UNIVERSITÉ DU QUÉBEC À MONTRÉAL

COPULA-BASED QUANTILE REGRESSION

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

MAHDI SHAHBAZI

NOVEMBRE 2020

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

COPULA-BASED QUANTILE REGRESSION

DISSERTATION

PRESENTED

AS PARTIAL REQUIREMENT

TO THE MASTERS IN MATHEMATICS

BY

MAHDI SHAHBAZI

MAY 2020

CONTENTS

LIST OF TABLES

LIST OF FIGURES

RÉSUMÉ

La régression linéaire, qui a comme objectif la modelisation de la moyenne conditionnelle, est une méthode standard pour étudier la relation entre une variable réponse et un ensemble de prédicteurs. Elle explique principalement la relation linéaire entre les variables au milieu de la distribution conditionnelle de la variable réponse et, généralement, la méthode nécessite plusieurs hypothèses principales à satisfaire en pratique. En plus, cette méthode est sensible aux observations aberrantes.
En utilisant les fonctions quantiles conditionnels proposées par Koenker et Bassett (1978), on peut expliquer la relation entre la variable réponse et les prédicteurs dans des quantiles souhaités afin de capturer une image globale de l'effet des covariables sur la distribution de la variable réponse. La régression quantile nécessite aucune pré-hypothèse et c'est un outil robuste pour analyser les données influencées par des valeurs aberrantes; cette technique performe bien également en présence de données hétérogènes. Cependant, la méthode décrit une relation linéaire entre les quantiles de la variable réponse et les prédicteurs.
Ainsi, l'étude de la structure de dépendance non linéaire entre la variable réponse et les prédicteurs lorsque la variable réponse suit une distribution asymétrique ou à queue lourde est un grand défi pour les modèles existants.
Les copules sont des modèles appropriés pour décrire l'association entre des variables aléatoires. Elles permettent de modéliser séparément la dépendance et les distributions marginales. D'autre part, le modèle d'association basé sur les copules peut être vu comme une régression quantile. Cette méthode permet d'étudier l'association non linéaire entre variables aléatoires et expliquer ainsi la structure de dépendance entre les variables dans les queues de la distribution jointe. Ainsi la régression quantile basée sur des copules est une alternative utile aux méthodes existantes telles que les moindres carrés ordinaires et la régression quantile.
Dans ce mémoire nous faisons une investigation globale de la méthode de régression quantile basée sur les copules et nous comparons cette méthode avec les approches existantes.

Mots-clés: régression, régression quantile, fonctions quantiles conditionnelles, copules, dépendance non linéaire.

ABSTRACT

The linear regression framework, which is based on conditional mean, is a standard method for investigating the relationship between a response variable and a set of predictors. It mainly explains the linear relationship between variables in the middle of the distribution of the outcome and, generally, the method requires principal assumptions and is sensitive to outliers.

By employing conditional quantile functions presented by Koenker and Bassett (1978), one can explain the relationship between the response variable and predictors in desired quantiles, and thus, one can capture an overall picture of the predictors' effects on the response variable distribution. The quantile regression does not require any pre-assumption, and it is a robust tool for analyzing data influenced by outliers. The method also has acceptable performance in heterogeneous situations. However, this method explains the linear relationship between the response quantiles and the predictors.

Thus, the investigation of nonlinear dependence structure between the response variable and the predictors, when the response follows an asymmetric or heavy-tailed distribution, is a big challenge for existing models, including both linear and quantile regression methods.

Copula functions are suitable models for describing the association between random variables. They allow modeling of dependence and the marginal distributions separately. On the other hand, the copula-based quantile regression is a quantile regression model based on copulas. This method provides a flexible way to investigate the nonlinear association between random variables and enables us to explain the dependence structure of variables in the tails of their joint distribution. Consequently, copula-based quantile regression is a useful alternative to existing methods such as the ordinary least squares and quantile regression.

This thesis provides a global investigation of the quantile regression method based on copulas, and we compare this method with existing approaches.

Key-words: Regression, Quantile regression, Conditional quantile functions, Copulas, non-linear dependence.

INTRODUCTION

Several relational approaches are available in the literature to examine the association between variables. The linear regression is one of the existing methods which studies the linear relationship between a response variable and a set of predictors. It estimates the mean of the response variable using these predictors (conditional mean). The underlying optimization problem of linear regression is to minimize the sum of squared errors, and it is thus labeled as ordinary least squares (OLS) regression.

OLS regression, generally, requires some principal assumptions such as outcome normality, homogeneity of variance, and independence of the errors. Violation of these primary assumptions has a substantial impact on model estimation. Also, it is notable that the method is sensitive to outliers.

The quantile regressions (QR) can be considered as an alternative approach to OLS regression. The method models the conditional quantile function instead of the outcome conditional mean. Koenker and Bassett Jr (1978) introduced the method to estimate linear models based on conditional quantile functions. These models are able to investigate the association between random variables, not only in the middle of the response conditional distribution but also in the tails of the distribution. Consequently, they can provide a comprehensive picture of the predictors' effects on the distribution of the response variable. An appealing feature of QR that has been frequently highlighted is that it describes the conditional distribution of the response variable while it does not require global distributional assumptions (Koenker, 2005). Moreover, QR is a robust tool for analyzing data influenced by outliers since the corresponding model reduces the influence of outliers (Koenker & Bassett Jr, 1982). Also, the method is known to have acceptable performance in heterogeneous situations.

A common limitation of OLS and QR techniques is that both methods are less convenient for describing nonlinear dependence structure between the response and the predictors. Moreover, capturing the association between these variables at a given quantile in the upper-right quadrant and lower-left quadrant, when the distribution of the response variable is asymmetric or a heavy-tailed distribution, is a big challenge for both methods.

Copula-based quantile regression (CBQR) models are flexible alternative approaches to existing techniques such as OLS and QR; they are quantile regression modeling approaches based on copula functions to study both linear and nonlinear associations between random variables (Briollais & Durrieu, 2014). Copulas have recently seen growing attention in the literature; this is mainly due to Sklar's Theorem, which states that a copula is a function that couples a multivariate distribution function to its one-dimensional marginal distributions (Nelsen, 2006). Thus, joint modeling of random variables can be achieved using a copula model and marginal distributions. Several classes of copulas exist; this allows for flexible modeling of random variables joint distribution.

In this thesis, we are interested in investigating CBQR models because they do not require strong pre-assumptions such as normality or homogeneity of variance, and they are robust to outlies.

A summary of the upcoming chapters is given as follows: in Chapter 1, we present a brief history of copula theory, provide definitions and essential properties of copulas, describe various methods of constructing copulas and introduce some prominent families of copulas. In Chapter 2, we describe quantiles and quantile check/loss function, introduce QR, and explain the basic properties of QR. Chapter 3 investigates the copula-based quantile models for Elliptical and Archimedean copulas; CBQR parameter estimation is also the focus of this chapter. In Chapter

4, we evaluate the performance of CBQR with respect to existing methods using simulated data generated from different models. Through these simulation studies, we compare the performance of OLS, QR, and CBQR techniques. We end this chapter with copula misspecification sensitivity analysis. Finally, the thesis terminates with some conclusions and a discussion.

CHAPTER I

BASICS OF COPULAS

The concept of copulas was first introduced in mathematical and statistical literature by Sklar (1959), who described a copula as a function that joins (or "couples") a multivariate distribution function to its one-dimensional margins which are uniform on the interval [0,1] (Nelsen, 2006).

As Fisher (1997) mentioned in the Encyclopedia of Statistical Sciences, "Copulas [are] of interest to statisticians for two main reasons: firstly, as a way of studying scale-free measures of dependence, and secondly, as a starting point for constructing families of bivariate distributions (Nelsen, 2003)." Thus, copulas allow us to study the dependence structure of the marginal distributions.

Some of a copulas' utilities are 1) a capability to investigate the nonlinear dependence, 2) an ability to measure dependencies for heavy-tailed distributions, and 3) flexibility to fit parametric, semi-parametric, and non-parametric models.

In this chapter, we will present some important theorems, definitions, and properties necessary to understand copulas' concept. A summary of this chapter is as follows. Section 1.1 provides definitions and important properties of copulas. Section 1.2 describes various methods of constructing copulas and Section 1.4 introduces some important families of copulas.

1.1 Definitions and basic properties of copulas

In this section, firstly, we introduce some new notation, definitions, theorems, properties, and concepts, based on Nelsen (2006).

Let $\mathbb{R}$, $\bar{\mathbb{R}}$ and $\bar{\mathbb{R}}^2$ denote the common real line $(-\infty, +\infty)$, extended real line $[-\infty, +\infty]$, and extended real plane $\bar{\mathbb{R}} \times \bar{\mathbb{R}}$, respectively. The Cartesian product of two closed intervals $([x_1, x_2] \times [y_1, y_2])$ is a rectangle, B, with the vertices $(x_1, y_1), (x_1, y_2), (x_2, y_1)$, and (x_2, y_2) in $\bar{\mathbb{R}}^2$. Finally, define the unit square $\mathbb{I}^2$ as the product $\mathbb{I} \times \mathbb{I}$ where $\mathbb{I} = [0, 1]$.

Consider two random variables X and Y with distribution functions $F(x)$ and $G(y)$, respectively. Let the joint distribution function of these two random variables be $H(x, y)$. For each point (x, y), we have three numbers: $F(x)$, $G(y)$ and $H(x, y)$. According to Nelsen (2006), a function like H such that whose domain, Dom$H$, is a subset of $\bar{\mathbb{R}}^2$ and whose range, Ran$H$, is a subset of $\mathbb{R}$, is called a 2-place real function.

Definition 1.1. *H-volume of rectangle $B = [x_1, x_2] \times [y_1, y_2]$ is given by*

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1),$$

where H is a 2-place real function such that Dom$H = S_1 \times S_2$ and S_1, S_2 are nonempty subsets of $\bar{\mathbb{R}}$ and the vertices of the rectangle B are in Dom$H$.

Definition 1.2. *Consider a 2-place real function H with DomH. For all rectangles B whose vertices fall in DomH, we say that H is 2-increasing if $V_H(B) \geq 0$.*

Assume that H is a nondecreasing function of x and y. The following example shows that each nondecreasing function, such as H, is not 2-increasing in its domain.

Example 1.1. Let $H(x, y) = max(x, y)$ be a function on $\mathbb{I}^2$, where $\mathbb{I} = [0, 1]$. H is a nondecreasing function of x and of y. By definition 1.2, we have

$$V_H(\mathbb{I}^2) = 1 - 1 - 1 + 0 = -1.$$

Thus, H is not a 2-increasing function. It should be noted that if H is a 2-increasing function, it does not imply that H is nondecreasing in each argument. The following definition is an additional hypothesis that shows a 2-increasing function is nondecreasing with respect to each argument.

Definition 1.3. *Let S_1 and S_2 be nonempty subsets of $\mathbb{R}$, and a_1 and a_2 are minimum elements of S_1 and S_2, respectively. The function H is grounded if $H(x, a_2) = 0 = H(a_1, y)$ for all $(x, y) \in S_1 \times S_2$.*

Example 1.2. Let $H(x, y) = \frac{(x+1)(e^y-1)}{x+2e^y-1}$ be a function with domain $[-1, 1] \times [0, \infty]$. The minimum values of $[-1, 1]$ and $[0, \infty]$ are -1 and 0 respectively, so that $H(x, 0) = 0 = H(-1, y)$ $\forall (x, y) \in [-1, 1] \times [0, \infty]$. Thus, H is grounded.

Definition 1.2 and 1.3 lead to the following Proposition proved in Nelsen (2006, p.9):

Proposition 1.1. *Let S_1 and S_2 be nonempty subsets of $\bar{\mathbb{R}}$. If H is a grounded and 2-increasing function with domain $S_1 \times S_2$, then it is nondecreasing for each argument.*

1.1.1 The definition of a copula

Definition 1.4. *A subcopula C' is a function with the following properties:*
1. The domain of C' is $S_1 \times S_2$, where S_1 and S_2 are subsets of $\mathbb{I}$ including 0 and 1;

2. C' is grounded and 2-increasing function;

3. $\forall(u,v) \in S_1 \times S_2$, $C'(u,1) = u$ and $C'(1,v) = v$.

Note that because C' is grounded, its minimum values in its domain equals zero. According to Property 3 of Definition 1.4, the maximum value of a subcopula C' for every (u,v), in DomC' is equal to one. Hence, $0 \leq C'(u,v) \leq 1$.

We can now present a copula function definition, which is a subcopula with domain $[0,1]^2$.

Definition 1.5. *A (two-dimensional) copula is a subcopula function $C : [0,1]^2 \to [0,1]$ with the following properties:*

1. For all $u,v \in \mathbb{I}$, C is grounded, that is, $C(u,0) = 0 = C(0,v)$;

2. For all $u,v \in \mathbb{I}$, C is such that $C(u,1) = u$ and $C(1,v) = v$;

3. For every $u_1, u_2, v_1, v_2 \in \mathbb{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, C is 2-increasing, that is, $V_C([u_1,u_2] \times [v_1,v_2]) = C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$.

Example 1.3. Let M be the function defined on $[0,1]^2$ by $M(u,v) = min(u,v)$, known as the "Fréchet-Hoeffding upper bound". We show that M is a copula function. That is, we have

1. For all $u,v \in \mathbb{I}$, M is grounded because, $min(u,0) = 0 = min(0,v)$.

2. For all $u,v \in \mathbb{I}$, $min(u,1) = u$ and $min(1,v) = v$.

3. In this step, we are going to demonstrate M is a 2-increasing function. To do so, for every u_1, u_2, v_1, v_2 in $\mathbb{I}$, we consider four different cases. First, $u_1 < v_1$ and $u_2 < v_2$; second, $u_1 < v_1$ and $u_2 > v_2$; third, $u_1 > v_1$ and $u_2 < v_2$; fourth, $u_1 > v_1$ and $u_2 > v_2$. We consider the first case ($u_1 < v_1$ and $u_2 < v_2$). For every $u_1, u_2, v_1, v_2 \in \mathbb{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, one can write $u_1 < v_1 < v_2$. In

this case, M is a 2-increasing function because

$$\begin{aligned} V_M([u_1,u_2]\times[v_1,v_2]) &= min(u_2,v_2) - min(u_2,v_1) - min(u_1,v_2) + min(u_1,v_1) \\ &= u_2 - min(u_2,v_1) - u_1 + u_1 \\ &= u_2 - min(u_2,v_1) \geq 0. \end{aligned}$$

Similarly, for the other cases, it can be proved that M is a 2-increasing function. So, from 1, 2, and 3 above, we conclude that M is a copula. In a similar manner, one can show that, W, the "Fréchet-Hoeffding lower bound" defined by $W(u,v) = max(u+v-1,0)$ with domain $[0,1]^2$, is also a copula function.

Example 1.4. The function, $\Pi(u,v) = uv$, is called the product copula. In what follows, we demonstrate that Π is indeed a copula in $[0,1]^2$.

1. For all u, v in $\mathbb{I}$, $\Pi(u,v)$ is grounded because the least value for u and v in $\mathbb{I}$ is zero. So, $\Pi(u,0) = 0 = \Pi(0,v)$.
2. For all u, v in $\mathbb{I}$, $\Pi(u,1) = u$ and $\Pi(1,v) = v$.
3. For every u_1, u_2, v_1, v_2 in $\mathbb{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$, $\Pi(u,v)$ is a 2-increasing functon because

$$\begin{aligned} V_\Pi([u_1,u_2]\times[v_1,v_2]) &= \Pi(u_2,v_2) - \Pi(u_2,v_1) - \Pi(u_1,v_2) + \Pi(u_1,v_1) \\ &= u_2v_2 - u_2v_1 - u_1v_2 + u_1v_1 \\ &= (v_2 - v_1)(u_2 - u_1) \geq 0. \end{aligned}$$

So, Π is a copula function.

As follows, we introduce an essential theorem about copulas, which is presented by Sklar in 1959. Sklar's theorem connects the joint distribution functions and their marginal distributions via copulas.

Theorem 1.1. *(Sklar's theorem): Let X and Y be random variables with distribution functions F and G, respectively, and joint distribution function H. Then* *there exists a copula C such that,*

$$H(x, y) = C(F(x), G(y)). \tag{1.1}$$

If F and G are continuous, then C is unique. Conversely, the function H in (1.1) is a joint distribution function with margins F and G if C is a copula and F and G are two distribution functions (Nelsen 2006, p.18). Next is the multivariate case of Sklar's theorem.

Theorem 1.2. ***(Multivariate case)*** *Let $X = (X_1, X_2, ..., X_d)$ be a multivariate random vector with d-dimensional joint distribution function H and continuous marginal distribution functions $F_1, F_2, ..., F_d$, respectively. There exists a d-copula C such that,*

$$H(x_1, x_2, ..., x_d) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d)). \tag{1.2}$$

Conversely, if C is a d-copula and $F_1, F_2, ..., F_d$ are distribution functions, then the function H in (1.2) is a d-dimensional joint distribution function with margins $F_1, F_2, \cdots, F_d$ (Cherubini et al., 2004).

1.1.2 Basic properties of copulas

Consider two random variables, X and Y, with continuous distribution functions $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$, respectively, and a joint distribution function $H(x, y) = P(X \leq x, Y \leq y)$. Let $U = F(X)$ and $V = G(Y)$, then according to Probability Integral Transformation, the random variables U and V have a uniform distribution on the interval $[0, 1]$.

Property 1.1. ***A copula is uniformly continuous***

Let C be a copula. Then for every (u_1, u_2) and (v_1, v_2) in domain of C, we have

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|.$$

Thus, C satisfies a Lipschitz condition in $[0,1]^2$. Thus, it is uniformly continious (Nelsen, 2003).

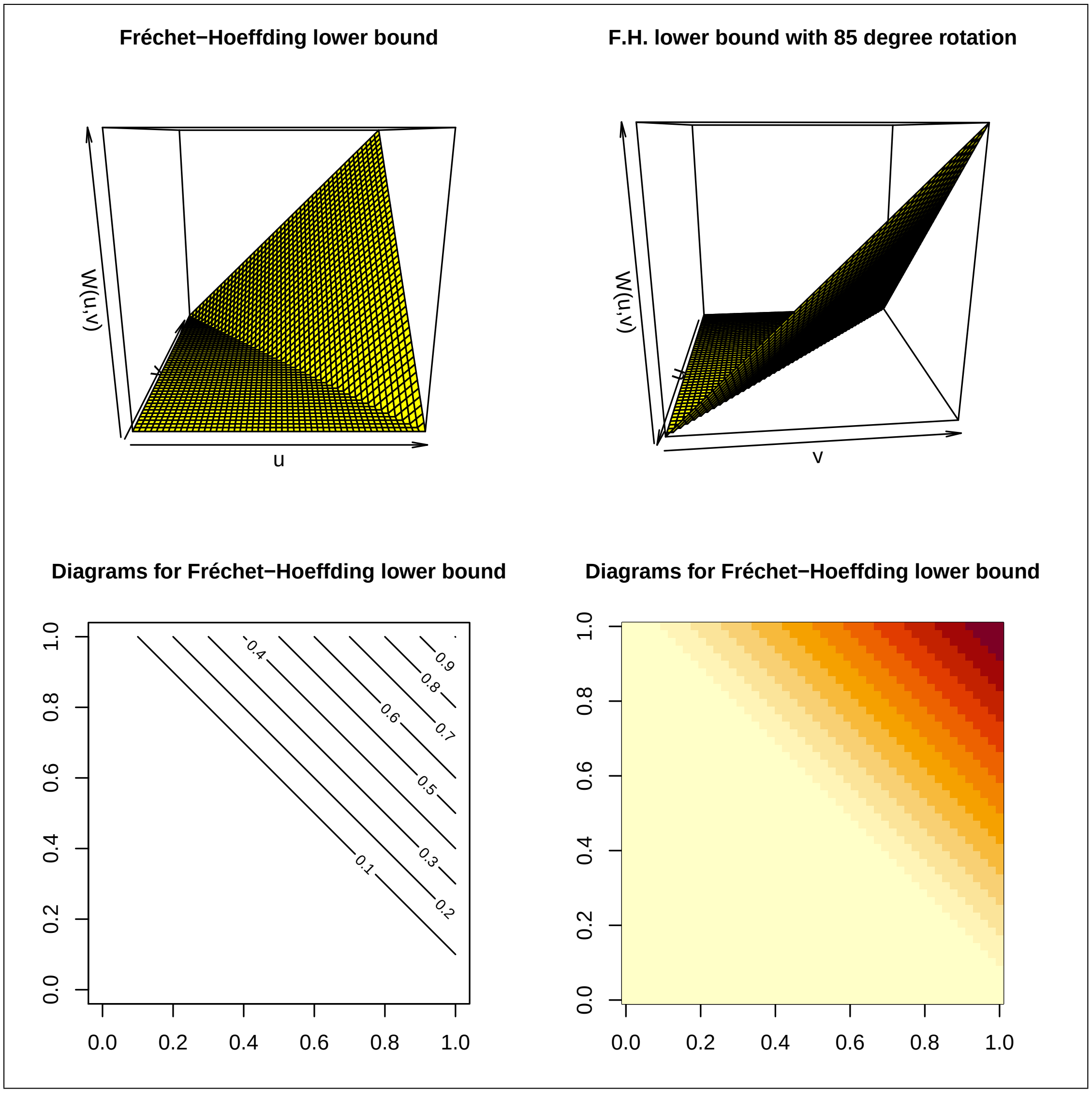

Figure 1.1: Three-dimensional graphs and contour diagrams for Fréchet-Hoeffding lower bound copula.

Property 1.2. ***A copula is between the Fréchet-Hoeffding bounds.***

Let C be a copula. For an arbitrary point (u, v) in Dom$C$, we can show that

$$W(u, v) \leq C(u, v) \leq M(u, v),$$

where $W(u,v) = max(u+v-1,0)$ is the Fréchet-Hoeffding lower bound and $M(u,v) = min(u,v)$ is the Fréchet-Hoeffding upper bound.

