5-31-2021

Asymmetric multivariate archimedean copula models and semi-competing risks data analysis

Ziyan Guo
New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

ASYMMETRIC MULTIVARIATE ARCHIMEDEAN COPULA MODELS AND SEMI-COMPETING RISKS DATA ANALYSIS

by
Ziyan Guo

Many multivariate models have been proposed and developed to model high dimensional data when the dimension of a data set is greater than 2 ($d \geqslant 3$). The existing multivariate models often force the "exchangeable" structure for part or the whole model, are not very flexible which tends to be of limited use in practice. There is a demand for developing and studying multivariate models with any pre-specified bivariate margins.

Suppose there exists such a class of flexible models with any pre-specified bivariate margins. Given a multivariate data, what is the distribution function and how to easily estimate the parameters from this multivariate model are often important issues to solve.

Dependent censoring has become an increasingly important issue in medical data analysis. Quite often failure times are subject to dependent censoring and how to model and quantify such dependence is also of great interest.

The research described in Chapter 2 of this dissertation has been motivated by the above challenging questions. Copula models are used to address these important problems.

The first result is to generalize the model construction approach proposed by Chakak (1993) to $d-$dimensional models with arbitrarily pre-specified bivariate margins. The second result is to give the distribution functions for models constructed using the construction approach proposed in the first result. The third result is to propose parameters estimation approach and new model selection approach for models

constructed using the construction approach proposed in the first result. Simulation studies show that the parameter estimate works very well.

The research described in Chapter 3 of this dissertation has been motivated by the dependent censoring. The copula-graphic estimator (Zheng and Klein 1996) is first derived in this dissertation for marginal survival functions using Archimedean copula models based on semi-competing risks data. And its uniform consistency and asymptotic properties are proved.

A parameter estimation strategy is given to analyze the semi-competing risks data using Archimedean copula models. The method described in this dissertation is important and flexible in that it allows us to determine dependence levels between competing risks when two dependent competing risks are subject to independent censoring.

Based on the parameter estimation strategy proposed above, a new model selection procedure is given. An easy way to accommodate possible covariates in data analysis using the strategies is discussed.

Simulation studies show that the parameter estimate outperforms the estimator proposed by Lakhal, Rivest and Abdous (2008) for the Hougaard model and the model selection procedure works quite well. A leukemia data set is fitted by using the proposed model selection procedure and this dissertation end with some discussion.

ASYMMETRIC MULTIVARIATE ARCHIMEDEAN COPULA MODELS AND SEMI-COMPETING RISKS DATA ANALYSIS

by
Ziyan Guo

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences
Department of Mathematics and Computer Science, Rutgers-Newark

May 2021

APPROVAL PAGE

ASYMMETRIC MULTIVARIATE ARCHIMEDEAN COPULA MODELS AND SEMI-COMPETING RISKS DATA ANALYSIS

Ziyan Guo

Dr. Antai Wang, Dissertation Advisor — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Wenge Guo, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Ji Meng Loh, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Zhi Wei, Committee Member — Date
Professor of Computer and Information Science, NJIT

Dr. Yixin Fang, Committee Member — Date
Director, GMA-Statistics, AbbVie, North Chicago, Illinois

BIOGRAPHICAL SKETCH

Author: Ziyan Guo

Degree: Doctor of Philosophy

Date: May 2021

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology and Rutgers, The State University of New Jersey - Newark, Newark, NJ, 2021
- Master of Science in Applied Mathematics,
  New Jersey Institute of Technology, Newark, NJ, 2019
- Master of Science in Computer Science,
  Jiangnan University, Wuxi, China, 2008
- Bachelor of Science in Mathematical Sciences,
  Jiangnan University, Wuxi, China, 2006

Major: Mathematical Sciences

Presentations and Publications:

Antai Wang, Yilong Zhang, Ziyan Guo and Jihua Wu, "The Analysis of Semi-Competing Risks Data Using Archimedean Copula Models," *Statistica Neerlandica*, in revision, 2021.

Jianfu Guo, Guangyang Tang and Ziyan Guo et al., "Advanced Mathematics II (in Chinese)," *Wuhan, P. R. China: Wuhan University Press*, ISBN:978-7-3070-6170-5, 2008.

Ziyan Guo and Zuhua Liao, "Anti-Fuzzy Rough Subsemigroup and Anti-Fuzzy Rough Subgroup," The 10th China Conference on Machine Learning (CCML06); also published in *Computer Science*, ISSN1002-137X, Volume 33, Issue 10 Supplement (9-771002-137063-99), Pages 340-341, 2006.

Zuhua Liao, Ziyan Guo and Jianfu Guo, "$(\in, \in \vee q_{(\lambda,\mu)})-$Fuzzy Vector Spaces (in Chinese)," *Fuzzy Systems and Mathematics*, ISSN1001-7402, Volume 20 Supplement, Pages 82-84, 2006.

"Ipsa scientia potestas est.
Knowledge itself is power."

—— Francis Bacon

ACKNOWLEDGMENT

I would like to express my deep and sincere gratitude to my dissertation advisor Dr. Antai Wang, for giving me the opportunity to do research with him and providing invaluable guidance and tremendous assistance throughout this research. His dynamism, vision, sincerity and motivation have deeply inspired me. From the beginning to the end, he has been a steadfast source of information, ideas and energy. He has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under his guidance. I am deeply grateful for his kind and informative guidance, patience, friendship, empathy and encouragement. I will be forever grateful for his trust and great support during this dissertation complete procedure.

My sincere thanks gose to the distinguished members of this dissertation committee: Dr. Wenge Guo, Dr. Ji Meng Loh, Dr. Zhi Wei and Dr. Yixin Fang, for their active participation, constant encouragement and indightful comments throughout this research work. Their ideas, skills and talents have always supported me.

Without the financial supports provided by the Department of Mathematical Sciences, this dissertation would not exist. You have given me the opportunity to be a researcher and PhD. Many of my present and former professors in the Department of Mathematical Sciences are deserving recognition for their direct or indirect help during my graduate study life.

Finally, I would like to acknowledge with gratitude, the support, encouragement and love of my family. All have been encouraging.

I hope you will enjoy reading my dissertation.

TABLE OF CONTENTS

Chapter **Page**

1 INTRODUCTION 1

1.1 Objective 1

1.2 Background Information 2

1.2.1 Definitions and properties of high dimensional copulas 3

1.2.2 Motivation 14

1.2.3 Fully nested Archimedean copulas 17

1.2.4 Partially nested Archimedean copulas 19

1.2.5 General nested Archimedean copulas 20

1.2.6 Pair copulas 22

1.2.7 Discussion 24

2 ASYMMETRIC MULTIVARIATE ARCHIMEDEAN COPULA MODELS 28

2.1 Introduction 28

2.2 Method of Constructing Asymmetric Multivariate Archimedean Copula Models 28

2.2.1 Three-dimensional structures 36

2.2.2 Four-dimensional structures 39

2.2.3 $d-$dimensional structures 46

2.2.4 Flexibility for the proposed structures 50

2.3 Examples Based on Proposed Structures 53

2.3.1 Clayton copulas 53

2.3.2 Gumbel - Hougaard copula 54

2.3.3 Frank copula 56

2.3.4 Three-dimensional structures based on Clayton copulas 57

2.3.5 Four-dimensional structures based on Clayton copulas 64

2.3.6 Three-dimensional structures based on different copulas 70

TABLE OF CONTENTS
(Continued)

Chapter **Page**

2.4 Survival Functions for Proposed Structures 76

2.5 Parameter Estimation for Proposed Structures 84

2.6 Model Selection for Proposed Structures 88

2.7 Numerical Studies 90

2.8 Discussion 92

3 ANALYSIS OF SEMI-COMPETING RISKS DATA USING ARCHIMEDEAN COPULA MODELS 94

3.1 Introduction 94

3.2 Copula-graphic Estimator for Marginal Survival Function of Y Based on Semi-competing Risks Data 97

3.3 A New Parameter Estimation Strategy Based on Semi-competing Risks Data 110

3.4 Model Selection 118

3.5 Accomodation of Covariates 119

3.6 Simulation Studies 119

3.7 An Illustrative Example 125

3.8 Discussion 126

REFERENCES 129

LIST OF TABLES

Table **Page**

2.1 Simulation Results using Pairwise Estimation vs. Maximum Likelihood Estimation for the Clayton Copula 92

3.1 Performance of $\hat{\tau}$ and $\tilde{\tau}$ for the Hougaard Model Based on Kendall's $\tau = 0.2, 0.4, 0.6, 0.8$ in Different Censoring Proportions (Scenario $1-8$). Here Kendall's τ is the True Value of the Parameter. $\hat{\tau}$ is the Proposed Parameter Estimator and $\tilde{\tau}$ is the Estimator Proposed by Lakhal, Rivest and Abdous (2008) [21]. 121

3.2 Performance of $\hat{\tau}$ and $\tilde{\tau}$ for the Hougaard Model Based on Kendall's $\tau = 0.2, 0.4, 0.6, 0.8$ in Different Censoring Proportions (Scenario $9-16$). Here Kendall's τ is the True Value of the Parameter. $\hat{\tau}$ is the Proposed Parameter Estimator and $\tilde{\tau}$ is the Estimator Proposed by Lakhal, Rivest and Abdous (2008) [21]. 122

3.3 Selection Percentages for Data from the Clayton Model 123

3.4 Selection Percentages for Data from the Hougaard Model 124

3.5 Selection Percentages for Data from the Frank Model 124

3.6 $Q_n(\theta)$ Value for the Leukemia Data Set (Included in R Package KMsurv). 125

3.7 Parameters for the Leukemia Data Set (Included in R Package KMsurv). Here the Proposed Parameter Estimator is Compared to the Estimator Proposed by Fine, Jiang and Chappel (2001) [5]. 126

3.8 Parameters for the Leukemia Data Set (Included in R Package KMsurv) by Variable Age (Using the Median Value as the Cut-off Point). 126

LIST OF FIGURES

Figure		Page
1.1	Fully nested Archimedean copulas.	18
1.2	Partially nested Archimedean copulas.	19
1.3	General nested Archimedean copulas.	23
1.4	Pair copulas with Canonical vine.	24
1.5	Pair copulas with $D-$vine.	25
2.1	Three-dimensional structure for proposed method.	37
2.2	Four-dimensional structure for proposed method.	42
2.3	Flexibility of proposed method (four-dimensional structure: one of the possibilities).	51
2.4	Flexibility of proposed method (four-dimensional structure: another possibility).	52
2.5	Three-dimensional structure based on Clayton copulas for proposed method.	58
2.6	Four-dimensional structure based on Clayton copulas for proposed method.	64
2.7	Flexibility of proposed method (three-dimensional structure: Clayton + Hougaard).	71
2.8	Flexibility of proposed method (three-dimensional structure: Clayton + Hougaard + Frank).	73

CHAPTER 1

INTRODUCTION

1.1 Objective

In multivariate analysis, it is often a very difficult problem to model non-normal multivariate data. These random variables might or might not be correlated. Many multivariate models have been proposed and developed to model high dimensional data when the dimension of a data set is greater than 2 ($d \geqslant 3$). "Copulas" are multivariate probability distributions which be used to describe the dependence between random variables. By Sklar's theorem [34], every multivariate cumulative distribution functions can be expressed in terms of its marginals and a copula. Many different ways to construct asymmetric Archimedean copulas were introduced. However, most of them are not very flexible. The existing multivariate models often forces the "exchangeable" structure for part or the whole model, which tends to be of limited use in practice. There is a demand for developing and studying multivariate models with any pre-specified bivariate margins.

Suppose there exists such a class of flexible models with any pre-specified bivariate margins. Given a multivariate data, what is the distribution function and how to easily estimate the parameters from this multivariate model are often important issues to solve.

Dependent censoring has become an increasingly important issue in medical data analysis. Quite often failure times are subject to dependent censoring and how to model and quantify such dependence is also of great interest.

The research described in this dissertation has been motivated by the above challenging questions. Copula models is used to address these important problems.

1.2 Background Information

The name of copula comes from Latin "cōpulāre" for "link" or "tie". Copulas are actually multivariate probability distributions and the marginal probability distribution of each variable is uniformly distributed on [0,1].

In the statistical literature, the idea of copulas can be traced back to the 20th century. They have been widely used in financial analysts, financial stability, credit risk management and insurance to model the dependence between variables and describe the joint probability distribution. Subsequently, over the last few decades, copulas have also been widely used in Biostatistics, Hydrology, Meteorology, Environmental Science and so on.

Copulas are multivariate dependence functions which used to describe the dependence between random variables. They are used for separating the marginal distributions from a given multivariate distribution (the dependency structure). Why are they so popular? The main reason is they help to understand the correlation and expose the various paralogisms associated with the correlation. Copula-based multivariate models allow separating the pre-specified marginal distributions from the dependence structure.

There are also some problems associated with the use of copulas due to the inappropriate application. Moreover, they are sometimes used in a "black-box" fashion with the development and popularization of machine learning and artificial intelligence. We aim to give a deeper insight for how to construct general models allowing arbitrary selection of pairwise correlation which is desired in our practical applications.

This dissertation is developed based upon the construction approach proposed by Chakak (1993) [2] which allows arbitrary selection of pairwise correlation. It is a big step forward since such structure have been thinking about for decades. In this

research, the structures were extended to more than two dimensions, and proved the feasibility of the structure to $d-$dimensions by mathematical induction method.

1.2.1 Definitions and properties of high dimensional copulas

Many multivariate models have been proposed and developed to model high dimensional data when the dimension of a data set is greater than 2 ($d \geqslant 3$). This section is to give the definitions of high dimensional copulas and show some properties of them [22].

Definition 1.2.1. ***Copula:*** *A $d-$dimensional copula $C : [0,1]^d \to [0,1]$ is a cumulative distribution function with the marginal probability distribution of each variable is uniform on the interval $[0,1]$.*

Property 1.2.2. ***Copula:*** *Let $\boldsymbol{u} = [u_1, u_2, \cdots, u_d], u_i \in [0,1]$ with $\forall i \in \{1, 2, \cdots, d\}$. Any $d-$dimensional copula $C(\boldsymbol{u}) = C(u_1, u_2, \cdots, u_d)$ has following properties*

1. *$C(u_1, u_2, \cdots, u_d)$ is a non-decreasing function in each variable u_i;*
2. *If at least one $u_i = 0$ in $\boldsymbol{u}$, then*

$$C(u_1, u_2, \cdots, u_d) = C(u_1, u_2, \cdots, u_{i-1}, 0, u_{i+1}, \cdots, u_d) = 0; \tag{1.1}$$

3. *The i^{th} marginal distribution is obtained by setting all the $u_j = 1, j \neq i$ in $\boldsymbol{u}$. Since all the marginal probability distribution are uniformly distributed, then*

$$C(u_1, u_2, \cdots, u_d) = C(1, 1, \cdots, 1, u_i, 1, \cdots, 1) = u_i; \tag{1.2}$$

4. *$C(u_1, u_2, \cdots, u_d)$ has a non-negative value for any $d-$dimensional interval, that is $\forall(a_1, a_2, \cdots, a_d) \in [0,1]^d$ and $\forall(b_1, b_2, \cdots, b_d) \in [0,1]^d$ with $a_i \leqslant b_i$ and $t_i \in \{1, 2\}$, $P(X_1 \in [a_1, b_1], X_2 \in [a_2, b_2], \cdots, X_d \in [a_d, b_d]) \geqslant 0$. When $t_i = 1, x_{it_i} = x_{i1} = a_i$. When $t_i = 2, x_{it_i} = x_{i2} = b_i$. This implies the rectangle inequality*

$$\sum_{t_1=1}^{2}\sum_{t_2=1}^{2}\cdots\sum_{t_d=1}^{2}(-1)^{t_1+t_2+\cdots+t_d}C(x_{1t_1}, x_{2t_2}, \cdots, x_{dt_d}) \geqslant 0. \tag{1.3}$$

Example 1.2.3. ***An example with*** $\mathbf{d = 2}$***:*** *When $(a_1, a_2) \in [0,1]^2$ and $(b_1, b_2) \in [0,1]^2$ with $a_1 \leqslant b_1$, $a_2 \leqslant b_2$, that is*

$$\begin{aligned}
&\sum_{t_1=1}^{2}\sum_{t_2=1}^{2}(-1)^{t_1+t_2}C(x_{1t_1}, x_{2t_2}) \\
=&\sum_{t_1=1}^{2}\left[(-1)^{t_1+1}C(x_{1t_1}, x_{21}) + (-1)^{t_1+2}C(x_{1t_1}, x_{22})\right] \\
=&(-1)^2C(x_{11}, x_{21}) + (-1)^3C(x_{11}, x_{22}) + (-1)^3C(x_{12}, x_{21}) + (-1)^4C(x_{12}, x_{22}) \\
=&C(x_{11}, x_{21}) - C(x_{11}, x_{22}) - C(x_{12}, x_{21}) + C(x_{12}, x_{22}) \\
=&C(a_1, a_2) - C(a_1, b_2) - C(b_1, a_2) + C(b_1, b_2) \geqslant 0
\end{aligned} \tag{1.4}$$

Example 1.2.4. ***An example with*** $\mathbf{d = 3}$***:*** *When* $(a_1, a_2, a_3) \in [0,1]^3$*, and* $(b_1, b_2, b_3) \in [0,1]^3$ *with* $a_1 \leqslant b_1$*,* $a_2 \leqslant b_2$ *and* $a_3 \leqslant b_3$*, that is*

$$
\begin{aligned}
&\sum_{t_1=1}^{2}\sum_{t_2=1}^{2}\sum_{t_3=1}^{2}(-1)^{t_1+t_2+t_3}C(x_{1t_1}, x_{2t_2}, x_{3t_3}) \\
=&\sum_{t_1=1}^{2}\sum_{t_2=1}^{2}\left[(-1)^{t_1+t_2+1}C(x_{1t_1}, x_{2t_2}, x_{31}) + (-1)^{t_1+t_2+2}C(x_{1t_1}, x_{2t_2}, x_{32})\right] \\
=&\sum_{t_1=1}^{2}[(-1)^{t_1+1+1}C(x_{1t_1}, x_{21}, x_{31}) + (-1)^{t_1+1+2}C(x_{1t_1}, x_{21}, x_{32}) + \\
&(-1)^{t_1+2+1}C(x_{1t_1}, x_{22}, x_{31}) + (-1)^{t_1+2+2}C(x_{1t_1}, x_{22}, x_{32})] \\
=&(-1)^{1+1+1}C(x_{11}, x_{21}, x_{31}) + (-1)^{1+1+2}C(x_{11}, u_{21}, x_{32}) + \\
&(-1)^{1+2+1}C(x_{11}, x_{22}, x_{31}) + (-1)^{1+2+2}C(x_{11}, x_{22}, x_{32}) + \\
&(-1)^{2+1+1}C(x_{12}, x_{21}, x_{31}) + (-1)^{2+1+2}C(x_{12}, x_{21}, x_{32}) + \\
&(-1)^{2+2+1}C(x_{12}, x_{22}, x_{31}) + (-1)^{2+2+2}C(x_{12}, x_{22}, x_{32}) \\
=&-C(x_{11}, x_{21}, x_{31}) + C(x_{11}, x_{21}, x_{32}) \\
&+C(x_{11}, x_{22}, x_{31}) - C(x_{11}, x_{22}, x_{32}) \\
&+C(x_{12}, x_{21}, x_{31}) - C(x_{12}, x_{21}, x_{32}) \\
&-C(x_{12}, x_{22}, x_{31}) + C(x_{12}, x_{22}, x_{32}) \\
=&-C(a_1, a_2, a_3) + C(a_1, a_2, b_3) \\
&+C(a_1, b_2, a_3) - C(a_1, b_2, b_3) \\
&+C(b_1, a_2, a_3) - C(b_1, a_2, b_3) \\
&-C(b_1, b_2, a_3) + C(b_1, b_2, b_3) \geqslant 0
\end{aligned}
\tag{1.5}
$$

5. *$C(u_1, u_2, \cdots, u_d)$ meets the boundary conditions* $0 \leqslant C(u_1, u_2, \cdots, u_d) \leqslant 1$*;*

6. *The following Fréchet–Hoeffding copula bounds hold*

$$
\max\left\{0, 1 - d + \sum_{i=1}^{d} u_i\right\} \leqslant C(u_1, u_2, \cdots, u_d) \leqslant \min\left\{u_1, u_2, \cdots, u_d\right\}; \tag{1.6}
$$

7. *$F_1, F_2, \cdots, F_d$ are given as marginal distributions for random variables $X_1, X_2, \cdots, X_d$. The joint distributions function is $F(x_1, x_2, \cdots, x_d)$ with $u_1 = F_1(x_1), u_2 = F_2(x_2), \cdots, u_d = F_d(x_d)$. If $X_1 \perp X_2 \perp \cdots \perp X_d$, then $F(x_1, x_2, \cdots, x_d) = \prod_{i=1}^{d} F_i(x_i)$. And*

$$C(u_1, u_2, \cdots, u_d) = \prod_{i=1}^{d} u_i \tag{1.7}$$

for this situation.

Abe Sklar (1959) [34] introduced the concept and name of copula into probability theory and proved the theorem named after him. It stated that every continuous multivariate distribution functions can be expressed in terms of its marginals and a copula.

Theorem 1.2.5. ***Sklar's Theorem:*** *For a $d-$dimensional distributions function $F(x_1, x_2, \cdots, x_d)$ with marginal distributions $F_1, F_2, \cdots, F_d$ for random variables $X_1, X_2, \cdots, X_d$. Then $\forall x_i \in [-\infty, \infty], i \in \{1, 2, \cdots, d\}$, there exists a copula C, such that*

$$\begin{aligned} F(x_1, x_2, \cdots, x_d) &= P(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_d \leqslant x_d) \\ &= C(F_1(x_1), F_2(x_2), \cdots, F_d(x_d)) \\ &= C(u_1, u_2, \cdots, u_d). \end{aligned} \tag{1.8}$$

Sklar's theorem provided the theoretical foundation for the application of copulas and linked the $d-$dimensional distributions function $F(x_1, x_2, \cdots, x_d)$ and

$C(u_1, u_2, \cdots, u_d)$. By Sklar's theorem, for any multivariate distribution function with continuous marginal distribution functions, a copula can be defined.

Copulas themselves can be generated in several different ways, including the method of inversion, geometric methods, and algebraic methods. For instance, given a known multivariate distribution $F(x_1, x_2, \cdots, x_d)$ with continuous margins $F_i(x_i), i \in \{1, 2, \cdots, d\}$, the inverse method to obtain a copula is:

Definition 1.2.6. ***Copula:*** *For* $(u_1, u_2, \cdots, u_d) \in [0,1]^d$, *a* $d-$*dimensional copula* $C : [0,1]^d \to [0,1]$ *is*

$$C(u_1, u_2, \cdots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \cdots, F_d^{-1}(u_d)). \tag{1.9}$$

Once a copula was developed, it is easy to develop new multivariate distributions with arbitrary univariate marginal distribution.

Let $c(u_1, u_2, \cdots, u_d)$ be the density function of $C(u_1, u_2, \cdots, u_d)$, and $f(x_i), i \in \{1, 2, \cdots, d\}$ be the density function of $F(x_i)$, then

$$\begin{aligned} f(x_1, x_2, \cdots, x_d) &= \frac{\partial^d F(x_1, x_2, \cdots, x_d)}{\partial x_1 \partial x_2 \cdots \partial x_d} \\ &= \frac{\partial^d C(u_1, u_2, \cdots, u_d)}{\partial x_1 \partial x_2 \cdots \partial x_d} \\ &= \frac{\partial^d C(u_1, u_2, \cdots, u_d)}{\partial u_1 \partial u_2 \cdots \partial u_d} \frac{\partial u_1}{\partial x_1} \frac{\partial u_2}{\partial x_2} \cdots \frac{\partial u_d}{\partial x_d} \\ &= \frac{\partial^d C(u_1, u_2, \cdots, u_d)}{\partial u_1 \partial u_2 \cdots \partial u_d} \frac{\partial F_1(x_1)}{\partial x_1} \frac{\partial F_2(x_2)}{\partial x_2} \cdots \frac{\partial F_d(x_d)}{\partial x_d} \\ &= c(u_1, u_2, \cdots, u_d) \prod_{i=1}^{d} \frac{\partial F_i(x_i)}{\partial x_i} \\ &= c(u_1, u_2, \cdots, u_d) \prod_{i=1}^{d} f_i(x_i), \end{aligned} \tag{1.10}$$

that is

$$c(u_1, u_2, \cdots, u_d) = \frac{f(x_1, x_2, \cdots, x_d)}{\prod_{i=1}^{d} f_i(x_i)}. \tag{1.11}$$

Let T denote a survival time with distribution F. The survival function is given by

$$S(t) = P(T > t) = 1 - F(t). \tag{1.12}$$

Extend the previous definitions to the multivariate case. The multivariate survival function $S(t_1, t_2, \cdots, t_d)$ is defined by

$$S(t_1, t_2, \cdots, t_d) = P(T_1 > t_1, T_2 > t_2, \cdots, T_d > t_d), \tag{1.13}$$

where the $T_1, T_2, \cdots, T_d$ are survival times with univariate survival functions $S_i(t_i), i \in \{1, 2, \cdots, d\}$. Then

$$\begin{aligned} S_i(t_i) &= P(T_i > t_i) \\ &= P(T_1 > 0, T_2 > 0, \cdots, T_{i-1} > 0, T_i > t_i, T_{i+1} > 0, \cdots, T_d > 0) \\ &= S(0, 0, \cdots, 0, t_i, 0, \cdots, 0). \end{aligned} \tag{1.14}$$

Please note that the relationship between the multivariate survival function S and the multivariate distribution function F is not direct same as the univariate case, which is

$$S(t_1, t_2, \cdots, t_d) \neq 1 - F(t_1, t_2, \cdots, t_d). \tag{1.15}$$

The following properties and theorems of survival copulas have been studied and discussed by P. Georges, A-G. Lamy, et al. (2001) [11].

Definition 1.2.7. ***Survival Copula*** *With marginal survival functions* $S_i(t_i), i \in \{1, 2, \cdots, d\}$*, a* $d-$*dimensional survival copula* $\tilde{C}$ *can be defined as*

$$S(t_1, t_2, \cdots, t_d) = \tilde{C}(S_1(t_1), S_2(t_2), \cdots, S_d(t_d)). \tag{1.16}$$

Example 1.2.8. ***An example for survival copula with*** $\mathbf{d = 2}$***:*** *Let* C *be the copula function of the bivariate distribution of* (T_1, T_2) *with* $u_1 = S_1(t_1), u_2 = S_2(t_2)$*, then*

$$\begin{aligned} S(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) \\ &= 1 - F_1(t_1) - F_2(t_2) + F(t_1, t_2) \\ &= 1 - F_1(t_1) + (1 - F_2(t_2)) - 1 + F(t_1, t_2) \\ &= S_1(t_1) + S_2(t_2) - 1 + C(1 - S_1(t_1), 1 - S_2(t_2)) \\ &= \tilde{C}(S_1(t_1), S_2(t_2)). \end{aligned} \tag{1.17}$$

That is

$$\tilde{C}(S_1(t_1), S_2(t_2)) = S_1(t_1) + S_2(t_2) - 1 + C(1 - S_1(t_1), 1 - S_2(t_2)), \tag{1.18}$$

which is same as

$$\tilde{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2). \tag{1.19}$$

The survival copula $\tilde{C}$ is a copula function which satisfied:

1. *The marginal distribution function of $\tilde{C}$ are uniform;*

$$\begin{aligned} \tilde{C}(u_1, 1) &= u_1 + 1 - 1 + C(1 - u_1, 1 - 1) \\ &= u_1 + C(1 - u_1, 0) \\ &= u_1 + 0 \\ &= u_1 \end{aligned} \tag{1.20}$$

and

$$\begin{aligned} \tilde{C}(1, u_2) &= 1 + u_2 - 1 + C(1 - 1, 1 - u_2) \\ &= u_2 + C(0, 1 - u_2) \\ &= u_2 + 0 \\ &= u_2 \end{aligned} \tag{1.21}$$

2. *If any* $u_i = 0$ *in* $\boldsymbol{u}$*, then*

$$\begin{aligned}\tilde{C}(u_1, 0) &= u_1 + 0 - 1 + C(1 - u_1, 1 - 0) \\ &= u_1 - 1 + C(1 - u_1, 1) \\ &= u_1 - 1 + 1 - u_1 \\ &= 0\end{aligned} \tag{1.22}$$

and

$$\begin{aligned}\tilde{C}(0, u_2) &= 0 + u_2 - 1 + C(1 - 0, 1 - u_2) \\ &= u_2 - 1 + C(1, 1 - u_2) \\ &= u_2 - 1 + 1 - u_2 \\ &= 0\end{aligned} \tag{1.23}$$

3. $\tilde{C}(u_1, u_2)$ *has a non-negative value for any two-dimensional interval, that is* $\forall (v_1, v_2) \in [0, 1]^2$ *and* $\forall (w_1, w_2) \in [0, 1]^2$ *with* $v_i \geqslant w_i$ *and* $i \in \{1, 2\}$*,* $P(U_1 \in [v_1, w_1], U_2 \in [v_2, w_2]) \geqslant 0$*. This implies the rectangle inequality*

$$\begin{aligned}&\tilde{C}(v_1, v_2) - \tilde{C}(v_1, w_2) - \tilde{C}(w_1, v_2) + \tilde{C}(w_1, w_2) \\ =&[v_1 + v_2 - 1 + C(1 - v_1, 1 - v_2)] - [v_1 + w_2 - 1 + C(1 - v_1, 1 - w_2)] - \\ &[w_1 + v_2 - 1 + C(1 - w_1, 1 - v_2)] + [w_1 + w_2 - 1 + C(1 - w_1, 1 - w_2)] \\ =&C(1 - v_1, 1 - v_2) - C(1 - v_1, 1 - w_2) - \\ &C(1 - w_1, 1 - v_2) + C(1 - w_1, 1 - w_2) \geqslant 0\end{aligned} \tag{1.24}$$

In the general case, similar results will be obtained.

