}\tau - \tau + 2\tau^2)\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-} + exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right]$$

$$(R_{13})_j = (R_{23})_j$$

$$\begin{aligned}
(C_2^*V_2R_{12}^TV_1C_1^{*T})_j &= \frac{\mathbf{x}_j\mathbf{x}_j^T}{f_{Y^1}^2 f_{Y^2}^2}\left(\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-} + exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)^2 \\
&\left[4F_{Y^1,Y^2}\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - 2\tau F_{Y^1,Y^2}\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds\right. \\
&- 4\tau^2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds + 2\tau^3\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds \\
&\left.- 2\tau F_{Y^1,Y^2}\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt + \tau^2F_{Y^1,Y^2} + 2\tau^3\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - \tau^4\right]
\end{aligned}$$

$$(C_1^*V_1R_{12}V_2C_2^{*T})_j = (C_2^*V_2R_{12}^TV_1C_1^{*T})_j$$

$$\begin{aligned}
(R_{13}^TV_1C_1^{*T})_j &= \frac{-\mathbf{x}_j\mathbf{x}_j^T}{f_{Y^1}^2}\left(\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-} + exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)^2\left[2F_{Y^1,Y^2}\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt\right. \\
&- \tau F_{Y^1,Y^2} - 4\tau F_{Y^1,Y^2}\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt + 2\tau^2F_{Y^1,Y^2} \\
&\left.- 2\tau\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt + 4\tau^2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - 2\tau^3 + \tau^2\right]
\end{aligned}$$

$$\begin{aligned}
(R_{23}^TV_2C_2^{*T})_j &= \frac{-\mathbf{x}_j\mathbf{x}_j^T}{f_{Y^2}^2}\left(\frac{\exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}{\sigma_{dep-} + exp(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)}\right)^2\left[2F_{Y^1,Y^2}\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds\right. \\
&- \tau F_{Y^1,Y^2} - 4\tau F_{Y^1,Y^2}\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds + 2\tau^2F_{Y^1,Y^2} \\
&\left.- 2\tau\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds + 4\tau^2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - 2\tau^3 + \tau^2\right]
\end{aligned}$$

$$(C_1^*V_1R_{13})_j = (R_{13}^TV_1C_1^{*T})_j \quad \text{and} \quad (C_2^*V_2R_{23})_j = (R_{23}^TV_2C_2^{*T})_j$$

$$(C_1^*V_{S1}C_1^{*T} + C_2^*V_2R_{12}^TV_1C_1^{*T} - R_{13}^TV_1C_1^{*T} + C_1^*V_1R_{12}V_2C_2^{*T} + C_2^*V_{S2}C_2^{*T} - R_{23}^TV_2C_2^{*T}$$

$$
\begin{aligned}
- &\quad C_1^* V_1 R_{13} + C_2^* V_2 R_{23})_j = (C_1^* V_{S1} C_1^{*T} + 2(C_2^* V_2 R_{12}^T V_1 C_1^{*T}) + C_2^* V_{S2} C_2^{*T})_j \\
= &\quad \mathbf{x}_j \mathbf{x}_j^T \left(\frac{\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{\sigma_{dep-} + exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \right)^2 \left\{ \frac{\tau(1-\tau)}{f_{Y^1}^2} \left(2 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt - f_{Y^1} \right)^2 \right. \\
+ &\quad \frac{2}{f_{Y^1}^2 f_{Y^2}^2} \left[4F_{Y^1,Y^2} \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds - 2\tau F_{Y^1,Y^2} \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds \right. \\
- &\quad 4\tau^2 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds + 2\tau^3 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds - 2\tau F_{Y^1,Y^2} \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt \\
+ &\quad \left. \left. \tau^2 F_{Y^1,Y^2} + 2\tau^3 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt - \tau^4 \right] + \frac{\tau(1-\tau)}{f_{Y^2}^2} \left(2 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds - f_{Y^2} \right)^2 \right\}
\end{aligned}
$$

To conclude we get the corrected matrix of variance for the j-th subject as:

$$
\begin{aligned}
(V_S(\boldsymbol{\gamma}_\tau))_j &= [V_{S3} + V_3(C_1^* V_{S1} C_1^{*T} + C_2^* V_2 R_{12}^T V_1 C_1^{*T} - R_{13}^T V_1 C_1^{*T} + C_1^* V_1 R_{12} V_2 C_2^{*T} \\
&+ C_2^* V_{S2} C_2^{*T} - R_{23}^T V_2 C_2^{*T} - C_1^* V_1 R_{13} + C_2^* V_2 R_{23}) V_3]_j \\
&= (V_{S3} + V_3(C_1^* V_{S1} C_1^{*T} + 2(C_2^* V_2 R_{12}^T V_1 C_1^{*T}) + C_2^* V_{S2} C_2^{*T}) V_3)_j \\
&= \frac{1}{\mathbf{x}_j \mathbf{x}_j^T \left[-\frac{\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{1+exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} + \frac{\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{\sigma_{dep-}+exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} (2F_{Y^1,Y^2} + 1 - 2\tau) \right]^2} \\
&\left\{ \frac{4F_{Y^1,Y^2} + 1 - 2\tau}{(\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau))^2} - \frac{4F_{Y^1,Y^2} + 2 - 4\tau}{(1 + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau))(\sigma_{dep-} + \exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau))} \right. \\
&+ \frac{1}{(1 + exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau))^2} + \left(\frac{\exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)}{\sigma_{dep-} + exp(\mathbf{x}_j^T \boldsymbol{\gamma}_\tau)} \right)^2 \left[\frac{\tau(1-\tau)}{f_{Y^1}^2} \left(2 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt \right. \right. \\
&- \left. f_{Y^1} \right)^2 + \frac{2}{f_{Y^1}^2 f_{Y^2}^2} \left(4F_{Y^1,Y^2} \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds \right. \\
&- 2\tau F_{Y^1,Y^2} \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds - 4\tau^2 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds \\
&+ 2\tau^3 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds - 2\tau F_{Y^1,Y^2} \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt + \tau^2 F_{Y^1,Y^2} \\
&+ \left. \left. \left. 2\tau^3 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2} dt - \tau^4 \right) + \frac{\tau(1-\tau)}{f_{Y^2}^2} \left(2 \int_{-\infty}^{\mathbf{x}_j^T \boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2} ds - f_{Y^2} \right)^2 \right] \right\}
\end{aligned}
$$

5.3.2 Asymptotic variance for the constrained polynomial model

Let us rewrite the system of moment equations related to the second regression method proposed in section 5.2 as

$$\begin{cases} \Psi_1(\mathbf{y}^1, \boldsymbol{\beta}^1_\tau) = \frac{\partial}{\partial \boldsymbol{\beta}^1_\tau}(\boldsymbol{\omega}^1 - \tau)(\mathbf{y}^1 - \mathbf{X}\boldsymbol{\beta}^1_\tau) = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^1_\tau}(\omega^1_j - \tau)(y^1_j - \mathbf{x}_j^T \boldsymbol{\beta}^1_\tau) \\ \Psi_2(\mathbf{y}^2, \boldsymbol{\beta}^2_\tau) = \frac{\partial}{\partial \boldsymbol{\beta}^2_\tau}(\boldsymbol{\omega}^2 - \tau)(\mathbf{y}^2 - \mathbf{X}\boldsymbol{\beta}^2_\tau) = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\beta}^2_\tau}(\omega^2_j - \tau)(y^2_j - \mathbf{x}_j^T \boldsymbol{\beta}^2_\tau) \\ \Psi_3(\mathbf{z}, \boldsymbol{\gamma}_\tau | \boldsymbol{\beta}^1_\tau, \boldsymbol{\beta}^2_\tau) = \frac{\partial}{\partial \boldsymbol{\gamma}_\tau}(\mathbf{z} - (1 - \sigma_{dep-})(\mathbf{X}\boldsymbol{\gamma}_\tau)^2(3 - 2\mathbf{x}^T\boldsymbol{\gamma}_\tau - \sigma_{dep-})^2 = \\ \qquad = \sum_{j=1}^n \frac{\partial}{\partial \boldsymbol{\gamma}_\tau}(z_j - (1 - \sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3 - 2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - \sigma_{dep-})^2 \end{cases}$$

We recall that the last estimation step here requires a nonlinear regression model. After differentiation, the expression of the j-th elements in the system above is

$$(\Psi_1)_j = \mathbf{x}_j(\omega^1_j - \tau) \quad (\Psi_2)_j = \mathbf{x}_j(\omega^2_j - \tau)$$

$$(\Psi_3)_j = -12(1-\sigma_{dep-})\mathbf{x}_j[z_j-(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-\sigma_{dep-}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2].$$

Analogously to section 5.3.1, we report the calculations of the sandwich estimator of variance for the polynomial regression parameter $\boldsymbol{\gamma}_\tau$, corrected for the errors of estimation of $\boldsymbol{\beta}^1_\tau$ and $\boldsymbol{\beta}^2_\tau$. We have

$$\begin{aligned} (V_3^{-1})_j &= \mathbb{E}\left(-\frac{\partial \Psi_3}{\partial (\gamma_\tau)^T}\right)_j \\ &= 12(1-\sigma_{dep-})\mathbf{x}_j \frac{\partial}{\partial \boldsymbol{\gamma}_\tau}\mathbb{E}\{[z_j - (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) - \sigma_{dep-}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]\} \\ &= 12(1-\sigma_{dep-})\mathbf{x}_j\mathbf{x}_j^T\{-6(1-\sigma_{dep-})[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2 \\ &+ [2F_{Y^1,Y^2} + 1 - 2\tau - (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) - \sigma_{dep-}][1 - 2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau] \end{aligned}$$

$$\begin{aligned} (V_3^{*-1})_j &= \mathbb{E}(\Psi_3\Psi_3^T)_j \\ &= \mathbb{E}\{(-12)(1-\sigma_{dep-})\mathbf{x}_j[z_j - (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) - \sigma_{dep-}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2] \\ &\quad (-12)(1-\sigma_{dep-})\mathbf{x}_j^T[z_j - (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) - \sigma_{dep-}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau) - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]\} \\ &= 144(1-\sigma_{dep-})^2\mathbf{x}_j\mathbf{x}_j^T[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\{4F_{Y^1,Y^2}[1 - \sigma_{dep-} - (1-\sigma_{dep-})^2 - \tau^2 \\ &- (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)] + 2\tau^2\sigma_{dep-} + [2\sigma_{dep-} + 2\tau^2 - 2][(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2 \\ &\quad (3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)] + [(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)]^2\} \end{aligned}$$

$$(V_{S3})_j = (V_3 V_3^{*-1} V_3)_j$$

$$= \frac{[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\{4F_{Y^1,Y^2}[1-\sigma_{dep-}-\mathbf{H}] + (1-\sigma_{dep-})^2 - \tau^2 + 2\sigma_{dep-}\tau^2}{\mathbf{x}_j\mathbf{x}_j^T\{-6(1-\sigma_{dep-})[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2 + (1-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)[2F_{Y^1,Y^2}+1-2\tau-\sigma_{dep-}-\mathbf{H}]\}}$$
$$+ \frac{[2\sigma_{dep-}+2\tau^2-2][\mathbf{H}]+\mathbf{H}^2\}}{\mathbf{x}_j\mathbf{x}_j^T\{-6(1-\sigma_{dep-})[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2 + (1-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)[2F_{Y^1,Y^2}+1-2\tau-\sigma_{dep-}-\mathbf{H}]\}}.$$

Note that in $(V_{S3})_j$ we have made the substitution $\mathbf{H} = (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)$.

Because of the nondifferentiability of the indicator Z with respect to $\boldsymbol{\beta}_\tau^1$ and $\boldsymbol{\beta}_\tau^2$, as we have already done for 5.12 and 5.13, we exchange differentiation and expectation operators to obtain correct calculations of $(C_1^*)_j$ and $(C_2^*)_j$. Indeed,

$$\begin{aligned}
(C_1^*)_j &= \mathbb{E}\left(\frac{\partial\Psi_3}{\partial\boldsymbol{\beta}_\tau^1}\right)_j = \frac{\partial}{\partial\boldsymbol{\beta}_\tau^1}\mathbb{E}\{-12(1-\sigma_{dep-})\mathbf{x}_j[z_j-(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-\sigma_{dep-}] \\
&\quad [(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]\} \\
&= -12(1-\sigma_{dep-})\mathbf{x}_j\frac{\partial}{\partial\boldsymbol{\beta}_\tau^1}[2F_{Y^1,Y^2}+1-F_{Y^1}-F_{Y^2}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2] \\
&= -12(1-\sigma_{dep-})\mathbf{x}_j\mathbf{x}_j^T[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - f_{Y^1}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]
\end{aligned}$$

$$\begin{aligned}
(C_2^*)_j &= \mathbb{E}\left(\frac{\partial\Psi_3}{\partial\boldsymbol{\beta}_\tau^2}\right)_j = \frac{\partial}{\partial\boldsymbol{\beta}_\tau^2}\mathbb{E}\{-12(1-\sigma_{dep-})\mathbf{x}_j[\mathbf{z_j}-(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-\sigma_{dep-}] \\
&\quad [(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau^2]\} \\
&= -12(1-\sigma_{dep-})\mathbf{x}_j\frac{\partial}{\partial\boldsymbol{\beta}_\tau^2}[2F_{Y^1,Y^2}+1-F_{Y^1}-F_{Y^2}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2] \\
&= -12(1-\sigma_{dep-})\mathbf{x}_j\mathbf{x}_j^T[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}][(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]
\end{aligned}$$

$$(C_1^*V_{S1}C_1^{*T})_j = (-12)^2(1-\sigma_{dep-})^2\mathbf{x}_j\mathbf{x}_j^T\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - f_{Y^1}\right]^2[(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\frac{\tau(1-\tau)}{f_{Y^1}^2}$$

$$(C_2^*V_{S2}C_2^{*T})_j = (-12)^2(1-\sigma_{dep-})^2\mathbf{x}_j\mathbf{x}_j^T\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}\right]^2[(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\frac{\tau(1-\tau)}{f_{Y^2}^2}$$

$$\begin{aligned}
(R_{12})_j &= \mathbb{E}(\Psi_1\Psi_2^T)_j = \mathbb{E}[\mathbf{x}_j\mathbf{x}_j^T(\omega^1-\tau)(\omega^2-\tau)] = \mathbb{E}[\mathbf{x}_j\mathbf{x}_j^T(\mathbb{I}(\mathbf{y}^1 \le \mathbf{x}_j^T\boldsymbol{\beta}_\tau^1)-\tau)(\mathbb{I}(\mathbf{y}^2 \le \mathbf{x}_j^T\boldsymbol{\beta}_\tau^2)-\tau)] \\
&= \mathbf{x}_j\mathbf{x}_j^T[F_{Y^1,Y^2}(\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1, \mathbf{x}_j^T\boldsymbol{\beta}_\tau^2)-\tau^2]
\end{aligned}$$

$$\begin{aligned}
(R_{13})_j &= \mathbb{E}(\Psi_1\Psi_3^T)_j = \mathbb{E}\{\mathbf{x}_j(\omega^1-\tau)12(1-\sigma_{dep-})\mathbf{x}_j[\mathbf{z_j}-(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-\sigma_{dep-}] \\
&\quad [(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]\}
\end{aligned}$$

$$
\begin{aligned}
&= 12(1-\sigma_{dep-})\mathbf{x}_j\mathbf{x}_j^T[(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]\{F_{Y^1,Y^2}+1-2\tau-2\tau\sigma_{dep-}+2\tau^2\\
&- 2\tau[(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)]-2\tau F_{Y^1,Y^2}\}
\end{aligned}
$$

$$(R_{13})_j = (R_{23})_j$$

$$
\begin{aligned}
(C_2^*V_2R_{12}^TV_1C_1^{*T})_j &= \frac{\mathbf{x}_j\mathbf{x}_j^T}{f_{Y^1}^2f_{Y^2}^2}144(1-\sigma_{dep-})^2[(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2[F_{Y^1,Y^2}-\tau^2]\\
&\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}ds - f_{Y^1}\right]\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}\right]
\end{aligned}
$$

$$(C_1^*V_1R_{12}V_2C_2^{*T})_j = (C_2^*V_2R_{12}^TV_1C_1^{*T})_j$$

$$
\begin{aligned}
(R_{13}^TV_1C_1^{*T})_j &= -\frac{\mathbf{x}_j\mathbf{x}_j^T}{f_{Y^1}^2}144(1-\sigma_{dep-})^2[(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}ds - f_{Y^1}\right]\\
&[(F_{Y^1,Y^2}+1)(1-2\tau)-2\tau\sigma_{dep-}+2\tau^2-2\tau(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)]
\end{aligned}
$$

$$
\begin{aligned}
(R_{23}^TV_2C_2^{*T})_j &= -\frac{\mathbf{x}_j\mathbf{x}_j^T}{f_{Y^2}^2}144(1-\sigma_{dep-})^2[(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}\right]\\
&[(F_{Y^1,Y^2}+1)(1-2\tau)-2\tau\sigma_{dep-}+2\tau^2-2\tau(1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)]
\end{aligned}
$$

$$(C_1^*V_1R_{13})_j = (R_{13}^TV_1C_1^{*T})_j \quad \text{and} \quad (C_2^*V_2R_{23})_j = (R_{23}^TV_2C_2^{*T})_j$$

$$
\begin{aligned}
&(C_1^*V_{S1}C_1^{*T}+C_2^*V_2R_{12}^TV_1C_1^{*T}-R_{13}^TV_1C_1^{*T}+C_1^*V_1R_{12}V_2C_2^{*T}+C_2^*V_{S2}C_2^{*T}-R_{23}^TV_2C_2^{*T}\\
&- C_1^*V_1R_{13}+C_2^*V_2R_{23})_j = (C_1^*V_{S1}C_1^{*T}+C_2^*V_{S2}C_2^{*T}+2(C_2^*V_2R_{12}^TV_1C_1^{*T}))_j\\
&= \mathbf{x}_j\mathbf{x}_j^T144(1-\sigma_{dep-})^2[(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\left\{\frac{\tau(1-\tau)}{f_{Y^1}^2}\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}ds - f_{Y^1}\right]^2\right.\\
&+ \left.\frac{2}{f_{Y^1}^2f_{Y^2}^2}\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}ds - f_{Y^1}\right]\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{\mathbf{y}^1,\mathbf{y}^2}ds - f_{\mathbf{y}^2}\right][F_{Y^1,Y^2}-\tau^2]\right\}\\
&+ \frac{\tau(1-\tau)}{f_{Y^2}^2}\left[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}\right]^2
\end{aligned}
$$

All of the calculations above yeld the expression of the corrected matrix of variance, for the j-th subject, below:

$$
\begin{aligned}
(V_S(\boldsymbol{\gamma}_\tau))_j &= \frac{[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2\{4F_{Y^1,\mathbf{y}^2}[1-\sigma_{dep-}-\mathbf{H}+(1-\sigma_{dep-})^2-\tau^2+2\sigma_{dep-}\tau^2}{\mathbf{x}_j\mathbf{x}_j^T\{-6(1-\sigma_{dep-})[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2+(1-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)[2F_{Y^1,Y^2}+1-2\tau-\sigma_{dep-}-\mathbf{H}]\}}\\
&+ \frac{[2\sigma_{dep-}+2\tau^2-2]\mathbf{H}+\mathbf{H}^2}{\mathbf{x}_j\mathbf{x}_j^T\{-6(1-\sigma_{dep-})[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau-(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2+(1-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)[2F_{Y^1,Y^2}+1-2\tau-\sigma_{dep-}-\mathbf{H}]\}}
\end{aligned}
$$

$$+ \quad \frac{\tau(1-\tau)([2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - f_{Y^1}]^2/f_{Y^1} + [2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}]^2/f_{Y^2})}{\mathbf{x}_j\mathbf{x}_j^T\{-6(1-\sigma_{dep-})[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2 + (1-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)[2F_{Y^1,Y^2}+1-2\tau-\sigma_{dep-}-\mathbf{H}]\}}$$

$$+ \quad \frac{[2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^2} f_{Y^1,Y^2}dt - f_{Y^1}][2\int_{-\infty}^{\mathbf{x}_j^T\boldsymbol{\beta}_\tau^1} f_{Y^1,Y^2}ds - f_{Y^2}][F_{Y^1,Y^2}-\tau^2]2/(f_{Y^1}f_{Y^2})\}}{\mathbf{x}_j\mathbf{x}_j^T\{-6(1-\sigma_{dep-})[\mathbf{x}_j^T\boldsymbol{\gamma}_\tau - (\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2]^2 + (1-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)[2F_{Y^1,Y^2}+1-2\tau-\sigma_{dep-}-\mathbf{H}]\}}$$

as in 5.14, $\mathbf{H} = (1-\sigma_{dep-})(\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)^2(3-2\mathbf{x}_j^T\boldsymbol{\gamma}_\tau)$.

