ASSOCIATION ANALYSIS FOR MULTIVARIATE TIME-TO-EVENT DATA

by

Hao Wang

B.S. Applied Mathematics, Beijing Institute of Technology, Beijing, China 2004

M.S. Mathematical Statistics, Beijing Institute of Technology, Beijing, China 2006

M.A. Applied Statistics, University of Pittsburgh, Pittsburgh, PA 2008

Submitted to the Graduate Faculty of
the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF STATISTICS

This dissertation was presented

by

Hao Wang

It was defended on

December 8th 2011

and approved by

Dr. Satish Iyengar, Department of Statistics

Dr. Yu Cheng, Deptartment of Statistics

Dr. Leon Gleser, Department of Statistics

Dr. John Chadam, Department of Mathematics

Dissertation Advisors: Dr. Satish Iyengar, Department of Statistics,

Dr. Yu Cheng, Deptartment of Statistics

ASSOCIATION ANALYSIS FOR MULTIVARIATE TIME-TO-EVENT DATA

Hao Wang, PhD

University of Pittsburgh, 2011

Association analyses are performed for two types of multivariate time-to-event data: multivariate clustered competing risks data and bivariate recurrent events data.

In the first part, we extend the bivariate hazard ratio [Cheng and Fine, 2008] to multivariate competing risks data and show it is equivalent to the cause-specific cross hazard ratio in Cheng et al. [2010]. Two nonparametric approaches are proposed. One extends the plug-in estimator in Cheng and Fine [2008] and the other adapts the pseudo likelihood estimator for bivariate survival data [Clayton, 1978] to multivariate competing risks data. The asymptotic properties are established by using empirical process techniques. We compare the extended plug-in and pseudo likelihood estimators with the existing U statistic [Cheng et al., 2010] by simulations and show that the three methods have comparable performance when no tied events exist. However, the plug-in estimator underestimate and the other two overestimate positive associations in the presence of rounding errors. Hence, we propose a modified U statistic for tied observations, which outperforms the other estimators by simulation studies. All methods are applied to the Cache County Study to examine mother-child and sibship associations in dementia among this aging population. The modified U essentially lies between the plug-in estimate and the original U statistic. We therefore recommend using the straightforward plug-in estimator for untied data, and using the modified U statistic when there are rounding errors.

In the second part, bivariate recurrent events data are modeled by a compound Poisson process, whose dependence structure is then modeled by a Lévy copula. When only the parameter of dependence structure is of primary interest, we proposed two methods to estimate the dependence parameter of the Lévy copula. One uses Kendall's τ assuming the Clayton Lévy copula while the other uses two-stage strategy to propose a semiparametric estimator. Consistency and asymptotic normality are also established. Simulation studies show that the proposed semi-parametric estimator is less efficient than the full likelihood estimator but superior to the nonparametric one. The proposed methods are also applied to Danish fire data to examine the relationship between loss to a building and loss to its contents.

Keywords: Association analysis, multivariate competing risk data, cause-specific cross hazard ratio, bivariate cause-specific hazard ratio, pseudolikelihood estimator, bivariate compound Poisson process, Lévy copula, Lévy process, two-stage estimator.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The research in this thesis focuses on association analysis of two types of data from: survival analysis and stochastic processes. The first type of data is the multivariate competing risks data, an example of which, the Cauchy county data, is shown to me by Dr. Cheng two years ago, when we realized that association analysis of high dimensional case is rarely investigated by other researchers, so that it is a promising direction to develop nonparametric inference methods for such data. The second type of data is the aggregated insurance losses, which can be modeled by a bivariate compound Poisson process. This is motivated by discussions with Dr. Iyengar and Dr. Chadam in financial area three years ago. At that time, I just attended Dr. Iyengar's course "Stochastic process and SDE" for the fundamentals of Brownian motion and Dr. Chadam's course "Stochastic calculus for finance" for applications of stochastic processes in finance. Since I was very interested in this area, I tried to search for research topics and Dr. Chadam recommended Lévy process and the famous book on it by Cont and Tankov [2004]. In this book, I found the Lévy copula and its potential application in the dependence modeling of the high dimensional insurance data.

I'd like to thank the people who have in various way helped me out in the searching of topics, discussions with results and writing of the thesis. Among these are Dr. Iyengar, Dr. Cheng, Dr. Chadam and Dr. Gleser who kindly participate in my PhD. defense committee and provide insights to my research results. I would also like to thank my advisor in China, Dr. Xinzhong Xu, who led to the research of statistics. And finally, thanks to my parents and friends, who's been supporting my research these years.

1.0 INTRODUCTION

We consider association analysis of two types of data: multivariate clustered time-to-event data and bivariate recurrent events time data. The first type arises as failure time data in follow-up epidemiological studies of family members. In this case, subjects are grouped into families leading to dependence within groups, while each subject may experience two or more types of dependent events, of which only the first can be observed; second type of failure time data arises when individual study subjects experience multiple events. Events for different individuals may occur simultaneously and the accompanying quantities are usually dependent. In this case, the co-occurrence rate and relationship between accompanying quantities make up the dependence structure of individuals.

In this thesis, we will focus on the association analysis of two data sets that led to our study. First, we consider a study conducted in Cache County, Utah on dementia in an aging population. One goal of this study was to characterize the association in dementia among family members. For each member of a family, the researcher recorded his or her onset age of dementia, age at death, or age at the termination of the study whichever occurred at first along with an indicator of each type of outcome. This falls into the paradigm of multivariate competing risks, where there is within-subject dependence between the risk due to event of interest and other competing risks as well as between-subject dependence among family members included by common genetic or environmental factors. Multivariate competing risks data virtually occur frequently as in many applications. Association analysis of such data focuses on quantifying the between-subjects dependence of a major event of interest, while accounting for the within-subject dependence.

The second part of this thesis will focus on analyzing the association of bivariate aggregate insurance losses data. McNeil [1997] split the Danish fire insurance claims data into three categories: damage to the building, damage to furniture and personal property, and loss of profits. Each type of insurance loss is made up of arrival times and accompanying losses; the aggregate losses up to a time t is the sum of losses before t. When two or more types of losses are considered, the losses may occur at the same time or independently. This requires the association analysis of recurrent failure times. On the one hand, if all losses occur simultaneously and the accompanying losses are fully dependent, then the aggregate losses of different types are also completely dependent. On the other hand, if the losses never occur together, the aggregate losses of different types are independent. Actual aggregate losses data are usually between these two extreme cases. The special feature of this problem is how to decompose the aggregate losses process and how to quantify the dependence based on frequency of co-occurrences and the relationship between the accompanying loss sizes.

The first part of my dissertation will extend an association measure of bivariate competing risks data to multivariate clustered competing risks data and then propose two non-parametric estimation methods for the multivariate case. The second part models the aggregate losses processes by compound Poisson processes with dependence structure modeled by the Lévy copula. Then a non-parametric estimation based on Kendall's τ and a two-stage semi-parametric estimator method are proposed and compared to the existing full likelihood method. Simulation studies were performed in both parts to assess the proposed methods, which are then applied to the corresponding motivating real data.

2.0 CAUSE-SPECIFIC ASSOCIATION MEASURES FOR MULTIVARIATE COMPETING RISKS DATA

2.1 INTRODUCTION

Many researchers have studied association analysis for bivariate survival data with independent right censoring. An overview of their work was given by Hougaard [2000]. The previous work, however, cannot be applied directly to multivariate competing risks data because the event of interest may be dependently censored by the occurrence of a competing event. One limitation of competing risks data is that the within-subject dependence between the target event and a competing event is nonparametrically untestable, and their marginal distributions are nonparametrically nonidentifiable [Tsiatis, 1975, Pruitt, 1993]. The nonidentifiability issue raises difficulties in analyzing the association in dementia in the presence of possibly dependent censoring by death, which occurred frequently among this aging population.

Some of the association analyses for bivariate survival data with independent censoring have been recently adapted to bivariate competing risks data [Bandeen-Roche and Liang, 2002, Bandeen-Roche and Ning, 2008, Cheng and Fine, 2008]. Bandeen-Roche and Liang [2002] modified Oakes [1989]'s cross hazard ratio to the bivariate competing risks setting. They defined a cause-specific association measure $\theta_{CS}(s, t; k, l)$ which gives the factor by which a child's risk of having cause k event at time s compares if his/her mother has experienced the cause l event by time t and has not yet experienced any event by time t. To estimate $\theta_{CS}(s, t; k, l)$, Bandeen-Roche and Liang [2002] and Bandeen-Roche and Ning [2008] proposed a U-type estimator based on a local version of Kendall's τ with inference based on standard theories for U statistics. Cheng and Fine [2008] defined a ratio $\zeta(s, t; k, l)$ of the bivariate hazard with cause-specific events from both subjects to the product of the bivariate hazards with a single event from each subject as an association measure. They established the equivalence of $\zeta(s,t;k,l)$ to the association measure $\theta_{CS}(s,t;k,l)$, and developed a plug-in estimator with an explicit expression for its variance.

More recently, Shih and Albert [2010] examined familial association in breast cancer based on a cross hazard ratio for paired times of first event and an odds ratio for two competing causes (cancer and non-cancer deaths), and estimated the two association measures by adopting the two-stage estimation method [Shih and Louis, 1995]. They noted that the cause-specific hazard ratio $\theta_{CS}(s,t;k,l)$ may be more clinically relevant than their cross hazard ratio for the first event times of paired members. Because of the instantaneous nature of $\theta_{CS}(s,t;k,l)$, they also suggested examining an association measure based on cumulative incidence functions [Shih and Albert, 2010, Cheng et al., 2007]. In this thesis, we focus on the cause-specific association measures θ and ζ, as the two are essentially a local correlation between indicator functions and hence have an interpretation that is easy to understand. In addition, they do not require estimation of bivariate cumulative incidence functions.

All previous works related to θ_{CS} or ζ focused on bivariate competing risks data except Cheng et al. [2010], who adapted the cause-specific cross hazard ratio θ_{CS} for bivariate data to multiple sibship data and multiple mother-child pairs and greatly broadened its scope of application. In line with this work, we first extend the association measure ζ from bivariate competing risks data to clustered data and develop plug-in estimators as is done in the bivariate case. We next adapt a pseudo likelihood estimator for bivariate survival data [Clayton, 1978, Oakes, 1982] to multivariate competing risks data. The extended plug-in estimators and pseudo maximum likelihood estimators are compared with the U-statistics given in Cheng et al. [2010] via simulation studies. In this thesis we also investigate the impact of rounding errors on the performance of the three estimators. A modified U statistic is proposed to specifically handle tied event data.

2.2 METHODS

2.2.1 Multivariate Competing Risks Data

In this thesis, we consider clustered competing risks data. Suppose that in each cluster there are $M+1$ failure times $(T_0, T_1, \ldots, T_M)$ and their corresponding cause indicators $(\epsilon_0, \epsilon_1, \ldots, \epsilon_M)$ from a mother and her M children, where M is random. We only consider two causes here $\epsilon_j = 1, 2, j = 0, \ldots, M$, as we are primarily interested in association in cause 1 events and can always group multiple competing events together. In addition, there may also be independent censoring times $(C_0, C_1, \ldots, C_M)$. Hence one observes $Y_j = \min(T_j, C_j)$, and $\eta_j = \epsilon_j I\{T_j > C_j\}$, for $j = 0, 1, \ldots, M$, where I is the indicator function. This notation incorporates two types of multivariate competing risks data. One is the sibship data, where pairwise exchangeability can be assumed among siblings, i.e., all siblings have the same marginals and the pairwise joint densities are symmetric. The other consists of multiple mother and child pairs where we do not impose the exchangeability assumption between a mother and her child, but assume exchangeability among all children. In this thesis, we use the maternal data as an example; the same method can be readily applied to quantify the association between a father and his children. The observed data containing n i.i.d. clusters are denoted as $\{(Y_{i0}, \ldots, Y_{i,m_i}, \eta_{i0}, \ldots, \eta_{i,m_i}, m_i), i = 1, \ldots, n\}$. The following two association measures are considered for each of the two types of multivariate competing risks data.

2.2.2 Cause-specific Cross Hazard-ratio for Clustered Data

For mother and all children data, under the exchangeability assumption for the children, the child-mother cause-specific density is defined as

$$g(s, t; k, l) = \lim_{(\Delta s, \Delta t) \downarrow 0} \frac{1}{\Delta s \Delta t} \times P(s \le T_j \le s + \Delta s, t \le T_0 \le t + \Delta t, \epsilon_j = k, \epsilon_0 = l),$$

for $k, l = 1, 2$ and $j = 1, \ldots, M$. Assuming g is absolutely continuous, a child-mother pair (T_j, T_0) has a joint overall survival function

$$G(s, t) = \int_s^\infty \int_t^\infty \sum_{k=1}^{2} \sum_{l=1}^{2} g(u, v; k, l) du dv.$$

Cheng et al. [2010] defined a child-mother cause-specific association measure denoted by $\theta_{CP}(s,t;k,l)$ as:

$$\theta_{CP}(s,t;k,l) = \frac{G(s,t)g(s,t;k,l)}{\{\int_s^\infty \sum_{m=1}^2 g(u,t;m,l)du\}\{\int_t^\infty \sum_{h=1}^2 g(s,v;k,h)dv\}}, \tag{2.1}$$

which shares a similar interpretation as the conditional cross hazard-ratio proposed by Bandeen-Roche and Liang [2002]. That is, the risk of a child experiencing a cause k event by time s would be accelerated for $\theta_{CP} > 1$ or decelerated for $\theta_{CP} < 1$ had his mother developed a cause l event by time t as compared to the case that his mother had no event by time t. This extended conditional cross hazard-ratio was estimated piecewisely by a U statistic [Cheng et al., 2010]. In this thesis, we will extend an equivalent association measure and its plug-in estimator to clustered competing risks data and also propose a pseudo likelihood estimator and compare it with this U statistic. We now briefly describe the notation and methods given in Cheng et al. [2010]. Similar notation will be used in this paper and a modified U statistic will be proposed in section 2.3.2 based on the following development.