Theorem 1.2.9. *The relationship between the copula C and the survival copula $\tilde{C}$ is given by*

$$\tilde{C}(u_1, u_2, \cdots, u_d) = \sum_{i=0}^{d} \left[(-1)^i \sum_{\mathbf{v}(u_1,u_2,\cdots,u_d)\in\mathcal{Z}(d-i,d,1)} C(v_1, v_2, \cdots, v_d) \right] \tag{1.25}$$

where $\mathcal{Z}(M, N, 1)$ denotes the set $\{\mathbf{v} \in [0,1]^d \mid v_i \in [u_i, 1], \sum_{i=1}^{d} \mathcal{X}_1(v_i) = M\}$. And

$$C(u_1, u_2, \cdots, u_d) = \sum_{i=0}^{d} \left[(-1)^i \sum_{\mathbf{v}(u_1,u_2,\cdots,u_d)\in\mathcal{Z}(d-i,d,0)} \tilde{C}(1-v_1, 1-v_2, \cdots, 1-v_d) \right] \tag{1.26}$$

where $\mathcal{Z}(M, N, 0)$ denotes the set $\{\mathbf{v} \in [0,1]^d \mid v_i \in [0, u_i], \sum_{i=1}^{d} \mathcal{X}_0(v_i) = M\}$.

In the general case, the survival copulas are not same as copulas except in some cases. For example, it can be shown that for elliptical copulas $C = \tilde{C}$ (Gaussian, student's t). It is also true for the Clayton copulas [3] and Frank copulas [7].

Property 1.2.10. *The copula is radially symmetric if and only if*

$$C = \tilde{C} \tag{1.27}$$

Then it is equivalent to work with the copula or to work with the survival copula. It is a very interesting property especially for the computational purpose.

Please note that $C(u_1, u_2, \cdots, u_d)$ is a non-decreasing function in each variable $u_i, i \in \{1, 2, \cdots, d\}$. And $\tilde{C}(u_1, u_2, \cdots, u_d)$ is a non-increasing function in each variable $u_i, i \in \{1, 2, \cdots, d\}$ which is different.

1.2.2 Motivation

A rich set of copula families have been developed. Such as the Farlie-Gumbel-Morgenstern (FGM) copulas, Archimedean copulas, quadratic copulas, cubic copulas, meta-elliptical copulas (including Gaussian copulas, student's t copulas) and plackett copulas are all commonly used copulas. Because of an easy and explicit formula, the simple way of construction and the nice properties associated, Archimedean copulas are popular and widely used.

This dissertation will focus on Archimedean copulas. Based on the features of exchangeability and non-exchangeability, copulas can be divided into two categories: symmetric copulas and asymmetric copulas. For example, the one-parameter Archimedean copulas are symmetric copulas. Nested Archimedean copulas and the vine copulas through pair-copula construction are asymmetric copulas. Details will be discussed in the next few sections.

Symmetric Archimedean copulas which also called exchangeable Archimedean copulas are popular due to they allow modeling dependence in arbitrarily high dimensions and result in a straightforward calculation. Most commonly used symmetric Archimedean copulas have an easy and explicit formula and one parameter setting, for instance, Clayton symmetric Archimedean copulas [3], Gumbel-Hougaard copulas [13, 14] and Frank copulas [7].

It is too strict that symmetric Archimedean copulas require equal dependence among different pairs. In practice, the data could be positive correlated, negative correlated or independent. There are also many asymmetric data need to be considered, especially for the high dimensional data.

In general, bivariate distribution can be handled easily. The joint distribution can be described by symmetric Archimedean copulas. Due to the preconditions of symmetric Archimedean copulas, which require a symmetric dependence between different pairs of variables and result in only one parameter is allowed, there is a limitation for high dimensional data when the dimension of a data set is greater than 2 ($d > 3$). More detailed explanation is discussed below.

The construction of multivariate model via the frailty model is one important strategy to extend bivariate models to higher dimensions ($d \geqslant 3$). Let $S_1, S_2, \cdots, S_d$ be univariate survival functions of $T_1, T_2, \cdots, T_d$, respectively and W be a positive random variable with Laplace transform $\psi(s) = E[\exp(-sW)]$. Let $T_1, T_2, \cdots, T_d$ be dependent random variables that are conditional independent given $W = w$ such that $\Pr(T_i > t_i|w) = B_i(t_i)^w$ where $B_i(t)$ is defined as the baseline distribution function for T_i. Then it can be easily shown that

$$\begin{aligned} S(t_1, t_2, \cdots, t_d) &= \Pr(T_1 > t_1, T_2 > t_2, \cdots, T_d > t_d) \\ &= \psi\left\{\psi^{-1}\left[S_1(t_1)\right] + \psi^{-1}\left[S_2(t_2)\right] + \cdots + \psi^{-1}\left[S_d(t_d)\right]\right\}, \end{aligned} \tag{1.28}$$

where ψ^{-1} is the inverse function of ψ.

When my adviser Dr. Wang was a PhD student of Dr. David Oakes at University of Rochester, he took the course "Frailty Models" of Dr. Oakes. In the

lecture notes of it [25], Dr. Oakes made the following comments about the frailty models:

1. The frailty representation is very convenient for two survival times, but we may have $d-$variate data $(T_1, T_2, \cdots, T_d)$;
2. An "exchangeable" multivariate joint distribution is easily derived if all the components depend on the same frailty W. It can be obtained by

$$S(t_1, t_2, \cdots, t_d) = \psi\left\{\psi^{-1}\left[S_1(t_1)\right] + \psi^{-1}\left[S_2(t_2)\right] + \cdots + \psi^{-1}\left[S_d(t_d)\right]\right\}; \tag{1.29}$$

3. Although this gives total flexibility as to the marginal distribution $S_i(t_i)$, $i = 1, 2, \cdots, d$, it forces exchangeability of the dependence structure;
4. For example, if W follows gamma distribution, then the Clayton odds ratio θ are the same for all pairs of components;
5. It would be nice to have a model allowing $\theta_{i,j}$ for T_i and T_j to differ. To date, it is not even known what conditions need to impose on the $\theta_{i,j}$, $1 \leqslant i < j \leqslant d$ to ensure the existence of a $d-$variate distribution with Clayton [3] type bivariate margin.

From above comments, it is obvious that an "exchangeable" multivariate model is not very useful in practice and an asymmetric (nonexchangeable) multivariate copula model with any pre-specified bivariate margins are preferred.

Other from the construction method for symmetric multivariate copulas which was just described, there exist mainly two approaches of constructing asymmetric multivariate copulas: Nested Archimedean Copula Construction (NACC) and the vine copulas through Pair-Copula Construction (PCC) as pointed out by Zhang and Singh (2019) [46].

Recently a series of studies have been introduced, for instance, fully nested Archimedean copulas were introduced and studied by Joe (1997) [15], Embrechts et

al. (2001) [4], Whelan (2004) [45], Savu and Trede (2010) [30]. Partially nested Archimedean copulas were first introduced by Joe (1997) [15]. General nested Archimedean copulas also called hierarchical Archimedean copulas are a combination of the two nested Archimedean copulas methods. Kurowicka and Cooke (2005) [17, 18], introduced the canonical vine as a model construction method. Aas and Berg (2009) [1] introduced $D-$vine as another model construction method. More details will be discussed in the next few sections.

1.2.3 Fully nested Archimedean copulas

Fully nested Archimedean copulas were introduced and studied by Joe (1997) [15], Embrechts et al. (2001) [4], Whelan (2004) [45], Savu and Trede (2010) [30]. Take the four-dimensional structure of fully nested Archimedean copulas for example, the Figure 1.1 shows how to construct it from two-dimension to four-dimension.

Variables u_1 and u_2 are dependent with each other, the joint distribution function of (u_1, u_2) can be modeled by a copula $C_3(u_1, u_2)$. Next, variable u_3 and $C_3(u_1, u_2)$ are dependent, the joint distribution function of (u_1, u_2, u_3) can be modeled by a copula $C_2(u_3, C_3(u_1, u_2))$. Last, variable u_4 and $C_2(u_3, C_3(u_1, u_2))$ are dependent, the joint distribution function of (u_1, u_2, u_3, u_4) can be modeled by an overall copula $C_1(u_4, C_2(u_3, C_3(u_1, u_2)))$ with distribution function

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= C_1\left(u_4, C_2(u_3, C_3(u_1, u_2))\right) \\ &= \varphi_1^{-1}\left(\varphi_1(u_4) + \varphi_1(\varphi_2^{-1}(\varphi_2(u_3) + \varphi_2(\varphi_3^{-1}(\varphi_3(u_1) + \varphi_3(u_2)))))\right) \\ &= \varphi_1^{-1}\left(\varphi_1(u_4) + \varphi_1 \circ \varphi_2^{-1}(\varphi_2(u_3) + \varphi_2 \circ \varphi_3^{-1}(\varphi_3(u_1) + \varphi_3(u_2)))\right). \end{aligned} \tag{1.30}$$

Symbol "$\circ$" represent the product of two functions, such four-dimensional copula was constructed by three copulas C_1, C_2, and C_3 with generator φ_1, φ_2, and φ_3 respectively.

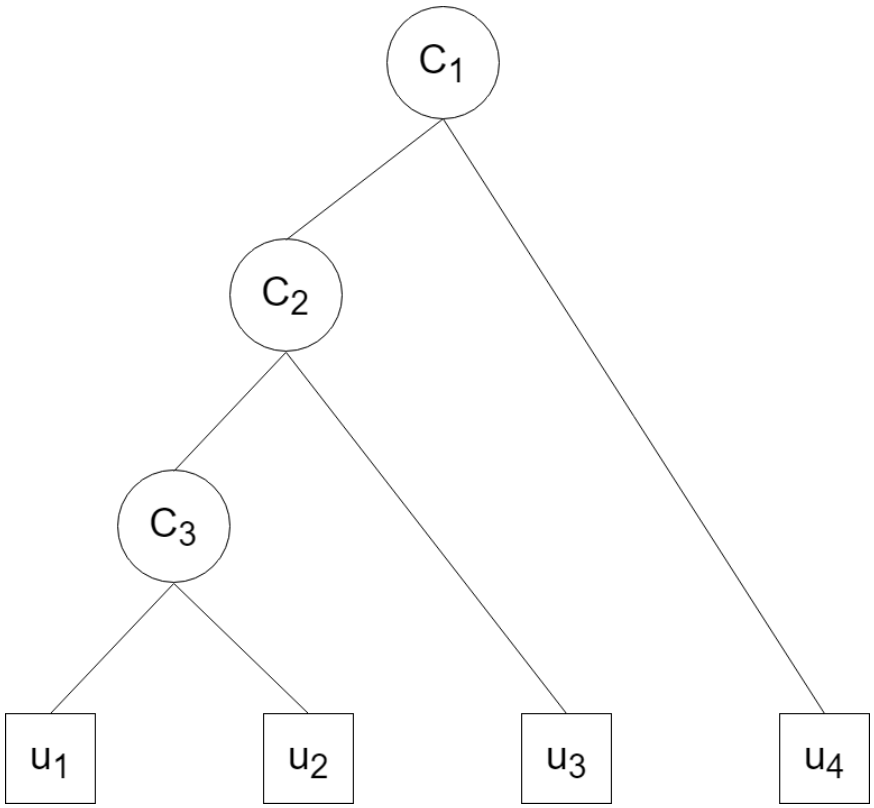

Figure 1.1 Fully nested Archimedean copulas (four-dimensional structure).

.

That is

- (u_1, u_2) is modeled by a copula C_3 with parameter θ_3 and generator φ_3;
- (C_3, u_3) is modeled by a copula C_2 with parameter θ_2 and generator φ_2;
- (C_2, u_4) is modeled by a copula C_1 with parameter θ_1 and generator φ_1.

In general, a $d-$ dimensional copula can be constructed by a two-dimensional copula (total $\frac{d(d-1)}{2}$ different options for the starting two-dimensional copula) and $d-1$ generators with parameters $\theta_1 \geqslant \theta_2 \geqslant \cdots \geqslant \theta_{d-1}$, and all the inverse functions $\varphi_1^{-1}, \varphi_2^{-1}, \cdots \varphi_{d-1}^{-1}$ are monotonic functions to make sure the final copula is a cumulative distribution function by Aas and Berg (2009) [1].

1.2.4 Partially nested Archimedean copulas

Partially nested Archimedean copulas were first introduced by Joe (1997) [15]. Take the four-dimensional structure of partially nested Archimedean copulas for example, the Figure 1.2 shows how to construct it from two-dimension to four-dimension.

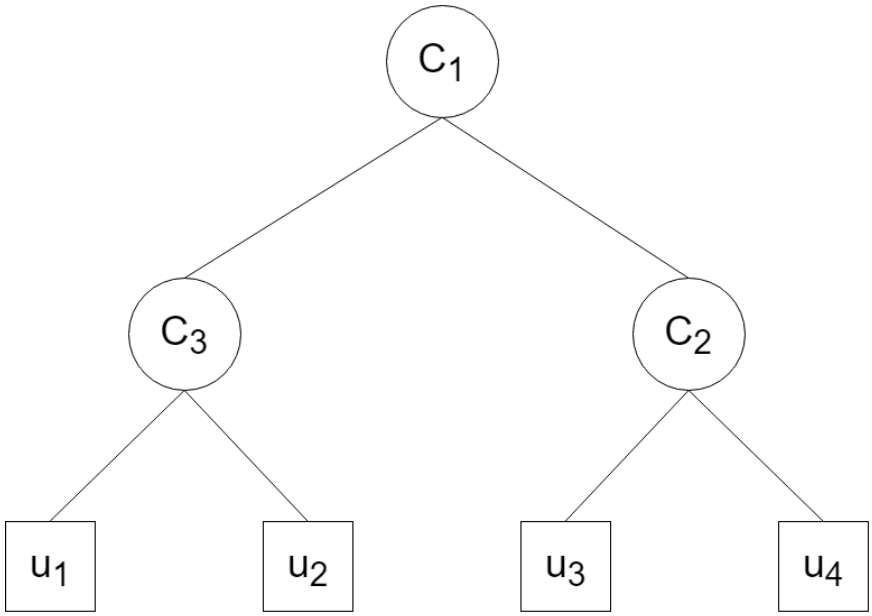

Figure 1.2 Partially nested Archimedean copulas (four-dimensional structure).

.

Variables u_1 and u_2 are dependent with each other and variable u_3 and u_4 are dependent with each other. The joint distribution function of (u_1, u_2) and (u_3, u_4) can be modeled by copula $C_3(u_1, u_2)$ and $C_2(u_3, u_4)$ respectively. Next, copula $C_3(u_1, u_2)$

and copula $C_2(u_3, u_4)$ are dependent with each other, the joint distribution function of (u_1, u_2, u_3, u_4) can be modeled by an overall copula $C_1(C_3(u_1, u_2), C_2(u_3, u_4))$. That is

- (u_1, u_2) is modeled by a copula C_3 with parameter θ_3 and generator φ_3;
- (u_3, u_4) is modeled by a copula C_2 with parameter θ_2 and generator φ_2;
- (C_2, C_3) is modeled by a copula C_1 with parameter θ_1 and generator φ_1.

Aas and Berg [1], pointed out that the exchangeability between u_1 and u_2 within copula C_3, and the exchangeability between u_3 and u_4 within copula C_2 in such a four-dimensional partially nested Archimedean copula. Conclude that the partially nested Archimedean copulas can be understand as a combination of exchangeable Archimedean copulas and fully nested Archimedean copulas.

1.2.5 General nested Archimedean copulas

General nested Archimedean copulas are also called hierarchical Archimedean copulas. The general expressions for them are mainly based on the nested Archimedean copulas, and the l level Archimedean copulas is generated by $(l-1) \geqslant 0$ level Archimedean copulas. Then, the top-level copula is constructed by hierarchical structures. The $d-$dimensional jointed distribution has an estimation at point $u = (u_1, u_2, \cdots u_d) \in [0,1]^d$ and total L levels with level number l, where $l = 0, 1, 2, \cdots, L$. At level l, there are n_l groups with group number $j = 1, 2, \cdots, n_l$.

At the lowest level $l = 0$, there are variables $u_1, u_2, \cdots u_d$. They have been divided into n_1 different groups in the next level $l = 1$, each group is modeled by a

regular Archimedean copula $C_{1,j}$, where $j = 1, 2, \cdots, n_1$ with formula

$$C_{1,j}(u_{j,dj}) = \varphi_{1,j}^{-1}\left(\sum_{\{u_{j,dj}\}} \varphi_{1,j}(u_{j,dj})\right). \tag{1.31}$$

- Each copula $C_{1,j}$ has its generator $\varphi_{1,j}, j = 1, 2, \cdots, n_1$;
- $\{u_{j,dj}\}$ is the set contains all the variables which are belong to the copula $C_{1,j}, j = 1, 2, \cdots, n_1$;
- $\sum_{\{u_{j,dj}\}} \varphi_{1,j}(u_{j,dj}) = \varphi_{1,j}(u_{j,1}) + \varphi_{1,j}(u_{j,2}) + \cdots + \varphi_{1,j}(u_{j,dj})$;
- $C_{1,1}, C_{1,2}, \cdots, C_{1,n_1}$ may have different kind of Archimedean copula, for instant, Clayton copula [3], Gumbel - Hougaard copula [13, 14], Frank copula [7] and so on.

At the level $l = 2$, copulas $C_{1,j}(j = 1, 2, \cdots, n_1)$ have been divided into n_2 different groups in the next level $l = 2$, each group is modeled by a regular Archimedean copula $C_{2,i}$, where $i = 1, 2, \cdots, n_2$ with partial exchangeability based on their structures with formula

$$C_{2,i}(C_{1,k_i}) = \varphi_{2,i}^{-1}\left(\sum_{\{C_{1,k_i}\}} \varphi_{2,i}(C_{1,k_i})\right). \tag{1.32}$$

- Each copula $C_{2,i}$ has its generator $\varphi_{2,i}, i = 1, 2, \cdots, n_2$;
- $\{C_{1,k_i}\}$ is the set contains all the copulas in level $l = 1$ which belong to the copula $C_{2,i}, i = 1, 2, \cdots, n_2$;
- $C_{2,1}, C_{2,2}, \cdots, C_{2,n_2}$ may have different kind of Archimedean copula, for instant, Clayton copula [3], Gumbel - Hougaard copula [13, 14], Frank copula [7] and so on.

Repeat these steps, a L levels hierarchical Archimedean copula $C_{L,1}$ can be generated. To make sure the $d-$ dimensional copula $C_{L,1}$ is a cumulative distribution function, the following three criteria must be met:

- Each inverse function $\varphi_{l,j}^{-1}$ must be a strict monotonic functions, where $l = 1, 2, \cdots, L$ and $j = 1, 2, \cdots, n_l$;
- The number of copulas in each level is decreasing when l increased which is $n_{l+1} < n_l$, where $l = 0, 1, 2, \cdots, L$.

The Figure 1.3 shows how a $L-$level hierarchical Archimedean copula $C_{L,1}$ looks like.

Ideally, there are $\frac{d(d-1)}{2}$ different bivariate combinations for a $d-$ dimensional data. But the nested Archimedean structure cannot satisfy the requirement for including all the $\frac{d(d-1)}{2}$ different dependent combinations in one structure by Aas and Berg (2009) [1].

1.2.6 Pair copulas

After Joe (1997) [15] first introduced pair copulas, a series of papers studied them. Compare to exchangeable Archimedean copulas, there are two major improvements for pair copulas. One improvement is pair copulas allow $\frac{d(d-1)}{2}$ assigned combinations of copulas. The idea is break down a multivariate density function into $\frac{d(d-1)}{2}$ bivariate copulas. Among them, there are $d-1$ unconditional copulas and the rest are conditional copulas by Aas and Berg (2009) [1]. And the other

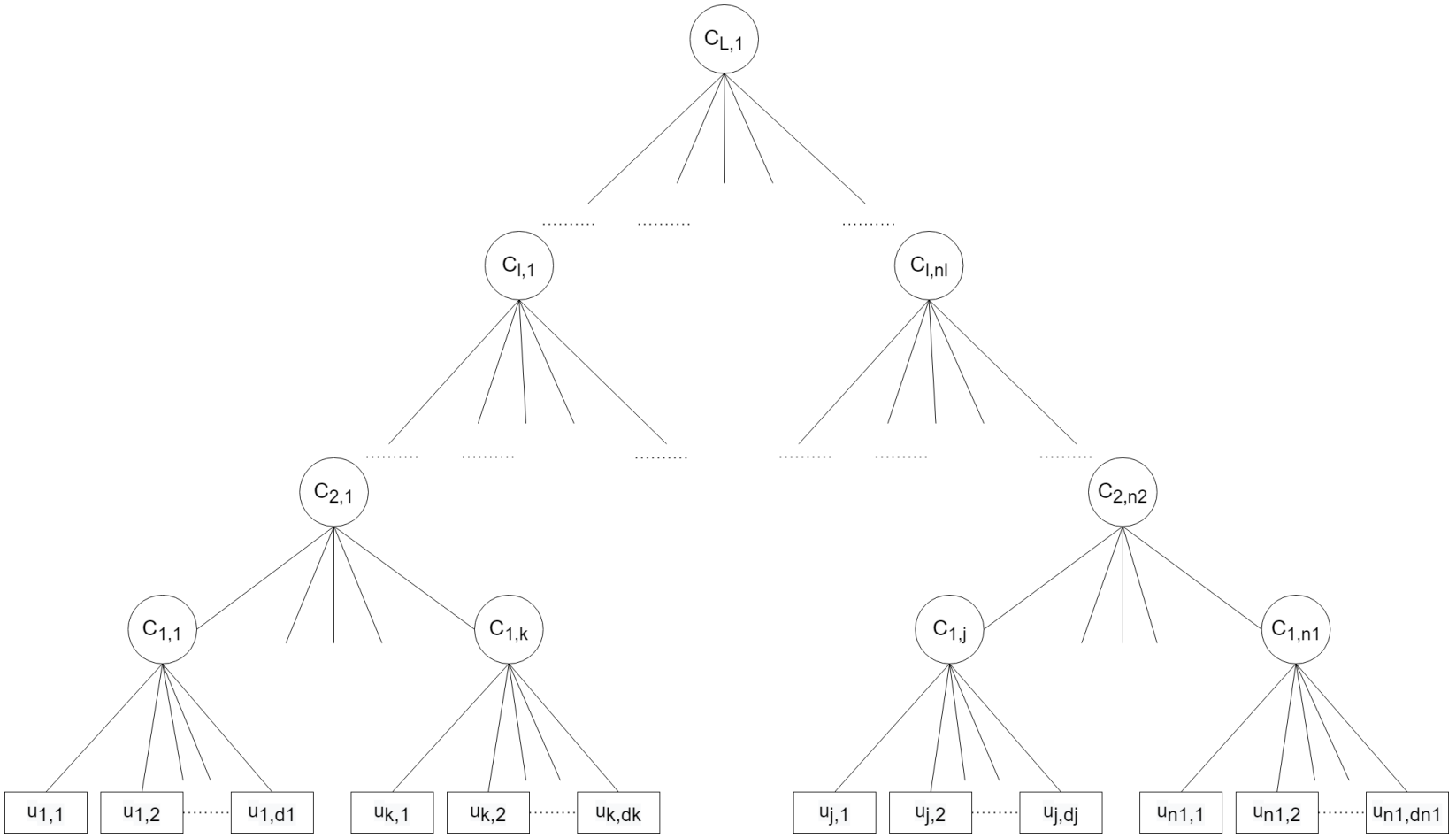

Figure 1.3 General nested Archimedean copulas.

.

major improvement is the $\frac{d(d-1)}{2}$ bivariate copulas are not limited to Archimedean copulas.

Kurowicka and Cooke (2005) [17, 18], introduced the Canonical vine as construction method. Take the four-dimensional structure of pair copulas for example, the Figure 1.4 shows how to construct it from two-dimension to four-dimension.

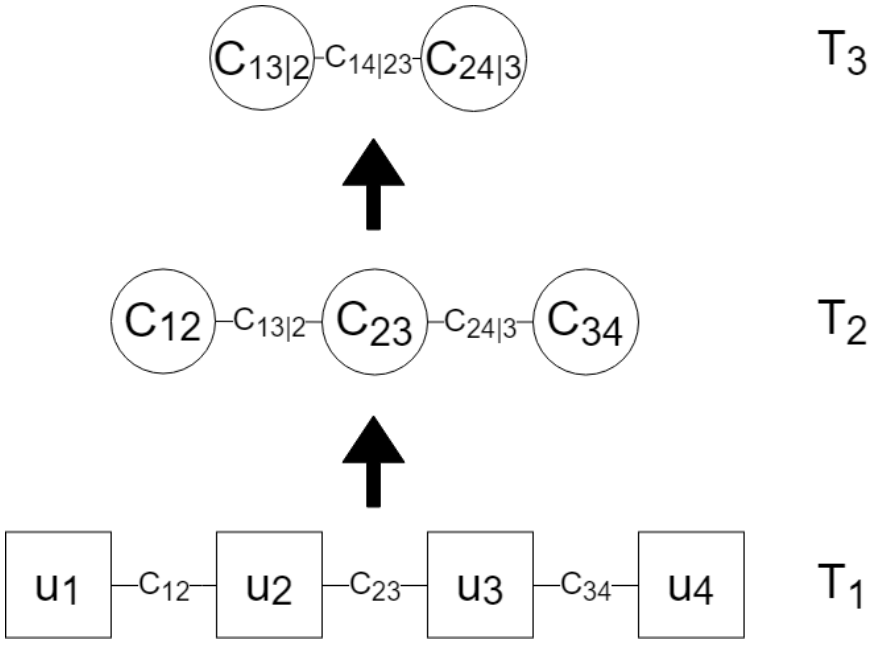

Figure 1.4 Pair copulas (four-dimensional structure with Canonical vine).

.

Aas and Berg (2009) [1] introduced $D-$vine as another construction method. Take the four-dimensional structure of pair copulas for example, Figure 1.5 shows how to construct it from two-dimension to four-dimension.