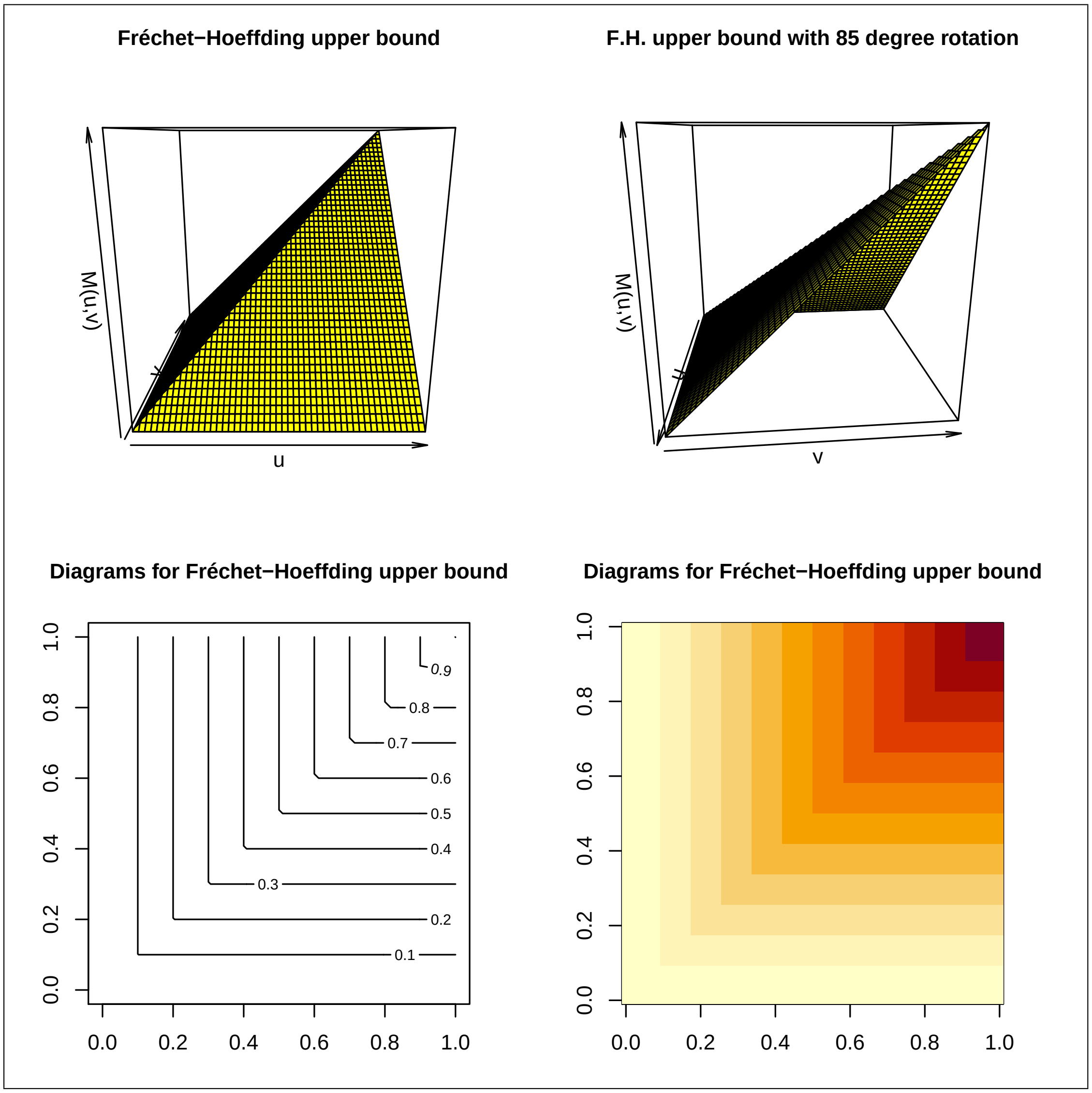

Figure 1.2: Three-dimensional graphs and contour diagrams for Fréchet-Hoeffding upper bound copula.

Definition 1.5 implies that $C(u,v) \leq C(u,1) = u$ and $C(u,v) \leq C(1,v) = v$. Thus, $C(u,v) \leq min(u,v)$. Also, the copula C is a 2-increasing function, and

$V_C([u,1] \times [v,1]) \geq 0$ implies

$$\begin{aligned} V_C([u,1] \times [v,1]) &= C(u,v) + C(1,1) - C(u,1) - C(1,v) \\ &= C(u,v) + 1 - u - v \geq 0. \end{aligned}$$

So, $C(u,v) \geq u+v-1$, and by considering $C(u,v) \geq 0$ yields $C(u,v) \geq max(u+v-1, 0)$. Therefore, for every copula C and every (u,v) in $\mathbb{I}^2$, a copula is between the Fréchet-Hoeffding bounds. On the other hand, any copula is a continuous surface situated between Fréchet-Hoeffding lower and upper bounds (Nelsen 2006, p.12).

Proposition 1.2. *Let C be the copula of the continuous random variables X and Y. Let Π be the product copula. Then X and Y are independent if and only if $C = \Pi$* (Nelsen, 2006).

Proof. C is a copula. So, for the random variables X and Y with distribution functions $u = F(x)$ and $v = G(y)$, there exists a joint distribution function H such that, $H(x,y) = C(F(x), F(y))$. By considering $C(F(x), F(y)) = \Pi = uv$, we have $H(x,y) = uv = F(x)G(y)$. That is, X and Y are independent. On the other hand, if X and Y are independent, we conclude that $H(x,y) = F(x)G(y) = uv$. Also, we know that $H(x,y) = C(F(x), F(y))$, thus, $C(F(x), F(y)) = uv$ and $C = \Pi$.

In Figure 1.1, Figure 1.2, and Figure 1.3, we illustrate the Fréchet-Hoeffding bounds and the product copula by three-dimensional graphs and contour diagrams. A contour diagram is a way to show the specific three-dimensional graph in a two-dimensional plate: i.e.: $C(u,v) = cte$.

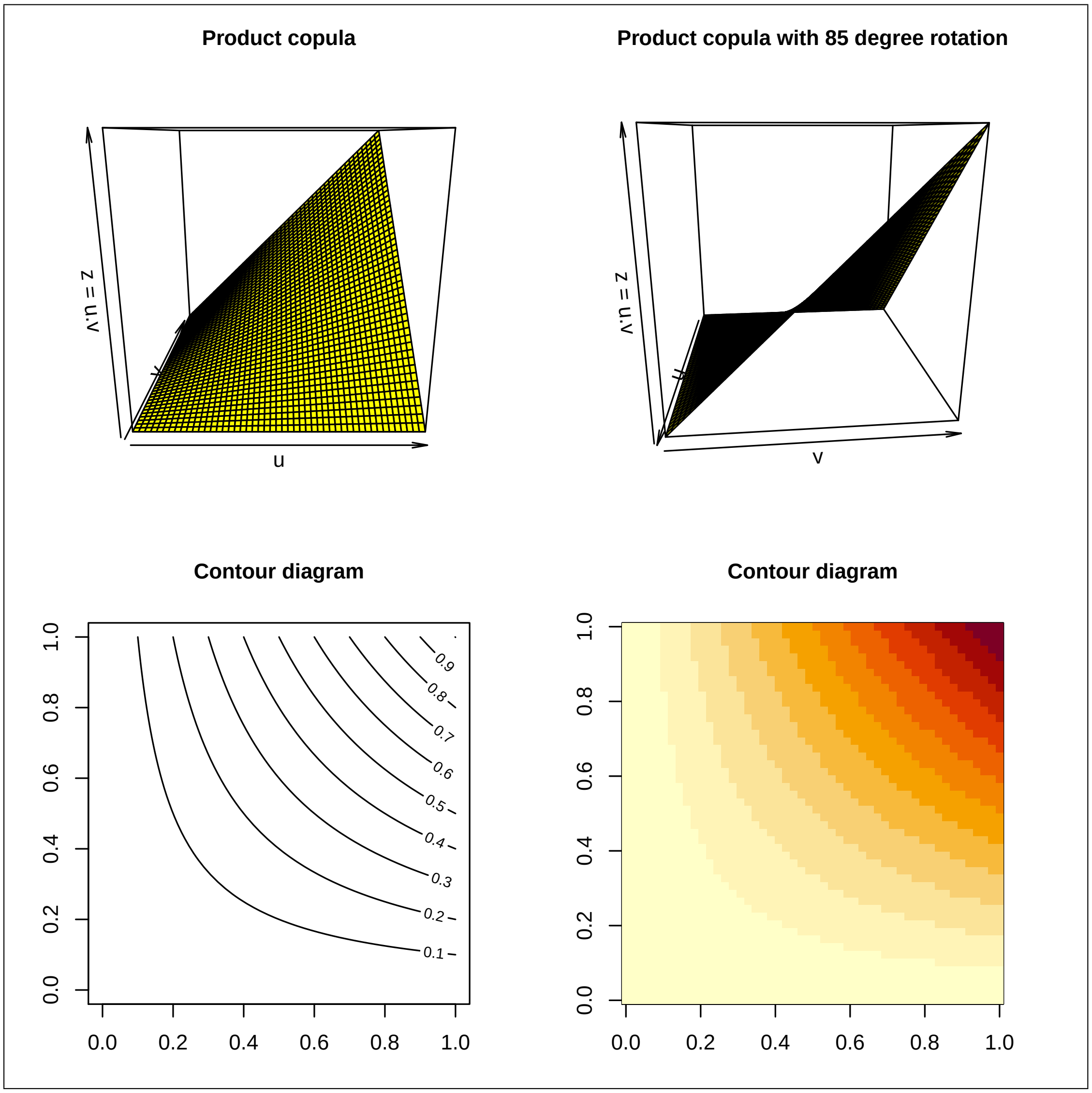

Figure 1.3: Three-dimensional graphs and contour diagrams for the product copula.

Proposition 1.3. *The partial derivatives of the copula C, with respect to variables, $U = F(X)$ or $V = G(Y)$, exist and are defined as*

$$0 \leq \frac{\partial}{\partial u} C(u, v) = P(V \leq v | U = u) \leq 1,$$
$$0 \leq \frac{\partial}{\partial v} C(u, v) = P(U \leq u | V = v) \leq 1.$$

In addition, the functions u $\mapsto \partial C(u, v)/\partial v$ and v $\mapsto \partial C(u, v)/\partial u$ are well-defined and nondecreasing almost everywhere on $[0, 1]$, see page 13 of Nelsen (2006) for proof. We can use the partial derivatives of the copula for determining conditional distribution functions. See Chapter 3 to explain how the conditional copula quantiles arise.

Property 1.3. ***The copula density***
The copula density, $c(.,.)$, exists everywhere in $\mathbb{I}^2$ and is defined as

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}.$$

Proof (Nelsen 2006, p.14).

Let X and Y be two continuous random variables with joint density function h, marginal distributions F and G, and density functions f and g, respectively, then, we have

$$h(x, y) = c(F_X(x), G_Y(y))f(x)g(y)$$

1.2 Methods of constructing copulas

There are several methods for creating copula. In this section, we introduce some techniques to generate bivariate and multivariate copulas. We access a bivariate or multivariate distribution function and its corresponding marginal distribution functions by defining a copula. It is useful for both simulation and modeling.

1.2.1 The inversion method

Let X and Y be random variables with joint distribution function, H, and known continuous marginal distributions $u = F(x)$ and $v = G(y)$. Their inverses exist and are F^{-1} and G^{-1}, respectively. When F and G are continuous, then for any

(u, v) in the domain of C, we can find a unique copula defined as

$$\begin{aligned} C(u, v) &= H(F^{-1}(u), G^{-1}(v)) \\ &= P(X \leq F^{-1}(u), Y \leq G^{-1}(v)). \end{aligned} \tag{1.3}$$

The above result provides a method of constructing copulas from joint distribution functions. Similar to the two-dimensional case, we present the result in the case of a random vector. Thus, for $(u_1, u_2, \ldots, u_d)^\top \in \mathbb{R}^d$, one can write

$$\begin{aligned} C(u_1, u_2, \ldots, u_d) &= H(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots F_d^{-1}(u_d)) \\ &= P(X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), \ldots, X_n \leq F_d^{-1}(u_d)). \end{aligned}$$

Example 1.5. Let X and Y be random variables with Gumbel bivariate joint distribution function defined by

$$H(x, y) = [1 + \exp(-x) + \exp(-y)]^{-1}, \quad x, y \in [-\infty, \infty].$$

The marginal distributions of X and Y are

$$F(x) = [1 + \exp(-x)]^{-1} \quad \text{and} \quad F(y) = [1 + \exp(-y)]^{-1}.$$

Hence,

$$F^{-1}(u) = -\ln\left(\frac{1-u}{u}\right) \quad \text{and} \quad F^{-1}(v) = -\ln\left(\frac{1-v}{v}\right).$$

Then, according to relationship (1.3), we have

$$\begin{aligned} C(u, v) &= H\left(-\ln\left(\frac{1-u}{u}\right), -\ln\left(\frac{1-v}{v}\right)\right) \\ &= \left[1 + \exp\left(-\left(-\ln\left(\frac{1-u}{u}\right)\right)\right) + \exp\left(-\left(-\ln\left(\frac{1-v}{v}\right)\right)\right)\right]^{-1} \\ &= \frac{uv}{u + v - uv}, \quad u, v \in [0, 1]. \end{aligned}$$

1.2.2 Generator function method

In this method, we use a specific function called the "Generator" function to construct a copula. It is denoted by φ and has the following properties:

1. φ is a continuous and strictly decreasing function with domain $[0,1]$ and range $[0,\infty]$;

2. φ is such that $\varphi(1)=0$;

3. The pseudo-inverse (see Appendix A) of φ exist (φ^{-1}) and is a function from $[0,\infty]$ to $[0,1]$ (Cherubini et al., 2004).

Theorem 1.3. *Let φ be a generator function and φ^{-1} be its pseudo-inverse. Then, the function C from $\mathbb{I}^2$ to $\mathbb{I}$ given by*

$$C(u,v)=\varphi^{-1}\left(\varphi(u)+\varphi(v)\right), \tag{1.4}$$

is a copula if and only if φ is convex.

Proof (Nelsen 2006, p.111).

Example 1.6. (Clayton copula) suppose the generator φ is given by $\varphi_\delta(t)=\frac{1}{\delta}(t^{-\delta}-1)$, $\delta\in[-1,\infty)\setminus\{0\}$. Then, its inverse is $\varphi_\delta^{-1}(t)=(\delta t+1)^{\frac{-1}{\delta}}$. By consideration of φ, φ^{-1} and relationship (1.4), the Clayton copula is defined as

$$\begin{aligned} C(u,v)&=\varphi^{-1}(\varphi(u)+\varphi(v))\\ &=[\delta(\frac{1}{\delta}(u^{-\delta}-1))+\delta(\frac{1}{\delta}(v^{-\delta}-1))+1]^{\frac{-1}{\delta}}\\ &=[(u^{-\delta}+v^{-\delta}-1)]^{\frac{-1}{\delta}}. \end{aligned}$$

Table 1.1 lists the generator function for well-known copulas of the Archimedean family. A complete list of these generator functions is presented in Nelsen (2006), page 116.

Name	$C_\delta(u,v)$	$\delta \in$	**Generator** $\varphi_\delta(t)$
Gumbel	$\exp(-[(-\ln u)^\delta + [(-\ln v)^\delta]^{\frac{1}{\delta}})$	$[1,\infty)$	$(-\ln(t))^\delta$
Clayton	$[\max(u^{-\delta} + v^{-\delta} - 1, 0)]^{-\frac{1}{\delta}}$	$[-1,\infty) \setminus \{0\}$	$\frac{1}{\delta}(t^{-\delta} - 1)$
Frank	$-\frac{1}{\delta}\ln(1 + \frac{(e^{-\delta u}-1)(e^{-\delta v}-1)}{e^{-\delta}-1})$	$(-\infty,+\infty) \setminus \{0\}$	$-\ln(\frac{e^{-\delta t}-1}{e^{-\delta}-1})$
Joe	$1 - [(1-u)^\delta + (1-v)^\delta - (1-u)^\delta(1-v)^\delta]^{\frac{1}{\delta}}$	$[1,\infty)$	$-\ln(1-(1-t)^\delta)$

Table 1.1 The list of well-known Archimedean copulas with their generator functions.

1.3 Dependence measures

This thesis introduced dependence modeling between X and Y via copulas, which are so far, parametric models. In mathematical statistics, there are several non-parametric association measures. Kendall's tau and Spearman's rho are two examples used to measure the strength of dependence implied by copulas (Mai & Scherer 2014, p.44).

In this section, we describe Kendall's tau as a concordance-based dependence measure.

Let (X, Y) be an independent and identically bivariate random vector with a continuous joint distribution function H. Also, let (x_1, y_1) and (x_2, y_2) be two pair observations of this vector. If $(x_1 - x_2)(y_1 - y_2) > 0$, these pair observations are concordant and if $(x_1 - x_2)(y_1 - y_2) < 0$, they are disconcordant. In fact, the Kendall's tau refers to differences between the concordant probability and the disconcordant probability (Mai & Scherer 2014, p.41).

More precisely, the Kendall's tau for two random copies of (X, Y), (X_1, Y_1) and

(X_2, Y_2), is defined as

$$\tau_{X,Y} = P\left[(X_1 - X_2)(Y_1 - Y_2) > 0\right] - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0\right].$$

Let U = F(X) and V = G(Y) be two random variables with uniform distribution function on $[0, 1]$. The link between Kendall's tau and the copula C of the two random variables X and Y is given as

$$\tau_{X,Y} = \iint_{\mathbb{I}^2} \mathrm{C}(u, v)\mathrm{dC}(u, v) - 1. \tag{1.5}$$

Demonstration of (1.5) is given in details in Nelsen (2006), page 159.

Table 1.2 demonstrates the relationship between Kendall's tau and some well-known copulas. Note that there is no closed-form to reach the $\tau_{X,Y}$ for the Frank copula, and it is computed by statistical software such as R.

Measure	Gaussian $-1 < \delta < 1$	Gumbel $(\delta \geq 1)$	Clayton $(\delta > 0)$	Frank $(\delta > 0)$
$\tau_{X,Y}$	$\frac{2}{\pi}\arcsin(\delta)$	$1 - \frac{1}{\delta}$	$\frac{\delta}{\delta+2}$	$1 + \frac{4(D_1(\delta)-1)}{\delta}$

Table 1.2 Relationship between Kendall's tau and copula parameter for different copulas. Note that, $D_1(\delta) = \frac{1}{\delta}\int_0^{\delta} \frac{\mathrm{t}}{1-\mathrm{e}^{\mathrm{t}}}\mathrm{dt}$.

1.4 Families of copulas

In this section, we introduce the two most important families of copulas: Elliptical and Archimedean copulas.

1.4.1 Elliptical copulas

Two important Elliptical copulas are the Gaussian and the Student-t copulas.

Gaussian copula: Let $\Phi_2(x, y; \rho)$ be a bivariate standard normal distribution function with correlation coefficient ρ and let $\Phi(x)$ and $\Phi(y)$ be univariate standard normal margin. Then, according to the inversion method, the general form of the Gaussian (Normal) copula is introduced as follows (Briollais & Durrieu, 2014):

$$C(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v)), \quad (u, v) \in [0, 1]^2.$$

```
## Warning in library(package, lib.loc = lib.loc, character.only =
TRUE, logical.return = TRUE, :  there is no package called 'copula'
## Error in ellipCopula("normal", param = 0.8, dim = 2, dispstr =
"un"):  could not find function "ellipCopula"
## Error in persp(normal, pCopula, theta = 30, phi = 30, expand =
0.8, col = "yellow", :  object 'normal' not found
## Error in persp(normal, dCopula, theta = 65, phi = 30, expand =
0.8, col = "yellow", :  object 'normal' not found
## Error in contour(normal, dCopula, col = "red", main = "Contour
diagram for Gaussian's density"):  object 'normal' not found
## Error in rCopula(2000, normal):  could not find function
"rCopula"
```

Figure 1.4: The Gaussian copula and its density function for $\rho = 0.8$. The corresponding Kendall's tau is $\tau_{X,Y} = 0.59$.

Remember that if X and Y have a bivariate normal distribution function, we can also conclude that their marginal distribution functions are normal distributions. The inverse of this property is not true. So, if the marginal distribution function of X and Y are normally distributed, we cannot conclude that the joint distribution function of X and Y is bivariate normal distribution (Casella & Berger, 2002).

```
## Warning in library(package, lib.loc = lib.loc, character.only =
TRUE, logical.return = TRUE, :  there is no package called 'copula'
## Warning in library(package, lib.loc = lib.loc, character.only =
TRUE, logical.return = TRUE, :  there is no package called 'gumbel'
## Error in ellipCopula("t", param = 0.8, dim = 2, df = 2, dispstr
= "un"):  could not find function "ellipCopula"
## Error in diff(y):  object 'pCopula' not found
## Error in diff(y):  object 'dCopula' not found
## Error in diff(y):  object 'dCopula' not found
## Error in rCopula(2000, t):  could not find function "rCopula"
```

Figure 1.5: The Student-t copula with 2 degrees of freedom and its density function for $\rho = 0.8$. The corresponding Kendall's tau is $\tau_{X,Y} = 0.59$.

For example, let $F_x = \Phi(x)$ and $G_y = \Phi(y)$ be standard normal distribution functions. So, any joint distribution function H, which is equal to a copula $C(\Phi(x), \Phi(y))$, is not a Gaussian copula. It can be used to illustrate that H is not bivariate Gaussian distribution but F and G are univariate normal distributions.

