2018

Copula Theory and Regression Analysis

Mayooran Thevaraja
Minnesota State University, Mankato

Copula Theory and Regression Analysis

By

Mayooran, Thevaraja

A Thesis submitted in Partial Fulfillment of the Requirement for the Degree of Masters in Mathematics and Statistics

Department of Mathematics and Statistics
Minnesota State University, Mankato, Minnesota

Spring 2018

Date:24 April 2018

Title:Copula Theory and Regression Analysis.

Student's Name: Mayooran,Thevaraja .

This thesis has been examined and approved by the following members of the student's committee.

.....

Advisor/Chair Person,
Dr Mezbahur Rahman,
Professor of Statistics,
Minnesota State University, Mankato.

.....

Committee Member,
Dr Han Wu,
Professor of Statistics,
Minnesota State University, Mankato.

.....

Committee Member,
Dr Deepak Sanjel,
Professor of Statistics,
Minnesota State University, Mankato.

I would like to dedicate this thesis to my loving wife veshane. . . . . .

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Mayooran, Thevaraja
Spring 2018

Acknowledgements

It was the best of time. It was the worst of time. For Charles Dickinson, it was French Revolution, but for me, it was my life in the past two years as a Masters student in Minnesota State University. When I first came here, I was with curiosity, with fear, with doubt, and with uncertainty. I wish I can leave here with confidence, with independence, and with hope, to a bright future. First and foremost, I am deeply grateful to my wonderful current advisor, Dr. Mezbahur Rahman , who encouraged me, inspired me, and helped me with great tolerance and patience. Without him, this dissertation is never made possible. I was moved by his rigorous thinking, broad vision, and his passion to statistics. Moreover, Dr. Han Wu is not just a former advisor, but also a mentor, and a role model. I learned a lot from his thoughtfulness and kindness. My thanks go out to Dr. Mezbahur Rahman, Dr. Han Wu, and Dr. Deepak Sanjel for serving on my dissertation committee and providing valuable suggestions and support on my research. Specially i received broad of programming knowledge from Dr Sanjel's class during my Masters studies. My thanks also go out to Dr. Charles Waters, who is my teaching assistantship supervisor and Chair / Department of Mathematics and Statistics at Minnesota State University. I am grateful for his supervision and guidance, which provides a valuable life experience in my first career in USA. In addition, I would like to thank the faculty and staff from Department of Mathematics and Statistics, Minnesota State University Mankato, USA for their help throughout my graduate studies. I am thankful for all my best friends in Mankato and elsewhere around the world. With them, I had a great time, and never felt lonely. The most important of all, I must thank my wife, who love me unconditionally, support me no matter what, and always have absolute faith in me. This thesis is dedicated to my beloved wife.

Abstract

Researchers are often interested to study in the relationships between one variable and several other variables. Regression analysis is the statistical method for investigating such relationship and it is one of the most commonly used statistical Methods in many scientific fields such as financial data analysis, medicine, biology, agriculture, economics, engineering, sociology, geology, etc. But basic form of the regression analysis, ordinary least squares (OLS) is not suitable for actuarial applications because the relationships are often nonlinear and the probability distribution of the response variable may be non-Gaussian distribution. One of the method that has been successful in overcoming these challenges is the generalized linear model (GLM), which requires that the response variable have a distribution from the exponential family. In this research work, we study copula regression as an alternative method to OLS and GLM. The major advantage of a copula regression is that there are no restrictions on the probability distributions that can be used. First part of this study, we will briefly discuss about copula regression by using several variety of marginal copula functions and copula regression is the most appropriate method in non Gaussian variable(violated normality assumption) regression model fitting. Also we validated our results by using real world example data and random generated (50000 observations) data. Second part of this study, we discussed about multiple regression model based on copula theory, and also we derived multiple regression line equation for Multivariate Non-Exchangeable Generalized Farlie-Gumbel-Morgenstern (FGM) copula function.

Keywords: *Regression, ordinary least squares (OLS), multivariate Gaussian copula, copula regression, generalized linear models(GLM).*

Table of contents

List of figures ix

List of tables x

Nomenclature xi

1 Introduction 1
- 1.1 Objectives of the study . . . 2
- 1.2 Literature Review . . . 3
- 1.3 Layout . . . 3

2 Preliminaries on Copula Theory 4
- 2.1 History of copula theory . . . 4
- 2.2 Definitions and properties . . . 6
- 2.3 Some extended concepts of copulas . . . 13
- 2.4 Methods of Constructing Copulas . . . 16
- 2.5 Important families of copulas . . . 19
  - 2.5.1 Farlie-Gumbel-Morgenstern's (FGM) Family . . . 19
  - 2.5.2 Archimedean Copulas . . . 20
  - 2.5.3 Elliptical Copulas . . . 22
  - 2.5.4 Frechet's Family . . . 25
- 2.6 Dependence Measures . . . 25
  - 2.6.1 Measure of concordance . . . 27
  - 2.6.2 Measure of dependence . . . 29
  - 2.6.3 Tail Dependence . . . 29

3 Statistical inference of copulas 31
- 3.1 Estimation and Asymptotic Properties . . . 31
  - 3.1.1 Maximum likelihood Estimation (MLE) . . . 32

3.1.2 Inference Functions for Margins (IFM) method 33
3.1.3 Canonical Maximum Likelihood (CML) method 34
3.1.4 Nonparametric estimation 35
3.2 Choice of Copula Model 37
3.3 Simulation from Copulas 39

4 Copula Regression Theory **40**
4.1 Properties of Copula Regression function 41
4.2 Linear Copula Regression Functions 43
4.3 Multiple linear Copula Regression Function 45
4.4 Multivariate Non-Exchangeable FGM Copula 46
4.5 Gaussian copula marginal regression models 48
4.6 Implementation by using R Programming 48

5 Results and Discussions **51**
5.1 Comparing copula,OLS and GLM Regressions 51

6 Conclusions and Future study **72**

References **74**

Appendix A Programming codes **78**

Index **83**

List of figures

2.1 surface plot of the independence copula 8
2.2 contour plot of the independence copula 9
2.3 Bivariate random samples of size 250 from various Clayton copulas 21
2.4 Bivariate random samples of size 250 from various Frank copulas. 23
2.5 Bivariate random samples of size 250 from various joe copulas 24
2.6 Spearman's rho and Kendall's tau for normal copulas. 28

5.1 Diagnostic plots - OLS Model-Example 1 56
5.2 Diagnostic plots - GLM Model-Example 1 57
5.3 Diagnostic plots - Copula Model Neg.Binomial as Marginal-Example 1 . . 58
5.4 Diagnostic plots - OLS Model-Example 2 62
5.5 Diagnostic plots - GLM Model-Example 2 63
5.6 Diagnostic plots - Copula Model Poisson as Marginal-Example 2 64
5.7 Diagnostic plots - OLS Model-Example 3 69
5.8 Diagnostic plots - GLM Model-Example 3 70
5.9 Diagnostic plots - Copula Model Gaussian as Marginal-Example 3 71

List of tables

4.1 Parameter Estimation 43
4.2 Marginals models available in gcmr with the default link function. 49
4.3 Correlation models available in gcmr package. 49
4.4 Functions and methods available for objects of class gcmr. 50

5.1 Parameter Estimations for Example 1 55
5.2 Parameter Estimations for Example 2 65
5.3 Parameter Estimations for Example 3 68

Nomenclature

Greek Symbols

$\bar{\Re}$	Extended real line
$\Im$	Sigma field
λ_L	Coefficient of lower tail dependence
λ_u	Coefficient of upper tail dependence
$\phi(\cdot)$	Univariate standard normal margin
$\phi^{[-1]}$	Pseudo inverse
$\prod(u,v)$	Product copula
ρ	Coefficient of correlation
ρ_s	Spearman's rho
τ	Kendall's tau

Other Symbols

S_X, S_Y	Marginal survival functions
$\bar{F}, \bar{G}$	Marginal survival functions
$\bar{H}$	Bivariate survival function
$\hat{C}$	Survival copula
ϕ^{-1}	Inverse of ϕ
AIC	Akaike Information Criterion

B_X, B_Y	Baseline survival functions
BIC	Bayesian Information Criterion
C	Copula distribution function
c	Copula density function
$C^+(u,v)$	Copula characterizes perfect positive dependence
$C^-(u,v)$	Copula characterizes perfect negative dependence
$C^0(u,v)$	Two-dimensional independence copula
C_B	Bernstein copula
C_n	Empirical copula
$F(x), F(y)$	Marginal distribution functions
F^{-1}	Quantile function
$H(x,y)$	Joint distribution function
$h(x,y)$	Joint density function
I	Indicator function
$P_{k_i,m_i}(u_i)$	Bernstein polynomials
r	Pearson correlation coefficient
$r_c(u)$	Copula regression function
$S_{XY}(x,y)$	joint survival function
$U = F(x)$	Probability Integral Transform of X
$V = F(y)$	Probability Integral Transform of Y
X, Y	Random variables

Chapter 1

Introduction

The term "Regression" and the methods for investigating the relationships between two variables may date back to about hundred years ago. It was first introduced by Francis Galton in 1908, the renowned British biologist, when he was engaged in the study of heredity. One of his study was that the children of tall parents to be taller than average but not as tall as their parents. This "Regression toward mediocrity" gave this statistical methodology as their name. The term regression and its evolution primarily describe statistical relationship between the variables. In particular, the simple regression is the regression method to discuss the relationship between one response/dependent variable (y) and set of explanatory/independent variable (x). The basic ordinary least squares (OLS) regression model presents a specific model for the relationship. The distribution of Y given the co-variates assumed to be normal with a variance that is constant (that is, not related to the co-variates) and a mean that is related to the co-variates as $E(Y|X_1 = x_1, \cdots, X_k = x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$. The equivalent (in this case) techniques of maximum likelihood and least-squares are used to estimate the unknown coefficients. In order to get more flexible approaches to apply in real problems, different models have been proposed overcoming the less realistic assumptions of the Linear Models, where a Gaussian distribution was assumed for the dependent variable, with a constant variance and linear relationship between the predictor and the dependent variable. The development of the Generalized Linear Models (GLM) McCullagh and Nelder, 1989 [29] relaxed the distributional assumption from Gaussian to any other distributions from the exponential family and the proposal of the Generalized Additive Models (GAM) Hastie and Tibshirani,1990 [38]; Wood, 2006 [49] allowed to model the relation between co-variates and the response variable in a non-linear way using additive predictors.

During a long time, researchers have been interested on the relationship between a multivariate distribution function and its lower dimensional margins. In 1950s, M. Frechet and G. Dallaglio studied about this matter, studying the bi-variate and tri-variate distribution

functions with given uni variate marginal distribution. The answer to this problem for the uni-variate marginal case was introduced by A. Sklar in 1959 [42] creating a new class of functions which he called copulas and also he proved the useful theorem that now bears his name; In Thomas Mikosch (2006) [30] stated that in September 2005, a Google search on the term "copulas" yielded 650,000 results. Then, in June 2017 and April 2018, this same query generated more than 3.81 million and 4.51 million, respectively. Given the number of publications in scientific journals and the number of scholarly articles available on the Internet, it is undeniable that the passion for copula theory continues to boom. The word 'copula' is a Latin noun, which means "a link, tie or bond". In this study we deal with the concept of copula theory and Regression that was first introduced in a mathematical or statistical sense by Sklar (1959) [42] in a theorem that describes a copula as a mathematical function, which joins or "couples" a multivariate distribution function to its one-dimensional marginal distribution functions. Equivalently, a copula is defined as a multivariate distribution function whose one-dimensional marginals uniform on the interval$[0,1]$. In the words of Fisher (1997) [16] as noted in his paper in the first update volume of the Encyclopedia of Statistical Sciences, "Copulas [are] of interest to the statistician 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, sometimes with a view to simulation". In the following chapter we will provide a short introduction of the historical development of copulas from these two perspectives.

1.1 Objectives of the study

In this research study the following objectives were considered,

1. Comparing linear regression, Generalized linear model and Copula based regression by using minimized AIC value. In this case we will calculate AIC values for each situations and finally we compared that AIC values and we proposed copula regression is best fitting method for Regression analysis when we are using OLS and GLM assumptions violated data.

2. We propose a new multivariate copula regressions function, which is developed from sungur (2005) [47] research paper, he studied just bi-variate case Farlie-Gumbel-Morgenstern (FGM) family. In this study also we also consider same family but multivariate case.

1.2 Literature Review

Copula theory and it's applications in statistics has been done by the scholars in different locations all over the world. Most of recently published works on the copula regression models have been focused on analyzing fully observed data; for example, Song (2007) [44]; Czado (2010) [9]; Joe et al. (2010) [23]; Genest et al. (2011) [20]; Masarotto et al. (2012) [28]; Acar et al. (2012) [1].Gaussian copula regression model Song (2000) [45] is an useful probability model for the correlated data. In 2005, Engin Sungur [47] was introduced an alternative way of looking at regression analysis by using copula theory. We will discuss more literature reviews in following chapter section 2.1.

1.3 Layout

The goal of this research is to show that copula regression could be extensively used in regression analysis. The report is organized as follows. In chapter 2, we present copula functions and some related theoretical properties, in particular the concept of dependence properties. After we consider the theoretical results of statistical inference of copulas in chapter 03 and discuss about copula regression theory in chapter 04. Chapter 05, includes the main results and the interpretations of our investigation by using some useful data set and randomly generated database. Chapter 06 presents our conclusions and some direction of future studies.

Chapter 2

Preliminaries on Copula Theory

In this chapter we provide a brief introduction to the ever-growing copula theory the results that are immediately usable in the subsequent chapters and to deeply discuss some useful developments in copula theory. Moreover, it is our understanding that the inclusion of this introductory chapter on review of copula theory makes the thesis to stand by itself. Therefore in this chapter is as follows. In Section 2.1 we present brief history of copula theory and Section 2.2 introduce some definitions and important properties of copulas. Describe some extended concepts of copulas in section 2.3. In Section 2.4, we discuss various methods of constructing copulas and in Section 2.5 we discuss several important families of copulas in section 2.6. Section 2.7 explains with defining the notions of dependence measures in terms of copulas such as measures of concordance, dependence and also discuss other dependence concepts. We will discuss several methods of estimations and selection for copula models and describe a method of simulating copulas in next chapter. The final section presents copula model application on the bi-variate characterization of drought events aiming at the investigation of the regression analysis.

2.1 History of copula theory

The word copula is a Latin noun that means “a link, tie, bond” (Cassell’s Latin Dictionary) and is used in grammar and logic to describe “that part of a proposition which connects the subject and predicate” (Oxford English Dictionary). The word copula was first employed in a mathematical or statistical sense by Abe Sklar (1959) [42] in the theorem (which now bears his name) describing the functions that “join together” one-dimensional distribution functions to form multivariate distribution functions (later we will discuss that Theorem). In (Sklar 1996) [43] we have the following account of the events leading to this use of the term copula:

"Féron (1956), in studying three-dimensional distributions had introduced auxiliary functions, defined on the unit cube, that connected such distributions with their one-dimensional marginals. I saw that similar functions could be defined on the unit n-cube for all $n \geq 2$ and would similarly serve to link n-dimensional distributions to their one-dimensional marginals. Having worked out the basic properties of these functions, I wrote about them to Fréchet, in English. He asked me to write a note about them in French. While writing this, I decided I needed a name for these functions. Knowing the word "copula" as a grammatical term for a word or expression that links a subject and predicate, I felt that this would make an appropriate name for a function that links a multidimensional distribution to its one-dimensional margins, and used it as such. Fréchet received my note, corrected one mathematical statement, made some minor corrections to my French, and had the note published by the Statistical Institute of the University of Paris as Sklar (1959) [42]."

But as Sklar notes, the functions themselves predate the use of the term copula. They appear in the work of Fréchet, Dall'Aglio, Féron, and many others in the study of multivariate distributions with fixed univariate marginal distributions. Indeed, many of the basic results about copulas can be traced to the early work of Wassily Hoeffding. After Hoeffding, Fréchet, and Sklar, the functions now known as copulas were rediscovered by several other authors. Kimeldorf and Sampson (1975b) [26] referred to them as uniform representations , and Galambos (1978) [17] and Deheuvels (1980) [13] called them dependence functions. At the time that Sklar wrote his 1959 paper with the term "copula," he was collaborating with Berthold Schweizer(1991) [39] in the development of the theory of probabilistic metric spaces, or PM spaces. During the period from 1958 through 1976, most of the important results concerning copulas were obtained in the course of the study of PM spaces.Though similar ideas and results, for example, can be traced back to the works of Hoeffding (1940) [48] who studied the link between bivariate "standardized distributions'" whose support is contained in the square $[-1/2, 12]^2$ and whose margins are uniform on the interval $[-1/2, 12]$, Sklar (1959) [42] in his break-through work answered the question about the link between a multivariate distribution and its one-dimensional margins by introducing the concept of copula. Schweizer (1991) [39] noted that had "Hoeffding chosen the unit square $[0,1]^2$ instead of $[-1/2, 12]^2$for his normalization, he would have discovered copulas before Sklar". Even after Hoeffding and Sklar, the functions now known as copulas were rediscovered by several other authors. Kimeldorf and Sampson (1975) [26] referred to them as "uniform representations", Galambos (1978) [17] and Deheuvels (1978) [13] called them "dependence functions" and Cook and Johnson (1981) [7] called them "standard forms". In the 1960's and 70's most of the results about copulas were obtained in the course of the development of the probabilistic metric spaces, mainly in the study of binary operations in the space of

probability distribution functions. The introduction of copulas to the statistical literature has been a recent phenomenon. In this regard Schweizer (1991) [39] in Thirty Years of Copulas noted that "Those of us working on these matters had no formal training in Statistics. Thus, we were only tangentially aware of possible statistical applications. Moreover, with the notable exception of Sklar's original paper, our results were presented in a novel context and published in journals not generally read by statisticians. Thus the statistical community took little note of our work...." In the last two decades remarkable advances have been made in the field of probability distributions with given or fixed marginals that led to the organization of four international conferences in Rome, Seattle, Prague and Barcelona in 1990, 1993, 1996 and 2000, respectively, under the main theme Distributions with Given Marginals and Related Topics. As a result more advances have been achieved in the development of copula theory and its broad applications in probability theory, Bio-statistics, Finance, Insurance, Economics, Data-mining, Hydrology, Environment and many other fields. The literature on the copula theory has kept growing as the interest in the application of copulas increases. Many extensions have been made to the concept of copula first introduced by Sklar (1959) [42]. Survival copula appears as a function, which relates the joint survival function of multivariate distribution with its one-dimensional survival functions. Extreme value copulas are discussed in Joe (1997) [22]. Time dependent or conditional copulas are introduced by Patton (2006a, b) [35]) [34] with application in time series. Sancetta and Satchell (2004) [37] defined Bernstein copula using Bernstein polynomial approximation. For review of more recent developments on copulas we refer to Kolev et al. (2006) [27]. It is also noted that along the development of copulas, their application in defining measures of dependence between random variables appear implicitly in earlier works on dependence by many authors. Hoeffding (1940) [48] used "standard distributions" to define Pearson's coefficient of correlation. Spearman's rho, Hoeffding's dependence index, and Pearson's coefficient of mean square contingency. Deheuvels (1979) [10] used "empirical dependence function" to estimate the population dependence function and to construct various nonparametric tests of independence. The earliest paper, as noted by Nelsen (2006) [32] that explicitly relating copulas to the study of dependence among random variables appeared in Schweizer and Wolff (1981) [40]. They defined many of the measures of dependence in terms of copulas and also established the basic invariance

2.2 Definitions and properties

In order to keep the main ideas in focus and for sake of brevity, first we restrict our attention to the two-dimensional case and later we provide brief note on extension to the d -dimensional

case ($d > 2$). The following notations and definitions are introduced. Let X and Y be two real-valued random variables on a common probability space $(\Omega, \Im, P)$ with distribution functions $F(X) = P(X \leq x)$ and $G(y) = P(Y \leq y)$, respectively and a joint distribution function $H(x,y) = P(X \leq x, Y \leq y)$. We further assume that the distribution function H has all the first and second order derivatives. Since we will make use of quantile functions and probability integral transforms, we give their definitions as follows.

Definition 2.1 (Quantile Function):

Let X and F be as defined above. Then the quantile function (or generalized inverse) of F is any function F^{-1} such that,

$$F^{-1}(t) = \inf\{x | F(x) \geq t\} \qquad t \in (0,1) \tag{2.1}$$

For a continuous random variable X denote the quantile function of F by F^{-1}.

Definition 2.2 (Probability Integral Transform):
Let X, F and F^{-1} as defined above. Then,

1. For any standard uniformly distributed $U \sim U(0,1)$ we have $F^{-1}(U) \sim F$.
2. If F is continuous then the random variable $F(X)$ is standard uniformly distributed $(F(X) \sim U(0,1))$,

Further, let $U = F(X)$ and $V = G(Y)$ denote the probability integral transforms of X and Y, respectively. Now we give two equivalent definitions of copula.

Definition 2.3 (Copula):
The copula of (X,Y) denoted by C is the joint distribution function of U and V , i.e., a copula is a bivariate distribution function with Uniform$(0,1)$ margins. Alternatively, an equivalent definition that provides some properties of a copula can be given as follows.

Definition 2.4 (Copula):
A (two-dimensional) copula is a function $C : [0,1]^2 \to [0,1]$ with the following properties:

1. C is grounded, that is, $C(u,0) = 0$ and $C(0,v) = 0$ for all $u, v \in [0,1]$;
2. C is such that $C(u,1) = u$ and $C(1,v) = v$ for all $u, v \in [0,1]$;
3. C is increasing function; that is for every u_1, u_2, v_1, v_2 in $[0,1]$ such hat $u_1 \leq u_2$ and $v_1 \leq v_2$, $V_c([u_1,v_1] \times [u_2,v_2]) = C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0$ where the function V_c is called the $C-$ volume of the rectangle $[u_1,v_1] \times [u_2,v_2]$.

According to above definition shows that C is a bivariate distribution function on the unit square $[0,1]^2$ whose margins are uniform on $[0,1]$. Both above definitions are connected by Sklar's (1959) [42] via his theorem stated below, which now bears his name. In fact, the use of copulas allows solving the fundamental problem of determining the relationship between the joint distribution functions and their one-dimensional distributions by performing two basic tasks. First, find out one-dimensional margins (not necessarily from the same family) and secondly, choose a copula to link them. We now present Sklar's theorem that justifies such a role.