1.2.7 Discussion

Take the four-dimensional data as an example. If the data is modeled by fully nested Archimedean copulas structure as illustrated in Figure 1.1 or the data is modeled by

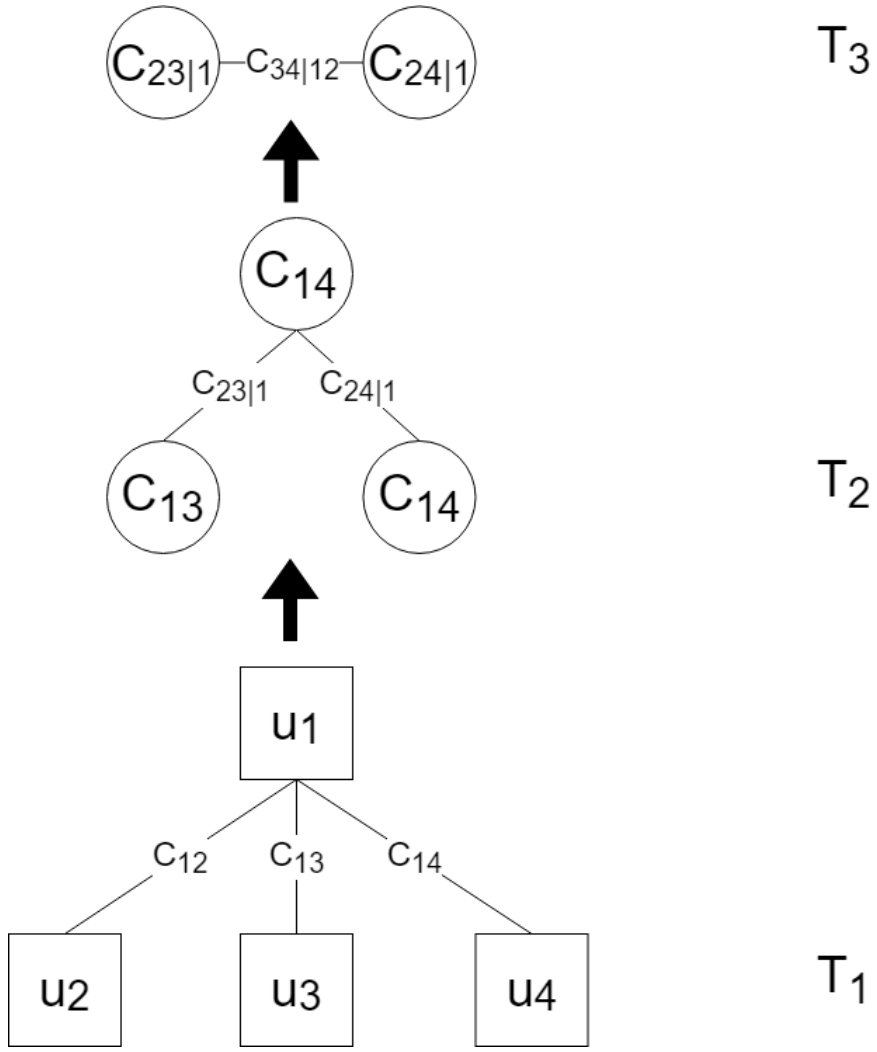

Figure 1.5 Pair copulas (four-dimensional structure with $D-$vine).

.

partially nested Archimedean copulas structure as illustrated in Figure 1.2, copula C_3 describes the dependence between u_1 and u_2. The question is how to model the dependence between u_2 and u_3? It seems that the question can be solved by using the four-dimensional pair copulas structure with Canonical vine as illustrated in Figure 1.4 which is copula C_{23}. The next question is how to model the dependence between u_1 and u_4? It seems the four-dimensional Pair copulas structure with $D-$vine as illustrated in Figure 1.5 which is copula C_{14} can provide a better answer to it. However, if the question change to how to model the dependence between u_2 and u_4, none of the methods mentioned above can solve this question. And these structures are quite complicated because of the complexity of the nest and vine.

Ideally, it is better to have models that can accommodate $\frac{d(d-1)}{2}$ different pre-specified and unconditional bivariate margins for a $d-$dimensional data. Both the nested Archimedean structure and the vine copula structure cannot satisfy this requirement by Aas and Berg (2009) [1]. Another limitation of the vine copula is that the dependence structure of the vine copula tends to be quite complicated. If d is large (for example, $d \geqslant 500$), the dependence structure of the vine copula is almost intractable.

With above considerations, most of the current structures are not very flexible [37, 43, 42], the research goal here is to construct general models allowing arbitrary selection of pairwise correlation which is desired in our practical applications. The proposed research is an extension of the model construction method proposed by Chakak (1993) and is a big step forward as researchers have searched for such a flexible class of models for decades.

In this time and era where each and every aspect of our day to day life has been technologized, there are many different types of data that generated from various sources. Needless to say, there are a lot of challenges in the analysis and study of such different kinds of multivariate distributed data with the methods mentioned

before. It is then necessary to consider any other structure which can provide more flexibility and better practicality. It must be easy to use, to extend and stands in line with the actual phenomenon. The detailed idea, approach and methods will be discussed in the next chapter.

CHAPTER 2

ASYMMETRIC MULTIVARIATE ARCHIMEDEAN COPULA MODELS

2.1 Introduction

This research is an extension to the survival copula of the model construction method proposed by Chakak (1993) [2] which allows arbitrary selection of pairwise correlation. It has a fixed or a pre-specified bivariate marginal distribution and allows for different coefficients for different pairs of random variables. It is a big step forward since such structure have been searched for decades.

The following discussion will focus on the survival copula which can be easily transferred to copula. All the copulas discussed in Chapter 2 of this dissertation are survival copulas, denoted as C instead of $\tilde{C}$ to simplify the notations.

In the following sections, a structure for more than two dimensions will be constructed, and the feasibility of the structure to d dimensions will be proved by mathematical induction method. For a better understanding, some demonstrations for the approach based upon the proposed structure will be given. The examples will focus on the Clayton copulas [3], Gumbel - Hougaard copulas [13, 14], and Frank copulas [7]. Methods regarding parameter estimation, model selection and data simulation will be discussed at this chapter.

2.2 Method of Constructing Asymmetric Multivariate Archimedean Copula Models

Suppose that all the univariate marginal survival functions are absolutely continuous. That is, for all $t_j, t_k (\forall j, k \in N)$ in the support of the survival functions, it is easy to

have

$$t_j = S_j^{-1}\left(S_j(t_j)\right), \tag{2.1}$$

and

$$t_k = S_k^{-1}\left(S_k(t_k)\right). \tag{2.2}$$

So that the bivariate joint survival functions $S_{jk}(t_j, t_k)$ with marginal survival functions $S_j(t_j)$ and $S_k(t_k)$ can be written as

$$\begin{aligned} S_{jk}(t_j, t_k) &= S_{jk}\left(S_j^{-1}\left(S_j(t_j)\right), S_k^{-1}\left(S_k(t_k)\right)\right) \\ &= C_{jk}\left(S_j(t_j), S_k(t_k)\right) \end{aligned} \tag{2.3}$$

where C_{jk} is a copula which is determined uniquely by S_{jk}. If S_{jk} is absolutely continuous, the copula C_{jk} is differentiable which means $\dfrac{\partial^2}{\partial t_j \partial t_k} C_{jk}(t_j, t_k)$ exists.

The following properties were reviewed by Schweizer and Wolff (1981) [32], Schweizer and Sklar (1983) [31]:

Property 2.2.1. *For* $\forall\ t_j, t_k, C = C_{jk}$, *if* $u = S_j(t_j) \in [0,1]$, $v = S_k(t_k) \in [0,1]$, *and* $\forall\ 0 \leqslant u_1 \leqslant u_2 \leqslant 1$, $\forall\ 0 \leqslant v_1 \leqslant v_2 \leqslant 1$, *then*

1. $C(u,1) = u, C(1,v) = v$;
2. $C(u,0) = 0, C(0,v) = 0$;
3. $C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geqslant 0$.

For any $d-$dimensional copula ($d \geqslant 3$), $u_i = S_i(t_i) \in [0,1], i = \{1,2,\cdots,d\}$, the following properties have been proved:

Lemma 2.2.2. *Assuming that the joint survival function of* $(U_1, U_2, \cdots, U_d)$ *is defined on* $[0,1]^d$ *and the marginal distributions are all uniform* $[0,1]$, *if* C *is the corresponding copula function such that*

$$C(u_1, u_2, \cdots, u_d) = Pr(S(T_1) \leqslant u_1, S(T_2) \leqslant u_2, \cdots, S(T_d) \leqslant u_d)$$

where $S_i(t_i) = U_i, i \in \{1,2,\cdots,d\}$ *is the survival function. Then* C *has the following properties:*

1. $C(u_1, u_2, \cdots, u_{i-1}, 1, u_{i+1}, \cdots, u_d) = C(u_1, u_2, \cdots, u_{i-1}, u_{i+1}, \cdots, u_d)$.
2. $C(u_1, u_2, \cdots, u_{i-1}, 0, u_{i+1}, \cdots, u_d) = 0$.

Proof. For $\forall\ u_i = 1, i \in \{1, 2, \cdots, d\}$, then

$$
\begin{aligned}
&C(u_1, u_2, \cdots, u_{i-1}, 1, u_{i+1}, \cdots, u_d) \\
&= P(U_1 < u_1, U_2 < u_2, \cdots, U_{i-1} < u_{i-1}, U_i < 1, U_{i+1} < u_{i+1}, \cdots, U_d < u_d) \\
&= P(S_1(t_1) < u_1, S_2(t_2) < u_2, \cdots, S_{i-1}(t_{i-1}) < u_{i-1}, S_i(t_i) < 1, \\
&\qquad S_{i+1}(t_{i+1}) < u_{i+1}, \cdots, S_d(t_d) < u_d) \\
&= P(S_1(t_1) < u_1, S_2(t_2) < u_2, \cdots, S_{i-1}(t_{i-1}) < u_{i-1}, \\
&\qquad S_{i+1}(t_{i+1}) < u_{i+1}, \cdots, S_d(t_d) < u_d) \\
&= P(U_1 < u_1, U_2 < u_2, \cdots, U_{i-1} < u_{i-1}, U_{i+1} < u_{i+1}, \cdots, U_d < u_d) \\
&= C(u_1, u_2, \cdots, u_{i-1}, u_{i+1}, \cdots, u_d)
\end{aligned}
$$

For $\forall\ u_i = 0, i \in (1, 2, \cdots, d)$, then

$$
\begin{aligned}
&C(u_1, u_2, \cdots, u_{i-1}, 0, u_{i+1}, \cdots, u_d) \\
&= P(U_1 < u_1, U_2 < u_2, \cdots, U_{i-1} < u_{i-1}, U_i < 0, U_{i+1} < u_{i+1}, \cdots, U_d < u_d) \\
&= P(S_1(t_1) < u_1, S_2(t_2) < u_2, \cdots, S_{i-1}(t_{i-1}) < u_{i-1}, S_i(t_i) < 0, \\
&\qquad S_{i+1}(t_{i+1}) < u_{i+1}, \cdots, S_d(t_d) < u_d) \\
&= P(\emptyset) \\
&= 0
\end{aligned}
$$

□

Lemma 2.2.2 is useful in proving the properties for the proposed multivariate models. The research goal is to find a model with any pre-specified bivariate marginal survival function $S_{jk}(t_j, t_k)$. The idea underlying the method of construction by

Chakak (1993) [2] is that, the bivariate survival function

$$S_{1\cdot 2}(t_1, t_2) = P(T_1 > t_1 | T_2 > t_2) S_2(t_2) \tag{2.4}$$

can be determined completely by the conditional probability since the marginal distribution function is specified. The three-dimensional construction approach has joint survival function as

$$S_{12\cdot 3}(t_1, t_2, t_3) = P(T_1 > t_1, T_2 > t_2 | T_3 > t_3) S_3(t_3). \tag{2.5}$$

Given t_3, the bivariate conditional survival function $S_{12|3}(t_1, t_2 | T_3 > t_3)$ has continuous marginals denoted by

$$S_{1|3}(t_1 | T_3 > t_3) = P(T_1 > t_1 | T_3 > t_3), \tag{2.6}$$

and

$$S_{2|3}(t_2|T_3 > t_3) = P(T_2 > t_2|T_3 > t_3). \tag{2.7}$$

with densities

$$\begin{aligned} \partial_{t_1} S_{1|3}(t_1|T_3 > t_3) &= \frac{\partial}{\partial t_1}[P(T_1 > t_1|T_3 > t_3)] \\ &= \frac{\partial}{\partial t_1}\left[\frac{S_{13}(t_1, t_3)}{S_3(t_3)}\right] \\ &= \frac{\frac{\partial}{\partial t_1} S_{13}(t_1, t_3)}{S_3(t_3)}, \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} \partial_{t_2} S_{2|3}(t_2|T_3 > t_3) &= \frac{\partial}{\partial t_2}[P(T_2 > t_2|T_3 > t_3)] \\ &= \frac{\partial}{\partial t_2}\left[\frac{S_{23}(t_2, t_3)}{S_3(t_3)}\right] \\ &= \frac{\frac{\partial}{\partial t_2} S_{23}(t_2, t_3)}{S_3(t_3)} \end{aligned} \tag{2.9}$$

respectively. Using Sklar's Theorem (1959) [34], there exists a copula $C_{12|3}$ such that

$$\begin{aligned} & C_{12|3}\left(S_{1|3}(t_1|T_3 > t_3), S_{2|3}(t_2|T_3 > t_3)\right) \\ & = P(T_1 > t_1, T_2 > t_2|T_3 > t_3) \\ & = S_{12|3}\left(S_{1|3}^{-1}\left(S_{1|3}(t_1|T_3 > t_3)\right), S_{2|3}^{-1}\left(S_{2|3}(t_2|T_3 > t_3)\right)\right) \end{aligned} \tag{2.10}$$

where $S_{12|3}$ is the joint conditional survival function of T_1 and T_2 given $T_3 > t_3$.

Similar results of copula $C_{23|1}$ and $C_{13|2}$ can be derived on condition $T_1 > t_1$ and on condition $T_2 > t_2$ respectively. Then for any (t_1, t_2, t_3) in the support of joint survival function, the following joint survival functions can be derived

$$S_{23\cdot1}(t_1, t_2, t_3) = C_{23|1}(S_{2|1}(t_2), S_{3|1}(t_3))S_1(t_1), \tag{2.11}$$

$$S_{13\cdot2}(t_1, t_2, t_3) = C_{13|2}(S_{1|2}(t_1), S_{3|2}(t_3))S_2(t_2), \tag{2.12}$$

$$S_{12\cdot3}(t_1, t_2, t_3) = C_{12|3}(S_{1|3}(t_1), S_{2|3}(t_2))S_3(t_3). \tag{2.13}$$

And for $t_k = 0$, where $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k \neq i$ the copulas should satisfy

$$C_{ij|k}(u, v) = C_{ij}(u, v) \tag{2.14}$$

to preserve its bivariate margins in the joint survival function. And C_{ij} is the copula corresponding to the joint survival function of (T_i, T_j) with $u = S_i(t_i)$ and $v = S_j(t_j)$.

Based on above considerations, if the $C_{ij|k}$ is replaced by C_{jk} (the idea was first proposed by Chakak in 1993 [2]), the corresponding tri-variate survival functions can be expressed as:

$$\begin{aligned} S_{23\cdot1}(t_1, t_2, t_3) &= C_{23|1}\left(\frac{C_{12}(S_1(t_1), S_2(t_2))}{S_1(t_1)}, \frac{C_{13}(S_1(t_1), S_3(t_3))}{S_1(t_1)}\right) S_1(t_1) \\ &= C_{23}\left(\frac{C_{12}(S_1(t_1), S_2(t_2))}{S_1(t_1)}, \frac{C_{13}(S_1(t_1), S_3(t_3))}{S_1(t_1)}\right) S_1(t_1) \end{aligned} \tag{2.15}$$

$$\begin{aligned} S_{13\cdot2}(t_1, t_2, t_3) &= C_{13|2}\left(\frac{C_{12}(S_1(t_1), S_2(t_2))}{S_2(t_2)}, \frac{C_{23}(S_2(t_2), S_3(t_3))}{S_2(t_2)}\right) S_2(t_2) \\ &= C_{13}\left(\frac{C_{12}(S_1(t_1), S_2(t_2))}{S_2(t_2)}, \frac{C_{23}(S_2(t_2), S_3(t_3))}{S_2(t_2)}\right) S_2(t_2) \end{aligned} \tag{2.16}$$

$$
\begin{aligned}
S_{12\cdot 3}(t_1,t_2,t_3) &= C_{12|3}\left(\frac{C_{13}(S_1(t_1),S_3(t_3))}{S_3(t_3)},\frac{C_{23}(S_2(t_2),S_3(t_3))}{S_3(t_3)}\right)S_3(t_3)\\
&= C_{12}\left(\frac{C_{13}(S_1(t_1),S_3(t_3))}{S_3(t_3)},\frac{C_{23}(S_2(t_2),S_3(t_3))}{S_3(t_3)}\right)S_3(t_3)
\end{aligned}
\tag{2.17}
$$

2.2.1 Three-dimensional structures

Based on the discussion above, the three-dimensional structures have the following expressions:

$$
\begin{aligned}
S_{23\cdot 1}(t_1,t_2,t_3) &= C_{23|1}\left(\frac{S_{12}(t_1,t_2)}{S_1(t_1)},\frac{S_{13}(t_1,t_3)}{S_1(t_1)}\right)S_1(t_1)\\
&= C_I\left(S_1(t_1),S_2(t_2),S_3(t_3)\right)
\end{aligned}
\tag{2.18}
$$

$$
\begin{aligned}
S_{13\cdot 2}(t_1,t_2,t_3) &= C_{13|2}\left(\frac{S_{12}(t_1,t_2)}{S_2(t_2)},\frac{S_{23}(t_2,t_3)}{S_2(t_2)}\right)S_2(t_2)\\
&= C_{II}\left(S_1(t_1),S_2(t_2),S_3(t_3)\right)
\end{aligned}
\tag{2.19}
$$

$$
\begin{aligned}
S_{12\cdot 3}(t_1,t_2,t_3) &= C_{12|3}\left(\frac{S_{13}(t_1,t_3)}{S_3(t_3)},\frac{S_{23}(t_2,t_3)}{S_3(t_3)}\right)S_3(t_3)\\
&= C_{III}\left(S_1(t_1),S_2(t_2),S_3(t_3)\right)
\end{aligned}
\tag{2.20}
$$

It is easy to check that all the three models above allow the bivariate margins to have different parameter values (not exchangeable). These structures (see Figure 2.1) can provide arbitrary pair selections which give the maximize flexibility and better practicality. They can be easily used, extended and also stand in line with the pre-specified bivariate structures.

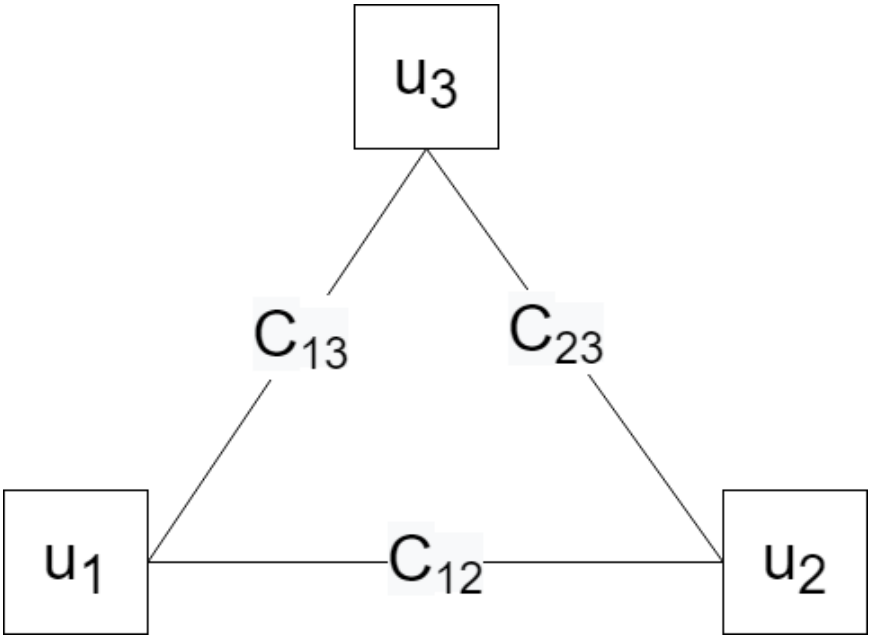

Figure 2.1 Three-dimensional structure for proposed method.

.

Take the $S_{12.3}$ as an example, the checking procedure is as below:

1. When $t_1 = 0$, then

$$\begin{aligned} S_{12\cdot 3}(0, t_2, t_3) &= C_{12|3}\left(\frac{S_{13}(0, t_3)}{S_3(t_3)}, \frac{S_{23}(t_2, t_3)}{S_3(t_3)}\right) S_3(t_3) \\ &= C_{12|3}\left(1, \frac{S_{23}(t_2, t_3)}{S_3(t_3)}\right) S_3(t_3) \\ &= \frac{S_{23}(t_2, t_3)}{S_3(t_3)} S_3(t_3) \\ &= S_{23}(t_2, t_3). \end{aligned} \tag{2.21}$$

The result $S_{23}(t_2, t_3)$ is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

2. When $t_2 = 0$, then

$$\begin{aligned} S_{12\cdot 3}(t_1, 0, t_3) &= C_{12|3}\left(\frac{S_{13}(t_1, t_3)}{S_3(t_3)}, \frac{S_{23}(0, t_3)}{S_3(t_3)}\right) S_3(t_3) \\ &= C_{12|3}\left(\frac{S_{13}(t_1, t_3)}{S_3(t_3)}, 1\right) S_3(t_3) \\ &= \frac{S_{13}(t_1, t_3)}{S_3(t_3)} S_3(t_3) \\ &= S_{13}(t_1, t_3). \end{aligned} \tag{2.22}$$

The result $S_{13}(t_1, t_3)$ is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

3. When $t_3 = 0$, then

$$\begin{aligned} S_{12\cdot 3}(t_1, t_2, 0) &= C_{12|3}(S_{13}(t_1, 0), S_{23}(t_2, 0)) \\ &= C_{12|3}(S_1(t_1), S_2(t_2)) \\ &= S_{12}(t_1, t_2). \end{aligned} \tag{2.23}$$

The result $S_{12}(t_1, t_2)$ is also the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

The other two models $S_{23\cdot1}$ and $S_{13\cdot2}$ have the similar properties. The checking procedures are omitted here.

2.2.2 Four-dimensional structures

Generalization to four-dimensions can be performed as below

$$S_{123\cdot4}(t_1, t_2, t_3, t_4) = P(T_1 > t_1, T_2 > t_2, T_3 > t_3 | T_4 > t_4) S_4(t_4) \tag{2.24}$$

By Sklar's theorem [34], then

$$\begin{aligned} &S_{123\cdot4}(t_1, t_2, t_3, t_4) \\ &= C_{123|4}\left(P(T_1 > t_1 | T_4 > t_4), P(T_2 > t_2 | T_4 > t_4), P(T_3 > t_3 | T_4 > t_4)\right) S_4(t_4) \end{aligned} \tag{2.25}$$

To preserves the pre-specified bivariate marginal distribution, the copula $C_{123|4}$ is required to equal to C_I or C_{II} or C_{III} from the three-dimensional structures from

Section 2.2.1. And for $i \in \{1, 2, 3\}$,

$$\begin{aligned} P(T_i > t_i | T_4 > t_4) &= \frac{P(T_i > t_i, T_4 > t_4)}{P(T_4 > t_4)} \\ &= \frac{S_{i4}(t_i, t_4)}{S_4(t_4)}. \end{aligned} \tag{2.26}$$

Then

$$\begin{aligned} &S_{123\cdot4}(t_1, t_2, t_3, t_4) \\ &= C_{123|4}\left(\frac{C_{14}(S_1(t_1), S_4(t_4))}{S_4(t_4)}, \frac{C_{24}(S_2(t_2), S_4(t_4))}{S_4(t_4)}, \frac{C_{34}(S_3(t_3), S_4(t_4))}{S_4(t_4)}\right) S_4(t_4) \end{aligned} \tag{2.27}$$

Similar results of copula $S_{234\cdot1}$, $S_{134\cdot2}$ and $S_{124\cdot3}$ can be derived on condition $T_1 > t_1$, $T_2 > t_2$ and on condition $T_3 > t_3$ respectively. Then for any (t_1, t_2, t_3, t_4) in the support of joint survival function, the following joint survival functions can be derived

$$S_{234\cdot1}(t_1, t_2, t_3, t_4) = C_{123|1}\left(\frac{S_{12}(t_1, t_2)}{S_1(t_1)}, \frac{S_{13}(t_1, t_3)}{S_1(t_1)}, \frac{S_{14}(t_1, t_4)}{S_1(t_1)}\right) S_1(t_1) \tag{2.28}$$

$$S_{134\cdot 2}(t_1,t_2,t_3,t_4) = C_{134|2}\left(\frac{S_{12}(t_1,t_2)}{S_2(t_2)}, \frac{S_{23}(t_2,t_3)}{S_2(t_2)}, \frac{S_{24}(t_2,t_4)}{S_2(t_2)}\right) S_2(t_2) \tag{2.29}$$

$$S_{124\cdot 3}(t_1,t_2,t_3,t_4) = C_{124|3}\left(\frac{S_{13}(t_1,t_3)}{S_3(t_3)}, \frac{S_{23}(t_2,t_3)}{S_3(t_3)}, \frac{S_{34}(t_3,t_4)}{S_3(t_3)}\right) S_3(t_3) \tag{2.30}$$

$$S_{123\cdot 4}(t_1,t_2,t_3,t_4) = C_{123|4}\left(\frac{S_{14}(t_1,t_4)}{S_4(t_4)}, \frac{S_{24}(t_2,t_4)}{S_4(t_4)}, \frac{S_{34}(t_3,t_4)}{S_4(t_4)}\right) S_4(t_4) \tag{2.31}$$

These structures (see Figure 2.2) can be easily proved that they stand in line with the three-dimensional structures from Section 2.2.1 and the pre-specified bivariate structures.

Take the $S_{123\cdot 4}$ as an example, the checking procedure is as below:

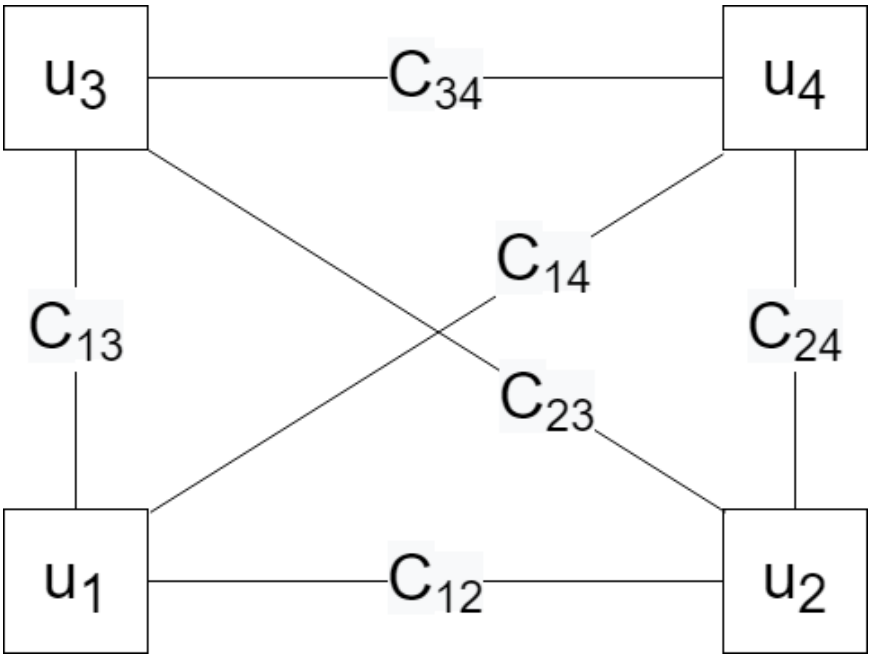

Figure 2.2 Four-dimensional structure for proposed method.

.

1. When $t_1 = 0$, then

$$\frac{S_{14}(t_1, t_4)}{S_4(t_4)} = \frac{S_{14}(0, t_4)}{S_4(t_4)} = \frac{S_4(t_4)}{S_4(t_4)} = 1. \tag{2.32}$$

Hence,

$$\begin{aligned} S_{123\cdot4}(0, t_2, t_3, t_4) &= C_{123|4}\left(1, \frac{S_{24}(t_2, t_4)}{S_4(t_4)}, \frac{S_{34}(t_3, t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= C_{23|4}\left(\frac{S_{24}(t_2, t_4)}{S_4(t_4)}, \frac{S_{34}(t_3, t_4)}{S_4(t_4)}, \right) S_4(t_4) \\ &= S_{23\cdot4}(t_2, t_3, t_4) \end{aligned} \tag{2.33}$$

which is the three-dimensional structure from Section 2.2.1 by applying Lemma

2.2.2.

2. When $t_2 = 0$, then

$$\frac{S_{24}(t_2, t_4)}{S_4(t_4)} = \frac{S_{24}(0, t_4)}{S_4(t_4)} = \frac{S_4(t_4)}{S_4(t_4)} = 1. \tag{2.34}$$

Hence,

$$\begin{aligned} S_{123\cdot4}(t_1, 0, t_3, t_4) &= C_{123|4}\left(\frac{S_{14}(t_1, t_4)}{S_4(t_4)}, 1, \frac{S_{34}(t_3, t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= C_{13|4}\left(\frac{S_{14}(t_1, t_4)}{S_4(t_4)}, \frac{S_{34}(t_3, t_4)}{S_4(t_4)}, \right) S_4(t_4) \\ &= S_{13\cdot4}(t_1, t_3, t_4) \end{aligned} \tag{2.35}$$

which is the three-dimensional structure from Section 2.2.1 by applying Lemma 2.2.2.