Figure 1.4 shows the Gaussian copula and its density for $\rho = 0.8$. Also, we have presented a complete picture of this copula by contour diagram and scatter plot. Note that ρ measure dependency between the random variables.

Student-t copula: Student-t copula is defined as

$$C(u, v; \rho, \nu) = t_{\rho,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v)), \quad (u, v) \in [0, 1]^2,$$

where t_ν is the distribution function of Student with ν degree of freedom, ρ is correlation coefficient and $t_{\rho,\nu}$ is bivariate Student-t function (Mai & Scherer 2014, p56). If ρ and ν are close to 0 then we fall back in the independence case.

In Figure 1.5 the Student-t copula with 2 degrees of freedom and its density for $\rho = 0.8$ are presented by contour diagram and scatter plot.

1.4.2 Archimedean copula

In this section, we discuss Archimedean copulas. As earlier explained, generally, an Archimedean copula is defined by the following relationship:

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)),$$

where φ is generator function. The well-known Archimedean copulas are the Gumbel, Clayton, Frank, and Joe copulas.

Gumbel copula: the Gumbel copula is generated using the generator function, $\varphi_\delta(t) = (-\ln t)^{\frac{1}{\delta}}$ and its corresponding inverse function, $\varphi_\delta^{-1}(t) = exp(-(t))^{\frac{1}{\delta}}$. Thus, the Gumbel copula is derived as follows:

$$C_\delta(u,v) = \exp(-[(-\ln u)^\delta + (-\ln v)^\delta]^{\frac{1}{\delta}}), \quad \delta \in [1,\infty).$$

The Gumbel copula's measure of dependence depends only on the parameter delta $\delta \in [1,\infty)$. If δ is equal to 1, then we have the independence case (Bernard & Czado, 2015). The Gumbel copula is useful for modeling strong upper and weak lower tail dependence. The notion of tail dependence relates to the study of the association between random variables X and Y at extreme values in the upper-right and lower-left quadrants, which are known as upper and lower tail dependence, respectively. In Figure 1.6 the Gumbel copula with $\delta = 2$ and its density function presented with the 3D plot, contour diagram, and scatter plot. These figures are generated by the "copula" package in R.

Clayton copula: the Clayton generator function is $\varphi_\delta(t) = \frac{1}{\delta}(t^{-\delta} - 1)$. The domain of the Clayton copula's parameter is $\delta \in [-1,\infty) \setminus \{0\}$. By using this generator function, the Clayton copula can be given as

$$C_\delta(u,v) = [\max(u^{-\delta} + v^{-\delta} - 1, 0)]^{-\frac{1}{\delta}}.$$

Small values of δ imply the independence case. The Clayton copula is suitable for modeling lower tail dependence. Figure 1.7 illustrates the Clayton copula and its density for $\delta = 2$.

Frank copula: The Frank copula's generator function is $-ln\left(\frac{e^{-\delta t}-1}{e^{-\delta}-1}\right)$, where $\delta \in (-\infty, +\infty) \setminus \{0\}$. It follows that the Frank copula is defined as

$$C_\delta(u, v) = -\frac{1}{\delta} \ln\left(1 + \frac{(e^{-\delta u} - 1)(e^{-\delta v} - 1)}{e^{-\delta} - 1}\right).$$

```
## Error in archmCopula("gumbel", 2):  could not find function
"archmCopula"
## Error in persp(gumbel, pCopula, theta = 30, phi = 30, expand =
0.8, col = "yellow", :  object 'gumbel' not found
## Error in persp(gumbel, dCopula, theta = 30, phi = 30, expand =
0.8, col = "yellow", :  object 'gumbel' not found
## Error in contour(gumbel, dCopula, col = "red", main = "Contour
diagram for Gumbel's density"):  object 'gumbel' not found
## Error in rCopula(2000, gumbel):  could not find function
"rCopula"
```

Figure 1.6: The Gumbel copula and its density function for $\delta = 2$. As scatter plot shows it is useful to model strong upper and weak lower tail dependence. The corresponding Kendall's tau is $\tau_{X,Y} = 0.5$.

Figure 1.8 presents the Frank copula for $\delta = 2$. This copula can model lower and upper tail independence. Of note, Clayton and Gumbel copula allow only for positive modeling association. On the other hand, Frank copula can model both positive and negative associations, and this is a valuable property of the Frank copula.

In the next chapter, we give definitions, crucial properties, and some brief quantile regression concepts.

```
## Warning in library(package, lib.loc = lib.loc, character.only =
TRUE, logical.return = TRUE, :  there is no package called 'copula'
## Warning in library(package, lib.loc = lib.loc, character.only =
TRUE, logical.return = TRUE, :  there is no package called 'gumbel'
## Error in archmCopula("clayton", 2):  could not find function
"archmCopula"
## Error in persp(clayton, pCopula, theta = 30, phi = 30, expand =
0.8, col = "yellow", :  object 'clayton' not found
## Error in persp(clayton, dCopula, theta = 50, phi = 30, expand =
0.8, col = "yellow", :  object 'clayton' not found
## Error in contour(clayton, dCopula, col = "red", main = "Contour
diagram for Clayton's density"):  object 'clayton' not found
## Error in rCopula(2000, clayton):  could not find function
"rCopula"
```

Figure 1.7: The Clayton copula and its density function for $\delta = 2$. The Clayton copula is adequate for modeling strong lower tail dependence. The corresponding Kendall's tau is $\tau_{X,Y} = 0.5$.

```
## Warning in library(package, lib.loc = lib.loc, character.only =
TRUE, logical.return = TRUE, :  there is no package called 'copula'
## Warning in library(package, lib.loc = lib.loc, character.only =
TRUE, logical.return = TRUE, :  there is no package called 'gumbel'
## Error in archmCopula("frank", 2):  could not find function
"archmCopula"
## Error in persp(frank, pCopula, theta = 30, phi = 30, expand =
0.8, col = "yellow", :  object 'frank' not found
## Error in persp(frank, dCopula, theta = 130, phi = 30, expand =
0.8, col = "yellow", :  object 'frank' not found
## Error in contour(frank, dCopula, col = "red", main = "Contour
diagram for Frank's density"):  object 'frank' not found
## Error in rCopula(2000, frank):  could not find function
"rCopula"
```

Figure 1.8: The Frank copula and its density function for $\delta = 2$. The corresponding Kendall's tau is $\tau_{X,Y} = 0.21$.

CHAPTER II

QUANTILE REGRESSION

Quantile regression is a statistical approach capable to estimate conditional quantile functions (Koenker, 2012). The method is based on minimizing a hinge loss function to estimate the conditional quantile function at different locations, which provides a global overview of the distribution of a dependent variable.

Robustness to outliers, handling heteroskedasticity, flexible distribution assumptions, capturing a complete picture of the relationship between a response variable and predictors are essential characteristics of the quantile regression.

In this chapter, firstly, we describe a basic concept of quantile regression in Example 2.1. Then, quantile function, check function, multiple quantile regression, basic properties, and quantile regression application are also discussed in the upcoming sections.

Example 2.1. The information of 60 Households in 1995 has been collected. The data consist of weekly household's consumption and their weekly incomes disposable (Gujarati, 1995). Suppose that we want to investigate the relationship between weekly household's consumption and their weekly income. We consider weekly consumption as the dependent variable (Y) and weekly income as an independent variable (X). In other words, we are going to predict householder weekly consumption by knowing their weekly incomes.

```
Warning:  Computation failed in `stat_quantile()`:
Package `quantreg` required for `stat_quantile`.
Please install and try again.
Warning:  Computation failed in `stat_quantile()`:
Package `quantreg` required for `stat_quantile`.
Please install and try again.
Warning:  Computation failed in `stat_quantile()`:
Package `quantreg` required for `stat_quantile`.
Please install and try again.
```

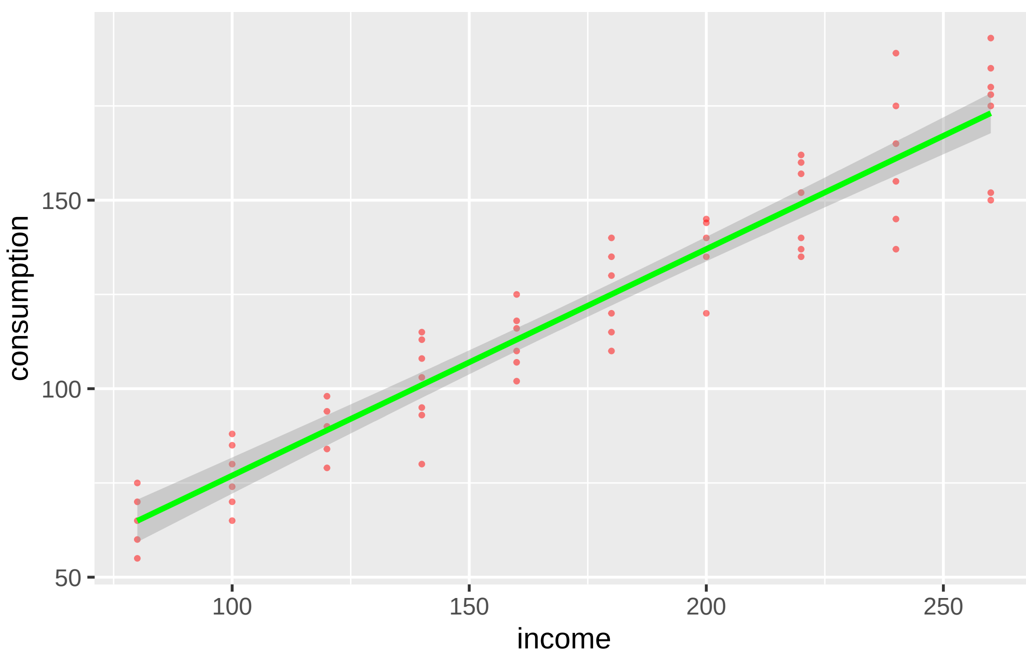

Figure 2.1: This Figure illustrates the ordinary least-squares (OLS) regression and three quantile regression lines in 20% (solid blue line), 50% (dashed line), 90% (dotted line). The strong green line (OLS) is conditional means of Y at different level of x.

Figure 2.1 provides the ordinary least-squares (OLS) regression line (the strong green line) and three quantile regression lines at 20% (solid blue line), 50% (dashed line), and 90% (dotted line). The OLS regression line tends to explain the middle of data and it also passes through $(\bar{x}, \hat{\beta}_0 + \hat{\beta}_1\bar{x})$. It is the conditional mean of Y in different levels of x. The quantile regression lines explain how the 20 percent (50 and 90) of the consumption (dependent variable) increases with income (independent variable). The dashed line is the conditional median of Y. Moreover, the three quantile lines can explain the behavior of the data in the lower, center, and upper locations of the dependent variable.

In Example 2.1, the residuals are normally distributed. Thus, as Figure 2.1 shows, there are striking resemblances between median quantile or OLS regression line. In fact, when X and Y follow a bivariate normal distribution, the conditional distribution of $Y|X = x$ is normal with constant variance; also, the conditional mean function $E(Y|X = x)$ is linear in the conditioning variable X. If we intend to investigate relationship between Y and X and the distribution of the dependent variable is not normal or there are outliers in the data, the conditional mean cannot provide an optimal model. In such situations, the conditional median, or in general, the conditional quantiles can be appropriate choices.

2.1 Quantile function

The empirical θ^{th} quantile for a continuous distribution function is the $n\theta$ order statistic of the sample $Y_1, Y_2, \cdots, Y_n$. In other words, θ^{th} quantile divides the data set into two parts such that $100 \times \theta$ percent of values lies below and $100 \times (1 - \theta)$ percent of values lies above it.

In mathematical statistics, any real-valued random variable Y can be defined by its cumulative distribution function $F(y) = P(Y \leq y) = \theta$. The quantile function returns the value y such that $F(y) = \theta$. Additional definition of quantile function

(θ^{th} quantile of Y), which is based on the generalized inverse of F, is defined as

$$Q_Y(\theta) := F^{-1}(\theta) = \inf\{y : F(y) \geq \theta\}, \quad 0 < \theta < 1. \tag{2.1}$$

An important characteristic of the quantile is equivariance to monotone transformations. Formally, let $h(.)$ be a nondecreasing function on $\mathbb{R}$ and Y be a random variable with quantile function $Q_Y(\theta)$. Then, we have

$$Q_{h(Y)}(\theta) = h(Q_Y(\theta)).$$

2.2 Check function and optimization

Let Y be a random variable with cumulative distribution function F. The θ^{th} quantile is the solution to the following optimization problem

$$\arg\min_{m\in\mathbb{R}} \left\{ \mathbb{E}(\rho_\theta(Y - m)) = \int \rho_\theta(y - m)dF(y) \right\}, \tag{2.2}$$

where $\rho_\theta(.)$ is the check function defined as follows

$$\rho_\theta(u) = u(\theta - \mathbb{I}(u < 0)) = \begin{cases} \theta u, & \text{if } u \geq 0, \\ (\theta - 1)u, & \text{if } u < 0, \end{cases}$$

with $\theta \in (0, 1)$, $\mathbb{I}(u < 0)$ is 1 for $u < 0$ and 0 otherwise.

The function $\rho_\theta(u)$ is described as a two-piecewise linear function and is illustrated in Figure 2.2 for $\theta = 0.3$, 0.6, 0.9. We can demonstrate that the expected loss in (2.2) attains its minimum at $\hat{m}$, such that $F(\hat{m}) = \theta$. To do so, we can write

$$\begin{aligned} \mathbb{E}(\rho_\theta(Y - m)) &= (\theta - 1) \int_{-\infty}^{m} (y - m)dF(y) + \theta \int_{m}^{\infty} (y - m)dF(y) \\ &= (\theta - 1) \int_{-\infty}^{m} (y - m)f(y)dy + \theta \int_{m}^{\infty} (y - m)f(y)dy. \end{aligned} \tag{2.3}$$

By differentiating (2.3) with respect to m and setting it equals to zero, we have

$$\begin{aligned}\frac{d\,\mathbb{E}(\rho_\theta(Y-m))}{dm} &= (\theta-1)\frac{d}{dm}\int_{-\infty}^{m}(y-m)f(y)dy + \theta\frac{d}{dm}\int_{m}^{\infty}(y-m)f(y)dy \\ &= (\theta-1)\int_{-\infty}^{m}\frac{d}{dm}(y-m)f(y)dy + \theta\int_{m}^{\infty}\frac{d}{dm}(y-m)f(y)dy \\ &= (1-\theta)\int_{-\infty}^{m} f(y)dy - \theta\int_{m}^{\infty} f(y)dy \\ &= (1-\theta)\int_{-\infty}^{m} dF(y) - \theta\int_{m}^{\infty} dF(y) \\ &= F(m)-\theta.\end{aligned}$$

Thus, $F(\hat{m}) = \theta$. Consequently, the quantiles are the result of a simple optimization problem.

As F is monotone, each element of $\{y : F(y) = \theta\}$ minimizes the expected loss (Koenker 2005, p.6). When F is a continuous and strictly increasing function, it is an one-to-one function and the solution of (2.2) is unique and $\hat{m} = F^{-1}(\theta)$.

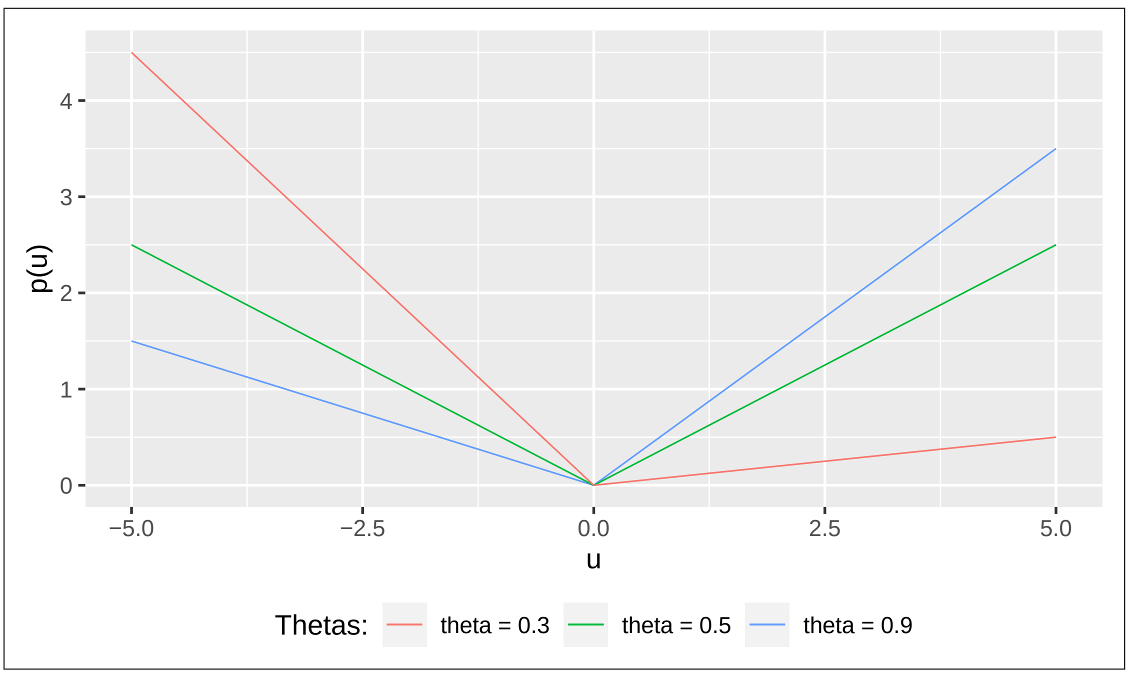

Figure 2.2: Check function for $\theta = 0.3$ (red line), $\theta = 0.5$ (green line line) and $\theta = 0.9$ (blue line).

Given a sample $y_1, y_2, \ldots, y_n$, the empirical version of the optimization problem (2.2) is

$$\hat{m} = \arg\min_{m\in\mathbb{R}} \int \rho_\theta(y-m)dF_n(y),$$

where $F_n(y)$ is empirical cdf of Y. It is equivalent to

$$\hat{m} = \arg\min_{m\in\mathbb{R}} \sum_{i}^{n} \rho_\theta(y_i - m). \tag{2.4}$$

For instance, the sample median solves the optimization problem

$$\hat{Q}_Y(0.5) = \arg\min_{m\in\mathbb{R}} \sum_{i}^{n} |y_i - m|.$$

The sample mean solves the optimization problem

$$\hat{\mu} = \bar{y} = \arg\min_{\mu\in\mathbb{R}} \sum_{i}^{n} (y_i - \mu)^2.$$

Example 2.2. Let Y be a discrete random variable with support $S = \{1, 2, 3, 4, 5, 6, 7\}$ and equal probability. We are going to find $\hat{m} = \hat{Q}_Y(0.6)$ by using the loss function defined in (2.4). To do so, we have

$$\hat{m} = \arg\min_{m\in S} \left\{ \left[R(m) := \frac{(\theta-1)}{7} \sum_{y_i<m} (y_i - m) + \frac{\theta}{7} \sum_{y_i\geq m} (y_i - m) \right] \right\}.$$

By considering $\theta = 0.6$, if we let $m = 5$, for instance, we have

$$\begin{aligned} R(5) &= \frac{6}{70} \sum_{i=5}^{7} (y_i - 5) - \frac{4}{70} \sum_{i=1}^{4} (y_i - 5) \\ &= \frac{6}{70}[(0+1+2)] - \frac{4}{70}[(-1-2-3-4)] = 0.83 \end{aligned}$$

Table 2.1 shows calculated values of $R(m)$ for all elements of Y.

m	1	2	3	4	5	6	7
$R(m)$	1.8	1.34	1.03	0.85	0.83	0.94	1.2

Table 2.1 Calculated values of the loss function R(m) for all elements of Y.

So, the minimum value of $R(m)$ occurs when $m = 5$. Thus, $\hat{Q}_Y(0.6) = 5$.

2.3 Quantile regression

Consider a simple regression model with response variable Y and explanatory variable X. Assume that errors are independent and identically distributed (iid) and F_ϵ is the distribution function of these errors. We can write

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i. \tag{2.5}$$

According to equivariance to monotone transformation property of the quantiles, the conditional quantile function of the variable Y given X is often denoted as $Q_Y(\theta|X = x)$ and given by

$$\begin{aligned} Q_Y(\theta|X = x) &= \beta_0 + \beta_1 x + Q_\epsilon(\theta) \\ &= \beta_0 + \beta_1 x + F_\epsilon^{-1}(\theta). \end{aligned} \tag{2.6}$$

Equation (2.6), for $0 < \theta < 1$, constructs several lines with the same slope and different intercepts. When the errors are *iid*, the conditional quantile functions build a family of parallel lines (Koenker & Bassett Jr, 1982). The vector $\beta(\theta) = (\beta_0 + F^{-1}(\theta), \beta_1)^\top$ is the lines' vector parameters and their estimates are $\hat{\beta}_0(\theta)$ and $\hat{\beta}_1(\theta)$. The parameters' estimation procedure for the vector of parameters $\beta(\theta)$, is given next.

Simple quantile regression: The relationship $Q_Y(\theta|X = x) = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x$

is defined as the θ^{th} quantile model, where

$$\hat{\beta}(\theta) = \underset{\beta(\theta)\in\mathbb{R}^2}{\arg\min}\, \mathbb{E}[\rho_\theta(Y - \beta_0(\theta) - \beta_1(\theta)x)|X = x].$$

Similarly, given $(x_i, y_i), i = 1, \ldots, n$, the vector of quantile regression coefficients, $\beta(\theta)$, can be estimated empirically by solving

$$\hat{\beta}(\theta) = \underset{\beta(\theta)\in\mathbb{R}^2}{\arg\min} \sum_{i=1}^{n} \rho_\theta(y_i - \beta_0(\theta) - \beta_1(\theta)x_i). \tag{2.7}$$

This optimization problem can be solved using linear programming techniques such as simplex algorithms (Koenker 2005, p.173).