5.4 Computational methods

To fit fit the models of probability of concordance, we need to use iterative processes. The maximum likelihood maximization for logistic quantile constrained regression, as well as the non least square estimation in the polynomial constrained regression, cannot be expressed in closed form. In the following we refer to Davidson and McKinnon (2003) [9] and Greene (2011) [15].

For the constrained logistic regression model in section 5.1 a Newton-Raphson algorithm is used. Such a method is based on asecond order approximation of the objective function to be minimized (if a maximization is needed instead one could consider the function with negative sign). At each iteration the objective function of interest is approximated by a quadratic function, and a step towards the maximum/minimum of that quadratic function is taken. Let $S(\boldsymbol{\gamma})$ be the function to be minimized, where $\boldsymbol{\gamma}$ is a q-vector and $S(\boldsymbol{\gamma})$ is twice differentiable. Given any initial value of $\boldsymbol{\gamma}$, say $\boldsymbol{\gamma}_0$, we can perform a second-order Taylor expansion of $S(\boldsymbol{\gamma})$ about the starting point $\boldsymbol{\gamma}_0$ in order to obtain an approximation $S^*(\boldsymbol{\gamma})$ of $S(\boldsymbol{\gamma})$

$$S^*(\boldsymbol{\gamma}) = S(\boldsymbol{\gamma}_0) + g_0^T(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) + \frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T H_0(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)$$

where $g(\boldsymbol{\gamma})$ is the gradient of $S(\boldsymbol{\gamma})$, which is a column vector of length q with entries $\frac{\partial S(\gamma)}{\partial \gamma_j}$. We denote by $H(\boldsymbol{\gamma})$ the Hessian matrix of $S(\boldsymbol{\gamma})$, which has the dimension $q \times q$ and is composed of elements of the form $\frac{\partial^2 S(\gamma)}{\partial\gamma_j\partial\gamma_\ell}$. By g_0 and H_0 we denote the gradient and the Hessian evaluated at the starting point $\boldsymbol{\gamma}_0$. Differentiating $S^*(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ we get the first order condition of minimization

$$g_0 + H_0(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0) = 0$$

which suggests the iterative procedure

$$\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0 - \alpha H_0^{-1} g_0$$

where α is a step size parameter chosen at each iteration to improve convergence.

For the minimization in the non linear regression method, presented in section 5.2, we prefer to use a first order method such as the gradient search or descent. This method works for simple differentiable objective functions. Given an initial vector $\boldsymbol{\gamma}_0$ and $S(\boldsymbol{\gamma})$ a differentiable objective function to minimize, this algorithm iteratively moves toward lower values of the function by taking steps in the direction of the negative gradient, which is the direction of steepest descent. Let us consider a first order Taylor approximation of $S(\boldsymbol{\gamma})$ about the starting point $\boldsymbol{\gamma}_0$

$$S^*(\boldsymbol{\gamma}) = S(\boldsymbol{\gamma}_0) + g_0^T(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)$$

where g_0 denotes the gradient in the form already introduced for Newton's method evaluated at the starting point. Differentiating with respect to $\boldsymbol{\gamma}$ we obtain that the iterative steps are of the form

$$\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0 - \delta g_0$$

where δ is the step length of moving and is allowed to change at every iteration.

One important feature of iterative algorithms is that they must start with an initial starting point for the parameter $\boldsymbol{\gamma}$ to estimate. A wrong choice of the starting value may significantely affect the performances of the algorithm. In some cases it could cause non convergence of the minimization iterations.

5.5 Simulation studies

We present a series of simulation studies to evaluate finite sample prediction properties of the two proposed estimators, under a variety of different scenarios. A selection of the found results is presented in this section. For the remaining results, which will lead to equivalent conclusions, we refer to appendix A of this thesis.

For each simulation setting we generate bivariate datasets, with different degrees of association structures, and estimate the probability of concordance of quantile regressions residuals along all the distribution of the bivariate outcome. The two models of logistic constrained and polynomial constrained regressions are considered and compared as to their performances. We performed and analyzed univariate quantile residuals for a grid of quantiles $\tau = \{0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90\}$. Different sample sizes were considered ($N = 200$, $N = 500$, and $N = 1000$). Based on $B = 2000$ Monte Carlo replications, we evaluated the absolute bias, standard error, and mean squared error of $\hat{\sigma}_{\mathbf{x}}$ in comparison to the true probability $\sigma_{\mathbf{x}}$. This last can be obtained from theoretical univariate and bivariate distribution probabilities.

In order to facilitate the theoretical derivation and comparison of probabilities, in all simulation scenarios a single binary covariate, generated as $x \sim Binom(1, 0.5)$, is used. This produces two estimates for each simulation corresponding to the two groups with $x = 0$ and $x = 1$. The covariate used is the same in the univariate quantile regression models and for the estimation of conditional probabilities.

The two probabilities $\sigma_{x=0}$ and $\sigma_{x=1}$ are calculated considering joint probability distributions and survival distributions of the bivariate generated datasets. We used the fact that the probability of concordance of the sign of residuals can be expressed as a sum of joint probabilities. This, jointly to the application of some basic results on bivariate distributions, resulted in

$$\sigma_{\mathbf{X}} = P(\mathbf{y}^1 \leq \mathbf{X}\boldsymbol{\beta}_\tau^1 \wedge \mathbf{y}^2 \leq \mathbf{X}\boldsymbol{\beta}_\tau^2) + P(\mathbf{y}^1 > \mathbf{X}\boldsymbol{\beta}_\tau^1 \wedge \mathbf{y}^2 > \mathbf{X}\boldsymbol{\beta}_\tau^2) = \\ 2F_{\mathbf{y}^1,\mathbf{y}^2}(\mathbf{X}\boldsymbol{\beta}_\tau^1, \mathbf{X}\boldsymbol{\beta}_\tau^2) + 1 - F_{\mathbf{y}^1}(\mathbf{X}\boldsymbol{\beta}_\tau^1) - F_{\mathbf{y}^2}(\mathbf{X}\boldsymbol{\beta}_\tau^2).$$

All simulations were implemented with the aid of the statistical software `R`. For univariate quantile regression we used the `rq` function in the library `quantreg`. As to the estimates of concordance probability parameters, we used the `nlm` routine to maximize the loglikelihood in the logistic constrained regression and implemented a gradient search algorithm to minimize the objective function in polynomial constrained regression. The programs can be found in appendix B.

Simulation 1. For the first study we generated bivariate samples from a normal distribution whose vector of the means was (a_0+a_1x, b_0+b_1x) and matrix of variance covariance

$$\begin{pmatrix} 1 & 0.7 \\ 0.7 & 2 \end{pmatrix}.$$

Note that the covariance, which is the parameter measuring the strength of association, was set to the positive quite high value of 0.7.

Simulation 2. For the second study we generated samples of a bivariate eteroskedastic normal variable with mean (a_0+a_1x, b_0+b_1x) and covariance matrix

$$\begin{pmatrix} 1+\frac{1}{2}x & -0.5+\frac{1}{3}x \\ -0.5+\frac{1}{3}x & 2-x \end{pmatrix}.$$

For this setup we chose a negative structure of association between the sample vectors $\mathbf{y}^1$ and $\mathbf{y}^2$. The covariance was set to -0.5 for units with $x = 0$ and to $-0.5+\frac{1}{3}$ when $x = 1$.

Simulation 3. The last study involves copula distribution functions discussed in 3.5. We generated bivariate samples from a bivariate random variable $(Y^1, Y^2) \sim C^{Frank}(Y^1, Y^2)$, where C^{Frank} indicates the Frank copula. This function is in the class of Archimedean associative copulas and has dependence parameter defined on the entire real line $(-\infty, \infty)$ (see [30]). Its dependence parameter θ was set to 7.92 which corresponds to a positive kendall's tau, $\rho_\tau(Y^1, Y^2) = 0.6$. We considered as marginal distributions for the two variables a normal distribution with characteristics $Y^1 \sim N(a_0 + a_1x, 1, 1)$ and a logistic distribution with parameters $Y^2 \sim Logistic(b_0 + b_1x, 2)$. We used the `copula` library in `R` to generate the chosen bivariate distribution.

In all simulations we set $a_0 = a_1 = 2$ and $b_0 = b_1 = 3$.

The performances of the proposed estimators proved to be consistent with the parameters of prediction considered, in the same way for constrained logistic and polynomial models, in all sample sizes for all quantiles considered. None of the settings used for the generation of samples showed disadvantages over the others.

For simplicity, in the tables we denote by LCR the constrained logistic method and by PCR the polynomial regression method.

Simulation 1, Method:PCR, N=200									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=0}$	0.8534	0.7606	0.7015	0.6660	0.6555	0.6682	0.7030	0.7638	0.8542
std.err.	0.0257	0.0359	0.0429	0.0462	0.0472	0.0464	0.0435	0.0369	0.0246
bias	-0.0107	-0.0125	-0.0105	-0.0104	-0.0093	-0.0083	-0.0090	-0.0094	-0.0100
mse	0.0008	0.0014	0.0019	0.0022	0.0023	0.0022	0.0020	0.0014	0.0007
$\sigma_{x=1}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=1}$	0.8552	0.7648	0.7009	0.6668	0.6553	0.6644	0.7020	0.7610	0.8534
std.err.	0.0255	0.0373	0.0430	0.0448	0.0471	0.0459	0.0427	0.0365	0.0250
bias	-0.0090	-0.0083	-0.0111	-0.0096	-0.0095	-0.0121	-0.0100	-0.0122	-0.0108
mse	0.0007	0.0015	0.0020	0.0021	0.0023	0.0023	0.0019	0.0015	0.0007

Table 5.1: Estimated standard error, bias and mean squared error of probability predictions, relative to the first smulation setting: binormal omoskedastic distribution of (Y^1, Y^2), size 200, constrained polynomial regression method.

Simulation 1, Method:PCR, N=500									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=0}$	0.8598	0.7683	0.7072	0.6721	0.6612	0.6730	0.7087	0.7708	0.8604
std.err.	0.0159	0.0235	0.0276	0.0299	0.0298	0.0290	0.0271	0.0232	0.0160
bias	-0.0043	-0.0048	-0.0048	-0.0044	-0.0036	-0.0035	-0.0033	-0.0024	-0.0038
mse	0.0003	0.0006	0.0008	0.0009	0.0009	0.0009	0.0007	0.0005	0.0003
$\sigma_{x=1}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=1}$	0.8609	0.7707	0.7078	0.6723	0.6603	0.6726	0.7069	0.7681	0.8596
std.err.	0.0160	0.0227	0.0271	0.0288	0.0300	0.0291	0.0270	0.0233	0.0161
bias	-0.0033	-0.0025	-0.0042	-0.0042	-0.0045	-0.0039	-0.0051	-0.0050	-0.0046
mse	0.0003	0.0005	0.0008	0.0008	0.0009	0.0009	0.0008	0.0006	0.0003

Table 5.2: Estimated standard error, bias and mean squared error of probability predictions, relative to the first smulation setting: binormal omoskedastic distribution of (Y^1, Y^2), size 500, constrained polynomial regression method.

Simulation 1, Method:PCR, N=1000									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=0}$	0.8619	0.7717	0.7103	0.6745	0.6621	0.6744	0.7101	0.7705	0.8616
std.err.	0.0113	0.0165	0.0195	0.0207	0.0211	0.0207	0.0193	0.0158	0.0112
bias	-0.0022	-0.0027	-0.0015	-0.0014	-0.0017	-0.0024	-0.0017	-0.0017	-0.0020
mse	0.0001	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0001
$\sigma_{x=1}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=1}$	0.8609	0.7707	0.7078	0.6723	0.6603	0.6726	0.7069	0.7681	0.8596
std.err.	0.0115	0.0160	0.0194	0.0213	0.0213	0.0206	0.0194	0.0157	0.0116
bias	-0.0023	-0.0015	-0.0017	-0.0020	-0.0027	-0.0021	-0.0019	-0.0026	-0.0026
mse	0.0001	0.0003	0.0004	0.0005	0.0005	0.0004	0.0004	0.0003	0.0001

Table 5.3: Estimated standard error, bias and mean squared error of probability predictions, relative to the first smulation setting: binormal omoskedastic distribution of (Y^1, Y^2), size 1000, constrained polynomial regression method.

Simulation 1, Method:LCR, N=200									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=0}$	0.8534	0.7606	0.7016	0.6661	0.6556	0.6681	0.7030	0.7638	0.8542
std.err.	0.0257	0.0359	0.0430	0.0462	0.0473	0.0464	0.0435	0.0371	0.0247
bias	-0.0107	-0.0125	-0.0104	-0.0104	-0.0093	-0.0084	-0.0090	-0.0094	-0.0100
mse	0.0008	0.0014	0.0020	0.0022	0.0023	0.0022	0.0020	0.0015	0.0007
$\sigma_{x=1}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=1}$	0.8553	0.7649	0.7009	0.6669	0.6553	0.6644	0.7020	0.7610	0.8534
std.err.	0.0255	0.0372	0.0429	0.0447	0.0471	0.0459	0.0427	0.0365	0.0250
bias	-0.0089	-0.0083	-0.0111	-0.0096	-0.0095	-0.0121	-0.0100	-0.0122	-0.0108
mse	0.0007	0.0014	0.0020	0.0021	0.0023	0.0023	0.0019	0.0015	0.0007

Table 5.4: Estimated standard error, bias and mean squared error of probability predictions, relative to the first smulation setting: binormal omoskedastic distribution of (Y^1, Y^2), size 200, constrained logistic regression method.

Simulation 1, Method:LCR, N=500

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=0}$	0.8598	0.7683	0.7072	0.6721	0.6612	0.6730	0.7087	0.7708	0.8604
std.err.	0.0159	0.0235	0.0276	0.0299	0.0298	0.0290	0.0271	0.0232	0.0160
bias	-0.0043	-0.0048	-0.0048	-0.0044	-0.0036	-0.0035	-0.0033	-0.0024	-0.0038
mse	0.0003	0.0006	0.0008	0.0009	0.0009	0.0009	0.0007	0.0005	0.0003
$\sigma_{x=1}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=1}$	0.8609	0.7707	0.7078	0.6723	0.6603	0.6726	0.7069	0.7681	0.8596
std.err.	0.0160	0.0227	0.0271	0.0288	0.0300	0.0291	0.0270	0.0233	0.0161
bias	-0.0033	-0.0025	-0.0042	-0.0042	-0.0045	-0.0039	-0.0051	-0.0050	-0.0046
mse	0.0003	0.0005	0.0008	0.0008	0.0009	0.0009	0.0008	0.0006	0.0003

Table 5.5: Estimated standard error, bias and mean squared error of probability predictions, relative to the first smulation setting: binormal omoskedastic distribution of (Y^1, Y^2), size 500, constrained logistic regression method.

Simulation 1, Method:LCR, N=1000

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=0}$	0.8619	0.7705	0.7105	0.6751	0.6631	0.6741	0.7103	0.7714	0.8622
std.err.	0.0113	0.0165	0.0195	0.0207	0.0211	0.0207	0.0193	0.0158	0.0112
bias	-0.0022	-0.0027	-0.0015	-0.0014	-0.0017	-0.0024	-0.0017	-0.0017	-0.0020
mse	0.0001	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003	0.0001
$\sigma_{x=1}$	0.8642	0.7732	0.7120	0.6765	0.6648	0.6765	0.7120	0.7732	0.8642
$\hat{\sigma}_{x=1}$	0.8619	0.7717	0.7103	0.6745	0.6621	0.6744	0.7101	0.7705	0.8616
std.err.	0.0115	0.0160	0.0194	0.0213	0.0213	0.0206	0.0194	0.0157	0.0116
bias	-0.0023	-0.0015	-0.0017	-0.0020	-0.0027	-0.0021	-0.0019	-0.0026	-0.0026
mse	0.0001	0.0003	0.0004	0.0005	0.0005	0.0004	0.0004	0.0003	0.0001

Table 5.6: Estimated standard error, bias and mean squared error of probability predictions, relative to the first smulation setting: binormal omoskedastic distribution of (Y^1, Y^2), size 1000, constrained logistic regression method.