3. When $t_3 = 0$, then

$$\frac{S_{34}(t_3, t_4)}{S_4(t_4)} = \frac{S_{34}(0, t_4)}{S_4(t_4)} = \frac{S_4(t_4)}{S_4(t_4)} = 1. \tag{2.36}$$

Hence,

$$\begin{aligned} S_{123\cdot4}(t_1,t_2,0,t_4) &= C_{123|4}\left(\frac{S_{14}(t_1,t_4)}{S_4(t_4)},\frac{S_{24}(t_2,t_4)}{S_4(t_4)},1\right)S_4(t_4) \\ &= C_{12|4}\left(\frac{S_{14}(t_1,t_4)}{S_4(t_4)},\frac{S_{24}(t_2,t_4)}{S_4(t_4)},\right)S_4(t_4) \\ &= S_{12\cdot4}(t_1,t_2,t_4) \end{aligned} \tag{2.37}$$

which is the three-dimensional structure from Section 2.2.1 by applying Lemma 2.2.2.

4. When $t_4 = 0$, then

$$S_{123\cdot4}(t_1,t_2,t_3,0) = C_{123}\left(S_1(t_1),S_2(t_2),S_3(t_3)\right) \tag{2.38}$$

where $C_{123} = C_I$ or C_{II} or C_{III} which are the three-dimensional structures from Section 2.2.1.

5. When $t_1 = 0$, $t_2 = 0$, $t_3 \neq 0$, $t_4 \neq 0$, then

$$\begin{aligned} S_{123\cdot4}(0,0,t_3,t_4) &= C_{123|4}\left(1,1,\frac{S_{34}(t_3,t_4)}{S_4(t_4)}\right)S_4(t_4) \\ &= \frac{S_{34}(t_3,t_4)}{S_4(t_4)}S_4(t_4) \\ &= S_{34}(t_3,t_4) \end{aligned} \tag{2.39}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

6. When $t_1 = 0$, $t_2 \neq 0$, $t_3 = 0$, $t_4 \neq 0$, then

$$\begin{aligned} S_{123\cdot 4}(0, t_2, 0, t_4) &= C_{123|4}\left(1, \frac{S_{24}(t_2, t_4)}{S_4(t_4)}, 1\right) S_4(t_4) \\ &= \frac{S_{24}(t_2, t_4)}{S_4(t_4)} S_4(t_4) \\ &= S_{24}(t_2, t_4) \end{aligned} \tag{2.40}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

7. When $t_1 \neq 0$, $t_2 = 0$, $t_3 = 0$, $t_4 \neq 0$, then

$$\begin{aligned} S_{123\cdot 4}(t_1, 0, 0, t_4) &= C_{123|4}\left(\frac{S_{14}(t_1, t_4)}{S_4(t_4)}, 1, 1\right) S_4(t_4) \\ &= \frac{S_{14}(t_1, t_4)}{S_4(t_4)} S_4(t_4) \\ &= S_{14}(t_1, t_4) \end{aligned} \tag{2.41}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

8. When $t_1 = 0$, $t_2 \neq 0$, $t_3 \neq 0$, $t_4 = 0$, then

$$\begin{aligned} S_{123\cdot 4}(0, t_2, t_3, 0) &= C_{123|4}\left(1, \frac{S_{24}(t_2, t_4)}{S_4(0)}, \frac{S_{34}(t_3, t_4)}{S_4(0)}\right) S_4(0) \\ &= C_{23}\left(S_2(t_2), S_3(t_3)\right) \\ &= S_{23}(t_2, t_3) \end{aligned} \tag{2.42}$$

which is the pre-specified bivariate marginal distribution by applying Lemma

2.2.2.

9. When $t_1 \neq 0$, $t_2 = 0$, $t_3 \neq 0$, $t_4 = 0$, then

$$\begin{aligned} S_{123\cdot 4}(t_1, 0, t_3, 0) &= C_{123|4}\left(\frac{S_{14}(t_1, t_4)}{S_4(0)}, 1, \frac{S_{34}(t_3, t_4)}{S_4(0)}\right) S_4(0) \\ &= C_{13}\left(S_1(t_1), S_3(t_3)\right) \\ &= S_{13}(t_1, t_3) \end{aligned} \tag{2.43}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

10. When $t_1 \neq 0$, $t_2 \neq 0$, $t_3 = 0$, $t_4 = 0$, then

$$\begin{aligned} S_{123\cdot 4}(t_1, t_2, 0, 0) &= C_{123|4}\left(\frac{S_{14}(t_1, t_4)}{S_4(0)}, \frac{S_{24}(t_2, t_4)}{S_4(0)}, 1\right) S_4(0) \\ &= C_{12}\left(S_1(t_1), S_2(t_2)\right) \\ &= S_{12}(t_1, t_2) \end{aligned} \tag{2.44}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

The other three models $S_{234\cdot 1}$, $S_{134\cdot 2}$ and $S_{124\cdot 3}$ have the similar properties. The checking procedures are omitted here.

2.2.3 $d-$dimensional structures

Now the goal is to extend the above model construction approach to any $d-$dimensional data ($d \geqslant 3$) so that any bivariate margins are pre-specified and the pairwise dependence can be distinct for any $i \neq i'$ and $j \neq j'$. The idea is simple. Suppose

that any $d-$dimensional structure as below

$$
\begin{aligned}
&S_{12\cdots(d-1)\cdot d}(t_1, t_2, \cdots, t_{d-1}, t_d) \\
&= C_{12\cdots(d-1)|d}\left(\frac{S_{1d}(t_1, t_d)}{S_d(t_d)}, \frac{S_{2d}(t_2, t_d)}{S_d(t_d)}, \cdots, \frac{S_{(d-1)d}(t_{d-1}, t_d)}{S_d(t_d)}\right) S_d(t_d)
\end{aligned} \tag{2.45}
$$

satisfying the requirements that the survival function of any bivariate margin of the $d-$dimensional (and also the model of lower dimensions constructed in the same way) is a pre-specified bivariate copula model with parameter θ_{ij} for $1 \leqslant i < j \leqslant d$ where θ_{ij} are not necessarily the same for different subscripts (i, j) (this fact has just been proved for $d = 3$). Then the $(d+1)-$dimensional structure can be constructed as:

$$
\begin{aligned}
&S_{12\cdots d\cdot(d+1)}(t_1, t_2, \cdots, t_{d-1}, t_d, t_{d+1}) \\
&= C_{12\cdots d|(d+1)}\left(\frac{S_{1(d+1)}(t_1, t_{d+1})}{S_{d+1}(t_{d+1})}, \frac{S_{2(d+1)}(t_2, t_{d+1})}{S_{d+1}(t_{d+1})}, \cdots,\right. \\
&\qquad \left.\frac{S_{(d-1)(d+1)}(t_{d-1}, t_{d+1})}{S_{d+1}(t_{d+1})}, \frac{S_{d(d+1)}(t_d, t_{d+1})}{S_{d+1}(t_{d+1})}\right) S_{d+1}(t_{d+1})
\end{aligned} \tag{2.46}
$$

This structure can be easily proved that it stands in line with the $d-$dimensional structure and the bivariate structures. The checking procedure is as below:

1. When $t_i = 0, i \in \{1, 2, \cdots, d\}$, then the $(d+1)-$dimensional structure is reduced to

$$
\begin{aligned}
&S_{12\cdots d\cdot(d+1)}(t_1, \cdots, t_{i-1}, 0, t_{i+1}, \cdots, t_{d-1}, t_d, t_{d+1}) \\
&= C_{12\cdots(i-1)i(i+1)\cdots d|(d+1)}\left(\frac{S_{1(d+1)}(t_1, t_{d+1})}{S_{d+1}(t_{d+1})}, \cdots, \frac{S_{(i-1)(d+1)}(t_{i-1}, t_{d+1})}{S_{d+1}(t_{d+1})}, 1,\right. \\
&\qquad\qquad \frac{S_{(i+1)(d+1)}(t_{i+1}, t_{d+1})}{S_{d+1}(t_{d+1})}, \cdots, \frac{S_{(d-1)(d+1)}(t_{d-1}, t_{d+1})}{S_{d+1}(t_{d+1})}, \\
&\qquad\qquad \left.\frac{S_{d(d+1)}(t_d, t_{d+1})}{S_{d+1}(t_{d+1})}\right) S_{d+1}(t_{d+1}) \\
&= C_{12\cdots(i-1)(i+1)\cdots d|(d+1)}\left(\frac{S_{1(d+1)}(t_1, t_{d+1})}{S_{d+1}(t_{d+1})}, \cdots, \frac{S_{(i-1)(d+1)}(t_{i-1}, t_{d+1})}{S_{d+1}(t_{d+1})},\right. \\
&\qquad\qquad \frac{S_{(i+1)(d+1)}(t_{i+1}, t_{d+1})}{S_{d+1}(t_{d+1})}, \cdots, \frac{S_{(d-1)(d+1)}(t_{d-1}, t_{d+1})}{S_{d+1}(t_{d+1})}, \\
&\qquad\qquad \left.\frac{S_{d(d+1)}(t_d, t_{d+1})}{S_{d+1}(t_{d+1})}\right) S_{d+1}(t_{d+1}) \\
&= S_{12\cdots(i-1)(i+1)\cdots d\cdot(d+1)}(t_1, \cdots, t_{i-1}, t_{i+1}, \cdots, t_{d+1})
\end{aligned}
\tag{2.47}
$$

   which is the $d-$dimensional structure by applying Lemma 2.2.2. Here the $C_{12\cdots(i-1)(i+1)\cdots d|(d+1)}$ is the $(d-1)-$dimensional structure satisfying the desired nonexchangeable property and $S_{12\cdots(i-1)(i+1)\cdots d\cdot(d+1)}$ is the $d-$dimensional structure built based on it, which also satisfies the desired nonexchangeable property as previous assumed.

2. Similarly, when $t_{d+1} = 0$, it is easy to show that the $(d+1)-$dimensional structure is also reduced to a $d-$dimensional structure satisfying the desired nonexchangeable property.

$$
S_{12\cdots d\cdot(d+1)}(t_1, \cdots, t_d, 0) = C_{12\cdots d}\left(S_1(t_1), \cdots, S_d(t_d)\right) \tag{2.48}
$$

   where $C_{12\cdots d}$ is the $d-$dimensional structure satisfying the desired nonexchangeable property.

3. The $(d+1)-$dimensional model stands in line with any pre-specified bivariate margins can be proved as follows: when $t_j \neq 0$, $t_{d+1} \neq 0$ and every other $t_i = 0, (i, j \in \{1, 2, \cdots, d\}, i \neq j)$, the $(d+1)-$dimensional structure is reduced to

$$\begin{aligned}
&S_{12\cdots d\cdot(d+1)}(0, \cdots, 0, t_j, 0, \cdots, 0, t_{d+1}) \\
&= C_{12\cdots d|(d+1)}\left(1, \cdots, 1, \frac{S_{j(d+1)}(t_j, t_{d+1})}{S_{d+1}(t_{d+1})}, 1, \cdots, 1\right) S_{d+1}(t_{d+1}) \\
&= \frac{S_{j(d+1)}(t_j, t_{d+1})}{S_{d+1}(t_{d+1})} S_{d+1}(t_{d+1}), \\
&= S_{j(d+1)}(t_j, t_{d+1})
\end{aligned} \tag{2.49}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

4. When $t_j \neq 0$, $t_k \neq 0$, every other $t_i = 0, (i, j, k \in \{1, 2, \cdots, d\}, i \neq j \neq k \neq i)$ and $t_{d+1} = 0$, the $(d+1)-$dimensional structure is reduced to

$$\begin{aligned}
&S_{12\cdots d|(d+1)}(0, \cdots, 0, t_j, 0, \cdots, 0, t_k, 0, \cdots, 0) \\
&= C_{12\cdots d|(d+1)}\left(1, \cdots, 1, S_j(t_j), 1, \cdots, 1, S_k(t_k), 1, \cdots, 1\right) \\
&= C_{jk}(S_j(t_j), S_k(t_k)) \\
&= S_{jk}(t_j, t_k)
\end{aligned} \tag{2.50}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

The feasibility of the structure to $d-$dimensions was proved by mathematical induction method.

Based on above discussions, the way to construct a $d-$dimensional model for $d \geqslant 3$ is that you can start from a tri-variate model and add one additional variable into the model to build a four-dimensional model according to the same method which

is used to get the $(d+1)-$dimensional model from a $d-$dimensional model, then you can build four such models based on different variables for four-dimensional structure. Continuing this way, then a five-dimensional model can be builded, $\cdots$, and so on.

The primary advantage of this high dimensional dependence structure is that it is purely generated from the formulas used to define bivariate survival functions with the specified copula structure. Secondly, it allows for non-identical levels of associations (actually can be totally different) for different pairs of random variables which is desirable in many practical situations. Thirdly, it stands in line with the existing vine copulas through pair-copula construction, which means the vine copulas through pair-copula construction will become a special case of it. Some examples will be demonstrated in the next section the third point mentioned above.

2.2.4 Flexibility for the proposed structures

Take the four-dimensional copulas structures illustrated in Figure 2.2 as examples, the structures satisfy every possible dependence combination, and you can increase or decrease the dependence anytime as needed.

Figure 2.3 illustrate one of the possibilities for the four-dimensional structure which result the same structure as four-dimensional pair copula with Canonical vine structure (see Figure 1.4).

Figure 2.4 illustrate another possibility for the four-dimensional structure which result the same structure as four-dimensional pair copula with D-vine structure (see Figure 1.5).

There are more possibilities exist, you can choose according to the practical use.

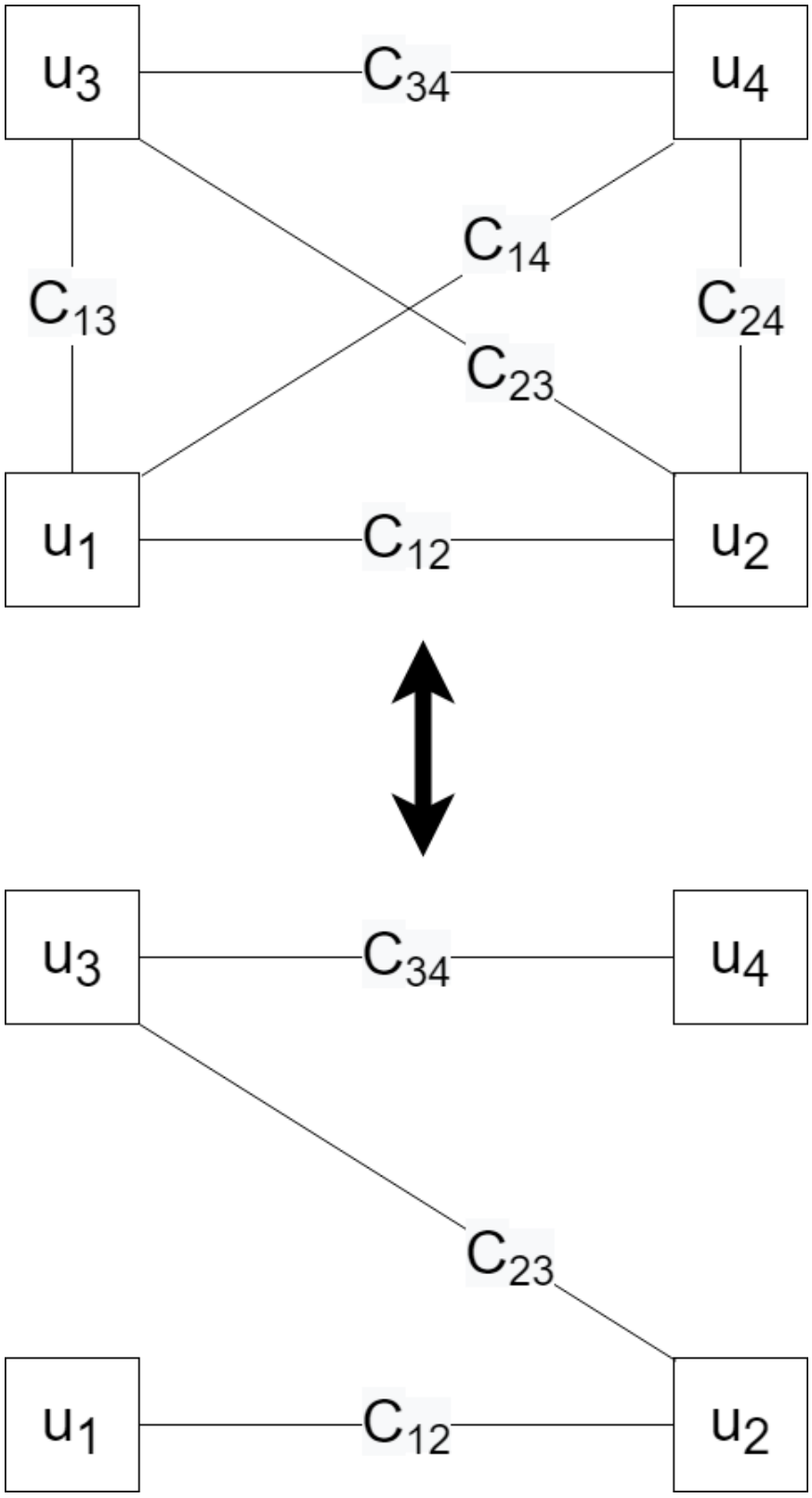

Figure 2.3 Flexibility of proposed method (Decrease the dependences which result the same structure as four-dimensional pair copula with Canonical vine).

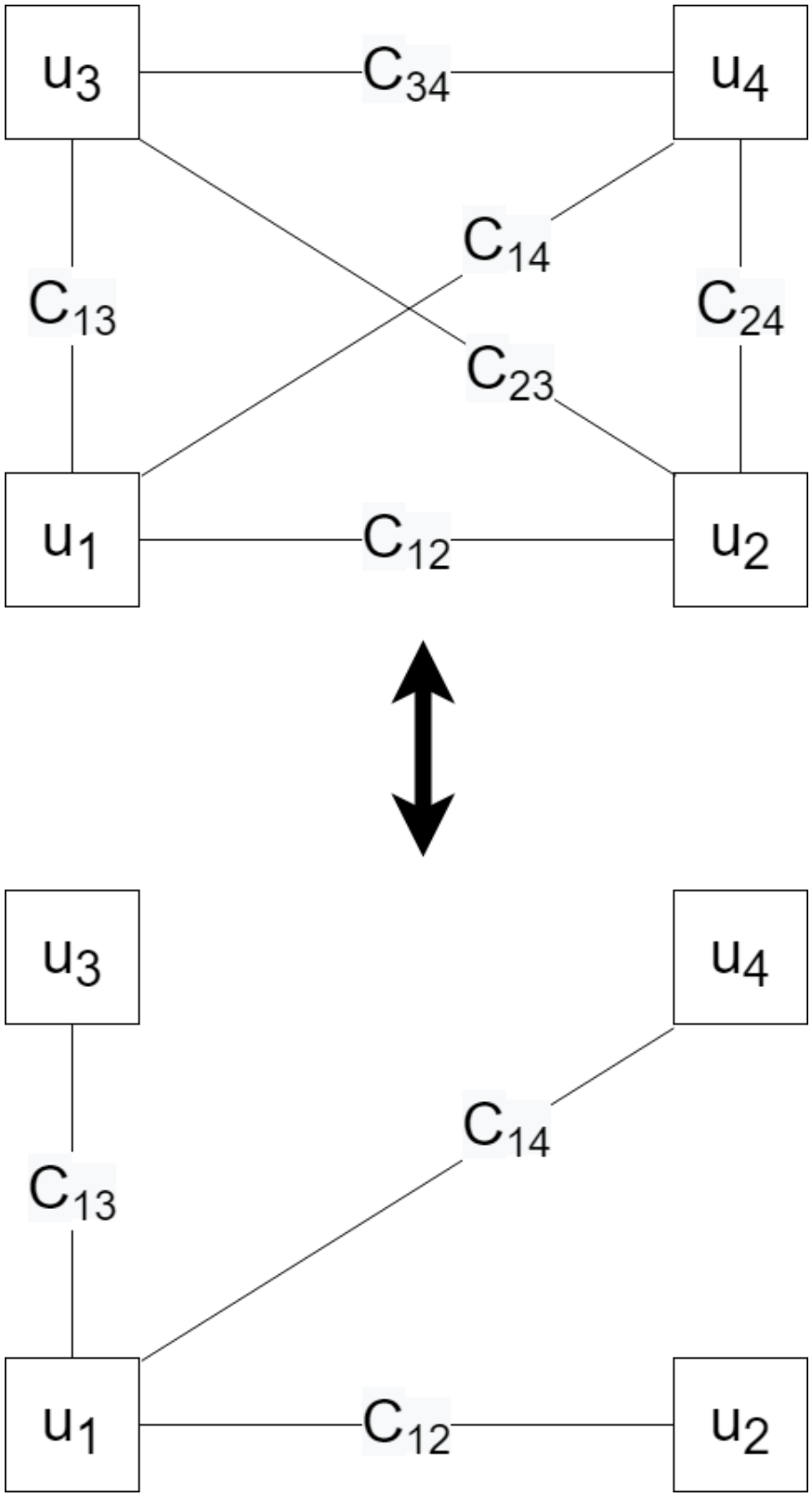

Figure 2.4 Flexibility of proposed method (Decrease the dependences which result the same structure as four-dimensional pair copula with D-vine).

2.3 Examples Based on Proposed Structures

In order to make a better understanding for the proposed methods and structures, some examples will be given here. The examples in this section will focus on the Clayton copulas [3], Gumbel - Hougaard copulas [13, 14] and Frank copulas [7]. And you are free to replace them with any other Archimedean copulas.

2.3.1 Clayton copulas

Clayton (1978) [3] introduced the model name after him described the transmission of coronary disease in human population form a father to his son based on following assumptions:

- It is straightforward to estimate the association parameter given censored observation on either or both the survival times for pairs of farther and his son;
- The effect of the parental history is expressible as a constant ratio of age specific rates;
- The model is symmetric with respect to the father and his son survival times.

The copula proposed by Clayton (1978) [3] has the expression as

$$C_{\psi}(t_1, t_2) = (S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1)^{-\frac{1}{\theta}} \tag{2.51}$$

where $\theta \in (0, \infty)$ with generator

$$\psi^{-1}(s) = (1 + s)^{-\frac{1}{\theta}}. \tag{2.52}$$

The $d-$dimensional Clayton copula has the expression as

$$C_\psi(t_1, t_2, \cdots, t_d) = (S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} + \cdots + S_d(t_d)^{-\theta} - d + 1)^{-\frac{1}{\theta}}. \tag{2.53}$$

The relationship between Kendall's tau and the Clayton copula parameter θ is given by:

$$\theta = \frac{2\tau}{1-\tau}. \tag{2.54}$$

2.3.2 Gumbel - Hougaard copula

Consider the Archimedean copula

$$C_\phi(t_1, t_2, \cdots, t_d) = \phi^{-1}(\phi(S_1(t_1)) + \phi(S_2(t_2)) + \cdots + \phi(S_d(t_d)), \tag{2.55}$$

with $(S_1(t_1), S_2(t_2), \cdots, S_d(t_d)) \in [0,1]^d$ and generator

$$\phi^{-1}(s) = \exp(-s^\beta) \tag{2.56}$$

where $\beta \in (0, 1)$.

The bivariate asymmetric Gumbel–Hougaard copula [13, 14] with expression

$$C(t_1, t_2) = \exp\left\{-[(-\log S_1(t_1))^\beta + (-\log S_2(t_2))^\beta]^{\frac{1}{\beta}}\right\}. \tag{2.57}$$

And the $d-$ demensional copula associated is

$$\begin{aligned} &C(t_1, t_2, \cdots, t_d) \\ &= \exp\left\{-[(-\log S_1(t_1))^\beta + (-\log S_2(t_2))^\beta + \cdots + (-\log S_d(t_d))^\beta]^{\frac{1}{\beta}}\right\} \end{aligned} \tag{2.58}$$

which known as the Gumbel–Hougaard copula [13, 14]. It derived from the work by Gumbel (1960) [13] and has been further considered by Hougaard (1986) [14]. For this reason, this copula is named as Gumbel-Hougaard copula. It was discovered independently in survival analysis.

The relationship between Kendall's tau and the Hougaard copula parameter β is given by:

$$\beta = 1 - \tau. \tag{2.59}$$

In the rest of this dissertation, Gumbel-Hougaard copula will be called as Hougaard copula for short.

2.3.3 Frank copula

Consider the Archimedean copula

$$C_\phi(t_1, t_2, \cdots, t_d) = \phi^{-1}(\phi(S_1(t_1)) + \phi(S_2(t_2)) + \cdots + \phi(S_d(t_d)), \tag{2.60}$$

with $(S_1(t_1), S_2(t_2), \cdots, S_d(t_d)) \in [0,1]^d$ and generator

$$\varphi^{-1}(s) = -\frac{1}{\gamma}\log\big(1 + \exp(-s)(\exp(-\gamma) - 1)\big) \tag{2.61}$$

where $\gamma \neq 0$.

The bivariate Frank copula [7] has expression

$$C(t_1, t_2) = -\frac{1}{\gamma}\log\left(1 + \frac{(\exp(-\gamma S_1(t_1)) - 1)(\exp(-\gamma S_2(t_2)) - 1)}{\exp(-\gamma) - 1}\right). \tag{2.62}$$

The relationship between Kendall's tau and the Frank copula parameter γ is given by:

$$D(\gamma) = 1 + \frac{\gamma}{4}(1 - \tau) \tag{2.63}$$

where

$$D(\gamma) = \frac{1}{\gamma} \int_0^{\gamma} \frac{t}{e^t - 1} dt. \tag{2.64}$$

2.3.4 Three-dimensional structures based on Clayton copulas

Let the pre-specified bivariate marginal distribution S_{12}, S_{13}, S_{23} be Clayton copulas [3] and have Clayton parameter θ_{12}, θ_{13}, θ_{23} respectively. The associated parameters between t_1, t_2 and t_3 are demonstrated in Figure 2.5. Using $S_{ij} = S_{\theta_{ij}}, (i, j \in \{1, 2, 3\}, i \neq j)$ is to emphasize that S_{ij} has one parameter θ_{ij} based on Clayton copulas.

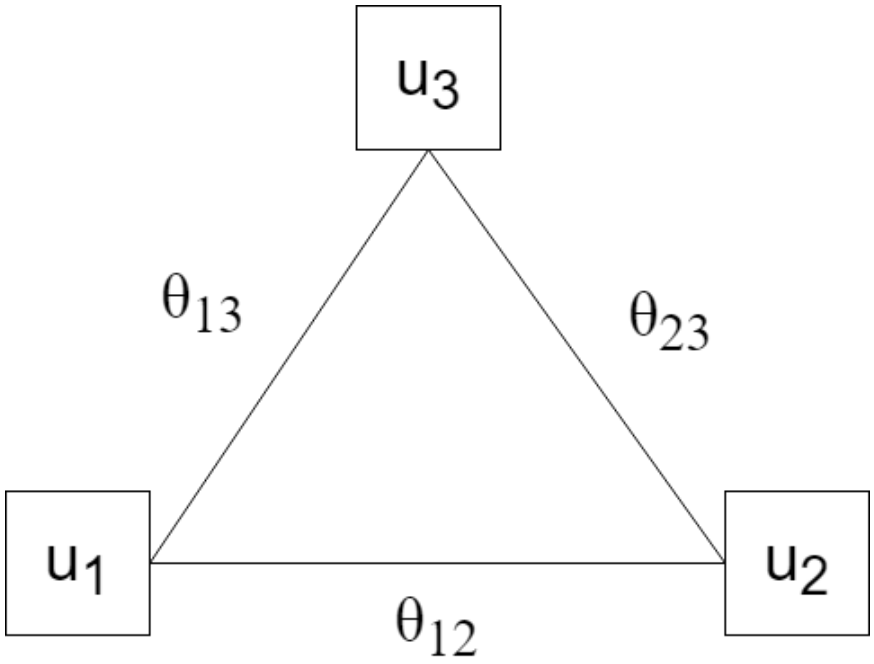

Figure 2.5 Three-dimensional structure based on Clayton copulas for proposed method.

.