Let $Y_{i,a}, Y_{i,b}, 1 \le a < b \le m_i$ be the failure times of two distinct members of the ith cluster, and $Y_{j,c}, Y_{j,d}, 1 \le c < d \le m_j$ be the failure times from the jth cluster. Cheng et al. [2010] divided the time scale as follows: $0 = \tau_{1,0} < \tau_{1,1} < \tau_{1,2} < \cdots < \tau_{1,n_1} = \tau_1$, $0 = \tau_{2,0} < \tau_{2,1} < \tau_{2,2} < \cdots < \tau_{2,n_2} = \tau_2$ and assumed that $\theta_{CP}(s,t;k,l)$ be a constant over each grid $\Omega_{qr} \equiv (\tau_{1,q}, \tau_{1,q+1}) \times (\tau_{2,r}, \tau_{2,r+1})$ with $0 \le q \le n_1 - 1, 0 \le r \le n_2 - 1$. They defined $Y_{(iajc)} = \min(Y_{ia}, Y_{jc}), \eta_{(iajc)} = \eta_{ia} I(Y_{ia} < Y_{jc}) + \eta_{jc} I(Y_{ia} > Y_{jc})$ and similarly $Y_{(ibjd)}$ and $\eta_{(ibjd)}$, and defined the concordant indicator

$$\phi_{ij,acbd}^{qr} = I\{(Y_{ia} - Y_{jc})(Y_{ib} - Y_{jd}) > 0, (Y_{(iajc)}, Y_{(ibjd)}) \in \Omega_{qr}, (\eta_{(iajc)}, \eta_{(ibjd)}) = (k,l)\},$$

and the discordant indicator

$$\psi_{ij,acbd}^{qr} = I\{(Y_{ia} - Y_{jc})(Y_{ib} - Y_{jd}) < 0, (Y_{(iajc)}, Y_{(ibjd)}) \in \Omega_{qr}, (\eta_{(iajc)}, \eta_{(ibjd)}) = (k,l)\}.$$

The extended child-mother association measure was estimated by

$$\hat{\theta}_{CP}^U(\Omega_{qr};k,l) = \frac{\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \sum_{1\le a\le m_i} \sum_{1\le c\le m_j} \phi_{ij,ac00}^{qr}}{\binom{n}{2}^{-1} \sum_{1\le i<j\le n} \sum_{1\le a\le m_i} \sum_{1\le c\le m_j} \psi_{ij,ac00}^{qr}} \tag{2.2}$$

which was the ratio of the number of concordant pairs to the number of discordant pairs, amongst those pairs where concordance status was determinable, and the times and causes of failure matched those in θ_{CP}.

For any $1 \leq j < j^{'} \leq M$ in the sibship data, Cheng et al. [2010] defined the bivariate cause-specific densities $f_{jj'}(s,t;k,l) = \lim_{(\Delta s,\Delta t)\downarrow 0} P(s \leq T_j \leq s+\Delta s, \epsilon_j = k, t \leq T_{j'} \leq t+\Delta t, \epsilon_{j'} = l)/(\Delta s \Delta t)$, for $k,l = 1,2$ and the overall survival function $S_{jj'}(s,t) = \int_s^\infty \int_t^\infty \sum_{k=1}^2 \sum_{l=1}^2 f_{jj'}(u,v;k,l)dudv$ assuming that $f_{jj'}$ is absolutely continuous. The subscripts j and $j^{'}$ were dropped under the exchangeability assumption among siblings, so that $f(s,t;k,l) = f(t,s;l,k)$ for any pair of causes k,l at any time point (s,t). The sibship cause-specific association measure assuming exchangeability was defined as

$$\theta_{CC}(s,t;k,l) = \frac{f(s,t;k,l)}{\int_s^\infty \sum_{m=1}^2 f(u,t;m,l)du} \frac{S(s,t)}{\int_t^\infty \sum_{h=1}^2 f(s,v;k,h)dv}. \tag{2.3}$$

The estimation of the sibship association θ_{CC} was more complicated than that of the child mother association θ_{CP}, as under pairwise exchangeability it was meaningful to define concordance and discordance indicators based on two different pairings (Y_{ia}, Y_{jc}) with (Y_{ib}, Y_{jd}) as well as (Y_{ia}, Y_{jd}) with (Y_{ib}, Y_{jc}) [Fine and Jiang, 2000]. Hence θ_{CC} was estimated by

$$\hat{\theta}^U_{CC}(\Omega_{qr};k,l) = \frac{\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \sum_{1\leq a<b\leq m_i} \sum_{1\leq c<d\leq m_j} (\phi^{qr}_{ij,acbd} + \phi^{qr}_{ij,adbc})}{\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \sum_{1\leq a<b\leq m_i} \sum_{1\leq c<d\leq m_j} (\psi^{qr}_{ij,acbd} + \psi^{qr}_{ij,adbc})}. \tag{2.4}$$

The estimators (2.2) and (2.4) are U statistics whose asymptotic properties can be obtained with U-statistic theories [Cheng et al., 2010].

2.2.3 Bivariate Cause-specific Hazard Ratio for Clustered Data and its Plug-in Estimator

Cheng and Fine [2008] proposed a cause-specific association measure based on bivariate cause-specific hazard functions and demonstrated its equivalence to the conditional cross hazard-ratio [Bandeen-Roche and Liang, 2002, Bandeen-Roche and Ning, 2008]. We now extend their association measure to clustered competing risks data. To simplify the notation we focus on association analysis of cause 1 events, though the methods proposed here are applicable to other causes. Analogous to Cheng and Fine [2008], for any child-mother pair

$(T_j, T_0), 1 \leq j \leq M$, we define a bivariate cause-specific hazard function for double cause 1 events from both subjects under the exchangeability assumption for all children

$$\lambda_{11}(s,t) = \lim_{(\Delta s, \Delta t)\downarrow 0} P(T_j \in [s, s+\Delta s], \epsilon_j = 1, T_0 \in [t, t+\Delta t], \epsilon_0 = 1 | T_j \geq s, T_0 \geq t)/(\Delta s \Delta t),$$

and define bivariate hazard functions of a single cause 1 event from the child or mother with the other member being at risk as the following:

$$\lambda_{10}(s,t) = \lim_{\Delta s \downarrow 0} P(T_j \in [s, s+\Delta s], \epsilon_j = 1 | T_j \geq s, T_0 \geq t)/\Delta s, \text{ and}$$

$$\lambda_{01}(s,t) = \lim_{\Delta t \downarrow 0} P(T_0 \in [t, t+\Delta t], \epsilon_0 = 1 | T_j \geq s, T_0 \geq t)/\Delta t.$$

Then the association measure for cause 1 events based on these bivariate hazard functions is defined as

$$\zeta_{CP}(s,t;1,1) = \frac{\lambda_{11}(s,t)}{\lambda_{10}(s,t)\lambda_{01}(s,t)}. \tag{2.5}$$

A simple plug-in estimator of $\zeta_{CP}(s,t;1,1)$ can be constructed analogously to Cheng and Fine (2008) based on all child-mother pairs. We first need to extend the event processes to the more complicated family structure with multiple children in a family. Let $N_{11}^*(s,t) = \sum_j I(Y_j \leq s, \eta_j = 1, Y_0 \leq t, \eta_0 = 1)$, $N_{10}^*(s,t) = \sum_j I(Y_j \leq s, \eta_j = 1, Y_0 \geq t)$, $N_{01}^*(s,t) = \sum_j I(Y_j \geq s, Y_0 \leq t, \eta_0 = 1)$, $H^*(s,t) = \sum_j I(Y_j \geq s, Y_0 \geq t)$. Assuming constant association measure $\zeta_{CP}(\Omega_{qr})$ over the region Ω_{qr}, we have

$$\int_{(s,t)\in\Omega_{qr}} w(s,t)\{EN_{11}^*(ds,dt)EH^*(s,t) - \zeta_{CP}(\Omega_{qr})EN_{10}^*(ds,t)EN_{01}^*(s,dt)\} = 0,$$

where E is the expectation operator and $w(s,t)$ is some bounded deterministic function. Define the empirical process $\mathbb{P}_n N_{11}^*(s,t) = \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^{m_i} I(Y_{ij} \leq s, Y_{i0} \leq t, \eta_{ij} = 1, \eta_{i0} = 1)$, and similarly for $\mathbb{P}_n H^*(s,t)$, $\mathbb{P}_n N_{10}^*(s,t)$ and $\mathbb{P}_n N_{01}^*(s,t)$. Plugging the empirical processes into the above equation yields a plug-in estimator

$$\hat{\zeta}_{CP}(\Omega_{qr}) = \frac{\int_{(s,t)\in\Omega_{qr}} \hat{w}(s,t)\mathbb{P}_n N_{11}^*(ds,dt)\mathbb{P}_n H^*(s,t)}{\int_{(s,t)\in\Omega_{qr}} \hat{w}(s,t)\mathbb{P}_n N_{10}^*(ds,t)\mathbb{P}_n N_{01}^*(s,dt)}, \tag{2.6}$$

where $\hat{w}$ is a weight function satisfying $\int_{(s,t)\in\Omega_{qr}} \hat{w}(s,t)dsdt = 1$, $\|\hat{w} - w\|_\infty \to 0$ and $\sqrt{n}(\hat{w} - w) = O_P(1)$. The consistency and asymptotic normality of $\hat{\zeta}_{CP}(\Omega_{qr})$ can be established using empirical processes techniques [Kosorok, 2007]. Given $EM^2 < \infty$, which is reasonable for familial studies where the family size M is random but bounded, we can show that all the above indicator processes are P-Donsker. Since $\hat{\zeta}_{CP}$ is a compact differential map of empirical processes indexed by Donsker classes, its consistency and asymptotic normality follow. In addition, there is an asymptotically linear representation $n^{1/2}(\hat{\zeta}_{CP} - \zeta_{CP}) = n^{-1/2}\sum_i I_i + o_P(1)$, where I_i are influence functions, upon which the variance of $\hat{\zeta}_{CP}$ can be estimated explicitly. Detailed proofs are given in Appendix A.1.

The association measure $\zeta(s, t; 1, 1)$ for bivariate competing risks data can be similarly adapted to sibship data denoted as $\zeta_{CC}(s, t; 1, 1)$. To estimate ζ_{CC}, we define the empirical processes $\mathbb{P}_n N_{11}^{\#}(s, t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{1\leq j<j'\leq m_i} I(Y_{ij} \leq s, \eta_{ij} = 1, Y_{ij'} \leq t, \eta_{ij'} = 1)$ and similarly $\mathbb{P}_n H^{\#}(s, t)$, $\mathbb{P}_n N_{10}^{\#}(s, t)$ and $\mathbb{P}_n N_{01}^{\#}(s, t)$ based on all within-cluster pairs. The asymptotic property of the corresponding plug-in estimator for ζ_{CC} can be established using similar arguments to those in Appendix A.1 except that now we assume $EM^4 < \infty$ as a sufficient condition, which is still reasonable due to finiteness of family size M.

2.2.4 Maximum Pseudo (Composite) Likelihood Estimator

It can be shown that ζ_{CC} is equivalent to θ_{CC} defined in (2.3), and that ζ_{CP} is equivalent to θ_{CP} defined in (2.1) [Cheng and Fine, 2008]. Hence the two sets of estimators, the U statistics and the plug-in estimators, are estimating the same quantities. We now discuss another estimating procedure for θs adapted from the Clayton [1978]'s pseudo likelihood function for bivariate survival data. The likelihood was also called a composite likelihood as discussed in Lindsay [1988]. Bandeen-Roche and Ning (manuscript) recently adapted the pseudo (composite) likelihood to bivariate competing risks data under a regression setting. We now further extend the likelihood to multivariate competing risks data by using a stratified method.

We first divide mother-multiple children data into several strata, each containing pairs of mothers and their d-th children. Here d ranges from 1 to m_{max}, where $m_{max} \equiv \max_{i=1}^{n} m_i$ is the maximum number of children among all families. Without loss of generality, we consider the first stratum of mothers and the corresponding eldest children. Define the cause specific set $\mathcal{A}(s, t; k, l)$ of size $a(s, t; k, l)$ pairs satisfying 1) both members are not censored by time

(s,t), and $a(s,t;k,l) \geq 1$; 2) there is at least one type l event that occurred at time t from one of the mothers and no event by time s from her eldest child and 3) there is at least one type k event that happened at time s from one of the eldest children and no event by time t from the corresponding mother. Let $\Delta(s,t;k,l) = 1$ if the types k and l events were from the same pair and 0 otherwise. It can be shown that $P\{\Delta(s,t;k,l) = 1|\mathcal{A}(s,t;k,l)\} = \frac{\theta_{CP}(s,t;k,l)}{a(s,t;k,l)-1+\theta_{CP}(s,t;k,l)}$, and $P\{\Delta(s,t;k,l) = 0|\mathcal{A}(s,t;k,l)\} = \frac{a(s,t;k,l)-1}{a(s,t;k,l)-1+\theta_{CP}(s,t;k,l)}$. See appendix A.3.

We now construct the pseudo likelihood function based on the observed mother-multiple children data $\{(Y_{i0}, \ldots, Y_{i,m_i}, \eta_{i0}, \ldots, \eta_{i,m_i}, m_i), i = 1, \ldots, n\}$ from the Cache County Study. To incorporate the multiple children structure, we adopt a stratified method. Let D_d denote the set containing all the families that have a d-th child, where $d = 1, \ldots, m_{max}$.Then mother-child pairs belonging to D_d form a bivariate competing risks data set $\{(Y_{id}, \eta_{id}, Y_{i0}, \eta_{i0}), i \in D_d\}$ for which the pseudo likelihood can be constructed. As discussed before, we assume a constant $\theta_{CP}(\Omega_{qr})$ over the region Ω_{qr}. To find all (s,t) pairs in $\mathcal{A}(s,t;k,l) \cap \Omega_{qr}$ such that $\Delta(s,t;k,l) = 1$, we count each child-mother pair in D_d as a concordant pair, if two event times from a pair are contained in Ω_{qr} and their causes are (k,l), denoted by the indicator function $\phi^{qr}_{i,d0} = I\Big\{(Y_{id}, Y_{i0}) \in \Omega_{qr}, (\eta_{id}, \eta_{i0}) = (k,l),\Big\}$, $i \in D_d$. The contribution of such pair to the log likelihood function is $\phi^{qr}_{i,d0} \log(\frac{\theta_{CP}}{a(Y_{id},Y_{i0},k,l)-1+\theta_{CP}})$. Next, to find (s,t) such that $\Delta(s,t;k,l) = 0$, we check any two different pairs $(Y_{id}, \eta_{id}, Y_{i0}, \eta_{i0})$ and $(Y_{jd}, \eta_{jd}, Y_{j0}, \eta_{j0})$ in D_d. If the times of the two pairs are discordant, s will be the minimum of children's failure times $Y_{(idjd)}$ and t will be the minimum of mothers' failure times $Y_{(i0j0)}$. The pair of the minimum failure times $(Y_{(idjd)}, Y_{(i0j0)})$ needs to be in Ω_{qr} and the cause indicators of the minimums $(\eta_{(idjd)}, \eta_{(i0j0)})$ need to match (k,l). We call two such pairs "discordant pairs" and define the discordant indicator $\psi^{qr}_{ij,dd00} = I\Big\{(Y_{id} - Y_{jd})(Y_{i0} - Y_{j0}) < 0$, $(Y_{(idjd)}, Y_{(i0j0)}) \in \Omega_{qr}, (\eta_{(idjd)}, \eta_{(i0j0)}) = (k,l)\Big\}$, where $i < j$ and $i,j \in D_d$. Their contribution to the log likelihood function is $\psi^{qr}_{ij,dd00} \log(\frac{a(Y_{(idjd)},Y_{(i0j0)},1,1)-1}{a(Y_{(idjd)},Y_{(i0j0)},1,1)-1+\theta_{CP}})$. Combining all the bivariate competing risks data from different strata D_d together, $d = 1, \ldots, m_{max}$, we have the following pseudo log-likelihood function:

$$L_n(\theta_{CP}(\Omega_{qr})) = \sum_{d=1}^{m_{max}} \sum_{i \in D_d} \phi_{i,d0}^{qr} \log\Big\{\frac{\theta_{CP}}{a(Y_{id}, Y_{i0}; k, l) - 1 + \theta_{CP}}\Big\} + \sum_{d=1}^{m_{max}} \sum_{i<j \& i,j \in D_d} \psi_{ij,dd00}^{qr} \log\Big\{\frac{a(Y_{(idjd)}, Y_{(i0j0)}; k, l) - 1}{a(Y_{(idjd)}, Y_{(i0j0)}; k, l) - 1 + \theta_{CP}}\Big\} \quad (2.7)$$