Chapter 6

Application and interpretation of results

In this chapter we report the results relative to the application of the proposed method to the analysis of a selection of 1996 subjects from the Po River Delta study.

We introduce graphicals tool developed to illustrate and interpret the outputs deriving from the application of our method. Furthermore, we suggest the use of a statistical test to verify independence of conditional residuals signs.

The discussed example can be useful to understand the importance of evaluating the residual correlation among outcomes after removal of the effect of covariates.

The Po River Delta study is a prospective study conducted to investigate obstructive pulmonary diseases on the general population of a rural area in northern Italy (near Venice) [31].

Different spirometric indexes are usually evaluated to undertake lung function impairement (we refer to the website of the Global Lung Function Initiative for a complete review [16]). In the present application we consider the effect of a set of variables on two of the most common parameters measured in spirometry.

The first one is Forced Expiratory Volume (FEV1), the volume exhaled during the first second after a full inspiration. The second one is the ratio $\frac{FEV1}{FVC}$ where (FVC) indicates the Forced Vital Capacity, a measure of the volume change of the lung between a full inspiration to total lung capacity and a maximal expiration to residual volume. The ratio measures the proportion of a person vital capacity that they are able to expire in the first second of expiration.

Two covariates are considered: height (in centimeters), measured in standing position without shoes, and age, recorded at last birthday. It is well known from medical literature ([16], [37]) that lung function measurements vary with age, standing height, sex and ethnicity. The effect of height and age on the two lung functions were therefore analyzed separately for men and women: for the same height and age, females tend to have smaller lungs than males.

The dataset consists of 1996 subjects divided into 956 males (48%) and 1040 females. All subjects are adults (age range 25-64).

Some preliminary statistical analysis was performed to describe behaviours and relationships among the variables considered. Figure 6 suggests that there is a difference in the association between FEV1 and the ratio for males and females. FEV1 grows along with FEV1/FVC, but this association seems stronger for males than for females. This is confirmed by correlation which is positive for both sexes but higher in males (cor= 0.65 for males, cor= 0.45 for females). The existing correlation between the two dependent variables considered requires the use of multivariate methods to evaluate if this dependence has an impact on the effects of age and height.

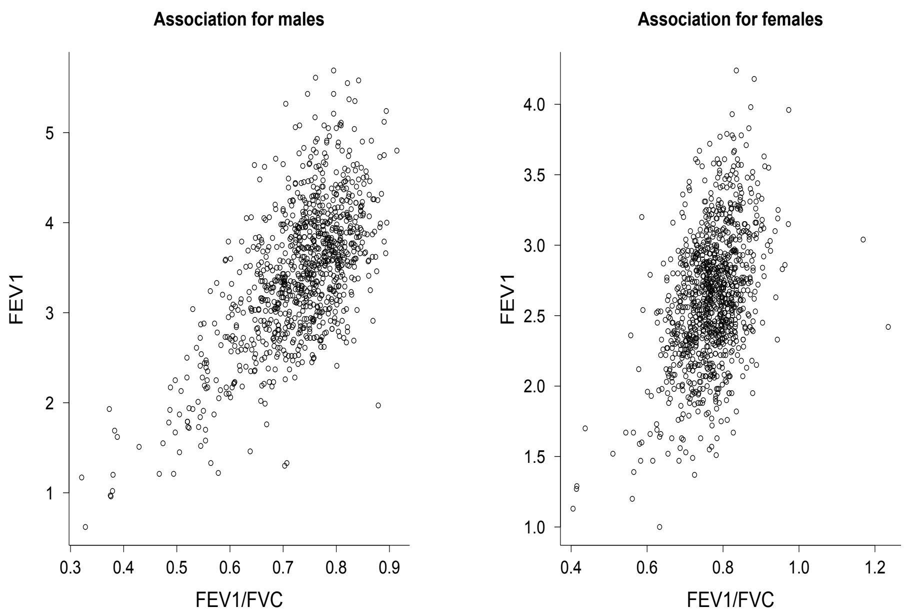

Figure 6.1: Scatterplot showing association between the ratio FEV1/FVC and FEV1, for males and females.

Pulmonary disfunctions are diagnosed by considering the extreme tails of the distribution of measured parameters in a "normal" population of reference, see [37]. Usually subjects with spirometrical parameters below the 5th percentile are considered respiratory deficient. Analysing and interpreting percentiles of their distribution rather than the simple mean might then be more informative in the determination of factors related to respiratory impairement.

As described in chapter 4, as a first step we estimate a quantile regression model for either outcome. We do it for a grid of 11 quantiles

$$\tau = \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95\}$$

and the two models estimated are the following

$$FEV1 = \beta^1_{\tau,0} + \beta^1_{\tau,1} age + \beta^1_{\tau,2} height$$

$$FEV_1/FVC = \beta^2_{\tau,0} + \beta^2_{\tau,1} age + \beta^2_{\tau,2} height.$$

The *rq* function of the *quantreg* package in R was used to estimate regression parameters. We extracted regression residuals in both models and used them to define the indicator variable of concordance Z introduced in chapter 4. We used the same set of covariates, either to fit univariate quantile regressions or to model the probability of concordance ($\sigma_{\mathbf{X}} = P(Z = 1)$).

The effects of height and age on $\sigma_{\mathbf{X}}$ are evaluated by applying either constrained logistic or constrained polynomial regression. The results derived from the two are then discussed and compared.

For the logistic constrained regression the fitted model is

$$logit^* \left(\frac{\sigma_{(age,height)} - \sigma_{dep-}}{\sigma_{dep+} - \sigma_{dep-}} \right) = \gamma_{\tau,0} + \gamma_{\tau,1} age + \gamma_{\tau,2} height.$$

For the nonlinear regression approach we estimate the model

$$\sigma_{(age,height)} = (1 - \sigma_{dep-})(\gamma_{\tau,0} + \gamma_{\tau,1} age + \gamma_{\tau,2} height)^2 [3 - 2(\gamma_{\tau,0} + \gamma_{\tau,1} age + \gamma_{\tau,2} height)] + \sigma_{dep-}.$$

Table 6.1: Parameters estimates of constrained logistic regression model for females, outcomes were FEV1 and FEV1/FVC. 200 Bootstrap tilted replications are used to estimate standard errors.

	Intercept	Age	Height
$\tau = 0.05$	4.365 (-12.499, 21.229)	0.045 (-2.953, 3.044)	-0.045 (-11.42, 11.33)
$\tau = 0.10$	-0.345 (-10.577, 9.888)	0.051 (-0.561, 0.663)	-0.017 (-2.35, 2.316)
$\tau = 0.20$	-0.156 (-6.519, 6.207)	0.027* (0.004, 0.051)	-0.009 (-0.049, 0.032)
$\tau = 0.30$	0.046 (-4.781, 4.874)	0.013 (-0.004, 0.029)	-0.004 (-0.034, 0.025)
$\tau = 0.40$	0.292 (-3.238, 3.822)	0.005 (-0.009, 0.019)	-0.002 (-0.024, 0.020)
$\tau = 0.50$	0.526 (-2.730, 3.781)	0.000 (-0.013, 0.013)	-0.000 (-0.020, 0.019)
$\tau = 0.60$	0.142 (-4.056, 4.340)	-0.001 (-0.014, 0.012)	-0.000 (-0.026, 0.026)
$\tau = 0.70$	-3.379 (-8.727, 1.968)	0.005 (-0.013, 0.023)	0.019 (-0.015, 0.052)
$\tau = 0.80$	-5.645 (-14.656, 3.365)	-0.004 (-0.028, 0.020)	0.032 (-0.023, 0.086)
$\tau = 0.90$	-0.760 (-21.156, 19.635)	0.003 (-3.039, 3.044)	-0.004 (-11.072, 11.064)
$\tau = 0.95$	-1.003 (-14.998, 12.993)	-4.566* (-8.667, -0.465)	-16.766* (-31.460, -2.073)

* Significantly different from zero at the 5 per cent level.

Table 6.2: Parameters estimates of constrained logistic regression model for males, outcomes were FEV1 and FEV1/FVC. 200 Bootstrap exponentially tilted replications are used to estimate standard errors.

	Intercept	Age	Height
$\tau = 0.05$	0.198 (-47.731, 48.128)	0.021 (-0.622, 0.665)	-0.009 (-2.477, 2.460)
$\tau = 0.10$	0.273 (-11.065, 11.612)	0.039* (0.007, 0.071)	-0.011 (-0.075, 0.053)
$\tau = 0.20$	0.369 (-6.151, 6.890)	0.044* (0.016, 0.073)	-0.012 (-0.049, 0.026)
$\tau = 0.30$	0.625 (-4.537, 5.786)	0.036* (0.018, 0.053)	-0.009 (-0.039, 0.020)
$\tau = 0.40$	0.645 (-3.379, 4.670)	0.025* (0.009, 0.041)	-0.007 (-0.029, 0.016)
$\tau = 0.50$	0.750 (-2.327, 3.827)	0.012 (-0.002, 0.025)	-0.003 (-0.021, 0.014)
$\tau = 0.60$	4.875 (-0.131, 9.882)	0.002 (-0.014, 0.019)	-0.028 (-0.055, -0.000)
$\tau = 0.70$	0.121 (-4.470, 4.712)	0.014 (-0.004, 0.032)	-0.004 (-0.030, 0.022)
$\tau = 0.80$	-3.798 (-10.633, 3.037)	0.024 (-0.003, 0.052)	0.013 (-0.024, 0.050)
$\tau = 0.90$	-0.806 (-61.564, 59.951)	0.019 (-1.574, 1.612)	-0.008 (-6.680, 6.664)
$\tau = 0.95$	-0.985 (-1773.722, 1771.752)	-4.169 (-22.983, 14.645)	-16.869 (-34.332, 0.594)

* Significantly different from zero at the 5 per cent level.

The interpretation of regression coefficients is beyond our scope, the interest of the analysis being more on the interpretation of predicted probabilities. However, we briefly comment on them. We are aware that, contrarily to the logistic transformed method, which preserves interpretability in terms of odd ratios, we cannot provide a meaningful interpretation of coefficients for the nonlinear alternative.

Coefficient estimates of the two approaches, together with their confidence intervals, are reported in separate tables for males and females (Table 6.1, Table 6.2, Table 6.3 and Table 6.4). For the calculations of standard errors we used 200 exponentially tilted bootstrap replicates ([10], [12]). In fact, although having derived asymptotic approximations of corrected standard errors (see section 5.3), their analytical expression is too complicated and involves the calculation of bivariate unknowns *cdf*s. We therefore prefered to use sampling techniques for the inference.

The results, in terms of statistical significance of estimates, are quite similar for both methods considered. We found significant estimates only in correspondence to few quantiles, and almost always relatively to the effect of age.

For the logistic approach, when $\tau = \{0.10, 0.20, 0.30, 0.40\}$ the estimates of the coefficients associated with age for males are statistically significant, at a level of 95%, showing an increasing effect of concordance between the residuals of FEV1 and the ratio FEV1/FVC along with age. The effect of height is not significant in any of the quantiles considered. A similar consideration can be made for females. Here the only significant estimates are in correspondence to the effect of age in the tails

Table 6.3: Parameters estimates of constrained polynomial regression model for females, outcomes were FEV1 and FEV1/FVC. 200 Bootstrap exponentially tilted replications are used to estimate standard errors.

	Intercept	Age	Height
$\tau = 0.05$	0.907 (-6.463, 8.277)	0.008 (-0.013, 0.029)	-0.006 (-0.056, 0.045)
$\tau = 0.10$	0.258 (-1.84, 2.356)	0.009* (0.003, 0.015)	-0.002 (-0.015, 0.011)
$\tau = 0.20$	0.510 (-0.785, 1.804)	0.005* (0.0003, 0.009)	-0.002 (-0.01, 0.006)
$\tau = 0.30$	0.755 (-0.177, 1.688)	0.002 (-0.001, 0.005)	-0.002 (-0.008, 0.003)
$\tau = 0.40$	0.346 (-0.399, 1.092)	0.001 (-0.001, 0.003)	0.001 (-0.004, 0.005)
$\tau = 0.50$	0.376 (-0.246, 0.999)	0.0002 (-0.002, 0.003)	0.001 (-0.003, 0.005)
$\tau = 0.60$	0.241 (-0.450, 0.932)	0.0001 (-0.002, 0.003)	0.002 (-0.003, 0.006)
$\tau = 0.70$	-0.074 (-0.913, 0.766)	0.001 (-0.002, 0.004)	0.003 (-0.002, 0.008)
$\tau = 0.80$	-0.406 (-1.595, 0.783)	-0.001 (-0.005, 0.003)	0.005 (-0.002, 0.012)
$\tau = 0.90$	-1.849 (-5.041, 1.342)	0.004 (-0.003, 0.011)	0.012 (-0.006, 0.031)
$\tau = 0.95$	-0.491 (-7.756, 6.774)	0.002 (-0.019, 0.024)	0.004 (-0.037, 0.044)

* Significantly different from zero at the 5 per cent level.

of the distribution ($\tau = \{0.20, 0.95\}$). The results about the nonlinear regression coefficients are almost identical, except for two small differences. The effect of age is significant also at the 10th percentile for females, estimates when $\tau = 0.95$ are not significant.

We remark that the estimates are less accurate for extreme quantiles. Confidence intervals are larger. In particular, we observe that, in correspondence to the 95th percentile for the constrained logistic approach the estimates of the effects of age and height are unusually large, besides the extreme largeness of confidence intervals. We believe that, in these two situations, there was a failure in the maximization of the likelihood due to the observation of true probabilities outside of the boundaries. Therefore, as higlighted in section 5.1, estimates at this quantiles cannot be considered valid.

We have already observed that the aim of the modelization is the analysis of changes in probabilities along with the effect of covariates. Therefore, it is important to provide tools that can be of support for the interpretation and visualization of such changes, if there are any.

The residual dependence on outcomes after removing the effect of the covariates can be graphically described. To do so, after the estimation of the models, we construct a figure in which the predicted probability of concordance is plotted against covariates and directly compared with values of the largest possible dependence or independence. We report the graphics for a selection of five quantiles out of the eleven analyzed, for the two sex separated, and for both methods of probability modelization.

Table 6.4: Parameters estimates of constrained polynomial regression model for males, outcomes were FEV1 and FEV1/FVC. 200 Bootstrap exponentially tilted replications are used to estimate standard errors.

	Intercept	Age	Height
$\tau = 0.05$	0.379 (-5.971, 6.729)	0.005 (-0.007, 0.017)	-0.001 (-0.037, 0.035)
$\tau = 0.10$	0.399 (-1.505, 2.302)	0.008* (0.002, 0.015)	-0.001 (-0.012, 0.009)
$\tau = 0.20$	0.445 (-0.838, 1.729)	0.008* (0.003, 0.012)	-0.001 (-0.009, 0.006)
$\tau = 0.30$	0.176 (-0.750, 1.103)	0.007* (0.003, 0.010)	0.001 (-0.004 ,0.006)
$\tau = 0.40$	0.593 (-0.401, 1.587)	0.004* (0.001, 0.007)	-0.001 (-0.006, 0.005)
$\tau = 0.50$	0.542 (-0.168, 1.253)	0.002 (-0.000, 0.004)	-0.00005 (-0.004, 0.004)
$\tau = 0.60$	1.383 (0.591, 2.174)	0.0003 (-0.002, 0.003)	-0.005* (-0.009, -0.001)
$\tau = 0.70$	0.526 (-0.44, 1.492)	0.002 (-0.001, 0.005)	-0.001 (-0.006, 0.005)
$\tau = 0.80$	-0.105 (-1.254, 1.045)	0.004 (-0.001, 0.008)	0.002 (0.004, 0.008)
$\tau = 0.90$	-0.444 (-4.328, 3.441)	0.004 (-0.009, 0.017)	0.003 (-0.016, 0.023)
$\tau = 0.95$	0.036 (-9.711, 9.783)	0.001 (-0.026, 0.029)	0.0004 (-0.050, 0.051)

* Significantly different from zero at the 5 per cent level.

We observe that the predictions are almost the same for both methods, showing that there is coherence in the estimation of probabilities. When all the assumption of the models are respected both models could be applied indisctinctly.

From figures 6.2, 6.4, 6.6, 6.8, 6.10, 6.12, 6.14, 6.16 we can see that, both for age and height, either for males and females, the association structure between residuals is in general of positive dependence. We can say that the strenght of positive association is higher for elderly patients, closer to 1 which is the value of maximum positive dependence, both in males and females at the tails of the distribution; whereas there is no effect of age on the probability from the median above.

When looking at height we observe a slightly different situation. The concordance decreases with height in the low tails, going from values close to 1 almost until the threshold of independence. There is no change in the concordance for the median, and we observe an increase with height when looking at the superior part of the distribution.

The case of $\tau = 0.95$ must be analyzed separately. In fact, we observe different predictions of probabilities for the constrained logistic and the constrained polynomial regressions. In the first one (see figures 6.2, 6.4, 6.6, 6.8) the probability predicted is constant and equal to the minimal value attainable (σ_{dep-}) for all the combinations (males and females, height and age). On the other hand, for the second one (see figures 6.10, 6.12, 6.14) we observe, almost always, a probability which is slighlty increasing in all the covariates for females, and for age in males.

The curve representing the probability of concordance passes through the line of independence, going from the negative dependence region toward the region of positive dependence. In the case of the effect of height for males there is a difference. The probability predicted is constant, falls in the upper region of positive dependence, and is really close to the independence border (figure 6.16).

The situation observed for the modified approach logistic, derives from the failure in the likelihood maximization discussed when commenting about coefficient estimates. The assumptions of the model are not satisfied; true conditional probabilities fall outside the range $(\sigma_{dep-}, \sigma_{dep+})$. In particular, we have that the true probability is smaller than the minimum attainable. This induces high negative estimates of the coefficients that result in a constant prediction of probability ($\hat{\sigma}_{age,height} = \sigma_{dep-}$).