Based on the method from Section 2.2, the three-dimensional structure based on Clayton copulas [3] under condition $T_3 > t_3$ is given below:

$$
\begin{aligned}
& S_{12\cdot 3}(t_1, t_2, t_3) \\
& = P(T_1 > t_1, T_2 > t_2 | T_3 > t_3) S_3(t_3) \\
& = C_{12|3}\left(\frac{S_{13}(t_1, t_3)}{S_3(t_3)}, \frac{S_{23}(t_2, t_3)}{S_3(t_3)}\right) S_3(t_3) \\
& = C_{12|3}\left(\frac{S_{\theta_{13}}(t_1, t_3)}{S_3(t_3)}, \frac{S_{\theta_{23}}(t_2, t_3)}{S_3(t_3)}\right) S_3(t_3) \\
& = \left(\left(\frac{S_{\theta_{13}}(t_1, t_3)}{S_3(t_3)}\right)^{-\theta_{12}} + \left(\frac{S_{\theta_{23}}(t_2, t_3)}{S_3(t_3)}\right)^{-\theta_{12}} - 1\right)^{-\frac{1}{\theta_{12}}} S_3(t_3) \\
& = \left(S_{\theta_{13}}(t_1, t_3)^{-\theta_{12}} + S_{\theta_{23}}(t_2, t_3)^{-\theta_{12}} - S_3(t_3)^{-\theta_{12}}\right)^{-\frac{1}{\theta_{12}}}
\end{aligned}
\tag{2.65}
$$

The other two three-dimensional structures based on Clayton copulas [3] under condition $T_1 > t_1$ or $T_2 > t_2$ with different parameters for bivariate copulas can also be derived from the same procedure. All three different three-dimensional structures are summarized below:

$$\begin{aligned} & S_{23\cdot 1}(t_1, t_2, t_3) \\ & = \left(S_{\theta_{12}}(t_1, t_2)^{-\theta_{23}} + S_{\theta_{13}}(t_1, t_3)^{-\theta_{23}} - S_1(t_1)^{-\theta_{23}} \right)^{-\frac{1}{\theta_{23}}} \\ & = C_I(S_1(t_1), S_2(t_2), S_3(t_3)) \end{aligned} \tag{2.66}$$

$$\begin{aligned} & S_{13\cdot 2}(t_1, t_2, t_3) \\ & = \left(S_{\theta_{12}}(t_1, t_2)^{-\theta_{13}} + S_{\theta_{23}}(t_2, t_3)^{-\theta_{13}} - S_2(t_2)^{-\theta_{13}} \right)^{-\frac{1}{\theta_{13}}} \\ & = C_{II}(S_1(t_1), S_2(t_2), S_3(t_3)) \end{aligned} \tag{2.67}$$

$$\begin{aligned} & S_{12\cdot 3}(t_1, t_2, t_3) \\ & = \left(S_{\theta_{13}}(t_1, t_3)^{-\theta_{12}} + S_{\theta_{23}}(t_2, t_3)^{-\theta_{12}} - S_3(t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}} \\ & = C_{III}(S_1(t_1), S_2(t_2), S_3(t_3)) \end{aligned} \tag{2.68}$$

If $\theta_{12} = \theta_{13} = \theta_{23} = \theta$, the above three different copulas are equal to each other and result a common expression. Details will be discussed below:

1. For $C_I(S_1(t_1), S_2(t_2), S_3(t_3))$, when $\theta_{12} = \theta_{13} = \theta_{23} = \theta$, the following expression can be derived:

$$
\begin{aligned}
& C_I(S_1(t_1), S_2(t_2), S_3(t_3)) \\
& = S_{23\cdot 1}(t_1, t_2, t_3) \\
& = \left(S_\theta(t_1, t_2)^{-\theta} + S_\theta(t_1, t_3)^{-\theta} - S_1(t_1)^{-\theta} \right)^{-\frac{1}{\theta}} \\
& = \Bigg(\left(S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1 \right)^{-\frac{1}{\theta} \times (-\theta)} + \left(S_1(t_1)^{-\theta} + S_3(t_3)^{-\theta} - 1 \right)^{-\frac{1}{\theta} \times (-\theta)} \\
& \qquad - S_1(t_1)^{-\theta} \Bigg)^{-\frac{1}{\theta}} \\
& = \left(\left(S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1 \right) + \left(S_1(t_1)^{-\theta} + S_3(t_3)^{-\theta} - 1 \right) - S_1(t_1)^{-\theta} \right)^{-\frac{1}{\theta}} \\
& = \left(S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} + S_3(t_3)^{-\theta} - 2 \right)^{-\frac{1}{\theta}};
\end{aligned}
\tag{2.69}
$$

2. For $C_{II}(S_1(t_1), S_2(t_2), S_3(t_3))$, when $\theta_{12} = \theta_{13} = \theta_{23} = \theta$, the following expression can be derived:

$$
\begin{aligned}
&C_{II}(S_1(t_1), S_2(t_2), S_3(t_3)) \\
&= S_{13\cdot 2}(t_1, t_2, t_3) \\
&= \Big(S_\theta(t_1, t_2)^{-\theta} + S_\theta(t_2, t_3)^{-\theta} - S_2(t_2)^{-\theta}\Big)^{-\frac{1}{\theta}} \\
&= \Bigg(\Big(S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1\Big)^{-\frac{1}{\theta}\times(-\theta)} + \Big(S_2(t_2)^{-\theta} + S_3(t_3)^{-\theta} - 1\Big)^{-\frac{1}{\theta}\times(-\theta)} \\
&\qquad - S_2(t_2)^{-\theta}\Bigg)^{-\frac{1}{\theta}} \\
&= \Big(\Big(S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1\Big) + \Big(S_2(t_2)^{-\theta} + S_3(t_3)^{-\theta} - 1\Big) - S_2(t_2)^{-\theta}\Big)^{-\frac{1}{\theta}} \\
&= \Big(S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} + S_3(t_3)^{-\theta} - 2\Big)^{-\frac{1}{\theta}};
\end{aligned}
\tag{2.70}
$$

3. For $C_{III}(S_1(t_1), S_2(t_2), S_3(t_3))$, when $\theta_{12} = \theta_{13} = \theta_{23} = \theta$, the following expression can be derived:

$$
\begin{aligned}
&C_{III}(S_1(t_1), S_2(t_2), S_3(t_3)) \\
&= S_{12\cdot 3}(t_1, t_2, t_3) \\
&= \left(S_\theta(t_1, t_3)^{-\theta} + S_\theta(t_2, t_3)^{-\theta} - S_3(t_3)^{-\theta} \right)^{-\frac{1}{\theta}} \\
&= \Bigg(\left(S_1(t_1)^{-\theta} + S_3(t_3)^{-\theta} - 1 \right)^{-\frac{1}{\theta}\times(-\theta)} + \left(S_2(t_2)^{-\theta} + S_3(t_3)^{-\theta} - 1 \right)^{-\frac{1}{\theta}\times(-\theta)} \\
&\qquad - S_3(t_3)^{-\theta} \Bigg)^{-\frac{1}{\theta}} \\
&= \left(\left(S_1(t_1)^{-\theta} + S_3(t_3)^{-\theta} - 1 \right) + \left(S_2(t_2)^{-\theta} + S_3(t_3)^{-\theta} - 1 \right) - S_3(t_3)^{-\theta} \right)^{-\frac{1}{\theta}} \\
&= \left(S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} + S_3(t_3)^{-\theta} - 2 \right)^{-\frac{1}{\theta}}.
\end{aligned}
\tag{2.71}
$$

So the final expression of three different copulas are equal to each other when $\theta_{12} = \theta_{13} = \theta_{23} = \theta$, which is an desired property for the proposed structures.

Furthermore, these structures can be easily proved that they stand in line with the pre-specified bivariate copulas. Take the $S_{12\cdot 3}$ as an example, the checking procedure is as below:

1. When $t_1 = 0$, then

$$\begin{aligned} S_{12\cdot 3}(0, t_2, t_3) &= \left(S_{\theta_{13}}(0, t_3)^{-\theta_{12}} + S_{\theta_{23}}(t_2, t_3)^{-\theta_{12}} - S_3(t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}} \\ &= \left(S_3(t_3)^{-\theta_{12}} + S_{\theta_{23}}(t_2, t_3)^{-\theta_{12}} - S_3(t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}} \\ &= \left(S_{\theta_{23}}(t_2, t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}} \\ &= S_{\theta_{23}}(t_2, t_3) \end{aligned} \tag{2.72}$$

   which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

2. When $t_2 = 0$, then

$$\begin{aligned} S_{12\cdot 3}(t_1, 0, t_3) &= \left(S_{\theta_{13}}(t_1, t_3)^{-\theta_{12}} + S_{\theta_{23}}(0, t_3)^{-\theta_{12}} - S_3(t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}} \\ &= \left(S_{\theta_{13}}(t_1, t_3)^{-\theta_{12}} + S_3(t_3)^{-\theta_{12}} - S_3(t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}} \\ &= \left(S_{\theta_{13}}(t_1, t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}} \\ &= S_{\theta_{13}}(t_1, t_3) \end{aligned} \tag{2.73}$$

   which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

3. When $t_3 = 0$, then

$$\begin{aligned} S_{12\cdot 3}(t_2, t_2, 0) &= \left(S_{\theta_{13}}(t_1, 0)^{-\theta_{12}} + S_{\theta_{23}}(t_2, 0)^{-\theta_{12}} - 1 \right)^{-\frac{1}{\theta_{12}}} \\ &= \left(S_1(t_1)^{-\theta_{12}} + S_2(t_2)^{-\theta_{12}} - 1 \right)^{-\frac{1}{\theta_{12}}} \\ &= S_{\theta_{12}}(t_1, t_2) \end{aligned} \tag{2.74}$$

which is the pre-specified bivariate marginal distribution by applying Lemma 2.2.2.

The other two models $S_{23\cdot1}$ and $S_{13\cdot2}$ have the similar properties. The checking procedures are omitted here.

2.3.5 Four-dimensional structures based on Clayton copulas

Let the pre-specified bivariate marginal distribution S_{12}, S_{13}, S_{14}, S_{23}, S_{24}, S_{34} be Clayton copulas [3] and have Clayton parameter θ_{12}, θ_{13}, θ_{14}, θ_{23}, θ_{24}, θ_{34} respectively. In order to be more clear, the associated parameters between t_1, t_2, t_3 and t_4 are demonstrated in Figure 2.6. Using $S_{ij} = S_{\theta_{ij}}, (i, j \in \{1, 2, 3\}, i \neq j)$ is to emphasize that S_{ij} has one parameter θ_{ij} based on Clayton copulas [3].

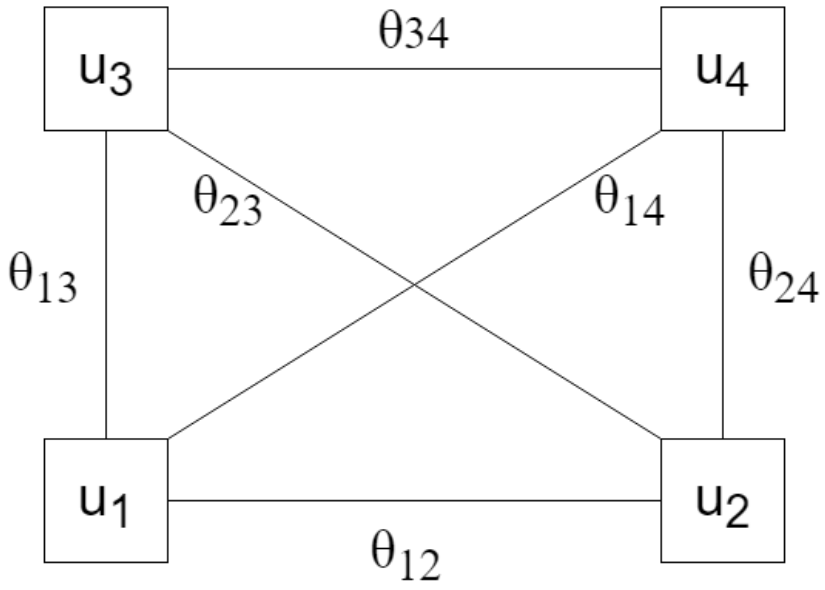

Figure 2.6 Four-dimensional structure based on Clayton copulas for proposed method.

.

Based on the method from Section 2.2, the four-dimensional structures based on Clayton copulas [3] are given below:

$$S_{234\cdot 1}(t_1,t_2,t_3,t_4) = C_{234|1}\left(\frac{S_{\theta_{12}}(t_1,t_2)}{S_1(t_1)}, \frac{S_{\theta_{13}}(t_1,t_3)}{S_1(t_1)}, \frac{S_{\theta_{14}}(t_1,t_4)}{S_1(t_1)}\right) S_1(t_1) \tag{2.75}$$

$$S_{134\cdot 2}(t_1,t_2,t_3,t_4) = C_{134|2}\left(\frac{S_{\theta_{12}}(t_1,t_2)}{S_2(t_2)}, \frac{S_{\theta_{23}}(t_2,t_3)}{S_2(t_2)}, \frac{S_{\theta_{24}}(t_2,t_4)}{S_2(t_2)}\right) S_2(t_2) \tag{2.76}$$

$$S_{124\cdot 3}(t_1,t_2,t_3,t_4) = C_{124|3}\left(\frac{S_{\theta_{13}}(t_1,t_3)}{S_3(t_3)}, \frac{S_{\theta_{23}}(t_2,t_3)}{S_3(t_3)}, \frac{S_{\theta_{34}}(t_3,t_4)}{S_3(t_3)}\right) S_3(t_3) \tag{2.77}$$

$$S_{123\cdot 4}(t_1,t_2,t_3,t_4) = C_{123|4}\left(\frac{S_{\theta_{14}}(t_1,t_4)}{S_4(t_4)}, \frac{S_{\theta_{24}}(t_2,t_4)}{S_4(t_4)}, \frac{S_{\theta_{34}}(t_3,t_4)}{S_4(t_4)}\right) S_4(t_4) \tag{2.78}$$

To preserves the pre-specified bivariate marginal distribution, the copulas $C_{234|1}$, $C_{134|2}$, $C_{124|3}$ and $C_{123|4}$ is required to equal to one of the three-dimensional structures from formula (2.66), (2.67) and (2.68). Take the copula $C_{123|4}$ as an example, it means:

$$
\begin{aligned}
C_{123|4} &= C_I(S_1(t_1), S_2(t_2), S_3(t_3)) \\
&= S_{23\cdot 1}(t_1, t_2, t_3) \\
&= \left(S_{\theta_{12}}(t_1, t_2)^{-\theta_{23}} + S_{\theta_{13}}(t_1, t_3)^{-\theta_{23}} - S_1(t_1)^{-\theta_{23}} \right)^{-\frac{1}{\theta_{23}}}
\end{aligned} \tag{2.79}
$$

or

$$
\begin{aligned}
C_{123|4} &= C_{II}(S_1(t_1), S_2(t_2), S_3(t_3)) \\
&= S_{13\cdot 2}(t_1, t_2, t_3) \\
&= \left(S_{\theta_{12}}(t_1, t_2)^{-\theta_{13}} + S_{\theta_{23}}(t_2, t_3)^{-\theta_{13}} - S_2(t_2)^{-\theta_{13}} \right)^{-\frac{1}{\theta_{13}}}
\end{aligned} \tag{2.80}
$$

or

$$
\begin{aligned}
C_{123|4} &= C_{III}(S_1(t_1), S_2(t_2), S_3(t_3)) \\
&= S_{12\cdot 3}(t_1, t_2, t_3) \\
&= \left(S_{\theta_{13}}(t_1, t_3)^{-\theta_{12}} + S_{\theta_{23}}(t_2, t_3)^{-\theta_{12}} - S_3(t_3)^{-\theta_{12}} \right)^{-\frac{1}{\theta_{12}}}
\end{aligned} \tag{2.81}
$$

These structures can be easily proved that they stand in line with the three-dimensional structures from formula (2.66), (2.67) and (2.68) and the pre-specified

bivariate structures. Take the $S_{123\cdot4}$ as an example, the checking procedure is as below:

1. When $t_1 = 0$, then

$$\begin{aligned} S_{123\cdot4}(0,t_2,t_3,t_4) &= C_{123|4}\left(1, \frac{S_{\theta_{24}}(t_2,t_4)}{S_4(t_4)}, \frac{S_{\theta_{34}}(t_3,t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= C_{23|4}\left(\frac{S_{\theta_{24}}(t_2,t_4)}{S_4(t_4)}, \frac{S_{\theta_{34}}(t_3,t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= S_{23\cdot4}(t_2,t_3,t_4) \end{aligned} \tag{2.82}$$

   which is same as the three-dimension structure by applying Lemma 2.2.2.

2. When $t_2 = 0$, then

$$\begin{aligned} S_{123\cdot4}(t_1,0,t_3,t_4) &= C_{123|4}\left(\frac{S_{\theta_{14}}(t_1,t_4)}{S_4(t_4)}, 1, \frac{S_{\theta_{34}}(t_3,t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= C_{13|4}\left(\frac{S_{\theta_{14}}(t_1,t_4)}{S_4(t_4)}, \frac{S_{\theta_{34}}(t_3,t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= S_{13\cdot4}(t_1,t_3,t_4) \end{aligned} \tag{2.83}$$

   which is same as the three-dimension structure by applying Lemma 2.2.2.

3. When $t_3 = 0$, then

$$\begin{aligned} S_{123\cdot4}(t_1,t_2,0,t_4) &= C_{123|4}\left(\frac{S_{\theta_{14}}(t_1,t_4)}{S_4(t_4)}, \frac{S_{\theta_{24}}(t_2,t_4)}{S_4(t_4)}, 1\right) S_4(t_4) \\ &= C_{12|4}\left(\frac{S_{\theta_{14}}(t_1,t_4)}{S_4(t_4)}, \frac{S_{\theta_{24}}(t_2,t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= S_{12\cdot4}(t_1,t_2,t_4) \end{aligned} \tag{2.84}$$

   which is same as the three-dimension structure by applying Lemma 2.2.2.

4. When $t_4 = 0$, then

$$S_{123\cdot 4}(t_1, t_2, t_3, 0) = C_{123}(S_1(t_1), S_2(t_2), S_3(t_3)) \tag{2.85}$$

where $C_{123} = C_I$ or $C_{123} = C_{II}$ or $C_{123} = C_{III}$ from the three-dimensional structures.

5. When $t_1 = 0, t_2 = 0, t_3 \neq 0, t_4 \neq 0$, then

$$\begin{aligned} S_{123\cdot 4}(0, 0, t_3, t_4) &= C_{123|4}\left(1, 1, \frac{S_{\theta_{34}}(t_3, t_4)}{S_4(t_4)}\right) S_4(t_4) \\ &= \frac{S_{\theta_{34}}(t_3, t_4)}{S_4(t_4)} S_4(t_4) \\ &= S_{\theta_{34}}(t_3, t_4) \\ &= (S_3(t_3)^{-\theta_{34}} + S_4(t_4)^{-\theta_{34}} - 1)^{-\frac{1}{\theta_{34}}} \end{aligned} \tag{2.86}$$

which is the pre-specified two-dimensional Clayton copula [3] by applying Lemma 2.2.2.

6. When $t_1 = 0$, $t_2 \neq 0$, $t_3 = 0$, $t_4 \neq 0$, then

$$\begin{aligned} S_{123\cdot 4}(0, t_2, 0, t_4) &= C_{123|4}\left(1, \frac{S_{\theta_{24}}(t_2, t_4)}{S_4(t_4)}, 1\right) S_4(t_4) \\ &= \frac{S_{\theta_{24}}(t_2, t_4)}{S_4(t_4)} S_4(t_4) \\ &= S_{\theta_{24}}(t_2, t_4) \\ &= (S_2(t_2)^{-\theta_{24}} + S_4(t_4)^{-\theta_{24}} - 1)^{-\frac{1}{\theta_{24}}} \end{aligned} \tag{2.87}$$

which is the pre-specified two-dimensional Clayton copula [3] by applying Lemma 2.2.2.

7. When $t_1 \neq 0$, $t_2 = 0$, $t_3 = 0$, $t_4 \neq 0$, then

$$\begin{aligned} S_{123\cdot 4}(t_1, 0, 0, t_4) &= C_{123|4}\left(\frac{S_{\theta_{14}}(t_1, t_4)}{S_4(t_4)}, 1, 1\right) S_4(t_4) \\ &= \frac{S_{\theta_{14}}(t_1, t_4)}{S_4(t_4)} S_4(t_4) \\ &= S_{\theta_{14}}(t_1, t_4) \\ &= (S_1(t_1)^{-\theta_{14}} + S_4(t_4)^{-\theta_{14}} - 1)^{-\frac{1}{\theta_{14}}} \end{aligned} \tag{2.88}$$

which is the pre-specified two-dimensional Clayton copula [3] by applying Lemma 2.2.2.

8. When $t_1 = 0$, $t_2 \neq 0$, $t_3 \neq 0$, $t_4 = 0$, then

$$\begin{aligned} S_{123\cdot 4}(0, t_2, t_3, 0) &= C_{123|4}\left(1, \frac{S_{\theta_{24}}(t_2, t_4)}{S_4(0)}, \frac{S_{\theta_{34}}(t_3, t_4)}{S_4(0)}\right) S_4(0) \\ &= C_{23}\left(S_2(t_2), S_3(t_3)\right) \\ &= S_{23}(t_2, t_3) \\ &= (S_2(t_2)^{-\theta_{23}} + S_3(t_3)^{-\theta_{23}} - 1)^{-\frac{1}{\theta_{23}}} \end{aligned} \tag{2.89}$$

which is the pre-specified two-dimensional Clayton copula [3] by applying Lemma 2.2.2.

9. When $t_1 \neq 0$, $t_2 = 0$, $t_3 \neq 0$, $t_4 = 0$, then

$$\begin{aligned} S_{123\cdot 4}(t_1, 0, t_3, 0) &= C_{123|4}\left(\frac{S_{\theta_{14}}(t_1, t_4)}{S_4(0)}, 1, \frac{S_{\theta_{34}}(t_3, t_4)}{S_4(0)}\right) S_4(0) \\ &= C_{13}\left(S_1(t_1), S_3(t_3)\right) \\ &= S_{13}(t_1, t_3) \\ &= (S_1(t_1)^{-\theta_{13}} + S_3(t_3)^{-\theta_{13}} - 1)^{-\frac{1}{\theta_{13}}} \end{aligned} \tag{2.90}$$

which is the pre-specified two-dimensional Clayton copula [3] by applying Lemma 2.2.2.

10. When $t_1 \neq 0$, $t_2 \neq 0$, $t_3 = 0$, $t_4 = 0$, then

$$\begin{aligned} S_{123\cdot 4}(0, t_2, t_3, 0) &= C_{123|4}\left(\frac{S_{\theta_{14}}(t_1, t_4)}{S_4(0)}, \frac{S_{\theta_{24}}(t_2, t_4)}{S_4(0)}, 1\right) S_4(0) \\ &= C_{12}\left(S_1(t_1), S_2(t_2)\right) \\ &= S_{12}(t_1, t_2) \\ &= \left(S_1(t_1)^{-\theta_{12}} + S_2(t_2)^{-\theta_{12}} - 1\right)^{-\frac{1}{\theta_{12}}} \end{aligned} \tag{2.91}$$

which is the pre-specified two-dimensional Clayton copula [3] by applying Lemma 2.2.2.

The other three models $S_{234\cdot 1}$, $S_{134\cdot 2}$ and $S_{124\cdot 3}$ have the similar properties. The checking procedures are omitted here.

2.3.6 Three-dimensional structures based on different copulas

The flexibility of the proposed method allows arbitrary selection of pairwise correlation. It is not only the different associated parameters between variables but also different copulas can be chosen. The examples in this section will focus on the Clayton copulas [3], Hougaard copulas [13, 14], and Frank copulas [7]. And you are free to replace them with any other Archimedean copulas.

An example of the three-dimensional structures based on proposed method which include two Clayton copulas [3] and a Hougaard copula [13, 14] is shown in Example 2.3.1 with Figure 2.7.

Example 2.3.1. ***One possible three-dimensional structure for Clayton + Hougaard:***

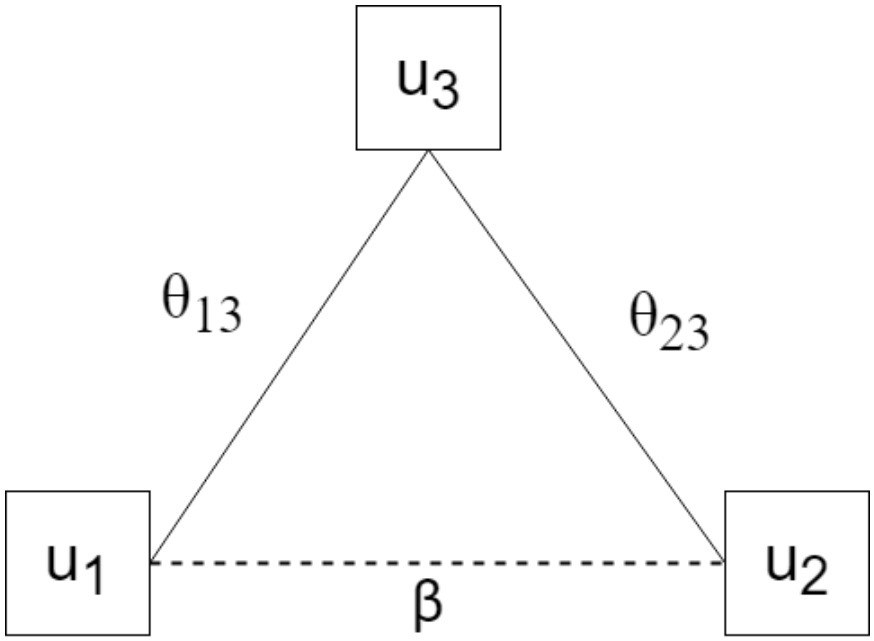

Figure 2.7 Flexibility of proposed method (three-dimensional structure: Clayton + Hougaard).

.

- (t_1, t_2) *is modeled by a Hougaard copula [13, 14] with parameter* β*, that is*

$$S_\beta(t_1, t_2) = \exp\left\{-[(-\log S_1(t_1))^\beta + (-\log S_2(t_2))^\beta]^{\frac{1}{\beta}}\right\}; \tag{2.92}$$

- (t_1, t_3) *is modeled by a Clayton copula [3] with parameter* θ_{13}*, that is*

$$S_{\theta_{13}}(t_1, t_3) = (S_1(t_1)^{-\theta_{13}} + S_3(t_3)^{-\theta_{13}} - 1)^{-\frac{1}{\theta_{13}}}; \tag{2.93}$$

- (t_2, t_3) *is modeled by a Clayton copula [3] with parameter* θ_{23}*, that is*

$$S_{\theta_{23}}(t_2, t_3) = (S_2(t_2)^{-\theta_{23}} + S_3(t_3)^{-\theta_{23}} - 1)^{-\frac{1}{\theta_{23}}}. \tag{2.94}$$

All the three-dimensional structures based on proposed structures for Figure 2.7 are given below:

$$
\begin{aligned}
&S_{23\cdot1}(t_1,t_2,t_3)\\
&=\Big(S_\beta(t_1,t_2)^{-\theta_{23}}+S_{\theta_{13}}(t_1,t_3)^{-\theta_{23}}-S_1(t_1)^{-\theta_{23}}\Big)^{-\frac{1}{\theta_{23}}}\\
&=\Bigg(\Bigg[\exp\Big\{-[(-\log S_1(t_1))^\beta+(-\log S_2(t_2))^\beta]^{\frac{1}{\beta}}\Big\}\Bigg]^{-\theta_{23}}\\
&\quad+\Bigg[(S_1(t_1)^{-\theta_{13}}+S_3(t_3)^{-\theta_{13}}-1)^{-\frac{1}{\theta_{13}}}\Bigg]^{-\theta_{23}}-S_1(t_1)^{-\theta_{23}}\Bigg)^{-\frac{1}{\theta_{23}}}
\end{aligned}
\tag{2.95}
$$

$$
\begin{aligned}
&S_{13\cdot2}(t_1,t_2,t_3)\\
&=\Big(S_\beta(t_1,t_2)^{-\theta_{13}}+S_{\theta_{23}}(t_2,t_3)^{-\theta_{13}}-S_2(t_2)^{-\theta_{13}}\Big)^{-\frac{1}{\theta_{13}}}\\
&=\Bigg(\Bigg[\exp\Big\{-[(-\log S_1(t_1))^\beta+(-\log S_2(t_2))^\beta]^{\frac{1}{\beta}}\Big\}\Bigg]^{-\theta_{13}}\\
&\quad+\Bigg[(S_2(t_2)^{-\theta_{23}}+S_3(t_3)^{-\theta_{23}}-1)^{-\frac{1}{\theta_{13}}}\Bigg]^{-\theta_{13}}-S_2(t_2)^{-\theta_{13}}\Bigg)^{-\frac{1}{\theta_{13}}}
\end{aligned}
\tag{2.96}
$$

$$
\begin{aligned}
&S_{12\cdot3}(t_1,t_2,t_3)\\
&=\exp\Big\{-[(-\log S_{\theta_{13}}(t_1,t_3))^\beta+(-\log S_{\theta_{23}}(t_2,t_3))^\beta]^{\frac{1}{\beta}}\Big\}\\
&=\exp\Bigg\{-[(-\log(S_1(t_1)^{-\theta_{13}}+S_3(t_3)^{-\theta_{13}}-1)^{-\frac{1}{\theta_{13}}})^\beta\\
&\quad+(-\log(S_2(t_2)^{-\theta_{23}}+S_3(t_3)^{-\theta_{23}}-1)^{-\frac{1}{\theta_{23}}})^\beta]^{\frac{1}{\beta}}\Bigg\}
\end{aligned}
\tag{2.97}
$$

Another example of the three-dimensional structures based on proposed method which include a Clayton copula [3], a Hougaard copula [13, 14] and a Frank copula [7] is shown in Example 2.3.2 with Figure 2.8.