Multiple quantile regression: An extended form of the relationship (2.7) to multiple quantile regression is given as follows

$$\begin{aligned}\hat{\beta}(\theta) &= \underset{\beta\in\mathbb{R}^p}{\arg\min}\{\sum \rho_\theta(y_i - x_i^\top \beta)\},\\ &= \underset{\beta\in\mathbb{R}^p}{\arg\min}\{\sum_{i\in y_i \geq x_i^\top \beta} \theta|y_i - x_i^\top \beta| + \sum_{i\in y_i < x_i^\top \beta} (1-\theta)|y_i - x_i^\top \beta|\}. \end{aligned} \tag{2.8}$$

Equation (2.8) consists of two terms, the first one is the penalized positive absolute error with penalty rate θ and the second one is the penalized negative absolute error with penalty rate $1 - \theta$. Note that for $\theta = 0.5$ the penalty rate of absolute errors are the same and is equal to 0.5.

```
Warning:  Computation failed in ‘stat_quantile()‘:
Package ‘quantreg‘ required for ‘stat_quantile‘.
Please install and try again.
```

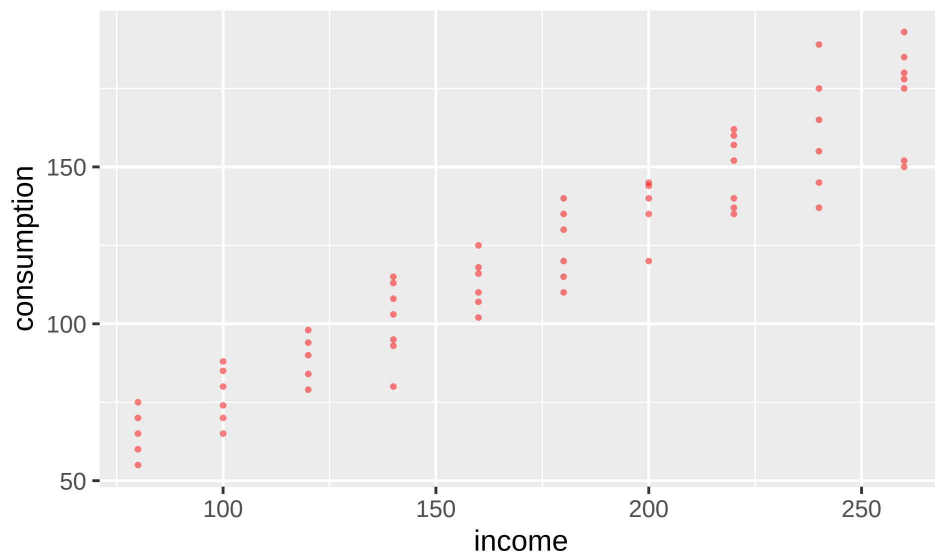

Figure 2.3: Quantile regression lines in $\theta = 0.05, 0.25, 0.75, 0.95$ for 60 Households (Example 2.1). The data consist of weekly household consumption and their weekly disposable income. The dependent variable (Y) is weekly consumptions and independent variable (X) is weekly incomes.

2.4 Basic properties

In the following, we describe the noticeable properties of quantile regression. These properties explicitly demonstrate the usefulness of quantile regression. Robustness to outliers, flexible distribution assumptions, handling heteroskedasticity, and capturing a complete view of the distribution are objectives of this section.

Property 2.1. ***Robustness to outliers***

Quantile Regression is more accurate than OLS regression when there are extreme values of the data set. Its performance is better than the OLS in presence of heavy-tailed distribution (Briollais & Durrieu, 2014). In fact, since it is based on the absolute deviation loss function, extreme values may not affect the model parameter estimates. Moreover, the OLS regression estimates the response variable's conditional mean, which characterizes the data's central tendency.

Property 2.2. ***Flexible distribution assumption***

An appealing feature of quantile regression that has been frequently highlighted is its ability to provide an opportunity to describe part of the conditional distribution without requiring global distributional assumptions (Koenker 2005, p.55). In fact, distribution assumption refers to homogeneity of the residuals' variances in the regression framework. Violation of homogeneity of variances leads to inappropriate estimated models. However, quantile regression has been shown to perform well in heteroskedastic situations.

Property 2.3. ***Capability of detecting heteroskedasticity***

Quantile regression is an effective way to detect heteroskedasticity. Consider, for instance, the location-scale shift model described as

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + (\gamma_0 + \gamma_1 X_i)\epsilon_i, \\ &= \beta_0 + \beta_1 X_i + \tilde{\epsilon}_i, \end{aligned}$$

where γ_0 and γ_1 are scale parameters and the errors ϵ_i has distribution function F_ϵ. When $\gamma_0 = 0$ and $\gamma_1 > 0$, the errors $\tilde{\epsilon}_i$ are correlated with the explanatory variable. In this case, the variances of the errors are not homogeneous and the slopes are different from one quantile to another. The true quantile regression lines $Q(\theta|X_i = x_i) = \beta_0 + (\beta_1 + \gamma_1 F^{-1}(\theta))x_i$, for different θs are not parallel and pass through the point $(0, \hat{\beta}_0)$. Testing for presence of heteroskedasticity can be formally established by testing for slopes equality at different quantiles (parallel lines). More details of such hypotheses testing are given in Koenker and Bassett (1978).

2.5 Application of quantile regression

Let us show the utility of the method with an example presented by Koenker (2012). Figure 2.4 illustrates the scatter plot of the Engel data on food expenditure (Y) vs. household income (X) for a sample of 335 nineteenth century working class Belgian households. As Figure 2.4 shows, the variation in food expenditure increases with increasing the value of income. This dispersion violates homogeneity of variance. In this situation, two outlier observations have significantly influenced the ordinary least squares regression estimator.
Table 2.2 shows the fit of OLS model and different quantile regression models for Engel data. As the Table 2.2 shows, the estimated coefficient by OLS method is larger than the estimated coefficients by the quantile regression method in lower quantiles; conversely, the estimated coefficient by OLS method is smaller than the estimated coefficients by quantile regression method at upper quantiles.

Coeffcient	**OLS**	**Q0.1**	**Q0.2**	**Q0.4**	**Q0.5**	**Q0.75**	**Q0.9**
Intercept	147.5	110	102	101	81.5	62	69
Incomes	0.5	0.4	.45	0.5	0.56	.56	0.68

Table 2.2 Quantile regression and OLS coefficient estimation.

Generally, when the estimated coefficients in lower quantiles are different to those upper quantiles, the quantile regression method is preferred to the OLS method. The coefficients' plots are useful tools to choose the regression method (Figure 2.5). Figure 2.5 illustrates the estimated parameters of the quantile regression models versus the quantiles (θs) for the Engel data.

```
Warning:  Computation failed in ‘stat_quantile()‘:
Package ‘quantreg‘ required for ‘stat_quantile‘.
Please install and try again.
Warning:  Computation failed in ‘stat_quantile()‘:
Package ‘quantreg‘ required for ‘stat_quantile‘.
Please install and try again.
```

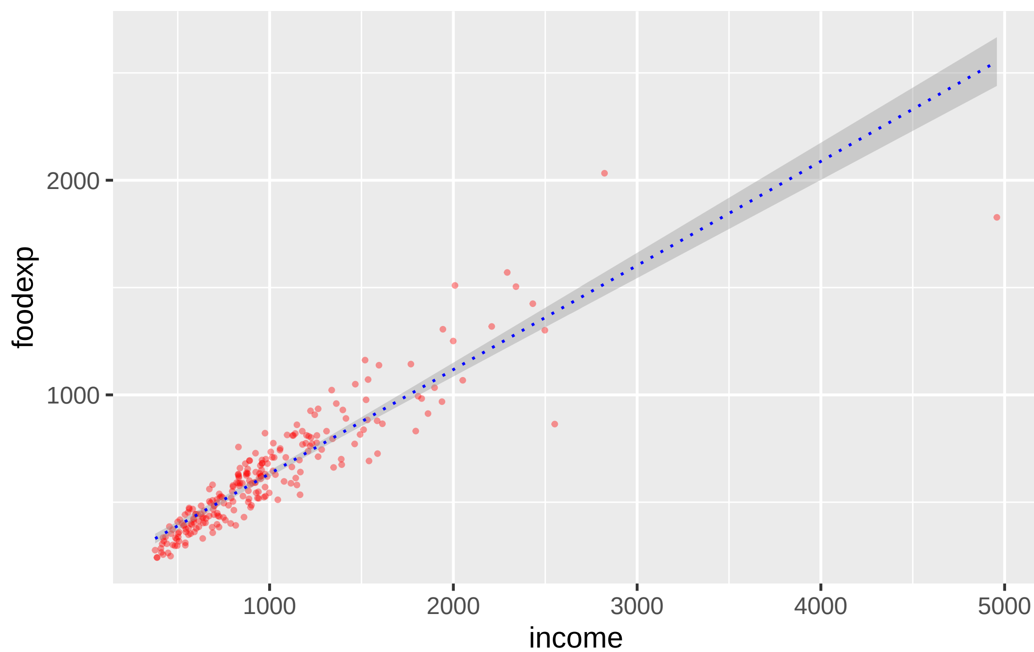

Figure 2.4: Food expenditure scatter plot with OLS regression line (dashed green) and quantile regression lines for $\theta = 0.05, 0.25, 0.5, 0.75, 0.95$. Y and X are annual Household Food Expenditure and annual Household Income in Belgian Francs, respectively. As one can see, the median quantile (dotted blue line) does not fit the OLS regression line. The figure depicted with our R code; it is similar to Figure 1 in the vignette of the quantile regression R package, "quantile regression in R: A vignette" Koenker (2012).

The red horizontal line shows the OLS regression coefficient, and red dashed lines show its confidence interval. The dark line shows the quantiles regression coefficient, and the area around shows its confidence interval. These confidence intervals are calculated by the rank inversion method described in Koenker(2005, p.91). As the right panel of Figure 2.5 shows, the lower quantiles for Incomes variable are under the red line, and the upper quantiles are above the red line. In this case, due to heteroskedasticity using the quantile regression is suggested for data analysis. One can interpret that since the estimated parameters of quantile regression models are between the red dashed lines, the quantile regression and OLS performance are the same.

```
## Error in rq(foodexp ~ income, data = engel, tau = 1:99/100):
could not find function "rq"
## Error in plot(fit1, mfrow = c(1, 2), pin = c(2, 2)):  object
'fit1' not found
```

Figure 2.5: The red horizontal line indicates the OLS regression coefficient, the red dashed lines show the confidence interval for the coefficient estimates, the dark points show the quantile regression coefficients estimates and the shaded grey area shows the corresponding confidence intervals.

The next chapter introduces copula-based quantiles which consists of conditional Elliptical and Archimedean copula quantiles.

CHAPTER III

COPULA-BASED QUANTILE

In this chapter, we outline mathematical details of the relationship between copulas and quantile functions. To do so, we demonstrate first that for a copula $C(.,.;\rho)$, which links the joint distribution H of (X, Y) to its marginals F and G, the following relationship

$$\theta = \frac{\partial C(F(x), G(y); \rho)}{\partial F(x)}, \quad \theta \in (0,1), \tag{3.1}$$

connects the copula function to the θ^{th} conditional quantile function $Q_{Y|X=x}(\theta)$. To prove this fact, we follow Sklar's theorem results in equation (1.1), which leads to

$$\begin{aligned} C(F(x), G(y)) &= H(x,y) \\ &= P(X \leq x, Y \leq y) \\ &= \int_{t \in (-\infty, x)} P(t, Y \leq y) \mathrm{d}t \\ &= \int_{-\infty}^{x} P(Y \leq y | t) f(t) \mathrm{d}t. \end{aligned} \tag{3.2}$$

Using chain-rule for the partial derivative of equation (3.2) with respect to $F(x)$, one has

$$\begin{aligned}\frac{\partial}{\partial F(x)}C(F(x),G(y)) &= \frac{\partial x}{\partial F(x)}\frac{\partial}{\partial x}\int_{-\infty}^{x} P(Y \leq y|t)f(t)\mathrm{d}t \\ &= \frac{1}{\frac{\partial F(x)}{\partial x}}\frac{\partial}{\partial x}\int_{-\infty}^{x} P(Y \leq y|t)f(t)\mathrm{d}t \\ &= \frac{1}{f(x)}f(x)P(Y \leq y|X = x) \\ &= G_{Y|X=x}(y|x). \end{aligned} \tag{3.3}$$

Now, if we assume that y is the θ^{th} conditional quantile of $Y|X = x$, then from (3.3), y satisfies

$$\frac{\partial C(F(x),G(y))}{\partial F(x)} = \theta. \tag{3.4}$$

Thus, solving for y in (3.4) leads to

$$y = G^{-1}(D^{-1}(\theta; \delta, u)),$$

where $u = F(x)$, and D^{-1} is the inverse of the function $v \mapsto D(\theta; \delta, u) = \frac{\partial C(u,v)}{\partial u}_{|u=F(x)}$. In some references the partial derivative $\frac{\partial C(u,v)}{\partial u}$ is denoted as $C_{2|1}(v|u)$ and is called an h-function (Bernard & Czado, 2015).

3.1 Conditional Gaussian-copula quantiles

In this section, we give an explicit form of the θ^{th} conditional copula quantile in the case of Gaussian copula model. Consider two random variables X and Y with marginal distribution functions F and G, respectively. Assume that (X, Y) are distributed following the bivariate Gaussian copula given by

$$\begin{aligned}H(x,y) &= C(F(x),G(y)) \\ &= \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v)|\Gamma), \quad u = F(x), v = G(y) \\ &= \int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{(1-\rho^2)}} exp\{-\frac{t^2 - 2\rho tr + r^2}{2(1-\rho^2)}\}\mathrm{d}t\mathrm{d}r, \end{aligned} \tag{3.5}$$

where Φ_2 and Φ are bivariate and univariate standard normal distribution function, respectively, and Γ is 2×2 correlation matrix, with off-diagonal element ρ. Assume that $h(t,r) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} exp\{-\frac{t^2-2\rho tr+r^2}{2(1-\rho^2)}\}$, $a = \Phi^{-1}(v)$ and $b = \Phi^{-1}(u)$. It can be showed that (Kjersti et al., 2015)

$$\begin{aligned}
C_{2|1}(u,v) &= \frac{\partial}{\partial u} C(u,v) \\
&= \frac{\partial}{\partial u} \int_{-\infty}^{a} \int_{-\infty}^{b} h(t,r) \mathrm{d}t \mathrm{d}r \\
&= \frac{\partial b}{\partial u} \frac{\partial}{\partial b} \int_{-\infty}^{a} \int_{-\infty}^{b} h(t,r) \mathrm{d}t \mathrm{d}r \\
&= \frac{1}{\phi(b)} \frac{\partial}{\partial b} \int_{-\infty}^{a} \int_{-\infty}^{b} h(t,r) \mathrm{d}t \mathrm{d}r \\
&= \frac{1}{\phi(b)} \int_{-\infty}^{a} \left[\frac{\partial}{\partial b} \int_{-\infty}^{b} h(t,r) \mathrm{d}t \right] \mathrm{d}r \\
&= \frac{1}{\phi(b)} \int_{-\infty}^{a} h(t,b) \mathrm{d}t \\
&= \frac{1}{\phi(b)} \int_{-\infty}^{a} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\{-\frac{t^2 - 2\rho tb + b^2}{2(1-\rho^2)}\} \mathrm{d}t \\
&= \frac{1}{\phi(b)} \int_{-\infty}^{a} \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\{-\frac{(t-\rho b)^2 + (b^2 - \rho^2 b^2)}{2(1-\rho^2)}\} \mathrm{d}t \\
&= \frac{1}{\phi(b)} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-b^2}{2}\right) \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(t-\rho b)^2}{2(1-\rho^2)}\right) \mathrm{d}t \\
&= \frac{1}{\phi(b)} \frac{\phi(b)}{1} \Phi\left(\frac{a-\rho b}{\sqrt{1-\rho^2}}\right) \\
&= \Phi\left(\frac{a-\rho b}{\sqrt{1-\rho^2}}\right).
\end{aligned}$$

Then, the link between x and y is defined by the following relationship

$$\theta = \Phi\left(\frac{\Phi^{-1}(v) - \rho\Phi^{-1}(u)}{\sqrt{(1-\rho^2)}}\right). \tag{3.6}$$

Thus, one has

$$\Phi^{-1}(\theta) = \frac{\Phi^{-1}(v) - \rho\Phi^{-1}(u)}{\sqrt{1-\rho^2}},$$

which leads to

$$\begin{aligned}
&\sqrt{1-\rho^2}\Phi^{-1}(\theta) = \Phi^{-1}(v) - \rho\Phi^{-1}(u) \\
&\Leftrightarrow \Phi^{-1}(v) = \rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(\theta) \\
&\Leftrightarrow v = \Phi\left(\rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(\theta)\right) \\
&\Leftrightarrow G(y) = \Phi\left(\rho\Phi^{-1}(F(x)) + \sqrt{1-\rho^2}\Phi^{-1}(\theta)\right) \\
&\Leftrightarrow y = G^{-1}\left(\Phi\left(\rho\Phi^{-1}(F(x)) + \sqrt{1-\rho^2}\Phi^{-1}(\theta)\right)\right). \qquad (3.7)
\end{aligned}$$

The relationship between x and y in (3.7) can be denoted by $y = q(x, \theta; \rho)$, which is the θ^{th} conditional of $Y|X = x$. Equation (3.7) can be expressed as a linear model if X and Y have marginal standard normal distributions (Bernard & Czado, 2015). That is, assuming $F = G = \Phi$, we have

$$\begin{aligned}
y &= G^{-1}\left(\Phi\left(\rho\Phi^{-1}(F(x)) + \sqrt{1-\rho^2}\Phi^{-1}(\theta)\right)\right) \\
&= G^{-1}\left(\Phi\left(\rho x + \sqrt{1-\rho^2}\Phi^{-1}(\theta)\right)\right) \\
&= \rho x + \sqrt{1-\rho^2}\Phi^{-1}(\theta).
\end{aligned}$$

Moreover, if X and Y have non-standard normal marginal distributions $F(x)$ and $G(y)$, with $\mu_x = \mathrm{E(X)}$, $\mu_y = \mathrm{E(Y)}$, $\sigma_x^2 = \mathrm{var(X)}$, $\sigma_y^2 = \mathrm{var(Y)}$ and ρ, then the relationship between X and Y in (3.7) can be again reduced to the following linear model

$$y = q(x, \theta; \rho) = b_0 + b_1 x, \qquad (3.8)$$

where $b_0 = (\mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x) + \sigma_y\sqrt{(1-\rho^2)}\Phi^{-1}(\theta)$ and $b_1 = \rho\frac{\sigma_y}{\sigma_x}$.
In conclusion, if (X, Y) are normally distributed, the θ^{th} conditional quantile of $Y|X = x$ is reduced to a line with only the intercept that depends on θ. In other words, the quantile functions at different locations are parallel lines. This is equivalent to the standard quantile regression model given in (2.6).

Figure 3.1 shows the true 1^{st} (red), 2^{nd} (green) and 3^{rd} (blue) quartiles; also, 5^{th} (black) and 95^{th} (cadet blue) percentiles for the conditional Gaussian-copula quantile model; these lines are parallel.

Now, if one considers chi-square marginal distribution functions with 5 and 9 degrees of freedom for X and Y, respectively, then the conditional Gaussian-copula quantile with $\rho = 0.8$ is not a line anymore. This fact is presented in Figure 3.2.