The maximum pseudo likelihood estimator $\tilde{\theta}_{CP}^{L}$ can be obtained by using some standard minimization procedure, e.g., function "nlminb" in R, for the negative log likelihood. As pointed out by Oakes [1982], the variance of $\tilde{\theta}_{CP}^{L}$ cannot be obtained from the second derivative of the pseudo log-likelihood function as $\psi_{ij,dd00}^{qr}$ may be dependent when $\theta_{CP}(\Omega_{qr}) \neq 1$. Instead, we use some results for Z-estimators in empirical processes [Kosorok, 2007] to derive the asymptotic properties of $\tilde{\theta}_{CP}^{L}$ based on the estimation equation $\Psi_n(\theta_{CP}) = \frac{\theta_{CP}}{\binom{n}{2}} \dot{L}_n(\theta_{CP}) = 0$, where the superscript dot denotes the derivative with respect to θ_{CP}.

The above framework can be further extended to clustered sibship data for which exchangeability can be assumed among all siblings. The adaptation of the maximum pseudo likelihood estimation to the sibship data is more complicated as there are two meaningful ways to form a pair, (Y_{ia}, Y_{ib}) and (Y_{ib}, Y_{ia}), for any two children a, b from the i-th family, under the exchangeability assumption. In addition, there may be too many different combinations of a and b from multiple-children families. To simplify the problem, we continue to use our stratified method and denote the set D_{ab}, $1 \leq a < b \leq m_{max}$, which consists of all families having a-th and b-th children. To adapt the exchangeability assumption, we consider the bivariate competing risks set consisting of all child-child pairs $\{(Y_{ia}, \eta_{ia}, Y_{ib}, \eta_{ib}), i \in D_{ab}\}$ together with its switched set containing all pairs $\{(Y_{ib}, \eta_{ib}, Y_{ia}, \eta_{ia}), i \in D_{ab}\}$.

The cause-specific set $\mathcal{A}(s, t, k, l)$ is naturally extended to this bivariate set and its switched set. The size of $\mathcal{A}(s, t, k, l)$ is now $a(s, t; k, l) = \sum_{i \in D_{ab}} \Big\{ I(Y_{ia} \geq s, Y_{ib} \geq t) + I(Y_{ib} \geq s, Y_{ia} \geq t) \Big\} \stackrel{d}{=} a_1(s, t; k, l) + a_2(s, t; k, l)$. As before let $\Delta(s, t; k, l) = 1$ if cause k event at time s and cause l event at time t were from the same pair and $\Delta(s, t; k, l) = 0$ otherwise. We can still show that $P\{\Delta(s, t; k, l) = 1 | \mathcal{A}(s, t; k, l)\} = \frac{\theta_{CC}(s,t;k,l)}{a(s,t;k,l) - 1 + \theta_{CC}(s,t;k,l)}$, and $P\{\Delta(s, t; k, l) = 0 | \mathcal{A}(s, t; k, l)\} = \frac{a(s,t;k,l)-1}{a(s,t;k,l) - 1 + \theta_{CC}(s,t;k,l)}$. See appendix A.3.

We now construct the pseudo log-likelihood function based on the stratum D_{ab}. As before we assume constant sibship association θ_{CC} within time region Ω_{qr}. To find all (s, t)

pairs in $\mathcal{A}(s,t;k,l)\cap\Omega_{qr}$ such that $\Delta(s,t;k,l)=1$, we count each concordant sibship pair in the bivariate set and its reversed set, and define concordant indicators $\phi^{qr}_{i,ab}=I\Big\{(Y_{ia},Y_{ib})\in\Omega_{qr},(\eta_{ia},\eta_{ib})=(k,l)\Big\}$ and $\phi^{qr}_{i,ba}=I\Big\{(Y_{ib},Y_{ia})\in\Omega_{qr},(\eta_{ib},\eta_{ia})=(k,l)\Big\}$, $i\in D_{ab}$. It is more complicated to find all time points $(s,t)\in\Omega_{qr}$ such that $\Delta(s,t;k,l)=0$, because for any two pairs, $i<j$, from D_{ab}, we may have four different pairings: (Y_{ia},Y_{ib}) with (Y_{ja},Y_{jb}), (Y_{ia},Y_{ib}) with (Y_{jb},Y_{ja}), (Y_{ib},Y_{ia}) with (Y_{ja},Y_{jb}) and (Y_{ib},Y_{ia}) with (Y_{jb},Y_{ja}). For the first pairing we define the discordant indicator $\psi^{qr}_{ij,aabb}=I\Big\{(Y_{ia}-Y_{ja})(Y_{ib}-Y_{jb})<0,(Y_{(iaja)},Y_{(ibjb)})\in\Omega_{qr},(\eta_{(iaja)},\eta_{(ibjb)})=(k,l)\Big\}$. Similarly, we define discordant indicators $\psi^{qr}_{ij,abba}$, $\psi^{qr}_{ij,baab}$ and $\psi^{qr}_{ij,bbaa}$. Combining all the strata D_{ab} together we have the following pseudo log-likelihood function:

$$\begin{aligned}
L_n(\theta_{CC}) = & \sum_{1\le a<b\le m_{max}}\sum_{i\in D_{ab}}\Big[\phi^{qr}_{i,ab}\log\Big\{\frac{\theta_{CC}}{a(Y_{ia},Y_{ib};k,l)-1+\theta_{CC}}\Big\} \\
& +\phi^{qr}_{i,ba}\log\Big\{\frac{\theta_{CC}}{a(Y_{ib},Y_{ia};k,l)-1+\theta_{CC}}\Big\}\Big] \\
& +\sum_{1\le a<b\le m_{max}}\sum_{i<j\&i,j\in D_{ab}}\Big[\psi^{qr}_{ij,aabb}\log\Big\{\frac{a(Y_{(iaja)},Y_{(ibjb)};k,l)-1}{a(Y_{(iaja)},Y_{(ibjb)};k,l)-1+\theta_{CC}}\Big\} \\
& +\psi^{qr}_{ij,abba}\log\Big\{\frac{a(Y_{(iajb)},Y_{(ibja)};k,l)-1}{a(Y_{(iajb)},Y_{(ibja)};k,l)-1+\theta_{CC}}\Big\} \\
& +\psi^{qr}_{ij,baab}\log\Big\{\frac{a(Y_{(ibja)},Y_{(iajb)};k,l)-1}{a(Y_{(ibja)},Y_{(iajb)};1,1)-1+\theta_{CC}}\Big\} \\
& +\psi^{qr}_{ij,bbaa}\log\Big\{\frac{a(Y_{(ibjb)},Y_{(iaja)};1,1)-1}{a(Y_{(ibjb)},Y_{(iaja)};1,1)-1+\theta_{CC}}\Big\}\Big] \quad (2.8)
\end{aligned}$$

The maximum likelihood estimator $\tilde{\theta}^L_{CC}$ for the constant association measure θ_{CC} can be similarly obtained by using the function "nlminb" in R. The consistency and asymptotic property of $\tilde{\theta}^L_{CC}$ can be proved analogously as for $\tilde{\theta}^L_{CP}$ in Appendix A.2.

For each extended association measures $\theta_{CC}(\zeta_{CC})$ and $\theta_{CP}(\zeta_{CP})$, we have discussed three sets of estimators. The plug-in estimators $\hat{\zeta}_{CC}$ and $\hat{\zeta}_{CP}$ are computationally straightforward to obtain. The extended pseudo maximum likelihood estimators proposed in this chapter $(\tilde{\theta}^L_{CP},\tilde{\theta}^L_{CC})$ and the U-type estimators discussed in Cheng et al. [2010] $(\hat{\theta}^U_{CP},\hat{\theta}^U_{CC})$ involve imputation and hence are more computationally intensive. Asymptotic normality can be established for all three estimators. Their relative efficiencies are examined in the following numerical studies.

2.3 SIMULATION STUDIES

2.3.1 Data without rounding errors

We conducted simulation studies to evaluate the small-sample performance of the plug-in estimator $\hat{\zeta}$, the pseudo maximum likelihood estimator $\tilde{\theta}^L$ and the U statistic $\hat{\theta}^U$. We first simulated mother-multiple children data by extending Bandeen-Roche and Liang [2002]'s frailty model. They modeled cause-specific between-subject dependency by an overall frailty A and proportions of cause allocations B and $1-B$ which were between 0 and 1 and independent of A. As the Archimedean copula is associative, we extended their frailty model for the bivariate competing risks naturally to multivariate settings in our simulations. Three different strengths of association were considered. For each scenario, we generated 500 replicates of 500 families with varying family sizes. More details of our simulation are given below.

`Step 1`. For each family, we first drew the familywise propensity A from a Gamma$(1,1)$.

`Step 2`. The number of children for a family M was determined from a multinomial distribution ranging from 1 to 5 with probabilities 0.09, 0.18, 0.24, 0.24, 0.25 which were estimated from the Cache County Data.

`Step 3`. Next we generated $M+1$ failure times from Exp(A) for the mother and her M children. For simplicity's sake we used the same distribution for mother and children.

`Step 4`. For each subject, we generated cause indicators from a Bernoulli distribution with probability B, where B was generated from a Beta$(R_1, 1-R_1)$. Three values of R_1, $0.2, 0.5$ and 0.8, were considered with corresponding theoretical values of $\theta_{CM}(s,t;1,1)$ being $6.0, 3.0$ and 2.25 for all s and t.

`Step 5`. Finally, independent censoring times were generated from Uniform$(5,10)$ for each subject in this family, imposing about 12% censoring.

We applied the three methods to each of the 500 simulated data sets. The mean of estimates, mean of standard error (for the plug-in and U estimators) or bootstrap standard error (for the likelihood based estimator), empirical standard error, and coverage rate of 95% confidence intervals are summarized in Table 1. With moderate sample size and not heavy censoring, the three estimators all perform well. The estimates are very close to the true values. The model-based or bootstrap standard errors agree well with the empirical standard errors. The coverages are close to the nominal level 0.95. The performance for $R_1 = 0.8$ is slightly better than that for $R_1 = 0.2$ as the latter has fewer cause one events.

Similarly, we simulated 500 sibship data sets by using the same simulation strategy. Each data set contained 250 families. The results are summarized in Table 2. When the sample size is smaller (250 families instead of 500) and the number of cause one events is small ($R_1 = 0.2$), there is some discrepancy between the estimates and the true value of association.

2.3.2 Data with tied observations

Real data are often rounded as in the Cache County Study, resulting in tied observations. We hence rounded the simulated data to one decimal place and evaluated how rounding errors would affect the performance of the three estimators. The adaptation of the plug-in estimator to the tied data is straightforward. For the U statistic, we followed Cheng et al. [2010] and excluded those pairs where one or both event times were tied. For the pseudo likelihood function (2.7), the tied observations do not affect the concordant pairs much but tend to underestimate the number of discordant pairs. Suppose we have two discordant pairs $(Y_{i0}, \eta_{i0}, Y_{i1}, \eta_{i1})$ and $(Y_{j0}, \eta_{j0}, Y_{j1}, \eta_{j1})$ from two distinct families i and j, where $Y_{i0} \leq Y_{j0}, \eta_{i0} = k$ and $Y_{i1} \geq Y_{j1}, \eta_{j1} = l$. When ties exist, e.g. $Y_{i1} = Y_{j1}$, and if $\eta_{i1} = l$, then the pairing of these two pairs would not be considered as discordant any longer.