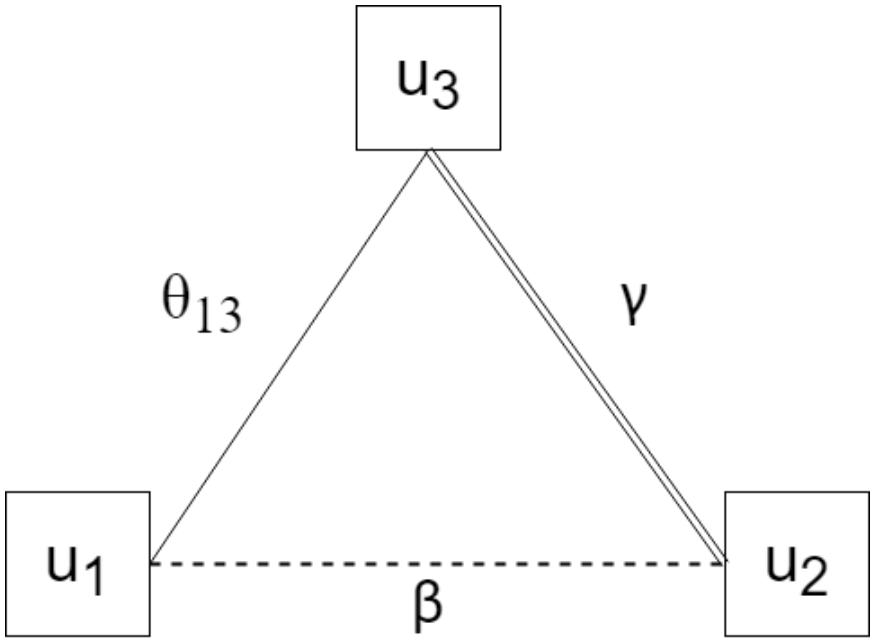

Figure 2.8 Flexibility of proposed method (three-dimensional structure: Clayton + Hougaard + Frank).

.

Example 2.3.2. ***One possible three-dimensional structure for Clayton + Hougaard + Frank:***

- (t_1, t_2) *is modeled by a Hougaard copula [13, 14] with parameter* β*, that is*

$$S_\beta(t_1, t_2) = \exp\left\{-[(-\log S_1(t_1))^\beta + (-\log S_2(t_2))^\beta]^{\frac{1}{\beta}}\right\}; \tag{2.98}$$

- (t_1, t_3) *is modeled by a Clayton copula [3] with parameter* θ_{13}*, that is*

$$S_{\theta_{13}}(t_1, t_3) = (S_1(t_1)^{-\theta_{13}} + S_3(t_3)^{-\theta_{13}} - 1)^{-\frac{1}{\theta_{13}}}; \tag{2.99}$$

- (t_2, t_3) *is modeled by a Frank copula [7] with parameter* γ*, that is*

$$S_\gamma(t_2, t_3) = -\frac{1}{\gamma}\log\left(1 + \frac{(\exp(-\gamma S_2(t_2)) - 1)(\exp(-\gamma S_3(t_3)) - 1)}{\exp(-\gamma) - 1}\right). \tag{2.100}$$

All the three-dimensional structures based on proposed structures for Figure 2.8 are given below:

$$\begin{aligned}
&S_{23\cdot1}(t_1, t_2, t_3) \\
&= -\frac{1}{\gamma}\log\left\{1 + \frac{[\exp(-\gamma S_\beta(t_1, t_2)) - 1][\exp(-\gamma S_{\theta_{13}}(t_1, t_3)) - 1]}{\exp(-\gamma) - 1}\right\} \\
&= -\frac{1}{\gamma}\log\left\{1 + \left[\exp\left(-\gamma\left[\exp\left\{-[(-\log S_1(t_1))^\beta + (-\log S_2(t_2))^\beta]^{\frac{1}{\beta}}\right\}\right]\right) - 1\right] \right. \\
&\quad \left.\left[\exp\left\{-\gamma\left[(S_1(t_1)^{-\theta_{13}} + S_3(t_3)^{-\theta_{13}} - 1)^{-\frac{1}{\theta_{13}}}\right]\right\} - 1\right][\exp(-\gamma) - 1]^{-1}\right\}
\end{aligned} \tag{2.101}$$

$$
\begin{aligned}
&S_{13\cdot 2}(t_1,t_2,t_3)\\
&=\left(S_\beta(t_1,t_2)^{-\theta_{13}}+S_\gamma(t_2,t_3)^{-\theta_{13}}-S_2(t_2)^{-\theta_{13}}\right)^{-\frac{1}{\theta_{13}}}\\
&=\Bigg(\left[\exp\left\{-[(-\log S_1(t_1))^\beta+(-\log S_2(t_2))^\beta]^{\frac{1}{\beta}}\right\}\right]^{-\theta_{13}}\\
&\quad+\left[-\frac{1}{\gamma}\log\left(1+\frac{(\exp(-\gamma S_2(t_2))-1)(\exp(-\gamma S_3(t_3))-1)}{\exp(-\gamma)-1}\right)\right]^{-\theta_{13}}-S_2(t_2)^{-\theta_{13}}\Bigg)^{-\frac{1}{\theta_{13}}}
\end{aligned}
\tag{2.102}
$$

$$
\begin{aligned}
&S_{12\cdot 3}(t_1,t_2,t_3)\\
&=\exp\left\{-\left[\left(-\log S_{\theta_{13}}(t_1,t_3)\right)^\beta+\left(-\log S_\gamma(t_2,t_3)\right)^\beta\right]^{\frac{1}{\beta}}\right\}\\
&=\exp\Bigg\{-\Bigg[\left(-\log(S_1(t_1)^{-\theta_{13}}+S_3(t_3)^{-\theta_{13}}-1)^{-\frac{1}{\theta_{13}}}\right)^\beta\\
&\quad+\left(-\log\left[-\frac{1}{\gamma}\log\left(1+\frac{(\exp(-\gamma S_2(t_2))-1)(\exp(-\gamma S_3(t_3))-1)}{\exp(-\gamma)-1}\right)\right]\right)^\beta\Bigg]^{\frac{1}{\beta}}\Bigg\}
\end{aligned}
\tag{2.103}
$$

These are two examples for reference. You are free to choose any other Archimedean copulas and construct the model based on the proposed method.

2.4 Survival Functions for Proposed Structures

By constructing the models using the method proposed in Section 2.2, Oakes's question has been answered in a satisfactory way. Now a class of flexible models with any pre-specified bivariate margins have been obtained. This class of models can model any non-normal data and have more complicated structures for more than two dimensions. The research goal in the next few sections is to explore more properties of this class of models and apply them to model high dimensional data more effectively.

The following discussion based on a basic assumption that the data can be modeled by the Archimedean copulas. The joint bivariate survival functions of (T_1, T_2), (T_1, T_3) and (T_2, T_3) follow Archimedean copulas (can be three different copulas). That is

$$S(t_1, t_2) = \psi_{\theta_{12}}^{-1}\left\{\psi_{\theta_{12}}[S_1(t_1)] + \psi_{\theta_{12}}[S_2(t_2)]\right\} \tag{2.104}$$

$$S(t_1, t_3) = \phi_{\theta_{13}}^{-1}\left\{\phi_{\theta_{13}}[S_1(t_1)] + \phi_{\theta_{13}}[S_3(t_3)]\right\} \tag{2.105}$$

and

$$S(t_2, t_3) = \varphi_{\theta_{23}}^{-1}\left\{\varphi_{\theta_{23}}[S_2(t_2)] + \varphi_{\theta_{23}}[S_3(t_3)]\right\} \tag{2.106}$$

where $S_1(t_1)$, $S_2(t_2)$ and $S_3(t_3)$ are marginal survival functions of T_1, T_2 and T_3 respectively. $\psi_{\theta_{12}}^{-1}$, $\phi_{\theta_{13}}^{-1}$ and $\varphi_{\theta_{23}}^{-1}$ are the inverse function of $\psi_{\theta_{12}}$, $\phi_{\theta_{13}}$ and $\varphi_{\theta_{23}}$ respectively. $\psi_{\theta_{12}}$, $\phi_{\theta_{13}}$ and $\varphi_{\theta_{23}}$ are defined as the copula generators [22]. θ_{12}, θ_{13} and θ_{23} are the unknown parameters.

Based on the property of Archimedean copula, T_1, T_2 and T_3 follow uniform distribution $U[0,1]$. That is

$$S_1(t_1) = 1 - t_1, \tag{2.107}$$

$$S_2(t_2) = 1 - t_2, \tag{2.108}$$

and

$$S_3(t_3) = 1 - t_3. \tag{2.109}$$

Theorem 2.4.1. *Let $i, j, k \in \{1, 2, 3\}$ and $i \neq j \neq k \neq i$, and T_i, T_j and T_k be uniform random variables. The joint survival function of (T_i, T_j, T_k) can be modeled*

by the proposed structure $S_{ij\cdot k}$*. That is*

$$\begin{aligned} S_{ij\cdot k}\left(S_i(t_i), S_j(t_j), S_k(t_k)\right) &= C_{ij|k}\left(S_{i|k}(t_i|t_k), S_{j|k}(t_j|t_k)\right) S_k(t_k) \\ &= q_{\theta_{ij}}^{-1}\left\{q_{\theta_{ij}}[S_{i|k}(t_i|t_k)] + q_{\theta_{ij}}[S_{j|k}(t_j|t_k)]\right\} S_k(t_k) \end{aligned}$$

q *is the copula generator. Let*

$$V = C(T_i, T_j, T_k) = S_{ij\cdot k}(S_i(T_i), S_j(T_j), S_k(T_k)),$$

Then V *has survival function on* $[0,1]$ *as*

$$S_k^*(v) = v + \int_0^{1-v} \int_0^1 \int_{L_k(t_j,t_k,v)}^{U_k(t_j,t_k,v)} s_{ij\cdot k}(t_i, t_j, t_k)\, dt_i\, dt_j\, dt_k,$$

with

$$U_k(t_j, t_k, v) = \min\left\{1, 0 \leqslant C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant \frac{v}{1-t_k}\right\},$$

$$L_k(t_j, t_k, v) = \max\left\{0, 0 \leqslant C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant \frac{v}{1-t_k}\right\},$$

and

$$s_{ij\cdot k}(t_i, t_j, t_k) = \frac{\partial^3}{\partial t_i \partial t_j \partial t_k} S_{ij\cdot k}(S_i(T_i), S_j(T_j), S_k(T_k)).$$

Proof. The survival function of V is

$$
\begin{aligned}
& S_k^*(v) \\
=& P[C(T_i, T_j, T_k) \leqslant v] \\
=& E\{P[C(T_i, T_j, T_k) \leqslant v \mid T_k = t_k]\} \\
=& E\{P[S_{ij\cdot k}(S_i(T_i), S_j(T_j), S_k(T_k)) \leqslant v \mid S_k(T_k) = S_k(t_k)]\} \\
=& E\{P[C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k))S_k(t_k) \leqslant v \mid S_k(T_k) = S_k(t_k)]\} \\
=& E\{\mathbf{1}[S_k(t_k) \leqslant v]P[C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant 1 \mid S_k(T_k) = S_k(t_k)]\}+ \\
& E\left\{\mathbf{1}[S_k(t_k) > v]P\left[C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant \frac{v}{S_k(t_k)} \;\middle|\; S_k(T_k) = S_k(t_k)\right]\right\} \\
=& E\{\mathbf{1}[1 - t_k \leqslant v] * 1 \mid S_k(T_k) = 1 - t_k)]\}+ \\
& E\left\{\mathbf{1}[1 - t_k > v]P\left[C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant \frac{v}{1 - t_k} \;\middle|\; S_k(T_k) = 1 - t_k\right]\right\} \\
=& \int_{1-v}^{1} 1\, dt_k + \int_0^{1-v} E\left\{\mathbf{1}\left[C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant \frac{v}{1 - t_k}\right]\right\} dt_k \\
=& v + \int_0^{1-v} \int_0^1 \int_{L_k(t_j,t_k,v)}^{U_k(t_j,t_k,v)} s_{ij\cdot k}(t_i, t_j, t_k)\, dt_i\, dt_j\, dt_k
\end{aligned}
\tag{2.110}
$$

with

$$
U_k(t_j, t_k, v) = \min\left\{1, 0 \leqslant C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant \frac{v}{1 - t_k}\right\}, \tag{2.111}
$$

$$L_k(t_j, t_k, v) = \max\left\{0, 0 \leqslant C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \leqslant \frac{v}{1-t_k}\right\}, \tag{2.112}$$

and

$$s_{ij\cdot k}(t_i, t_j, t_k) = \frac{\partial^3}{\partial t_i \partial t_j \partial t_k} S_{ij\cdot k}(S_i(T_i), S_j(T_j), S_k(T_k)). \tag{2.113}$$

□

Theorem 2.4.1 is the survival functions for the three-dimensional proposed structures. And you can use the same method to get the survival functions for any $d-$dimensional proposed structures. Details are ommited here.

Since there are many nice properties for the bivariate structures. It is easy to think of the application of these nice properties for the proposed structures. Can you use these nice bivariate properties directly to the proposed $d-$dimensional structures? The answer is no. Take the Kendall's procedure proposed by Genest and Rivest (1993) [10] for example. This approach can not apply for the proposed $d-$dimensional structures. The reason will be discussed below.

The discussion below based on a possible pitfall for the proposed three-dimensional structures. Let $i, j, k \in \{1, 2, 3\}$ with $i \neq j \neq k \neq i$, T_i, T_j and T_k are uniform random variables. The joint survival function of (T_i, T_j, T_k) can be modeled

by the proposed structure $S_{ij\cdot k}$.That is

$$\begin{aligned} S_{ij\cdot k}\left(S_i(t_i), S_j(t_j), S_k(t_k)\right) &= C_{ij|k}\left(S_{i|k}(t_i|t_k), S_{j|k}(t_j|t_k)\right) S_k(t_k) \\ &= q_{\theta_{ij}}^{-1}\left\{q_{\theta_{ij}}[S_{i|k}(t_i|t_k)] + q_{\theta_{ij}}[S_{j|k}(t_j|t_k)]\right\} S_k(t_k) \end{aligned} \tag{2.114}$$

q is the copula generator. Let

$$V = C(T_i, T_j, T_k) = S_{ij\cdot k}(S_i(T_i), S_j(T_j), S_k(T_k)), \tag{2.115}$$

and for $0 \leqslant v \leqslant 1$, let

$$U_k = \frac{q_{\theta_{ij}}[S_{i|k}(T_i|T_k)]}{q_{\theta_{ij}}[S_{i|k}(T_i|T_k)] + q_{\theta_{ij}}[S_{j|k}(T_j|T_k)]}, \tag{2.116}$$

and

$$V_k = \frac{V}{S_k(T_k)} \tag{2.117}$$

Then V is distributed on $[0,1]$ as

$$
\begin{aligned}
& W_k^*(v) \\
=& P[C(T_i,T_j,T_k) \leqslant v] \\
=& E\{P[C(T_i,T_j,T_k) \leqslant v \mid T_k = t_k]\} \\
=& E\{P[C(T_i,T_j,T_k) \leqslant v \mid S_k(T_k) = S_k(t_k)]\} \\
=& E\{P[S_{ij\cdot k}(S_i(T_i),S_j(T_j),S_k(T_k)) \leqslant v \mid S_k(T_k) = S_k(t_k)]\} \\
=& E\{P[C_{ij|k}(S_{i|k}(T_i|t_k),S_{j|k}(T_j|t_k))S_k(t_k) \leqslant v \mid S_k(T_k) = S_k(t_k)]\} \\
=& E\{\mathbf{1}[S_k(t_k) \leqslant v]P[C_{ij|k}(S_{i|k}(T_i|t_k),S_{j|k}(T_j|t_k)) \leqslant 1 \mid S_k(T_k) = S_k(t_k)]\} \\
& + E\left\{\mathbf{1}[S_k(t_k) > v]P\left[C_{ij|k}(S_{i|k}(T_i|t_k),S_{j|k}(T_j|t_k)) \leqslant \frac{v}{S_k(t_k)} \;\middle|\; S_k(T_k) = S_k(t_k)\right]\right\} \\
=& E\{\mathbf{1}[1-t_k \leqslant v] * 1 \mid S_k(T_k) = S_k(t_k)]\} \\
& + E\left\{\mathbf{1}[1-t_k > v]G_k\left(\frac{v}{S_k(t_k)}\right) \;\middle|\; S_k(T_k) = S_k(t_k)\right\} \\
=& E\{\mathbf{1}[t_k \geqslant 1-v] * 1 \mid S_k(T_k) = S_k(t_k)]\} \\
& + E\left\{\mathbf{1}[t_k < 1-v]G_k\left(\frac{v}{1-t_k}\right) \;\middle|\; S_k(T_k) = S_k(t_k)\right\} \\
=& \int_{1-v}^{1} 1\,dt_k + \int_0^{1-v} G_k\left(\frac{v}{1-t_k}\right)\,dt_k \\
=& v + \int_0^{1-v} G_k\left(\frac{v}{1-t_k}\right)\,dt_k
\end{aligned}
\tag{2.118}
$$

So V is distributed on $[0,1]$ as

$$W_k^*(v) = v + \int_0^{1-v} G_k\left(\frac{v}{1-t_k}\right) dt_k, \tag{2.119}$$

with

$$G_k(v_k) = v_k - \frac{q_{\theta_{ij}}(v_k)}{q'_{\theta_{ij}}(v_k)}, v_k = \frac{v}{1-t_k}. \tag{2.120}$$

And

$$\begin{aligned} V_k &= \frac{V}{S_k(T_k)} \\ &= \frac{V}{1-T_k} \\ &= C_{ij|k}\left(S_{i|l}(T_i|T_k), S_{j|k}(T_j|T_k)\right) \\ &= q_{\theta_{ij}}^{-1}\left\{q_{\theta_{ij}}\left[S_{i|k}(T_i|T_k)\right] + q_{\theta_{ij}}\left[S_{j|k}(T_j|T_k)\right]\right\} \end{aligned} \tag{2.121}$$

Why the above discussion is incorrect? The reason is the assumption is not satisfied. A basic assumption for the Kendall's procedure is that the bivariate data must be the pre-specified Archimedean copula. Take the three-dimensional proposed

structure $S_{ij\cdot k}$ as an example:

$$
\begin{aligned}
&P(T_i > t_i, T_j > t_j | T_k = t_k) \\
=&\frac{P(T_i > t_i, T_j > t_j, T_k = t_k)}{P(T_k = t_k)} \\
=&\frac{\partial}{\partial t_k} S_{ij\cdot k}(t_i, t_j, t_k) \\
=&\frac{\partial}{\partial t_k} \left\{ C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) S_k(t_k) \right\} \\
=&\left\{ \frac{\partial}{\partial t_k} C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \right\} S_k(t_k) + C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k)) \left\{ \frac{\partial}{\partial t_k} S_k(t_k) \right\} \\
\neq& C_{ij|k}(S_{i|k}(T_i|t_k), S_{j|k}(T_j|t_k))
\end{aligned}
\tag{2.122}
$$

So it is not the Archimedean copula as pre-specified. Because the assumption is not satisfied which result a pitfall to use the Kendall's procedure directly for the proposed structure.

It is worth mentioning that you should always check the assumption before use any existing procedure for the proposed structures. Otherwise, you will get a incorrect conclusion.

2.5 Parameter Estimation for Proposed Structures

For parameter estimation of the proposed structures, there are two different ways to approach. The first one is to use the Maximum Likelihood Estimation (MLE). The other one is to use the pairwise estimation. Details will be discussed in this section.

For any given structures $S_{12\cdots(k-1)(k+1)\cdots d\cdot k}$ based on the proposed construction method, the parameter estimator from the maximum likelihood estimation can be

derived through following steps:

First, the likelihood function can be derived by

$$\mathcal{L} = \prod_{h=1}^{n} s_{12\cdots(k-1)(k+1)\cdots d\cdot k}(t_1, t_2, \cdots, t_d), \tag{2.123}$$

where n is the sample size, and

$$\begin{aligned} & s_{12\cdots(k-1)(k+1)\cdots d\cdot k}(t_1, t_2, \cdots, t_d) \\ =& \frac{\partial^d}{\partial t_1 \partial t_2 \cdots \partial t_d} S_{12\cdots(k-1)(k+1)\cdots d\cdot k}(S_1(T_1), S_2(T_2), \cdots, S_d(T_d)). \end{aligned} \tag{2.124}$$

Second, the log likelihood function can be derived by

$$\begin{aligned} \ell &= \log \mathcal{L} \\ &= \sum_{h=1}^{n} \log \left[s_{12\cdots(k-1)(k+1)\cdots d\cdot k}(t_1, t_2, \cdots, t_d) \right]. \end{aligned} \tag{2.125}$$

Next steps are taking the derivative of the log likelihood, and setting them equal to

0, that are

$$\frac{\partial \ell}{\partial \theta_{ij}} = 0, \tag{2.126}$$

where $i, j \in \{1, 2, \cdots, d\}, i \neq j$. Solve those equations, the parameter estimators θ_{ij} by maximum likelihood estimation can be derived.

It is worth mentioning that for this class of models, there is another approach available for parameter estimation. The dependence parameters for any pair of random variables can be estimated by the sample estimate of Kendall's τ. Kendall's τ_{ij} of (T_i, T_j) is defined as

$$\begin{aligned} \tau &= E[sign(T_{ih'} - T_{ih})(T_{jh'} - T_{jh})] \\ &= 4E[S(T_i, T_j)] - 1 \end{aligned} \tag{2.127}$$

where $i, j \in \{1, 2, \cdots, d\}, i \neq j$, and $h, h' \in \{1, 2, \cdots, n\}, h \neq h'$, n is the sample size.

And the joint survival function $S(T_i, T_j)$ of (T_i, T_j) can be estimated by

$$\hat{S}_{\theta_{ij},n}(t_{ih}, t_{jh}) = \hat{P}(T_i > t_{ih}, T_j > t_{jh}) = \frac{e_{ij,h}}{n} \tag{2.128}$$

where $e_{ij,h}$ denote the number of events satisfied $t_{ih'} > t_{ih}$ and $t_{jh'} > t_{jh}$ at the same time, which is

$$e_{ij,h} = \#_{h'}(t_{ih'} > t_{ih}, t_{jh'} > t_{jh}). \tag{2.129}$$

Next, the pairwise estimators of τ_{ij} can be estimated by

$$\hat{\tau}_{ij,n} = 4E(\hat{S}_{\theta_{ij}}(t_{ih}, t_{jh})) - 1 = 4E\left(\frac{e_{ij,h}}{n}\right) - 1 \tag{2.130}$$

It can be linked to the parameter θ_{ij} by $\tau = g(\theta)$. Then the pairwise estimator of θ_{ij} can be estimated by

$$\hat{\theta}_{ij,n} = g^{-1}(\hat{\tau}_{ij,n}) \tag{2.131}$$

where $\hat{\tau}_{ij,n}$ is the sample estimate of Kendall's τ. Here g^{-1} is the inverse function of g. And g is a one-to-one function which is uniquely determined by the copula used to model (T_i, T_j).

This proposed pairwise estimator of θ_{ij} offers a simple way for parameter estimation. Firstly, it only needs the information for (T_i, T_j) from $d-$dimensional data $(T_1, T_2, \cdots, T_d)$ which is much simpler than the maximum likelihood estimation. If d is large (for example, $d > 500$), the estimator of MLE is almost intractable. Secondly, it only uses the incomplete information from a $d-$dimensional data, but it still solves the problem quite well. Details will be given by numerical studies.

2.6 Model Selection for Proposed Structures

For bivariate copula, there are many goodness-of-fit or model selection procedures for copula models. Oakes (1989) [24] proposed a graphic diagnostic approach to check the goodness-of-fit for Archimedean copula models. Shih (1998) [33] proposed a goodness-of-fit test for the Clayton model. The test procedure is designed specifically for the Clayton model. Wang and Wells (2000) [44] proposed a model selection procedure within the Archimedean copula family for right-censored bivariate data based on the so-called L^2 norm of the Kendall distribution (basically a distance measure between the empirical and the estimated Kendall distribution). Genest, Quessy, and Rémillard (2006) [9] extended the idea in Wang and Wells (2000) [44] and proposed a general goodness-of-fit test procedure for models belonging to the Archimedean copula family. Wang (2010) [38] proposed goodness-of-fit tests for Archimedean copula models for both uncensored and censored bivariate models. The research goal will focus on a model selection procedure for the multivariate models proposed.

Based on the discussion from previous sections, the models from fomular $(2.18), (2.19)$ and (2.20) are different when $C_{12} \neq C_{13} \neq C_{23} \neq C_{12}$. Similarly, the models from fomular $(2.28), (2.29), (2.30)$ and (2.31) are also different when $C_{12}, C_{13}, C_{14}, C_{23}, C_{24}$ and C_{34} are different with each other. Here the differences between copulas include they are different types of copulas and also include they are same type of copulas with different parameters. How to select the best model for

our data? The answer will greatly help us for applications. In this dissertation, a simple goodness-of-fit test is given to check the best model for the given uncensored multivariate data.

The joint survival function $S(T_1, T_2, \cdots, T_d)$ of $(T_1, T_2, \cdots, T_d)$ can be estimated by

$$\hat{S}_n(t_{1h}, t_{2h}, \cdots, t_{dh}) = \hat{P}(T_1 > t_{1h}, T_2 > t_{2h}, \cdots, T_d > t_{dh}) = \frac{e_{12\cdots d,h}}{n} \tag{2.132}$$

where n is the sample size and $e_{12\cdots d,h}$ denote the number of events satisfied $t_{1h'} > t_{1h}$, $t_{2h'} > t_{2h}, \cdots, t_{dh'} > t_{dh}$ at the same time and $h, h' \in \{1, 2, \cdots, n\}, h \neq h'$, which is

$$e_{12\cdots d,h} = \#_{h'}(t_{1h'} > t_{1h}, t_{2h'} > t_{2h}, \cdots, t_{dh'} > t_{dh}). \tag{2.133}$$

The main idea of the proposed method is to use the least squares approach that the best model can be chosen to minimize the corresponding sum of squares:

$$R_{k,n} = \sum_{h=1}^{n} \left[\hat{S}_n(t_{1h}, t_{2h}, \cdots, t_{dh}) - S_k^*(t_{1h}, t_{2h}, \cdots, t_{dh})\right]^2. \tag{2.134}$$

where S_k^* is the survival function from Section 2.4 for the proposed structure $S_{12\cdots(k-1)(k+1)\cdots d\cdot k}$ and $k \in \{1, 2, \cdots, d\}$.

The model selection procedure is described as follows: suppose that there are several possible models of proposed structures to fit a $d-$dimensional data, for each possible model, the unknown parameters can be estimated by the estimators from Section 2.5. The model $S_{12\cdots(k-1)(k+1)\cdots d\cdot k}$ producing the smallest $R_{k,n}$ will be selected as the best model for analyzing the data set.

2.7 Numerical Studies

In this section, the parameters estimation procedures for three-dimensional data are demonstrated. The performance of the proposed pairwise estimator in different scenarios with sample size $N = 500$ and sample size $N = 1000$ are evaluated.

Suppose the data follows the copula structure illustrated in Figure 2.5 where

- (t_1, t_2) is modeled by a Clayton copula [3] with parameter θ_{12};
- (t_1, t_3) is modeled by a Clayton copula [3] with parameter θ_{13};
- (t_2, t_3) is modeled by a Clayton copula [3] with parameter θ_{23}.

The data generation begins from generating n random variables v_1, v_2 and v_3 respectively from Uniform $[0, 1]$ with $v_1 \perp v_2 \perp v_3 \perp v_1$. Next, you can transform v_1, v_2 and v_3 through CPI Rosenblatt transform [29, 26] to u_1, u_2 and u_3 by the proposed three-dimensional structures from fomula (2.66) or (2.67) or (2.68) as illustrated in Figure 2.5 with parameters $\theta_{12}, \theta_{13}, \theta_{23}$. And their marginal distributions are chosen to be exponentially distributed with rate 1, therefore the observation of data (t_1, t_2, t_3) can be generated with

$$t_1 = -\ln(u_1), \tag{2.135}$$

$$t_2 = -\ln(u_2), \tag{2.136}$$

and

$$t_3 = -\ln(u_3). \tag{2.137}$$

Here model $S_{12\cdot3}$ from fomula (2.68) is used for demonstration. The settings for parameters are $\theta_{12} = \frac{6}{7}, \theta_{13} = 2, \theta_{23} = \frac{14}{3}$ corresponding to Kendall's $\tau_{12} = 0.3, \tau_{13} = 0.5, \tau_{23} = 0.7$ respectively. In each scenario, 1000 times replications for the same procedure are performed and the estimated Kendall's τ are compared with the true value. In the simulations, the proposed pairwise estimators are also compared with the estimators by MLE. The standard deviation of the estimated Kendall's τ are calculated. Table 2.1 shows the simulation results from all scenarios.