To verify if there effectively is a change in the dependence structure between the signs of the residuals, relying on the normal distribution of predicted values, we performe Z-tests for all the different covariate patterns in the sample. We test the differences between predicted probabilities of the model against the independence value σ_{indep}. We consider a significance level of 95% and plot the values of the Z statistic against the covariates. We then provide a graphical tool similar to the one for the illustration of predictions. To simplify the interpretation we add to the graphs horizontal lines in correspondence to 1.96 and -1.96 which delimit the area representing 95% of the normal distribution. Values falling outside this area are considered, with a confidence of 95%, different from the value of independence. Furthemore, values above 1.96 correspond to covariate patterns in which a positive dependence is detected and values below -1.96 to covariate patterns corresponding to negative dependence. The results of Z-tests are reported in figures 6.3, 6.5, 6.7, 6.9, 6.11, 6.13, 6.15, 6.17. We observe that in general the structure of positive dependence is confirmed except in few occasions. For $\tau = 0.05$ the predictions, for females in polynomial regression along the covariate patterns in the dataset, are between the situations of positive dependence and independence; whereas for logistic constrained regression the situation of independence at this quantile was more pronounced. There is a variation even at the 70th percentile: in the group of females we have a confirmation of positive dependence, with sporadic observations falling in the region of non-rejection, along age. Along height there is independence in correspondence to low values and positive dependence for the highest ones in both methods tested.

In particular we observe that, for all the scenarios considered, at the 95th quantile there is always independence when the polynomial constrained regression is applied, whereas the results of the test are not available for the logistic modified model, which we have shown to be not valid at this level of the distribution.

Finally, in the extreme tails of the distribution of FEV1 and $\frac{FEV1}{FVC}$, corresponding to the threshold values used to diagnose respiratory disfunctions, the application of the method in the thesis showed that there is no change in the dependence structure caused by age and height after having removed their univariate effect on quantiles.

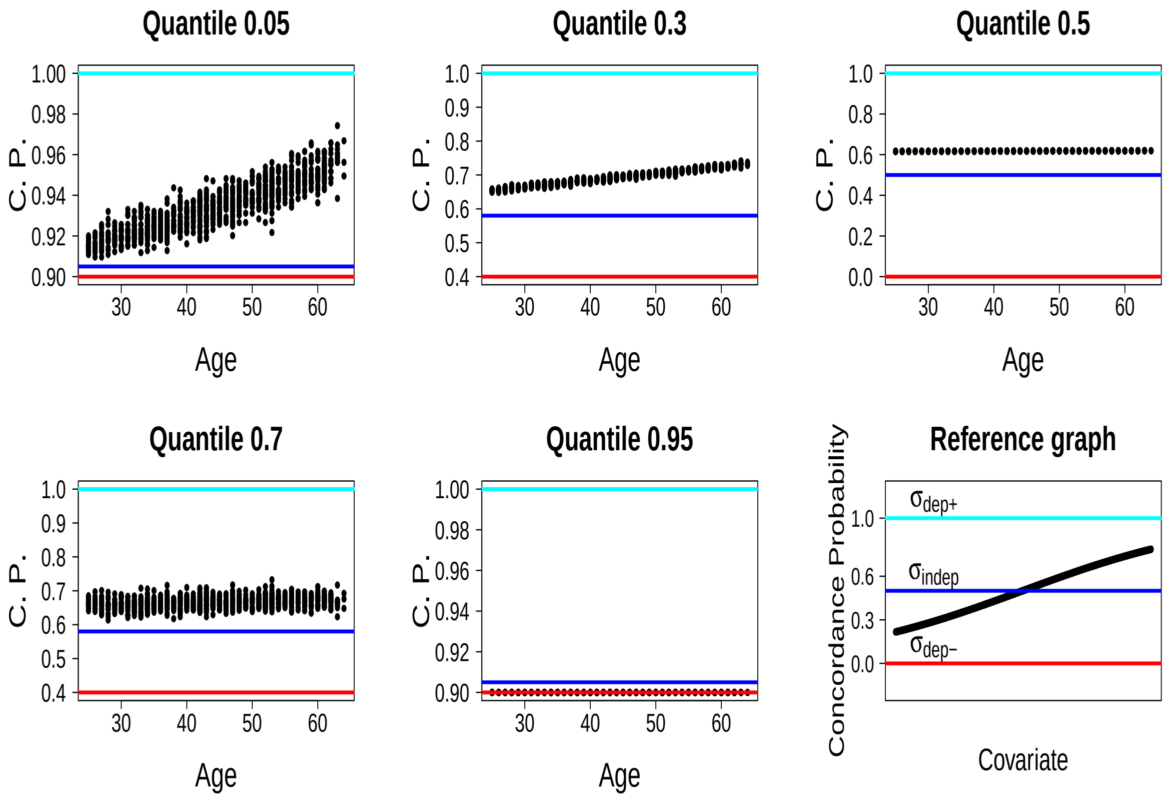

Figure 6.2: Probability of concordance predicted by constrained logistic regression along age, for females, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

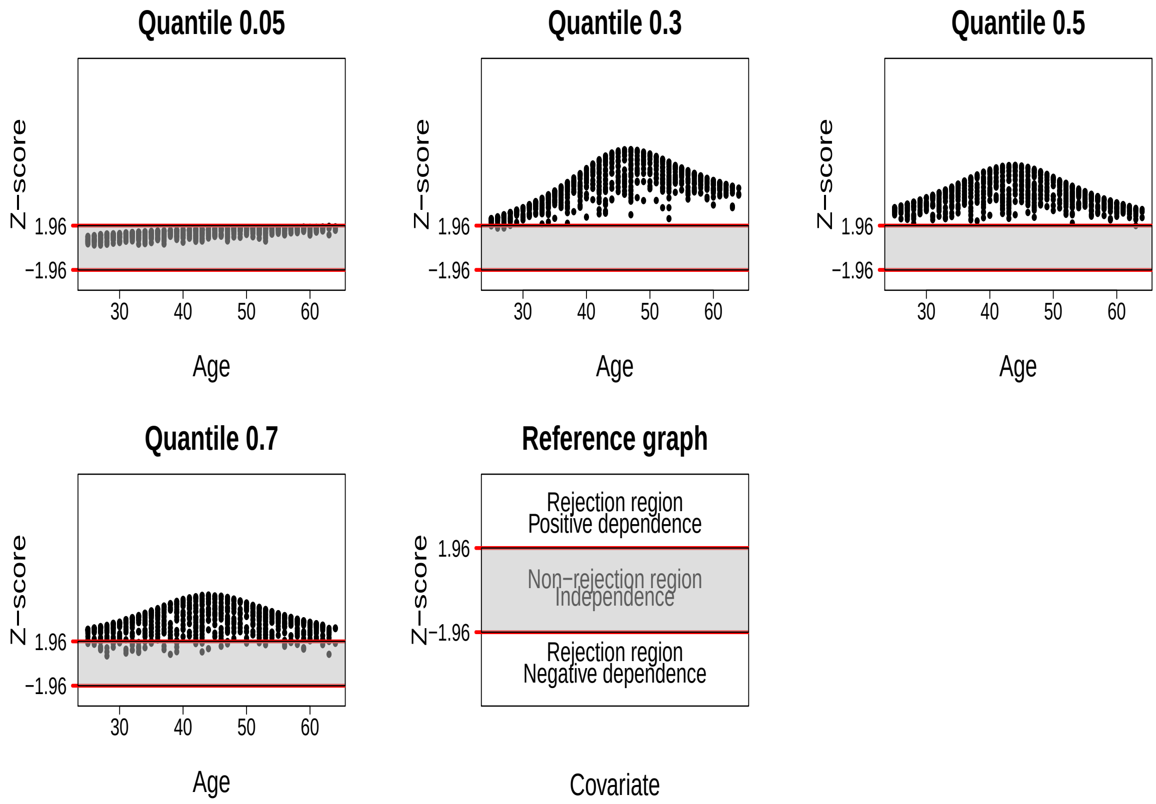

Figure 6.3: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained logistic regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against age for females, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%.

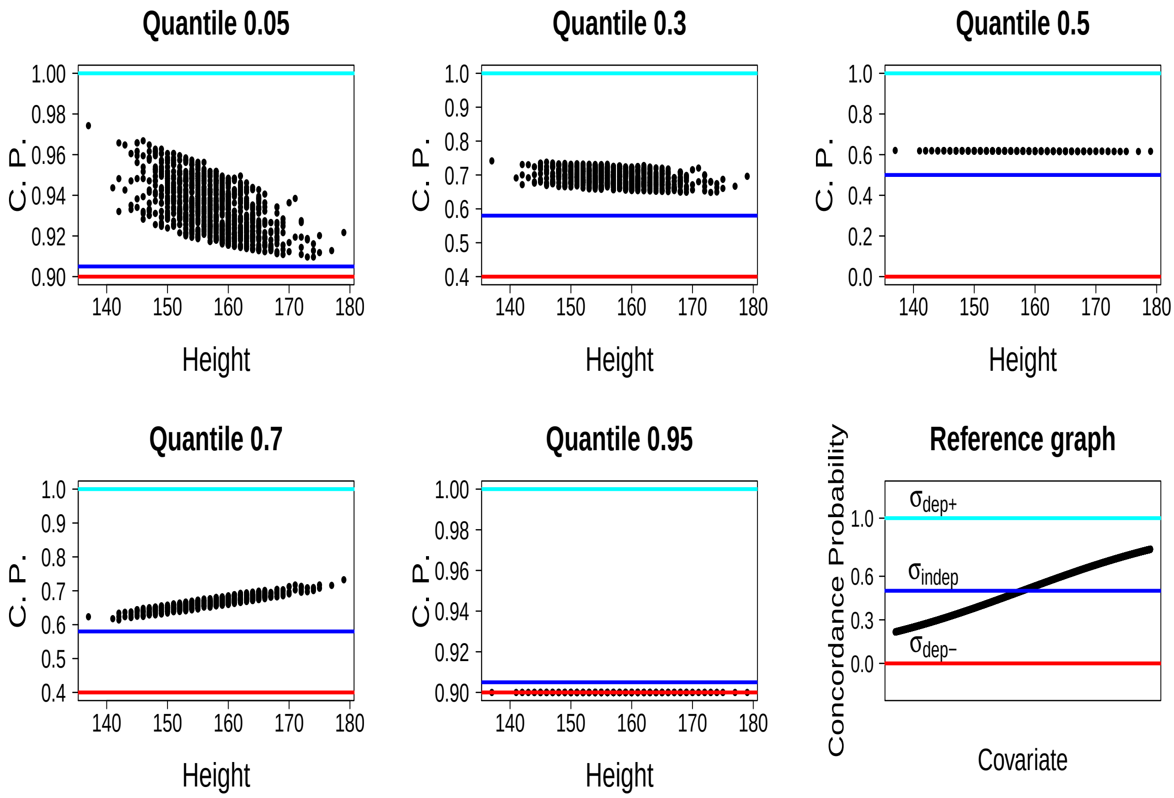

Figure 6.4: Probability of concordance predicted by constrained logistic regression along height, for females, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

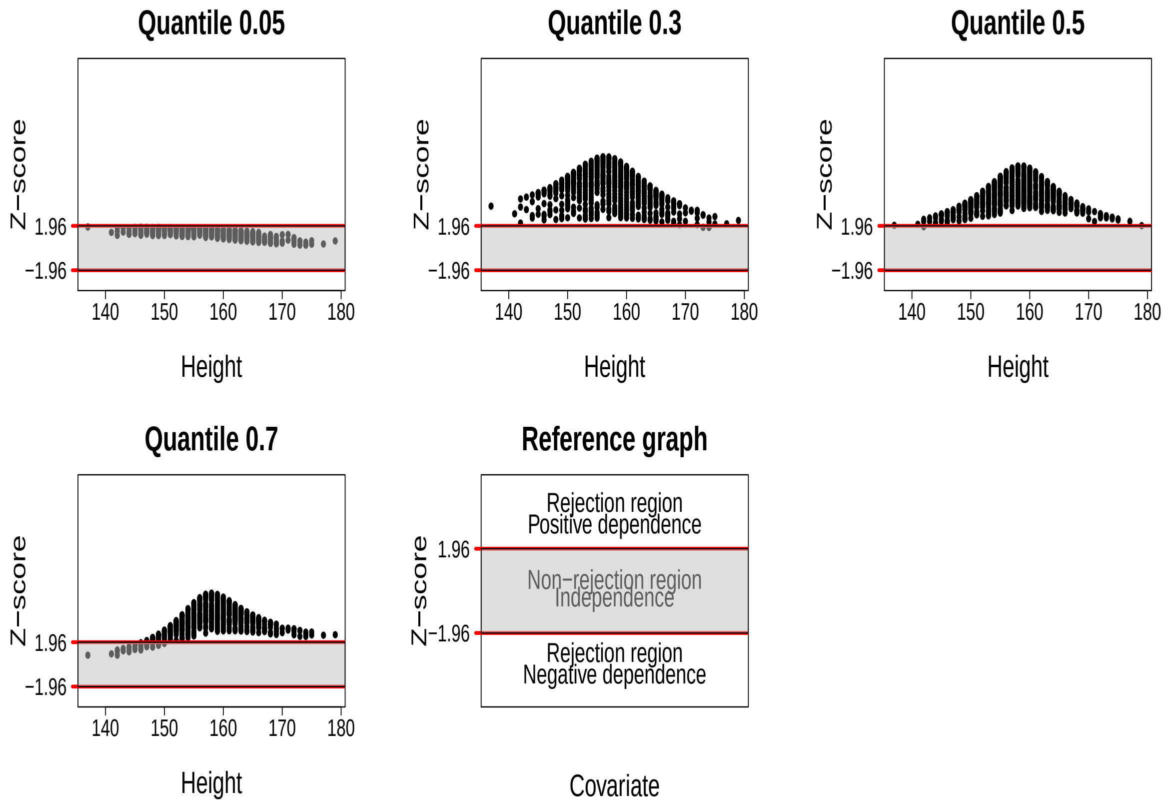

Figure 6.5: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained logistic regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against height for females, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%.

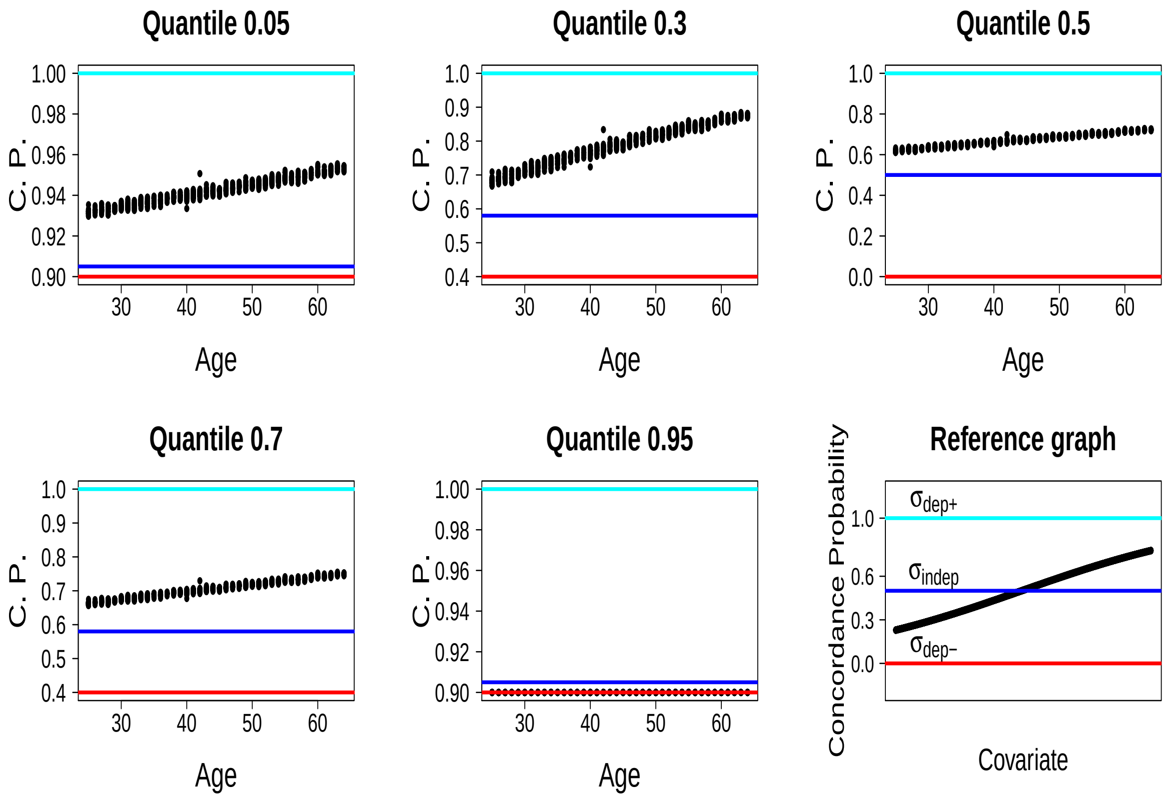

Figure 6.6: Probability of concordance predicted by constrained logistic regression along age, for males, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

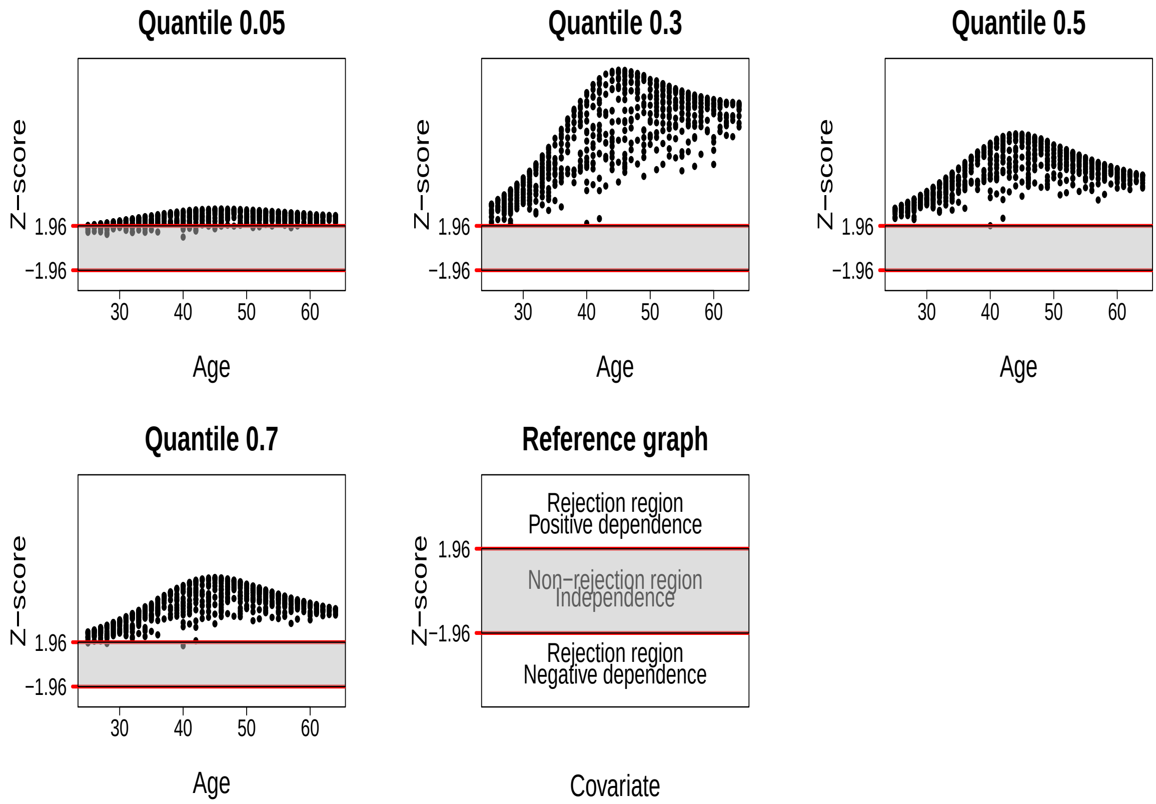

Figure 6.7: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained logistic regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against age for males, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%.