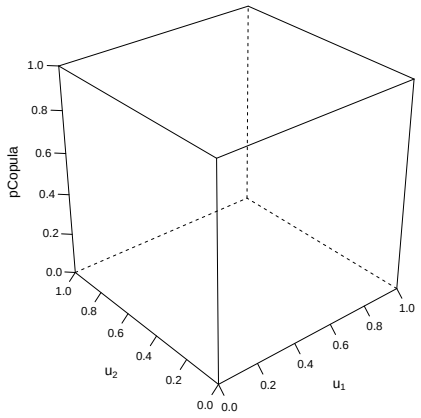

Fig. 2.1 surface plot of the independence copula

Theorem 2.1 (Sklar's Theorem):

Let H be a joint distribution function with margins F and G . Then there exists a copula C such that for all $(x,y) \in \bar{\Re}$ (extended real line),

$$H(x,y) = C(F(x),G(y)) \tag{2.2}$$

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on(Range of F) $\times$ (Range of G). Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (2.1) is a joint distribution function with margins F and G . See proof in Nelsen (2006) [32]. The following corollary provides a means how to determine a copula function from the joint distribution and the generalized inverse of its marginals, F^{-1} and G^{-1}.

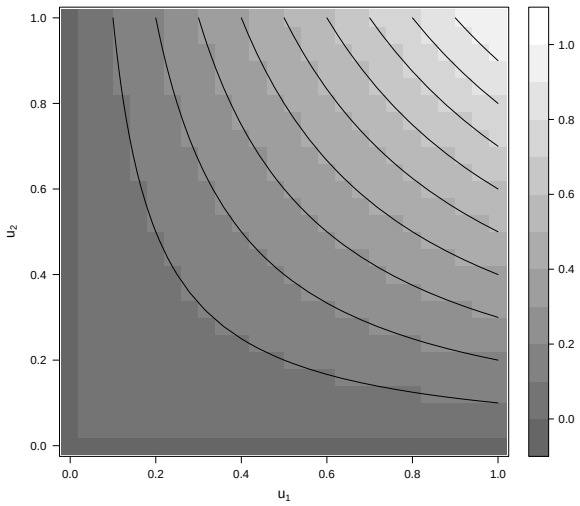

Fig. 2.2 contour plot of the independence copula

Corollary 2.1 (Sklar's Theorem):

Let H, F, G, and C be as defined above and let F^{-1} and G^{-1} be generalized inverses of F and G , respectively. Then for any(u,v) in $[0,1]^2$

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)) \tag{2.3}$$

See proof in Nelsen (2006) [32].

Note that when F and G are continuous this result provides a method of constructing copulas from knowledge of joint and marginal distribution functions (see Section 2.5). Next, we describe some of the basic properties of copulas. Proofs of these properties can be found, for example, in Nelsen (2006) [32] and Cherubini et al. (2004) [6].

Property:1 The Copula C is non-decreasing in each argument

Theorem 2.2:
Let C be a copula and $u_1 \leq u_2$ for $u_1, u_2 \in [0,1]$ such that for all $v \in [0,1]$

$$C(u_2,v) - C(u_1,v) \text{ non decreasing on } [0,1]$$

Similarly, for $v_1 \leq v_2$ for $v_1, v_2 \in [0,1]$ such that for all $u \in [0,1]$

$$C(u,v_2)-C(u,v_1) \text{ non decreasing on } [0,1]$$

Property:2 The Copula C is uniformly continuous in its domain.

Theorem 2.3:

Let C be a copula and for $u_1,u_2,v_1,v_2 \in [0,1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$ then

$$|C(u_2,v_2)-C(u_1,v_1)| \leq |u_2-u_1|+|v_2-v_1| \tag{2.4}$$

Property:3. All partial derivatives of the copula C exists.

Let C be a copula. For any v in $[0,1]$, the partial derivative $\partial C/\partial u$ exists for almost all u, and for such v and u, $0 \leq \partial C(u,v)/\partial u \leq 1$. Similarly, For any u in $[0,1]$, the partial derivative $\partial C/\partial v$ exists for almost all v, and for such u and v, $0 \leq \partial C(u,v)/\partial v \leq 1$. The partial derivatives of the copula can be used to define conditional distribution functions by the relationships;

$$P[X \leq x\,|Y=y] = \partial C(u,v)/\partial v \text{and} P[Y \leq y\,|X=x] = \partial C(u,v)/\partial u \tag{2.5}$$

where $u = F(x)$ and $v = G(y)$

Property:4. The Copula C satisfies the Frechet-Hoeffding bounds.

Theorem 2.4:

Let C be a copula. Then for every (u,v) in $[0,1]^2$

$$W(u,v) \leq C(u,v) \leq M(u,v) \tag{2.6}$$

where $W(u,v) = \max(u+v-1,0)$ and $M(u,v) = \min(u,v)$ In the two dimensions, both the Frechet-Hoeffding lower bound, W, and upper bound, M , are copulas. However, at higher dimensions $(d > 2)$ the lower bound is never a copula; see for example Nelsen (2006) [32].

Definition 2.5:

The Product copula denoted by H is given by $\prod(u,v) = uv$.

The product copula $(\prod)$, and the Frechet-Hoeffding bounds W and M are associated with important statistical interpretations of random variables (see properties Property:5 and Property:6 below). A copula that includes H, W and M is called a comprehensive copula.

Property:5. Independence of random variables

Theorem 2.4:

Let X and Y be continuous random variables. Denote by C the copula of X and Y . Then X and Y are independent if and only if $C = \prod$.

Property:6. Perfect dependence of random variables

Theorem 2.5:

Let X and Y be continuous random variables. Then the copula of X and Y is M if and only if each of X and Y is almost surely an increasing function of the other. Similarly, the copula of X and Y is W if and only if each of X and Y is almost surely a decreasing function of the other.

This theorem implies that the copulas M and W are associated with perfect positive dependence and perfect negative dependence, respectively. One very attractive property of copulas is that they are invariant or change in a predictable way under strictly monotone transformations of random variables.

Property:7. Invariance property of Copulas

Theorem 2.6 (Schweizer and Wolff, 1981 [40]):

Let X and Y be continuous random variables with copula C_{XY}. Then

1. If α and β are strictly increasing almost surely on Range of X and Range of Y , respectively, then $C_{\alpha(X),\beta(Y)} = C_{XY}$ Thus C_{XY} is invariant under strictly increasing transformations of X and Y

2. If α and β are strictly decreasing on almost surely on Range of X and Range of Y, then the copulas C_1, C_2 and C_3 of the pairs $(\alpha(X), Y), (X, \beta(Y))$, and $(\alpha(X), \beta(Y))$ respectively, are independent of the particular choices of α and β and are given by
   $C_1(u,v) = v - C_{XY}(1-u,v)$
   $C_2(u,v) = u - C_{XY}(u,1-v)$
   $C_3(u,v) = u + v - 1 + C_{XY}(1-u,1-v)$

See proof in Schweizer and Wolff (1981).

This interesting invariance property of copulas is very useful. Suppose we know the form of the copula for two variables X and Y but due to some practical reasons it is necessary to transform the data to $\log(x)$ and $\log(y)$. Then, because of the invariance property of copulas, the copula for the logarithm transformation $\log(x)$ and $\log(y)$ remains unchanged.

Property:8. The copula density exists everywhere in $[0,1]^2$.

Definition 2.6:
The copula density $c(\cdot,\cdot)$ is defined as $c(u,v) = \partial^2 C(u,v)/\partial u \partial v$.
Theorem 2.7:
The copula density c exists almost everywhere in $[0,1]^2$ and is non negative.

Note that for continuous random variables X and Y, the copula density c can be related to the density of the distribution functions H denoted by h and the marginal distributions F and G with densities denoted by f and g, respectively. From the probability integral transforms we have $U = F(x)$ and $V = G(Y)$, as a result $X = F^{-1}(U)$ and $Y = G^{-1}(V)$. Since for continuous random variables these transformations are strictly increasing and continuous we have

$$c(u,v) = h(F^{-1}(U), G^{-1}(V)) * \left| \begin{pmatrix} \partial X/\partial U & \partial X/\partial V \\ \partial Y/\partial U & \partial Y/\partial V \end{pmatrix} \right| \tag{2.7}$$

$$= \frac{h(F^{-1}(U), G^{-1}(V))}{f(F^{-1}(U))g(G^{-1}(V))}$$

It follows that,

$$c(F(x), G(y)) = \frac{h(x,y)}{f(x)g(y)} \tag{2.8}$$

Thus, the joint density of X and Y is expressed as the product of the marginal and copula densities:

$$h(x,y) = f(x)g(y)c(F(x), G(y)) \tag{2.9}$$

Next we state the d-dimensional $(d > 2)$ extension of Sklar's theorem. Many properties of the d -dimensional copula can be analogously defined as in the two-dimensional case. Here we just give the d -dimensional extension of the Sklar's theorem, further details can be found in Nelsen (2006) [32] and Cherubini et al. (2004) [6].

Theorem 2.8 (Sklar's Theorem in d -dimensions):

Let H be a d -dimensional distribution function with margins $F_1, F_2, ..., F_d$. Then there exists a d-copula C such that for all $(x_1, x_2, ..., x_d)$ in $\bar{\Re}^d$

$$H(x_1, x_2, ..., x_d) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d)) \tag{2.10}$$

If $F_1, F_2, ..., F_d$ are all continuous, then C is unique; otherwise, C is uniquely determined on Range $F_1 \times \cdots \times$ Range F_d . Conversely, if C is a d -copula and $F_1, F_2, ..., F_d$ are distribution functions, then the function H defined by (2.3) is a d -dimensional distribution function with margins $F_1, F_2, ..., F_d$

Corollary 2.2 (Sklar's Theorem in d-dimensions):

Let $H, C, F_1, F_2, ..., F_d$, be as defined above and let $F_1^{-1}, F_2^{-1}, ..., F_d^{-1}$ be generalized inverses of $F_1, F_2, ..., F_d$, respectively. Then for any $u_1, u_2, ...u_d$ in $[0,1]^2$

$$C(u_1, u_2, ...u_d) = H(F_1^{-1}(u_1), F_2^{-1}(u_2), ..., F_d^{-1}(u_d)) \tag{2.11}$$

2.3 Some extended concepts of copulas

In this section we summarize some important extended copula concepts, which have made remarkable contribution towards the growing interest in copula theory.

Survival Copula

For two random variables X and Y with marginal distributions F and G , respectively and bivariate distribution function H, the marginal survival functions $\bar{F}$ and $\bar{G}$ and the bivariate survival function $\bar{H}$ are given by $\bar{F}(x) = P(X > x), \bar{G}(y) = P(Y > y)$ and $\bar{H}(x,y) = P(X > x, Y > y)$ respectively.

Definition 2.7:

The survival copula $\hat{C}$ is a function, which relates the bivariate survival function to its marginal survival functions, i.e.,

$$\bar{H}(x,y) = \hat{C}(\bar{F}(x), \bar{G}(y)) \tag{2.12}$$

The survival copula $\hat{C}$ is a copula, and is related to the copula C of X and Y via the equation

$$\hat{C}(u,v) = u + v - 1 + C(1-u, 1-v) \tag{2.13}$$

Conditional or Time Dependent Copulas

In practice, in the presence of temporal dependence, which is a common feature of time series data, a powerful tool to specify the underlying models is conditional information with respect to past observations. Here we state the extended version of Sklar's Theorem given in Patton (2006a) [35] and Fermanian and Scaillet (2005a [15]).

Let X, Y and W be continuous random variables with W be the conditioning variable. Let the joint distribution of (X,Y,W) be H_{XYW}, denote the conditional distribution of (X,Y) given W as $H_{XY|W}$ and let the conditional marginal distributions of $X|W$ and $Y|W$ be denoted by $F_{X|W}$ and $G_{Y|W}$, respectively. Note that $F_{X|W}(x|w) = H_{XY|W}(x,\infty|w)$ and $G_{Y|W}(y|w) = H_{XY|W}(\infty,y|w)$. Assume that H_{XYW} has all the required derivatives, and that $F_{X|W}, G_{Y|W}$, and $H_{XY|W}$ are continuous.

Definition 2.8 (Conditional copula , Patton (2006a) [35]):

The conditional copula of $(X,Y)|W = w$, where $X|W = w \sim F_{X|W}(\cdot|w)$ and $Y|W = w \sim G_{Y|W}(\cdot|w)$is the conditional joint distribution function of $U = F_{X|W}(X|w)$ and $V = G_{Y|W}(Y|w)$ given $W = w$, where the variables U and V are known as conditional probability integral transforms of X and Y given W .

Patton (2006a) [35] has shown that the conditional copula function satisfies similar properties as that of the unconditional copula defined by Sklar (1959) [42]. Hence, the following version of Sklar's theorem holds.

Sklar's Theorem 2.9 (Conditional copula, Patton (2006a) [35]):

Let $F_{X|W}(\cdot|w)$ be the conditional distribution of $X|W = w, G_{Y|W}(\cdot|w)$ be the conditional distribution of $Y|W = w, H_{XY|W}(\cdot|w)$ be the joint conditional distribution function of $(X,Y)|W = w$, and $\bar{\omega}$ be the support of W . Assume that $F_{X|W}(\cdot|w)$ and $G_{Y|W}(\cdot|w)$ are continuous in x and v for all $w \in \bar{\omega}$. Then there exists a unique conditional copula $C(\cdot|w)$ such that $H_{XY|w} = C(F_{X|W}(x|w), G_{Y|W}(y|w)|w)$ for all $(x,y) \in \bar{\Re}^2$ and each $w \in \bar{\omega}$. (2.4) Conversely, if we let $F_{X|W}(\cdot|w)$ be the conditional distribution of $X|W = w, G_{Y|W}(\cdot|w)$ be the conditional distribution of $Y|W = w$, and $C(\cdot|w)$ be a family of conditional copulas that is measurable in w, then the function $H_{XY|W}(\cdot|w)$ defined by (2.4) is a conditional bivariate distribution

function with conditional marginal distributions $F_{X|W}(\cdot|w)$ and $G_{Y|W}(\cdot|w)$.

Remark: This version of Sklar's theorem for conditional distributions demands that the conditioning variable, W, must be same for both marginal distributions and the copula. Otherwise the function H_W may not be a joint conditional distribution function. Consider $X|W_1, Y|W_2$ and $C_{XY|w_1,w_2}$ and specify

$$H_{XY|w_1,w_2}(x,y|w_1,w_2) = C\left(F_{X|W_1}(x|w_1), G_{Y|W_2}(y|w_2)|w_1,w_2\right) \tag{2.14}$$

Then, $H_{XY|w_1,w_2}(x,\infty|w_1,w_2) = C\left(F_{X|W_1}(x|w_1),1|w_1,w_2\right) = F_{X|W_1}(x|w_1)$ which is the conditional marginal distribution of $X|W_1$, which is the conditional marginal distribution of $(X,Y)|W_1$. Similarly,$H_{XY|w_1,w_2}(\infty,y|w_1,w_2) = C\left(1,G_{Y|W_2}(y|w_2)|w_1,w_2\right) = F_{Y|W_2}(y|w_2)$ which is the conditional marginal distribution of $Y|W_2$, which is the conditional marginal distribution of $(X,Y)|W_2$. Thus the function $H_{XY|w_1,w_2}$ cannot be the joint distribution of $(X,Y)|(W_1,W_2)$. The only case $H_{XY|W_1,W_2}$ will be joint distribution of $(X,Y)|(W_1,W_2)$ is when $F_{X|W_1}(x|w_1) = F_{X|w_1,w_2}(x|w_1,w_2)$ and $F_{Y|W_1}(y|w_1) = F_{Y|w_2,w_2}(y|w_1,w_2)$ that is when some conditioning variables affect the conditional distribution of one variable but not the other.

Remark.

Corollary 2.1 can be formulated in a manner that can help to extract the conditional copula from any bivariate conditional distribution. Fermanian and Scaillet (2005a) [15] provided the definition of conditional copula in terms of sigma field $\Im$.

Empirical Copula

Deheuvels (1979) [10] introduced empirical copulas. He called them empirical dependence functions.

Definition 2.9: Let $(x_t,y_t),\ \ t = 1,2,...,n$ denote a sample of size n from a continuous bivariate distribution. The empirical copula function C_n is given by

$$C_n\left(\frac{t_1}{n},\frac{t_2}{n}\right) = \frac{1}{n}\sum_{t=1}^{n} I\left(x_t \le x_{t_1}, y_t \le y_{t_2}\right) \tag{2.15}$$

where x_{t_i} and y_{t_i}, $1 \le t_1 \cdots \le t_i \le n$ are order statistics from the sample and I is the usual indicator function. Note that the sample versions of several measures of dependence/association can be expressed in terms of empirical copula analogous to the population versions of measures of dependence that can be expressed in terms of copulas.

2.4 Methods of Constructing Copulas

First we note that there are several methods of constructing copulas, see for example Nelsen (2006) [32] and in this section we considered a few selected methods.

1. **Inversion of Marginals:** Let H he a two-dimensional distribution function with known continuous marginal distributions F and G and their inverses F^{-1} and G^{-1}, respectively. Then we can find the unique copula as

$$C(u,v) = H(F^{-1}(u), G^{-1}(v)) \tag{2.16}$$

Example: Gaussian Copula. Let $\phi_\rho(x,y)$ be a standard bivariate normal distribution function with coefficient of correlation ρ and let $\phi(\cdot)$ represents univariate standard normal margin. Then, the Gaussian (Normal) copula is given by

$$C(u,v;\rho) = \phi_\rho\left(\phi^{-1}(u), \phi^{-1}(v)\right) \tag{2.17}$$

2. **Method of Frailties:** The concept of frailty has been extensively used in survival analysis. Frailties can be used to construct various families of copulas (Oakes, 1989 [33]). Denote the frailty by Z and assume that it is non negative with density $f(z)$, distribution function $F(z)$, and Laplace transform $L_Z(t) = E_Z(exp(-z))$. Now, consider random variables X and Y with survival functions S_X and S_Y respectively. Let B_X and B_Y. be two continuous baseline survival functions. Assume that the random variables X and Y are conditionally independent given the frailty Z . That is

$$P(X > x, Y > y|Z = z) = P(X > x|Z = z)P(Y > y|Z = z) \tag{2.18}$$
$$= S_X(x|Z = z)S_Y(y|Z = z) \tag{2.19}$$

and the joint survival function $S_{XY}(x,y)$ can be derived from

$$S_{XY}(x,y) = \int_0^\infty S_X(x|Z = z)S_Y(y|Z = z)f(z)dz \tag{2.20}$$

Let us consider a special case where we assume that

$$S_X(x|Z = z) = [B_X(x)]^z \text{ and } S_Y(y|Z = z) = [B_Y(y)]^z \tag{2.21}$$

such that the unconditional survival functions are given by,

$$S_X(x) = L_Z[-\log B_X(x)] \text{ and } S_Y(y) = L_Z[-\log B_Y(y)] \tag{2.22}$$

It follows that,
$S_{XY}(x,y) = \int_0^\infty S_X(x|Z=z)S_Y(y|Z=z)f(z)dz$
$= \int_0^\infty [B_X(x)]^z[B_Y(y)]^z f(z)dz$
$= E_Z\{B_X(x)B_Y(y)^z\}$
$= E_Z\{exp(z[\log B_X(x) + \log B_Y(y)])\}$
$= L_Z[-(\log B_X(x) + \log B_Y(y))]$

Further we have,

$$-\log(B_X(x)) = L_Z^{-1}(S_X(x)) \text{ and } -\log(B_Y(y)) = L_Z^{-1}(S_Y(y)) \tag{2.23}$$

Thus, we can write the joint survival function is given by,

$$(S_{XY}(x,y) = L_Z\left[L_Z^{-1}(S_X(x)) + L_Z^{-1}(S_Y(y))\right] \tag{2.24}$$

Finally, it follows that the copula is given by

$$C(u,v) = L_Z\left[L_Z^{-1}(u) + L_Z^{-1}(v)\right] \tag{2.25}$$

where $u = S_X(x)$ and $v = S_Y(y)$

Example: Gumbel-Hougaard Copula. Suppose the frailty Z has a positive stable distribution with Laplace transform given by

$$L_Z(t) = exp(-t^{1/\delta})$$

Hence, using (2.8) we have the Gumbel-Hougaard family of copulas given by

$$C(u,v) = exp\left\{-\left[(-\log(u))^\delta + (-\log(v))^\delta\right]^{1/\delta}\right\}$$

Remark: The advantage of the frailty concept is that it allows us to interpret the dependence of the random variables in such away that a frailty that contributes to the dependence of the random variables may exist.These class of copulas generated by frailty models are a subclass of the Archimedean copulas defined on a more general functions called "Generators" discussed below.

3. **Method of Generators:** Another important way of constructing copulas is using a function ϕ called a "Generator" that satisfies the following properties

   (a) $\phi : [0,1] \to [0,\infty]$

   (b) ϕ is continuous and strictly decreasing function,

   (c) ϕ is such that $\phi(1) = 0$

   The pseudo inverse of ϕ is the function $\phi^{[-1]}[0,\infty] \to [0,1]$ given by,

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & 0 \leq t \leq \phi(0) \\ 0 & \phi(0) \leq t \leq \infty \end{cases}$$

   If $\phi(0) = \infty$, then the pseudo inverse $\phi^{[-1]} = \phi^{-1}$, inverse of ϕ.

   Theorem 2.10: Let the Generator ϕ be as defined above and let $\phi^{[-1]}$ be the pseudo inverse of ϕ. Then, the function C from $[0,1]^2 \to [0,1]$ given by

$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \tag{2.26}$$

   is a copula if and only if ϕ is convex. Copulas of the form (2.9) are called Archimedean copulas. Some examples and properties are discussed in subsequent sections; see details in Nelsen (2006) [32], Chapter 4).

   Example: (Frank Copula) Suppose the generator is given by

$$\phi(t) = -log\left(\frac{1-e^{-\delta t}}{1-e^{-\delta}}\right) \text{ and inverse } \phi^{-1}(u) = -\frac{1}{\delta}\log\left[1-\left(1-e^{-\delta}\right)e^{-u}\right]$$

   Thus, the resulting copula is the Frank's family given by

$$C(u,v) = -\frac{1}{\delta}\log\left(1 - \frac{\left(1-e^{-\delta u}\right)\left(1-e^{-\delta v}\right)}{1-e^{-\delta}}\right)$$

4. **Polynomial Approximations to copulas** Polynomials are used in the approximation of distribution functions. To this end, Hoeffding (1940) [48] used Legendre polynomials to approximate bivariate "standard distributions" later called copulas. Recently, Sancetta and Satchell 2004 [37]) introduced Bernstein copula defined by Bernstein polynomials that are closed under differentiation.