The simulation results have shown that the mean values of the proposed pairwise parameter estimates are very close to the true values even when the Kendall's τ value are small. Overall the proposed pairwise estimator works very well under the model $S_{12\cdot3}$ for all scenarios. When the sample size N increased from 500 to 1000, the proposed pairwise estimators are closer to the true values. Also from Tables 2.1, the estimators from MLE is less biased and outperforms the proposed pairwise estimator under the model $S_{12\cdot3}$ when the dependence level is high. Because the estimators

Table 2.1 Simulation Results using Pairwise Estimation vs. Maximum Likelihood Estimation for the Clayton Copula

Sample Size	N=500					
τ Setting	$\tau_{12} = 0.3$		$\tau_{13} = 0.5$		$\tau_{23} = 0.7$	
Method	Pairwise	MLE	Pairwise	MLE	Pairwise	MLE
Mean of $\hat{\tau}$	0.304	0.185	0.438	0.520	0.635	0.677
SD of $\hat{\tau}$	0.0283	0.0277	0.0249	0.0211	0.0209	0.0176

Sample Size	N=1000					
τ Setting	$\tau_{12} = 0.3$		$\tau_{13} = 0.5$		$\tau_{23} = 0.7$	
Method	Pairwise	MLE	Pairwise	MLE	Pairwise	MLE
Mean of $\hat{\tau}$	0.303	0.184	0.438	0.519	0.632	0.675
SD of $\hat{\tau}$	0.0196	0.0226	0.0181	0.0153	0.0143	0.0108

from MLE uses more information than the proposed pairwise estimators. Overall, the proposed pairwise estimator is comparable with the estimators from MLE. And the proposed pairwise estimator is much simpler than the estimators from MLE especially when d is big.

It is worth mentioning that the simulation studies for the model $S_{23\cdot1}$ and $S_{13\cdot2}$ also have been conducted and find that the proposed pairwise estimators works very well (the simulation results for the model $S_{23\cdot1}$ and $S_{13\cdot2}$ are omitted here).

2.8 Discussion

The intellectual merit of this proposed structures should be assessed on (at least) three levels. On the first level, this research contributes directly to high dimensional data analysis and survival analysis. The proposed high dimensional models can be

used to model multivariate data with any bivariate margins. It offers the modelling flexibility many models such as the hierarchical models, frailty models or vine models cannot provide. The properties of the proposed models have also been explored to provide guidance for estimating the parameters of models for a given multivariate data. The association between competing risks can be decided and quantified using the copula model assumption and the knowledge of the marginal distributions.

On the second level, multivariate normal distribution has been a dominant multivariate model because of its simplicity and flexibility for bivariate margins. The proposed class of models have these nice properties and can be used to model multivariate non-normal data. The exploration of the properties of this class of models will not only deepen the understanding of this type of models but also reveal any advantages of applying these models. The identifiability property established using the proposed structures has shown the advantages to model dependent competing risks data using the Archimedean copula models.

On the highest level, the results of this research are especially important in analyzing multivariate data using copula models. High dimensional data analysis now has become a hot area with the quick development of the computer science and the emergence of big data. The proposed modelling techniques will definitely help to lay a solid foundation for future development and success of multivariate analysis and survival analysis. The research will contribute to the advancement of the statistical theory on correlation studies and greatly improve the understanding of the dependence structure in multivariate data.

CHAPTER 3

ANALYSIS OF SEMI-COMPETING RISKS DATA USING ARCHIMEDEAN COPULA MODELS

3.1 Introduction

In medical research, the Disease Free Survival time (DFS) or the Disease Progression Free survival time (PFS) and the Overall Survival time (OS) are usually observable for the same subject. Understanding the dependency of DFS/PFS and OS plays an important role in clinical practice.

For example, DFS/PFS can be used as a surrogate primary endpoint for OS if the association between DFS/PFS and OS are high [35, 27]. DFS and OS are often dependent and the OS (denotes by Y) can censor DFS (denoted by X), but not vice versa. This type of survival data is called semi-competing risks data [5] and has received lots of attention recently.

Lagakos (1976) [19, 20] applied a parametric distribution for (X, Y) in the region $X < Y$. Fine, Jiang and Chappel (2001) [5] proposed to model the dependence structure between X and Y using the Clayton copula [3] model for pairs falling into the upper wedge (i.e, $X < Y$) and applied a procedure proposed by Oakes (1986) [23] to estimate the association parameter. Lakhal, Rivest and Abdous (2008) [21] extended their approach and proposed estimation strategies based on an estimating equation derived from the conditional tau. Their approach is essentially a moment estimation approach and tends to be quite complicated in setting up their estimating equation according to different scenarios from the simulation studies.

In this dissertation, target is to determine the true dependence level between X and Y in semi-competing risks data setting. A novel and effective strategy is proposed to analyze this type of data using different Archimedean copula models [8].

A bivariate random vector (X, Y) follows an Archimedean copula if the joint survival function of (X, Y) can be expressed as:

$$S(x, y) = \psi_\theta^{-1} \left\{ \psi_\theta \left[S_X(x) \right] + \psi_\theta \left[S_Y(y) \right] \right\}, \tag{3.1}$$

where S_X and S_Y are marginal survival functions of X and Y respectively, ψ_θ^{-1} is defined on $[0, \infty]$ so that

$$\psi_\theta^{-1}(0) = 1, \tag{3.2}$$

$$[\psi_\theta^{-1}]'(s) < 0, \tag{3.3}$$

$$[\psi_\theta^{-1}]''(s) > 0. \tag{3.4}$$

ψ_θ is the inverse function of ψ_θ^{-1}, defined as a copula generator [22] and θ is the unknown parameter.

The first Archimedean copula model was proposed by Clayton (1978) [3]. Another important frailty model, the Hougaard model [13, 14], has

$$\psi_\beta^{-1}(s) = exp(-s^\beta). \tag{3.5}$$

Its bivariate survivor function is

$$S(x,y) = exp\left(- \left[\left\{ -log\left[S_X(x) \right] \right\}^{\frac{1}{\beta}} + \left\{ -log\left[S_Y(y) \right] \right\}^{\frac{1}{\beta}} \right]^{\beta} \right) \tag{3.6}$$

for $\beta \in (0,1)$. Besides the Clayton model [3] and the Hougaard model [13, 14], some well-known models such as the Frank model [7] and the Log-copula model also belong to this family.

Assuming that (X,Y) follows an Archimedean copula model, this dissertation first derive the copula graphic estimator of marginal survival function of Y based on a semi-competing risks data and prove its asymptotic properties, the copula graphic estimator was proposed by Zheng and Klein (1995) [47, 16] initially.

Applying the copula-graphic estimator derived in this dissertation, a new estimation procedure is proposed for association parameters in Archimedean copula models based on an observed semi-competing risks data set. Using the proposed parameter estimates, the marginal survival functions of X and Y can be consistently estimated.

This chapter is organized in the following way: in Section 3.2, the copula-graphic estimator for survival function of Y based on a semi-competing risks data is derived. In Section 3.3, a new estimation strategy is presented and the asymptotic properties of the new parameter estimates is discussed. A model selection procedure is proposed in Section 3.4 and how to accommodate covariates is described in Section 3.5. The simulation studies are reported in Section 3.6 and an illustrative example is gived to demonstrate the usefulness and effectiveness of the porposed strategies in Section 3.7. The chapter is ended with some discussion in Section 3.8.

3.2 Copula-graphic Estimator for Marginal Survival Function of Y Based on Semi-competing Risks Data

Using the similar notation as in Fine, Jiang and Chappel (2001) [5] and Lakhal, Rivest and Abdous (2008) [21], define

$$Z = \min\{X, Y\}, \tag{3.7}$$

$$\delta_Z = I(Z < C), \tag{3.8}$$

$$\delta_X = I(X < Y), \tag{3.9}$$

$$\delta_Y = I(Y < C), \tag{3.10}$$

$$S = \min\{Z, C\}, \tag{3.11}$$

$$R = \min\{Y, C\}. \tag{3.12}$$

The observable part of a semi-competing risks data can be expressed as

$$(S, \delta_Z, \delta_Z\delta_X, R, \delta_Y). \tag{3.13}$$

Suppose that (X, Y) can be modelled by an Archimedean copula with the dependence parameter θ, one way of estimating the survival function of Y is to apply the copula-

graphic estimator proposed by Zheng and Klein (1995) [47] to the setting. It is worth mentioning that the proposed copula-graphic estimator in this dissertation is different from the one proposed by Lakhal, Rivest and Abdous (2008) [21] as they proposed the estimator for the marginal survival function of X while the proposed copula-graphic estimator in this dissertation is developed to estimate the marginal survival function of Y.

This dissertation has also established uniform consistency and weak convergence for the copula-graphic estimator of the marginal survival function of Y. The main purpose of deriving this copula-graphic estimator for survival function of Y is to use it to develop a new parameter estimation approach for semi-competing risks data.

The idea is described as follows:

- Because C is independent of (X, Y), the survival function of $Z = \min\{X, Y\}$ which is $\pi(z) = Pr(Z > z)$ can be estimated by the Kaplan- Meier estimate, denote it by $\hat{\pi}(z)$.
- At points (X, Y, C) where $Y < \min\{X, C\}$ (i.e., $\delta_Z = 1, \delta_Z\delta_X = 0$), the corresponding survival function of Y has a jump at Y while the corresponding survival function of X has no jump at X.
- At points (X, Y, C) where $C < \min\{X, Y\}$ (i.e., $\delta_Z = 0$), the corresponding survival function of X has no jump at X and the corresponding survival function of Y has no jump at Y, therefore $\hat{\pi}(z)$ has no change at these points.

For $i \in \{1, 2, \cdots, n\}$, a semi-competing risks data set includes n independent replications of

$$S_i = \min\{Z_i, C_i\}, \tag{3.14}$$

$$\delta_{Z_i} = I(Z_i < C_i), \tag{3.15}$$

$$\delta_{Z_i}\delta_{X_i} = I(Z_i < C_i)I(X_i < Y_i), \tag{3.16}$$

$$R_i = \min\{Y_i, C_i\}, \tag{3.17}$$

$$\delta_{Y_i} = I(Y_i < C_i). \tag{3.18}$$

For any given θ value,

$$\hat{\pi}(S_i) = \psi_\theta^{-1}\left\{\psi_\theta\left[\hat{S}_X(S_i)\right] + \psi_\theta\left[\hat{S}_Y(S_i)\right]\right\}, \tag{3.19}$$

if $Y_i < \min\{X_i, C_i\}$, then $\hat{S}_X(S_i) - \hat{S}_X(S_{i^-}) = 0$ and we must have

$$\psi_\theta\left[\hat{S}_Y(S_i)\right] - \psi_\theta\left[\hat{S}_Y(S_{i^-})\right] = \psi_\theta\left[\hat{\pi}(S_i)\right] - \psi_\theta\left[\hat{\pi}(S_{i^-})\right] \tag{3.20}$$

where $0 = S_0 < S_1 < S_2 < \cdots < S_n$, suppose that S_i are increasingly ordered. Summing above equality from 0 to t:

$$\psi_\theta\left[\hat{S}_Y(t)\right] = \sum_{S_i \leqslant t, Y_i < \min\{X_i, C_i\}} \left\{\psi_\theta\left[\hat{\pi}(S_i)\right] - \psi_\theta\left[\hat{\pi}(S_{i^-})\right]\right\} \tag{3.21}$$

using the property $\psi_\theta(1) = 0$. Equivalently,

$$\hat{S}_Y(t) = \psi_\theta^{-1}\left[\sum_{S_i \leqslant t, Y_i < \min\{X_i, C_i\}} \left\{\psi_\theta\left[\hat{\pi}(S_i)\right] - \psi_\theta\left[\hat{\pi}(S_{i^-})\right]\right\}\right]. \tag{3.22}$$

Using the martingale presentation, the above estimator can be expressed as:

$$\hat{S}_Y(t) = \psi_\theta^{-1}\left[\int_0^t \left\{\psi_\theta\left[\hat{\pi}(u)\right] - \psi_\theta\left[\hat{\pi}(u^-)\right]\right\} d\bar{N}(u)\right], \tag{3.23}$$

where

$$N_i(t) = I\{Y_i \leqslant t, Y_i < \min\{X_i, C_i\}\}, \tag{3.24}$$

$$\bar{N}(t) = \sum_{i=1}^{n} N_i(t), \tag{3.25}$$

$$Y_i(t) = I\{\min\{X_i, Y_i, C_i\} \geqslant t\} = I\{S_i \geqslant t\}, \tag{3.26}$$

$$\bar{Y}(t) = \sum_{i=1}^{n} Y_i(t). \tag{3.27}$$

For the Kaplan-Meier estimator $\hat{\pi}(u)$,

$$\hat{\pi}(u) = \hat{\pi}(u^-) \left\{ 1 - \frac{\Delta \bar{N}_1(u)}{\bar{Y}_1(u)} \right\}, \tag{3.28}$$

where

$$N_{1i}(t) = I\{\min\{X_i, Y_i\} \leqslant t, \min\{X_i, Y_i\} < C_i\}, \tag{3.29}$$

$$\bar{N}_1(t) = \sum_{i=1}^{n} N_{1i}(t), \tag{3.30}$$

$$Y_{1i}(t) = I\{\min\{X_i, Y_i\} \geqslant t, C_i \geqslant t\} = I\{S_i \geqslant t\}, \tag{3.31}$$

$$\bar{Y}_1(t) = \sum_{i=1}^{n} Y_{1i}(t). \tag{3.32}$$

Therefore

$$\begin{aligned} \psi_\theta \left[\hat{\pi}(u)\right] - \psi_\theta \left[\hat{\pi}(u^-)\right] &\approx \psi'_\theta \left[\hat{\pi}(u^-)\right] \left[\hat{\pi}(u) - \hat{\pi}(u^-)\right] \\ &= -\psi'_\theta \left[\hat{\pi}(u^-)\right] \hat{\pi}(u^-) \frac{\Delta \bar{N}_1(u)}{\bar{Y}_1(u)} \end{aligned} \tag{3.33}$$

by the Taylor expansion and the formulas given on page 97 in Fleming and Harrington (1991) [6]. When $Y_i < \min\{X_i, C_i\}$ and $Y_i \leqslant t$ for some i, then: $\min\{X_i, Y_i\} \leqslant Y_i \leqslant t$ and $\min\{X_i, Y_i\} = Y_i < \min\{X_i, C_i\} \leqslant C_i$, therefore if $dN_i(u) = 1$, then $\Delta N_{1i}(u) = 1$. Assuming that Y and X are both absolutely continuous failure time random variables, then from $d\bar{N}(u) = 1$, we have $\Delta\bar{N}_1(u) = 1$ (i.e., there are no tied failure times). Define

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda^\sharp(s)ds \tag{3.34}$$

and

$$\bar{M}(t) = \bar{N}(t) - \int_0^t \bar{Y}(s)\lambda^\sharp(s)ds \tag{3.35}$$

as the martingales with respect to the σ field

$$\mathcal{F}_t^i = \sigma\left\{I(Y_i \leqslant t, Y_i < \min\{X_i, C_i\}), I(Y_i \leqslant t, Y_i \geqslant \min\{X_i, C_i\})\right\} \tag{3.36}$$

and

$$\mathcal{F}_t = \bigvee_{i=1}^{n} \mathcal{F}_t^i \tag{3.37}$$

respectively. The $\lambda^\sharp(s)$ is defined as the crude hazard function of Y:

$$\begin{aligned}
\lambda^\sharp(s) &= \frac{-\dfrac{\partial}{\partial u} P(Y \geqslant u, \min\{X, C\} \geqslant t)|_{u=t}}{P(Y \geqslant t, \min\{X, C\} \geqslant t)} \\
&= \frac{-\dfrac{\partial}{\partial u} P(Y \geqslant u, X \geqslant t, C \geqslant t)|_{u=t}}{P(Y \geqslant t, X \geqslant t, C \geqslant t)} \\
&= \frac{-\dfrac{\partial}{\partial u} P(Y \geqslant u, X \geqslant t)|_{u=t}}{P(Y \geqslant t, X \geqslant t)}
\end{aligned} \tag{3.38}$$

because (X, Y) is independent of C. The corresponding crude cumulative function was denoted by $\Lambda^\sharp(s)$.

Based on above arguments and definitions, then

$$\psi_\theta \left[\hat{S}_Y(t)\right] = -\int_0^t \psi'_\theta \left[\hat{\pi}(u^-)\right] \hat{\pi}(u^-) \frac{\bar{N}(u)}{\bar{Y}(u)} \tag{3.39}$$

by considering the fact that

$$\begin{aligned} Y_{1i}(u) &= I\{\min\{X_i, Y_i\} \geqslant u, C_i \geqslant u\} \\ &= I\{S_i \geqslant u\} \\ &= I\{Y_i \geqslant u, \min\{X_i, C_i\} \geqslant u\} \\ &= Y_i(u) \end{aligned} \tag{3.40}$$

and hence

$$\bar{Y}_1(u) = \bar{Y}(u). \tag{3.41}$$

Now

$$\begin{aligned} &\psi_\theta\left[\hat{S}_Y(t)\right] \\ &= -\int_0^t \psi'_\theta\left[\hat{\pi}(u^-)\right]\hat{\pi}(u^-)\frac{d\bar{N}(u) - \bar{Y}(u)\lambda^\sharp(u)du}{\bar{Y}(u)} - \int_0^t \psi'_\theta\left[\hat{\pi}(u^-)\right]\hat{\pi}(u^-)\lambda^\sharp(u)du \\ &= -\int_0^t \psi'_\theta\left[\hat{\pi}(u^-)\right]\hat{\pi}(u^-)\frac{d\bar{M}(u)}{\bar{Y}(u)} - \int_0^t \psi'_\theta\left[\hat{\pi}(u^-)\right]\hat{\pi}(u^-)\lambda^\sharp(u)du. \end{aligned} \tag{3.42}$$

Also,

$$\sup_{0<u<t}\left|\frac{\bar{Y}(u)}{n}-P(S\geqslant u)\right|\rightarrow 0,\ when\ n\rightarrow\infty, \tag{3.43}$$

$$\hat{\pi}(u^{-})\rightarrow\pi(u)=P(\min\{X,Y\}>u),\ when\ n\rightarrow\infty. \tag{3.44}$$

Based on above derivations, the estimator estimates a survival function

$$S_Y(t)=\psi_\theta^{-1}\left[-\int_0^t\psi_\theta'\left[\pi(u)\right]\pi(u)\lambda^\sharp(u)du\right]. \tag{3.45}$$

This is because

$$\begin{aligned}&\psi_\theta\left[\hat{S}_Y(t)\right]-\psi_\theta\left[S_Y(t)\right]\\ =&\frac{1}{n}\int_0^t\frac{-\psi_\theta'\left[\hat{\pi}(u^{-})\right]\hat{\pi}(u^{-})}{\bar{Y}(u)/n}d\bar{M}(u)-\int_0^t\left\{\psi_\theta'\left[\hat{\pi}(u^{-})\right]\hat{\pi}(u^{-})-\psi_\theta'\left[\pi(u)\right]\pi(u)\right\}\lambda^\sharp(u)du.\end{aligned} \tag{3.46}$$

It is easy to show that the first term goes to zero by following the similar arguments in proving Theorem 3.4.2 in Fleming and Harrington (1991) [6] (the Lenglart inequality will be applied). The second term goes to zero by the uniform consistency of the Kaplan-Meier estimate and the boundedness of the derivative function of $\Phi(s)=$

$-\psi'_\theta(s)s$. It has thus proved that the copula-graphic estimator derived is still a consistent estimator of the survival function of Y given θ.

Now the asymptotic distribution of $\hat{S}_Y(t)$ is derived as

$$\begin{aligned}
&\psi_\theta\left[\hat{S}_Y(t)\right] - \psi_\theta\left[S_Y(t)\right] \\
=&\frac{1}{n}\int_0^t \frac{-\psi'_\theta\left[\pi(u)\right]}{P(C>u)} d\bar{M}(u) - \int_0^t \left\{\psi'_\theta\left[\hat{\pi}(u^-)\right]\hat{\pi}(u^-) - \psi'_\theta\left[\pi(u)\right]\pi(u)\right\}\lambda^\sharp(u)du + o_p(1).
\end{aligned} \tag{3.47}$$

Using Gill's representation (1980) [12] for the Kaplan-Meier estimate, then

$$\sqrt{n}\left[\hat{\pi}(u^-) - \pi(u)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[-\pi(u)\int_0^u \frac{dM_{1i}(w)}{\pi_1(w)}\right] + o_p(1) \tag{3.48}$$

can be derived by noticing the fact that

$$\pi(u) = \pi(u^-) = P(\min\{X, Y\} > u) \tag{3.49}$$

and the absolute continuity of X and Y. Here $\pi_1(w)$ is defined as

$$\pi_1(w) = P(S > w) = P(X > w, Y > w, C > w) \tag{3.50}$$

and $M_{1i}(u)$ is defined as

$$M_{1i}(u) = N_{1i}(u) - \int_0^u I\{S \geqslant w\}\lambda_1(w)dw, \tag{3.51}$$

where

$$N_{1i}(u) = I\left\{\min\{X, Y\} \leqslant u, \min\{X, Y\} < C\right\}. \tag{3.52}$$

$M_{1i}(u)$ is the martingale with respect to the σ field,

$$\mathcal{F}_{1t}^i = \sigma\left\{I(\min\{X_i, Y_i\} \leqslant t, \min\{X_i, Y_i\} < C_i), I(\min\{X_i, Y_i\} \leqslant t, \min\{X_i, Y_i\} \geqslant C_i)\right\} \tag{3.53}$$

and

$$\mathcal{F}_{1t} = \bigvee_{i=1}^{n} \mathcal{F}_{1t}^{i} \tag{3.54}$$

respectively. And $\lambda_1(t)$ is defined as the crude hazard function of $\min\{X_i, Y_i\}$:

$$\begin{aligned}
\lambda_1(t) =& \frac{-\frac{\partial}{\partial u} P\left(\min\{X,Y\} \geqslant u, C \geqslant t\right)|_{u=t}}{P\left(\min\{X,Y\} \geqslant t, C \geqslant t\right)} \\
=& \frac{-\frac{\partial}{\partial u} P\left(X \geqslant u, Y \geqslant u, C \geqslant t\right)|_{u=t}}{P\left(X \geqslant t, Y \geqslant t, C \geqslant t\right)} \\
=& \frac{-\frac{\partial}{\partial u} P\left(X \geqslant u, Y \geqslant u\right)|_{u=t}}{P\left(X \geqslant t, Y \geqslant t\right)}
\end{aligned} \tag{3.55}$$

by the independence of (X, Y) and C. Then

$$\begin{aligned}
&\sqrt{n}\left\{\psi_\theta\left[\hat{S}_Y(t)\right] - \psi_\theta\left[S_Y(t)\right]\right\} \\
=& \frac{1}{\sqrt{n}} \int_0^t \frac{-\psi'_\theta\left[\pi(u)\right]}{P(C>u)} d\bar{M}(u) + \frac{1}{\sqrt{n}} \int_0^t \Phi'\left[\pi(u)\right] \left[-\pi(u) \int_0^u \frac{d\bar{M}_1(w)}{\pi_1(w)}\right] \lambda^\sharp(u) du + o_p(1) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \int_0^t \frac{-\psi'_\theta\left[\pi(u)\right]}{P(C>u)} dM_i(u) + \int_0^t \Phi'\left[\pi(u)\right] \left[-\pi(u) \int_0^u \frac{dM_{1i}(w)}{\pi_1(w)}\right] \lambda^\sharp(u) du \right\} \\
&+ o_p(1)
\end{aligned} \tag{3.56}$$

where

$$\Phi(s) = -s\psi'_\theta(s). \tag{3.57}$$

Overall it can be concluded that

$$\sqrt{n}\left[\hat{S}_Y(t) - S_Y(t)\right] = \frac{1}{\sqrt{n}\psi'_\theta\left[S_Y(t)\right]} \times$$
$$\sum_{i=1}^{n}\left\{\int_0^t \frac{-\psi'_\theta\left[\pi(u)\right]}{P(C>u)} dM_i(u) + \int_0^t \Phi'\left[\pi(u)\right]\left[-\pi(u)\int_0^u \frac{dM_{1i}(w)}{\pi_1(w)}\right]\lambda^\sharp(u)du\right\} + o_p(1). \tag{3.58}$$

Therefore, the copula-graphic estimator derived above is asymptotically normal with finite variance $v(t)$ for fixed t. And here the analytic form of $v(t)$ is very complicated and use bootstrap method to estimate its variance is recommended.

Theorem 3.2.1. *Let $t_0 \in (0,\infty)$ be such that $\pi(t_0) = P(\min\{X,Y\} > t_0) > 0$. Suppose (X,Y) follows an Archimedean copula model and the derivatives of $\psi_\theta(s)$ and $\Phi(s)$ are bounded for $s \in (\pi(t_0),1)$, $\hat{S}_Y(t)$ is a uniformly consistent estimator of $S_Y(t)$ and $\sqrt{n}[\hat{S}_Y(t) - S_Y(t)]$ converges weakly on $D[0,t_0)$ to a Gaussian process with mean zero and a finite variance $v(t)$.*

3.3 A New Parameter Estimation Strategy Based on Semi-competing Risks Data

For a semi-competing risks data $(S, \delta_Z, \delta_Z\delta_X, R, \delta_Y)$, where Y as the OS can only be censored by an independent censoring time C. The true survival function $S(Y)$ of Y can be estimated by the Kaplan-Meier curve consistently based on (R, δ_Y), the Kaplan-Meier estimator denote by $\hat{S}_K(y)$ here. Based on the copula-graphic estimator $\hat{S}_Y$ derived in previous section, a new parameter estimation strategy for semi-competing risks data is proposed in this section.

The main idea of the proposed method is to use the least squares approach that determines the unknown parameter θ value by minimizing the corresponding sum of squares:

$$Q_n(\theta) = \sum_{i=1}^{n} \eta_i \left[\hat{S}_Y(Y_i) - \hat{S}_K(Y_i)\right]^2 \tag{3.59}$$

where

$$\eta_i = I(Y_i < \min\{X_i, C_i\}). \tag{3.60}$$

Or equivalently,

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta). \tag{3.61}$$

The estimator as the minimizer of $Q_n(\theta)$ is obtained by solving the estimating equation

$$\frac{\partial Q_n}{\partial \theta} = 0 \tag{3.62}$$

where

$$\frac{\partial Q_n}{\partial \theta} = 2\sum_{i=1}^{n} \eta_i \left[\hat{S}_Y(Y_i) - \hat{S}_K(Y_i)\right] \frac{\partial \hat{S}_Y(Y_i)}{\partial \theta}. \tag{3.63}$$

The following regularity conditions are needed to establish the asymptotic normality of the proposed parameter estimate $\hat{\theta}_n$:

Condition 3.3.1. *The parameter space Θ is compact and the true parameter $\theta_0 \in Int(\Theta)$ (the interior of parameter space Θ).*

Condition 3.3.2. $\frac{\partial S_Y(y)}{\partial \theta}$ *is continuous in y and bounded by a constant K.*

Condition 3.3.3. $\frac{\psi'_{\theta_1}(u)}{\psi'_{\theta_2}(u)}$ *is an increasing function of u when $\theta_2 > \theta_1$.*

Remark: the regularity condition 3.3.3 is satisfied by most Archimedean copula models such as the Clayton model [3] and the Frank model [7]. It has also been presented as a condition to prove the Proposition 2 in Rivest and Wells (2001) [28].

Condition 3.3.4. *Integration and differentiation operators are interchangeable.*

The following theorem can be shown

Theorem 3.3.5. *Under regularity conditions 3.3.1-3.3.4, the parameter estimate* $\hat{\theta}_n$ *is consistent and* $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *is asymptotic normal with zero mean and variance* $\frac{\sigma^2}{\gamma^2}$, *where* σ^2 *and* γ^2 *are defined in proof.*

Proof.