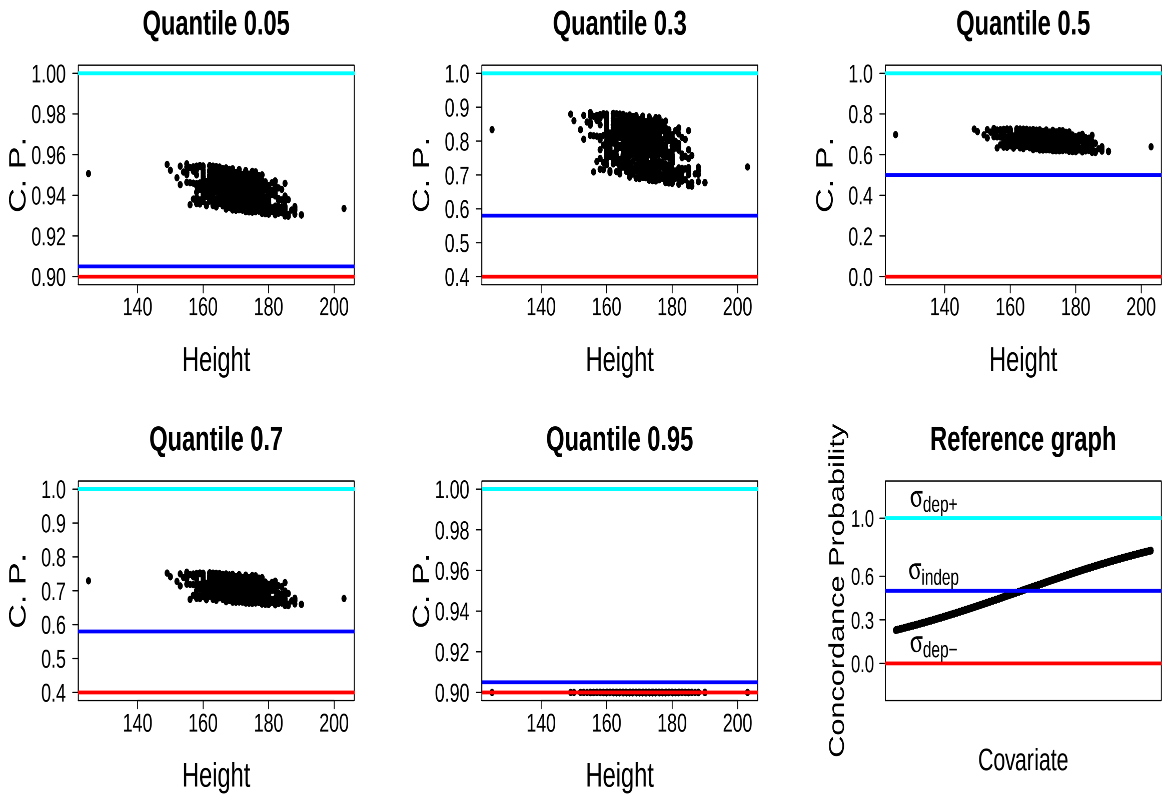

Figure 6.8: Probability of concordance predicted by constrained logistic regression along height, for males, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

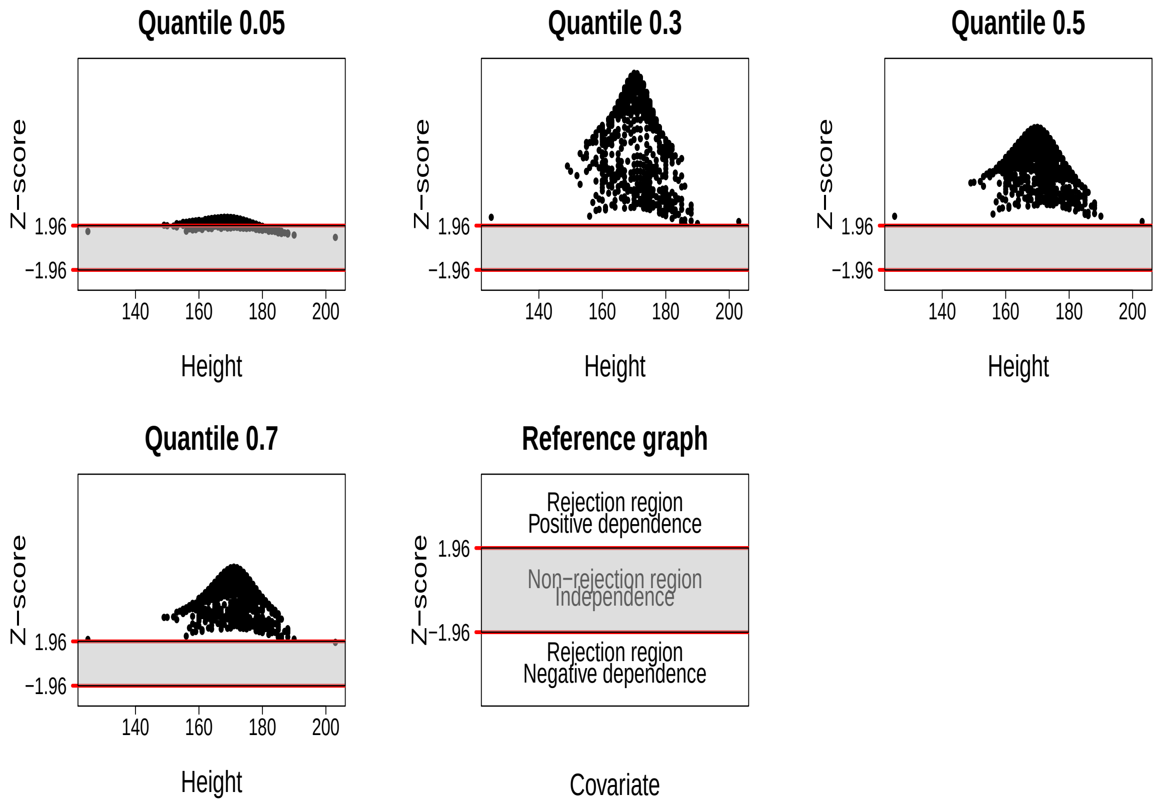

Figure 6.9: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained logistic regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against height for males, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%.

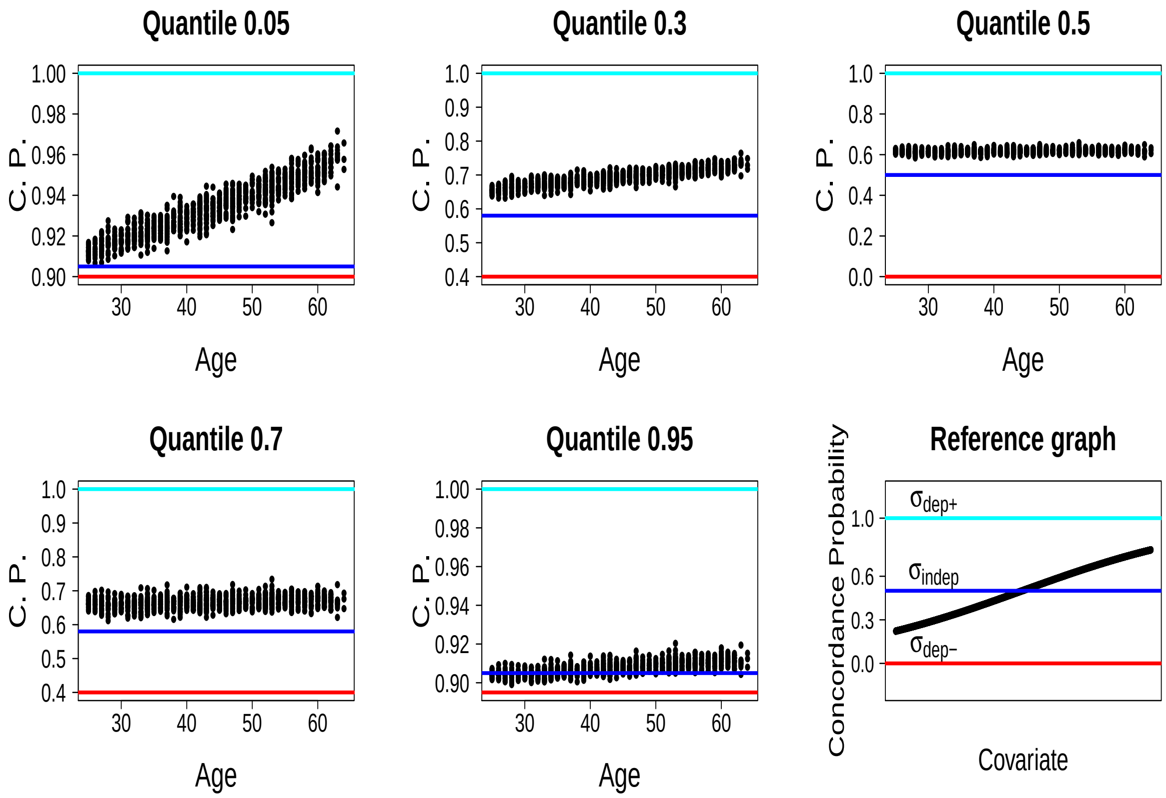

Figure 6.10: Probability of concordance predicted by constrained polynomial regression along age, for females, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

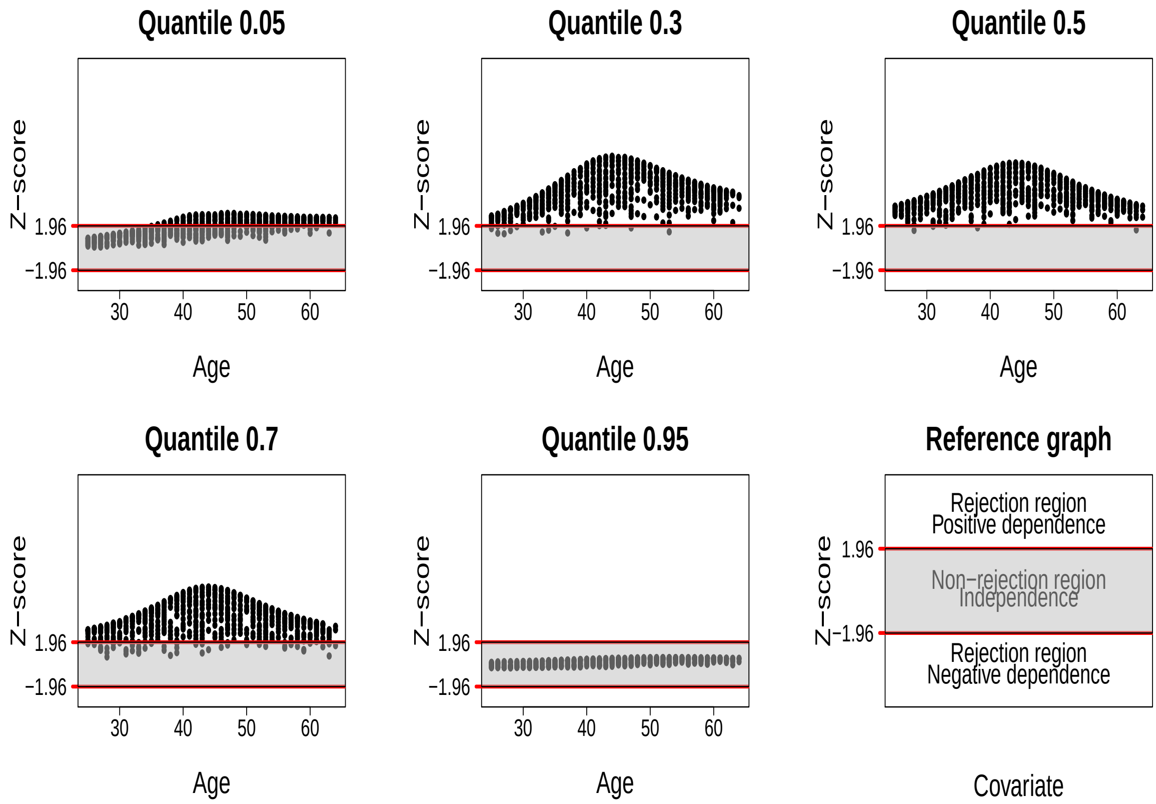

Figure 6.11: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained polynomial regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against age for females, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%

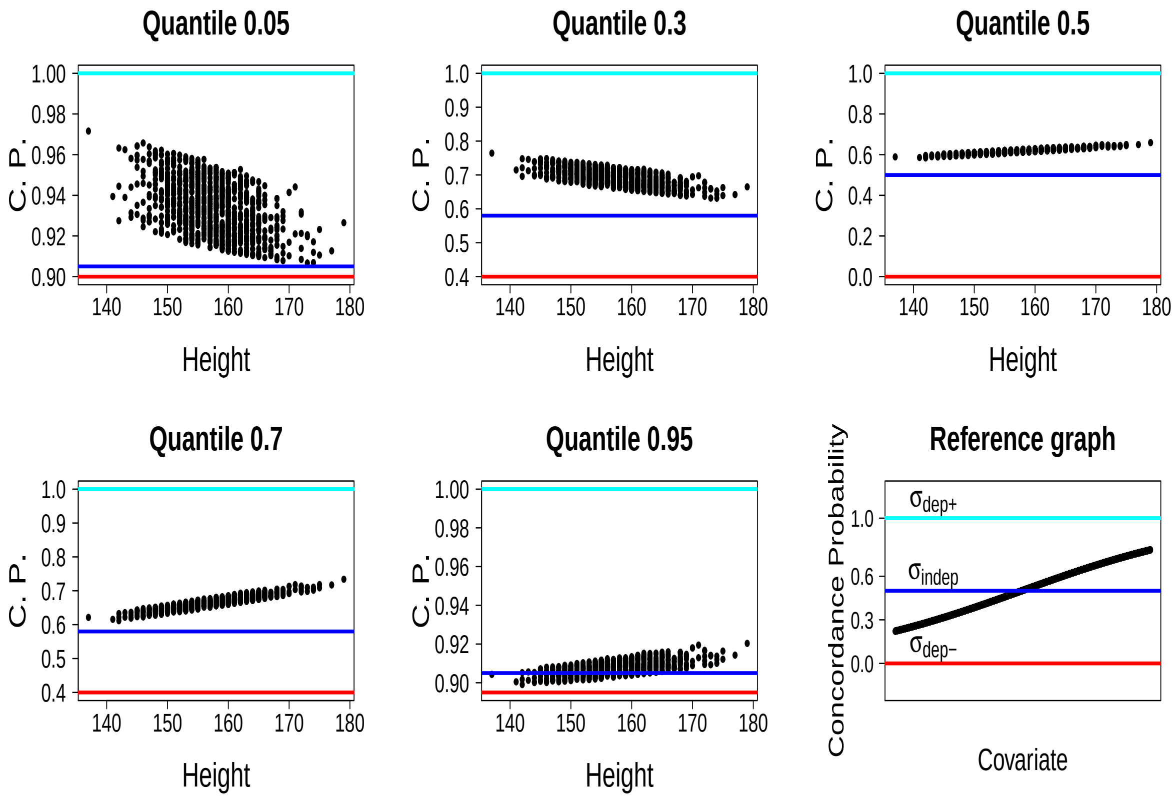

Figure 6.12: Probability of concordance predicted by constrained polynomial regression along height, for females, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

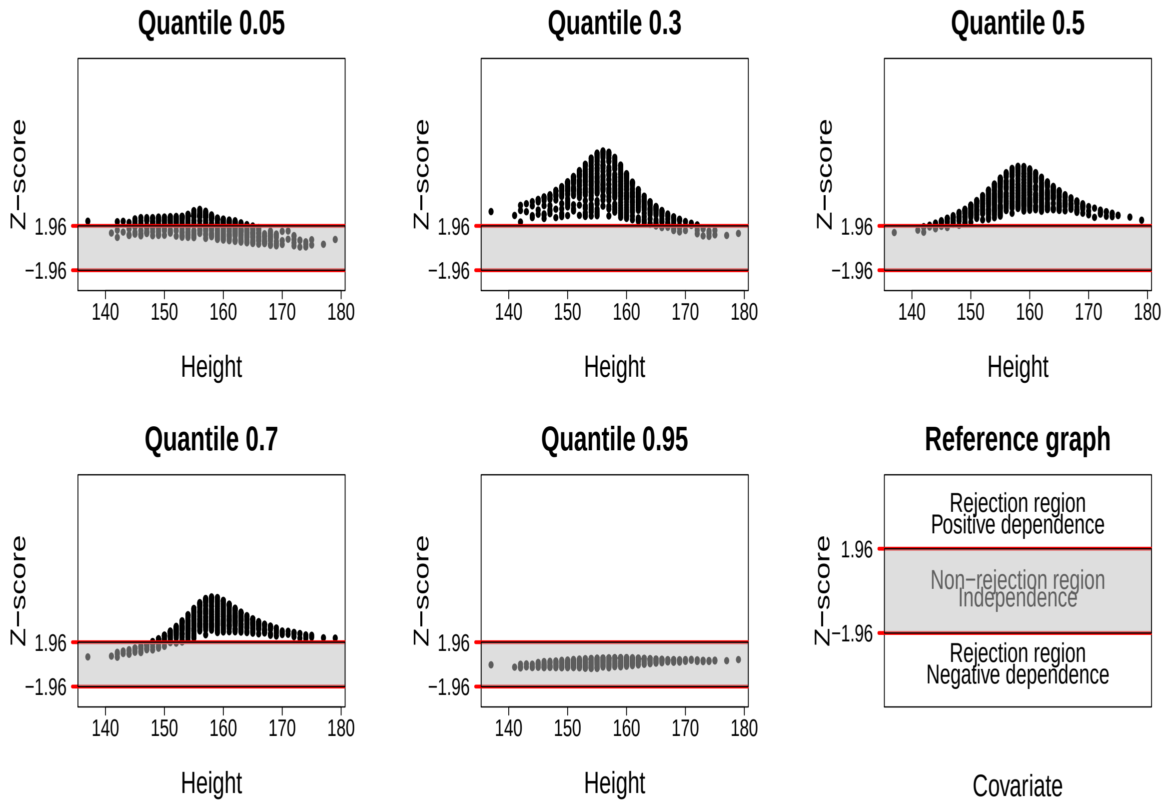

Figure 6.13: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained polynomial regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against height for females, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%.