Definition 2.10 (Bernstein copula, Sancetta and Satchell (2004) [37]):
Let $\alpha\left(\frac{k_1}{m_1},\frac{k_2}{m_2}\right)$ valued constant, $k \in \{1,2\}$ such that $1 \leq k \leq m_i, i = 1,2$ and the Bernstein polynomials given by

$$P_{k_i,m_i}(u_i) = \binom{m_i}{k_i} u_{ii}^{k}(1-u_i)_i^k \text{ If } C_B : [0,1]^2 \to [0,1] \tag{2.27}$$

where

$$C_B(u_1,u_2) = \sum_{k_1=1}^{m_1}\sum_{k_2=1}^{m_2} \alpha\left(\frac{k_1}{m_1},\frac{k_2}{m_2}\right) P_{k_1,m_1}(u_1)P_{k_2,m_2}(u_2) \tag{2.28}$$

satisfies the properties of the copula function in Definition (2.4), then C_B is called the Bernstein copula. Some statistical properties of Bernstein copula are studied by Sancetta and Satchell (2004) [37]. In particular, they have shown that the coefficients of the Bernstein copula C_B have a direct interpretation as the points of some arbitrary approximated copula, C, i.e.,

$$\alpha\left(\frac{k_1}{m_1},\frac{k_2}{m_2}\right) = C\left(\frac{k_1}{m_1},\frac{k_2}{m_2}\right) \tag{2.29}$$

2.5 Important families of copulas

In the copula literature several examples of copula functions are introduced. Most of the copulas belong to members of families with one or more real valued parameters. In this section we present a very brief overview of some parametric families of copulas. Extensive surveys of families of copulas can be found in Joe (1997) [22] and Nelsen (2006) [32].

2.5.1 Farlie-Gumbel-Morgenstern's (FGM) Family

The FGM family is a symmetric and one parameter family of copulas whose functional form is a polynomial in u and in v. That is, for $\delta \in [-1,1]$ then the function is given by

$$C_\delta(u,v) = uv + \delta uv(1-u)(1-v) \tag{2.30}$$

is the FGM family of copulas.

2.5.2 Archimedean Copulas

The Archimedean copulas are given by the general expression

$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)) \tag{2.31}$$

where ϕ is a "Generator" defined in Section 2.5 above.
The Archimedean copulas have a wide range of applications as indicated in Nelsen (2006) [32] because of the ease with which they can be constructed, the many parametric families of copulas belonging to this class, the great variety of dependence structures offered by this class and the nice properties possessed by the members of this class such as extension to higher dimensions, convenient in developing criteria for copula model selection and establishing relationship with nonparametric measures of associations, etc. Genest and MacKay (1986a,b) [19] have presented many properties of this class of copulas, which made them extremely suitable for statistical applications. Next, we present brief descriptions to four members of Archimedean family of copulas that are used in the subsequent chapters.

1. **Gumbel's Family.**
   The Gumbel copula is given by

   $$C(u,v;\delta) = exp\left(-(\tilde{u}^{\delta} + \tilde{v}^{\delta})^{\frac{1}{\delta}}\right) \tag{2.32}$$

   where $\tilde{u} = -\log(u), \tilde{v} = -\log(v)$
   with copula density,

   $$c(u,v;\delta) = C(u,v;\delta)[uv]^{-1} \frac{(\tilde{u}\tilde{v})^{\delta-1}}{(\tilde{u}^{\delta} + \tilde{v}^{\delta})^{2-\frac{1}{\delta}}} \left[(\tilde{u}^{\delta} + \tilde{v}^{\delta})^{\frac{1}{\delta}} + \delta - 1\right] \tag{2.33}$$

   This family has properties like upper tail dependence, product copula for $\delta = 1$, Frechet-Hoeffding upper bound copula for $\delta \to \infty$ (see Joe, 1997, Family B6, p. 142). These properties are the subject of the next section.

2. **Clayton's (or Kimeldrof and Sampson's) Family.**
   This copula is referred as Clayton's copula, for example, in Nelsen (2006) [32] and Kimeldrof and Sampson's copula in Joe (1997) [22]. This copula is given by

   $$C(u,v;\delta) = (\tilde{u}^{\delta} + \tilde{v}^{\delta}) - 1)^{\frac{-1}{\delta}} \qquad \delta \geq 0 \tag{2.34}$$

with copula density,

$$c(u,v;\delta) = (1+\delta)[uv]^{-\delta-1}(\tilde{u}^{\delta}+\tilde{v}^{\delta})^{-2-\frac{1}{\delta}} \tag{2.35}$$

This family has properties like lower tail dependence, product copula for $\delta \to 0$, Frechet Hoeffding upper bound copula for $\delta \to \infty$ (see Joe, 1997 [22], Family B4, p. 141).

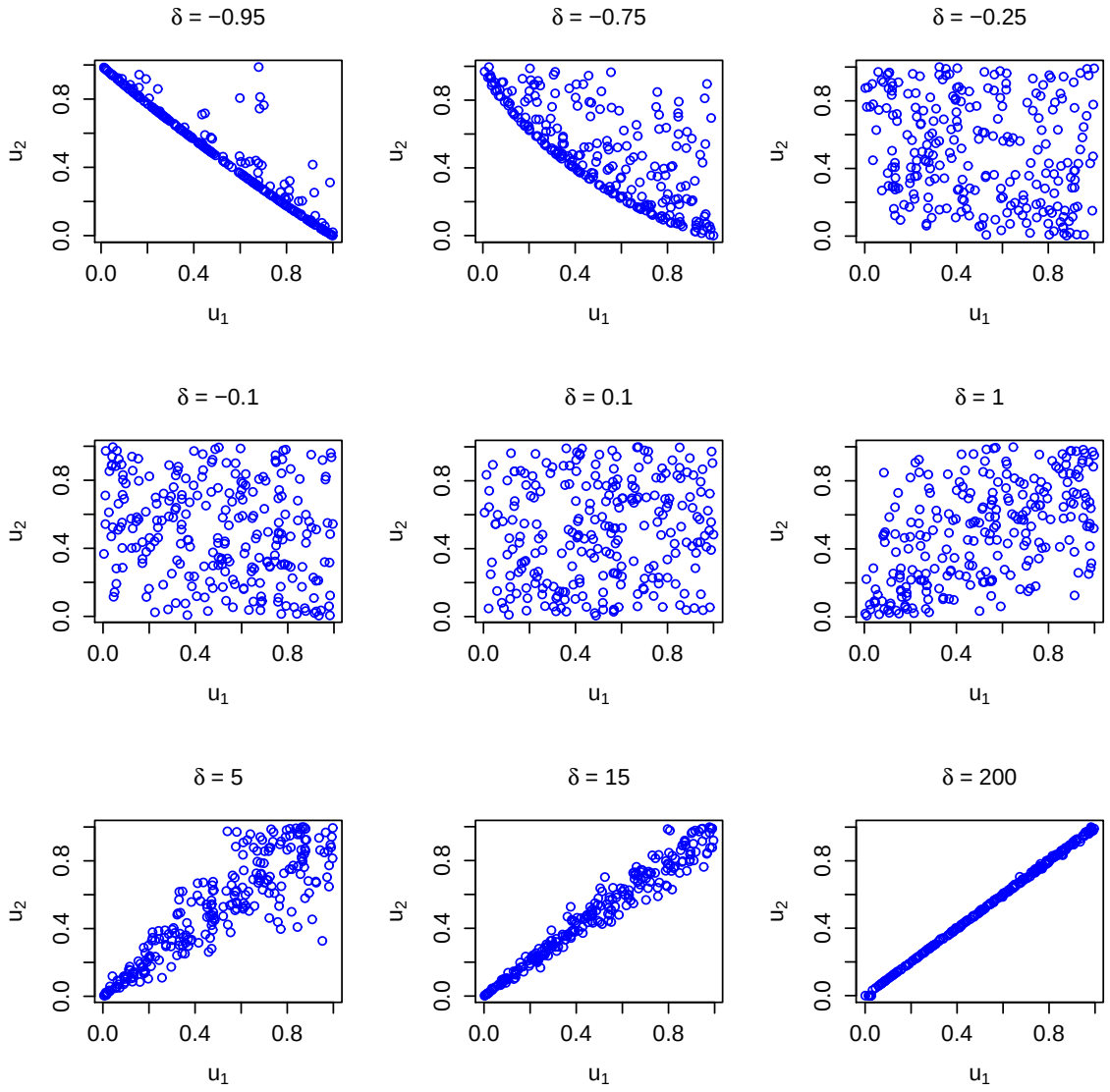

Fig. 2.3 Bivariate random samples of size 250 from various Clayton copulas

3. **Frank's Family.**
   The Frank copula is given by

$$C(u,v;\delta) = -\delta^{-1}\log\left(\frac{\left[(1-e^{-\delta})-(1-e^{-\delta u})(1-e^{-\delta v})\right]}{(1-e^{-\delta})}\right) \tag{2.36}$$

   with copula density,

$$c(u,v;\delta) = \frac{\left[\delta(1-e^{-\delta})e^{-\delta(u+v)}\right]}{\left[(1-e^{-\delta})-(1-e^{-\delta u})(1-e^{-\delta v})\right]^2} \tag{2.37}$$

   This family has properties like reflection symmetry, product copula for $\delta \to 0$, Frechet Hoeffding lower and upper bound copulas for $\delta \to -\infty$ and $\delta \to \infty$, respectively (see Joe, 1997 [22], Family B3, p. 141).

4. **Joe's Family**
   The Joe's copula is given by

$$C(u,v;\delta) = 1-\left(\bar{u}^{\delta}+\bar{v}^{\delta}-\bar{u}^{\delta}\bar{v}^{\delta}\right)^{1/\delta} \qquad \delta \geq 1 \tag{2.38}$$

   with copula density,

$$c(u,v;\delta) = \left(\bar{u}^{\delta}+\bar{v}^{\delta}-\bar{u}^{\delta}\bar{v}^{\delta}\right)^{-2+1/\delta}\bar{u}^{\delta-1}\bar{v}^{\delta-1}\left(\delta-1+\bar{u}^{\delta}+\bar{v}^{\delta}-\bar{u}^{\delta}\bar{v}^{\delta}\right) \tag{2.39}$$

   where $\bar{u} = 1-u$ and $\bar{v} = 1-v$ This family has properties like upper tail dependence, product copula for $\delta = 1$, Frechet Hoeffding upper bound copula for $\delta \to \infty$ (see Joe, 1997 [22],Family B5, pp. 141-142).

2.5.3 Elliptical Copulas

Elliptical copulas are copulas associated with elliptical distributions. For definitions of elliptical distributions see for example Embrechts et al. (2002) [14]. The Gaussian copula and the t-copula are examples of elliptical copulas. For example, the Gaussian copula is given by

$$C(u,v;\delta) = \phi_{\delta}\left(\phi^{-1}(u),\phi^{-1}(v)\right) \qquad -1 \leq \rho \leq 1 \tag{2.40}$$

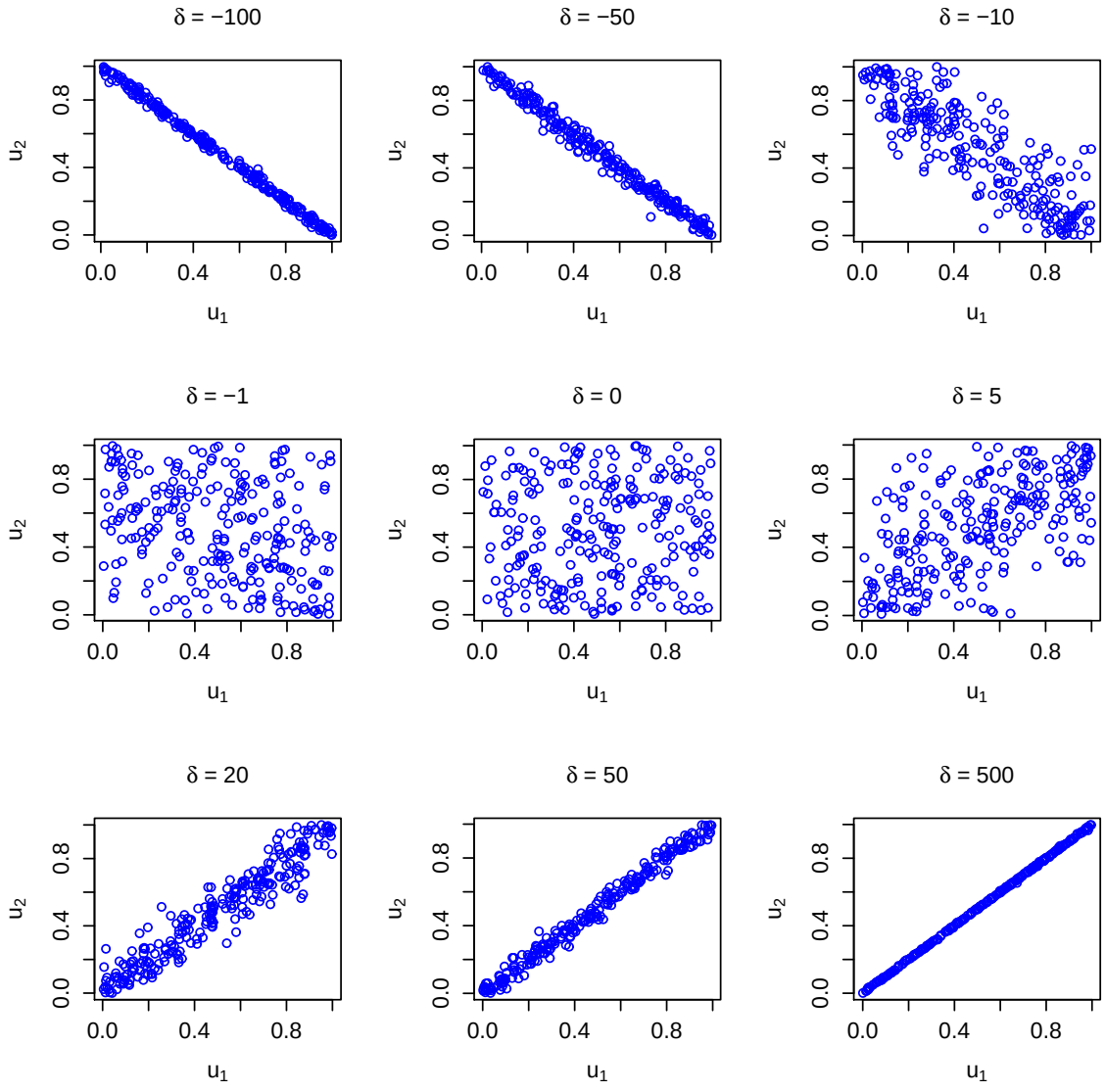

Fig. 2.4 Bivariate random samples of size 250 from various Frank copulas.

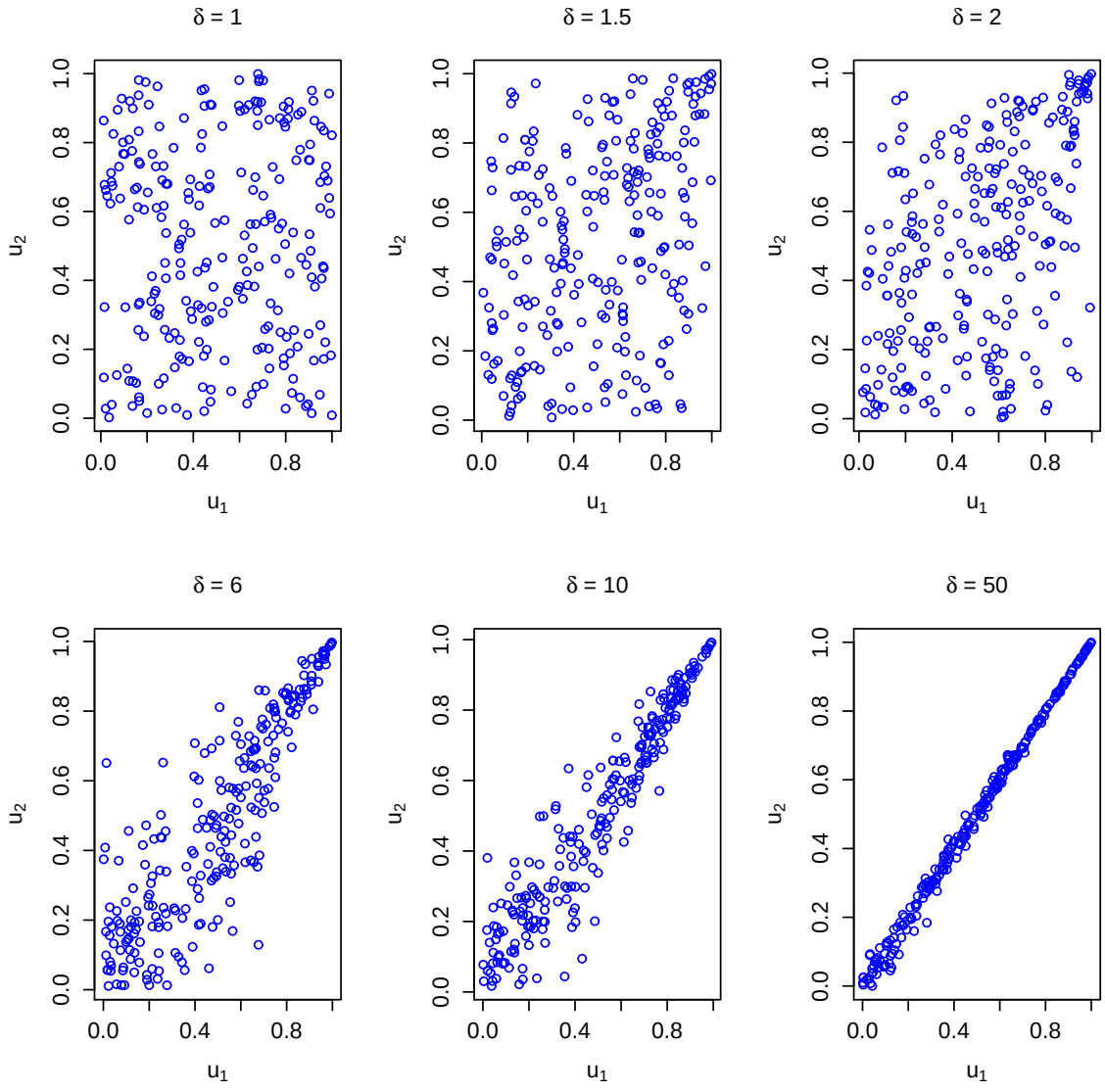

Fig. 2.5 Bivariate random samples of size 250 from various joe copulas

with copula density,

$$c(u,v;\delta) = \frac{1}{(1-\delta^2)} \left\{ e^{\frac{-1}{2(1-\delta^2)}\left(X^2+Y^2-2\delta XY\right)} \right\} \left\{ e^{\frac{1}{2}\left(X^2+Y^2\right)} \right\} \tag{2.41}$$

where $X = \phi^{-1}(u)$ and $Y = \phi^{-1}(v)$ We note that multi-parametric families of copulas are extensively discussed in Joe (1997) [22]. Here we give one example from two parametric families of copulas.

2.5.4 Frechet's Family

The Frechet's family is a two-parameter family of copulas given as a convex linear combination of the copulas π, W and M , i.e.,

$$C_{\alpha\beta}(u,v) = \alpha M(u,v) + (1-\alpha-\beta)\pi(u,v) + \beta W(u,v) \tag{2.42}$$

where $\alpha, \beta \in [0,1]$ with $\alpha + \beta \leq 1$

2.6 Dependence Measures

In this section we deal with different ways in which copulas can be used in the study of dependence between random variables. There are a variety of ways to describe and measure the dependence or association between random variables. The multivariate normal distribution and linear correlation have been the basis for most dependence modeling in practice. In fact, linear correlation is a good measure of dependence in the context of multivariate normal distributions or elliptical distributions in general but it has several problems if applied to distributions other than elliptical distributions (Embrechts et al, 2002 [14]). Alternative measures of dependence using nonparametric methods are also common in practice. Copulas are capable of describing these measures of dependence and also capturing any form of dependence structure. Many measures of dependence have been introduced and studied in the literature. Among them the most widely used measures are: the Pearson's coefficient of correlation (r) , Spearman's rho (ρ_s) introduced by Spearman(1904) [46], and Kendall's tau (τ)introduced by kendall (1938) [25]. Definitions of these measures and their relation to copulas can be found, for example, in Nelsen (2006) [32] and Joe (1997) [22]. Let X and Y be two random variables and let F, G, H, and C be defined as in Section 2.2.

Pearson correlation coefficient

$$r(x,y) = \frac{1}{\sigma(x)\sigma(y)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H(x,y) - F(x)G(y)]\, dxdy \tag{2.43}$$

where $\sigma(\cdot)$ represent for standard deviation.

Spearman's rho

$$\rho(x,y) = 12 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H(x,y) - F(x)G(y)]\, dF(x)dG(y) \tag{2.44}$$

Kendall's tau

$$\tau(x,y) = 4\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y)dH(x,y)\right] - 1 \tag{2.45}$$

Moreover, Schweizer and Wolff (1981) studied the following three nonparametric measures of association σ_{SW}, γ and K based on L_1, L_2 and L_∞ distances, respectively. These measures are given by,

1. $\sigma_{SW}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |H(x,y) - F(x)G(y)|\, dF(x)dG(y)$
2. $\gamma(X,Y) = \left(90 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [H(x,y) - F(x)G(y)]^2 dF(x)dG(y)\right)^{\frac{1}{2}}$
3. $\kappa(X,Y) = 4 \sup_{x,y\in R} |H(x,y) - F(x)G(y)|$

Now, it is important to note that these measures of dependence can be expressed in terms of copulas (see Schweizer and Wolff, 1981 [40]). Let $U = F(X)$ and $V = G(Y)$be probability integral transformations. Using Sklar's framework $[H(x,y) = C(F(X), G(y))]$ we have

- $r(X,Y) = \frac{1}{\sigma(X)\sigma(Y)} \int_0^1\int_0^1 [C(u,v) - uv]\, dF^{-1}(u)dG^{-1}(v)$
- $\rho(X,Y) = 12 \int_0^1\int_0^1 [C(u,v) - uv]\, dudv$
- $\tau(X,Y) = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1$
- $\sigma_{SW}(X,Y) = \int_0^1\int_0^1 |C(u,v) - uv|\, dudv$
- $\gamma(X,Y) = \left(90 \int_0^1\int_0^1 [C(u,v) - uv]^2 dudv\right)^{\frac{1}{2}}$

- $\kappa(X,Y) = 4 \sup_{x,y \in R} |C(u,v) - uv|$

Remark: Using the copula transformations at least two things are apparent. First, integrals over the plane are transformed into integrals in the unit square. Second, the nonparametric measures, $\rho_S, \tau, \sigma_{SW}, \gamma$ and κ, distinguished from the coefficient of correlation, r, in that they are functions of the copula alone, i.e., it is only the coefficient of correlation that depends on the marginals but all others are scale free measures.

Furthermore, in the study of properties of Archimedean copulas, Genest and Mac Kay (1986b [19]) provided a simplified version of Kendall's tau, which is stated in the following theorem.

Theorem 2.11 (Kendall's Tau for Archimedean Copulas, Genest and MacKay (1986b) [19]):

Let X and Y be random variables with an Archimedean copula C having generator ϕ.The population version of Kendall's tau , r, for the random variables X and Y is given by

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt \tag{2.46}$$

where ϕ' is the first derivative of ϕ. See Genest and MacKay (1986b) [19] for proof.