To evaluate how the three estimators perform when the data are subject to rounding errors, we rounded simulated mother-children to one decimal place and computed the three estimators for each rounded data set. The results are summarized in Table 3. The plug-in estimator $\hat{\zeta}$ tends to underestimate θ_{CP} while the pseudo maximum likelihood estimator $\tilde{\theta}^L$ and the U statistic $\hat{\theta}^U$ tend to overestimate it. As we mentioned before, the pseudo likelihood estimator tends to overestimate the association for the rounded data. For the

Table 1: Simulation results for **mother-multiple children data**: mean of estimates (EST), mean of standard error/bootstrap standard error (MSE/BSE), empirical standard error (ESE), coverage rate of 95% confidence interval (COV)

R_1	$\theta(\zeta)_{CP}$	Stat	$\hat{\zeta}_{CP}$	$\tilde{\theta}^L_{CP}$	$\hat{\theta}^U_{CP}$
		EST	6.03	6.16	6.15
0.2	6	MSE/BSE	0.82	0.84	0.84
		ESE	0.83	0.79	0.86
		COV	0.93	0.96	0.95
		EST	2.99	3.01	3.01
0.5	3	MSE/BSE	0.24	0.23	0.24
		ESE	0.23	0.23	0.23
		COV	0.95	0.95	0.96
		EST	2.24	2.26	2.25
0.8	2.25	MSE/BSE	0.13	0.13	0.13
		ESE	0.13	0.13	0.13
		COV	0.94	0.96	0.94

Table 2: Simulation results for **sibship data**: mean of estimates (EST), mean of standard error/bootstrap standard error (MSE/BSE), empirical standard error (ESE), coverage rate of 95% confidence interval (COV)

R_1	$\theta(\zeta)_{CC}$	Stat	$\hat{\zeta}_{CC}$	$\tilde{\theta}^L_{CC}$	$\hat{\theta}^U_{CC}$
		EST	5.98	6.23	6.22
0.2	6	MSE/BSE	1.11	1.15	1.18
		ESE	1.21	1.23	1.31
		COV	0.90	0.93	0.92
		EST	3.00	3.04	3.05
0.5	3	MSE/BSE	0.32	0.33	0.33
		ESE	0.32	0.33	0.33
		COV	0.95	0.94	0.96
		EST	2.24	2.26	2.26
0.8	2.25	MSE/BSE	0.18	0.18	0.18
		ESE	0.16	0.16	0.17
		COV	0.97	0.96	0.97

Table 3: Simulation results for **rounded mother-multiple children data**: mean of estimates (EST), mean of standard error/bootstrap standard error (MSE/BSE), empirical standard error (ESE), coverage rate of 95% confidence interval (COV)

R_1	$\theta(\zeta)_{CP}$	Stat	$\hat{\zeta}_{CP}$	$\tilde{\theta}^{L}_{CP}$	$\hat{\theta}^{U}_{CP}$	$\hat{\theta}^{MU}_{CP}$
		EST	5.74	6.47	6.45	5.98
0.2	6	MSE/BSE	0.75	0.91	0.93	0.81
		ESE	0.76	0.89	0.94	0.81
		COV	0.90	0.95	0.96	0.96
		EST	2.84	3.24	3.15	2.98
0.5	3	MSE/BSE	0.21	0.26	0.26	0.23
		ESE	0.21	0.26	0.27	0.22
		COV	0.87	0.88	0.93	0.94
		EST	2.14	2.49	2.37	2.23
0.8	2.25	MSE/BSE	0.12	0.15	0.15	0.13
		ESE	0.12	0.15	0.15	0.13
		COV	0.82	0.67	0.91	0.93

plug-in estimator, when data are rounded and ties are present, we evaluate its numerator and denominator over fewer time grids. Suppose there are two pairs (Y_{i0}, Y_{i1}) and (Y_{j0}, Y_{j1}) that have both events of interest and the events times become tied after rounding, denoted as (Y_{i0}^*, Y_{i1}^*). The numerator would increase as $1 \cdot \mathbb{P}_n H(Y_{i0}, Y_{i1}) + 1 \cdot \mathbb{P}_n H(Y_{j0}, Y_{j1}) < 2 \cdot \mathbb{P}_n H(Y_{i0}^*, Y_{i1}^*)$. Similarly, the denominator would also increase but at a faster rate as the rounding affects both $\mathbb{P}_n N_{10}$ and $\mathbb{P}_n N_{01}$. Hence the plug-in estimator underestimates the strength of association when ties are present. For the U statistic, rounding causes the removal of roughly similar amounts of concordant and discordant pairs. When the association is positive, i.e. the numerator is larger than the denominator, subtracting the same positive value from both the numerator and denominator makes the ratio larger. Therefore, the U statistic overestimates the positive association when ties are present. We also evaluated the performance of the three estimators for sibship data with tied observations. The results are given in Table 4. We observed similar trends as in the mother-multiple children data. The plug-in estimators are likely to underestimate the association while the other two are prone to overestimate the association.

To address potential loss of accuracy in the three estimators for tied data, we now propose a modified U statistic which essentially adds back those removed concordant and discordant pairs due to tied observations. When two mother-child pairs have tied observations — for example, when both mothers had the same event time — we consider their corresponding cause indicators. If mother 1 died and mother 2 had dementia at the same age, we assume that mother 1 would have dementia later than mother 2, had mother 1 not died. If child event times were not tied, we would be able to determine concordant status. For those tied pairs that can be concordant or discordant with equal probability, we add half a pair to both concordant and discordant pairs. It becomes more complicated for sibship data under the exchangeability assumption. Appendix A.4 provides details on how to add back all missing concordant and discordant pairs for sibship data. We evaluated the performance of the modified U statistics, denoted by $\hat{\theta}_{CP}^{MU}$ and $\hat{\theta}_{CC}^{MU}$, based on the simulated rounded data and listed the results at the last column of Table 3 and Table 4. With the presence of rounding

Table 4: Simulation results for **rounded sibship data**: mean of estimates (EST), mean of standard error/bootstrap standard error (MSE/BSE), empirical standard error (ESE), coverage rate of 95% confidence interval (COV)

R_1	$\theta(\zeta)$	Stat	$\hat{\zeta}_{CC}$	$\tilde{\theta}^L_{CC}$	$\hat{\theta}^U_{CC}$	$\hat{\theta}^{MU}_{CC}$
		EST	5.66	6.54	6.50	6.12
0.2	6	MSE/BSE	1.01	1.23	1.28	1.14
		ESE	0.99	1.23	1.29	1.14
		COV	0.88	0.97	0.96	0.95
		EST	2.89	3.32	3.24	2.99
0.5	3	MSE/BSE	0.30	0.36	0.38	0.32
		ESE	0.33	0.40	0.41	0.31
		COV	0.89	0.89	0.92	0.94
		EST	2.13	2.49	2.36	2.23
0.8	2.25	MSE/BSE	0.16	0.20	0.20	0.17
		ESE	0.17	0.21	0.21	0.18
		COV	0.84	0.81	0.92	0.94

errors, the modified U statistics clearly outperform the other three estimators, as they are closest to the true values. There is also some efficiency gain over the original U because the latter loses information by discarding all tied observations.

2.4 CACHE COUNTY STUDY

The data from the Cache County Study on Memory in Aging have been analyzed in previous works. Bandeen-Roche and Liang [2002] and Bandeen-Roche and Ning [2008] applied their U-type estimator and Cheng and Fine [2008] used a plug-in estimator to investigate the child-mother association in dementia based on 3,635 pairs of mothers and their eldest children. In order to utilize the information on other siblings, Cheng et al. [2010] adapted the U-type estimator for bivariate competing risks data to multivariate cases, assuming exchangeability among all siblings. In this thesis, we were interested in both the child-mother and sibship associations. Hence we extended the plug-in estimator from bivariate competing risks data and the pseudo likelihood estimator from a typical bivariate survival setting to more complicated data structures such as mother-multiple children data and sibship data. The two extended estimators were applied to 3,635 families containing mothers and all her children for the mother-child association, and to the sibship data containing 4,770 families of siblings for the sibship association. As the data had tied observations due to ages being rounded to the closest integer, we also computed the modified U statistic and compared with the two extended estimators and the original U statistic.

For comparison, we examined the strengths of association over the same time regions as reported in Cheng et al. [2010]. We first divided the age ranges into three intervals ≤ 70, $71-80$, and > 80 for both dimensions. For each age region, we computed the plug-in estimator ($\hat{\zeta}_{CP}$), the pseudo likelihood estimator ($\tilde{\theta}^{L}_{CP}$), the original U estimator ($\hat{\theta}^{U}_{CP}$) and the modified U ($\hat{\theta}^{MU}_{CP}$) of the child-mother association according to (2.6), (2.7), (2.2) and an equation similar to ($A.5$). The model-based standard errors (SE) were available for the plug-in and U-type estimators and calculated based on their influence functions. We also computed the boostrap standard errors (BSE) for the four estimators based on 500 bootstrap samples. The results were summarized in the top panel of Table 5. Similarly, for the sibship association, we computed the four estimators $\hat{\zeta}_{CC}$, $\tilde{\theta}^{L}_{CC}$, $\hat{\theta}^{U}_{CC}$ and $\hat{\theta}^{MU}_{CC}$ according to (2.5), (2.8), (2.4) and (*A*.5). The model-based standard errors were given for the plug-in estimator and U estimators. The bootstrap standard errors were also calculated for all four estimators based on 500 bootstrap samples, each containing 2,000 clusters of siblings drawn from 4,770 families with replacement. The reported BSEs were adjusted by a factor of $\sqrt{\frac{2{,}000}{4{,}770}}$. The results on sibship association are given in the lower panel of Table 5. The same analyses were performed on four regions $\leq (>)75 \times \leq (>)80$ and the results are shown in Table 6.

The original U estimates for mother-child associations are identical to those reported in Cheng et al. [2010]. However, there are some that differ in the sibship analysis because we are now using not only the pairs whose event times belong to the region of interest but also the pairs whose switched times under exchangeability belong to the region. For mother-child associations, the four estimates are consistent on most regions. The modified U estimates almost always lie between the plug-in estimates and the original U estimates except for the region $s > 80, t \leq 70$ where there are only three mother-child pairs both having dementia ($N_{11} = 3$). We observe a similar trend for sibship associations in the lower panel of Table 5. This is consistent with our simulation studies in Section 2.3.2 where we have found that the plug-in estimator tends to underestimate and the original U statistic tends to overestimate, while the modified U is closest to the true association. However, when the number of double cause-1 events N_{11} is small (i.e. $N_{11} = 2$), there is a noticeable discrepancy between the pseudo likelihood estimate and the other three. For sibship associations the four sets of estimates are generally close even when N_{11} is small, and the standard errors are smaller than those for mother-child associations. As more sibship pairs were used in the sibship analysis than child-mother pairs in the child-mother analysis, the sibship estimates tend to be more robust than the child-mother estimates. This suggests that it would be important to extend original estimators for bivariate data to clustered data and to incorporate as much information as possible in quantifying mother-child associations and sibship associations when the event is rare.

When we examine the mother-child associations on broader regions [Cheng et al., 2007, 2009], as given in Table 6, the plug-in estimates, the modified U estimates and the original

U become closer and always follow the same order. A similar trend is observed in the sibship analysis. However, the discrepancy between the likelihood-based estimates and the other three still exists, especially on the region $s > 75, t > 80$, for both mother-child and sibship associations, in spite of the fact that the N_{11}s get bigger in the four-region analysis. One possible explanation is that the association may be time-varying on these broader regions, and the estimated association measure can be thought of as some weighted average over each region. The likelihood-based method may weight the time-varying association differently as compared to the plug-in and U statistics.

2.5 DISCUSSION

In this chapter we have discussed three types of nonparametric approaches to estimating two equivalent and easily interpreted association measures for multivariate competing risks data. The plug-in estimator is most straightforward to implement with an explicit variance estimator. The U-statistic requires imputation and needs more computing power. The pseudo likelihood estimator lacks an explicit variance estimator and relies on bootstrapping, hence it is most computationally intensive. On the other hand, the pseudo likelihood estimator has the potential to be most efficient. In this chapter we use a stratified approach which always matches up children with the same sibling order. It is possible to check concordance status for two pairs; for example, a mother and her eldest child from one family, and another mother and her second child from another family. There are actually many more combinations than those used in (2.7) and (2.8). Based on our simulation studies, the stratified pseudo likelihood estimator has comparable efficiency to the other two estimators. If we were able to use all possible combinations, which of course would be very computationally intensive, we would further improve the efficiency of the pseudo likelihood estimator. In addition, the likelihood-based approach is more adaptive to a regression setting or to a parametric model. For example, one may extend the regression model for bivariate competing risks data (Ning and Bandeen-Roche, manuscript) to multivariate competing risks data based on the pseudo likelihood functions in (2.7) and (2.8). Recently, Hu et al. [2011] used a polynomial function to approximate cross-hazard ratio and estimated the unknown parameters based on a pseudo-partial likelihood. Similar parametric approximation may be used for the cause-spefic cross-hazard ratio and the pseudo likelihood functions considered in this chapter can again be used as the base of estimation. These will be future research topics.

Moreover, we investigate how rounding errors affect the performance of the three non-parametric estimators and propose a modified U statistic to handle tied multivariate competing risks data. For the modified U statistic, we simply adjust for concordant and discordant pairs for the tied events. The idea may be used analogously for the pseudo likelihood estimator. In Appendix A.2, we show that the pseudo likelihood estimator is a Z-estimator based on the estimating equation which is a U statistic of order 2 with the kernel function $h(X_i, X_j; \theta)$ given in (A.1). One may let both $\phi^{qr}_{ij,dd00}$ and $\psi^{qr}_{ij,dd00}$ in the numerator of (A.1) be 0.5 for tied events following the same scheme as in Appendix A.4. However, it is not clear how to fix the plug-in estimator when there are rounding errors. Based on our current simulation studies and real data analysis, we suggest using the plug-in estimator for untied data as it is simplest to compute, and the modified U statistic when the data are subject to rounding errors.

Table 5: Child-mother association and sibship association estimated within 3 × 3 time regions Ω_{qr}, number of pairs at risk in each time region (N), number of cause 1 pairs in each time region (N_{11}), mean estimate (EST), model-based standard error (MSE), bootstrap standard error (BSE).

	Obs		$\hat{\zeta}_{CP}$			$\tilde{\theta}^L_{CP}$		$\hat{\theta}^U_{CP}$			$\hat{\theta}^{MU}_{CP}$		
Child-mother Association													
$(s,t) \in \Omega_{qr}$	N	N_{11}	EST	MSE	BSE	EST	BSE	EST	MSE	BSE	EST	MSE	BSE
$s \leq 70, t \leq 70$	1338	10	4.03	1.33	1.40	2.57	1.22	4.21	1.42	1.51	4.06	1.36	1.34
$s \leq 70, 70 < t \leq 80$	2171	12	2.27	0.69	0.72	3.13	1.16	2.34	0.73	0.76	2.29	0.70	0.69
$s \leq 70, t > 80$	3907	7	1.24	0.46	0.48	1.31	0.56	1.30	0.49	0.51	1.25	0.47	0.49
$70 < s \leq 80, t \leq 70$	1258	18	4.86	1.26	1.30	4.83	1.51	5.14	1.39	1.44	4.93	1.30	1.36
$70 < s \leq 80, 70 < t \leq 80$	2271	23	2.64	0.58	0.60	3.00	0.84	2.76	0.64	0.66	2.68	0.60	0.58
$70 < s \leq 80, t > 80$	3757	15	2.13	0.63	0.64	1.85	0.63	2.27	0.70	0.72	2.15	0.65	0.67
$s > 80, t \leq 70$	650	3	0.96	0.49	0.48	1.83	0.71	0.94	0.49	0.49	0.96	0.49	0.51
$s > 80, 70 < t \leq 80$	961	10	2.44	0.89	0.95	2.55	0.89	2.55	0.98	1.08	2.47	0.92	0.96
$s > 80, t > 80$	1715	9	2.40	1.04	1.12	1.88	0.91	2.66	1.19	1.15	2.43	1.08	1.07
Sibship Association													
	Obs		$\hat{\zeta}_{CC}$			$\tilde{\theta}^L_{CC}$		$\hat{\theta}^U_{CC}$			$\hat{\theta}^{MU}_{CC}$		
$(s,t) \in \Omega_{qr}$	N	N_{11}	EST	MSE	BSE	EST	BSE	EST	MSE	BSE	EST	MSE	BSE
$s \leq 70, t \leq 70$	11239	13	3.40	1.01	1.08	3.16	1.23	3.42	1.03	1.12	3.44	1.03	1.09
$s \leq 70, 70 < t \leq 80$	14312	29	3.35	0.58	0.56	3.87	0.76	3.45	0.61	0.59	3.37	0.59	0.57
$s \leq 70, t > 80$	3738	20	2.42	0.60	0.60	2.42	0.63	2.56	0.66	0.68	2.44	0.62	0.66
$70 < s \leq 80, t \leq 70$	3213	15	3.35	0.58	0.56	3.87	0.72	3.45	0.61	0.59	3.37	0.59	0.63
$70 < s \leq 80, 70 < t \leq 80$	10750	52	2.95	0.49	0.48	3.16	0.55	3.11	0.53	0.53	3.00	0.51	0.52
$70 < s \leq 80, t > 80$	8377	36	2.58	0.44	0.40	2.41	0.42	2.80	0.50	0.47	2.61	0.45	0.44
$s > 80, t \leq 70$	1518	2	2.42	0.60	0.63	2.42	0.68	2.56	0.66	0.71	2.44	0.62	0.68
$s > 80, 70 < t \leq 80$	2154	20	2.58	0.54	0.40	2.41	0.47	2.80	0.50	0.47	2.61	0.45	0.42
$s > 80, t > 80$	4992	30	1.65	0.45	0.42	1.43	0.27	1.81	0.52	0.52	1.67	0.45	0.46

Table 6: Child-mother association and sibship association estimated within 2× 2 time regions Ω_{qr}, number of pairs at risk in each time region (N), number of cause 1 pairs in each time region (N_{11}), mean estimate (EST), model-based standard error (MSE), bootstrap standard error (BSE).