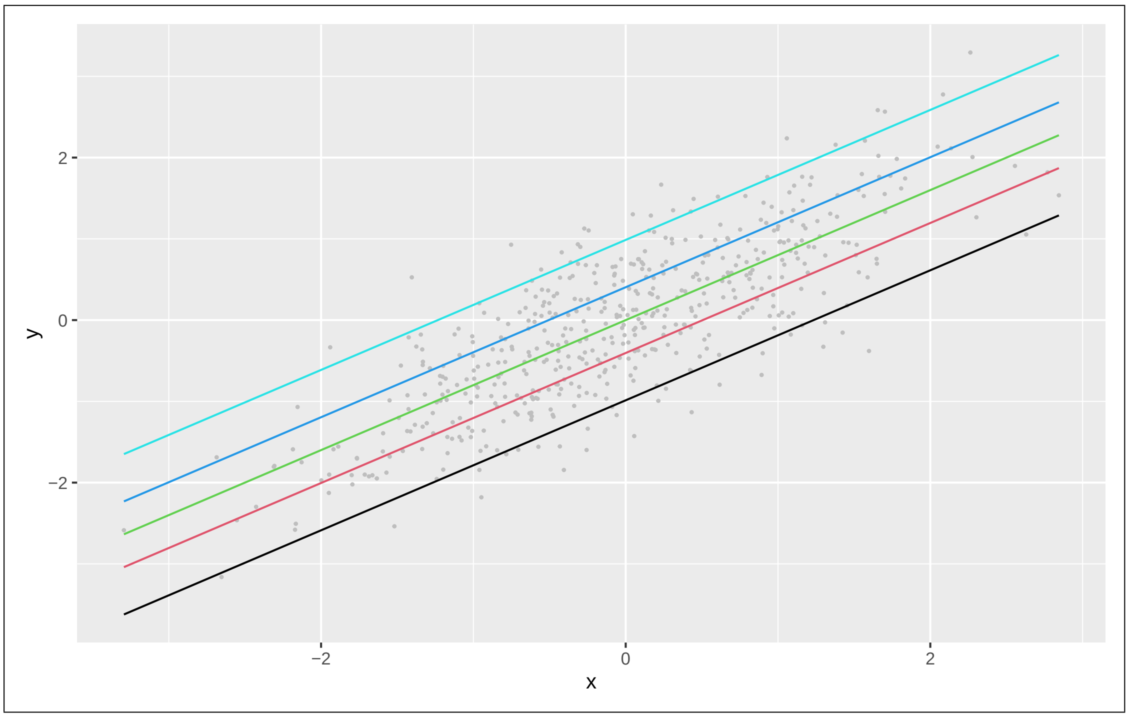

Figure 3.1: Conditional Gaussian-copula quantile curves (lines) for 1^{st} (red), 2^{nd} (green) and 3^{rd} (blue) quartiles; also, 5^{th} (black) and 95^{th} (cadet blue) percentiles. The data points in gray (n = 500), are generated from Gaussian copula model with $\rho = 0.8$. X and Y have bivariate Normal distribution function with normal margins as well.

```
Error in library(quantreg):  there is no package called 'quantreg'
```

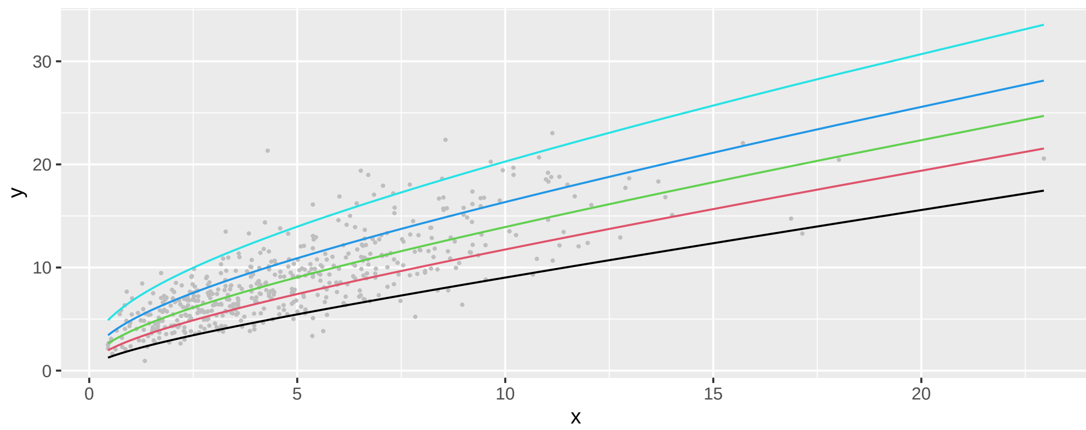

Figure 3.2: True conditional Gaussian-copula quantile curves for 1^{st} (red), 2^{nd} (green) and 3^{rd} (blue) quartiles; also, 5^{th} (black) and 95^{th} (cadet blue) percentiles. The data points in gray (n = 500), are generated from Gaussian copula model with $\rho = 0.8$. X and Y have Chi-Square marginal distribution functions with 5 and 9 degrees of freedom, respectively.

```
Error in library(quantreg):  there is no package called 'quantreg'
```

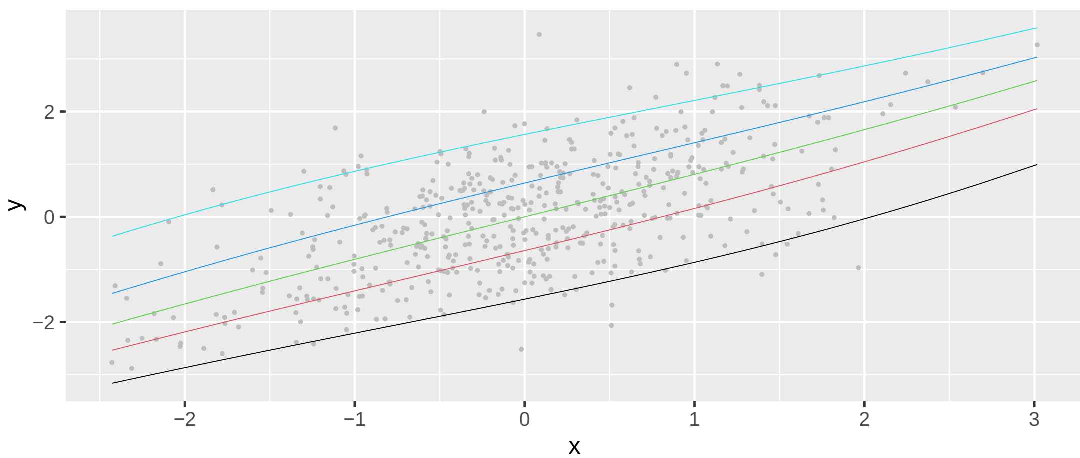

Figure 3.3: True conditional Student-t-copula quantile curves (with 10 df) for 1^{st} (red), 2^{nd} (green) and 3^{rd} (blue) quartiles; also, 5^{th} (black) and 95^{th} (cadet blue) percentiles. The data points in gray (n = 500), are generated from relationship (3.9) for $\rho = 0.8$. X and Y have standard normal marginal distribution functions.

3.1.1 Conditional Student-t-copula quantiles

For two random variables X and Y with marginal distribution functions $u = F(x)$, $v = G(y)$, the student-t copula is defined as

$$C(u, v) = t_{\rho,\nu}\left(t_\nu^{-1}(u), t_\nu^{-1}(v)\right), \tag{3.9}$$

where t_ν is the cumulative distribution function of the Student-t with ν degree of freedom, $t_{\rho,\nu}$ is the cumulative distribution function of the bivariate Student-t and $\rho \in [-1, 1]$ is the correlation coefficient between X and Y. The conditional Student-t-copula quantiles is introduced as (Mai & Scherer, 2014)

$$\theta = t_{\nu+1}\left(\frac{t_\nu^{-1}(v) - \rho t_\nu^{-1}(u)}{\sqrt{\left(\frac{(\nu+(t_\nu^{-1}(u))^2)(1-\rho^2)}{\nu+1}\right)}}\right). \tag{3.10}$$

By transforming the relationship (3.10) with respect to v, the link between x and y is given as follows

$$\begin{aligned}
\theta &= t_{\nu+1}\left(\frac{t_{\nu}^{-1}(v) - \rho t_{\nu}^{-1}(u)}{\sqrt{(\frac{(\nu+(t_{\nu}^{-1}(u))^2)(1-\rho^2)}{\nu+1})}}\right) \\
\Leftrightarrow t_{\nu}^{-1}(v) &= t_{\nu+1}^{-1}(\theta)\sqrt{(\frac{(\nu+(t_{\nu}^{-1}(u))^2)(1-\rho^2)}{\nu+1})} + \rho t_{\nu}^{-1}(u) \\
\Leftrightarrow v &= t_{\nu}\left(t_{\nu+1}^{-1}(\theta)\sqrt{(\frac{(\nu+(t_{\nu}^{-1}(u))^2)(1-\rho^2)}{\nu+1})} + \rho t_{\nu}^{-1}(u)\right) \\
\Leftrightarrow y &= G^{-1}\left[t_{\nu}\left(t_{\nu+1}^{-1}(\theta)\sqrt{(\frac{(\nu+(t_{\nu}^{-1}(F(x)))^2)(1-\rho^2)}{\nu+1})} + \rho t_{\nu}^{-1}(F(x))\right)\right].
\end{aligned}$$

Illustration of conditional Student-t-copula quantile for $\rho = 0.8$ is given in Figure 3.3.

3.2 Conditional Archimedean-copula quantiles

Earlier, we explained that when the marginal distributions are not normal, the conditional Gaussian-copula quantiles are not parallel lines anymore. This fact leads us to copula-based quantile regression, which is a class of nonlinear quantile regression. Here, we continue the investigation of copula-based quantile regression for Archimedean copulas.

Referring to Chapter 1, an Archimedean copula is generally defined by

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)), \tag{3.11}$$

where φ is the generator function of this copula, presented in Table 1.1. The Gumbel, Clayton, Frank and Joe copulas are well-known Archimedean copulas. These copulas are the focus of this section. Equation (3.11) can be rewritten as

$$\begin{aligned}
C(u, v) &= \varphi^{-1}(\varphi(u) + \varphi(v)) \\
\Leftrightarrow \varphi(C(u, v)) &= \varphi(u) + \varphi(v). \qquad (3.12)
\end{aligned}$$

Therefore, by differentiating (3.12) with respect to u, we have

$$\frac{\partial\left[\varphi(C(u,v))\right]}{\partial u}=\frac{\partial\left[\varphi(u)+\varphi(v)\right]}{\partial u}$$
$$\Leftrightarrow \varphi'(C(u,v))\frac{\partial C(u,v)}{\partial u}=\varphi'(u)$$
$$\Leftrightarrow \frac{\partial C(u,v)}{\partial u}=\frac{\varphi'(u)}{\varphi'(C(u,v))}.$$

Considering $\theta=\frac{\partial C(u,v)}{\partial u}$ and solving for v leads to

$$v=\varphi^{-1}\left[\varphi\left(\varphi'^{-1}\left(\frac{1}{\theta}\varphi'(u)\right)\right)-\varphi(u)\right]. \tag{3.13}$$

Equation (3.13) is generally introduced as the conditional copula quantile curves for Archimedean copulas. By replacing $u=F(x)$ and $v=G(y)$, next equation gives the link between x and the θ^{th} conditional quantile $y=q(x,\theta;\delta)$

$$q(x,\theta;\delta)=y=G^{-1}\left[\varphi^{-1}\left\{\varphi\left(\varphi'^{-1}(\frac{1}{\theta}\varphi'(F(x)))-\varphi(F(x))\right)\right\}\right]. \tag{3.14}$$

Next sections illustrate (3.14) for Frank, Clayton, Gumbel and Joe copula models.

3.2.1 Conditional Frank-copula quantiles

Consider (X,Y) distributed following Frank copula with margins F and G respectively. The Frank copula with parameter δ, given by Table 1.1, is defined as follows

$$C(u,v:\delta)=-\frac{1}{\delta}\ln\left(1+\frac{(e^{-\delta u}-1)(e^{-\delta v}-1)}{e^{-\delta}-1}\right),\quad \delta\in\mathbb{R}\setminus\{0\}. \tag{3.15}$$

Frank's generator function is defined as $\varphi_\delta(t)=-ln(\frac{e^{-\delta t}-1}{e^{-\delta}-1})$. Using (3.14), the conditional Frank-copula quantiles are defined as

$$y=G^{-1}\left(\frac{-1}{\delta}\ln(1+\frac{e^{-\delta}-1}{1+e^{-\delta F(x)}(\theta^{-1}-1)})\right). \tag{3.16}$$

For a specific δ, the model (3.16) can provide a family of conditional Frank-copula quantile curves, depending on different $\theta\in(0,1)$.

```
Error in library(quantreg):  there is no package called 'quantreg'
```

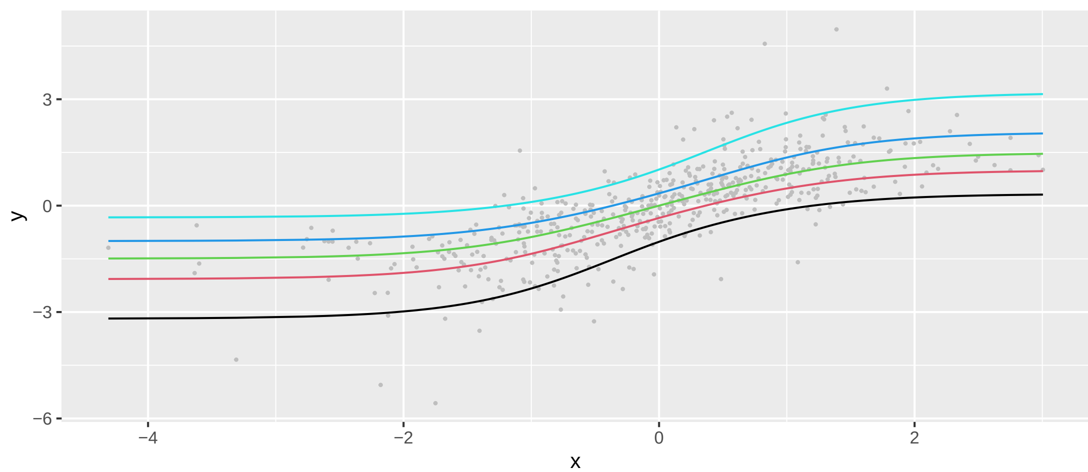

Figure 3.4: True conditional Frank-copula quantile curves for $\delta = 8$ when X and Y have Student-t marginal distributions with 8 degrees of freedom. The data point in gray (n = 500), are generated from model (3.15). The curves are shown for different θ =0.05 (black), 0.25 (red), 0.50 (green), 0.75 (blue), 0.95 (cadet blue).

Figure 3.4 illustrates the true conditional Frank-copula quantiles with $\delta = 8$ for 1^{st}, 2^{nd} and 3^{rd} quartiles; also, 5^{th} and 95^{th} percentiles are illustrated. X and Y follow the Student-t marginal distributions with 8 degrees of freedom. Figure 3.4 shows also 500 pairs of observations coming from the Frank copula model (3.15). As one can see again, the quantile curves are not lines.

3.2.2 Conditional Clayton-copula quantiles

Now, let (X, Y) distributed following Clayton copula with margins F and G, respectively. The Clayton copula and its generator function are defined as follows

$$C(u, v : \delta) = [\max\{(u^{-\delta} + v^{-\delta} - 1), 0\}]^{-\frac{1}{\delta}} \quad \delta \in [-1, \infty) \setminus \{0\}, \tag{3.17}$$

and

$$\varphi_\delta(t) = \frac{1}{\delta}(t^{-\delta} - 1).$$

Using the generator function and equation (3.14), the conditional Clayton-copula quantiles are presented as

$$y = G^{-1}\left[\left\{(\theta^{\frac{-\delta}{1+\delta}} - 1)F(x)^{-\delta} + 1\right\}^{\frac{-1}{\delta}}\right], \tag{3.18}$$

where δ is the copula parameter and $\theta \in (0, 1)$.

Figure 3.5 illustrates conditional Clayton-copula quantile curves with $\delta = 2$ for 1^{st}, 2^{nd} and 3^{rd} quartiles; 5^{th} and 95^{th} percentiles are also plotted. These curves are the true conditional Clayton-copula quantiles when X and Y have standard Normal marginal distributions. Figure 3.5 Plotted also 500 pairs observations generated from (3.17). Again, the curves are not linear lines.

3.2.3 Conditional Gumbel and Joe-copula quantiles

Let (X, Y) distributed following Gumbel and Joe copula with margins F and G, respectively. Although the first partial derivative of the Gumbel and Joe copulas $C_{2|1}(v|u)$ has an explicit form, it is hard to compute analytically its inverse function $C_{2|1}^{-1}(\theta|u)$ with respect to v. Therefore, in equation (3.14), $(\varphi')^{-1}$ does not have closed form and we have to use numerical programming for computing the conditional Gumbel and Joe-copula quantiles.

```
Error in library(quantreg):  there is no package called 'quantreg'
```

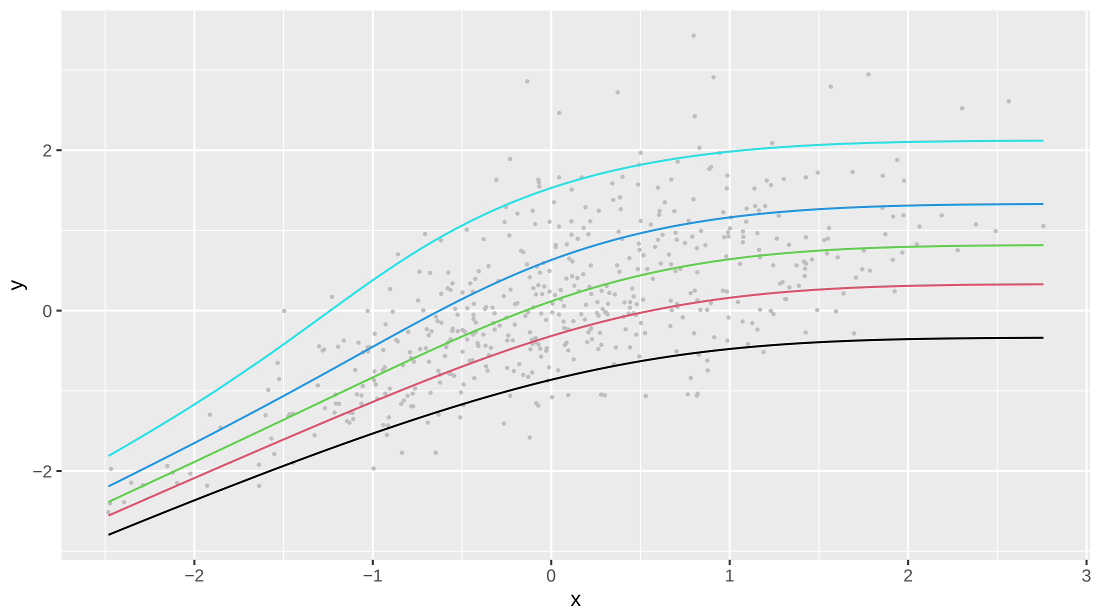

Figure 3.5: True conditional Clayton-copula quantile curves for $\delta = 2$ when X and Y have standard Normal marginal distributions. The data (n = 500) are generated from relationship (3.17). These curves are illustrating for $\theta = 0.05$ (black), 0.25 (red), 0.50 (green), 0.75 (blue), 0.95 (cadet blue).

The Gumbel copula: $C(u, v)$ introduced in Table 1.1 has generator function $\varphi_\delta(t) = (-ln\, t)^{\frac{1}{\delta}}$. The first partial derivative of $C(u, v)$, with respect to u, is defined as the conditional Gumbel-copula quantile:

$$\theta = \frac{1}{u}\left(-\ln(u)\right)^{\delta-1}((-\ln(u)^{\delta} + (-\ln(v))^{\delta}\right)^{\frac{1-\delta}{\delta}} C(u, v), \quad \delta \geq 1. \tag{3.19}$$

By solving numerically equation (3.19) with respect to v the conditional Gumbel-copula quantile curves for $\theta = 0.05, 0.25, 0.50, 0.75$ and 0.95 are calculated and presented in Figure 3.6. X and Y are distributed following standard normal distribution functions.

```
Error in library(VineCopula):  there is no package called
'VineCopula'
Error in BiCop(4, 2):  could not find function "BiCop"
Error in BiCopHinv1(u[i], t[i], gu.cop):  could not find function
"BiCopHinv1"
Error in data.frame(x, y):  arguments imply differing number of
rows:  500, 0
Error in library(VineCopula):  there is no package called
'VineCopula'
Error in BiCop(4, 2):  could not find function "BiCop"
Error in BiCopHinv1(u[i], tau, gu.cop):  could not find function
"BiCopHinv1"
```

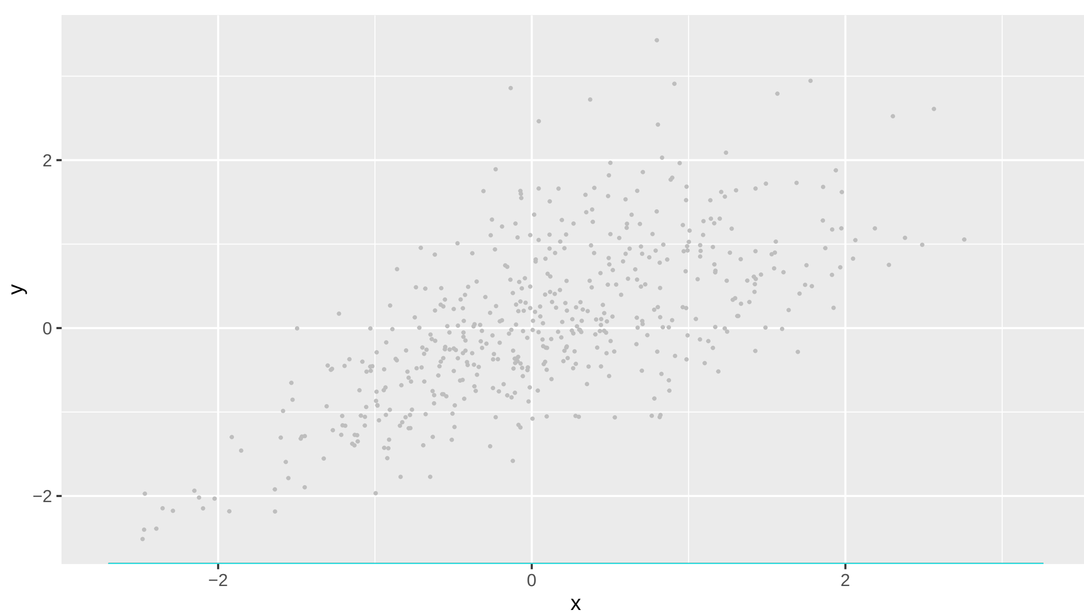

Figure 3.6: True conditional Gumbel-copula quantile curves are illustrated for $\delta = 2$ when X and Y have standard Normal distribution functions. These curves calculated for 1^{st} (red), 2^{nd} (green) and 3^{rd} (blue) quartiles; also, 5^{th} (black) and 95^{th} (cadet blue) percentiles. The data points in gray (n = 500), are generated from Gumbel copula model.