Example: Consider the Clayton copula with generator $\phi(t) = \frac{t^{-\delta}-1}{\delta}$ for $\delta > 0$. Then, Kendall's tau for this copula is

$$\tau = 1 + 4 \int_0^1 \frac{t^{\delta}+1}{\delta} dt = \frac{\delta}{\delta+2} \tag{2.47}$$

2.6.1 Measure of concordance

Definition: (Nelsen (2006) [32], page 136) A numeric measure of association between two continuous random variables X and Y whose copula is C is a measure of concordance if it satisfies the following properties:

1. κ is defined for every pair X, Y of continuous random variables;
2. $-1 = \kappa_{X,-X} \leq \kappa_C \leq \kappa_{X,X} = 1$;
3. $\kappa_{X,Y} = \kappa_{Y,X}$
4. if X and Y are independent, then $\kappa_{X,Y} = \kappa_{C^{\perp}} = 0$;
5. $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$;

6. if $C_1 \prec C_2$ then $\kappa_{C_1} \leq \kappa_{C_2}$;

7. if $\{(X_n, Y_n)\}$is a sequence of continuous random variables with copulas C_n, and if $\{C_n\}$ converges point wise to C, then $\lim_{n\to\infty} \kappa_{C_n} = \kappa_C$

Among all the measures of concordance, three famous measures play an important role in non-parametric statistics: the Kendall's tau, the Spearman's rho and the Gini indice. They could all be written with copulas, and we have (Schweitzer and Wolff [1981] [40]) Nelsen [1998] [31] presents some relationships between the measures τ and ρ , that can be summarized by a bounding region. In Figure 2.6 , we have plotted the links between τ and ρ for normal copulas. We note that the relationships are similar. However, some copulas do not cover the entire range $[-1, 1]$ of the possible values for concordance measures. For example, Kimeldorf-Sampson, Gumbel, Galambos and Hausler-Reiss copulas do not allow negative dependence.

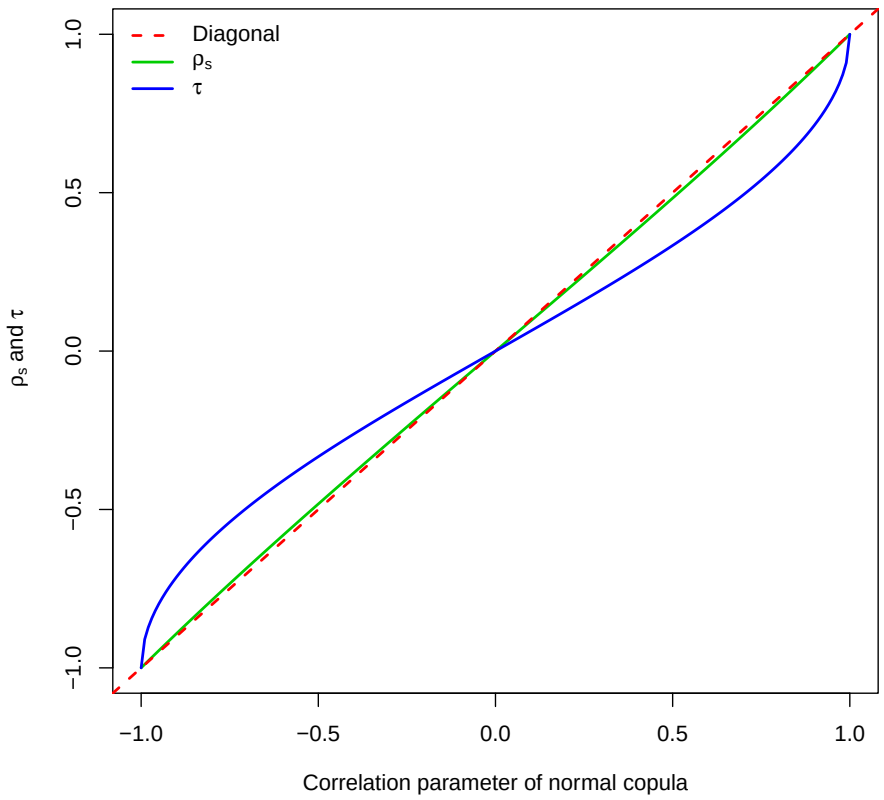

Fig. 2.6 Spearman's rho and Kendall's tau for normal copulas.

2.6.2 Measure of dependence

Definition 2.10: (Nelsen (1998) [31] A numeric measure of association between two continuous random variables X and Y whose copula is C is a measure of dependence if it satisfies the following properties:

1. δ is defined for every pair X, Y of continuous random variables;
2. $0 = \delta_{C^\perp} \leq \delta_C \leq \delta_{C^+} = 1$
3. $\delta_{X,Y} = \delta_{Y,X}$
4. $\delta_{X,Y} = \delta_{C^\perp} = 0$ if and only if X and Y are independent;
5. $\delta_{X,Y} = \delta_{C^+} = 1$ if and only if each of X and Y almost surely a strictly monotone function of the other;
6. if h_1 and h_2 are almost surely strictly monotone functions on $Im(X)$ and $Im(Y)$ respectively, then
$$\delta_{h_1(x),h_2(y)} = \delta_{X,Y}$$
7. if $\{(X_n, Y_n)\}$is a sequence of continuous random variables with copulas C_n, and if $\{C_n\}$ converges point wise to C, then $\lim_{n\to\infty} \delta_{C_n} = \delta_C$

2.6.3 Tail Dependence

Tail dependence measure refers to the dependence that arises between random variables from extreme observations. Upper tail dependence exists when large extreme values occur jointly, while lower tail dependence exists when small extreme values occur jointly. Another important feature of copulas is that the upper and lower tail dependence measures can be expressed in terms of copulas.

Definition 2.11: (Upper tail dependence):

Let X and Y be two continuous random variables with marginal distribution functions F and G, and copula C. Then the coefficient of upper tail dependence of X and Y is:

$$\lambda_u = \lim_{u\to 1^-} Pr\left[X > F^{-1}(u) | Y > G^{-1}(u)\right] \tag{2.48}$$

provided that a limit $\lambda_u \in (0,1]$ exists. If $\lambda_u \in (0,1]$, X and Y are said to be asymptotically dependent in the upper tail; if $\lambda_u = 0$, X and Y are said to be asymptotically independent

in the upper tail. Using the probability integral transforms $U = F(X)$ and $V = G(Y)$ the coefficient of upper tail dependence can be rewritten as

$$\lambda_u = \lim_{u \to 1^-} Pr[U > u | V > v] \tag{2.49}$$

Further, this can be expressed in terms of copulas as follows:

$$\lambda_u = \lim_{u \to 1^-} \frac{1 - 2u - C(u,u)}{1-u} \tag{2.50}$$

Definition 2.12: (Lower tail dependence):

Let X and Y be two continuous random variables with marginal distribution functions F and G, and copula C. Then the coefficient of lower tail dependence of X and Y is:

$$\lambda_L = \lim_{u \to 0^+} Pr\left[X \leq F^{-1}(u) | Y \leq G^{-1}(u)\right] = \lim_{u \to 0^+} Pr[U \leq u | V \leq v] \tag{2.51}$$

provided that a limit $\lambda_L \in (0,1]$ exists. If $\lambda_L \in (0,1]$, X and Y are said to be asymptotically dependent in the lower tail; if $\lambda_L = 0$, X and Y are said to be asymptotically independent in the lower tail.This can be expressed using copulas as

$$\lambda_L = \lim_{u \to 0^+} \frac{C(u,u)}{u} \tag{2.52}$$

Independence

The most commonly used dependence property is the assumption that there is no dependence that is the random variables are independent. If two continuous random variables are independent, then their copula is the product copula, n , see Section (2.2) property 5.

Perfect Dependence

We have seen in Section (2.2) property 6 that one function is almost surely a monotone function of the other (perfect dependence) whenever the copula is either W or M . We note that other dependence forms that lie between the extremes independence and perfect dependence and their relationship with copulas are discussed in Joe (1997) [22] and Nelsen (2006) [32].

Chapter 3

Statistical inference of copulas

3.1 Estimation and Asymptotic Properties

In this section we present various methods related to estimation of copulas. Consider a correctly specified copula that belongs to a parametric family $C = \{C(\cdot, \delta), \delta \in \Re\}$. Consistent and asymptotically normally distributed estimates can be obtained through maximum likelihood methods (one-stage or two-stages) mainly using a fully parametric or a semi-parametric approach, see for details in Joe (1997) [22], Genest et al. (1995) [18] and Shih and Louis (1995) [41]. Alternatively, one can estimate copulas by nonparametric methods using empirical copula (Deheuvels, 1979) [10]. Recently, for multivariate time series models Patton (2006a, b) [35] [34] employed two-stage maximum likelihood estimation for conditional copulas. Similarly, Chen and Fan (2006) [4] have studied the two-stage approach for a Markov time series under the semiparametric setup. We now give a survey of various approaches of estimation suggested in the literature for the i.i.d. multivariate setup. Consider a random sample of d variables and n number of observations represented by the vector $x = (x_{1t}, x_{2t}, ...x_{dt}), t = l, ..., n$. Consider a copula-based model for the random vector X , with distribution function

$$H(x_1, x_2.....x_d; \theta_1, ...\theta_d, \delta) = C(F_1(x_1, \theta_1), ...F_d(x_d, \theta_d); \delta) \tag{3.1}$$

where $\theta_i, i = l, ..., d$, each being scalar or vector of parameter(s) of the marginal distributions $F_i, i = 1, 2, ..., d$ and δ is a scalar or vector of the copula parameter(s). Let $\eta = (\theta_1, ...\theta_d, \delta)$ be the parameter vector to be estimated.

3.1.1 Maximum likelihood Estimation (MLE)

Consider the log-likelihood function based on (2.2), that is given by :

$$l(\eta;x) = \sum_{j=1}^{n}\sum_{i=1}^{d} \log f_i(x_{it};\theta_i) + \sum_{i=1}^{n} \log c\,(F_1(x_{1i},\theta_1),.....F_d(x_{di},\theta_d);\delta) \tag{3.2}$$

Thus, the maximum likelihood estimator $\hat{\eta}$ of the parameter vector η is the solution of

$$\frac{\partial l(\eta,x)}{\partial \eta} = 0$$

Let η_0 be the true value of η. Under standard regularity conditions, consistency and asymptotic normality properties of the estimator $\hat{\eta}$ have been established; see, for example, Joe (1997) [22]. That is,

$\sqrt{n}(-\eta_0) \to N(0, I^{-1})$ in distribution,

where I is the Fisher Information matrix.

The estimation of Maximum likelihood Estimation (MLE) can be carried out by using R program's codes which is include in Appendix section,then that corresponding out put given below.

```
> summary(mle)
Call: fitMvdc(data = X, mvdc = mcc, start = start)
Maximum Likelihood estimation based on 2000
2-dimensional observations.
Copula:  claytonCopula
Margin 1 :
        Estimate Std. Error
m1.mean 0.004963      0.018
m1.sd   0.990165      0.007
Margin 2 :
        Estimate Std. Error
m2.rate   0.9853      0.021
Clayton copula, dim. d = 2
      Estimate Std. Error
alpha     4.99      0.124
The maximized loglikelihood is -2923
Optimization converged
Number of loglikelihood evaluations:
```

```
function gradient
      78       16
```

Note that as the number of parameters increases computational problems may arise when using MLE method. An alternative approach that helps to reduce computational problem is the use of two-stage method discussed below.

3.1.2 Inference Functions for Margins (IFM) method

For copula-based models this method is first followed by Joe and Xu (1996) [24], in order to exploit the fundamental idea of copula theory that allow the separation of the univariate margins from the dependence structure or copula. This method consists of estimating the model parameter by finding solutions for conveniently defined set of inference (estimating) functions. In this method the score functions of the margins and the copula constitute the set of inference functions. Thus, according to the two-stage estimation method, the parameters of the marginal distributions are estimated separately from the parameters of the copula. In other words, the estimation process is divided into two steps:

Step 1.

Estimating the parameters $\theta_i, i = 1,2,...d$, using maximum likelihood method from the respective marginal log-likelihoods, i.e., consider the marginal log-likelihoods

$$l_i(\theta_i; x_i) = \sum_{j=1}^{n} \log f_i(x_{ij}; \theta_i) \tag{3.3}$$

Then, the maximum likelihood estimator $\tilde{\theta}_i$ of the parameter vector θ_i is the solution of Thus, we have the estimates $\tilde{\theta}_1, ...\tilde{\theta}_d$ for the marginal parameters

Step 2.

To Estimate the vector of copula parameters δ, first substitute the marginal estimates $\tilde{\theta}_1, ...\tilde{\theta}_d$ to the copula log-likelihood

$$l_c(\delta; x, \tilde{\theta}_1, ...\tilde{\theta}_d) = \sum_{t=1}^{n} \log c\left(F_1(x_{1t}; \tilde{\theta}_1), ...F_d(x_{dt}; \tilde{\theta}_d); \delta\right) \tag{3.4}$$

Then, the pseudo maximum likelihood estimator S of the parameter vector $\tilde{\delta}$ is the solution to

$$\frac{\partial l_c\left(\delta; x, \bar{\theta}_1, ...\bar{\theta}_d\right)}{\partial \delta} = 0 \tag{3.5}$$

Therefore, the estimator $\bar{\eta} = \left(\tilde{\theta}_1, ...\tilde{\theta}_d, \tilde{\delta}\right)$ is referred as the two-stage maximum likelihood estimator of $\eta = (\theta_1, ...\theta_d, \delta)$. Joe and Xu (1996) [24] established asymptotic normality of this estimator. Let the inference functions be denote by the vector $g(x, \eta)$:

The estimation of Inference Functions for Margins (IFM) can be carried out by using R program's codes which is include in Appendix section,then that corresponding out put given below.

```
> summary(ifme)
Call: fitCopula(copula, data = data, method = "ml")
Fit based on "maximum likelihood" and 2000 2-dimensional observations.
Clayton copula, dim. d = 2
      Estimate Std. Error
alpha    4.932      0.116
The maximized loglikelihood is 1886
Optimization converged
Number of loglikelihood evaluations:
function gradient
       4        4
```

3.1.3 Canonical Maximum Likelihood (CML) method

Both the MLE and IFM methods are based on some specified parametric form of the univariate margins. The choice of the best possible fit distributions for the margins is of course crucial. Hence, to avoid the risk involved in choosing parametric marginal models and without much information loss on the dependence structure, one can consider non-parametric marginal models. The semi parametric copula-based model estimation procedure also involves two steps:

Step 1.

Transform the observed data vector $x_t = (x_{1t},,,...,x_{dt}), \quad t = 1,2,...,n$ into uniform values called pseudo-observations using rescaled empirical distributions defined by

$$F_{in}(x) = \frac{1}{n}\sum_{t=1}^{n}[X_{it} \leq x] \text{ for } i = 1,2,.....d \tag{3.6}$$

where $1[\cdot]$ represents indicator function. Let the transformed observation be denoted by $(\tilde{u_1}, ..., \tilde{u}_{dt}) = (F_{1n}(x_{1t}), ..., F_{dn}(x_{dt}))$ for $t = 1, ...n$.

Step 2.

To Estimate the vector of copula parameters δ, first substitute the transformed observations $(\tilde{u}_1, ..., \tilde{u}_{dt})$ into the copula log-likelihood

$$l_c(\delta; \tilde{u}_1, ..., \tilde{u}_{dt}) = \sum_{t=1}^{n} \log c(\tilde{u}_1, ..., \tilde{u}_{dt}; \delta) \tag{3.7}$$

Step 3. Then, the pseudo maximum likelihood estimator $\tilde{\delta}$ of the parameter vector δ is the solution to

$$\frac{\partial l_c}{\partial \delta} = 0 \tag{3.8}$$

Genest et al. (1995) established consistency and asymptotic normality of the semi-parametric estimator $\tilde{\delta}_n$.

The estimation of the Canonical Maximum Likelihood (CML) can be carried out by using R program's codes which is include in Appendix section,then that corresponding out put given below.

```
summary(CML)
Call: fitCopula(copula, data = data, method = "mpl")
Fit based on "maximum pseudo-likelihood" and 2000
2-dimensional observations.
Clayton copula, dim. d = 2
      Estimate Std. Error
alpha    4.932      0.173
The maximized loglikelihood is 1886
Optimization converged
Number of loglikelihood evaluations:
function gradient
       4        4
```

3.1.4 Nonparametric estimation

Nonparametric estimation of copulas can be obtained by using the empirical copula discussed in Section (2.5.3). Here we present how the empirical copulas can be used to estimate dependence measures like Spearman's ρ_s and Kendall's τ and how these in turn can be used to estimate the copula parameter. The sample version of Spearman's rho ρ_s in terms of the

empirical copula C_n, is

$$\hat{\rho}_s = \frac{12}{n^2-1}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\left\{C_n\left(\frac{t_1}{n},\frac{t_2}{n}\right)-\frac{t_1 t_2}{n^2}\right\} \tag{3.9}$$

Similarly, the sample version of Kendall's τ is

$$\bar{\tau} = \frac{2n}{n-1}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\left\{C_n\left(\frac{t_1}{n},\frac{t_2}{n}\right)C_n\left(\frac{t_1-1}{n},\frac{t_2-1}{n}\right)-C_n\left(\frac{t_1-1}{n},\frac{t_2}{n}\right)C_n\left(\frac{t_1}{n},\frac{t_2-1}{n}\right)\right\} \tag{3.10}$$

Thus, using the relationship between the copula parameter and the population versions of either Kendall's tau or Spearman's rho, we can obtain nonparametric estimate for the copula parameter. For example, in the case of Clayton copula, a nonparametric estimator of the copula parameter d can be obtained from the sample version of Kendall's tau $\bar{\tau}$ a.s $\bar{\delta} = \frac{2\bar{\tau}}{(1-\bar{\tau})}$.

The Nonparametric estimation of the copulas based on Kendall's tau can be carried out by using R program's codes which is include in Appendix section,then that corresponding out put given below.

```
Call: fitCopula(copula, data = data, method = "itau")
Fit based on "inversion of Kendall's tau" and 1000
2-dimensional observations.
Clayton copula, dim. d = 2
      Estimate Std. Error
alpha    4.842      0.274
```

The Nonparametric estimation of the copulas based on Spearman's rho can be carried out by using R program's codes which is include in Appendix section,then that corresponding out put given below.

```
Call: fitCopula(copula, data = data, method = "irho")
Fit based on "inversion of Spearman's rho" and 1000
2-dimensional observations.
Clayton copula, dim. d = 2
      Estimate Std. Error
alpha    4.961      0.294
```

3.2 Choice of Copula Model

Several varieties of copula models exist in the literature. An important issue in the implementation of any copula model in practice is the choice of an appropriate parametric copula. Several studies have attempted to provide graphical and formal statistical model selection procedures. Breymann, et al. (2003) [2] applied the Akaike Information Criterion (AIC) to several parametric copulas and selected the one with the minimum AIC value. Genest and Rivest (1993) [21] proposed a method based on the distribution function of the copula in order to select the best possible fit copula model among the Archimedean copulas. Chen et al. (2004) [5] proposed simple tests for models of dependence between multiple financial time series. Chen and Fan (2005) [3] used pseudo likelihood ratio tests for copula selection for semi-parametric copula-based multivariate models under copula misspecification. Fermanian (2005) [15] and Dobric and Schmid (2005) [12] proposed goodness of fit tests for copula models. Next, we present two procedures that are used for choosing the best-fit copula model.

1. **Graphical approach based on the distribution function of the copula**
   Genest and Rivest (1993) [21] proposed copula selection for Archimedean copulas based on parametric and nonparametric estimates of the distribution function of copulas. The idea is to choose among the Archimedean copulas, the one that most closely resembles the nonparametric copula estimate.

   The Parametric Estimate: Let the parametric distribution function of a copula, K, be given by $K(w) = P_1\left[C(u,v;\delta) \leq w\right]$ for we $w \in (0,1)$. For Archimedean copulas this can be expressed in terms of the generator ϕ, as

$$K(w) = w - \frac{\phi(w)}{\phi'(w)} \text{ (see Genest and Rivest, 1993 [21])}. \tag{3.11}$$

   Letting $\lambda(w) = \frac{\phi(w)}{\phi'(w)}$ we have $\lambda(w) = w - K(w)$
   Then, the estimated distribution function for the copula is given by

$$\hat{K}(w) = Pr\left[C(u,v;\hat{\delta}) \leq w\right], \quad w \in (0,1) \text{ such that } \hat{\lambda}(w) = w - \hat{K}(w) \tag{3.12}$$

   where $\hat{\delta}$ can be obtained by any one of the methods discussed in Section 2.4 Note that the parametric estimate $\hat{\lambda}(\cdot)$ can be computed for several Archimedean copulas.

   The Nonparametric Estimate: A nonparametric distribution function estimate, K_n of K can be obtained using the procedure proposed by Genest and Rivest (1993) [21] in two steps:

(a) Let W_i be defined for the bivariate data $(x_t, y_t), t = 1, ..., n$, as follows:

$$W_t = \frac{\#\{(x_s,y_s): x_s < x_t, y_s < y_t\}}{n+1} \quad 1 \leq t \leq n$$

(b) The nonparametric distribution function estimate K_n is then given by

$$K_n(w) = \frac{\sum_{t=1}^{n} 1\,[w - W_t]}{n}, \quad w \in (0,1)$$

Similarly, a nonparametric estimate of $\lambda(\cdot)$ is given by

$$\lambda_n(w) = w - K_n(w)$$

Then, plot the parameter estimates $\hat{\lambda}(w)$ for several Archimedean copulas and the non-parameter estimate $\lambda_n(w)$ against w and choose the copula that provides a curve closely resemble to the nonparametric case (see for further details and example in Genest and Rivest, 1993 [21]).

2. **Akaike Information Criterion (AIC)**
In case where maximum likelihood estimation is used the choice of the best possible fit copula can be done by comparing the likelihood contributions of the various copulas under consideration, using the Akaike information Criterion (AIC). The AIC is given by

$$AIC = 2(-\text{loglikelihood})/n + 2p/n,$$

where p is the number of parameters in the model. Thus, one can choose the "best" fit copula that corresponds to the minimum AIC.

3. **Bayesian Information Criterion (BIC)**
BIC is another criteria based on log-likelihood and is often used to choose between a finite set of models. Kole et al (2006) [27]. used BIC to asses the likelihood of their copula model. Tibishrani et al. define BIC as

$$BIC = k\log(n) - 2l$$

where k is the number of parameters of the model, l is the log-likelihood of the fitted model and n is the number of observations. Evidently, if the BIC is chosen, the penalty for two parameter families is stronger than when using the AIC.