Child-mother Association													
	Obs			$\hat{\zeta}_{CP}$		$\tilde{\theta}^{L}_{CP}$			$\hat{\theta}^{U}_{CP}$			$\hat{\theta}^{MU}_{CP}$	
$(s,t) \in \Omega_{qr}$	N	N_{11}	EST	MSE	BSE	EST	BSE	EST	MSE	BSE	EST	MSE	BSE
$s \le 75, t \le 80$	5319	41	3.41	0.60	0.59	3.57	0.68	3.55	0.65	0.65	3.44	0.61	0.62
$s \le 75, t > 80$	5891	12	1.30	0.39	0.40	1.25	0.40	1.36	0.42	0.44	1.30	0.39	0.45
$s > 75, t \le 80$	3330	35	2.40	0.48	0.52	2.45	0.49	2.49	0.53	0.58	2.43	0.49	0.50
$s > 75, t > 80$	3488	19	2.71	0.87	0.55	1.80	0.41	2.90	0.99	0.63	2.75	0.90	0.95
Sibship Association													
	Obs			$\hat{\zeta}_{CC}$		$\tilde{\theta}^{L}_{CC}$			$\hat{\theta}^{U}_{CC}$			$\hat{\theta}^{CU}_{CC}$	
$(s,t) \in \Omega_{qr}$	N	N_{11}	EST	MSE	BSE	EST	BSE	EST	MSE	BSE	EST	MSE	BSE
$s \le 75, t \le 80$	34517	68	3.28	0.52	0.51	3.17	0.50	3.36	0.55	0.54	3.33	0.53	0.52
$s \le 75, t > 80$	7345	36	2.61	0.47	0.52	2.41	0.45	2.76	0.52	0.58	2.63	0.48	0.49
$s > 75, t \le 80$	8669	63	2.94	0.35	0.34	2.77	0.34	3.11	0.39	0.39	2.97	0.36	0.36
$s > 75, t > 80$	9762	50	2.03	0.34	0.33	1.67	0.25	2.24	0.40	0.40	2.06	0.35	0.34

3.0 ASSOCIATION ANALYSIS OF MULTIVARIATE RECURRENT EVENTS WITH MARKS: WITH APPLICATION TO MEASURE THE TAIL DEPENDENCE OF INSURANCE LOSSES

Dynamics of insurance losses are mutually dependent via the dependence between their two components: interarrival times and accompanying losses. Many multivariate models have been developed to capture the dependence of insurance risks [Lindskog and McNeil, 2003, Pfeifer and Nešlehvová, 2004, Chavez-Demoulin et al., 2005, Bäuerle and Grübel, 2005]. Some previous works focused only on dependence of arrival times, while the others modeled the dependence of arrival times and losses separately. Since the marginal insurance loss process can be modeled as a compound Poisson process (CPP), which is a special case of Lévy process, the Lévy copula proposed by Tankov [2003] can then be used to model the dependence of arrival times and losses simultaneously.

More specifically, insurance losses fall into the category of multivariate recurrent events with marks, which can be modeled by multivariate dependent compound Poisson processes (CPP). To study the dependence structures of multivariate CPP, it is convenient to consider it in a broader class of processes: Lévy processes. The dependence structure of several special cases of Lévy processes have been studied, e.g. Brownian subordinator and multivariate tempered stable processes [Rosinski, 2007]. These models are designed for specific classes and can only handle limited range of dependence. We note that a Lévy process is completely determined by its characteristic triplet (ν, A, γ), where covariance matrix A completely characterizes the continuous part of a Lévy process and is easy to specify. However, the jump part of a Lévy process determined by the Lévy measure ν is hard to model. Tankov [2003] proposed the Lévy copula to combine the tails integral of marginal Lévy measures. Then Esmaeili and Klüppelberg [2010a] proposed to model tail dependence of aggregate losses by a bivariate CPP and investigated the parameter inference of the Lévy copula by a full likelihood method.

In this part of thesis, a new two-stage semiparametric inference method is proposed to estimate the marginal and dependence parameters, which is shown to be superior than purely non-parametric method, e.g. Kendall's τ estimator, but inferior to the full likelihood. Although less efficient than full likelihood method, the proposed semi-parametric method is superior in several other aspects: First, it is easier to expand to the multivariate case, for which a full likelihood method becomes very complicated. Second, the proposed method does not assume a specific distribution for marginal loss, instead, it uses a nonparametric estimator and so is more robust. Last, by using the new method, inference of the dependence parameter is margin-independent, which is important when the dependence, instead of the margins, is the major concern.

In the following sections, the structure of insurance loss data and the corresponding bivariate CPP model are introduced first. Then two existing inference methods, Kendall τ estimator and full likelihood method are presented. Next, the new two-stage estimator is proposed, and the consistency and asymptotic normality of the estimator are also studied. Finally, a numerical study to compare these methods and an analysis of Danish fire data are conducted. An introduction to Lévy process can be found in the appendix B.2.2.

3.1 DATA AND MODEL

Assume that the aggregate loss process can be observed at a high frequency, so that all the arrival times and claim sizes are observable. This is different from the assumptions for decompounding CPP[Buckmann and Rudolf, 2003], for which only a sample of the CPP is observable. Instead, consider a bivariate aggregate loss process $(S_1(t), S_2(t)), 0 \leq t \leq T$, which can be represented by:

$$S_1 = \sum_{i=1}^{N_1(t)} X_i \quad \text{and} \quad S_2 = \sum_{j=1}^{N_2(t)} Y_j \tag{3.1}$$

where $N_k(t),\ k = 1, 2$ is the underlying Poisson process associated with arrival times $0 = t_{k,0} < t_{k,1} < t_{k,2} < \ldots < t_{k,n_k} = T$ and n_k is the total number of losses up to T. The accompanying losses for $(S_1(t), S_2(t))$ are $X_m \in R^+, m = 1\ldots, n_1$ and $Y_m \in R^+, m = 1, \ldots, n_2$ respectively. In practice, $n_1 \neq n_2$, which means that S_1 and S_2 do not always jump together. One can model the marginal loss process as a CPP while understanding the arrival times as jump times and losses as jump sizes.

A d-dimensional CPP can be defined as:

$$\mathbf{S}(t) = \sum_{i=1}^{N(t)} \mathbf{X}_i$$

where $(N(t))_{t\geq 0}$ is a homogeneous Poisson process with intensity $\lambda > 0$ and $(\mathbf{X}_i)_{i\in(N)}$ is a sequence of iid random variable in $\mathbb{R}^d$, independent from $(N(t))_{t\geq 0}$. To accommodate the fact that $S_1(t)$ and $S_2(t)$ may not jump together, it is necessary to allow the marginal jump size distribution to have atom at 0. However, $\mathbf{X}_i$ has no atom at $\mathbf{0}$, i.e. $P(\mathbf{X}_i = \mathbf{0}) = 0$, where $\mathbf{0} \in \mathbb{R}^d$.

If the components of a CPP always jump together, we only need to model the dependence structure of jump sizes. On the other hand, if they never jump together, intuitively, they are mutually independent so that the dependence need not be modeled. However, in practice, the aggregate process is a mixture of single jumps and common jumps. The dependence of a bivariate CPP does not only rely on the joint jump sizes but also on the frequency of occurrences. To justify this theoretically, we use the Lévy Itŏ decomposition to separate the single and joint jumps of CPP and then obtain the decomposition of a bivariate CPP thus:

$$S_1(t) = S_1(t)^{\perp} + S_1(t)^{\parallel} \ \text{ and } \ S_2(t) = S_2(t)^{\perp} + S_2(t)^{\parallel}$$

To see this, first recall the Lévy Itŏ decomposition for CPP:

$$\begin{aligned} S(t) &= \sum_{i=1}^{N(t)} (Z_{1i}, Z_{2i}) = \int_0^t \int_{\mathbb{R}_+^2 \backslash \{\mathbf{0}\}} \mathbf{z} J(ds \times d\mathbf{z}) \\ &= \int_0^t \int_{(\mathbb{R}^+\backslash\{0\})\times\{0\}} \mathbf{z} J(ds \cdot d\mathbf{z}) + \int_0^t \int_{\{0\}\times(\mathbb{R}^+\backslash\{0\})} \mathbf{z} J(ds \cdot d\mathbf{z}) + \int_0^t \int_{(\mathbb{R}^+\backslash\{\mathbf{0}\})^2} \mathbf{z} J(ds \cdot d\mathbf{z}) \end{aligned}$$

where $J(ds \cdot d\mathbf{z})$ is the corresponding Poisson measure on $[0, \infty) \times (\mathbb{R}_+^2 \backslash \{\mathbf{0}\})$ with intensity $ds \cdot \nu(d\mathbf{z})$.

Note in the last equality, the three terms are independent, because they have mutually exclusive supports and so never jump together. The first two correspond to the single jumps of $S_1(t)$ and $S_2(t)$ respectively, while the last one models the joint jumps.

Another way of understanding this formula is by decomposing the Lévy measure:

$$\nu_1(A) = \nu(A \times \{0\}),\ \forall A \in \mathcal{B}(\mathbb{R}^+ \backslash \{0\})$$
$$\nu_2(A) = \nu(\{0\} \times A),\ \forall A \in \mathcal{B}(\mathbb{R}^+ \backslash \{0\})$$
$$\nu_3(A) = \nu(A) - \nu(A_X \times \{0\}) - \nu(\{0\} \times A_Y)$$

where $A_X = \{x : (x,0) \in A\}$ and $A_Y = \{y : (0,y) \in A\}$, $\forall A \in \mathcal{B}(\mathbb{R}^2_+ \backslash \{\mathbf{0}\})$, in which $\mathcal{B}$ indicates the Borel Sets.

So $\nu(A) = \nu_1(A) + \nu_2(A) + \nu_3(A)$, where $\nu_1(A)$ and $\nu_2(A)$ correspond to $S_1(t)^{\perp}$ and $S_2(t)^{\perp}$, respectively. Similarly, $\nu_1(\cdot)$ and $\nu_2(\cdot)$ have disjoint supports and so are independent of each other. $\nu_3(\cdot)$ corresponds to the common jumps $S_1(t)^{\parallel}$ and $S_2(t)^{\parallel}$. To justify this formula, recall that a d-dimensional CPP $S(t) = \sum_{i=1}^{N(t)} \mathbf{X}_i$ is a pure jump Lévy process following the Lévy-Khintchine representation:

$$E[e^{i(\mathbf{z},\mathbf{S}(t))}] = \exp\left\{ t \int_{\mathbb{R}^d} (e^{i(\mathbf{z},\mathbf{x})} - 1)\nu(d\mathbf{x}) \right\}, \quad \mathbf{z} \in \mathbb{R}^d$$

where the Lévy measure satisfies $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^d} \nu(dx) < \infty$. Then the formula for the bivariate CPP can be decomposed as

$$\begin{aligned}
&E[e^{iz_1 S_1(t) + iz_2 S_2(t)}] \\
&\quad = \exp\left\{ t \int_{\mathbb{R}^+ \backslash \{0\}} (e^{iz_1 x} - 1)\nu_1(dx) + t \int_{\mathbb{R}^+ \backslash \{0\}} (e^{iz_2 y} - 1)\nu_2(dy) + t \int_{\mathbb{R}^2_+ \backslash \{\mathbf{0}\}} (e^{iz_1 x + iz_2 y} - 1)\nu_3(dx \times dy) \right\} \\
&\quad = E\left[e^{iz_1 S_1(t)^{\perp}}\right] E\left[e^{iz_2 S_2(t)^{\perp}}\right] E\left[e^{iz_1 S_1(t)^{\parallel} + iz_2 S_2(t)^{\parallel}}\right]
\end{aligned}$$

3.2 DEPENDENCE STRUCTURE BY LÉVY COPULA

Researchers have proposed many methods to estimate parameters of the one-dimensional Lévy process under both low frequency and high frequency observation settings. However, inference methods for multivariate Lévy processes are not well investigated, except those for multivariate compound Poisson processes. Schicks [2009] proposed a nonparametric inference method for multivariate compound Poisson processes. Esmaeili and Klüppelberg [2010a] used a full likelihood method to estimate parameters of bivariate compound Poisson processes based on the Lévy copula. In this thesis, we use the Lévy copula to construct a parametrized dependence structure of the bivariate CPP while estimating the marginal jump sizes nonparametrically.

3.2.1 Lévy Copula

The Lévy copula for spectrally non-negative Lévy process, e.g. CPP with positive jumps only, is introduced below. It uses tail integral of Lévy measure, analogous to the cumulative distribution function of probability measure, to construct the dependence structure of Lévy process. A tail integral quantifying the frequency of jumps with sizes greater than a threshold is used to avoid the possibility that the Lévy process may explode when the jump size approaches 0, which is because the number of small jumps of a Lévy process with sizes smaller than a value may be infinite.