The Joe copula: Considering the Joe copula defined in Table 1.1 and its first partial derivative with respect to u, the conditional Joe-copula quantiles are specified as

$$\theta = (1 - C(u,v))^{1-\delta}(1-u)^{\delta-1}(1-(1-v)^{\delta}), \quad \delta \geq 1. \tag{3.20}$$

Figure 3.7 shows conditional Joe copula quantile curves for different θ =0.05 (black), 0.25 (red), 0.50 (green), 0.75 (blue), 0.95 (cadet blue). $\delta = 2.8$ and X and Y have standard Normal marginal distributions. To draw this Figure, we have used an algorithm similar to the one used for the Gumbel copula.

```
Error in library(VineCopula):  there is no package called
'VineCopula'
Error in BiCop(6, theta):  could not find function "BiCop"
Error in BiCopHinv1(u[i], t[i], gu.cop):  could not find function
"BiCopHinv1"
Error in data.frame(x, y):  arguments imply differing number of
rows:  500, 0
Error in library(VineCopula):  there is no package called
'VineCopula'
```

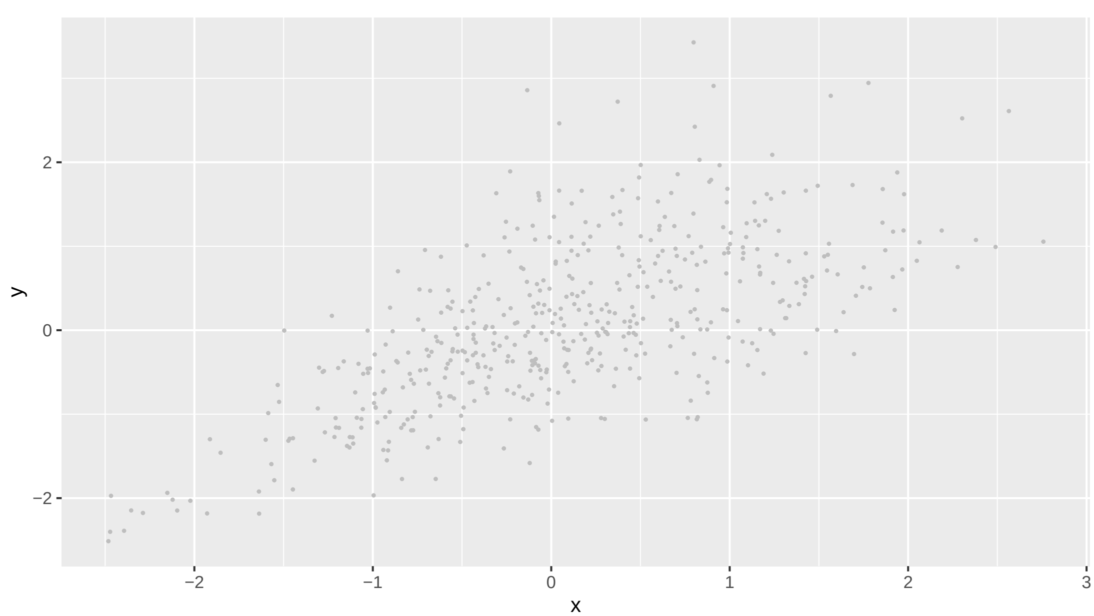

Figure 3.7: True conditional Joe-copula quantile curves are presented for $\delta = 2.8$ when X and Y have standard Normal distribution functions. These curves calculated for 1^{st} (red), 2^{nd} (green) and 3^{rd} (blue) quartiles; also, 5^{th} (black) and 95^{th} (cadet blue) percentiles. The data points in gray (n = 500), are generated from Joe copula model.

3.3 Model parameters estimation

In quantile regression, we minimize the loss function to estimate the vector of parameters $\beta(\theta)$. In nonlinear quantile regression, analogously we are willing to estimate the dependence parameter δ by minimizing the following loss function.

$$\hat{\delta}(\theta) = \arg\min\{\sum \rho_\theta(y_i - f_\delta(x_i))\}, \tag{3.21}$$

where ρ_θ is the check function given in (2.2) and f_θ is an unknown function.

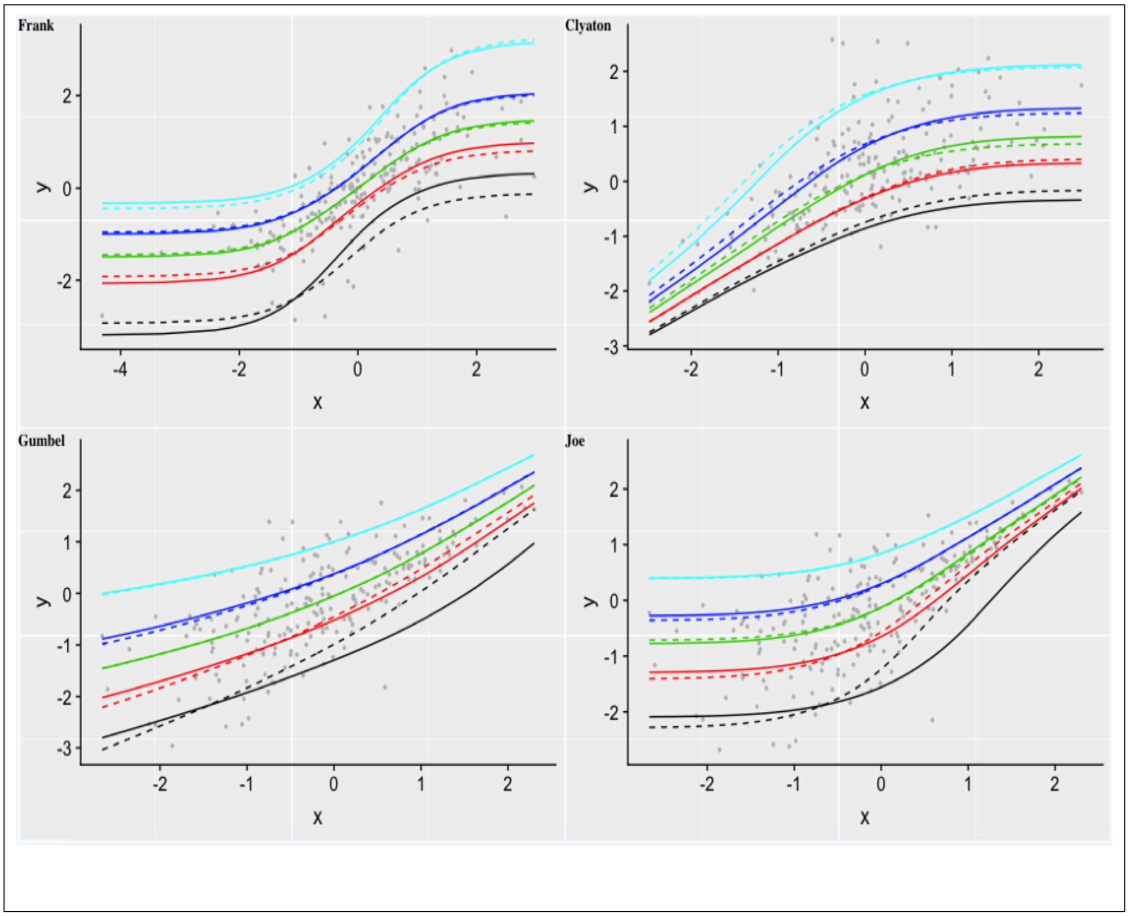

Figure 3.8: The true (solid) and estimated (dashed) conditional quantile curves for $\theta = 0.05$ (black), 0.25 (red), 0.50 (green), 0.75 (blue), 0.95 (cadet blue) based on one replication. The marginal distributions X and Y, for the Clayton, Gumbel, and Joe copulas follow standard normal and the marginal distributions X and Y for the Frank copula are Student-t distribution with 8 degrees of freedom.

By considering $f_\delta(x)$ as one of the conditional copula quantile functions described in the previous section, and minimizing (3.21), we are able to estimate the copula parameter. In fact, one can solve the following optimization problem for δ

$$\begin{aligned}\hat{\delta}(\theta) &= \arg\min_{\delta}\{\sum \rho_\theta(y_i - q(x_i,\theta;\delta))\} \\ &= \arg\min_{\delta}\{\sum_{\{i:y_i\geq q(x_i,\theta;\delta)\}} \theta|y_i - q(x_i,\theta;\delta)| + \sum_{\{i:y_i<q(x_i,\theta;\delta)\}} (1-\theta)|y_i - q(x_i,\theta;\delta)|\},\end{aligned} \tag{3.22}$$

where $q(x_i,\theta;\delta) = G^{-1}(C^{-1}_{2|1}(\theta|u_i))$ and $C^{-1}_{2|1}(\theta|u_i)$ is the inverse of the first partial derivative of the copula function with respect to its second argument, $u_i = F(x_i)$ (Briollais & Durrieu, 2014). The solution of (3.22) can be obtained by the "nlrq" function of the R package software "quantreg".

Figure 3.8 shows the true (solid curves) and estimated (dashed curves) copula quantile curves for 200 observations simulated from the bivariate copulas Frank, Clayton, Gumbel and Joe. These curves depicted for 1^{st} (red), 2^{nd} (green) and 3^{rd} (blue) quartiles; also, 5^{th} (black) and 95^{th} (cadet blue) percentiles. As Figure 3.8 shows, the estimated quantile curves are similar to the true copula curves at the middle quantiles for all copulas. However, there are differences in lower and upper quantiles, which may be due to few observations in the tails.

The next chapter compares the copula-based quantile regression models with the standard quantile and OLS regression models via simulations.

CHAPTER IV

SIMULATION STUDIES

This chapter aims to demonstrate the performance of the copula-based quantile regression approach in terms of prediction. The model performance is compared to standard OLS and quantile regression methods. To this aim, three simulation settings are investigated. That is, data are generated based on the location shift, location-scale shift and copula models. This chapter ends with a copula misspecification analysis.

In all simulation settings, firstly, we describe the data generation settings. Secondly, we present the evaluation criterion which are used for the methods comparison. Finally, we conclude based on the simulation results that are presented in Tables and Figures.

4.1 Data generation

The simple location-scale shift model is defined as

$$Y_i = \beta_0 + \beta_1 X_i + (\gamma_0 + \gamma_1 X_i)\epsilon_i. \tag{4.1}$$

In equation (4.1), when $\gamma_0 > 0$ and $\gamma_1 = 0$, the model (4.1) is a location shift model, and when $\gamma_0 \geq 0$ and $\gamma_1 > 0$, model (4.1) becomes a location-scale shift model. The procedure of data simulation for the different settings is given as follows:

Setting 1-Location shift:

To generate data form location shift model, in equation (4.1), we set $\gamma_0 = 1$, $\gamma_1 = 0$, $\beta_0 = -0.17$, $\beta_1 = -0.56$,

$$\begin{aligned}
\epsilon_i &\sim F_\epsilon, \\
X_i &\sim N(0,1), \\
Y_i &= -0.17 - 0.56X_i + \epsilon_i, \\
\rho &= -0.49.
\end{aligned}$$

In this setting, ϵ_i is assumed to follow three different distributions such as Normal, Student-t and Chi-square. Thus, three scenarios are considered in this setting corresponding to three error distributions. For each scenario, ρ is the true value of the population correlation coefficient between X and Y which is calculated by $\rho = \beta_1 \frac{\sigma_x}{\sigma_y}$ where σ_y is the standard deviation of the variable Y and varies in deferent scenarios and depends on the distribution of ϵ_i. It should be noted that in the asymmetric scenario (the errors follow the skewed Student-t distribution) the variance is calculated by the formula presented in Appendix B.

Setting 2-Location-scale shift:

To generate data from the location-scale shift setting, we set $\gamma_0 = 1$, $\gamma_1 = 1$, $\beta_0 = -0.17$, $\beta_1 = -0.56$ and

$$\begin{aligned}
\epsilon_i &\sim N(0,1), \\
X_i &\sim N(0,1), \\
Y_i &= -0.17 - 0.56X_i + (1 + X_i)\epsilon_i, \\
\rho &= -0.37.
\end{aligned}$$

As the previous setting, ρ is the true value of the population correlation coefficient between X and Y which is equal to $\rho = \beta_1 \frac{\sigma_x}{\sigma_y}$. For this setting, σ_y^2 is calculated as follows

$$\begin{aligned}\sigma_y^2 &= var(-0.17 - 0.56X_i + (1 + X_i)\epsilon_i)\\ &= (-0.56)^2 var(X_i) + var(\epsilon_i) + var(\epsilon_i X_i)\\ &= (-0.56)^2 var(X_i) + var(\epsilon_i) + var(\epsilon_i) var(X_i),\end{aligned}$$

since $\epsilon_i \perp X_i$ and ϵ_i, X_i continued, $\sigma_y^2 = 0.3136 + 1 + 1 = 2.3136$.

Setting 3-Copula-based quantile models:

Scenario 1: the conditional Clayton-copula quantile model introduced in equation (3.18) is used to the data simulation with standard normal margins. Data generating is described as follows:

we set

$$\begin{aligned}\theta_i &\sim unif(0, 1),\\ X_i &\sim N(0, 1),\\ Y_i &= \Phi^{-1}(\left[(\theta_i^{\frac{-2}{3}} - 1)F(x_i)^{-2} + 1\right]^{\frac{-1}{2}}),\\ \delta &= 2.\end{aligned}$$

Scenario 2: the conditional Gaussian-copula quantile model (3.7) with chi-square margins is used for data generation in this scenario. That is, in equation (3.7), we set

$$\begin{aligned}\theta_i &\sim unif(0, 1),\\ X_i &\sim \chi_5^2,\\ Y_i &= G_y^{-1}\left((-0.57)\Phi^{-1}(F(x_i)) + \Phi^{-1}(\theta_i)\sqrt{1 - (-0.57)^2}\right),\\ \rho &= -0.57,\end{aligned}$$

where $G(y)$ follows Chi-square distribution with 9 degrees of freedom.

Setting 4-Copula misspecification:

To evaluate the sensitivity of the copula-based quantile regression models and demonstrate the copula misspecification impact, we generate the data from the conditional Clayton-copula quantile model with standard normal margins as in Scenario 1 of Setting 3, however, we fit Frank and Student copula quantile models for these data. More precisely, the dependence between Y and X is generated from Clayton copula but we fit Frank and Student copulas to model (Y,X) dependence. In all settings, we set N = 1000 observations as the sample size of each data replication.

4.2 Evaluation criterion

To evaluate models' accuracy, the predicted values of different regression techniques $(\hat{y}_i)$ are compared to the observed values y_i. Also, evaluation of Bias of parameters is reported. Thus, the evaluation criterions employed are given as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2,$$

$$Bias = (\hat{\delta} - \delta),$$

where the errors are $(y_i - \hat{y}_i)$, δ is the true parameter of the models and $\hat{\delta}$ is parameter estimate. It should be noted that, the true parameter of interest in quantile and OLS regression is β_1, and for copula-based quantile regression the true parameter of interest is the copula parameter. Based on B = 1000 replications, average of *MSE*, standard deviation of *MSE* and the average of *Bias* are reported for comparison.

Throughout this chapter, "MSE" and "SD" indicate the average and standard deviation of *MSE* statistic, respectively. The "BIAS" indicates the average of the Bias of parameters over 1000 replications.

For all scenarios, the standard quantile and copula-based quantile regression are implemented for the following quantiles $\theta = 0.25$, 0.50, 0.75.

4.3 Results

The simulation results are presented in the upcoming sections.

4.3.1 Setting 1: Location shift

Scenario 1: Table 4.1 provides the MSE, SD and the BIAS (for $\hat{\beta}_1$ and $\hat{\rho}$) based on 1000 replications. Estimation is done by the standard quantile regression (SQR), the Gaussian copula-based quantile regression (GCBQR) with standard normal marginal distributions and OLS regression. As the results of Table 4.1 show, the MSE, SD and BIAS values of SQR and OLS models are smaller than GCBQR at the lower, middle and upper quantiles ($\theta = 0.25, 0.50, 0.75$). The boxplot of MSE values for OLS, SQR and GCBQR in $\theta = 0.25, 0.50, 0.75$ is presented in Figure 4.1.

θ	0.25		0.50		0.75		-
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR	OLS
MSE	0.003	0.072	0.003	0.031	0.004	0.004	0.002
SD	0.003	0.004	0.002	0.003	0.003	0.004	0.002
BIAS	0.001	-0.016	0.002	0.071	0.001	0.089	0.002

Table 4.1 Setting 1 (Scenario 1)-Location shift: MSE, SD and BIAS of the parameters' estimates ($\hat{\beta}_1$ and $\hat{\rho}$) based on 1000 replications. The fitted models are standard quantile regression (SQR), Gaussian copula-based quantile regression (GCBQR) with standard normal marginal distributions (for $\theta = 0.25, 0.50, 0.75$), and OLS regression when $\epsilon_i \sim N(0, 1)$ and $\rho = -0.49$.

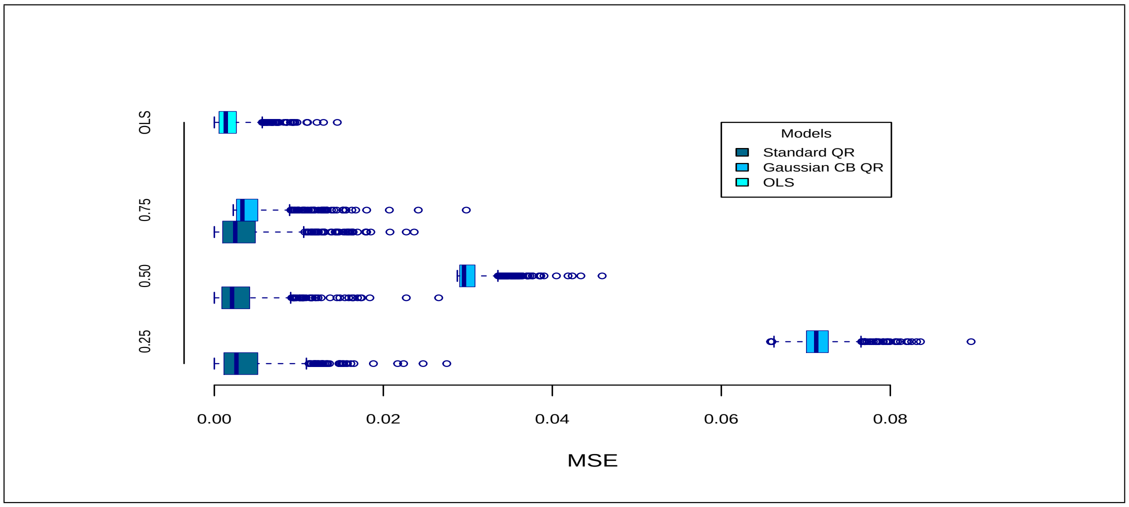

Figure 4.1: Setting 1 (Scenario 1)-Location shift: the boxplot of MSE values for OLS, SQR and GCBQR with standard normal margins in $\theta = 0.25, 0.50, 0.75$, when $\epsilon_i \sim N(0, 1)$.

θ	0.25		0.50		0.75		-
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR	OLS
MSE	0.471	0.093	0.013	0.036	0.469	0.041	0.010
SD	0.125	0.016	0.013	0.002	0.112	0.022	0.008
BIAS	0.001	-0.009	0.002	0.010	0.001	0.019	0.001

Table 4.2 Setting 1 (Scenario 1)-Location shift: MSE, SD and BIAS based on 1000 replications. Data are generated from location shift setting, when $\epsilon_i \sim N(0, 4)$ and $\rho = -0.27$. The fitted models are OLS, SQR and GCBQR regression.

Also, Table 4.2 shows the results with $\epsilon_i \sim N(0, 4)$ and $\rho = -0.27$. As one can see, by increasing the errors' variance, GCBQR models have acceptable performance in lower and upper quantiles since the copula are less sensitive to scale changes in data. However, OLS and SQR have appropriate performance at the middle quantiles. The boxplot of MSE values for these models is provided in Figure 4.2.