3.3 Simulation from Copulas

In this section, we present one simulation algorithm that helps to generate bivariate observations from a copula-based distribution function with given marginals. Other algorithms can be found, for example, in Nelsen (2006) [32]. Thus, the algorithm to simulate from a two-dimensional copula is as follows:

1. First generate two independent Uniform$(0,1)$ random variates v_1 and v_2;
2. Set $u_1 = v_1$;
3. Compute $u_2 = C_{2|1}^{-1}(v_2|u_1)$, where $C_{2|1}^{-1}(\cdot|u_1)$ is the inverse of $C_{2|1}^{-1}(\cdot|u_1)$. In many cases closed forms for $C_{2|1}^{-1}(\cdot|u_1)$ are not available. Hence, a numerical solution for MT is required from $v_2 C_{2|1}^{-1}(u_2|u_1)$.
4. Then (u_1,u_2) is the simulated value from the copula C.
5. Repeat I-IV, say n times, to obtain n bivariate uniform observations $(U_{1t},U_{2t})'t = l,...,n$, from the copula C .Furthermore, to simulate from a copula-based bivariate distribution function H with given marginal distributions $F,,i = l,2$, one more additional step is required. That is,
6. Invert each $w(\cdot)$, using the marginal distributions as

$$x_{it} = F_i^{-1}(u_{it}) \quad i = 1,2 \text{ and } t = 1,.....n$$

Thus, $(x_{1t},x_{2t}),t = l,...,n$ are n bivariate observations from a copula-based distribution function H.

Chapter 4

Copula Regression Theory

In this chapter, we will briefly discuss about theoretical results of copula regression and some important properties.Nelsen (2006) [32]provides an excellent overview of recent theoretical results on the study of copulas in his text book. One of the other researches in this area closest to copula regression analysis in the work of Cuadras (1992) [8]. In Cuadras (1992) [8] research study, he proposed a method of constructing multivariate distributions where both univariate marginals and a correlation matrix are given. The proposed method yields totally linear regressive family of distributions. Recently, Engin A. Sungur(2005) [47] studied about copula regression theoretical ideas and he proposed a deeper meaning to regression equation than a simple functional form.

Definition 4.1. LetU,V be a random variables with uniform marginals on the$[0,1]$ and copula C. We will call $E_C(V|U=u)$ copula regression function of V on U and denote it by $r_c(u)$.

Definition 4.2

Suppose that X and Y are continuous with marginal distribution functions F and G, respectively, joint distribution function H, and copula C. Then $U=F(X)$ and $V=G(Y)$ are uniform $(0,1)$ random variables with joint distribution function C. Here are some facts that we will use in the rest of this article:

1. The conditional distribution function for V given $U=u$, say $C_u(v)$ is:

$$P(V \leqslant v|U=u) = \frac{\partial C(u,v)}{\partial u} = C_u(v) \tag{4.1}$$

2. The conditional distribution function for Y given $X=x$ is

$$P(Y \leqslant y|X=x) = P(V \leqslant G(y)|U=F(x)) = \frac{\partial C(u,v)}{\partial u} = C_u(v) \tag{4.2}$$

3. The copula regression function of V on U is:

$$E_C(V|U=u) = r_c(u) = 1 - \int_0^1 C_u(v)dv \tag{4.3}$$

4. The regression function of Y on X is:

$$E[Y|X=x] = \hat{y} = G^{-1}(1 - \int_0^1 C_u(v)dv) = G^{-1}r_c(F(x)) \tag{4.4}$$

where G^{-1} is the inverse distribution function of Y .

For the facts (4.1) and (4.2), please see Cherubini et al. (2004, pp. 177–178, 182) [6]. Facts (4.3) and (4.4) follow directly from application of integration by parts on the definition of expectation. Note that any monotone strictly increasing transformation of X and Y will only change the marginal distribution, leaving the joint behavior of X and Y untouched. The structure of the article is as follows. First we provide basic properties of the copula regression function. In the last section we discuss some of the implications of our findings in application, such as transformation for the linearity.

4.1 Properties of Copula Regression function

In this section, we have to state some important properties of Copula regression function those are important for our future proof.

Theorem:4.1 We define the two- dimensional independence copula $C^0(u,v)$ is the CDF of d mutually independent Uniform $(0,1)$ random variables.The two-dimensional co-monotonicity copula $C^+(u,v)$ characterizes perfect positive dependence. The two-dimensional counter-monotonicity copula $C^-(u,v)$ is defined as the CDF of $(u,1-u)$, which has perfect negative dependence. Then,

1. If $C^0(u,v) = uv$ then $r_0(u) = 1/2$
2. If $C^+(u,v) = \min(uv)$ then $r_+(u) = u$
3. If $C^-(u,v) = \max(u+v-1,0)$ then $r_-(u) = 1-u$

Proof.

From above result equation 4.3, when we substitute corresponding copula functions $C^0(u,v), C^+(u,v)$ and $C^-(u,v)$, after integrate with respect to v then we will get above results such as $r_0(u) = 1/2, r_+(u) = u$ and $r_-(u) = 1-u$.

Theorem:4.2

1. $r_C(u) = E_C(u) = 1 - \int_0^1 \left[C_{u_0}(v) + \sum_{l=1}^{n-1} \frac{C^l{}_{u_0}(v)}{(l)} (u-u_0)^l \right] dv$
   where $C^l{}_{u_0}(v) = \frac{\partial^l C_u(v)}{\partial u^l}|_{u=u_0}$ and u_r is an interior point to the interval joining u and u_0.
2. $r_C(u) = E_C(V_u) \geq r(1 - C_u(r))$ for any $r \in (0,1]$
3. $E(V) = \int_0^1 E_C(V_u)du = \int_0^1 r_C(u)du = \frac{1}{2}$
4. $\rho_C = 3\left\{1 - 4\int_0^1 \left[\int_0^u r_C(w)dw\right] du\right\}$where ρ_C is the Pearson's correlation.

Proof:

1. By using Taylor's expansion we can get

   $$E_C(V_u) = E_C(V_{u0}) + \sum_{l=1}^{n-1} \frac{E^{(l)}{}_C[V_{u0}]}{l!}(u-u_0)^l + \frac{E^{(n)}{}_C[V_{ur}]}{n!}(u-u_r)^n$$

   where u_r is an interior point to the interval joining u and u_0, and

   $$E^{(l)}{}_C[V_{u0}] = \frac{d^l E_C[V_u]}{du^l}|_{u=u_0}$$

   The remainder term will be represented by

   $$R_{n-1} = \frac{E^{(n)}{}_C[V_{ur}]}{n!}(u-u_r)^n$$

   From(3)
   $E_C[V_u] = r_C(u) = 1 - \int_0^1 C_{u_0}(v)dv - \sum_{l=1}^{n-1} \frac{\int_0^1 C^l{}_{u_0}(v)}{l!}(u-u_0)^l - \frac{\int_0^1 C^n{}_{u_r}(v)}{n!}(u-u_r)^n$
   $= 1 - \int_0^1 \left[C_{u_0}(v) + \sum_{l=1}^{n-1} \frac{C^l{}_{u_0}(v)}{(l)}(u-u_0)^l \right] dv$
   where

   $$C^l{}_{u_0}(v) = \frac{\partial^l C_u(v)}{\partial u^l}|_{u=u_0}$$

2. Directly follows from the Markov's inequality, see Dudewicz and Mishra (1988, p. 296).
3. Directly follows from the fact that$E[V] = E[E_C(V_u)]$.
4. See the appendices.

   Note that for the polynomial regression is $\frac{\int_0^1 C^i{}_{u_0}(v)}{i!}dv$the coefficient of u^i.

4.2 Linear Copula Regression Functions

Now let us take a closer look at linear copula regression functions (Sungur (2005) [47]) by using two families:

1. $C(u,v) = uv[1+\theta(1-u)(1-v)], \qquad \theta \in [-1,1]$

$$(\text{Farlie} - \text{Gumbel} - \text{Morgenstern}, \text{F} - \text{G} - \text{M}). \tag{4.5}$$

2. $C(u,v) = \theta_1 \min\{u,v\} + (1-\theta_1-\theta_2)uv + \theta_2 \max\{u+v-1,0\}$

$$\theta_1, \theta_2 \in [0,1] \text{and} \theta_1+\theta_2 \leq 1 (\text{Frechet and Mardia, F} - \text{M}) \tag{4.6}$$

The following table provides the form of the copula regression function both in terms of dependence parameter and in terms of the intercept parameter of the regression equation, referred as the modified , and the Pearson's correlation:

Table 4.1 Parameter Estimation

	Example 2.1: F-G-M family	Example 2.2: F-M family
$r_C(u)$	$\frac{3-\theta}{6} + \frac{\theta}{3}u$	$\frac{1-\theta_1+\theta_2}{2} + (\theta_1-\theta_2)u$
Modified $r_C(u)$	$\alpha + (1-2\alpha)u, \alpha = \frac{3-\theta}{6}$	$\alpha + (1-2\alpha)u, \alpha = \frac{1-\theta_1+\theta_2}{2}$
Pearson's correlation	$\frac{\theta}{3}$	$\theta_1 - \theta_2$

The patterns that can be observed in these two examples could be generalized. Let ξ_L be class of copulas with linear copula regression functions, i.e.,

$$\xi_L = \left\{ C : 1 - \int_0^1 \frac{\partial C(u,v)}{\partial u} dv = \alpha + \beta u \right\}$$

Note that,

$$\begin{aligned} 1 - \int_0^1 \frac{\partial C(u,v)}{\partial u} dv &= \alpha + \beta u \\ \Rightarrow \frac{\partial}{\partial u} \int_0^1 C(u,v) dv &= 1 - \alpha - \beta u \\ \Rightarrow \int_0^1 C(u,v) dv &= (1-\alpha)u - \beta \frac{u^2}{2} + k \end{aligned}$$

Since

$$\int_0^1 C(1,v)dv = \int_0^1 vdv = \frac{1}{2} = (1-\alpha) - \beta\frac{1}{2} + k$$
$$\Rightarrow k = \frac{1}{2} - (1-\alpha) + \frac{\beta}{2}$$

On the other hand, by using (3),

$$E(V) = \int_0^1 E_C[V_u]du = \frac{1}{2}$$
$$\Rightarrow \int_0^1 \left[1 - \int_0^1 C_u(v)dv\right] du = \frac{1}{2}$$
$$\Rightarrow \int_0^1 \int_0^1 C_u(v)dvdu = \frac{1}{2} = \int_0^1 (1-\alpha-\beta u)\, du$$
$$\Rightarrow 1 - \alpha - \frac{\beta}{2} = \frac{1}{2}$$
$$\Rightarrow \beta = 1 - 2\alpha$$

Therefore, we can write

$$\int_0^1 C(u,v) = (1-\alpha)u - (1-2\alpha)\frac{u^2}{2} \text{ and } E_C[V_u] = \alpha + (1-2\alpha)u$$

Theorem:4.3

A copula has a linear regression function, i.e., $C \in \xi_L$, if and only if $r_C(u) = \alpha + (1-2\alpha)u$ or $r_C(u) = \frac{1-\beta}{2} + \beta u$. Note that the particular relationship between the intercept and slope coefficient for the linear copula regression functions provides a way of testing for linearity. In statistically linear copula regression functions the coefficients will be related with the Pearson's correlation.

Theorem:4.4

If $C \in \xi_L$ then Pearson's correlation $\rho_C = 1 - 2\alpha$ and $r_C(u) = \frac{1-\rho_C}{2} + \rho_C$u
Proof:
Suppose we assume that $C \in \xi_L$. Since $\int_0^1 C(u,v)dv = (1-\alpha)u - (1-2\alpha)\frac{u^2}{2}$

$$\rho_C = 12\int_0^1 \int_0^1 [C(u,v) - uv]dudv = 12\int_0^1 \int_0^1 C(u,v)dudv - 3$$
$$= 12\int_0^1 \left[\int_0^1 C(u,v)dv\right] du - 3$$
$$= 12\int_0^1 \left[(1-\alpha)u - (1-2\alpha)\frac{u^2}{2}\right] du - 3 = 1 - 2\alpha$$

Suppose $r_C(0) = \frac{1-\rho_C}{2}, r_C(1/2) = \frac{1}{2}$ and $r_C(1) = \frac{1+\rho_C}{2}$.Therefore, one can observe the strength of linear relationship by checking the intercept.Suppose that $C(u,v:\theta)$ and $T(u,v:\alpha)$ are two copulas with linear regression functions. Define a function $h(\cdot)$ such that $\rho_C(\theta) = \rho_T(h(\theta))$ then $r_C(u;\theta) = r_T(u;h(\theta))$. One important implication of this observation on dependence model building is the following. Suppose that one wants to predict V given $U = u$. The optimal solution is to use regression function of the copula C that minimizes mean squared error $MSE = E[V - g(U)]^2$. This scheme has a practical limitation, since it requires the knowledge of the C. Under the assumption of linear copula regression function it is enough to work with any target copula with linear regression function. This approach does not require a complete knowledge on the form of the copula. The needed value of the Pearson's correlation coefficient can be estimated easily from the data.

4.3 Multiple linear Copula Regression Function

In 2005, Sungur [47] developed the approach to the directional dependence by using bivariate copula based linear regression, But in practically most of the regression analysis has more than one regressors so in this study we are going to discuss about multivariate copula based regression. In this upcoming section, we will expand Sungur's (2005) [47] bivariate result to multivariate case.

Definition 4.3 (Extended definition from Definition 4.2):

If $C(u_0,u_1,u_2\cdots\cdot u_d)$ of U_0 and U with $U = (U_1,U_2\cdots U_k)^T$ then we define multivariate copula regression of U_0 on $U = u$ is given by

$$r^C_{U_0|U}(u) = E(U_0|U=u) = \int_0^1 u_0 \frac{c(u_0,u_1,u_2\cdots\cdot u_d)}{C_U(u)} du_0$$

$$= 1 - \frac{1}{c_U(u)} \int_0^1 C_{U_0|U}(u_0) du_0 \tag{4.7}$$

where $C_U(u) = C(1,u_1,u_2\cdots\cdot u_d)$ is the marginal distribution of U and

$c(u_0,u_1,u_2\cdots\cdot u_d) = \dfrac{\partial^{k+1} C(u_0,u_1\cdots_k)}{\partial u_0,\partial u_1\cdots\cdot\partial u_k}$

$c_U(u) = \dfrac{\partial^k C_U(u)}{\partial u_1\cdots\cdot\partial u_k}$ are the joint copula density function of U_0 and U,

the marginal copula density of $U$and the conditional copula distribution of U_0 given U is given by,

$$C_{U_0|U}(u_0) = \frac{\partial^k C(u_0, u_1, u_2 \cdots u_d)}{\partial u_1 \cdots \partial u_k} \tag{4.8}$$

Similarly we can extend the theorem 4.2 for multivariate case,such as expected value of multivariate copula regression function $E\left[r^C_{U_0|U}(u)\right] = \frac{1}{2}$. In following section we propose two new multivariate copula regressions function, which is developed from sungur (2005) [47] research paper, he studied just bi-variate case Farlie-Gumbel-Morgenstern (FGM) family. In this study also we will consider same families but multivariate case.

4.4 Multivariate Non-Exchangeable FGM Copula

Úbeda-Flores et al (2005) [36] proposed an asymmetric bivariate copulas, which generalizes several families such as the known Farlie–Gumbel–Morgenstern (FGM) family of copulas and others. In this study we will discuss u_0 as response variable and u_1, u_2 are regressors (trivariate extension) as follows.

Theorem:4.5 -Extended theorm 2.3 in Úbeda-Flores et al (2005) [36]

Let f_0, f_1 and f_2 be a real non zero functions defined on uniform $[0,1]$ and C be a function on $[0,1]^3$ given by

$$C(u_0, u_1, u_2) = u_0u_1u_2 + f_0(u_0)f_1(u_1)u_2 + f_0(u_0)f_2(u_2)u_1 + f_1(u_1)f_2(u_2)u_0 \tag{4.9}$$

Then C is a copula if and only if

1. $f_i(0) = f_i(1) = 0$ for all $i = 0,1,2$
2. f_i absolutely continuous for all $i = 0,1,2$ and
3. $\min(\alpha_0\beta_1 + \alpha_0\beta_2 + \alpha_1\beta_2, \alpha_1\beta_0 + \alpha_0\beta_2 + \alpha_1\beta_2, \alpha_0\beta_1 + \alpha_2\beta_0 + \alpha_1\beta_2, \alpha_0\beta_1 + \alpha_0\beta_2 + \alpha_1\beta_1, \alpha_1\beta_0 + \alpha_2\beta_0 + \alpha_1\beta_2, \alpha_1\beta_0 + \alpha_0\beta_2 + \alpha_2\beta_1, \alpha_0\beta_1 + \alpha_2\beta_0 + \alpha_1\beta_2, \alpha_1\beta_0 + \alpha_2\beta_0 + \alpha_2\beta_1) \geq -1$

where $\alpha_i = \inf(f_i'(u_i); u_i \in A_i) < 0$ and $\beta_i = \sup(f_i'(u_i); u_i \in A_i) > 0$ with $A_i = \{u_i \in I; f_i'(u_i) \text{exist}\}$ for all $i = 0,1,2$.

Proof: The proof is straight forward,Bivariate case proof given in Úbeda-Flores et al (2005) [36] paper, in this case similarly we have to proof (follow same procedures) trivariate case.

By using the generalized Farlie–Gumbel–Morgenstern (FGM) family of copulas, given above

equation 4.9 and results from definition 4.3, we can derive copula based regression function, which is given below

Since the copula distribution function is given by,

$$C(u_0,u_1,u_2) = u_0u_1u_2 + f_0(u_0)f_1(u_1)u_2 + f_0(u_0)f_2(u_2)u_1 + f_1(u_1)f_2(u_2)u_0$$

In this we are going to consider, u_0 as response variable and u_1, u_2 are regressors. Then,

$C_U(u) = C(1,u_1,u_2) = u_1u_2 + f_0(1)f_1(u_1)u_2 + f_0(1)f_2(u_2)u_1 + f_1(u_1)f_2(u_2)1$

$= u_1u_2 + f_1(u_1)f_2(u_2)$

$c_U(u) = \dfrac{\partial^2 C_U(u)}{\partial u_1 \partial u_2}$

$= \dfrac{\partial^2 C(1,u_1,u_2)}{\partial u_1 \partial u_2} = 1 + f_1'(u_1)f_2'(u_2)$

Now we consider,

$C_{U_0|U}(u_0) = \dfrac{\partial^2 C(u_0,u_1,u_2)}{\partial u_1 \partial u_2}$

$= \dfrac{\partial^2 \left(u_0u_1u_2 + f_0(u_0)f_1(u_1)u_2 + f_0(u_0)f_2(u_2)u_1 + f_1(u_1)f_2(u_2)u_0\right)}{\partial u_1 \partial u_2}$

$= u_0 + f_0(u_0)f_1'(u_1) + f_0(u_0)f_2'(u_2) + u_0f_1'(u_1)f_2'(u_2)$

Then we consider,

$\displaystyle\int_0^1 C_{U_0|U}(u_0)du_0 = \int_0^1 \left(u_0 + f_0(u_0)f_1'(u_1) + f_0(u_0)f_2'(u_2) + u_0f_1'(u_1)f_2'(u_2)\right) du_0$

$= \dfrac{1}{2} + \left(f_1'(u_1) + f_2'(u_2)\right) \displaystyle\int_0^1 f_0(u_0)du_0 + f_1'(u_1)f_2'(u_2)\frac{1}{2}$

$= \dfrac{1}{2}\left(1 + f_1'(u_1)f_2'(u_2)\right) + \left(f_1'(u_1) + f_2'(u_2)\right) \displaystyle\int_0^1 f_0(u_0)du_0$

Then multivariate(trivariate) copula regression of U_0 on $U = u$ is given by

$r^C_{U_0|U}(u) = E(U_0|U=u) = 1 - \dfrac{1}{c_U(u)} \displaystyle\int_0^1 C_{U_0|U}(u_0)du_0$

$= 1 - \dfrac{\frac{1}{2}\left(1 + f_1'(u_1)f_2'(u_2)\right) + \left(f_1'(u_1) + f_2'(u_2)\right) \int_0^1 f_0(u_0)du_0}{1 + f_1'(u_1)f_2'(u_2)}$

$$r^C_{U_0|U}(u) = 1 - \frac{\frac{1}{2}\left(1 + f_1'(u_1)f_2'(u_2)\right) + \left(f_1'(u_1) + f_2'(u_2)\right) \int_0^1 f_0(u_0)du_0}{1 + f_1'(u_1)f_2'(u_2)} \tag{4.10}$$

Similarly we can derive multivariate copula based regression line equation for other families such that and Frechet and Mardia (FM) family, Archimedean family etc.

4.5 Gaussian copula marginal regression models

In very general terms, a regression model is expressed as

$$y_i = f(x_i, \varepsilon_i; \lambda)$$

where $f(\cdot)$ is a suitable function of the regressors x_i and of an unobserved stochastic variable ε_i, commonly denoted as the error component. It is assumed that the regression model is known up to a vector of parameters λ. Among the possible specifications for the function $f(\cdot)$, a useful choice is,

$$y_i = F_i^{-1}(\phi(\varepsilon_i); \lambda) \quad i = 1, \cdots, n,$$

where ε_i is a standard normal variable and $F_i(\cdot; \lambda) = F(\cdot | x_i; \lambda)$ and are the cumulative distribution functions of y_i given x_i and of a standard normal variate, respectively. By the integral transformation theorem, the regression model ensures the desired marginal distribution for the response y_i and specifies ε_i in the familiar terms of a normal error. Specification includes all possible parametric regression models for continuous and noncontinuous responses. For example, the Gaussian linear regression model $y_i = {x_i}^T \beta$

4.6 Implementation by using R Programming

In this study, we are going to use `gcmr` package in R programming. The package `gcmr()` which allows to fit Gaussian copula models by using maximum likelihood in the continuous case and by maximum simulated likelihood in the discrete case. The arguments of `gcmr()` are the following

```
gcmr(formula, data, subset, offset, marginal, cormat, start,
fixed, options = gcmr.options(...), model = TRUE, ...)
```

The function has standard arguments for model-frame specification (Chambers and Hastie 1993) such as a formula, the possibility to restrict the analysis to a subset of the data, to set an offset, or to fix contrasts for factors. The specific arguments of `gcmr()` include the two key arguments `marginal` and `cormat`, which specify the marginal part of the model and the copula correlation matrix, respectively. Finally, there are three optional arguments to supply starting values (start), fix the values of some parameters (fixed) and set the fitting options (options). The rest of this section describes the components of `gcmr()` and the related methods. According to Guido Masarotto et al (2012) [28] briefly discussed about this package usage in regression analysis.