Consistency of the parameter estimate $\hat{\theta}_n$*:*

Define $S(y)$ as the true survival function of Y when $\theta = \theta_0$. And here θ_0 is the true underlying parameter value of θ in assumed Archimedean copula model. The

estimating equation can be written as:

$$\begin{aligned}
0 =& M_n(\theta) \\
=& \frac{1}{n} \times \frac{\partial Q_n(\theta)}{\partial \theta} \\
=& \frac{2}{n} \sum_{i=1}^{n} \eta_i \left[\hat{S}_Y(Y_i) - \hat{S}_K(Y_i) \right] \frac{\partial \hat{S}_Y(Y_i)}{\partial \theta} \\
=& \frac{2}{n} \sum_{i=1}^{n} \eta_i \left[\hat{S}_Y(Y_i) - S(Y_i) + S(Y_i) - \hat{S}_K(Y_i) \right] \frac{\partial \hat{S}_Y(Y_i)}{\partial \theta} \\
=& \frac{2}{n} \sum_{i=1}^{n} \eta_i \left[\hat{S}_Y(Y_i) - S(Y_i) \right] \frac{\partial \hat{S}_Y(Y_i)}{\partial \theta} + o_p(1)
\end{aligned} \tag{3.64}$$

by the consistency of the Kaplan-Meier estimate $\hat{S}_K$ of survival function S and also the boundedness of $\frac{\partial \hat{S}_Y}{\partial \theta}$ defined on a compact set (this is true because of the continuity of $\frac{\partial \hat{S}_Y}{\partial \theta}$). By the Strong Law of Large Numbers (SLLN),

$$\frac{1}{n} \times \frac{\partial Q_n(\theta)}{\partial \theta} \to 2E \left\{ \eta \left[S_Y(Y) - S(Y) \right] \frac{\partial S_Y(Y)}{\partial \theta} \right\} = M(\theta) \tag{3.65}$$

in probability as $n \to \infty$ as $\frac{\partial \hat{S}_Y}{\partial \theta}$ converges uniformly to $\frac{\partial S_Y}{\partial \theta}$ in probability. When $\theta = \theta_0$ (θ_0 is true parameter value), $S_Y(y) = S(y)$ is the true survival function of Y and hence $M(\theta_0) = 0$ and when $\theta \neq \theta_0$, $S_Y(y) \neq S(y)$. Actually $S_Y(y) > S(y)$ or $S_Y(y) < S(y)$ for almost all y when $\theta \neq \theta_0$ by the Proposition 2 in Rivest and Wells (2001) [28] (i.e., $S_Y(Y)$ are stochastically ordered for different θ values as the limits of copula graphic estimators based on Archimedean copula models). Now consider

the limiting function of the copula-graphic estimator

$$S_Y(y) = \psi_\theta^{-1}\left[-\int_0^y \psi'_\theta\left[\pi(u)\right]\pi(u)\lambda^\sharp(u)du\right]. \tag{3.66}$$

At two different θ values say $\theta_1 \neq \theta_2$, then

$$S_Y^{(1)}(y) = \psi_{\theta_1}^{-1}\left[-\int_0^y \psi'_{\theta_1}\left[\pi(u)\right]\pi(u)\lambda^\sharp(u)du\right], \tag{3.67}$$

$$S_Y^{(2)}(y) = \psi_{\theta_2}^{-1}\left[-\int_0^y \psi'_{\theta_2}\left[\pi(u)\right]\pi(u)\lambda^\sharp(u)du\right], \tag{3.68}$$

respectively. Mimicking the proof of Proposition 2 in Rivest and Wells (2001) [28], it shows that $S_Y^{(1)}(y) > S_Y^{(2)}(y)$ under the assumption that $\dfrac{\psi'_{\theta_1}(u)}{\psi'_{\theta_2}(u)}$ is an increasing function of u when $\theta_1 < \theta_2$. Therefore, if $S_Y(y)$ is differentiable with respect to θ, it must be negative given y (i.e., $\dfrac{\partial S_Y(y)}{\partial\theta} < 0$). Under the regularity condition 3.3.3, it concludes that $\|M(\theta)\| > 0$ when $\theta \neq \theta_0$. By checking the formula (3.65), it implies that

$$\inf_{\theta: d(\theta,\theta_0)\geqslant\epsilon} \|M(\theta)\| > 0 = \|M(\theta_0)\|. \tag{3.69}$$

Furthermore, it is easy to show that

$$\sup_{\theta\in\Theta} \|M_n(\theta) - M(\theta)\| \xrightarrow{\text{p}} 0. \tag{3.70}$$

By Theorem 5.9 in van der Vaart, A. W. (1998) [36], it concludes that $\{\hat{\theta}_n\}$, such that $\{M_n(\hat{\theta}_n) = 0\}$, converges in probability to θ_0.

Asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$*:*
Using the results in Gill (1980) [12] and Rivest and Wells (2001) [28], $\sqrt{n}\left[\hat{S}_K(Y_i) - S(Y_i)\right]$ and $\sqrt{n}\left[\hat{S}_Y(Y_i) - S(Y_i)\right]$ can be represented as the summation of iid random functions respectively such that:

$$\begin{aligned}\sqrt{n}\left[\hat{S}_K(Y_i) - S(Y_i)\right] &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[-S(Y_i)\int_0^{Y_i}\frac{dM_{2i}(u)}{P(Y > u, C > u)}\right] + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} h_{ij}^{(K)} + o_p(1)\end{aligned} \tag{3.71}$$

where

$$N_{2i}(u) = I\{Y_i \leqslant u, Y_i < C_i\}, \tag{3.72}$$

$$\lambda_2(t) = \frac{-\frac{\partial}{\partial u} P(Y \geqslant u, C \geqslant t)|_{u=t}}{P(Y \geqslant t, C \geqslant t)}, \tag{3.73}$$

and

$$M_{2i}(u) = N_{2i}(u) - \int_0^u I\{S \geqslant w\} \lambda_2(w) dw. \tag{3.74}$$

Then

$$\begin{aligned} &\sqrt{n}\left[\hat{S}_Y(t) - S(t)\right] = \frac{1}{\sqrt{n}\psi'_\theta\left[S_Y(t)\right]} \times \\ &\sum_{i=1}^{n} \left\{ \int_0^t \frac{-\psi'_\theta\left[\pi(u)\right]}{P(C > u)} dM_i(u) + \int_0^t \Phi'\left[\pi(u)\right] \left[-\pi(u) \int_0^u \frac{dM_{1i}(w)}{\pi_1(w)} \right] \lambda^\sharp(u) du \right\} \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} h_{ij}^{(Y)} + o_p(1) \end{aligned} \tag{3.75}$$

respectively, where $h_{ij}^{(K)}$ and $h_{ij}^{(Y)}$ are corresponding influence functions. Using the Taylor expansion at $\theta = \theta_0$, the estimating equation can be expressed as:

$$
\begin{aligned}
0 =& \sum_{i=1}^{n} \eta_i \frac{\partial \hat{S}_Y}{\partial \theta}(Y_i) \left[\hat{S}_Y(Y_i) - \hat{S}_K(Y_i) \right] \\
=& \sum_{i=1}^{n} \eta_i \frac{\partial \hat{S}_Y}{\partial \theta}(Y_i) \left[\hat{S}_Y(Y_i) - \hat{S}_K(Y_i) \right] \\
&+ \sum_{i=1}^{n} \eta_i \left\{ \left[\frac{\partial \hat{S}_Y}{\partial \theta}(Y_i) \right]^2 + \frac{\partial^2 \hat{S}_Y}{\partial \theta^2}(Y_i) \left[\hat{S}_Y(Y_i) - \hat{S}_K(Y_i) \right] \right\} (\hat{\theta}_n - \theta_0) + o_p(1) \\
=& \sum_{i=1}^{n} \eta_i \frac{\partial \hat{S}_Y}{\partial \theta}(Y_i) \left[\hat{S}_Y(Y_i) - \hat{S}_K(Y_i) \right] + \sum_{i=1}^{n} \eta_i \left[\frac{\partial \hat{S}_Y}{\partial \theta}(Y_i) \right]^2 (\hat{\theta}_n - \theta_0) + o_p(1).
\end{aligned}
\tag{3.76}
$$

Because $\hat{S}_Y(Y_i) - \hat{S}_K(Y_i)$ converges to 0 in probability uniformly when $n \to \infty$ and the derivatives of $\hat{S}_Y$ with respect to θ are all bounded, then

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_n - \theta_0) = \\
&\frac{\frac{1}{n} \sum_{i=1}^{n} \eta_i \left[\frac{\partial \hat{S}_Y}{\partial \theta}(Y_i) \right] \left\{ \sqrt{n} \left[\hat{S}_K(Y_i) - S(Y_i) \right] - \sqrt{n} \left[\hat{S}_Y(Y_i) - S(Y_i) \right] \right\}}{\frac{1}{n} \sum_{i=1}^{n} \eta_i \left[\frac{\partial \hat{S}_Y}{\partial \theta}(Y_i) \right]^2} + o_p(1).
\end{aligned}
\tag{3.77}
$$

By using (3.71) and (3.75) and the fact that $\frac{\partial \hat{S}_Y}{\partial \theta} \to \frac{\partial S_Y}{\partial \theta}$ in probability uniformly when $n \to \infty$, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\frac{1}{n\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i \left[\frac{\partial \hat{S}_Y}{\partial \theta}(Y_i)\right] \left[h_{ij}^{(K)} - h_{ij}^{(Y)}\right]}{\frac{1}{n} \sum_{i=1}^{n} \eta_i \left[\frac{\partial \hat{S}_Y}{\partial \theta}(Y_i)\right]^2}. \tag{3.78}$$

The numerator is asymptotic normal as it can be written as a second order U statistic with some fixed variance σ^2. The denominator converges to a constant:

$$\gamma = E\left[\eta_i \left\{\left[\frac{\partial S_Y}{\partial \theta}(Y_i)\right]^2\right\}\right] \tag{3.79}$$

using the Strong Law of Large Numbers (SLLN).

In summary, it concludes that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotic normal with zero mean and variance $\frac{\sigma^2}{\gamma^2}$. □

After $\hat{\theta}_n$ is obtained, the marginal survival function of X can be consistently estimated by the copula-graphic estimator which was derived in Section 3.2 of this dissertation.

3.4 Model Selection

A very important issue in modelling the semi-competing risks data is the model selection. In previous work, this issue has not been addressed adequately. It turns

out that $Q_n(\theta)$ defined in Section 3.3 can be used as a model selection criterion. The best Archimedean copula model can be chosen to minimize the corresponding $Q_n(\theta)$ value. The model selection procedure is described as follow.

Suppose that there are several possible families of Archimedean copula models to fit a semi-competing risks data. For each family, the unknown parameter can be estimated by minimizing corresponding $Q_n(\theta)$ defined in Section 3.3. The family of models producing the smallest $Q_n(\theta)$ will be selected as the best model for analyzing the data set.

3.5 Accomodation of Covariates

The method given in Section 3.4 can also accommodate covariates. Suppose that there is a dichotomous covariate U such as the patient age group ($U = 1$ representing young age and $U = 2$ representing old age). The best models can be selected and fitted using the proposed strategy and the corresponding model parameters can be estimated in each age group. When a covariate is continuous, the covariate U could be categorized in some way to have enough observations in each category, then the proposed analyses can be performed.

The method given in Section 3.4 can also be modified to accommodate multiple covariates. The regression analysis can be started from the copula-graphic estimator then the proposed analyses can be performed.

3.6 Simulation Studies

To evaluate the performance of the proposed estimator, the simulation studies in different scenarios are given. Under the Houggard model with sample size $N = 500$, the DFS/PFS time X and OS time Y are simulated from a Hougaard copula [13, 14] with parameter $\beta = 0.8, 0.6, 0.4$ and 0.2 corresponding to Kendall's $\tau = 0.2, 0.4, 0.6$ and 0.8 respectively. The marginal distribution of X and Y are assumed

to be exponential distribution with rate 1. The censoring time C is independent of X and Y and generated from an exponential distribution with rates $\lambda = 0.33, 0.50, 0.67$ and 1.00 respectively. In these settings, about 54% (censoring rate $\lambda = 1.00$) to 83% of overall survival time Y (censoring rate $\lambda = 0.33$) can be observed. In each scenario, 1000 replications were performed for the same procedure and compare the estimated Kendall's tau with the true value. The bootstrap standard deviations of the estimated Kendall's tau were compared with the empirical standard deviations based on the 1000 replications.

In the simulations studies, the proposed estimators (denoted by $\hat{\tau}$) were also compared with the estimators proposed by Lakhal, Rivest and Abdous (2008) [21] (denoted by $\tilde{\tau}$) for the Hougaard model [13, 14]. The empirical (bootstrap) standard deviations of $\hat{\tau}$ and $\tilde{\tau}$ are denoted by $\widehat{SD}$ and $\widetilde{SD}$ accordingly in Tables 3.1 and 3.2.

The simulation results have shown that the mean values of the proposed parameter estimates are very close to the true values even when the censoring proportions are high. Overall the proposed estimator works very well under the Hougaard model assumption [13, 14] even when the DFS/PFS time X and OS time Y are both heavily censored.

Also from Tables 3.1 and 3.2, it concludes that the proposed estimator is less biased and outperforms the estimators proposed by Lakhal, Rivest and Abdous (2008) [21] under the Hougaard model assumption [13, 14]. Because the estimators proposed by Lakhal, Rivest and Abdous (2008) [21] are established based on a complicated estimating equation developed from the conditional tau, the proposed estimators seem to be simpler and more efficient for the Hougaard model [13, 14].

Once the dependence level between the competing risks has been determined using the proposed estimator, the marginal survival functions of X can be consistently estimated using the Wang estimator (2014) [39] or the copula-graphic estimator proposed by Zheng and Klein (1995) [47] and Rivest and Wells (2001) [28].

Table 3.1 Performance of $\hat{\tau}$ and $\tilde{\tau}$ for the Hougaard Model Based on Kendall's $\tau = 0.2, 0.4, 0.6, 0.8$ in Different Censoring Proportions (Scenario $1-8$). Here Kendall's τ is the True Value of the Parameter. $\hat{\tau}$ is the Proposed Parameter Estimator and $\tilde{\tau}$ is the Estimator Proposed by Lakhal, Rivest and Abdous (2008) [21].

Sample Size	N=500							
Replication	M=1000							
Scenario	**Censoring Rate True**	**Corresponding Kendall's True**	**Proposed Kendall's Estimate**	**Lakhal Kendall's Estimate**	**Proposed Empirical Estimate**	**Proposed Bootstrap Estimate**	**Lakhal Empirical Estimate**	**Lakhal Bootstrap Estimate**
Number	λ	τ	$\hat{\tau}$	$\tilde{\tau}$	$\widehat{SD}$	$\widehat{SD}$	$\widetilde{SD}$	$\widetilde{SD}$
1	0.33	0.20	0.19	0.10	0.048	0.048	0.036	0.036
2	0.50	0.20	0.19	0.12	0.049	0.049	0.042	0.044
3	0.67	0.20	0.19	0.15	0.050	0.051	0.049	0.051
4	1.00	0.20	0.19	0.19	0.054	0.053	0.063	0.064
5	0.33	0.40	0.39	0.22	0.041	0.042	0.034	0.035
6	0.50	0.40	0.39	0.27	0.043	0.044	0.038	0.041
7	0.67	0.40	0.39	0.32	0.044	0.046	0.045	0.046
8	1.00	0.40	0.39	0.40	0.048	0.049	0.050	0.052

Table 3.2 Performance of $\hat{\tau}$ and $\tilde{\tau}$ for the Hougaard Model Based on Kendall's $\tau = 0.2, 0.4, 0.6, 0.8$ in Different Censoring Proportions (Scenario $9 - 16$). Here Kendall's τ is the True Value of the Parameter. $\hat{\tau}$ is the Proposed Parameter Estimator and $\tilde{\tau}$ is the Estimator Proposed by Lakhal, Rivest and Abdous (2008) [21].

Sample Size	N=500							
Replication	M=1000							
Scenario	**Censoring Rate True**	**Corresponding Kendall's True**	**Proposed Kendall's Estimate**	**Lakhal Kendall's Estimate**	**Proposed Empirical Estimate**	**Proposed Bootstrap Estimate**	**Lakhal Empirical Estimate**	**Lakhal Bootstrap Estimate**
Number	λ	τ	$\hat{\tau}$	$\tilde{\tau}$	$\widehat{SD}$	$\widehat{SD}$	$\widetilde{SD}$	$\widetilde{SD}$
9	0.33	0.60	0.59	0.38	0.031	0.032	0.035	0.034
10	0.50	0.60	0.59	0.45	0.034	0.034	0.038	0.037
11	0.67	0.60	0.59	0.51	0.035	0.037	0.039	0.039
12	1.00	0.60	0.59	0.60	0.038	0.040	0.038	0.039
13	0.33	0.80	0.79	0.60	0.020	0.020	0.031	0.032
14	0.50	0.80	0.79	0.68	0.021	0.022	0.029	0.030
15	0.67	0.80	0.79	0.73	0.022	0.023	0.026	0.027
16	1.00	0.80	0.79	0.80	0.026	0.027	0.023	0.023

Table 3.3 Selection Percentages for Data from the Clayton Model

Sample Size	N=500			
Replication	M=1000			
	True Model: Clayton			
Fitting Model	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
Clayton	79.8%	95.7%	99.1%	99.5%
Hougaard	1.2%	0.0%	0.0%	0.0%
Frank	19.0%	4.3%	0.9%	0.5%

It is worth mentioning that the simulation studies for the Clayton model [3] have also been conducted. It finds that the proposed estimators and the estimators proposed by Lakhal, Rivest and Abdous (2008) [21] are comparable under the Clayton model [3] (the simulation results for the Clayton model are omitted here).

The simulations have also been conducted to evaluate the proposed model selection procedure based on $Q_n(\theta)$. The DFS/PFS time X and OS time Y from three different Archimedean copula models have been simulated with parameter values corresponding to Kendall's $\tau = 0.2, 0.4, 0.6$ and 0.8 respectively. The marginal distribution of X and Y are assumed to be exponential distribution with rate 1 and the censoring time C is independent of X and Y from an exponential distribution with rate $\lambda = 0.33$. Tables 3.3, 3.4 and 3.5 present the selection percentages according to different τ values.

From these results, it concludes that the percentages of selection increase when τ values increase if the correct model was fitted. The percentages of selection decrease when τ values increase if the incorrect model was fitted. Overall, the model selection procedure works quite well.

Table 3.4 Selection Percentages for Data from the Hougaard Model

Sample Size	N=500			
Replication	M=1000			
	True Model: Hougaard			
Fitting Model	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
Clayton	2.8%	0.0%	0.0%	0.0%
Hougaard	79.3%	91.0%	97.3%	98.8%
Frank	17.9%	9.0%	2.7%	1.2%

Table 3.5 Selection Percentages for Data from the Frank Model

Sample Size	N=500			
Replication	M=1000			
	True Model: Frank			
Fitting Model	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$
Clayton	2.4%	0.6%	0.3%	0.4%
Hougaard	12.5%	5.6%	2.4%	1.3%
Frank	85.1%	93.8%	97.3%	98.3%

Table 3.6 $Q_n(\theta)$ Value for the Leukemia Data Set (Included in R Package KMsurv).

Sample Size	N=137
Fitting Model	$\mathbf{Q_n}(\theta)$
Clayton	0.004
Hougaard	0.014
Frank	0.005

3.7 An Illustrative Example

Using the proposed approach, the leukemia data set used in Fine, Jiang and Chappel (2001) [5] and also Wang, Chandra, Xu and Sun (2015) [41] was fitted. The data set is included in R package KMsurv and the details about it can be found in the following web site: https://cran.r-project.org/web/packages/KMsurv/KMsurv.pdf.

Applying the proposed model selection procedure for this semi-competing risks data, it concludes that the Clayton model [3] has a $Q_n(\theta)$ value 0.004 which is smaller than the $Q_n(\theta)$ values under the Hougaard model [13, 14] and the Frank model [7], see Table 3.6. The Clayton model [3] has been chosen to fit this data set based on the model selection procedure proposed in Section 3.4.

Under the Clayton model assumption [3], the estimated association parameter is $\hat{\theta} = 6.9$ with a bootstrap standard deviation $\widehat{SD}_\theta = 1.81$. The corresponding tau estimate is $\hat{\tau}$ with a bootstrap standard deviation $\widehat{SD}$. The comparison results are summarized in Table 3.7. The result based on the proposed procedure is close to the tau estimate obtained using the approach in Fine, Jiang and Chappel (2001) [5] under the Clayton model assumption [3].

Next, the data set can be categorized by the age variable into two categories using the median value as the cut-off point. For the young age group, the Frank model [7] is the best model based on the proposed strategy. For the old age group,

Table 3.7 Parameters for the Leukemia Data Set (Included in R Package KMsurv). Here the Proposed Parameter Estimator is Compared to the Estimator Proposed by Fine, Jiang and Chappel (2001) [5].

Sample Size	N=137	
Fitting Model	Clayton	
Estimator	**Kendall's** $\hat{\tau}$	**Bootstrap** $\widehat{SD}$
Proposed	0.78	0.04
Fine, 2001	0.80	0.04

Table 3.8 Parameters for the Leukemia Data Set (Included in R Package KMsurv) by Variable Age (Using the Median Value as the Cut-off Point).

Sample Size	N=137	
Fitting Model	Frank	
Variable Age	**Kendall's** $\hat{\tau}$	**Bootstrap** $\widehat{SD}$
Young	0.662	0.09
Old	0.726	0.08

the Frank model [7] is also the best model. Results summarized in Table 3.8. The dependence between the DFS and the OS seems stronger for older patients.

3.8 Discussion

In this dissertation, a copula-graphic estimator has been derived for marginal survival functions of failure times based on semi-competing risks data. The uniform consistency and the weak convergence of the copula-graphic estimator have been proved.

Based on the copula-graphic estimator of the survival function, a new strategy to estimate the unknown parameter has been proposed in Archimedean copula models based on semi-competing risks data. The proposed estimation strategy in this dissertation is simpler and more general than the method proposed in Fine, Jiang and Chappel (2001) [5] as the proposed strategy in this dissertation can be applied under different Archimedean copula model assumptions while the method proposed by Fine, Jiang and Chappel (2001) [5] can only be applied under the Clayton model assumption [3].

From the simulation studies in this dissertation, it finds that the proposed strategy also tends to be simpler and more effective than the strategy proposed by Lakhal, Rivest and Abdous (2008) [21] for the Hougaard model [13, 14].

Another main advantage of applying the proposed estimation strategy in this dissertation is the effectiveness of the proposed model selection procedure. The proposed model selection procedure in this dissertation is new and important for semi-competing risks data because not much research has been done to select the best copula model before this work.

When there are covariates in a semi-competing risks data, applying the proposed estimation and model selection procedures still work after the categorization of the covariates as long as there are enough patients in each category as shown in the data analysis example in Section 3.7.

Alternatively, applying a strategy proposed by Wang et al. (2015) [41, 40] using frailty models also can estimate the dependence level between competing risks. It is worth mentioning that the data structure setting in Wang et al. (2015) [41] is a little different from the setting in this dissertation: here in this dissertation, Y can always be observed unless it is censored by an independent censoring time C. Meanwhile in Wang et al. (2015) [41], Y can also be censored by a dependent censoring time,

the extra information provided by some covariates in that setting can be applied to determine the dependence levels between competing event times effectively.

REFERENCES

[1] K. Aas and D. Berg. Models for construction of multivariate dependence – a comparison study. *The European Journal of Finance*, 15(7-8):639–659, 2009.

[2] A. Chakak. *Some methods of constructing multivariate distributions*. Iowa City, IA: Iowa State University, 1993.

[3] D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151, 1978.

[4] P. Embrechts, F. Lindskog, and A. McNeil. *Modelling Dependence with Copulas and Applications to Risk Management*. Zürich, Switzerland: Technical Report, Department of Mathematics, ETH Zürich, 2001.

[5] J. P. Fine, H. Jiang, and R. Chappell. On semi-competing risks data. *Biometrika*, 88(4):907–919, 2001.

[6] T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. Hoboken, NJ: John Wiley & Sons, Inc., 1991.

[7] C. Genest. Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555, 1987.

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

[9] C. Genest, J-F. Quessy, and B. Rémillard. Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics*, 33(2):337–366, 2006.

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

[11] P. Georges, A-G. Lamy, E. Nicolas, G. Quibel, and T. Roncalli. Multivariate survival modelling: A unified approach with copulas. *SSRN Electronic Journal*, 2001.

[12] R. D. Gill. Censoring and stochastic integrals. *Statistica Neerlandica*, 34(2):124–124, 1980.

[13] E. J. Gumbel. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292):698–707, 1960.

[14] P. Hougaard. A class of multivanate failure time distributions. *Biometrika*, 73(3):671–678, 1986.

[15] H. Joe. *Multivariate Models and Multivariate Dependence Concepts*. London, England; New York, NY: Chapman & Hall, 1997.

[16] J. P. Klein and M. L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. New York, NY: Springer-Verlag, 2003.

[17] D. Kurowicka and R. M. Cooke. Distribution-free continuous bayesian belief nets. *Modern Statistical and Mathematical Methods in Reliability*, pages 309–322, 2005.

[18] D. Kurowicka and R. M. Cooke. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Hoboken, NJ: John Wiley & Sons, Inc., 2006.

[19] S. W. Lagakos. A stochastic model for censored-survival data in the presence of an auxiliary variable. *Biometrika*, 32(3):551–559, 1976.

[20] S. W. Lagakos. Using auxiliary variables for improved estimates of survival time. *Biometrika*, 33(2):399–404, 1977.

[21] L. Lakhal, L-P. Rivest, and B. Abdous. Estimating survival and association in a semicompeting risks model. *Biometrics*, 64(1):180–188, 2008.

[22] R. B. Nelsen. *An Introduction to Copulas*. New York, NY: Springer-Verlag, 2006.

[23] D. Oakes. Semiparametric inference in a model for association in bivariate survival data. *Biometrika*, 73(2):353–361, 1986.

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

[25] D. Oakes. *Frailty Models in Survival Analysis*. Lecture Notes, Rochester, NY: University of Rochester, 1998.

[26] F. O'Reilly and C. P. Quesenberry. The conditional probability integral transformation and applications to obtain composite chi-square goodness-of-fit tests. *The Annals of Statistics*, 1(1):74–83, 1973.

[27] M. S. Pepe. Inference for events with dependent risks in multiple endpoint studies. *Journal of the American Statistical Association*, 86(415):770–778, 1991.

[28] L-P. Rivest and M.T. Wells. A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *Journal of Multivariate Analysis*, 79(1):138–155, 2001.

[29] M. Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, 1952.

[30] C. Savu and M. Trede. Hierarchies of archimedean copulas. *Quantitative Finance*, 10(3):295–304, 2010.

[31] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. North Holland, Netherlands: Elsevier Science Ltd., 1983.

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

[33] J. H. Shih. A goodness-of-fit test for association in a bivariate survival model. *Biometrika*, 85(1):189–200, 1998.

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

[35] S. Suciu, A. Eggermont, P. Lorigan, J. Kirkwood, S. Markovic, C. Garbe, D. Cameron, S. Kotapati, T. Chen, K. Wheatley, N. Ives, G. Schaetzen, A. Efendi, and M. Buyse. Relapse-free survival as a surrogate for overall survival in the evaluation of stage ii-iii melanoma adjuvant therapy. *Journal of the National Cancer Institute*, 110(1):87–96, 2018.

[36] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge, England: Cambridge University Press, 1998.

[37] A. Wang. The analysis of bivariate truncated data using the clayton copula model. *The International Journal of Biostatistics*, 3(8), 2007.

[38] A. Wang. Goodness-of-fit tests for archimedean copula models. *Statistica Sinica*, 20(1):441–453, 2010.

[39] A. Wang. Properties of the marginal survival functions for dependent censored data under an assumed archimedean copula. *Journal of Multivariate Analysis*, 129:57–68, 2014.

[40] A. Wang, K. Chandra, and X. Jia. The analysis of left truncated bivariate data using frailty models. *Scandinavian Journal of Statistics*, 45:847–860, 2018.

[41] A. Wang, K. Chandra, R. Xu, and J. Sun. The identifiability of dependent competing risks models induced by bivariate frailty models. *Scandinavian Journal of Statistics*, 42:427–437, 2015.

[42] A. Wang, X. Jia, and Z. Jin. Estimation of the cumulative baseline hazard function for dependently right-censored failure time data. *Journal of Applied Statistics*, 2020.

[43] A. Wang, Y. Zhang, and W. Shao. On the likelihood of mixture cure models. *Statistics & Probability Letters*, 131:51–55, 2017.

[44] W. Wang and M. T. Wells. Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association*, 95(449):62–72, 2000.

[45] N. Whelan. Sampling from archimedean copulas. *Quantitative Finance*, 4(3):339–352, 2004.

[46] L. Zhang and V. P. Singh. *Copulas and their Applications in Water Resources Engineering*. Cambridge, England: Cambridge University Press, 2019.

[47] M. Zheng and J. P. Klein. Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika*, 82(1):127–138, 1995.