Definition Tail Integral: Let ν be a Lévy measure on R_+^2. The tail integral is a function $U : [0, \infty]^2 \to [0, \infty]$ defined by:

$$U(x_1, x_2) = \begin{cases} \nu([x_1, \infty) \times [x_2, \infty)), (x_1, x_2) \in [0, \infty)^2 \backslash \{\mathbf{0}\} \\ 0, \text{ if } x_i = \infty \text{ for at least one } i \in \{1, 2\} \\ \infty, \ x_i = 0 \text{ for all } i \in \{1, 2\} \end{cases} \tag{3.2}$$

The marginal tail integrals are defined analogously for $i = 1, 2$ as $U_i(x) = \nu_i([x, \infty))$ for $x \geq 0$.

Definition Lévy Copula: The Lévy copula of a spectrally positive Lévy process is a 2-increasing grounded function $\mathcal{C} : [0, \infty]^2 \to [0, \infty]$ with margins $\mathcal{C}_k(u) = u$ for all $u \in [0, \infty]$ and $k = 1, 2$.

The notion of groundedness means $\mathcal{C}(u_1, u_2) = 0$ if $u_i = 0$ for at least one $i \in \{1, 2\}$, which guarantees that $\mathcal{C}$ defines a measure on $[0, \infty]^2$. Indeed, the volume of Lévy copula is a 2-dimensional measure with Lebesgue margins.

Theorem 3.2.1. *Let U denote the tail integral of a spectrally positive 2-dimensional Lévy process, whose components have Lévy measures ν_1, ν_2. Then there exists a Lévy copula $\mathcal{C}$ such that for all $(x_1, x_2) \in [0, \infty]^2$*

$$U(x_1, x_2) = \mathcal{C}(U_1(x_1), U_2(x_2)) \tag{3.3}$$

If the marginal tail integrals are continuous, then this Lévy copula is unique. Otherwise, it is unique on $RanU_1 \times RanU_2$, where Ran denotes range.

Conversely, if $\mathcal{C}$ is a Lévy copula and U_1, U_2 are marginal tail integrals of a 2-d spectrally positive Lévy process, then the above formula defines the tail integral of a 2-dimensional spectrally positive Lévy process and U_1, U_2 are tail integrals of its components.

This theorem shows that a bivariate tail integral and the corresponding marginal tail integrals can be connected by a Lévy copula function. Conversely, connecting two marginal tail integrals by a Lévy copula function leads to a bivariate Lévy process. However, it does not tell us how to construct the Lévy copula. One of the most popular methods is through a generator, analogous to an ordinary Archimedean copula:

$$\mathcal{C}(u, v) = \phi^{-1}(\phi(u) + \phi(v))) \tag{3.4}$$

where the generator $\phi(\cdot)$ is a strictly decreasing convex function from $[0, \infty]$ to $[0, \infty]$ such that $\phi(0) = \infty$ and $\phi(\infty) = 0$. In practice, Archimedean copulas are popular because they allow us to model dependence in arbitrarily high dimensions with only one parameter, governing the strength of dependence.

Using $\phi(u) = u^{-\theta}$ with $\theta > 0$, one obtains the Clayton Lévy copula:

$$\mathcal{C}_\theta(u, v) = (u^{-\theta} + v^{-\theta})^{-1/\theta}. \tag{3.5}$$

Note that θ tends to 0 implies independence and when $\theta \to \infty$ we have complete dependence. The extreme cases of dependence will be introduced later.

When the dependence is specified through the Lévy copula and both the copula and one-dimensional tail integrals are sufficiently smooth, the Lévy density can be computed by differentiation:

$$\nu(x, y) = \frac{\partial^2 \mathcal{C}(u, v)}{\partial u \partial v} |_{u=U_1(x), v=U_2(y)} \ \nu_1(x)\nu_2(y), \tag{3.6}$$

which is useful for constructing likelihood.

Applying Lévy copula to a bivariate CPP, we are now able to consider its dependence structure. First, the Lévy measure of a one-dimensional CPP is $\lambda f(dx)$, where λ is the constant intensity of the underlying Poisson process and $f(x)$ is the density of jump size distribution function $F(x)$, which is assumed continuous. The tail integral is $\lambda \bar{F}(x)$, where $\bar{F}(x)$ is the corresponding survival distribution function of the jump size. Next, by the decomposition of CPP, we have for $i = 1, 2$:

$$U_i = U_i^{\perp} + U_i^{\parallel}.$$

If we use the Lévy copula $\mathcal{C}(u, v)$ to represent the 2-d tail integral for CPP, then we have:

$$\lambda^{\parallel} = \lim_{x,y \to 0^+} U(x, y) = \mathcal{C}(\lambda_1, \lambda_2; \delta) \text{ and } \lambda_i^{\perp} = \lambda_i - \lambda^{\parallel}$$

and the following result:

$$\lambda_1^{\perp} \bar{F}_1^{\perp}(x) = \lambda_1 \bar{F}_1(x) - \lambda^{\parallel} \bar{F}_1^{\parallel}(x) = \lambda_1 \bar{F}_1(x) - \mathcal{C}(\lambda_1 \bar{F}_1(x), \lambda_2; \delta) \tag{3.7}$$

$$\lambda_2^{\perp} \bar{F}_2^{\perp}(y) = \lambda_2 \bar{F}_2(y) - \lambda^{\parallel} \bar{F}_2^{\parallel}(y) = \lambda_2 \bar{F}_2(y) - \mathcal{C}(\lambda_1, \lambda_2 \bar{F}_2(y); \delta) \tag{3.8}$$

$$\lambda^{\parallel} \bar{F}^{\parallel}(x, y) = \mathcal{C}(\lambda_1 \bar{F}_1(x), \lambda_2 \bar{F}_2(y); \delta) \tag{3.9}$$

To understand the first two equalities, note that the tail integral is the expected number of jumps with size greater than a certain number (e.g., x), so the the expected number of jumps of the marginal CPP is the sum of expected numbers of its independent part and joint jump part. The last equality is through Sklar's Theorem applied to a bivariate CPP. Note that the left hand side uses only the tail integral of jumps, as a result of the fact that the joint tail integral has nothing to do with the independent parts. This is because independent Lévy processes have 0 tail integral, as shown in the next section.

3.2.2 Boundary of Lévy copula

This section studies two extreme dependence cases of bivariate spectrally non-negative Lévy process. We also explain the range of dependence studied in this thesis.

Proposition 3.2.2. *Let(X_t, Y_t) be a Lévy process with independent components. Then its Lévy measure is $\nu(A) = \nu_1(A_x) + \nu_2(A_y)$, where ν_1 and ν_2 are the marginal Lévy measures for X_t and Y_t respectively, and $A_x = \{x : (x, 0) \in A\}, A_y = \{y : (0, y) \in A\}$. Then the tail integral of this Lévy measure is $U(x_1, x_2) = U_1(x_1)1_{x_2=0} + U_2(x_2)1_{x_1=0}$. The corresponding Lévy copula is:*

$$\mathcal{C}_\perp(u_1, u_2) = u_1 1_{u_2=\infty} + u_2 1_{u_1=\infty} \tag{3.10}$$

To define complete dependence of Lévy processes, we need the notion of an increasing set:

Definition A subset of $\bar{R}^2$, where $\bar{R}$ is the extended real number including both positive and negative infinity, is called increasing if for any two vectors $(v_1, v_2) \in S$ and $(u_1, u_2) \in S$ either $v_k < u_k\ \forall k$ or $v_k > u_k\ \forall k$.

From the above definition, we can see the order of an element in an increasing set is completely determined by one coordinate only.

Definition Let (X_t, Y_t) be a Lévy process with positive jumps. Its jumps are said to be **completely dependent** or **comonotonic** if there exists an increasing set S in $[0, \infty]^2$ such that every jump $(\triangle X_t, \triangle Y_t)$ is in **S**.

Proposition 3.2.3. *Let (X_t, Y_t) be a Lévy process with positive jumps. If its jumps are completely dependent, then (a possible) Lévy copula is the complete dependence Lévy copula, which is defined by*

$$\mathcal{C}_\parallel(u, v) = min(u, v) \tag{3.11}$$

Conversely, if the Lévy copula of (X_t, Y_t) is given by $\mathcal{C}_\parallel$ and the tail integrals of components of (X_t, Y_t) are continuous, then the jumps of them are completely dependent.

This complete dependence restricts positive jump sizes to be concordant and so is a type of positive dependence. Complete negative dependence is able to be constructed in another way: for example, modify the set K to be set of jumps with opposite directions and define a *decreasing set* $S \in R^2$, that is, a set S such that for every two vectors $(v_1, v_2) \in S$ and $(u_1, u_2) \in S$, either $v_1 < u_1$ and $v_2 > u_2$ or $v_1 > u_1$ and $v_2 < u_2$, in another word, they are discordant. The corresponding Lévy copula for a 2-d Lévy process becomes $\mathcal{C}_(u, v) := -\min(|u|, |v|)$.

It is well known that negative dependence is hard to model and has too many restrictions with high dimensional random variables. In this thesis, we will avoid such Lévy copula with such extreme negative dependence structures. Instead, the lower bound of Lévy copula is set as the independent copula and upper bound as complete positive dependence. This is why the dependence parameter of Clayton Lévy copula is restricted to be greater than 0.

3.3 INFERENCE FOR BIVARIATE COMPOUND POISSON PROCESS WITH LÉVY COPULA

There are several papers devoted to this topic. Esmaeili and Klüppelberg [2010a] studied the bivariate CPP with exponential marginal claim size and Clayton Lévy copula dependence structure; then they used the full likelihood method to estimate the marginal and dependence parameters. Esmaeili and Klüppelberg [2010b] extended the full likelihood method to model bivariate stable Lévy process, which is transferred to a bivariate CPP by truncating small jumps. Then Esmaeili and Klüppelberg [2010c] considered a two-step parametric estimation method for a bivariate stable Lévy process. This two-step method first estimates marginal parameters and then estimates the dependence parameter by the full likelihood with plug-in marginal parameters from the first step. In this thesis, the above parametric methods are extended to either a nonparametric or semi-parametric inference method, which aim to provide more robust estimations. The full likelihood method is introduced first and its advantages and disadvantages are also analyzed. Next, a non-parametric method based on Kendall's τ and a new semi-parametric two-stage estimator are proposed. The consistency and asymptotic variance of the new estimators are also studied.

3.3.1 Full likelihood method

To implement the full likelihood method, CPP is assumed to be observed at a high frequency, that is, all the jump times and sizes are available. Following the decomposition method in section 3.1, we denote $N(T) = n$ the total number of jumps occurring in [0,T], and decompose it into number $N_1^{\perp}(T) = n_1^{\perp}$ of jumps which only occurs in $S_1(t)$, the number $N_2^{\perp}(T) = n_2^{\perp}$ of jumps which only occurs in $S_2(t)$ and $N^{\parallel}(T) = n^{\parallel}$ jumps in $S^{\parallel}(t) = (S_1^{\parallel}(t), S_2^{\parallel}(t))$. The corresponding jump sizes which only occur in $S_1(t)$ are $\tilde{x}_1, \ldots, \tilde{x}_{n_1^{\perp}}$, jump sizes which only occur in $S_2(t)$ are $\tilde{y}_1, \ldots, \tilde{y}_{n_2^{\perp}}$, and the observed jumps which occurs in both components are $(x_1, y_1), \ldots, (x_{n^{\parallel}}, y_{n^{\parallel}})$. Esmaeili and Klüppelberg [2010a] proposed the following full likelihood for estimating marginal and dependence parameters:

Theorem 3.3.1. *Consider the above observation scheme. Assume that θ_1 is a parameter of the marginal density f_1 of the first jump component only, and θ_2 a parameter of the marginal density f_2 of the second jump component only, and δ is the Lévy copula. Assume further that $\frac{\partial^2}{\partial u \partial v}\mathcal{C}(u, v; \delta)$ exists for all $(u, v) \in (0, \lambda_1) \times (0, \lambda_2)$, which is the domain of (C). Then the full likelihood of the bivariate CPP is*

$$L(\lambda_1, \lambda_2, \theta_1, \theta_2, \delta)$$

$$= (\lambda_1)^{n_1^{\perp}} e^{-(\lambda_1^{\perp})T} \prod_{i=1}^{n_1^{\perp}} \left[f_1(\tilde{x}_i; \theta_1) \left(1 - \frac{\partial}{\partial u}\mathcal{C}(u, \lambda_2; \delta) \mid_{u=\lambda_1 \overline{F}_1(\tilde{x}_i;\theta_1)} \right) \right] \tag{3.12}$$

$$\times (\lambda_2)^{n_2^{\perp}} e^{-(\lambda_2^{\perp})T} \prod_{i=1}^{n_2^{\perp}} \left[f_2(\tilde{y}_i; \theta_2) \left(1 - \frac{\partial}{\partial v}\mathcal{C}(\lambda_1, v; \delta) \mid_{u=\lambda_2 \overline{F}_2(\tilde{y}_i;\theta_2)} \right) \right] \tag{3.13}$$

$$\times (\lambda_1 \lambda_2)^{n^{\parallel}} e^{-(\lambda^{\parallel})T} \prod_{i=1}^{n^{\parallel}} \left[f_1(x_i; \theta_1) f_2(y_i; \theta_2) \frac{\partial^2}{\partial u \partial v}\mathcal{C}(u, v; \delta) \mid_{u=\lambda_1 \overline{F}_1(x_i;\theta_1), v=\lambda_2 \overline{F}_2(y_i;\boldsymbol{\theta_2})} \right] \tag{3.14}$$

with $\lambda^{\parallel}(\delta) = \mathcal{C}(\lambda_1, \lambda_2, \delta)$ and $\lambda_i^{\perp}(\delta) = \lambda_i - \lambda^{\parallel}(\delta)$ for $i = 1, 2$.

This form of the full likelihood is derived based on the decomposition of the bivariate CPP, which helps us to understand the dependence structure of CPP and naturally leads to the three parts of the full likelihood, corresponding to $S_1^{\perp}(t)$, $S_1^{\perp}(t)$ and $S^{\parallel}(t) = (S_1^{\parallel}(t), S_2^{\parallel}(t))$ respectively. Each part is the product of exponential inter-arrival times and jump size densities, based on the likelihood method of one-dimensional compound Poisson process by Basawa and Prakasa Rao [1980]. Although the full likelihood is easy to implement and generally more efficient, it has several problems: first, the single jump likelihood involves with Lévy copula, that is, the dependence structure, which would make the estimates of marginal parameters be indirectly affected by the Lévy copula, and vice versa. Second, the numerical optimization often fails when extending the full likelihood method to high dimensional CPP, which usually has too many parameters to estimate. Last, it requires specification of marginal jump size distributions, leading to estimates that are not robust.