The fitting options in `gcmr()` are set by argument options or by a direct call to function,

```
gcmr.options(seed = round(runif(1, 1, 1e+05)), nrep = c(100, 1000),
no.se = FALSE, method = c("BFGS", "Nelder-Mead", "CG"), ...)
```

The quantile residuals are computed by method

```
residuals.gcmr(object, type = c("conditional", "marginal"),
method = c("random", "mid"), ...)
```

The profile log-likelihood can be obtained with a call to method

```
profile.gcmr(fitted, which, low, up, npoints = 10, display = TRUE,
alpha = 0.05, progress.bar = TRUE, ...)
```

Table 4.2 Marginals models available in gcmr with the default link function.

marginal.gcmr	Distribution	Dispersion
beta.marg(link = "logit")	beta	yes
binomial.marg(link = "logit")	binomial	no
Gamma.marg(link = "inverse")	gamma	yes
gaussian.marg(link = "identity")	Gaussian	yes
negbin.marg(link = "log")	negative binomial	yes
poisson.marg(link = "log")	Poisson	no
weibull.marg(link = "log")	Weibull	yes

Table 4.3 Correlation models available in gcmr package.

cormat.gcmr	Correlation
arma.cormat(p, q)	ARMA(p,q)
cluster.cormat(id, type)	longitudinal/clustered data
ind.cormat()	independence
matern.cormat(D, alpha)	Matern correlation

In following chapter we will use this R package for comparing copula based regression with OLS and GLM model and also we will provide all R programming codes and output in next section.

Table 4.4 Functions and methods available for objects of class gcmr.

Function	Description
`print()`	simple printed display of coefficient estimates
`summary()`	standard regression output
`coef()`	coefficient estimates
`vcov()`	covariance matrix of coefficient estimates
`fitted()`	fitted means for observed data
`residuals()`	quantile residuals
`estfun()`	estimating functions for sandwich estimators
`bread()`	bread matrix for sandwich estimators
`terms()`	terms of model components
`model.frame()`	model frame
`model.matrix()`	model matrix
`logLik()`	maximized log-likelihood
`plot()`	diagnostic plots of quantile residuals
`profile()`	profile likelihood for focus coefficients
`coeftest()`	partial Wald tests of coefficients
`waldtest()`	Wald tests of nested models
`lrtest()`	likelihood ratio tests of nested models
`AIC()`	information criteria

Chapter 5

Results and Discussions

In this chapter,we attempt to provide a deeper explanation to copula regression than a simple functional form and propose our new theoretical results. We will use R-package `gcmr` implements maximum likelihood inference for Gaussian copula marginal regression.

5.1 Comparing copula,OLS and GLM Regressions

The example 01 considers the well-known longitudinal study on epileptic seizures described in Diggle et al. (2013) [11]:

```
R> data("epilepsy", package = "gcmr")
```

The data comprise information about 59 individuals observed at five different occasions each. The baseline observation consists of the number of epileptic seizures in a eight-week interval, followed by four measurements collected at subsequent visits every two weeks. Available variables are the patient identifier `id`, the patient age, the indicator `trt` whether the patient is treated with progabide `(trt = 1)` or not `(trt = 0)`, the number of epileptic seizures counts, the observation period time in weeks, that is `time = 8` for baseline and `time = 2` for subsequent visits, and the indicator visit whether the observation corresponds to a `visit (visit = 1)` or the baseline `(visit = 0)`. Diggle et al. (2013) [11] analyzed the seizure data with the method of generalized estimating equations assuming a log-linear regression model for counts with the logarithm of time used as offset and covariates trt, visit and their interaction. Moreover, Diggle et al. (2013) [11] suggested to omit an "outlier" patient - here corresponding to patient id = 49 - with an extremely high seizure count at baseline (151 counts) that even double after treatment (302 counts after 8 weeks of measurement). Indeed, estimated model coefficients vary considerably if this patient is set aside. The corresponding Gaussian copula analysis described below assumes a negative binomial marginal distribution

with mean specified as in Diggle et al. (2013) [11]. We start the analysis assuming a working independence correlation matrix for the Gaussian copula:

```
> names(epilepsy)
[1] "id" "age" "trt" "counts" "time" "visit"
> str(epilepsy)
'data.frame': 295 obs. of 6 variables:
$ id : int 1 1 1 1 1 2 2 2 2 2 ...
$ age : int 31 31 31 31 31 30 30 30 30 30 ...
$ trt : int 0 0 0 0 0 0 0 0 0 0 ...
$ counts: int 11 5 3 3 3 11 3 5 3 3 ...
$ time : num 8 2 2 2 2 8 2 2 2 2 ...
$ visit : num 0 1 1 1 1 0 1 1 1 1 ...
> library(gcmr)
> library(stats)
> #************************************************************
> data("epilepsy", package = "gcmr")
>
> Gaussiancouplamodel <- gcmr(counts ~ offset(log(time)) + visit + trt
+visit:trt+count,data = epilepsy, subset = (id != 49),
marginal = gaussian.marg,+ cormat = cluster.cormat(id, type = "ind"))
Error in eval(predvars, data, env) : object 'count' not found
> OLSmodel <-lm(counts~ offset(log(time)) + visit + trt + visit:trt,
data = epilepsy)
> GLMmodel <-glm(counts~ offset(log(time)) + visit + trt + visit:trt,
family = gaussian,data = epilepsy)
>
> summary(Gaussiancouplamodel)

Call:
gcmr(formula = counts ~ offset(log(time)) + visit + trt + visit:trt,
    data = epilepsy, subset = (id != 49), marginal = negbin.marg,
cormat = cluster.cormat(id, type = "ind"))

Coefficients marginal model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.34759    0.16649   8.094 5.77e-16 ***
```

```
visit         0.11187    0.18802   0.595    0.552
trt          -0.10685    0.23057  -0.463    0.643
visit:trt    -0.30237    0.26118  -1.158    0.247
dispersion    0.73421    0.07153  10.264  < 2e-16 ***

No coefficients in the Gaussian copula
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

log likelihood = 948.06,  AIC = 1906.1
> summary(OLSmodel)

Call:
lm(formula = counts ~ offset(log(time)) + visit + trt + visit:trt,
    data = epilepsy)

Residuals:
    Min      1Q  Median      3Q     Max
-24.786  -6.607  -3.968   2.032 119.355

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.7063     3.0890   9.293  < 2e-16 ***
visit       -20.7923     3.4536  -6.020 5.23e-09 ***
trt           0.8594     4.2615   0.202    0.840
visit:trt    -1.4988     4.7645  -0.315    0.753
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.35 on 291 degrees of freedom
Multiple R-squared:  0.2428,    Adjusted R-squared:  0.235
F-statistic:  31.1 on 3 and 291 DF,  p-value: < 2.2e-16

> summary(GLMmodel)

Call:
```

```
glm(formula = counts ~ offset(log(time)) + visit + trt + visit:trt,
    family = gaussian, data = epilepsy)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-24.786   -6.607   -3.968    2.032  119.355

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  28.7063     3.0890   9.293  < 2e-16 ***
visit       -20.7923     3.4536  -6.020 5.23e-09 ***
trt           0.8594     4.2615   0.202    0.840
visit:trt    -1.4988     4.7645  -0.315    0.753
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 267.1697)

    Null deviance: 99762  on 294  degrees of freedom
Residual deviance: 77746  on 291  degrees of freedom
AIC: 2491.6

Number of Fisher Scoring iterations: 2

>
> res1 <- residuals(Gaussiancouplamodel)\% Usual residual
> res2 <- residuals(OLSmodel) \% Usual residual
> res3 <- residuals(GLMmodel) \% Quantile residual
>
> sum(res1*res1) \% Quantile residual sum of square
[1] 295.1503
> sum(res2*res2) \% Usual residual sum of square
[1] 77746.4
> sum(res3*res3) \% Usual residual sum of square
[1] 77746.4
>
```

```
>
> AIC(Gaussiancouplamodel)
[1] 1906.128
> AIC(OLSmodel)
[1] 2491.572
> AIC(GLMmodel)
[1] 2491.572
>
>
> BIC(Gaussiancouplamodel)
[1] NA
> BIC(OLSmodel)
[1] 2510.007
> BIC(GLMmodel)
[1] 2510.007
>
>
```

Table 5.1 Parameter Estimations for Example 1

Description	OLS Model	GLM Model	Copula Model
Coefficients			
Intercept	28.7063	28.7063	1.34759
visit	-20.7923	-20.7923	0.11187
trt	0.8594	0.8594	-0.10685
visit:trt	-1.4988	-1.4988	-0.30237
AIC	2491.572	2491.572	1906.128

According to above table 5.1, Copula based regression model has low AIC value and other two models(OLS and GLM) has same AIC value. So we can say if we use copula model is more appropriate method in this regression analysis. In this case we used Gaussian as marginal distribution in copula regression model. Next example 02, we are going to consider general random generated data set, each data sets contains 50000 observations. Here we consider x_3 as a response variable which is generated from Poisson distribution and x_1, x_2, x_4, x_5, x_6 as regressors generated from different distributions. Also in this example, our model has no interaction term just only consider regressors. Similarly we can consider any data sets with any distributions (Most of the real world databases violated normality assumption)

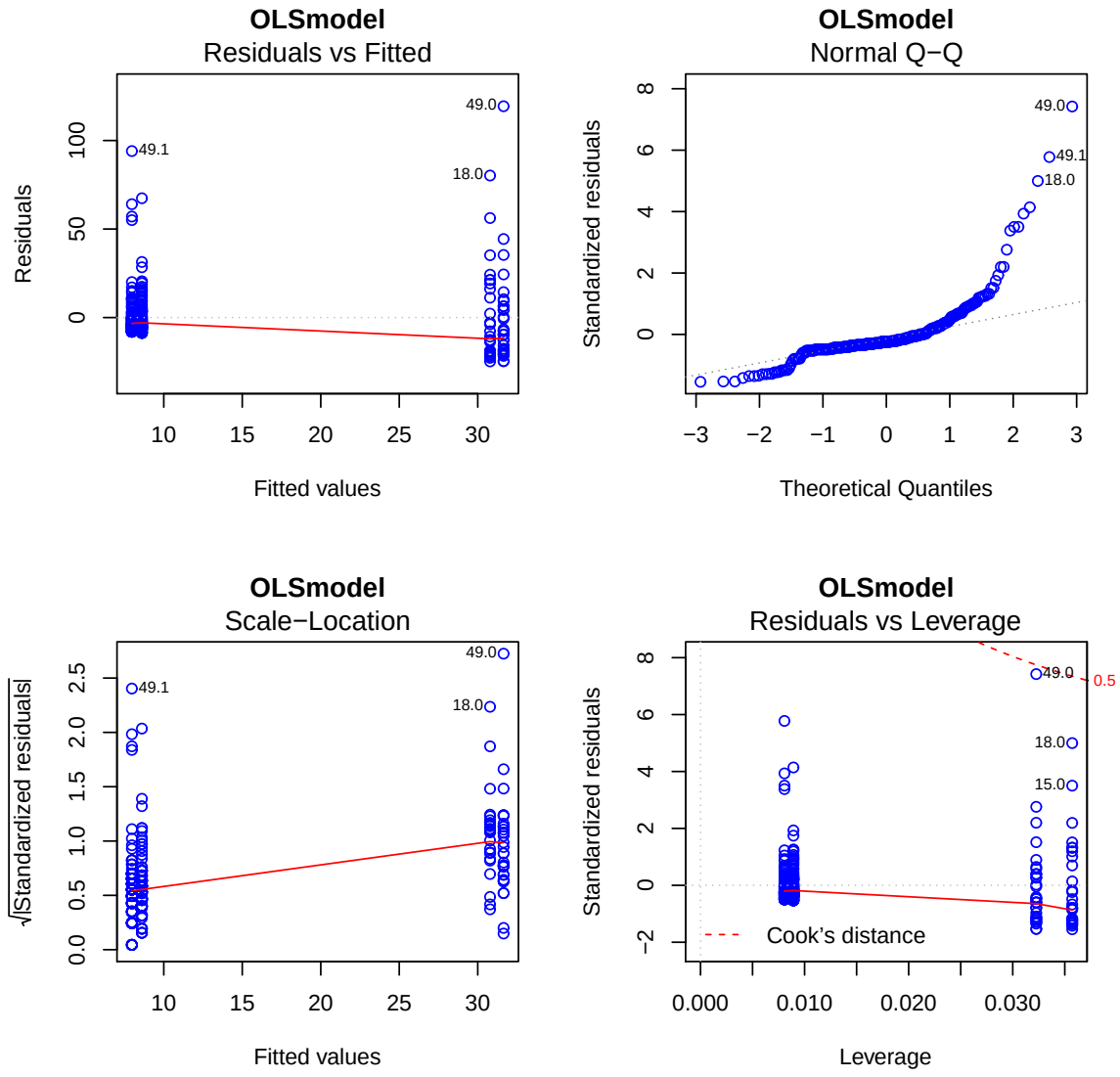

Fig. 5.1 Diagnostic plots - OLS Model-Example 1

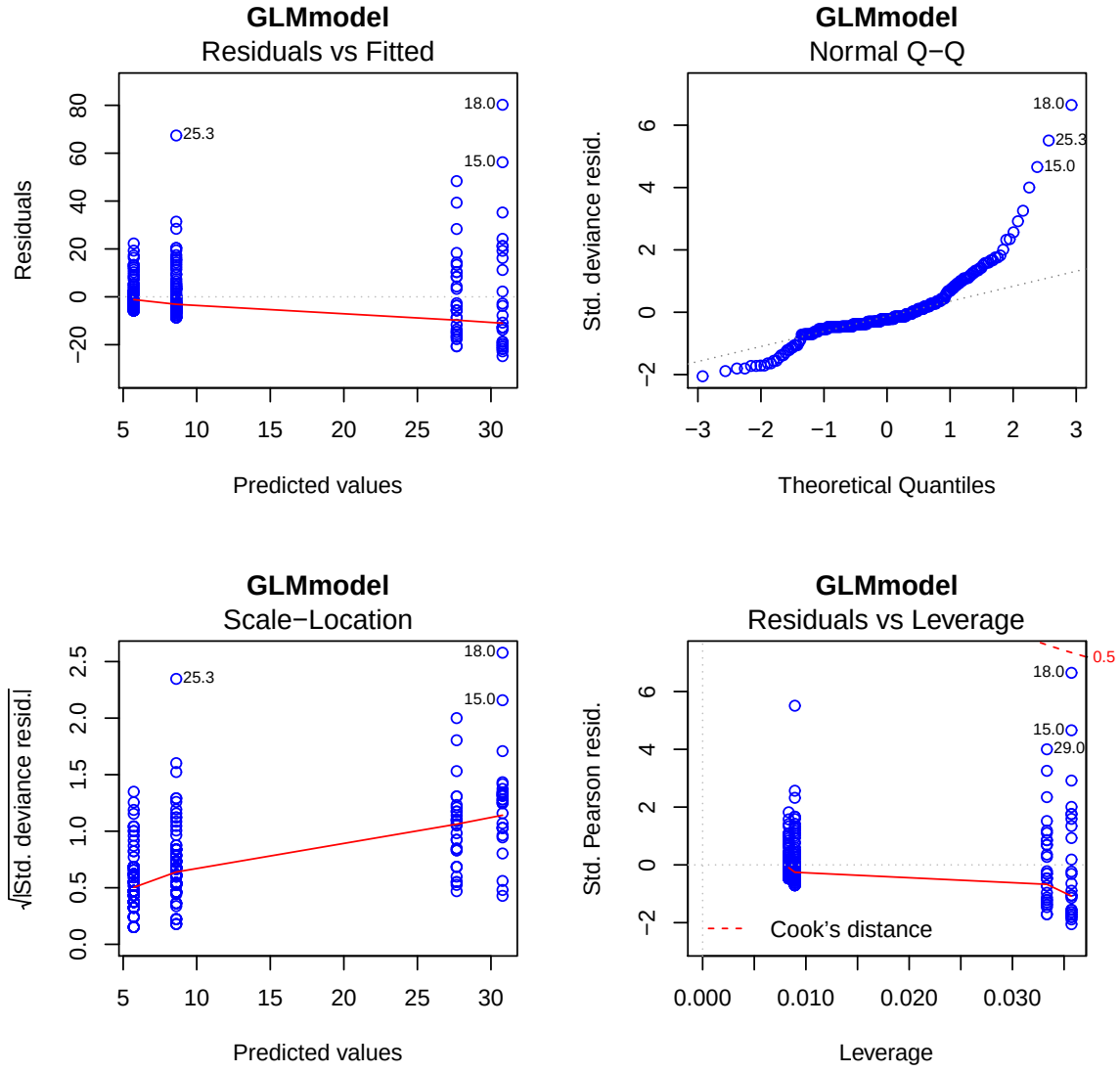

Fig. 5.2 Diagnostic plots - GLM Model-Example 1

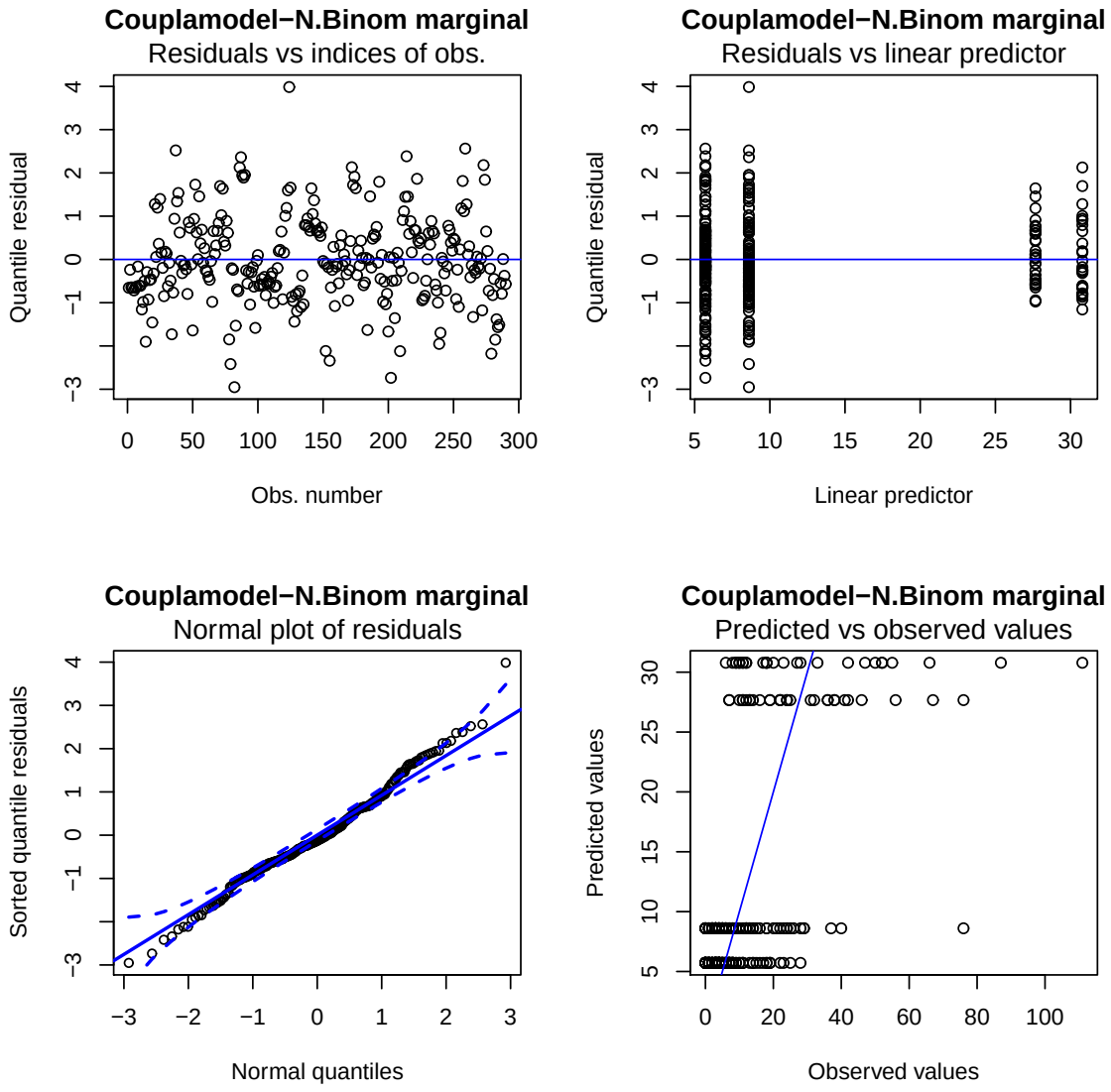

Fig. 5.3 Diagnostic plots - Copula Model Neg.Binomial as Marginal-Example 1

```
> #*********************************************************
> x1<-cbind(rpois(50000,6))
> x2<-cbind(rgamma(50000,2,1))
> x3<-cbind(rnorm(50000,5,8))
> x4<-cbind(rexp(50000,6))
> x5<-cbind(rexp(50000,3))
> x6<-cbind(rexp(50000,2))
>
> modelcoupula1 <- gcmr(x1 ~ x2+x3+x4+x5+x6, marginal = poisson.marg,
cormat = cluster.cormat(id, type= "ind"))
> modelsimple1 <-lm(x1 ~ x2+x3+x4+x5+x6)
> modelglm1 <-glm(x1 ~ x2+x3+x4+x5+x6,family = poisson)
>
> summary(modelcoupula1)

Call:
gcmr(formula = x1 ~ x2 + x3 + x4 + x5 + x6, marginal = poisson.marg,
    cormat = cluster.cormat(id, type = "ind"))


Coefficients marginal model:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.7908410  0.0046041 388.964   <2e-16 ***
x2          -0.0001669  0.0012889  -0.129    0.897
x3           0.0001294  0.0002278   0.568    0.570
x4          -0.0111978  0.0110000  -1.018    0.309
x5           0.0003424  0.0054662   0.063    0.950
x6           0.0018186  0.0036301   0.501    0.616

No coefficients in the Gaussian copula
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

log likelihood = 1.149e+05,  AIC = 229808
> summary(modelsimple1)
```

```
Call:
lm(formula = x1 ~ x2 + x3 + x4 + x5 + x6)

Residuals:
    Min      1Q  Median      3Q     Max
-6.0171 -1.9850  0.0018  1.9893 11.9928

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.9944805  0.0275575 217.526   <2e-16 ***
x2          -0.0009996  0.0077147  -0.130    0.897
x3           0.0007754  0.0013639   0.569    0.570
x4          -0.0669637  0.0657214  -1.019    0.308
x5           0.0020516  0.0327242   0.063    0.950
x6           0.0109068  0.0217507   0.501    0.616
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.446 on 49994 degrees of freedom
Multiple R-squared:  3.253e-05, Adjusted R-squared:  -6.748e-05
F-statistic: 0.3253 on 5 and 49994 DF,  p-value: 0.898

> summary(modelglm1)

Call:
glm(formula = x1 ~ x2 + x3 + x4 + x5 + x6, family = poisson)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4690  -0.8639   0.0007   0.7719   3.9397

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.7908410  0.0046041 388.966   <2e-16 ***
x2          -0.0001669  0.0012890  -0.129    0.897
x3           0.0001294  0.0002278   0.568    0.570
```

```
x4          -0.0111978  0.0109999  -1.018    0.309
x5           0.0003424  0.0054663   0.063    0.950
x6           0.0018186  0.0036301   0.501    0.616
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 51762  on 49999  degrees of freedom
Residual deviance: 51760  on 49994  degrees of freedom
AIC: 229808

Number of Fisher Scoring iterations: 4

> residualcop <- residuals(modelcoupula1) \% Quantile residual
> residualsim <- residuals(modelsimple1) \% Usual residual
> residualglm <- residuals(modelglm1) \% Usual residual
> sum(residualcop*residualcop) \% Quantile residual sum of square
[1] 49905.47
> sum(residualsim*residualsim) \% Usual residual sum of square
[1] 299000.6
> sum(residualglm*residualglm) \% Usual residual sum of square
[1] 51760.12
> AIC(modelcoupula1)
[1] 229808.1
> AIC(modelsimple1)
[1] 231329

> AIC(modelglm1)
[1] 229808.1
```

According to below table 5.2, GLM model and Copula based model has same AIC value and OLS model has high AIC value but in data analysis if we found any model has lower AIC then it is best fitting model. So we can conclude that GLM or Copula Models are better than OLS in this example. Next example 03, we are going to consider four Gaussian random variables each contains 50000 observations.