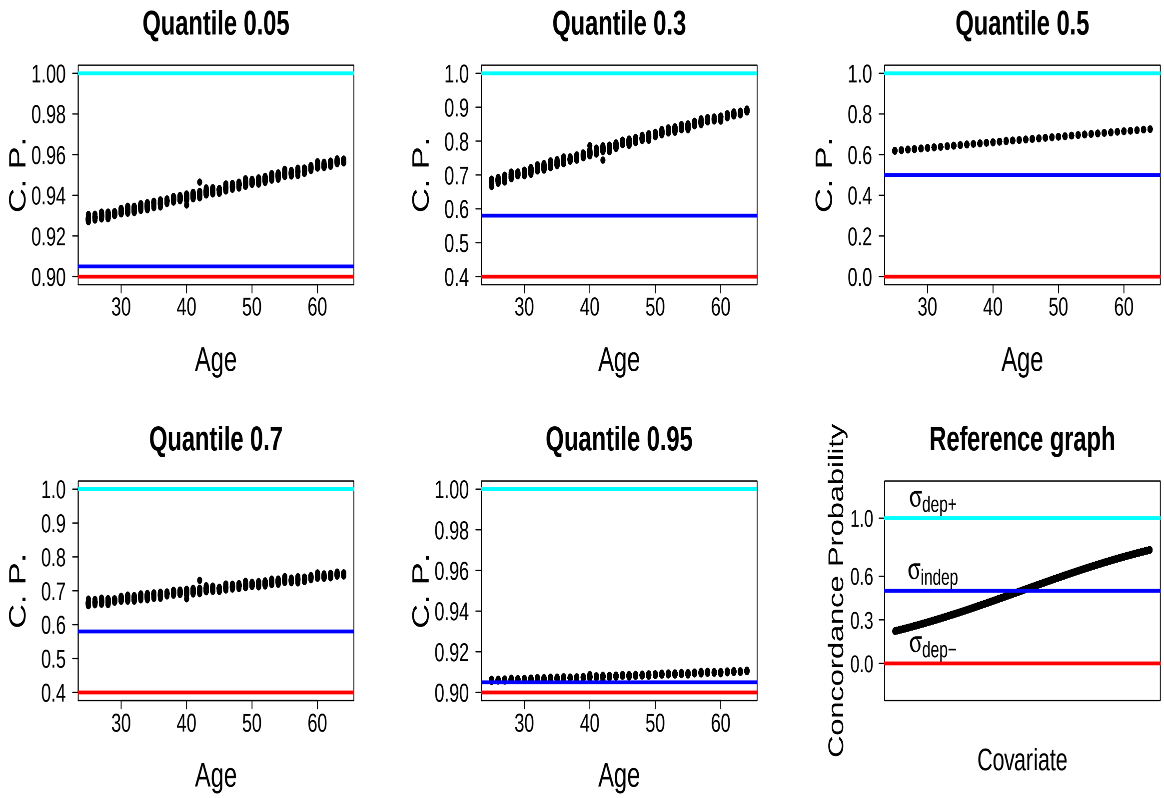

Figure 6.14: Probability of concordance predicted by constrained polynomial regression along age, for males, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

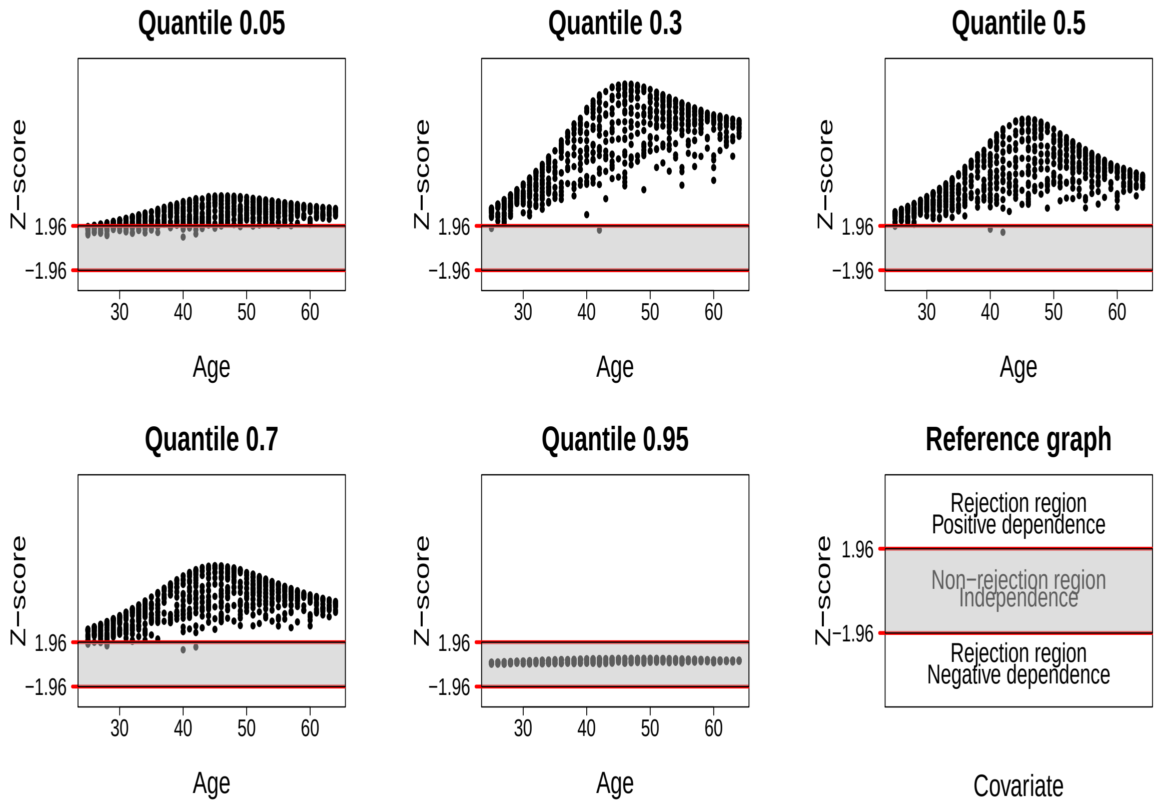

Figure 6.15: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained polynomial regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against age for males, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%.

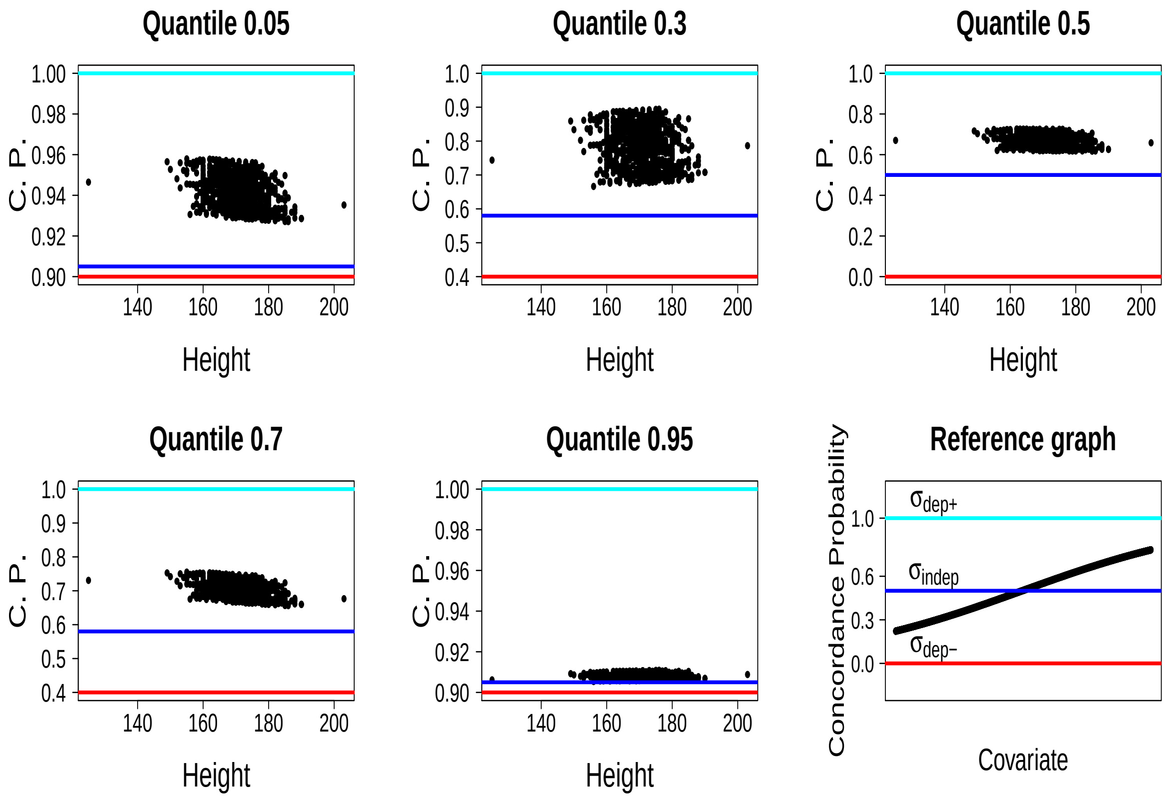

Figure 6.16: Probability of concordance predicted by constrained polynomial regression along height, for males, in correspondence to 5 selected quantiles. Prediction values are compared with limit values of $\hat{\sigma}$ statistic, as shown in the reference panel.

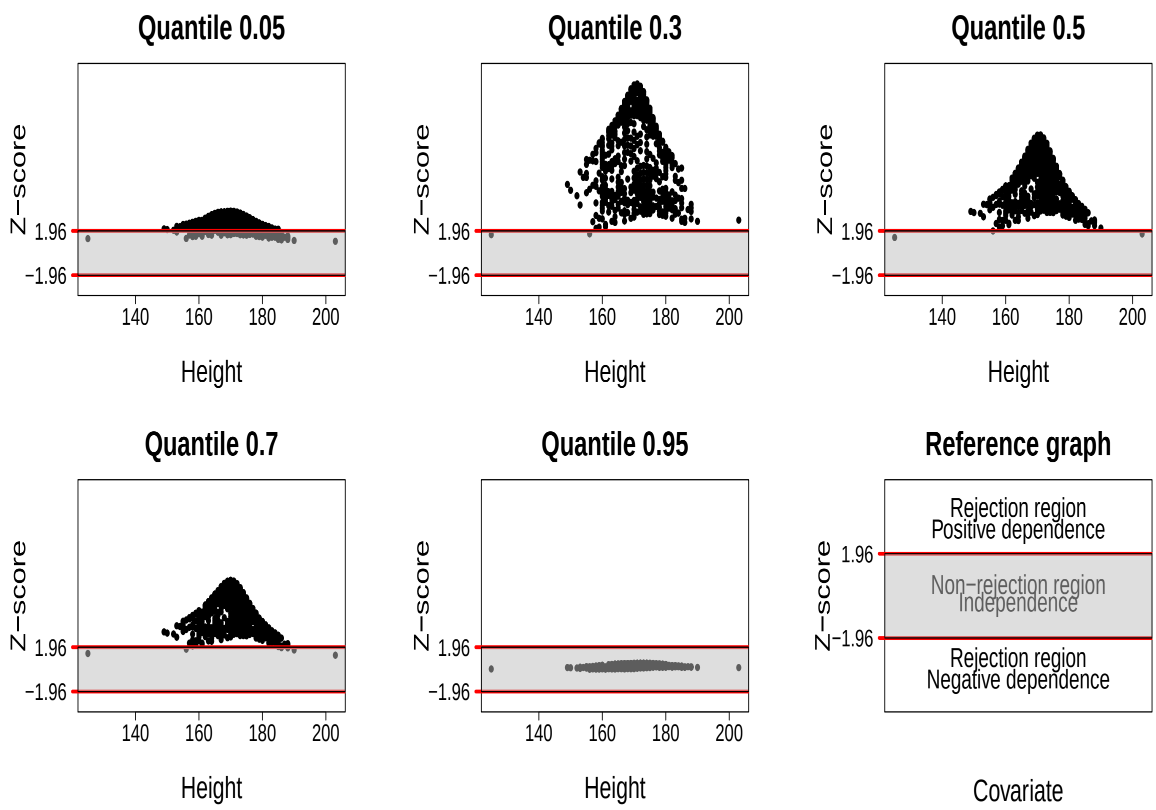

Figure 6.17: Z-score values resulting from testing the differences between prediction probabilities, relative to constrained polynomial regression, and σ_{indep} for all the covariate patterns in the dataset. The scores are plotted here against height for males, for 5 selected quantiles. A reference graph, illustrating the interpretation of figures, is also reported. Differences are tested at a level of 95%.

Chapter 7

Discussion and Future Work

7.1 Summary and Conclusions

The method presented in the thesis supplies an alternative solution to the analysis of multivariate structures in quantile regression frameworks. It preserves the dependency on quantiles, without defining a notion of multivariate quantile.

Although not being an inferential regression instrument, our method can be really powerful in real applications. It can be extensively used and understood by researchers and allows an intuitive interpretation typical of univariate quantiles.

The crucial point of the method is the interpretation of multivariate structures in terms of analyses of residual regressions. The avalaibility of appropriate regression instruments to study conditional probabilities in a restricted range of definition is an additional feature of the technique.

For what concerns modelization of probabilities, users could choose their own preferred regression model. Linear regression with the introduction of splines on covariates could also be a flexible tool for probability prediction. Among the two regression methods proposed, we think that the use of a nonlinear regression method, as the one described in section 5.2, is preferable to the logistic alternative discussed in section 5.1, for prediction purposes, since it overcomes the problem related to asymptotical definition of boundaries of the parameter space.

7.2 Future work

The method presented in the thesis offers the opportunity of many extensions. We will list some of them.

- *Comparison of more than two sets of residuals per time.* When considering dimensions higher than 2, instead of making all the comparisons in pairs, define an index that handles multidimensional comonoticities and countermonotonicities.

- *Consider distinct quantiles in each pair.* Comparing residuals of univariate quantile regressions coming from taking different quantiles in each regression, would give the opportunity to non-parametrically reconstruct the entire bivariate distribution of residuals.

- *Application to non-quantile settings.* The method could be applied to measure concordances in any situation in which compared data are divided in two sets based on different criteria; not only quantile repartition of the distribution.

- *Avalaibility of software instruments.* Last but not least, we aim at implementing and releasing a package, either on R either on Stata, to give the users the opportunity to apply the method without being forced to make the programming theirselves.

Bibliography

[1] Barrodale, I., and Roberts, F. Solution of an overdetermined system of equations in the ℓ_1 norm. *Communications of the ACM 17* (1974), 319–320.

[2] Boos, D. D., and Stefansky, L. A. *Essential Statistical Inference. Theory and Methods.* Springer, 2013.

[3] Boscovich, R. J. De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiensi Scientiarum et Artum Instituto Atque Academia Commentarii*, 4 (1757), 353–396.

[4] Bottai, M., Cai, B., and McKeown, R. E. Logistic quantile regression for bounded outcomes. *Statistics in Medicine*, 29 (2010), 309–317.

[5] Chakraborty, B. On multivariate quantile regression. *Journal of Statistical Planning and Inference 110* (2003), 109–132.

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

[7] Columbu, S., and Bottai, M. Logistic quantile regression to model cognitive impairment in sardinian cancer patients. In *Selected papers of the 47th Scientific meeting of the Italian Statistical Society*, R. Moreno, Ed. Springer-Verlag Series, (2015) to appear.

[8] Daniels, H. E. Rank correlation and population models. *J. Roy. Statist. Soc. Ser B 12* (1950).

[9] Davidson, R., and McKinnon, J. G. *Econometric Theory and Methods.* Oxford University Press, USA, 2003.

[10] Di Ciccio, T. J., and Romano, J. P. A review of bootstrap confidence intervals. *J R Statist Soc B 50* (1988), 338–354.

[11] Dudley, R. M., and Koltchinskii, V. I. The spatial quantiles. *Unpublished Manuscript* (1992).

[12] Efron, B., and Tibshirani, R. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science 1* (1986), 54–77.