3.3.2 Nonparametric Estimation by Kendall's Coefficient of Concordance

When the dependence parameter in Lévy copula is of primary interest, one may desire a nonparametric method, which avoids any specific assumption on the marginal jump size distribution and so is more robust than full likelihood method. Next, a simple nonparametric estimator based on Clayton Lévy copula is proposed. Our strategy is to connect Lévy copula with the copula of the jump size distribution and then estimate the dependence parameter of the Lévy copula by using the relationship between the ordinary copula and Kendall's τ estimator.

First we assume $\overline{C}(u, v)$ is the ordinary survival copula of the joint jumps sizes denoted by $(S_1(t)^{\parallel}, S_2(t)^{\parallel})_{t\geq 0}$, satisfying $\bar{F}^{\parallel}(x, y) = \overline{C}(\bar{F}_1^{\parallel}(x), \bar{F}_2^{\parallel}(y))$. Then assuming that Lévy copula $\mathcal{C}$ is continuous, we have the following equations:

$$\begin{aligned}
&\lambda^{\parallel}\bar{F}^{\parallel}(x, y) = \mathcal{C}(\lambda_1\bar{F}_1(x), \lambda_2\bar{F}_2(y)) \\
\Rightarrow\ &\bar{F}_1^{\parallel}(x) = \lim_{y\downarrow 0^+} \frac{1}{\lambda^{\parallel}}\mathcal{C}(\lambda_1\bar{F}_1(x), \lambda_2\bar{F}_2(y)) \\
\Rightarrow\ &\bar{F}_2^{\parallel}(y) = \lim_{x\downarrow 0^+} \frac{1}{\lambda^{\parallel}}\mathcal{C}(\lambda_1\bar{F}_1(x), \lambda_2\bar{F}_2(y)) \\
\Rightarrow\ &\bar{C}(\bar{F}_1^{\parallel}(x), \bar{F}_2^{\parallel}(y)) = \bar{C}\left(\frac{1}{\lambda^{\parallel}}\mathcal{C}(\lambda_1\bar{F}_1(x), \lambda_2), \frac{1}{\lambda^{\parallel}}\mathcal{C}(\lambda_1, \lambda_2\bar{F}_2(y))\right).
\end{aligned}$$

Thus we have

$$\bar{C}\left(\frac{1}{\lambda^{\|}}\mathcal{C}(\lambda_1\bar{F}_1(x),\lambda_2),\frac{1}{\lambda^{\|}}\mathcal{C}(\lambda_1,\lambda_2\bar{F}_2(y))\right)=\frac{1}{\lambda^{\|}}\mathcal{C}\left(\lambda_1\bar{F}_1(x),\lambda_2\bar{F}_2(y)\right). \tag{3.15}$$

Next, we apply the above results to the broader class of Archimedean Lévy copula, which includes the Clayton Lévy copula as a special case. To obtain Archimedean Lévy copula, we define the generator function ϕ, which is a strictly decreasing convex function from $[0,\infty]$ to $[0,\infty]$ such that $\phi(0)=\infty$ and $\phi(\infty)=0$. Then it can be shown that

$$F(u,v)=\phi^{-1}(\phi(u)+\phi(v))$$

defines a two dimensional spectrally positive positive Lévy copula. For example, letting $\phi(u)=u^{-\delta}$ where $\delta>0$, we obtain the Clayton Lévy copula:

$$F_\delta(u,v)=(u^{-\delta}+v^{-\delta})^{-1/\delta},$$

which includes as limiting cases complete dependence (when $\delta\to\infty$) and independence (when $\delta\to 0$). By 3.9, we have:

$$\lambda^{\|}=\phi^{-1}(\phi(\lambda_1)+\phi(\lambda_2)).$$

By using 3.15, we can represent the survival copula of jump sizes as:

$$\begin{aligned}&\bar{C}\left(\frac{1}{\lambda^{\|}}\phi^{-1}(\phi(\lambda_1\bar{F}(x))+\phi(\lambda_2)),\frac{1}{\lambda^{\|}}\phi^{-1}(\phi(\lambda_1)+\phi(\lambda_2\bar{F}_2(y)))\right)\\ &=\frac{1}{\lambda^{\|}}\phi^{-1}\left(\phi(\lambda_1\bar{F}_1(x)+\phi(\lambda_2\bar{F}_2(y)))\right).\end{aligned}$$

Equating first and second arguments of $\bar{C}$ to u and v respectively, we have:

$$\begin{aligned}\phi(\lambda_1\bar{F}_1(x))+\phi(\lambda_2)&=\phi(\lambda^{\|}u),\\ \phi(\lambda_1)+\phi(\lambda_2\bar{F}_2(y))&=\phi(\lambda^{\|}v).\end{aligned}$$

Thus, the survival copula of jump sizes can be simplified as:

$$\bar{C}(u,v)=\frac{1}{\lambda^{\|}}\phi^{-1}\left(\phi(\lambda^{\|}u)-\phi(\lambda_2)+\phi(\lambda^{\|}v)-\phi(\lambda_1)\right), \tag{3.16}$$

where $\lambda^{\|}=\phi^{-1}(\phi(\lambda_1)+\phi(\lambda_2))$

By 3.16, it is clear that if the marginal jump frequency is estimated and ϕ is specified, it is possible to estimate the dependence parameter by using the survival copula of jump sizes. Instead of pursuing the more complicated method, we find a simple method by noting that 3.16 is simplified assuming the Clayton Lévy copula, which is independent of the marginal jump frequencies:

$$\bar{C}(u, v) = (u^{-\delta} + v^{-\delta} - 1)^{-1/\delta},$$

where $\delta > 0$ is shared by both the Lévy copula and corresponding survival copula of jump sizes. The resulting survival copula is a special case of the ordinary Clayton copula with strict generator $\phi_\delta(t) = \frac{1}{\delta}(t^{-\delta} - 1)(\delta > 0)$. This is also easily extended to the multivariate Clayton Lévy copula with survival copula of jump sizes:

$$\bar{C}(u_1, u_2, \ldots, u_n) = (u_1^{-\delta} + u_2^{-\delta} + \ldots + u_n^{-\delta} - (n-1))^{-1/\delta},$$

which is still an ordinary Clayton copula.

This leads to the famous nonparametric estimation method for ordinary copula: Kendall's τ method [Nelsen, 1999], which is defined as:

$$t = \frac{c-d}{c+d} = (c-d) \Big/ \binom{n}{2},$$

where c is the number of concordant pairs and d is the number of discordant pairs. Equivalently, it can be understood as an estimate of the probability of concordance minus the probability of discordance for a pair of observations chosen randomly from the sample. Assume that $(X_1, Y_1), (X_2, Y_2)$ are independent and identically distributed random vectors, then the population version of Kendall's τ is defined as:

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

Copula have the following connection with Kendall's τ: Assume that (X_1, Y_1) and (X_2, Y_2) are continuous random variables with joint distribution $H_1(x, y)$ and $H_2(x, y)$ and common margins $F(x)$ and $G(y)$, then:

$$\begin{aligned} Q &= P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \\ &= Q(C_1, C_2) = 4 \int \int_{I^2} C_2(u, v) dC_1(u, v) - 1, \end{aligned}$$

where C_1 and C_2 are the copula of (X_1, Y_1) and (X_2, Y_2) respectively.

It is possible to show that the copula in above can be replaced by a survival copula Nelsen [1999], i.e. $Q(C_1, C_2) = Q(\bar{C}_1, \bar{C}_2)$. For the ordinary Clayton copula, it can be shown that $\tau_\delta = \frac{\delta}{\delta+2}$ and the estimate is $\hat{\delta} = \frac{2\tau_\delta}{1-\tau_\delta}$.

The variance of $\hat{\delta}$ can be calculated according to Oakes [1982], in which δ is reparameterized into $\theta = \delta - 1$, but:

$$var(\delta) = var(\theta) = \frac{(\theta+1)^4}{n}\gamma(\theta),$$

where

$$\gamma(\theta) = \frac{4}{3}\left\{\frac{\theta^3 + 4\theta^2 + 10\theta + 4}{(\theta+1)^2(\theta+2)} - 6J(\theta)\right\},$$

with

$$(\theta-1)^2 J(\theta) = \int_0^1 \int_0^1 \frac{u^p v^{p-1}}{(u+v-uv)^q} du dv,$$

where $p = 2/(\theta-1)$,$q = \theta/(\theta-1)$. The integration is evaluated by an infinite series expansion involving the gamma function. Values of $var(\theta)$ is given in Table 2 of Oakes [1982]. Values of $J(\theta)$ are also given for some sepcial cases, e.g. $J(2) = 3\frac{1}{2} - \pi^2/3 = 0.21013$ and $J(3) = 4\log 2 - 2 = 0.19315$.

Nonparametric estimation of dependence structure, such as the Kendall's τ method proposed above, may not work well due to small sample size. For example when we apply the ordinary method to intra-familial studies [Andersen, 2004], the sample size is sufficient to perform nonparametric estimation of the marginal process but insufficient for inference of dependence structure due to the presence of censoring. The same problem may also apply to the bivariate CPP, when the number of joint jumps is insufficient. Furthermore, the single jumps are not utilized, which is inefficient when the proportion of single jumps is substantial.

3.3.3 Two-Stage Estimator by Conditional Likelihood

The major inference method proposed in this thesis is based on the full likelihood method by Esmaeili and Klüppelberg [2010a]. The full likelihood method is the most efficient but requires specification of the jump size distribution. The nonparametric method can only utilize joint jumps, which is not efficient enough. Furthermore, the full likelihood method needs to specify the parametric form of both the marginal distribution and copula function. Genest et al. [1995] argue that the estimate of δ by using the likelihood method is margin-dependent, while the estimates of marginal parameters can also be affected by copula. The major idea of this thesis is to separate the inference for the marginals and dependence structure, which is usually called two-stage estimation in the copula literature [Andersen, 2004, Shih and Louis, 1995, Genest et al., 1995, Joe, 2005]. When only the dependence structure is of interest, marginal distributions are estimated nonparametrically, while the dependence parameters can be estimated by maximizing the likelihood function where the marginal parts are substituted with the empirical distribution functions. This two-stage strategy turns out to be more practical than the full likelihood method though some efficiency is lost. Unlike the nonparametric method, this strategy estimates marginal quantities nonparametrically while maintaining parametric structure for dependence, which may be able to avoid the case of insufficient information. It can also be easily extended to the multivariate case by constructing a composite likelihood. In this section, we will propose a similar semi-parametric method for Lévy copula based on the conditional likelihood.

3.3.3.1 Construct Conditional Likelihood from Lévy Copula As shown in Cont and Tankov [2004], Lévy copula is not a distribution function, but its derivative still has interesting probability properties. This is because the marginal of the Lévy copula is uniform. Lemma 4.2 in Tankov [2006] states that the derivative of a Lévy copula $\mathcal{C}(u_1, u_2, \ldots, u_d)$ w.r.t. u_1 is a conditional distribution function of $u_2, \ldots, u_d$ on u_1. A simpler version for a bivariate spectrally positive Lévy processes is from Cont and Tankov [2004]:

Lemma 3.3.2. *Let $\mathcal{C}$ be a two-dimensional positive Lévy copula. Then for almost all $u \in [0, \infty]$, the function $\mathcal{C}_u(v)$ exists and is continuous for all $v \in [0, \infty]$ outside a countable set. Moreover, it is a distribution function of a positive random variable, that is, it is increasing and satisfies $\mathcal{C}_u(0) = 0$ and $\mathcal{C}_v(\infty) = 1$.*

By plugging in specific form of the tail integrals into $\mathcal{C}_u(v)$, we have the following theorem:

Theorem 3.3.3. *Let (X_t, Y_t) be two Lévy processes with positive jumps, having marginal tail integrals U_1, U_2 and Lévy copula $\mathcal{C}$. Let $\Delta X_t, \Delta Y_t$ be the sizes of jumps at time t. Then* *if U_1 has a non-zero density at x, $\mathcal{C}_{U_1(x)}(\cdot)$ is the distribution function of $U_2(\Delta Y_t)$ conditional on $\Delta X_t = x$:*

$$\mathcal{C}_{U_1(x)}(y) = Pr\{U_2(\Delta Y_t) < y | \Delta X_t = x\}. \tag{3.17}$$

Next, differentiate the distribution function above with respect to y to get the conditional likelihood for a single observation $(U_1(x), U_2(y))$:

$$\left.\frac{\partial^2 \mathcal{C}}{\partial u \partial v}\right|_{u=U_1(x), v=U_2(y)}.$$

Note that this is a factor of (3.14) in the full likelihood, which is the likelihood of joint jumps and derived by using the relationship between bivariate Lévy density and Lévy copula [Cont and Tankov, 2004]:

$$\nu(x,y) = \left.\frac{\partial^2 \mathcal{C}(u,v)}{\partial u \partial v}\right|_{u=U_1(x), U_2(y)} \nu_1(x)\nu_2(y).$$

To construct the likelihood function of joint jumps from the Lévy density $\nu(x,y)$, note that the Lévy density can also be written as $\lambda^{\|} f^{\|}(x,y)$. Then the likelihood $f^{\|}(x,y)$ of joint jumps can be easily obtained after dividing $\nu(x,y)$ by the mean jump rate of joint jumps $\lambda^{\|}$. But (3.14) has two factors $\nu_1(x)$ and $\nu_2(y)$ containing no information about the dependence parameter. This leads us to use $\left.\frac{\partial^2 \mathcal{C}}{\partial u \partial v}\right|_{u=U_1(x), v=U_2(y)}$ only as the likelihood, which is exactly the conditional likelihood. Dividing the conditional density by $\lambda^{\|}$ and multiplying $\nu_1(x)$ and $\nu_2(y)$ leads to the full likelihood of joint jumps.

The conditional likelihood is actually based on the tail integrals $U_1(x)$ and $U_2(y)$:

$$L(\delta, U_1, U_2 | x, y) = \left.\frac{\partial^2 \mathcal{C}(u,v;\delta)}{\partial u \partial v}\right|_{u=U_1(x), v=U_2(y)}, \tag{3.18}$$

where U_1 and U_2, in the case of bivariate CPP, depend on empirical estimates of the marginal jump size distribution.