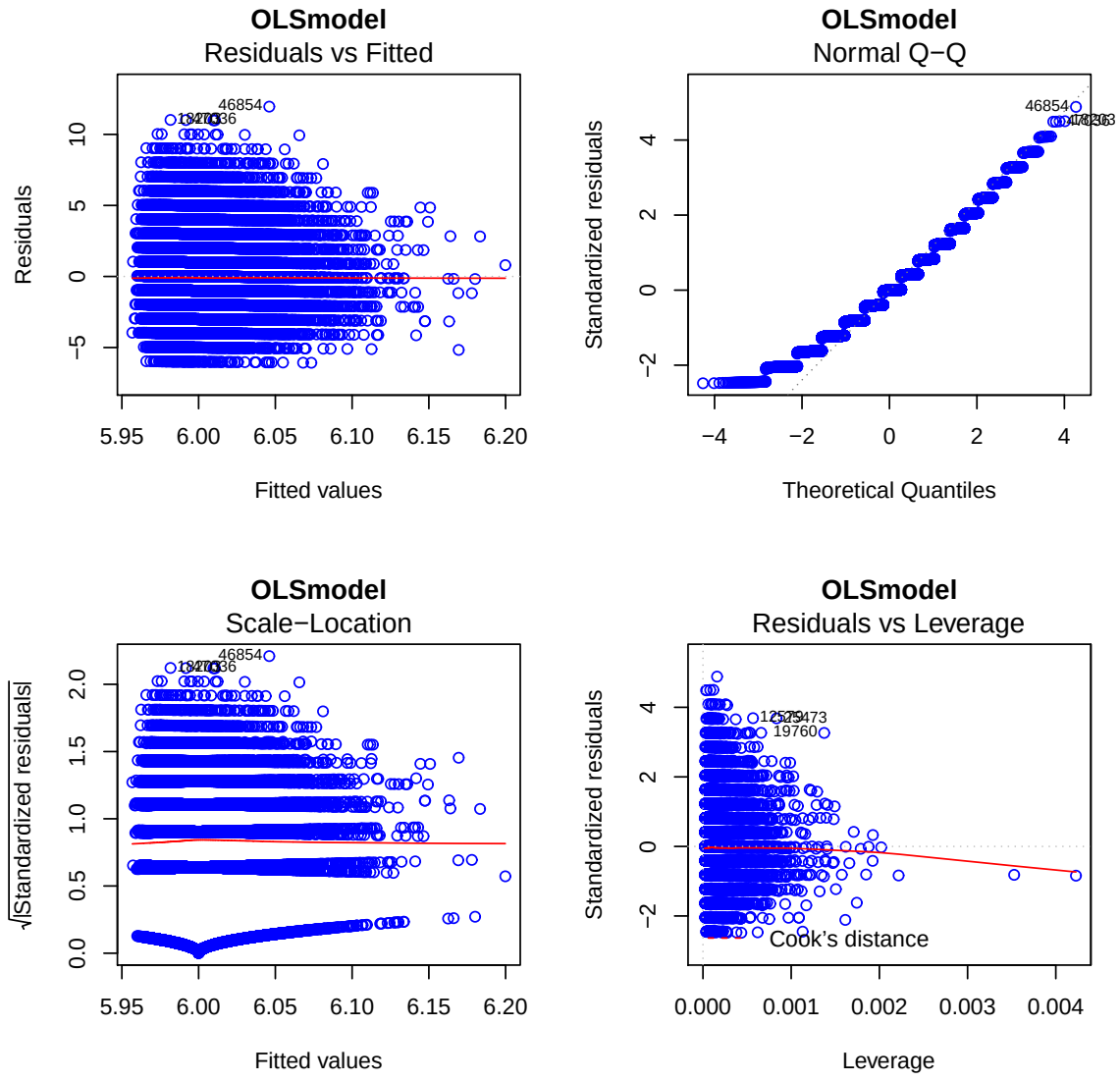

Fig. 5.4 Diagnostic plots - OLS Copula Model-Example 2

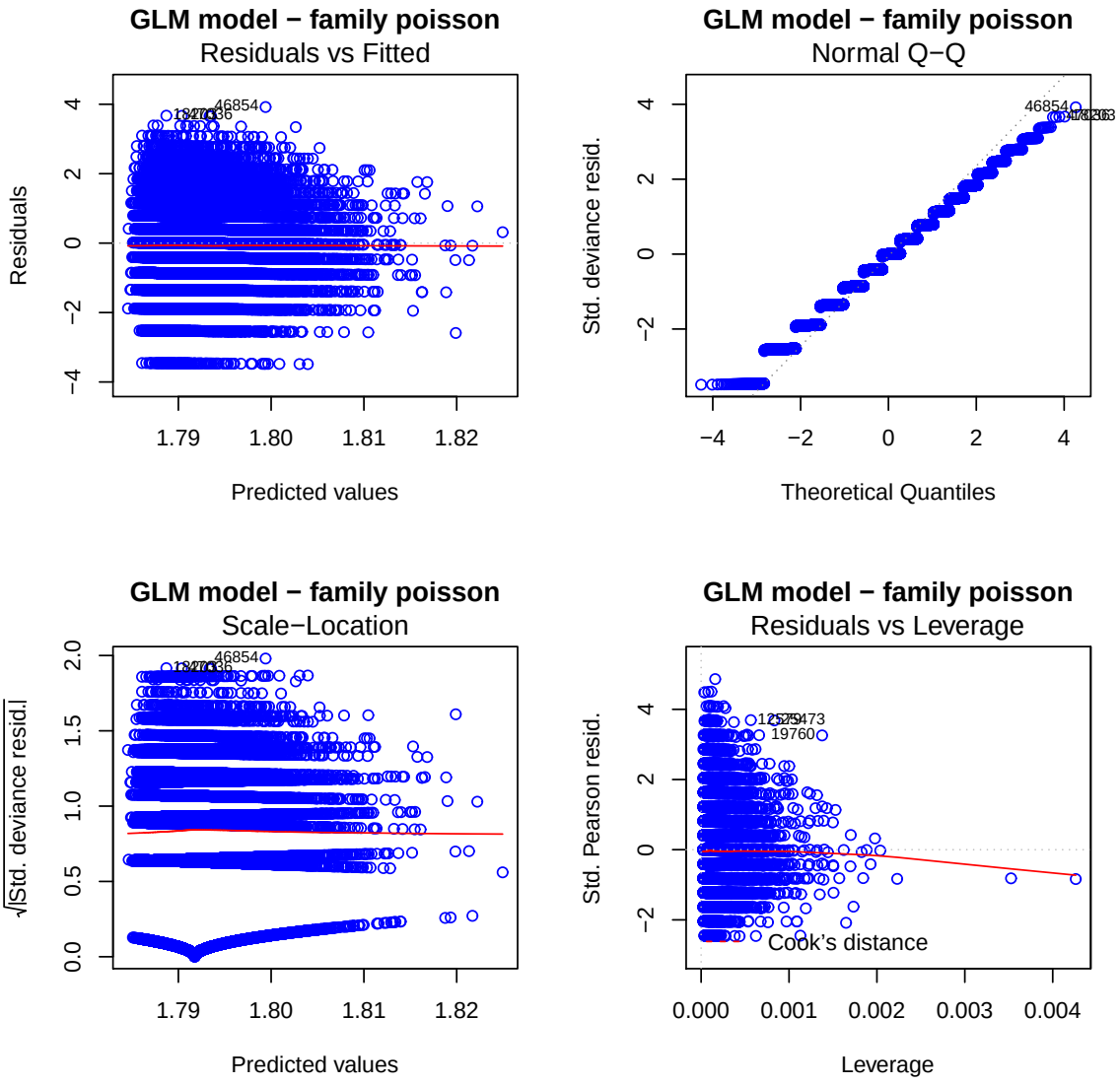

Fig. 5.5 Diagnostic plots - GLM Model-Example 2

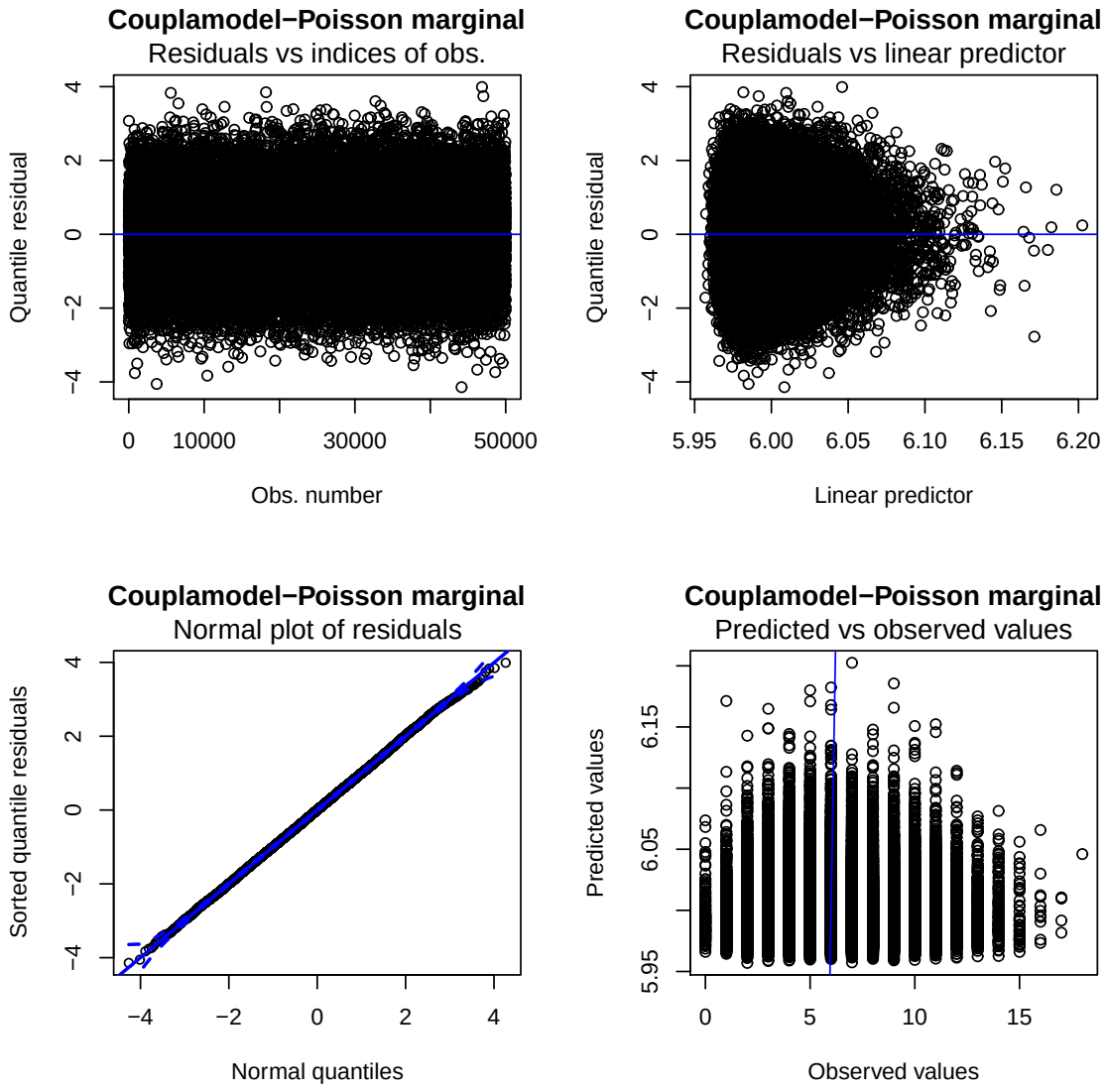

Fig. 5.6 Diagnostic plots - Copula Model Poisson as Marginal-Example 2

Table 5.2 Parameter Estimations for Example 2

Description	OLS Model	GLM Model	Copula Model
Coefficients			
Intercept	5.9944805	1.7908410	1.7908410
x_2	-0.0009996	-0.0001669	-0.0001669
x_3	0.0007754	0.0001294	0.0001294
x_4	-0.0669637	-0.0111978	-0.0111978
x_5	0.0020516	0.0003424	0.0003424
x_6	0.019068	0.0018186	0.0018186
AIC	231329	22908.1	22908.1

```
> #************************************************************
> x1<-cbind(rnorm(50000,6))
> x2<-cbind(rnorm(50000,5,6))
> x3<-cbind(rnorm(50000,6,8))
> x4<-cbind(rnorm(50000,2,9))
>
> modelcoup1 <- gcmr(x3 ~ x1+x2+x1:x2+x4, marginal = gaussian.marg,
cormat = cluster.cormat(id, type= "ind"))
> modelsim1 <-lm(x3 ~ x1+x2+x1:x2+x4)
> modelglm <-glm(x3 ~ x1+x2+x1:x2+x4,family = gaussian)
>
> summary(modelcoup1)

Call:
gcmr(formula = x3 ~ x1 + x2 + x1:x2 + x4, marginal = gaussian.marg,
cormat = cluster.cormat(id,
    type = "ind"))


Coefficients marginal model:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.9491964  0.2803725  21.219   <2e-16 ***
x1           0.0034283  0.0461106   0.074    0.941
x2           0.0091333  0.0355422   0.257    0.797
x4           0.0048751  0.0039652   1.229    0.219
x1:x2       -0.0009964  0.0058525  -0.170    0.865
```

```
sigma        7.9607947  0.0251745 316.225   <2e-16 ***

No coefficients in the Gaussian copula
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

log likelihood = 1.7467e+05,  AIC = 349358
> summary(modelsim1)

Call:
lm(formula = x3 ~ x1 + x2 + x1:x2 + x4)

Residuals:
    Min      1Q  Median      3Q     Max
-31.136  -5.415  -0.031   5.405  38.162

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.9491964  0.2803850  21.218   <2e-16 ***
x1           0.0034283  0.0461126   0.074    0.941
x2           0.0091333  0.0355439   0.257    0.797
x4           0.0048751  0.0039654   1.229    0.219
x1:x2       -0.0009964  0.0058527  -0.170    0.865
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.961 on 49995 degrees of freedom
Multiple R-squared:  3.672e-05, Adjusted R-squared:  -4.328e-05
F-statistic: 0.459 on 4 and 49995 DF,  p-value: 0.7659

> summary(modelglm)

Call:
glm(formula = x3 ~ x1 + x2 + x1:x2 + x4, family = gaussian)

Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-31.136   -5.415   -0.031    5.405   38.162

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.9491964  0.2803850  21.218   <2e-16 ***
x1           0.0034283  0.0461126   0.074    0.941
x2           0.0091333  0.0355439   0.257    0.797
x4           0.0048751  0.0039654   1.229    0.219
x1:x2       -0.0009964  0.0058527  -0.170    0.865
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 63.37996)

    Null deviance: 3168798  on 49999  degrees of freedom
Residual deviance: 3168681  on 49995  degrees of freedom
AIC: 349358

Number of Fisher Scoring iterations: 2

> res <- residuals(modelcoup1)
> res1 <- residuals(modelsim1)
> res2 <- residuals(modelglm)
>
> sum(res*res)\% Quantile residual sum of square
[1] 49999.5
> sum(res1*res1)\% Usual residual sum of square
[1] 3168681
> sum(res2*res2)\% Usual residual sum of square
[1] 3168681
>
> AIC(modelcoup1)
[1] 349358.2
> AIC(modelsim1)
[1] 349358.2
```

```
> AIC(modelglm)
[1] 349358.2
>
>
```

Table 5.3 Parameter Estimations for Example 3

Description	OLS Model	GLM Model	Copula Model
Coefficients			
Intercept	5.9491964	5.9491964	5.9491964
x_1	0.0034283	0.0034283	0.0034283
x_2	0.0091333	0.0091333	0.0091333
x_4	0.0048751	0.0048751	0.0048751
$x_1 : x_2$	-0.0009964	-0.0009964	-0.0009964
AIC	349358.2	349358.2	349358.2

According to above table 5.3, all model has same AIC values(349358.2) and satisfied normality assumption. In this data, we can choose any regression model. We can conclude that, if response variable follows any non exponential family distributions(violated GLM model assumption) or non Gaussian distribution (violated OLS model assumption), copula based regression is the best fitting model in regression analysis.