[13] Embrechts, P., McNeil, A., and Straumann, D. Correlation and dependency in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond, ed. M.A.H. Dempster, Cambridge University Press, Cambridge 12* (2002), 176–223.

[14] Gneiting, T. G., and Raftery, A. E. R. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*, 477 (2007), 359–378.

[15] Greene, W. H. *Econometric Analysis*, 7nd ed. Prentice Hall, 2011.

[16] H., Q. P., Stanojevic, S., Stocks, J., and Cole, T. J. Become an expert in spirometry. `http://www.spirxpert.com`, 2012.

[17] Hallin, M., Paindaveine, D., and Siman, M. Multivariate quantiles and multiple-output regression quantiles: from l1 optimization to halfspace depth. *The Annals of Statistics 110* (2010), 109–132.

[18] Hardin, J. W. The robust variance estimator for two-stage models. *Stata Journal 2* (2002), 253–266.

[19] Heinze, G., and Schemper, M. A solution to the problem of separation in logistic regression. *Statistics in Medicine 21* (2002), 2409–2419.

[20] Hillier, F. S., and Lieberman, G. J. *Introduction to Operations Research.* McGraw Hill, 2005.

[21] Huber, P. J. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Berkley, CA, 1967), U. of California Press, Ed., vol. 1, pp. 221–233.

[22] Koenker, R. *Quantile Regression.* 2004.

[23] Koenker, R., and Bassett, G. Regression quantiles. *Econometrica*, 46 (1978), 33–50.

[24] Koenker, R., and DOrey, V. Algorithm as 229: Computing regression quantiles. *Applied Statistics 36*, 3 (1987), 383–393.

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

[26] Koltchinskii, V. I. M-estimation, convexity and quantiles. *The Annals of Statistics 25*, 2 (1997), 435–477.

[27] Kong, L., and Mizera, I. Quantile tomography: using quantiles with multivariate data. *Statistics Sinica 22* (2010), 1589–1610.

[28] Liu, X., and Zuo, Y. Computing halfspace depth and regression depth. *Communications in Statistics - Simulation and Computation 43* (2014), 969–985.

[29] McCullagh, P., and Nelder, J. A. *Generalized Linear Models*. Chapman and Hall, 1989.

[30] Nelsen, R. B. *An introduction to Copulas.* Springer New York, 1999.

[31] Pistelli, F., Bottai, M., Viegi, G., Di Pede, F., Carrozzi, L., Baldacci, S., Pedreschi, M., and Giuntini, C. Smooth reference equations for slow vital capacity and flow-volume curve indexes. *Am J Respir Crit Care Med 161* (2000), 899–905.

[32] Rencher, A. C. *Methods of Multivariate Analysis*, 2nd ed. Wiley series in Probability and Statistics, 2002.

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

[34] Serfling, R. Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *Journal of Nonparametric Statistics 22*, 7 (2010), 915–936.

[35] Sklar, A. Fonctions de rèpartition á n dimensions et leurs merges. *Publ. Inst. Paris 8* (1959).

[36] Slutsky, E. Uber stochastische asymptoten und grenzwerte. *Metron 3*, 5 (1925).

[37] Stanojevic, S., Wade, A., and Stocks, J. Reference values for lungfunction: past, present and future. *Eur Respir J 36* (2010), 12–19.

[38] Stefansky, L. A., and Boos, D. D. The calculus of m-estimation. *The American Statistician 56*, 1 (2001), 29–38.

[39] Struyf, A. J., and Rousseeuw, P. J. Halfspace depth and regression depth characterize the empirical distribution. *Journal of Multivariate Analysis 69*, 1 (1999), 135–153.

[40] Trivedi, P. K., and Zimmer, D. M. Copula modeling: an introduction for practitioners. *Foundations and trends in econometrics 1*, 1 (2005), 1–111.

[41] Wagner, W. H. Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, 54 (1959), 206–212.

[42] Walter, S. *Defining Quantiles for Functional Data; with an Application to the Reversal of Stock Price Decreases.* PhD thesis, The University of Melbourne, Department of Mathematics and Statistics.

[43] Zuo, Y., and Serfling, R. General notions of statistical depth function. *The Annals of Statistics 28*, 2 (2000), 461–482.

Appendix A

Simulation tables

Simulation 2, Method:PCR, N=200									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8044	0.6318	0.4981	0.4137	0.3850	0.4137	0.4981	0.6318	0.8044
$\hat{\sigma}_{x=0}$	0.8022	0.6225	0.4881	0.4033	0.3776	0.4051	0.4906	0.6261	0.8027
std.err.	0.0057	0.0214	0.0363	0.0452	0.0490	0.0450	0.0363	0.0218	0.0063
bias	-0.0022	-0.0093	-0.0099	-0.0104	-0.0074	-0.0086	-0.0075	-0.0057	-0.0017
mse	0.0001	0.0012	0.0031	0.0050	0.0055	0.0048	0.0029	0.0011	0.0001
$\sigma_{x=1}$	0.8125	0.6597	0.5477	0.4794	0.4565	0.4794	0.5477	0.6597	0.8125
$\hat{\sigma}_{x=1}$	0.8088	0.6527	0.5382	0.4712	0.4471	0.4680	0.5367	0.6481	0.8068
std.err.	0.0122	0.0293	0.0401	0.0468	0.0500	0.0468	0.0410	0.0287	0.0110
bias	-0.0037	-0.0070	-0.0095	-0.0083	-0.0094	-0.0114	-0.0110	-0.0116	-0.0057
mse	0.0002	0.0011	0.0024	0.0038	0.0045	0.0037	0.0026	0.0010	0.0001

Table A.1: Estimated standard error, bias and mean squared error of probability predictions, relative to the second smulation setting: binormal eteroskedastic distribution of (Y^1, Y^2), size 200, polynomial constrained regression method.

Simulation 2, Method:PCR, N=500									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8044	0.6318	0.4981	0.4137	0.3850	0.4137	0.4981	0.6318	0.8044
$\hat{\sigma}_{x=0}$	0.8025	0.6277	0.4935	0.4101	0.3813	0.4116	0.4954	0.6291	0.8030
std.err.	0.0043	0.0151	0.0229	0.0293	0.0308	0.0287	0.0229	0.0147	0.0047
bias	-0.0019	-0.0041	-0.0046	-0.0036	-0.0037	-0.0021	-0.0027	-0.0027	-0.0015
mse	0.0001	0.0008	0.0020	0.0033	0.0038	0.0031	0.0019	0.0007	0.0001
$\sigma_{x=1}$	0.8125	0.6597	0.5477	0.4794	0.4565	0.4794	0.5477	0.6597	0.8125
$\hat{\sigma}_{x=1}$	0.8100	0.6577	0.5446	0.4759	0.4519	0.4757	0.5430	0.6557	0.8091
std.err.	0.0088	0.0192	0.0253	0.0297	0.0315	0.0297	0.0257	0.0186	0.0086
bias	-0.0025	-0.0020	-0.0030	-0.0036	-0.0047	-0.0038	-0.0047	-0.0040	-0.0034
mse	0.0001	0.0007	0.0017	0.0029	0.0032	0.0028	0.0017	0.0007	0.0001

Table A.2: Estimated standard error, bias and mean squared error of probability predictions, relative to the second smulation setting: binormal eteroskedastic distribution of (Y^1, Y^2), size 500, polynomial constrained regression method.

Simulation 2, Method:PCR, N=1000									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8044	0.6318	0.4981	0.4137	0.3850	0.4137	0.4981	0.6318	0.8044
$\hat{\sigma}_{x=0}$	0.8028	0.6294	0.4965	0.4122	0.3837	0.4119	0.4961	0.6305	0.8033
std.err.	0.0035	0.0103	0.0166	0.0201	0.0214	0.0201	0.0164	0.0102	0.0038
bias	-0.0016	-0.0024	-0.0016	-0.0015	-0.0013	-0.0019	-0.0020	-0.0014	-0.0012
mse	0.0001	0.0006	0.0016	0.0026	0.0031	0.0027	0.0016	0.0005	0.0001
$\sigma_{x=1}$	0.8125	0.6597	0.5477	0.4794	0.4565	0.4794	0.5477	0.6597	0.8125
$\hat{\sigma}_{x=1}$	0.8111	0.6589	0.5464	0.4786	0.4544	0.4777	0.5460	0.6574	0.8103
std.err.	0.0065	0.0131	0.0184	0.0216	0.0224	0.0213	0.0187	0.0129	0.0064
bias	-0.0014	-0.0008	-0.0013	-0.0008	-0.0021	-0.0018	-0.0016	-0.0023	-0.0022
mse	0.0001	0.0005	0.0015	0.0025	0.0029	0.0025	0.0015	0.0005	0.0001

Table A.3: Estimated standard error, bias and mean squared error of probability predictions, relative to the second smulation setting: binormal eteroskedastic distribution of (Y^1, Y^2), size 1000, polynomial constrained regression method.

Simulation 2, Method:LCR, N=200									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8044	0.6318	0.4981	0.4137	0.3850	0.4137	0.4981	0.6318	0.8044
$\hat{\sigma}_{x=0}$	0.8022	0.6225	0.4881	0.4033	0.3776	0.4051	0.4906	0.6261	0.8027
std.err.	0.0057	0.0214	0.0363	0.0452	0.0490	0.0450	0.0363	0.0217	0.0063
bias	-0.0022	-0.0093	-0.0099	-0.0104	-0.0074	-0.0086	-0.0075	-0.0057	-0.0017
mse	0.0001	0.0012	0.0031	0.0050	0.0055	0.0048	0.0029	0.0011	0.0001
$\sigma_{x=1}$	0.8125	0.6597	0.5477	0.4794	0.4565	0.4794	0.5477	0.6597	0.8125
$\hat{\sigma}_{x=1}$	0.8085	0.6527	0.5382	0.4712	0.4471	0.4680	0.5367	0.6481	0.8065
std.err.	0.0124	0.0294	0.0401	0.0468	0.0500	0.0468	0.0410	0.0287	0.0111
bias	-0.0040	-0.0070	-0.0095	-0.0083	-0.0094	-0.0115	-0.0110	-0.0116	-0.0060
mse	0.0002	0.0011	0.0024	0.0038	0.0045	0.0037	0.0026	0.0010	0.0001

Table A.4: Estimated standard error, bias and mean squared error of probability predictions, relative to the second smulation setting: binormal eteroskedastic distribution of (Y^1, Y^2), size 200, logistic constrained regression method.

Simulation 2, Method:LCR, N=500									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8044	0.6318	0.4981	0.4137	0.3850	0.4137	0.4981	0.6318	0.8044
$\hat{\sigma}_{x=0}$	0.8025	0.6277	0.4935	0.4101	0.3813	0.4116	0.4954	0.6291	0.8030
std.err.	0.0043	0.0151	0.0229	0.0293	0.0308	0.0287	0.0229	0.0147	0.0047
bias	-0.0019	-0.0041	-0.0046	-0.0036	-0.0037	-0.0021	-0.0027	-0.0027	-0.0015
mse	0.0001	0.0008	0.0020	0.0033	0.0038	0.0031	0.0019	0.0007	0.0001
$\sigma_{x=1}$	0.8125	0.6597	0.5477	0.4794	0.4565	0.4794	0.5477	0.6597	0.8125
$\hat{\sigma}_{x=1}$	0.8098	0.6577	0.5446	0.4759	0.4519	0.4757	0.5430	0.6557	0.8089
std.err.	0.0090	0.0192	0.0253	0.0297	0.0315	0.0297	0.0257	0.0186	0.0087
bias	-0.0027	-0.0020	-0.0030	-0.0036	-0.0047	-0.0038	-0.0047	-0.0040	-0.0036
mse	0.0001	0.0007	0.0017	0.0029	0.0032	0.0028	0.0017	0.0007	0.0001

Table A.5: Estimated standard error, bias and mean squared error of probability predictions, relative to the second smulation setting: binormal eteroskedastic distribution of (Y^1, Y^2), size 500, logistic constrained regression method.

Simulation 2, Method:LCR, N=1000									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8044	0.6318	0.4981	0.4137	0.3850	0.4137	0.4981	0.6318	0.8044
$\hat{\sigma}_{x=0}$	0.8006	0.6294	0.4965	0.4122	0.3837	0.4119	0.4961	0.6305	0.8007
std.err.	0.0014	0.0103	0.0166	0.0201	0.0214	0.0201	0.0164	0.0102	0.0016
bias	-0.0038	-0.0024	-0.0016	-0.0015	-0.0013	-0.0019	-0.0020	-0.0014	-0.0037
mse	0.0001	0.0006	0.0016	0.0026	0.0031	0.0027	0.0016	0.0005	0.0001
$\sigma_{x=1}$	0.8125	0.6597	0.5477	0.4794	0.4565	0.4794	0.5477	0.6597	0.8125
$\hat{\sigma}_{x=1}$	0.8010	0.6589	0.5464	0.4786	0.4544	0.4777	0.5460	0.6574	0.8007
std.err.	0.0034	0.0131	0.0184	0.0216	0.0224	0.0213	0.0187	0.0129	0.0027
bias	-0.0115	-0.0008	-0.0013	-0.0008	-0.0021	-0.0018	-0.0016	-0.0023	-0.0118
mse	0.0001	0.0005	0.0015	0.0025	0.0029	0.0025	0.0015	0.0005	0.0001

Table A.6: Estimated standard error, bias and mean squared error of probability predictions, relative to the second smulation setting: binormal eteroskedastic distribution of (Y^1, Y^2), size 1000, logistic constrained regression method.

Simulation 3, Method:PCR, N=200									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=0}$	0.8798	0.8404	0.8275	0.8214	0.8179	0.8181	0.8255	0.8421	0.8797
std.err.	0.0261	0.0338	0.0364	0.0372	0.0375	0.0370	0.0363	0.0333	0.0255
bias	-0.0101	-0.0121	-0.0100	-0.0101	-0.0120	-0.0134	-0.0120	-0.0105	-0.0102
mse	0.0008	0.0013	0.0014	0.0015	0.0015	0.0015	0.0015	0.0012	0.0008
$\sigma_{x=1}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=1}$	0.8809	0.8423	0.8269	0.8202	0.8187	0.8193	0.8254	0.8403	0.8796
std.err.	0.0258	0.0327	0.0360	0.0373	0.0379	0.0370	0.0355	0.0341	0.0265
bias	-0.0090	-0.0103	-0.0107	-0.0113	-0.0113	-0.0123	-0.0121	-0.0122	-0.0104
mse	0.0007	0.0012	0.0014	0.0015	0.0016	0.0015	0.0014	0.0013	0.0008

Table A.7: Estimated standard error, bias and mean squared error of probability predictions, relative to the third simulation setting: Frank copula distribution of (Y^1, Y^2), size 200, polynomial constrained regression method.

Simulation 3, PCR Model, N=500									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=0}$	0.8858	0.8477	0.8336	0.8267	0.8248	0.8266	0.8331	0.8477	0.8857
std.err.	0.0163	0.0211	0.0235	0.0236	0.0238	0.0238	0.0225	0.0212	0.0165
bias	-0.0041	-0.0049	-0.0040	-0.0048	-0.0052	-0.0049	-0.0044	-0.0049	-0.0042
mse	0.0003	0.0005	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0003
$\sigma_{x=1}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=1}$	0.8872	0.8486	0.8332	0.8274	0.8252	0.8265	0.8335	0.8476	0.8856
std.err.	0.0164	0.0213	0.0237	0.0239	0.0239	0.0235	0.0232	0.0213	0.0165
bias	-0.0027	-0.0039	-0.0043	-0.0041	-0.0048	-0.0050	-0.0041	-0.0050	-0.0043
mse	0.0003	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0003

Table A.8: Estimated standard error, bias and mean squared error of probability predictions, relative to the third simulation setting: Frank copula distribution of (Y^1, Y^2), size 500, polynomial constrained regression method.

Simulation 3, PCR Model, N=1000									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=0}$	0.8875	0.8504	0.8355	0.8293	0.8270	0.8293	0.8351	0.8502	0.8880
std.err.	0.0118	0.0152	0.0156	0.0166	0.0169	0.0169	0.0160	0.0149	0.0114
bias	-0.0024	-0.0021	-0.0020	-0.0022	-0.0029	-0.0023	-0.0025	-0.0024	-0.0019
mse	0.0001	0.0002	0.0002	0.0003	0.0003	0.0003	0.0003	0.0002	0.0001
$\sigma_{x=1}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=1}$	0.8886	0.8513	0.8353	0.8288	0.8277	0.8288	0.8345	0.8507	0.8876
std.err.	0.0119	0.0148	0.0160	0.0166	0.0169	0.0168	0.0163	0.0148	0.0118
bias	-0.0013	-0.0012	-0.0022	-0.0027	-0.0022	-0.0027	-0.0031	-0.0019	-0.0023
mse	0.0001	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0001

Table A.9: Estimated standard error, bias and mean squared error of probability predictions, relative to the third simulation setting: Frank copula distribution of (Y^1, Y^2), size 1000, polynomial constrained regression method.

Simulation 3, LCR Model, N=200									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=0}$	0.8798	0.8404	0.8275	0.8214	0.8179	0.8181	0.8255	0.8421	0.8797
std.err.	0.0261	0.0338	0.0364	0.0372	0.0375	0.0370	0.0363	0.0333	0.0255
bias	-0.0101	-0.0121	-0.0100	-0.0101	-0.0120	-0.0134	-0.0120	-0.0105	-0.0102
mse	0.0008	0.0013	0.0014	0.0015	0.0015	0.0015	0.0015	0.0012	0.0008
$\sigma_{x=1}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=1}$	0.8809	0.8423	0.8269	0.8202	0.8187	0.8193	0.8254	0.8403	0.8796
std.err.	0.0258	0.0327	0.0360	0.0373	0.0379	0.0370	0.0355	0.0341	0.0265
bias	-0.0090	-0.0103	-0.0107	-0.0113	-0.0113	-0.0123	-0.0121	-0.0122	-0.0104
mse	0.0007	0.0012	0.0014	0.0015	0.0016	0.0015	0.0014	0.0013	0.0008

Table A.10: Estimated standard error, bias and mean squared error of probability predictions, relative to the third simulation setting: Frank copula distribution of (Y^1, Y^2), size 200, logistic constrained regression method.