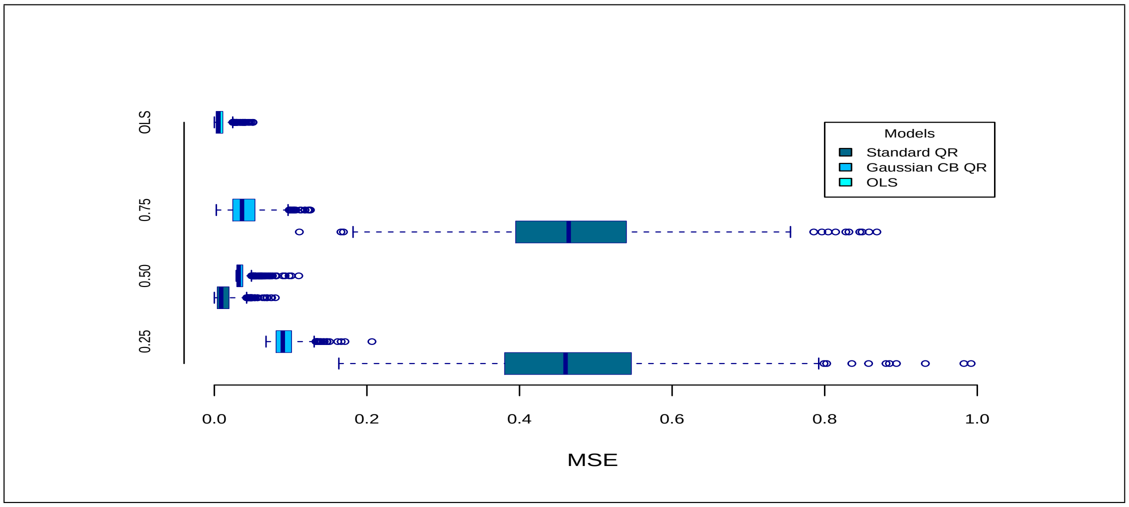

Figure 4.2: Setting 1 (Scenario 1)-Location shift: the boxplot of MSE values for OLS, SQR and GCBQR in $\theta = 0.25, 0.50, 0.75$, when $\epsilon_i \sim N(0, 4)$.

```
Error in library(quantreg): there is no package called 'quantreg'
Error in nlrq(df$y ~ q_normal(df$x, rho, tau), data = df, tau =
tau, start = list(rho = 0.1), : could not find function "nlrq"
Warning: Computation failed in `stat_quantile()`:
Package `quantreg` required for `stat_quantile`.
Please install and try again.
```

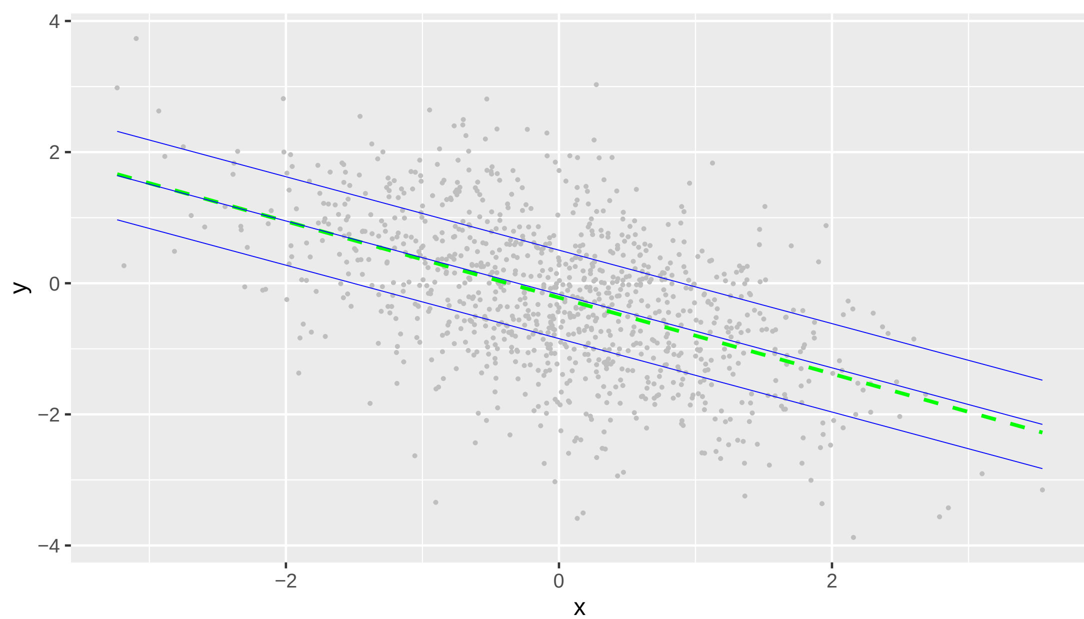

Figure 4.3: Setting 1 (Scenario 1)-Location shift: the fitted GCBQR with standard normal margins (black dotted lines), the fitted SQR (red dashed lines) and the true quantiles (solid blue lines) in $\theta = 0.25, 0.50, 0.75$, OLS regression (green dashed line) with $\epsilon_i \sim N(0, 1)$.

θ	0.25		0.50		0.75		-
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR	OLS
MSE	0.063	0.073	0.004	0.031	0.056	0.054	0.007
SD	0.005	0.003	0.004	0.003	0.006	0.004	0.006
BIAS	-0.047	-0.071	0.001	0.091	-0.016	-0.049	0.003

Table 4.3 Setting 1 (Scenario 2)-Location shift: MSE, SD and Bias based on 1000 replications generated from location shift model described in Setting 1; the errors follow a Student-t distribution with 3 degrees of freedom and $\rho = -0.31$. The fitted models are SQR, GCBQR with Student-t margin for Y for $\theta = 0.25, 0.50, 0.75$, and OLS regression.

Scenario 2: In this section, we assume that the errors follow the heavy-tails Student-t distribution function with 3 degrees of freedoms. As Table 4.3 shows, SQR models have better performance in lower and middle quantiles. However, GCBQR models have adequate performance at the upper quantiles. OLS performance seems to be not affected by heavy-tailed distribution.
Scenario 3: When the errors come from an asymmetric distribution such as skewed Student-t or Chi-square (see Tables 4.4 and 4.5), SQR method has acceptable performance in lower, middle and upper quantiles. However, the OLS regression is clearly inefficient when the errors follow an asymmetric distribution.
As earlier explained, the OLS regression is sensitive to errors' variance and it does not have appropriate performance in presence of both asymmetric and heavy-tailed situations. Also, the median quantile regression is the appropriate substitute for OLS in the middle of the data in such situations. Thus, in the upcoming sections, only standard quantile and copula-based quantile regression will be the focus of comparison.

θ	0.25		0.50		0.75		-
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR	OLS
MSE	0.005	0.098	0.008	0.056	0.021	0.044	0.339
SD	0.004	0.005	0.008	0.007	0.020	0.018	0.125
BIAS	0.012	0.138	-0.039	0.065	-0.032	-0.016	0.009

Table 4.4 Setting 1 (Scenario 3)-Location shift: MSE, SD and BIAS based on 1000 replications generated from location shift model described in Setting 1; the errors follow a skewed Student-t with 3 degrees of freedom and skewed parameter 2 and $\rho = -0.20$. The fitted models are SQR, GCBQR with skewed Student-t marginal distribution of Y, for $\theta = 0.25, 0.50, 0.75$, and OLS regression.

θ	0.25		0.50		0.75		-
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR	OLS
MSE	0.012	0.043	0.002	0.125	0.011	0.051	0.303
SD	0.0002	0.008	0.002	0.0008	0.013	0.069	0.050
BIAS	0.001	0.045	0.002	0.025	-0.018	-0.076	0.003

Table 4.5 Setting 1 (Scenario 3)-Location shift: MSE, SD and BIAS based on 1000 replications generated from location shift model described in Setting 1. The errors follow a chi-square distribution with 1 degree of freedom and $\rho = -0.37$. The fitted models are SQR, GCBQR with Chi-square marginal distribution for Y and normal marginal distribution for X for $\theta = 0.25, 0.50, 0.75$, and OLS regression.

4.3.2 Setting 2: Location-scale shift

Results of Table 4.6 and 4.7 refer to estimated models by SQR, GCBQR with standard normal margins (for $\theta = 0.25, 0.50, 0.75$) base on 1000 replications, when $\epsilon_i \sim N(0,1)$ and $\epsilon_i \sim N(0,4)$, respectively.

As Table 4.6 shows, the MSE values for SQR are smaller than the MSE values for GCBQR in the lower, middle and upper quantiles. Therefore, SQR method performs better in the location-scale shift situation. Figure 4.4 illustrates the MSE values of these two methods in $\theta = 0.25, 0.50, 0.75$ by boxplot. Furthermore, Table 4.7 shows the same results when $\epsilon_i \sim N(0,4)$.

As one can see, by increasing the errors' variance, GCBQR models are less sensitive to scale change compared to SQR models (based on MSE) in the upper quantiles (see Figure 4.5). Also, SQR provides good models in lower and middle quantiles. However, GCBQR models have acceptable performance in lower and middle quantiles as well. Figure 4.6 illustrates the fitted lines for these two methods.

θ	0.25		0.50		0.75	
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR
MSE	0.006	0.068	0.005	0.047	0.066	0.070
SD	0.008	0.005	0.007	0.006	0.030	0.020
BIAS	0.091	0.138	-0.003	0.062	-0.430	-0.281

Table 4.6 Setting 2-Location-scale shift: MSE, SD and BIAS based on 1000 replications generated from location-scale shift model described in Setting 2; the errors follow standard normal distribution and $\rho = -0.37$. The fitted models are SQR and GCBQR with standard normal margins for $\theta = 0.25, 0.50, 0.75$.

θ	0.25		0.50		0.75	
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR
MSE	0.018	0.029	0.011	0.031	0.289	0.042
SD	0.020	0.007	0.014	0.008	0.111	0.012
BIAS	0.185	0.064	-0.002	0.023	-0.721	-0.132

Table 4.7 Setting 2-Location-scale shift: MSE, SD and BIAS based on 1000 replications. Data are generated from location-scale shift model described in Setting 2, when $\epsilon_i \sim N(0, 4)$ and $\rho = -0.24$. The fitted models are SQR and GCBQR with standard normal margins for $\theta = 0.25, 0.50, 0.75$.

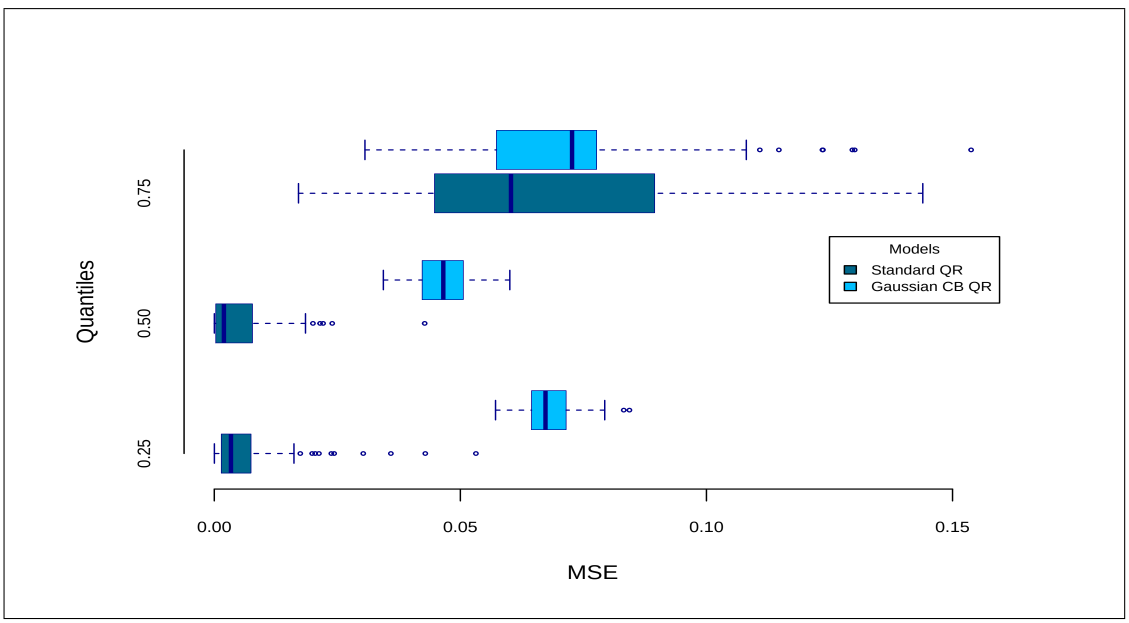

Figure 4.4: Setting 2-Location-scale shift: the boxplot of MSE values for the standard quantile, the Gaussian copula-based quantile regression with standard normal margins for $\theta = 0.25, 0.50, 0.75$, when $\epsilon_i \sim N(0, 1)$.

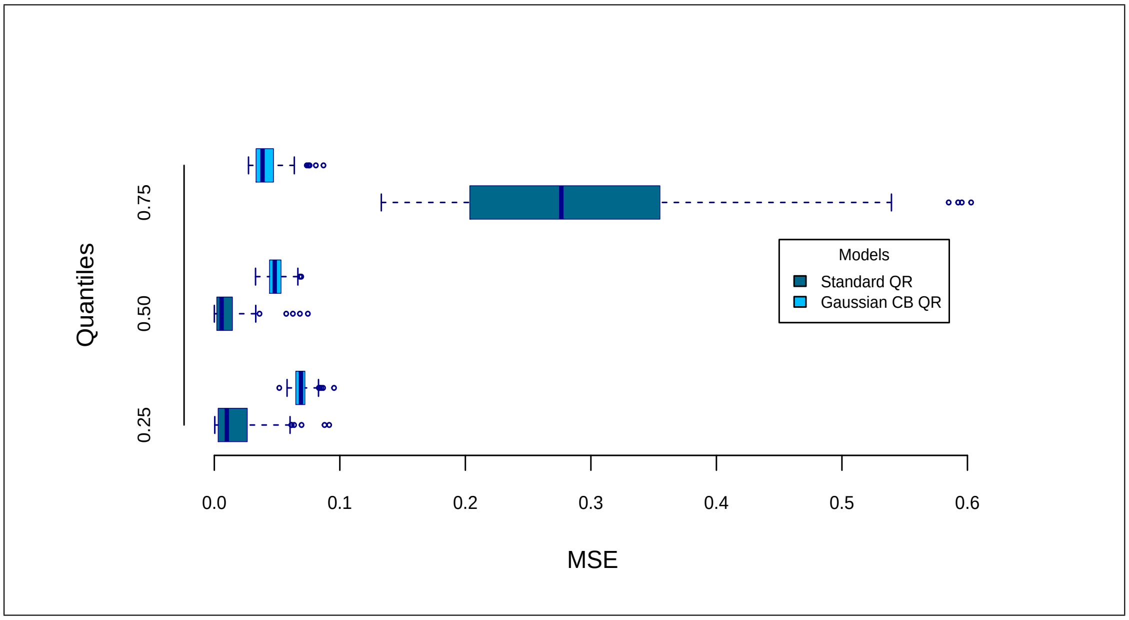

Figure 4.5: Setting 2-Location-scale shift: the boxplot of MSE values for SQR, GCBQR with standard normal margins for $\theta = 0.25, 0.50, 0.75$, when $\epsilon_i \sim N(0, 4)$.

We know that, for the location-scale shift models, the errors are correlated with the explanatory variables. In this case, the variance of the errors is heterogeneous. As Figure 4.6 shows, in the middle of the data, these two models are similar. At the lower and upper margins, there are differences between true quantile lines, quantile regression and copula-based quantile regression lines.

4.3.3 Setting 3: Copula-based quantile models

Scenario 1 (Clayton copula-based quantile regression): The results of models estimated by the standard quantile regression (SQR) and the Clayton copula-based quantile regression (CCBQR) with standard normal margins, based on 1000 replications are presented in Table 4.8. According to these results, at the lower, middle and upper quantiles, there are remarkable differences between the MSE and SD estimated by CCBQR and SQR. Thus, SQR has no appropriate performance in this scenario.

```
Error in library(quantreg):  there is no package called 'quantreg'
Error in nlrq(df$y ~ q_normal(df$x, rho, tau), data = df, tau =
tau, start = list(rho = -0.1), :  could not find function "nlrq"
Warning:  Computation failed in `stat_quantile()`:
Package `quantreg` required for `stat_quantile`.
Please install and try again.
```

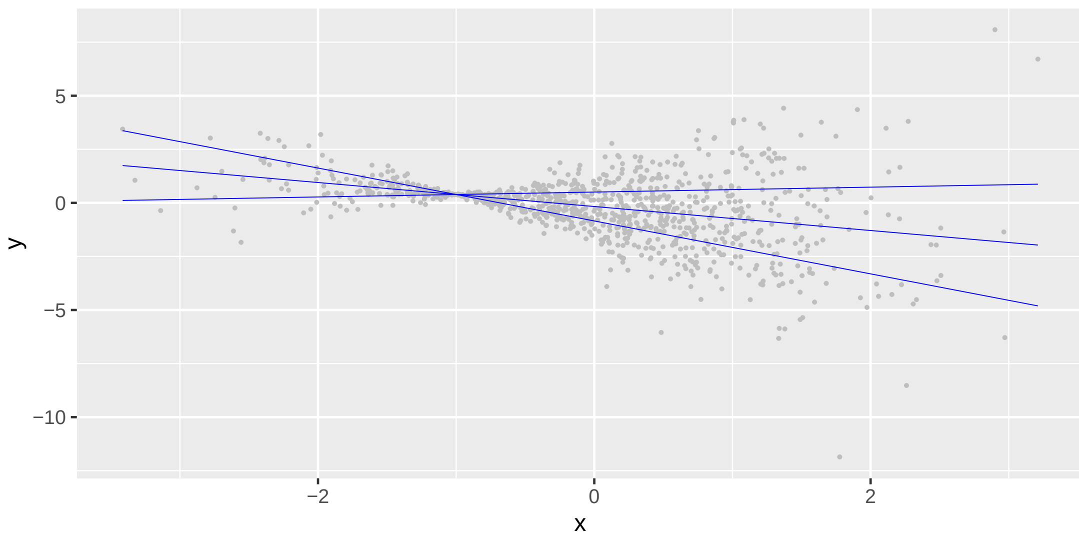

Figure 4.6: Setting 2-Location-scale shift: the fitted GCBQR with standard normal margins (black dotted lines), the fitted SQR (red dashed lines) and the true quantiles (solid blue lines) in $\theta = 0.25,\ 0.50,\ 0.75$.

Figure 4.7 depicts the MSE of these models by boxplot. As one can see, CCBQR models have the minimum values of MSE in $\theta = 0.25, 0.50, 0.75$. Also, Figure 4.8 shows the true and the estimated curves based on conditional Clayton-copula quantile functions.

Scenario 2 (Gaussian copula-based quantile regression with chi-square margins): Results in Table 4.9 includes the MSE, SD and BIAS of the models estimated by the standard quantile regression (SQR) and the Gaussian copula-based quantile regression (GCBQR) with chi-square margins, based on 1000 replications. As Table 4.9 shows, when the margins are assumed to be a chi-square distribution functions, the performance of models estimated by GCBQR is substantially better than the performance of models estimated by SQR, in lower, middle and upper quantiles (based on MSE and SD). Figure 4.9 shows that the minimum values of MSE belongs to GCBQR in all three percentiles 25^{th}, 50^{th}, and 75^{th}. Also, the estimated and true curves of GCBQR with chi-square margins along with the standard quantile regression are illustrated in Figure 4.10.

θ	0.25		0.50		0.75	
Methods	SQR	CCBQR	SQR	CCBQR	SQR	CCBQR
MSE	0.037	0.001	0.053	0.001	0.111	0.004
SD	0.005	0.002	0.005	0.001	0.013	0.006
BIAS	-	-0.002	-	-0.006	-	0.004

Table 4.8 Setting 3 (Scenario 1)-Clayton copula-based quantile regression: MSE, SD and BIAS based on 1000 replications generated from conditional Clayton-copula quantile function described in Setting 3. The fitted models are standard quantile (SQR) and Clayton copula-based quantile regression (CCBQR) with standard normal margins for $\theta = 0.25, 0.50, 0.75$. "-" means there is no explicit form of the true slope parameter of SQR based on copula.

θ	0.25		0.50		0.75	
Methods	SQR	GCBQR	SQR	GCBQR	SQR	GCBQR
MSE	0.721	0.021	0.974	0.022	1.314	0.026
SD	0.092	0.030	0.120	0.031	0.163	0.038
BIAS	-	-0.004	-	-0.001	-	-0.002

Table 4.9 Setting 3 (Scenario 2)-Gaussian copula-based quantile regression: MSE, SD and BIAS based on 1000 replications generated from conditional Gaussian-copula quantile function described in Setting 3. The fitted models are SQR and GCBQR with chi-square margins for $\theta = 0.25, 0.50, 0.75$. "-" means there is no explicit form of the true slope parameter of SQR based on copula.

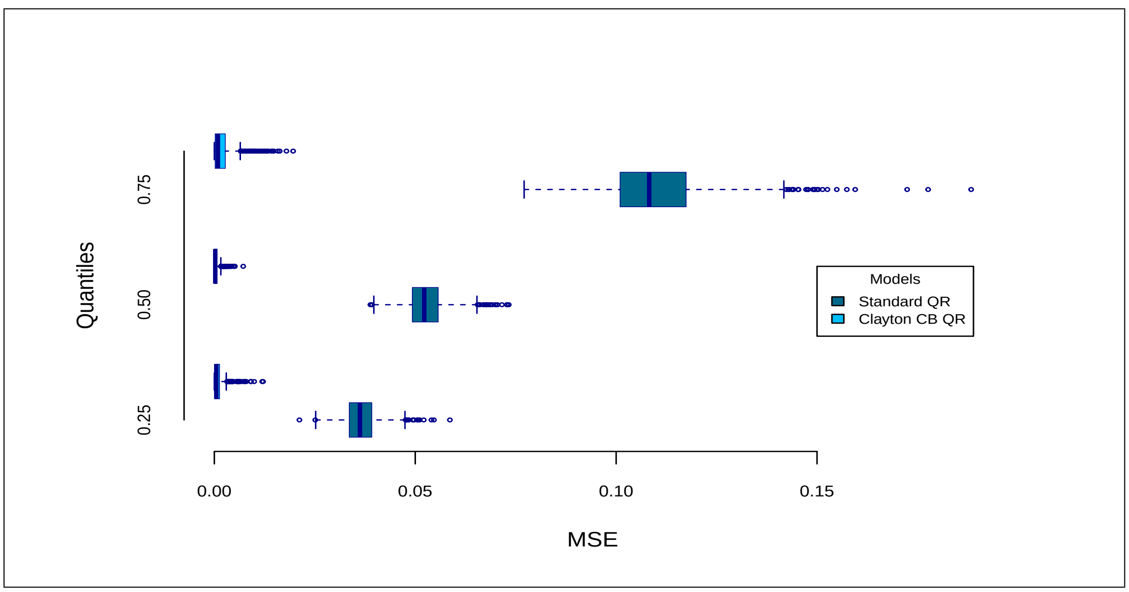

Figure 4.7: Setting 3 (Scenario 1)-Clayton copula-based quantile regression: the boxplot of MSE values of the standard quantile and the Clayton copula-based quantile regression in $\theta = 0.25, 0.50, 0.75$.

```
Error in library(quantreg):  there is no package called 'quantreg'
Error in nlrq(y ~ q_clayton(x, t, tau), data = df, tau = tau, start
= list(t = 2), :  could not find function "nlrq"
Warning:  Computation failed in `stat_quantile()`:
Package `quantreg` required for `stat_quantile`.
Please install and try again.
```

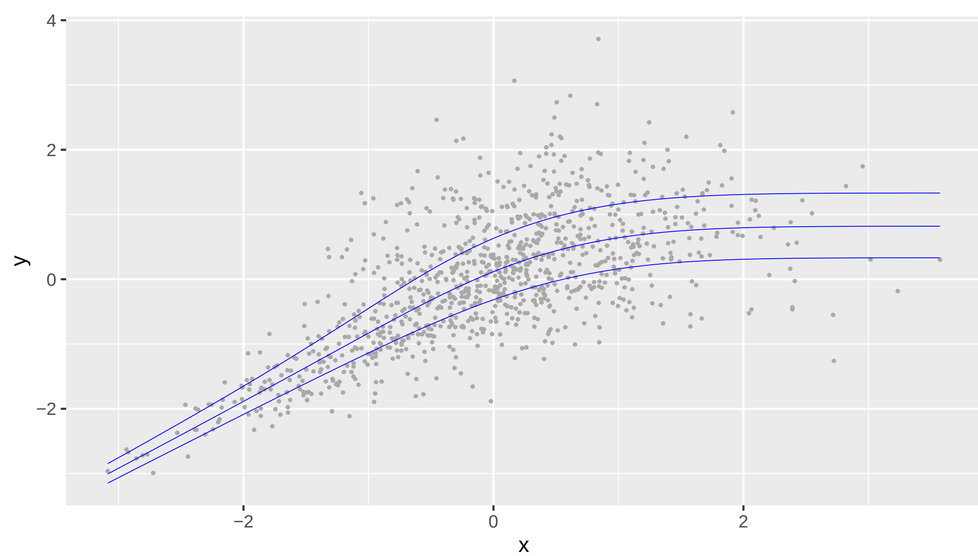

Figure 4.8: Setting 3 (Scenario 1)-Clayton copula-based quantile regression with standard normal margins: the fitted CCBQR (black dotted curves), the fitted SQR (red dashed lines) and true conditional Clayton-copula quantiles (solid blue curves) for $\theta = 0.25$, 0.50 and 0.75.