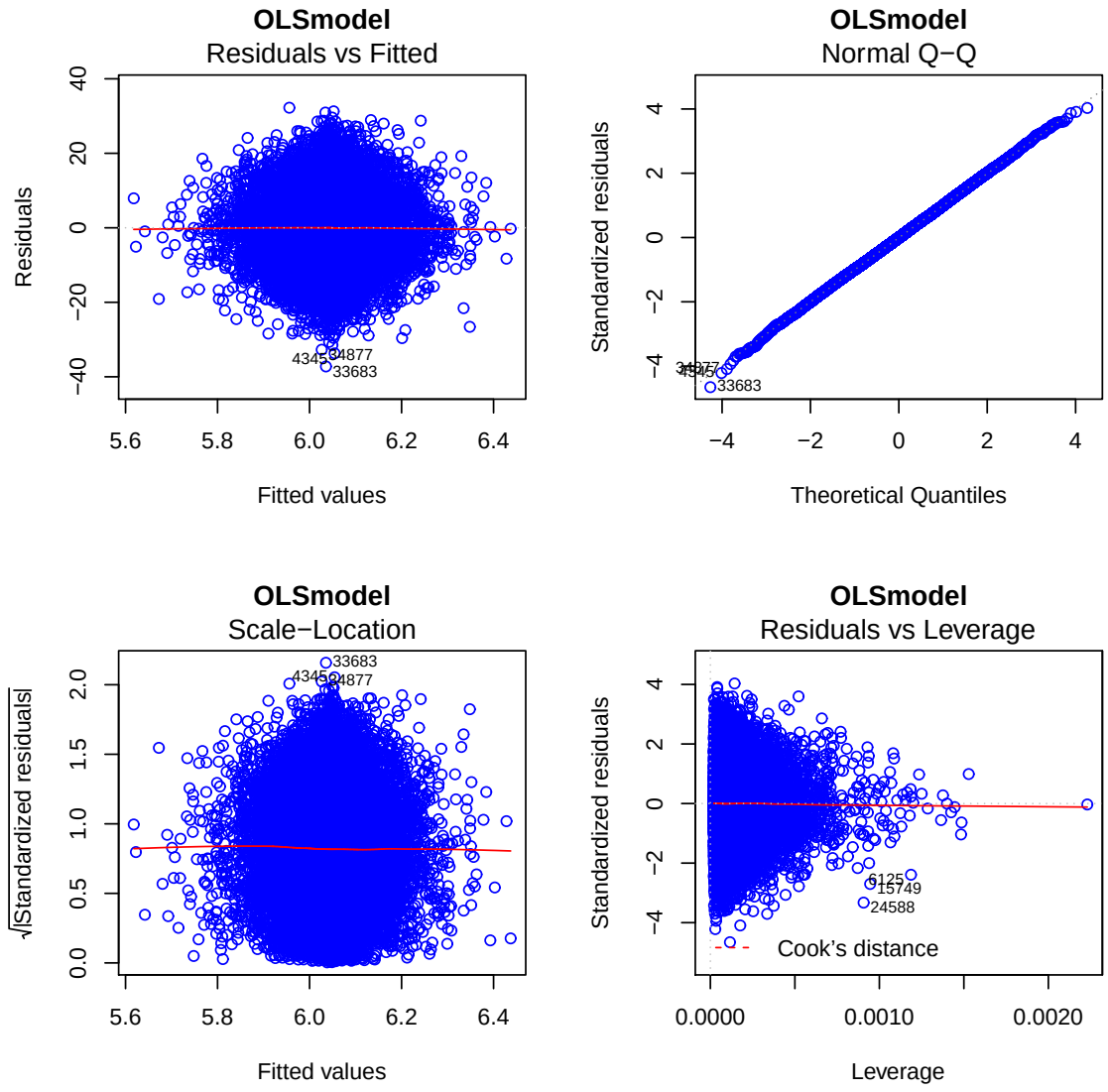

Fig. 5.7 Diagnostic plots - OLS Model-Example 3

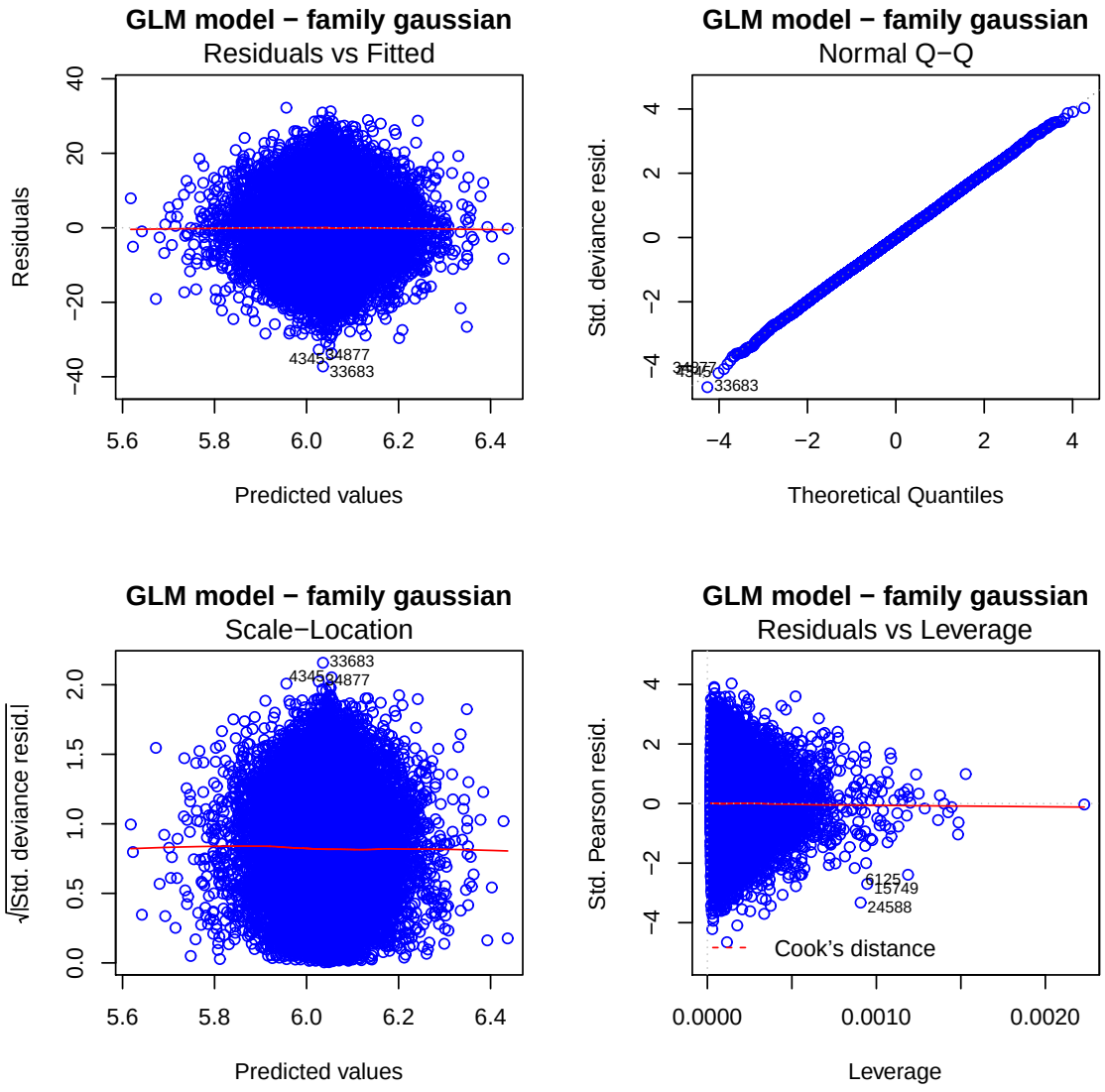

Fig. 5.8 Diagnostic plots - GLM Model-Example 3

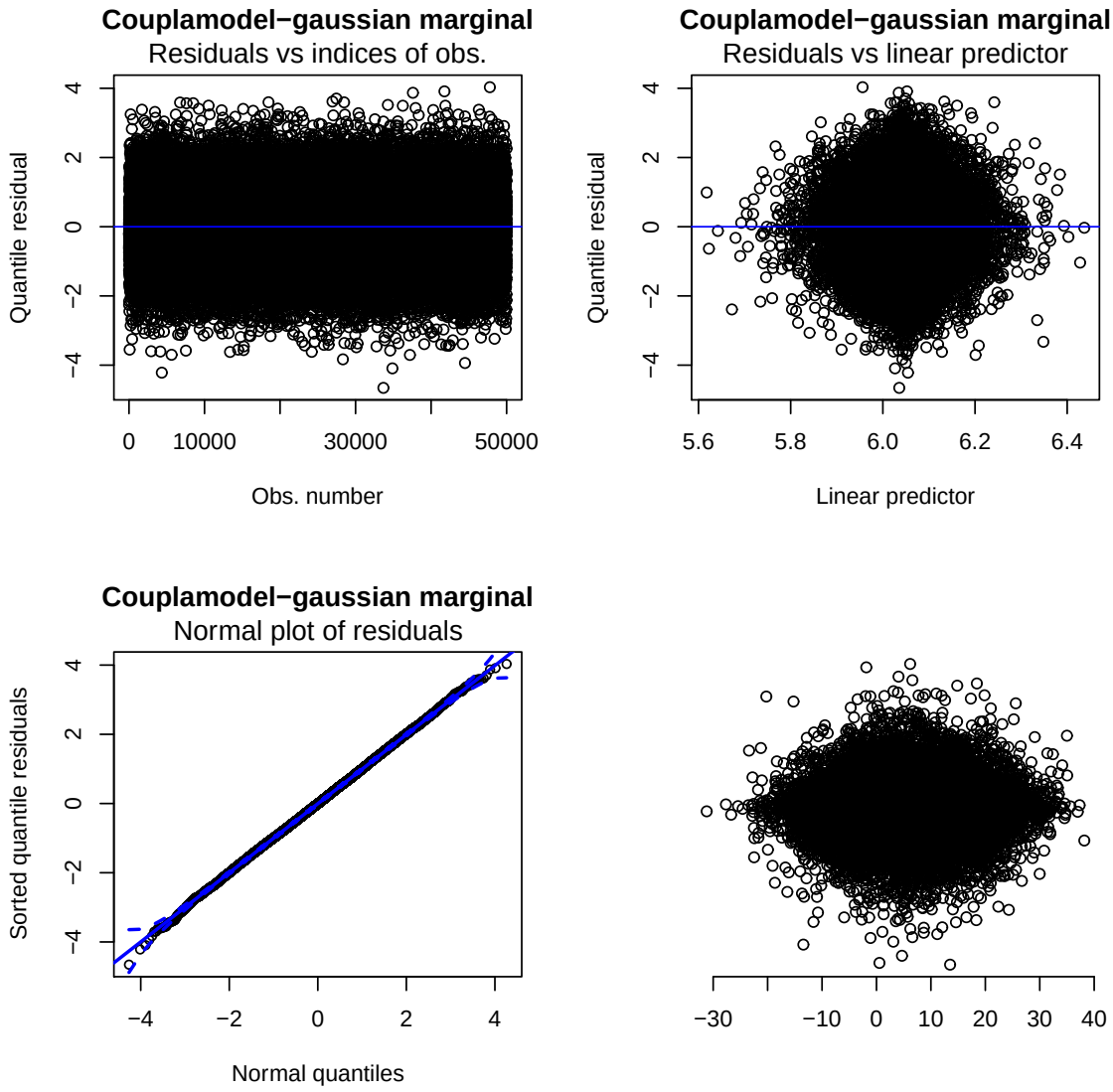

Fig. 5.9 Diagnostic plots - Copula Model Gaussian as Marginal-Example 3

Chapter 6

Conclusions and Future study

In this section we are going to propose our results of this study, Firstly we are going to propose our application result (Model Comparison- copula with OLS and GLM). Basic form of the regression analysis, ordinary least squares (OLS) is not suitable for actuarial applications because the relationships are often nonlinear and the probability distribution of the response variable may be non-Gaussian distribution. One of the method that has been successful in overcoming these challenges is the generalized linear model (GLM), which requires that the response variable have a distribution from the exponential family. In this research study, we study copula regression as an alternative method to OLS and GLM. The major advantage of a copula regression is that there are no restrictions on the probability distributions that can be used. We briefly discussed about copula regression by using several variety of marginal copula functions and copula regression is the most appropriate method in non Gaussian variable(violated normality assumption) regression model fitting. Also we validated our results by using real world example data and random generated (50000 observations) data. By using a simulated data set and other collected data, we found that the copula based regression approach performs well for OLS and GLM model assumption violated data and also copula based regression model has SSE (we don't have direct option in this R package, residual command provided Quantile residual), AIC and BIC values. We can conclude that, if response variable follows any non exponential family distributions(violated GLM model assumption) or non Gaussian distribution(violated OLS model assumption), copula based regression approach is the best model fitting in regression analysis.

Secondly, we are going to propose our theoretical result such as we proposed multiple regression line equation for Multivariate Non-Exchangeable Generalized Farlie-Gumbel-Morgenstern (FGM) copula function. In 2005, Sungur [47] developed the approach to the directional dependence by using bi-variate copula based linear regression, But in practically most of the regression analysis has more than one regressors so in this study we

discussed about multivariate copula based regression.We found,Multivariate (trivariate) Non-Exchangeable Generalized Farlie-Gumbel-Morgenstern (FGM) copula regression of U_0 on $U = u$ is given by

$$r^{C}_{U_0|U}(u) = 1 - \frac{\frac{1}{2}(1 + f_1'(u_1)f_2'(u_2)) + (f_1'(u_1) + f_2'(u_2))\int_0^1 f_0(u_0)du_0}{1 + f_1'(u_1)f_2'(u_2)}$$

There are many possibilities for extending the study domain as a part of future work. We can extend our result to copula based Ridge and Lasso regressions, there are no research works available in that topic. We couldn't find any literature related that topic. Ongoing research is focused on defining a new theoretical results such as copula based Ridge and Lasso regressions.

References

[1] Acar, E. F., Genest, C., and Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *Journal of Multivariate Analysis*, 110:74 – 90. Special Issue on Copula Modeling and Dependence.

[2] Breymann, W., Dias, A., and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3(1):1–14.

[3] Chen, X. and Fan, Y. (2005). Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection. *Canadian Journal of Statistics*, 33(3):389–414.

[4] Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154.

[5] Chen, X., Fan, Y., and Patton, A. J. (2004). Simple tests for models of dependence between multiple financial time series, with applications to u.s. equity returns and exchange rates (january 2004). *London Economics Financial Markets Group working paper number 483*.

[6] Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. The Wiley Finance Series. Wiley.

[7] Cook, R. and Johnson, M. (1981). A family of distributions for modeling non–elliptically symmetric multivariate data. *Journal of the Royal Statistical Society, Series B*, 43:210–218.

[8] Cuadras, C. (1992). Probability distributions with given multivariate marginals and given dependence structure. *Journal of Multivariate Analysis*, 42(1):51 – 66.

[9] Czado, C. (2010). *Pair-Copula Constructions of Multivariate Copulas*, volume 198. Springer New York.

[10] Deheuvels, P. (1979). La fonction de d ependance empirique et ses propri et es. un test non param etrique d'ind ependance. *Acad. Roy. Belg. Bull. Cl. Sci.*, 62(6)(03):274–292.

[11] Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (2013). *Analysis of Longitudinal Data*. Oxford Statistical Science Series. OUP Oxford.

[12] Dobric, J. and Schmid, F. (2005). Nonparametric estimation of the lower tail dependence l in bivariate copulas. *Journal of Applied Statistics*, 32(4):387–407.

[13] Dwass, M. (1980). The asymptotic theory of extreme order statistics (janos galambos). *SIAM Review*, 22(3):379–379.

[14] Embrechts, P., McNeil, A. J., and Straumann, D. (2002). *Correlation and Dependence in Risk Management: Properties and Pitfalls*, page 176–223. Cambridge University Press.

[15] Fermanian, J.-D. and Scaillet, O. (2005). Sensitivity analysis of VaR and Expected Shortfall for portfolios under netting agreements. *Journal of Banking & Finance*, 29(4):927–958.

[16] Fisher, N. I. (1997). Fonctions de répartition à n dimensions et leurs marges. *edited by S. Kotz, C. Read, and D. Banks-John Wiley and Sons, New York, 1997*, 1:159–164.

[17] Galambos, J. (1978). *The asymptotic theory of extreme order statistics / Janos Galambos*. Wiley New York.

[18] GENEST, C., GHOUDI, K., and RIVEST, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552.

[19] Genest, C. and MacKay, J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283.

[20] Genest, C., Nešlehová, J., and Ghorbal, N. B. (2011). Estimators based on kendall's tau in multivariate copula models. *Australian & New Zealand Journal of Statistics*, 53(2):157–177.

[21] Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88(423):1034–1043.

[22] Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

[23] Joe, H., Li, H., and Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101(1):252 – 270.

[24] Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models.

[25] KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1-2):81–93.

[26] Kimeldori, G. and Sampson, A. (1975). Uniform representations of bivariate distributions. *Communications in Statistics*, 4(7):617–627.

[27] Kolev, N., dos Anjos, U., and de M. Mendes, B. V. (2006). Copulas: A review and recent developments. *Stochastic Models*, 22(4):617–660.

[28] Masarotto, G. and Varin, C. (2012). Gaussian copula marginal regression. *Electron. J. Statist.*, 6:1517–1549.

[29] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall.

[30] Mikosch, T. (2006). Copulas: Tales and facts. *Extremes*, 9(1):3–20.

[31] Nelsen, R. B. (1998). Concordance and gini's measure of association. *Journal of Nonparametric Statistics*, 9(3):227–238.

[32] Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Science+Business Media, Inc.

[33] Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493.

[34] Patton, A. J. (2006a). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics*, 21(2):147–173.

[35] Patton, A. J. (2006b). Modelling asymmetric exchange rate dependence*. *International Economic Review*, 47(2):527–556.

[36] Rodrıx0301;guez-Lallena, J. A. and Úbeda Flores, M. (2004). A new class of bivariate copulas. *Statistics Probability Letters*, 66(3):315 – 325.

[37] Sancetta, A. and Satchell, S. (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20(03):535–562.

[38] Sasieni, P. (1990). Generalized additive models. t. j. hastie and r. j. tibshirani, chapman and hall, london, 1990. no. of pages: xv + 335. price: £25. isbn: 0-412-34390-8. *Statistics in Medicine*, 11(7):981–982.

[39] Schweizer, B. (1991). *Thirty Years of Copulas*, pages 13–50. Springer Netherlands, Dordrecht.

[40] Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *Ann. Statist.*, 9(4):879–885.

[41] Shih, J. and Louis, T. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399.

[42] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231.

[43] Sklar, A. (1996). *Random variables, distribution functions, and copulas—a personal look backward and forward*, volume Volume 28 of *Lecture Notes–Monograph Series*, pages 1–14. Institute of Mathematical Statistics, Hayward, CA.

[44] Song, P. (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer Series in Statistics. Springer New York.

[45] Song, P. X. (2000). Multivariate dispersion models generated from gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320.

[46] Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

[47] Sungur, E. (2005). Some observations on copula regression functions. *Communications in Statistics-theory and Methods -Taylor Francis, Inc.*, 34:1967–1978.

[48] W, H. (1940). Masstabinvariante korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für angewandte Mathematik der Université Berlin*, 5:179–233.

[49] Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman Hall/CRC, Boca Raton, Florida, first edition. ISBN 1-58488-474-6.

Appendix A

Programming codes

```
******************Bivariate random samples of size 200 from various
Frank copulas.**************
set.seed(5640)
delta = c(-100, -50, -10, -1, 0, 5, 20, 50, 500)
par(mfrow=c(3,3), cex.axis=1.2, cex.lab=1.2, cex.main=1.2)
for(i in 1:9){
U = rCopula(n=200,
copula=archmCopula(family="frank", param=delta[i]))
plot(U, xlab=expression(u[1]), ylab=expression(u[2]),
main=eval(substitute(expression(paste(delta," = ",j)),
list(j = as.character(theta[i])))))
}
```

```
*********************************Bivariate random samples of size 200
from various Clayton copulas.************
set.seed(5640)
delta = c(-0.95, -0.75, -0.25, -0.10, 0.10, 1, 5, 15, 200)
par(mfrow=c(3,3), cex.axis=1.2, cex.lab=1.2, cex.main=1.2)
for(i in 1:9){
U = rCopula(n=200,
copula=archmCopula(family="clayton", param=delta[i]))
plot(U, xlab=expression(u[1]), ylab=expression(u[2]),
main=eval(substitute(expression(paste(delta," = ",j)),
list(j = as.character(delta[i])))))
}
```

```
*****************************Bivariate random samples of size 200 from
various gumbel copulas.
set.seed(5640)
delta = c(1.0, 1.5, 2, 6, 10, 50)
par(mfrow=c(2,3), cex.axis=1.2, cex.lab=1.2, cex.main=1.2)
for(i in 1:6){
U = rCopula(n=200,
copula=archmCopula(family="gumbel", param=delta[i]))
plot(U, xlab=expression(u[1]), ylab=expression(u[2]),
main=eval(substitute(expression(paste(delta," = ",j)),
list(j = as.character(delta[i])))))
}
*****************************Bivariate random samples of size 200 from
various joe copulas.
set.seed(5640)
delta = c(1.0, 1.5, 2, 6, 10, 50)
par(mfrow=c(2,3), cex.axis=1.2, cex.lab=1.2, cex.main=1.2)
for(i in 1:6){
U = rCopula(n=200,
copula=archmCopula(family="joe", param=delta[i]))
plot(U, xlab=expression(u[1]), ylab=expression(u[2]),
main=eval(substitute(expression(paste(delta," = ",j)),
list(j = as.character(delta[i])))))
}

****************************Coefficients of tail dependence for bivariate
t-copulas as functions of ? for ? = 1, 4, 25, and 250.**

rho = seq(-1,1, by=0.01)
df = c(1, 5, 25, 100,250,500)
x1 = -sqrt((df[1]+1)*(1-rho)/(1+rho))
lambda1 = 2*pt(x1,df[1]+1)
x5 = -sqrt((df[2]+1)*(1-rho)/(1+rho))
lambda5 = 2*pt(x5,df[2]+1)
x25 = -sqrt((df[3]+1)*(1-rho)/(1+rho))
```

```
lambda25 = 2*pt(x25,df[3]+1)
x100 = -sqrt((df[4]+1)*(1-rho)/(1+rho))
lambda100 = 2*pt(x100,df[4]+1)
x250 = -sqrt((df[5]+1)*(1-rho)/(1+rho))
lambda250 = 2*pt(x250,df[4]+1)
x500 = -sqrt((df[6]+1)*(1-rho)/(1+rho))
lambda500 = 2*pt(x500,df[4]+1)

par(mfrow=c(1,1), lwd=2, cex.axis=1.2, cex.lab=1.2)
plot(rho, lambda1, type="l", lty=1, xlab=expression(rho),ylab=expression
(lambda[l]==lambda[u]))
lines(rho, lambda5, lty=2)
lines(rho, lambda25, lty=3)
lines(rho, lambda100, lty=4)
lines(rho, lambda250, lty=5)
lines(rho, lambda500, lty=6)
legend("topleft", c(expression(nu==1), expression(nu==5),
expression(nu==25), expression(nu==100), expression(nu==250),
expression(nu==500)), lty=1:6)

**********Chapter 03**********
### Estimation of copula parameters by using the MLE Method

## The "unknown" copula (a 2-dim. Clayton copula with parameter 5)
cc <- claytonCopula(5)
## The "unknown" distribution (N(0,1), Exp(6) margins)
mcc <- mvdc(cc, margins = c("norm", "exp"),
            paramMargins = list(list(mean = 0, sd = 1),
                                list(rate = 1)))
## Generate the "observed" sample
set.seed(712)
n <- 2000
X <- rMvdc(n, mvdc = mcc)
## The function fitMvdc() estimates all the parameters of the mvdc object
## mcc. Starting values need to be provided.
start <- c(mu0 = mean(X[,1]), sig0 = sd(X[,1]), lam0 = 1 / mean(X[,2]),
```

```
           th0 = 2)
mle <- fitMvdc(X, mvdc = mcc, start = start)

summary(mle)

### Estimation of copula parameters via the IFM Method

## Parametric pseudo-observations obtained from X by marginal MLE
U <- cbind(pnorm(X[,1], mean = mean(X[,1]),
                 sd = sqrt((n - 1) / n) * sd(X[,1])),
           pexp(X[,2], rate = 1 / mean(X[,2])))
ifme <- fitCopula(claytonCopula(), data = U, method = "ml")
summary(ifme)

### Estimation of copula parameters via the CML Method
CML <- fitCopula(claytonCopula(), data = U, method = "mpl")
summary(CML)


##Estimation of copula parameters via the method of moments based on
Kendall's tau

## The "unknown" copula (a 2-dim. clayton copula with parameter 5)
gc <- claytonCopula(5)
## The "unknown" distribution (N(0,1) margins)
mgc <- mvdc(gc, margins = c("norm", "norm"),
            paramMargins = list(list(mean = 0, sd = 1),
                                list(mean = 0, sd = 1)))
## Generate the "observed" sample
set.seed(49)
X <- rMvdc(1000, mvdc = mgc)
## The sample version of Kendall's tau
tau.n <- cor(X[,1], X[,2], method = "kendall")
## The corresponding copula parameter estimate
(itau <- iTau(gc, tau = tau.n))
```

```
stopifnot(all.equal(itau, 1 / (1 - tau.n))) # the same
## The same but with a standard error
summary(fitCopula(claytonCopula(), data = pobs(X), method = "itau"))


### Estimation of copula parameters via the method of moments based on
Spearman's rho

## The "unknown" copula (a 2-dim. normal copula with parameter 0.5)
nc <- claytonCopula(5)
## Generate the "observed" sample
set.seed(314)
X <- rCopula(1000, nc)
## The sample estimate of Spearman's rho
rho.n <- cor(X[,1], X[,2], method = "spearman")
## The corresponding copula parameter estimate
(irho <- iRho(nc, rho = rho.n))

stopifnot(all.equal(irho, 2 * sin(pi * rho.n / 6))) # the same
## The same but with a standard error
summary(fitCopula(claytonCopula(), data = pobs(X), method = "irho"))


### Spearman's rho and Kendall's tau for normal copulas
rho <- seq(-1, 1, by = 0.01) # correlation parameters of normal copulas
rho.s <- (6/pi) * asin(rho/2) # corresponding Spearman's rho
tau <- (2/pi) * asin(rho) # corresponding Kendall's tau
plot(rho, rho.s, type = "l", col = 3, lwd = 2,
     xlab = expression("Correlation parameter of normal copula"),
     ylab = expression(~rho[s]~"and"~tau))
abline(a = 0, b = 1, col = 2, lty = 2, lwd = 2)
lines(rho, tau, col = 4, lwd = 2)
legend("topleft", bty = "n", col = 2:4, lty = c(2, 1, 1), lwd = 2,
       legend = c("Diagonal", expression(rho[s]), expression(tau)))
```

Index

Akaike Information Criterion, 38
Archimedean Copulas, 20
Asymptotic Properties, 31

Bayesian Information Criterion, 38
Bernstein copula, 19
Bernstein polynomials, 19

Canonical Maximum Likelihood, 34
Clayton's Family, 20
Conditional copula, 14
Copula, 7
Copula log-likelihood, 33
Copula regression, 40

Dependence Measures, 25

Elliptical Copulas, 22

Fisher Information matrix, 32
Frailties, 16
Frank Copula, 18
Frank's Family, 22
Frechet's Family, 25
Frechet-Hoeffding bounds, 10

Gaussian copula regression, 48
Generalized Additive Models, 1
Gumbel's Family, 20
Gumbel-Hougaard, 17

IFM methods, 34
Independence, 30
Indicator function, 34
Inference Functions for Margins, 33
Invariance property, 11
Inversion of Marginals, 16

Joe's Family, 22

Kendall's tau, 26

Lower tail dependence, 30

Markov time series, 31
Maximum likelihood Estimation, 32
Measure of concordance, 27
Measure of dependence, 29
Multivariate copula regression, 45
Multivariate FGM Copula, 46

Nonparametric estimation, 35

Perfect Dependence, 30
Polynomial Approximations, 18
Probability Integral Transform, 7
Pseudo inverse, 18

Quantile Function, 7

Semi parametric approach, 34
Sklar's Theorem, 8
Sklar's Theorem in d -dimensions, 13
Spearman's rho, 26

Survival Copula, 13

Tail Dependence, 29
Time Dependent Copulas, 14

Uniform representations, 5
Upper tail dependence, 29