Simulation 3, LCR Model, N=500									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=0}$	0.8858	0.8477	0.8336	0.8267	0.8248	0.8266	0.8331	0.8477	0.8857
std.err.	0.0163	0.0211	0.0235	0.0236	0.0238	0.0238	0.0225	0.0212	0.0165
bias	-0.0041	-0.0049	-0.0040	-0.0048	-0.0052	-0.0049	-0.0044	-0.0049	-0.0042
mse	0.0003	0.0005	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0003
$\sigma_{x=1}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=1}$	0.8872	0.8486	0.8332	0.8274	0.8252	0.8265	0.8335	0.8476	0.8856
std.err.	0.0164	0.0213	0.0237	0.0239	0.0239	0.0235	0.0232	0.0213	0.0165
bias	-0.0027	-0.0039	-0.0043	-0.0041	-0.0048	-0.0050	-0.0041	-0.0050	-0.0043
mse	0.0003	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0003

Table A.11: Estimated standard error, bias and mean squared error of probability predictions, relative to the third simulation setting: Frank copula distribution of (Y^1, Y^2), size 500, logistic constrained regression method.

Simulation 3, LCR Model, N=500									
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\sigma_{x=0}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=0}$	0.8875	0.8504	0.8355	0.8293	0.8270	0.8293	0.8351	0.8502	0.8880
std.err.	0.0118	0.0152	0.0156	0.0166	0.0169	0.0169	0.0160	0.0149	0.0114
bias	-0.0024	-0.0021	-0.0020	-0.0022	-0.0029	-0.0023	-0.0025	-0.0024	-0.0019
mse	0.0001	0.0002	0.0002	0.0003	0.0003	0.0003	0.0003	0.0002	0.0001
$\sigma_{x=1}$	0.8899	0.8526	0.8376	0.8315	0.8299	0.8315	0.8376	0.8526	0.8899
$\hat{\sigma}_{x=1}$	0.8886	0.8513	0.8353	0.8288	0.8277	0.8288	0.8345	0.8507	0.8876
std.err.	0.0119	0.0148	0.0160	0.0166	0.0169	0.0168	0.0163	0.0148	0.0118
bias	-0.0013	-0.0012	-0.0022	-0.0027	-0.0022	-0.0027	-0.0031	-0.0019	-0.0023
mse	0.0001	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002	0.0001

Table A.12: Estimated standard error, bias and mean squared error of probability predictions, relative to the third simulation setting: Frank copula distribution of (Y^1, Y^2), size 1000, logistic constrained regression method.

Appendix B

Software development

We report a selection of the programs used to show results in the thesis. We have used the statistical software *R*. We indicate with τ the general quantile.

```
###########################################################################
######gradient search algorithm for non linear least squares,###########
######polynomial constrained regression##################################
###########################################################################

mom.gs.l <- function(y,r,w) -mean(w*(y - (r^2*(3-2*r)*(r>0)*(r<1) + (r>1))
              *(1-m)-m)^2) #non-linear polynomial function
mom.gs.s <- function(tx,y,r,w) tx%*%(w*2*(y - (r^2*(3-2*r)*(r>0)*(r<1) + (r>1))
  *(1-m)-m)*(6*(r-r^2)*(r>0)*(r<1)*(1-m)))/length(y) #gradient
mom.gs <- function(y,x,w=1,beta=lm.fit(x,y)$coef,delta=1,a=.25,b=1.25,tol=1e-7,
                 maxiter=1000){
tx <- t(x)
r <- x%*%beta
l <- mom.gs.l(y,r,w)
s <- mom.gs.s(tx,y,r,w)
for(i in 1:maxiter) {
nb <- beta + s*delta
r <- x%*%nb
nl <- mom.gs.l(y,r,w)
if(nl<l) {
delta <- a*delta
}
else {
if(abs(max((beta-nb)/beta))<tol) return(nb)
beta <- nb
l <- nl
s <- mom.gs.s(tx,y,r,w)
```

```
delta <- b*delta
}
}
print("Convergence not achieved")
nb
} #gradient search algorythm
#############################################################
##########loglikelihood function for the###################
##########logistic constrained model######################
#############################################################
loglik <- function(gamma, y, x, m, w=1){

n <- length(y)
eta <- x%*%cbind(gamma)
expeta <- exp(eta)
tx <-t(x)
ll <- log(1-m) - log(1+expeta) + y*log(expeta+m) - y*log(1-m)
g <- (expeta)*(-1/(1+expeta) + y/(m + expeta))
if(any(out <- which(eta > 709))){
yout <- y[out]
etaout <- eta[out]
ll[out] <- log(1-m) - etaout + yout*etaout - yout*log(1-m)
g[out] <- yout - 1
}
logl <- -sum(ll*w)
attr(logl, "gradient") <- -t(tx%*%(g*w))
logl
   }
```

For the simulation studies we report the program relative only to the first scenario in 5.5.

```
set.seed(123)

#####################################################
######binormal omoskedastik setting################
#####################################################
library(mnormt)
rMnorm <- function(n, mu1,mu2, V){
y <- NULL
for(i in 1:n){
y <- rbind(y,rmnorm(n = 1, mean = c(mu1[i],mu2[i]), varcov = V))
```

```
}
y
}

V <- matrix(c(1,0.7,0.7,2),2,2)

##########################################################################
#####Theoretical calculation of concordance probability ##################
##########################################################################
a0 <- 1
a1 <- 3
b0 <- b1 <-2
q10 <- qnorm(p=tau, a0, 1)
q11 <- qnorm(p=tau, a0+a1, 1)
q20 <- qnorm(p=tau, b0, sqrt(2))
q21 <- qnorm(p=tau, b0+b1, sqrt(2))

p0 <- pmnorm(c(q10, q20), mean = c(a0, b0), varcov = V)
sigma0 <- 1 + 2*(p0 - tau)

p1 <- pmnorm(c(q11, q21), mean = c(a0+a1, b0+b1), varcov = V)
sigma1 <- 1 + 2*(p1 - tau)

p.true <- c(sigma0, sigma1)
################################################
######Monte Carlo replications#################
################################################
library(quantreg)
B <- 2000
ppLC <- NULL
ppPC <- NULL
m <- abs(2*tau -1) #maximum negative dependence
for(i in 1:B){
#############################################
######generate binormal omoskedastic data####
#############################################
x <- rbinom(n,1,p=0.5)
mu1 <- a0 + a1*x
mu2 <- b0 + b1*x
y <- rMnorm(n, mu1, mu2, V)
y1 <- y[,1]
y2 <- y[,2]
```

```
m1 <- rq(y1 ~ x, tau = tau)
m2 <- rq(y2 ~ x, tau = tau)

res1 <- m1$residuals
res2 <- m2$residuals

s1 <- sign(res1)
s2 <- sign(res2)

Z <- as.numeric(s1 == s2)
xx <- cbind(1,x)

#############################################
#########logistic constrained regression#####
#############################################
p0 <- mean(Z)
int0 <- log((p0-m + 0.03)/(1-p0)) #starting intercept point

ModLC <- nlm(f = loglik, p = c(int0,0), y = Z, X = xx, m=m, stepmax=10000)
beta.hatLC <- ModLC$estimate
predLC <- cbind((exp(beta.hatLC[1])+m)/(1 + exp(beta.hatLC[1])),
           (exp(sum(beta.hatLC))+m)/(1 + exp(sum(sum(beta.hatLC)))) )
ppLC <- rbind(ppLC, predLC)

#########################################
###polynomial constrained regression#####
#########################################
X <- cbind(1,t <- scale(xx[,-1]))
ModPC <- mom.gs(Z,X)
beta.hatPC <- (c(modPC[1]-sum(modPC[-1]*attr(t,"scaled:center")/attr(t,"scaled:scale
               modPC[-1]/ attr(t,"scaled:scale")))
predPC <- cbind((beta.hatPC[1]^2*(3-2*beta.hatPC[1])*(beta.hatPC[1]>0)*(beta.hatPC[1
          + (beta.hatPC[1]>1))*(1-m)+m,
           (sum(beta.hatPC)^2*(3-2*sum(beta.hatPC))*(sum(beta.hatPC)>0)
           *(sum(beta.hatPC)<1) + (sum(beta.hatPC)>1))*(1-m)+m)

ppPC <- rbind(ppPC, predPC)

}

######Calculate standard error, bias and MSE of probability estimates###

#########Logistic constrained regression#####
```

```
ppestLC <- apply(ppLC,2,mean)
sdestpredLC <- apply(ppLC,2,sd)
biaspredLC <- ppestLC - p.true
MSEpredLC <- apply(((ppLC - p.true)^2),2,mean)

    ########Polynomial constrained regression####
ppestPC <- apply(ppPC,2,mean)
sdestpredPC <- apply(ppPC,2,sd)
biaspredPC <- ppestPC - p.true
MSEpredPC <- apply(((ppPC - p.true)^2),2,mean)
```

The following selection of program refers to the application shown in 6.

```
rm(list=ls())
library(quantreg)
library(epiR)
lung <- read.table(file="/home/silvia/Dropbox/RandomBivariate/ProgrammiR/
         Data_lungfun.txt", sep="\t", header=TRUE)


ratio <- round(lung$fev1/lung$fvc,3)
lungfun <- cbind(lung, ratio)
lungfunf <- lungfun[sex=='female',]
####################################################################
#########################estimation of the model##################
####################################################################

m1 <- rq(fev1 ~ age + hgt, tau = tau, data=lungfunf)
m2 <- rq(ratio ~ age + hgt, tau = tau, data=lungfunf)

res1 <- m1$residuals
res2 <- m2$residuals

sr1 <- sign(res1)
sr2 <- sign(res2)

indip <- 1 + 2*i^2- 2*i
m <- abs(2*i-1)

Z <- as.numeric(sr1 == sr2)
x <- cbind(1,lungfunf$age, lungfunf$hgt)
```

```
###############Logistic constrained regression######################
p0 <- mean(Z)
int0 <- log((p0-m+0.03 )/(1-p0))
ModLC <- nlm(f = loglik, p = c(int0,0,0), y = Z, X = X, m=m, stepmax=1000)
gamma.hatLC <- ModLC$estimate

prLC <- x%*%gamma.hatLC #predizioni modello
p.hatLC <- (exp(prLC)+m)/(1 + exp(prLC))


###############Polynomial constrained regression######################
X <- cbind(1,t <- scale(x[,-1]))
modPC <- mom.gs(Z,X)
gamma.hatPC <- (c(modPC[1]-sum(modPC[-1]
                *attr(t,"scaled:center")/attr(t,"scaled:scale")),
                 modPC[-1]/attr(t,"scaled:scale")))


######################################################################
#######bootstrap exponentially tilted replications###################
######################################################################
Gammab.hatLC <- NULL
Gammab.hatPC <- NULL
pb.hatLC <- NULL
pb.hatPC <- NULL
N <- nrow(lungfunf)
B <- 200
for(j in 1:B){

ww <- rexp(N)

m1b <- rq(fev1 ~ age + hgt, tau = tau, data = lungfunf, w=ww)
m2b <- rq(ratio ~ age + hgt, tau = tau, data = lungfunf, w=ww)

res1b <- m1b$residuals
res2b <- m2b$residuals

sr1b <- sign(res1b)
sr2b <- sign(res2b)

Zb <- as.numeric(sr1b == sr2b)
```

```
#########Logistic Constrained regression######
p0b <- mean(Zb)
int0b <- log((p0b-m+0.05)/(1-p0b))
ModbLC <- nlm(f = loglik, p = c(int0b,0,0), y = Zb, X = X, m=m, w=ww,
                stepmax = 1000)
gammab.hatLC <- ModbLC$estimate
Gammab.hatLC <- rbind(Gammab.hatLC, gammab.hatLC)
prbLC <- x%*%gammab.hatLC
pb.hatLCn <- (exp(prb)+m)/(1 + exp(prb))
pb.hatLC <- cbind(pb.hatLC,pb.hatLCn)

##########Polynomial constrained regression################
modbPC <- mom.gs(Zb,X, w=ww)
gammab.hatPC <- (c(modbPC[1]-sum(modbPC[-1]
              *attr(t,"scaled:center")/attr(t,"scaled:scale")),
                   modbPC[-1]/attr(t,"scaled:scale")))
prbPC <- x%*%gammab.hatPC
pb.hatbPCn <- (prbPC^2*(3-2*prbPC)*(prbPC>0)*(prbPC<1) + (prbPC>1))*(1-m)+m
pb.hatPC <- cbind(pb.hatLC, pb.hatbPCn)
Gammab.hatPC <- rbind(Gammab.hatPC, gammab.hatPC)

}

#####################################################################
################post estimation commands############################
#####################################################################

#############Logistic constrained regression######################
sd.gamma.hatLC <- apply(Gammab.hatLC,2,sd)
ci_lowLC <- gamma.hatLC - 1.96*sd.gamma.hatLC
ci_upLC <- gamma.hatLC + 1.96*sd.gamma.hatLC
ci.estLC <- cbind(gamma.hatLC, gamma_lowLC, gamma_upLC)

sd.p.hatLC <- apply(pb.hatLC,1,sd)

prLC <- x%*%gamma.hatLC
p.hatLC <- (exp(prLC)+m)/(1 + exp(prLC))
p.hat_lowLC <- p.hatLC - 1.96*sd.p.hatLC
p.hat_upPC <- p.hatPC + 1.96*sd.p.hatLC
ci_predLC <- cbind(p.hatPC, p.hat_lowLC, p.hat_upLC)

#############Polynomial constrained regression###################
```

```
sd.gamma.hatPC <- apply(Gammab.hatPC,2,sd)
gamma_lowPC <- gamma.hatPC - 1.96*sd.gamma.hatPC
gamma_upPC <- gamma.hatPC + 1.96*sd.gamma.hatPC
ci.estPC <- cbind(gamma.hatPC, gamma_lowPC, gamma_upPC)


sd.p.hatPC <- apply(pb.hatPC,1,sd)


prPC <- x%*%gamma.hatPC
p.hatPC <- (pr^2*(3-2*prPC)*(prPC>0)*(prPC<1) + (prPC>1))*(1-m)+m
p.hat_lowPC <- p.hatPC - 1.96*sd.p.hatPC
p.hat_upPC <- p.hatPC + 1.96*sd.p.hatPC
ci_predPC <- cbind(p.hatPC, p.hat_lowPC, p.hat_upPC)


##########################################################################
##################Graphs of concordance probability######################
##########################################################################
par(las=1, mai=c(.9,1.2,0.7,0.3))
plot(lungfunf$age, p.hatLC, ylim=c(m,1), main = "Concordance plot LC",
ylab = "Concordance Probability", xlab="Age",cex.main=2.5, cex.lab=2.5,
cex.axis=1.9, mgp = c(4.5,1.1,0), cex=1, pch=19)
abline(h=1, col=5, lwd = 3)
abline(h = indip, col="blue", lwd = 3)
abline(h = m, col="red", lwd = 3)

par(las=1, mai=c(.9,1.2,0.7,0.3))
plot(lungfunf$age, p.hatPC, ylim=c(m,1), main = "Concordance plot PC",
ylab = "Concordance Probability", xlab="Age", cex.main=2.5, cex.lab=2.5,
cex.axis=1.9, mgp = c(4.5,1.1,0), cex=1, pch=19)
abline(h=1, col=5, lwd = 3)
abline(h = indip, col="blue", lwd = 3)
abline(h = m, col="red", lwd = 3)


############################################################################
#######################Z-score independence test########################
############################################################################

data <- as.data.frame(cbind(lungfunf$age,lungfunf$hgt))
pat <- epi.cp(data)
cv <- pat$cov.pattern
```

```
pprLC <- unique(p.hatLC)
sddLC <- unique(sd.p.hatLC)

pr.cvLC <- NULL
zscoreLC <- NULL
for(h in 1:length(pprLC)){
  phat.cvLC <- pprLC[h]
  pr.cvLC <- rbind(pr.cvLC, phat.cvLC)
  zscore_pLC <- (phat.cvLC - indip)/ sddLC[h]
  zscoreLC <- rbind(zscoreLC, zscore_pLC)
 }

col2rgb("grey", alpha=TRUE)
greytrans <- rgb(190, 190, 190, 127, maxColorValue=255)
par(mai=c(.9,1.3,0.6,0.3), mgp = c(4.5,1,0))
plot(cv$V1,zscoreLC, xlab="Age", ylab="Z-score", ylim =c(-3,16), yaxt=’n’,
main= "Zscore test LC", cex.main=2.5, cex.lab =2.3, cex.axis=1.8, pch=19)
axis(2, at= -1.96, labels=T, lwd=3, col="red", cex.axis=1.8, las=1)
axis(2, at= 1.96, labels=T, lwd=3, col="red", cex.axis=1.8, las=1)
abline(h=-1.96, lwd=3, col="red")
abline(h=1.96, lwd=3, col="red")
xmin <- par("usr")[1]
xmax <- par("usr")[2]
polygon(c(xmin, xmin, xmax, xmax, xmin), c(-1.96, 1.96,1.96, -1.96,-1.96), col=gre

pprPC <- unique(p.hatPC)
sddPC <- unique(sd.p.hatPC)

pr.cvPC <- NULL
zscorePC <- NULL
for(h in 1:length(pprPC)){
  phat.cvPC <- pprPC[h]
  pr.cvPC <- rbind(pr.cvPC, phat.cvPC)
  zscore_pPC <- (phat.cvPC - indip)/ sddPC[h]
  zscorePC <- rbind(zscorePC, zscore_pPC)
 }

par(mai=c(.9,1.3,0.6,0.3), mgp = c(4.5,1,0))
plot(cv$V1,zscorePC, xlab="Age", ylab="Z-score", ylim =c(-3,16), yaxt=’n’,
main= "Zscore test PC", cex.main=2.5, cex.lab =2.3, cex.axis=1.8, pch=19)
axis(2, at= -1.96, labels=T, lwd=3, col="red", cex.axis=1.8, las=1)
axis(2, at= 1.96, labels=T, lwd=3, col="red", cex.axis=1.8, las=1)
```

```
abline(h=-1.96, lwd=3, col="red")
abline(h=1.96, lwd=3, col="red")
xmin <- par("usr")[1]
xmax <- par("usr")[2]
polygon(c(xmin, xmin, xmax, xmax, xmin), c(-1.96, 1.96,1.96, -1.96,-1.96), col=greyt
```