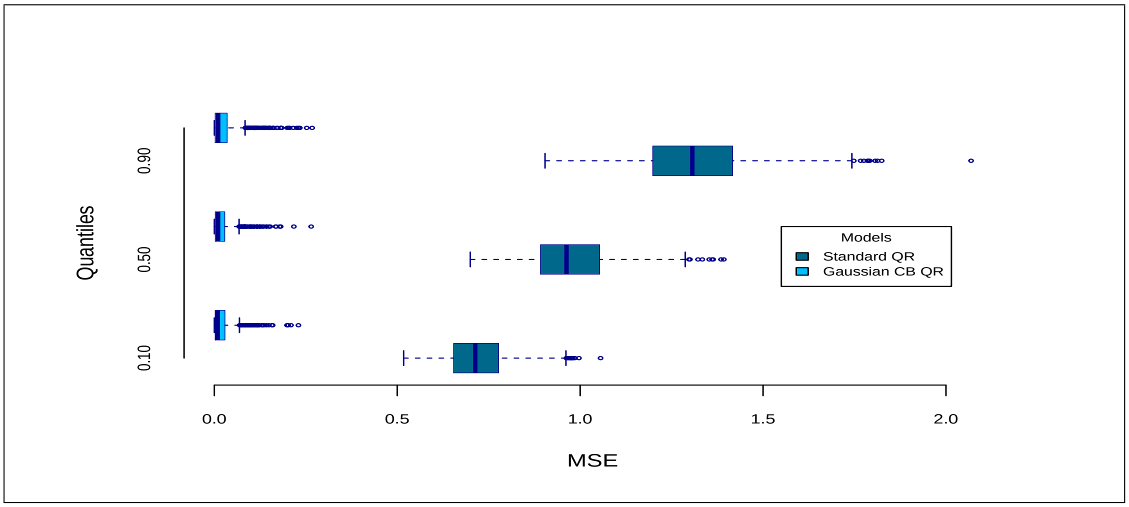

Figure 4.9: Setting 3 (Scenario 2)-Gaussian copula-based quantile regression with chi-square margins: boxplot of MSE values for SQR and GCBQR in $\theta = 0.25, 0.50, 0.75$.

```
Error in library(quantreg):  there is no package called 'quantreg'
Error in nlrq(df$y ~ q_normal(df$x, rho, tau, df1, df2), data = df,
tau = tau, :  could not find function "nlrq"
Warning:  Computation failed in `stat_quantile()`:
Package `quantreg` required for `stat_quantile`.
Please install and try again.
```

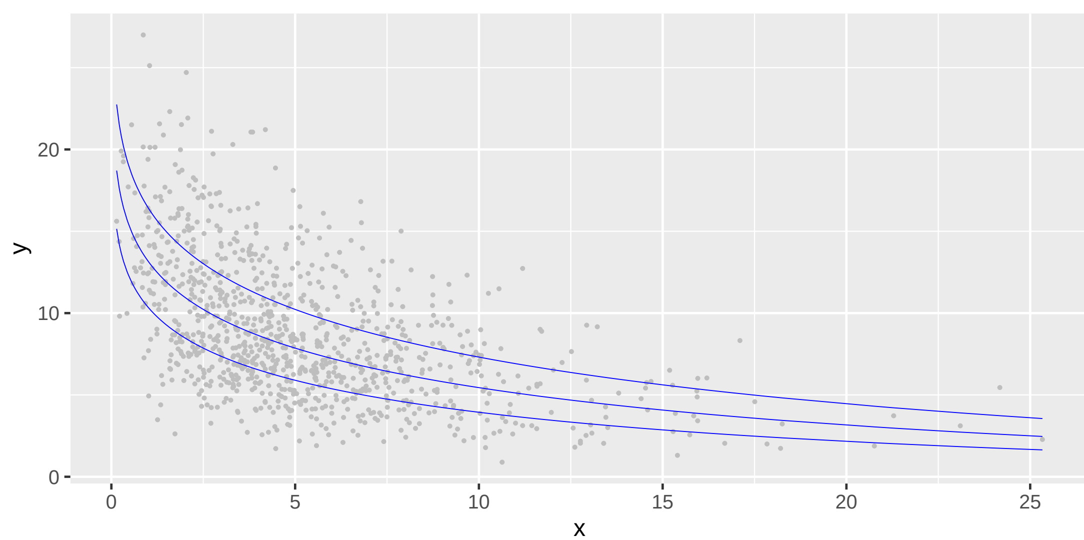

Figure 4.10: Setting 3 (Scenario 2)-Gaussian copula-based quantile regression with chi-square margins: the estimated curves of GCBQR with chi-square marginals distributions (dotted black), SQR (dashed red) lines and the true curves of conditional Gaussian-copula quantile with chi-square margins (solid blue) in $\theta = 0.25, 0.50, 0.75$.

4.3.4 Setting 4: Copula misspecification

In this section, the Frank and Student copula-based quantile regressions are employed to fit the models. These two models are compared to the Clayton copula-based quantile regression, from which the data are generated. The results of the analysis based on 1000 replications for Clayton (C), Frank (F) and Student-t (S) copula-based quantile regression models for $\theta = 0.25, 0.50, 0.75$, are presented in Table 4.10.

```
Error in library(quantreg):  there is no package called 'quantreg'
Error in nlrq(y ~ q_frank(x, p, tau), data = df, tau = tau, start =
list(p = 1), :  could not find function "nlrq"
```

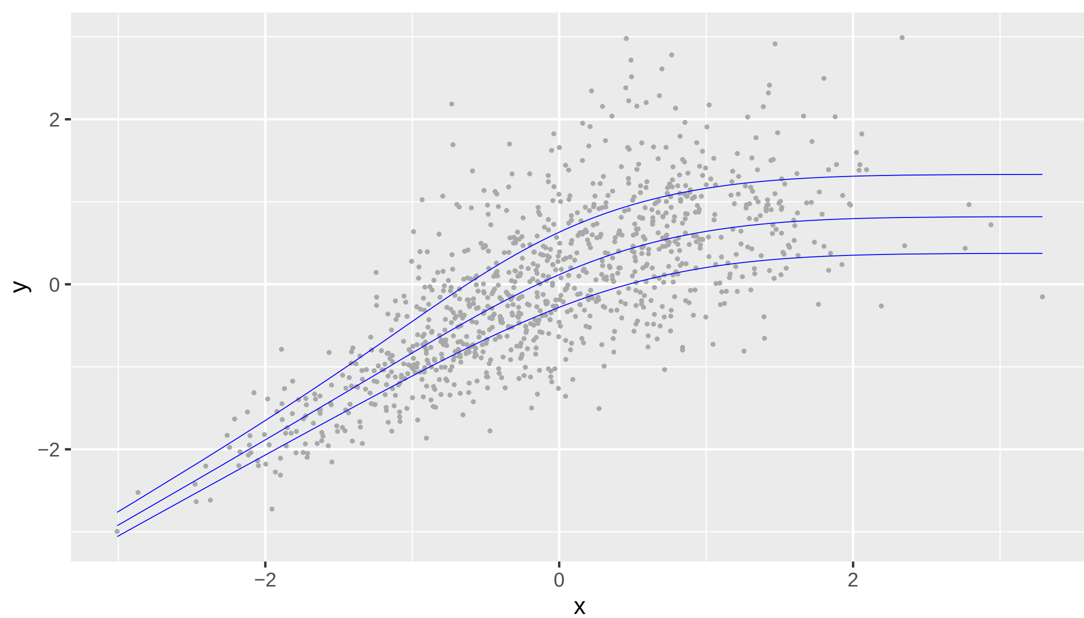

Figure 4.11: Setting 4-Copula misspecification: the estimated curves of FCBQR with standard normal margins (dotted black), the estimated curves of CCBQR with standard normal margins (dashed red) and the true curves of conditional Clayton-copula quantiles (solid blue) for $\theta = 0.25, 0.50, 0.75$.

For all models, standard normal distributions are used as marginal distribution of both X and Y. As one can see from this table, the MSE and SD for the Frank and Student-t copula-based quantile regression are larger than MSE and SD of the Clayton copula-based quantile regression model, for $\theta = 0.25, 0.50, 0.75$.

θ	0.25			0.50			0.75		
Methods	C	F	S	C	F	S	C	F	S
MSE	0.001	0.047	0.068	0.002	0.063	0.060	0.001	0.097	0.117
SD	0.001	0.006	0.003	0.001	0.007	0.007	0.001	0.010	0.011
BIAS	0.001	-0.029	-0.079	0.002	0.075	0.031	0.001	0.021	0.121

Table 4.10 Setting 4-Copula misspecification: MSE, SD and BIAS based on 1000 replications. The data generated from conditional Clayton-copula quantile described in Setting 4. The fitted models are Clayton (C), Frank (F) and Student-t (S) copula-based quantile regression models, with standard normal margins, for $\theta = 0.25, 0.50, 0.75$ quantiles.

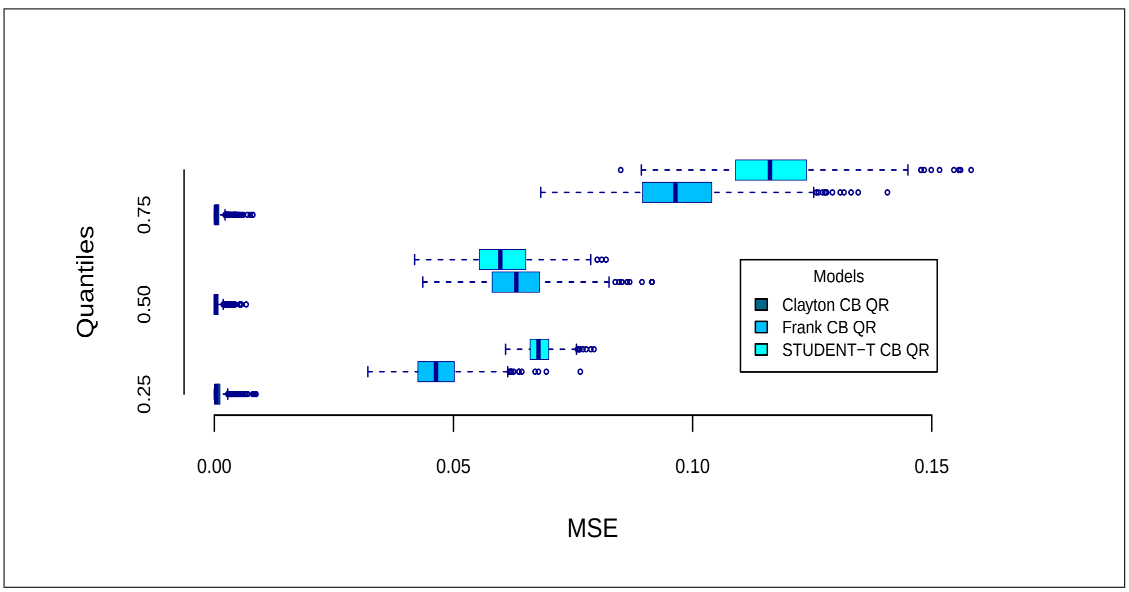

Figure 4.12: Setting 4-Copula misspecification: the boxplot of MSE values for the fitted Clayton copula-based quantile, the fitted Frank copula-based quantile and Student-t copula-based quantile regression with standard normal margins in $\theta = 0.25, 0.50, 0.75$

Furthermore, there are differences between the BIAS of the Frank, Student-t, and Clayton copula-based quantile regression in lower, middle and upper quantiles. Figure 4.12 shows the boxplot of the MSE values for these three models. As one can see, again the evidence of copula misspecification impact for Frank and Student-t copula-based quantile regression is clear (see also Figures 4.11 and 4.13).

```
Error in library(quantreg):  there is no package called 'quantreg'
Error in nlrq(y ~ q_student(x, rho, tau), data = df, tau = tau,
start = list(rho = 0.2), :  could not find function "nlrq"
```

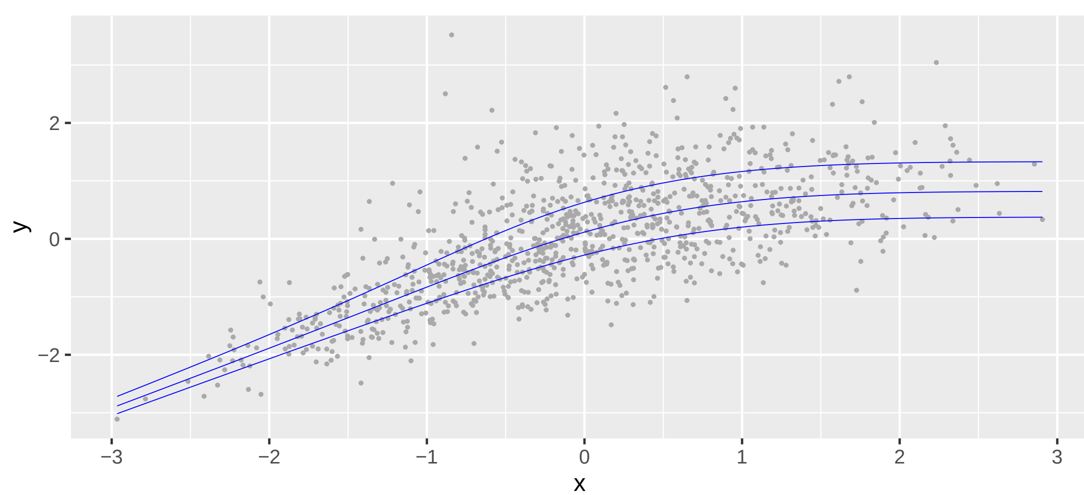

Figure 4.13: Setting 4-Copula misspecification: the estimated curves of SCBQR with standard normal margins (dotted black), the estimated curves of CCBQR with standard normal margins (dashed red) and the true curves of conditional Clayton-copula quantiles (solid blue) for $\theta = 0.25, 0.50, 0.75$.

4.4 Results summary

In the location shift setting, there are no significant differences between the performance of the models estimated by the standard quantile, copula-based quantile and OLS regression methods (refer again to Figure 4.1). By increasing the errors' variance, the copula-based quantile regression (CBQR) models are better than the standard quantile regression (see Figure 4.2). Also, SQR and OLS have appropriate performance in the middle quantiles. When the errors follow a heavy-tailed distribution, the SQR models show appropriate performance in the lower, middle and upper quantiles, however, the CBQR models perform better in upper quantiles. As Tables 4.4 and 4.5 show, when the errors are asymmetrically distributed, SQR and OLS have poor performance.

In the location-scale shift setting (heterogeneous), as Figure 4.4 shows, SQR models have acceptable performance in the lower, middle and upper quantiles. By increasing the errors' variance, SQR models have also appropriate performance. On the other hand, the CBQR models perform better in upper quantiles (see Figure 4.5).

In Setting 3, when the data is generated from conditional copula-based quantile models, CBQR have adequate performance rather than SQR (Figure 4.7).

4.5 Limitation

There are some restrictions to fit the multivariate copula-based quantile regression with the "quantreg" R package and "nlrq()" function. For instance, using "nlrq" to fit the CBQR model in tail of the distribution have shown numerical convergence issues and instability for the model parameters estimation. This might be due to the absence of observations falling in the extreme tails.

Moreover, fitting the Archimedean copula-based quantile regressions that have no closed form of the copula quantile function, such as Gumbel and Joe copula, can be time consuming.

CONCLUSION

This study aimed to investigate flexible modeling of the relationship between a response variable and a set of predictors. The performance of three modeling techniques, namely OLS, QR and copula-based quantile regression, based on MSE and BIAS statistics was explored.

We have shown that the copula-based quantile regression is a convenient alternative to both OLS and QR frameworks. For heavy-tailed outcome distribution, although the OLS and QR have shown appropriate performance in the center of distribution, we have demonstrated that the copula-based quantile regression has an acceptable performance in the tails. Also, in heteroscedastic situations, we have shown that the copula-based quantile model performs better than the quantile regression at upper quantiles.

By employing the copula-based quantile regression, there is a possibility to study linear or nonlinear relationships between a response and predictor variables at desired quantiles. Finally, it should be noted that choosing the appropriate copula for modeling the dependence between variables is relatively crucial for the copula-based quantile regression framework.

APPENDIX A

PSEUDO-INVERS OF GENERATOR FUNCTION

Definition 4.1.1 (Nelsen 2006, p110). Let φ be a continuous, strictly decreasing function with domain $[0,1]$ and range $[0,\infty]$ such that $\varphi(1)=0$. The pseudo-inverse of φ is the function $\varphi^{[-1]}$ with domain $[0,\infty]$ and range $[0,1]$ defined as

$$\varphi^{[-1]}(t)=\begin{cases}\varphi^{-1}(t), & 0\le t\le\varphi(0),\\ 0, & \varphi(0)\le t\le\infty.\end{cases}$$

Note that $\varphi^{[-1]}$ is continuous and nonincreasing on $[0,\infty]$, and strictly decreasing on $[0,\varphi(0)]$. Moreover, $\varphi^{[-1]}(\varphi(u))=u$ on $[0,1]$, and

$$\varphi\left(\varphi^{[-1]}(t)\right)=\begin{cases}t, & 0\le t\le\varphi(0),\\ \varphi(0), & \varphi(0)\le t\le\infty,\end{cases}$$

$$=min\left(0,\varphi(0)\right).$$

Finally, if $\varphi(0)=\infty$, then $\varphi^{[-1]}=\varphi^{-1}$.

APPENDIX B

SKEWED STUDENT-T DISTRIBUTION

For random variable T which follows a Skewed Student-t distribution, the mean and variance are defined as follows

$$E(T) = \begin{cases} \mu\sqrt{(\frac{\nu}{2})}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}, & \text{if } \nu > 1, \\ \text{Does not exist}, & \text{if } \nu \leq 1, \end{cases}$$

$$Var(T) = \begin{cases} \frac{\nu(1+\mu^2)}{\nu-2} - \frac{\mu^2\nu}{2}(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)})^2, & \text{if } \nu > 2, \\ \text{Does not exist}, & \text{if } \nu \leq 2, \end{cases}$$

where ν is degrees of freedom and μ is the skewed parameter of the Student-t distribution. It should be noted that when $\mu = 0$ the Student-t distribution is symmetric.

BIBLIOGRAPHY

Bernard, C. & Czado, C. (2015). Conditional quantiles and tail dependence. *Journal of Multivariate Analysis*, *138*, 104–126.

Bouyé, E. & Salmon, M. (2013). Dynamic copula quantile regressions and tail area dynamic dependence in forex markets. In *Copulae and Multivariate Probability Distributions in Finance* pp. 125–154. Routledge.

Briollais, L. & Durrieu, G. (2014). Application of quantile regression to recent genetic and-omic studies. *Human genetics*, *133*(8), 951–966.

Cade, B. S. & Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, *1*(8), 412–420.

Casella, G. & Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Cherubini, U., Luciano, E. & Vecchiato, W. (2004). *Copula methods in finance*. John Wiley & Sons.

Gujarati, D. N. (1995). Basic econometrics, 3-rd ed.

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., Maechler, M. M. & Suggests, M. (2014). Package 'copula'. *URL http://ie. archive. ubuntu. com/disk1/disk1/cran. r-project. org/web/packages/copula/copula. pdf*.

Kjersti, A., Czado, C., Frigessi, A. & Bakken, H. (2015). Pair-copula constructions of multiple dependence. *URL http://epub.ub.uni-muenchen.de*.

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

Koenker, R. (2012). Quantile regression in r: A vignette. *Retrieved November*.

Koenker, R. & Bassett Jr, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, pp. 43–61.

Koenker, R., Portnoy, S., Ng, P. T., Zeileis, A., Grosjean, P. & Ripley, B. D. (2019). Package 'quantreg'.

Mai, J. & Scherer, M. (2014). *Financial engineering with copulas explained.* Springer.

Nelsen, R. B. (2003). Properties and applications of copulas: A brief survey. In *Proceedings of the first brazilian conference on statistical modeling in insurance and finance*, pp. 10–28. Citeseer.

Nelsen, R. B. (2006). *An introduction to copulas.* Springer Science & Business Media.

Xie, Y. (2017). *Dynamic Documents with R and knitr.* Chapman and Hall/CRC.