The above results are based on the assumptions that $u > 0, v > 0$, but the tail integrals of the CPP are bounded as they are Lévy process with finite variation. More precisely, we have $u \in (0, \lambda_1), v \in (0, \lambda_2)$. Then the correct form of the likelihood should be a truncated version of conditional likelihood:

$$L(\delta, U_1, U_2 | x, y) = \frac{1}{\mathcal{C}(\lambda_1, \lambda_2; \delta)} \left.\frac{\partial^2 \mathcal{C}(u,v;\delta)}{\partial u \partial v}\right|_{u=U_1(x), v=U_2(y)}, \tag{3.19}$$

The above method can be easily adapted to the multivariate case. It can be shown that given the tail integral of the first component, the conditional likelihood of other tail integrals is:

$$
\begin{aligned}
L(\delta, U_1, U_2, \ldots, U_n | x_1, x_2, \ldots, x_n) &= \frac{1}{\mathcal{C}(\lambda_1, \lambda_2, \ldots, \lambda_n; \delta)} \cdot \\
&\quad \left. \frac{\partial^2 \mathcal{C}(u_1, u_2, \ldots, u_n; \delta)}{\partial u_1 \partial u_2, \ldots, \partial u_n} \right|_{u_1=U_1(x_1), u_2=U_2(x_2), \ldots, u_n=U_n(x_n)} .
\end{aligned}
$$

3.3.3.2 Two-Stage Strategy For Bivariate Case Based On Conditional Likelihood To obtain the two-stage estimate, the parameters in the margins are estimated at the first stage. For each dimension, recall the assumptions: the jump times are $t_{i,0} = 0, t_{i,1}, \ldots, t_{i,n_i} = T$ for i=1,2 and the jump sizes are $\mathbf{X} = (x_1, \ldots, x_{n_1})$ and $\mathbf{Y} = (y_1, \ldots, y_{n_2})$. The parameter of the underlying Poisson process can be estimated by using the mean inter-arrival time $T_{ij} = t_{i,j} - t_{i,j-1}$ for $i = 1, 2$ and $j = 1, 2, \ldots, n_i$. Then $\hat{\lambda}_i = (\bar{T}_i)^{-1}$. Its variance can be asymptotically estimated by the delta method to get: $\hat{\lambda}_i^2 / n_i$

The jump size distribution can be estimated nonparametrically. We use the empirical cumulative distribution function: $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$, where $n = n_1$ or n_2. For fixed x, $I(x_i \leq x)$ is a Bernoulli random variable with parameter p=F(x), hence $n\hat{F}_n(x)$ is a binomial random variable with mean $nF(x)$ and variance $nF(x)(1 - F(x))$.And we know by the central limit theorem: $\sqrt{n}(\hat{F}_n(x) - F(x)) \to N(0, F(x)(1 - F(x)))$. To avoid problems when $\hat{F}_n(x)$ equal to 1, $\frac{n}{n+1}\hat{F}_n(x)$ is used instead, which is asymptotically equivalent to $\hat{F}_n(x)$. The corresponding survival function $\hat{\bar{F}}_n(x) = 1 - \frac{n}{n+1}\hat{F}_n(x)$ is plugged into the conditional likelihood.

In the second stage, we will plug the estimated marginal components into the conditional likelihood, which is constructed from all the joint jumps (independent jumps will lead to 0 tail integrals). Assume that the joint jumps are $(x_1, y_1), (x_2, y_2), \ldots, (x_{n^{\parallel}}, y_{n^{\parallel}})$, where $n^{\parallel}$ is the number of joint jumps. Then we have the following profile likelihood function:

$$
L(\delta) = \prod_{i=1}^{n^{\parallel}} f(\delta, \hat{U}_1(x_i), \hat{U}_2(y_i)) = \frac{1}{\mathcal{C}(\lambda_1, \lambda_2; \delta)^{n^{\parallel}}} \prod_{i=1}^{n^{\parallel}} \left. \frac{\partial^2 \mathcal{C}(u, v; \delta)}{\partial u \partial v} \right|_{u=\hat{U}_1(x_i), v=\hat{U}_2(y_i)} . \tag{3.20}
$$

To estimate the dependence parameter δ, we will adopt the classical score method by taking the derivative of log likelihood with respect to δ. By equating score to 0, a maximum profile likelihood estimate of δ is finally obtained. Consistency and asymptotic normality of the estimator can be obtained under certain regularity conditions (appendix B.1.1). A variance estimator can be calculated based on a von Mises expansion.

3.3.3.3 Consistency and Asymptotic Normality The bivariate likelihood is a function of two types of parameters: $\beta = (\delta, \phi_1, \phi_2)$, including the dependence parameter δ and nuisance parameter $\phi = (\phi_1, \phi_2)$, where $\phi_i = (\bar{F}_i, \lambda_i)$, $i = 1, 2$. Assume that we already obtained the consistent estimate of the marginal frequency parameters $\hat{\lambda}_1, \hat{\lambda}_2$ and empirical estimator of marginal size distributions $\hat{\bar{F}}_1, \hat{\bar{F}}_2$. In addition, assume that the true values of parameters are δ_0 and $\phi_0 = (\phi_{10}, \phi_{20}) = ((\lambda_{10}, \bar{F}_{10}), (\lambda_{20}, \bar{F}_{20}))$.

The likelihood is $L(\delta, \hat{\phi}) = \prod_i f(\delta, \hat{\lambda}_1, \hat{\bar{F}}_1(x_i^{\|}), \hat{\lambda}_2, \hat{\bar{F}}_2(y_i^{\|}))$ and the corresponding log likelihood is

$$l(\delta, \hat{\phi}) = \log(L(\delta, \hat{\phi})) = \sum_i l_i(\delta, \hat{\phi}) = \sum_i \log f(\delta, \hat{\lambda}_1, \hat{\bar{F}}_1(x_i^{\|}), \hat{\lambda}_2, \hat{\bar{F}}_2(y_i^{\|})).$$

We need the following lemma to prove the consistency of semiparametric likelihood estimator; see proofs and regulation conditions **A0-A5** in appendix B.1.

Lemma 3.3.4. *Under regularity conditions **A0-A2** and **A4** and assume $\hat{\phi} \to \phi_0$ in probability. Then,*

$$P_{\delta_0,\phi_0}(l(\delta, \hat{\phi}) < l(\delta_0, \hat{\phi})) \to 1 \text{ as } n \to \infty$$

for any fixed $\delta \neq \delta_0$.

Then we can determine the MLE $\hat{\delta}$ of δ by equating the score $\frac{\partial l(\delta,\hat{\phi})}{\partial \delta}$ to 0. The following theorem proves the consistency of the MLE $\hat{\delta}$ by placing it in the neighborhood of true value δ_0, then using a Taylor expansion of likelihood function around δ_0, linearizing the estimator and applying the central limit theorem.

Theorem 3.3.5. (*Consistency) Under condition **A0-A4**, the equation:*

$$l_\delta(\delta, \hat{\phi}) = \sum_i \frac{f_i'(\delta, \hat{\phi})}{f_i(\delta, \hat{\phi})} = 0$$

has a root $\hat{\delta} = \hat{\delta}(X, Y)$ such that $\hat{\delta}$ tends to the true value δ_0 in probability.

By using the above lemma, the proof is the same as Theorem 3.7 of Lehmann [1998].

Asymptotic normality of the estimator is established based on the von Mises expansion of the likelihood function around $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\bar{F}}_1$ and $\hat{\bar{F}}_2$ which linearizes the estimator $\hat{\delta}$, and uses the asymptotic normality of the above marginal components. Define the Fisher information of δ as $I_{\delta\delta} = E[\frac{\partial l(\delta,\hat{\phi})}{\partial \delta}]^2$. We have the following theorem of asymptotically normality:

Theorem 3.3.6. Asymptotic normality (semiparametric form) *Assume $n_i \to \infty, n \to \infty$ and $n_i - n \to \infty$ for $i = 1, 2$. Under regularity conditions **A0-A5**, the estimate $\hat{\delta}$ from 3.3.5 is asymptotic normal, that is, $\sqrt{n}(\hat{\delta} - \delta_0)$ is asymptotically normal with mean 0 and variance v^2, where:*

$$v^2 = \frac{1}{I_{\delta\delta}} + \frac{\tau}{I_{\delta\delta}^2},$$

where $\tau = E[(\Phi_1)^2] + E[(\Phi_2)^2] + E[(\Phi)^2]$. Φ's are given in the Appendix B.1.

This also shows that the two-stage estimator is not as efficient as the full likelihood estimator. The proof and the variance estimator are also given in Appendix B.1.

3.4 SIMULATION

Simulation studies were conducted to assess the proposed method and compare it with other methods. Various simulation algorithms for Lévy processes are discussed in section 6 of Cont and Tankov [2004]. There are generally two methods to simulate bivariate dependent CPP: one uses Theorem 6.3 and Algorithm 6.15 in Cont and Tankov [2004], which extend the Rosinski's series approximation to a multivariate setting. Another extends Theorem 6.2 of Cont and Tankov [2004] by decomposition of the bivariate CPP for given λ_1, λ_2, marginal jump distribution functions F_1,F_2 and the Lévy copula $\mathcal{C}$, as shown by Esmaeili and Klüppelberg [2010a]. This method only works for the simple Lévy copula, e.g. Clayton Lévy copula, where the ordinary copula for jump sizes can be easily derived from the Lévy copula. The second method will be used for simulations in this thesis.

Given the intensity parameters λ_1,$\lambda_2 > 0$, marginal jump size distributions $F_1 = Exp(\theta_1)$ and $F_2 = Exp(\theta_2)$, dependence parameter δ and the time interval $[0, T]$, a bivariate CPP can be simulated in two steps:

First, generate the number of jump times (N_1, N_2) in $S_1(t)$ and $S_2(t)$ as Poisson random variables with parameters $\lambda_1 T$ and $\lambda_2 T$ respectively. Then generate the number of jump times $N^{\|}$ in the joint jump process $S^{\|}(t) = (S_1^{\|}(t), S_2^{\|}(t))$ as Poisson random variables with parameter $\lambda^{\|}T = \mathcal{C}(\lambda_1, \lambda_2; \delta)T$. This implies the number of single jumps of $S_1(t)$ and $S_2(t)$ are $N_1^{\perp} = N_1 - N^{\|}$ and $N_2^{\perp} = N_2 - N^{\|}$ respectively.

Conditional on these numbers, simulate $N_1^{\perp}$ independent uniform random variables on $[0, T]$: $T_{1,i}^{\perp}$, $i = 1, \ldots, N_1^{\perp}$, $N_2^{\perp}$ independent uniform random variables on $[0, T]$: $T_{2,i}^{\perp}$, $i = 1, \ldots, N_2^{\perp}$ and $N^{\|}$ independent uniform random variables on $[0, T]$: $T_i^{\|}$ for $i = 1, \ldots, N^{\|}$. $T_{1,i}, T_{2,i}$ and $T_i^{\|}$ are also mutually independent by the decomposition of the bivariate CPP. Based on the theorem that jump times on $[0, T]$ are uniformly distributed conditional on Poisson number of jumps, these jump uniform random variables are exactly the jump times.

Next simulate jump sizes: denote the jump time and size pair as $(T_{1,i}^{\perp}, X_i^{\perp})$, $i = 1, \ldots, N_1^{\perp}$, $(T_{2,i}^{\perp}, Y_i^{\perp})$, $i = 1, \ldots, N_2^{\perp}$ and $(T_i^{\perp}, (X_i^{\|}, Y_i^{\|}))$, $i = 1, \ldots, N^{\|}$ for $S_1^{\perp}(t)$, $S_2^{\perp}(t)$ and $S^{\|}(t)$ respectively, so that the trajectory of bivariate CPP on $[0, T]$ is given by:

$$\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N_1^{\perp}} 1_{\{T_{1,i}^{\perp} < t\}} X_i^{\perp} + \sum_{i=1}^{N^{\|}} 1_{\{T_i^{\|} < t\}} X_i^{\|} \\ \sum_{i=1}^{N_2^{\perp}} 1_{\{T_{2,i}^{\perp} < t\}} Y_i^{\perp} + \sum_{i=1}^{N^{\|}} 1_{\{T_i^{\|} < t\}} Y_i^{\|} \end{pmatrix}. \tag{3.21}$$

The marginal and joint jumps sizes are generated by the law derived from 3.7,3.8 and 3.9. $N_1^{\perp}$ independent jump sizes are generated by $\bar{F}_1^{\perp}(t)$ derived from 3.7, $N_2^{\perp}$ independent jump sizes are generated with $\bar{F}_2^{\perp}(y)$ derived from 3.8. These single jumps are generated by the probability inverse transformation. To simulate the $N^{\|}$ joint jumps with bivariate distribution $\bar{F}^{\|}(x, y)$, we need to derive the conditional distributions of x in $S_1^{\|}(x)$ given y in $S_2^{\|}(y)$ by 3.9. Given Clayton Lévy copula $\mathcal{C}(u, v; \delta)$, the survival copula of jump sizes is an ordinary Clayton copula with the same dependence parameter, denoted as $\bar{C}(u, v; \delta)$. It can be shown that

$$\lim_{\Delta x \to 0} P(Y^{\|} > y \big| x < X^{\|} \leq x + \Delta x) = \frac{\partial}{\partial u} \bar{C}(u, \bar{F}_2^{\|}(y); \delta)\big|_{u = \bar{F}_1^{\|}(x)} =: \bar{H}_x(y).$$

Based on the above results, we generate two sets of independent uniform variables $u_1, \ldots, u_{N^{\|}}$ and $v_1, \ldots, v_{N^{\|}}$ first. We then generate $X_j^{\|} = \bar{F}_1^{\|-}(u_j)$ and $Y_j^{\|} = H_{X_j}^{-}(v_j)$ conditional on X_j, where $j = 1, 2, \ldots, N^{\|}$, where $\bar{F}_1^{\|}(x)$ and $\bar{F}_2^{\|}(y)$ can be derived from $\bar{F}^{\|}(x, y)$ by equating the alternative variable to ∞.

For the bivariate CPP with exponential jump sizes and Clayton Lévy copula, the conditional likelihood is

$$\bar{H}_x(y) = \left(\frac{\lambda_1^{-\delta} e^{\theta_1 \delta x} + \lambda_2^{-\delta} e^{\theta_2 \delta y}}{\lambda_1^{-\delta} e^{\theta_1 \delta x} + \lambda_2^{-\delta}} \right)^{-\frac{1}{\delta}-1}.$$

Four different scenarios were simulated to assess the proposed method. Two different jump frequencies are implemented: $\lambda_1 = 200$, $\lambda_2 = 160$ or $\lambda_1 = 400$, $\lambda_2 = 320$, which will generate around 2700 and 5400 samples respectively, having the same magnitude as the 2157 losses in the Danish fire data. And two values of dependence parameter are used: $\delta = 1$ or $\delta = 2$. The jump size distribution parameters θ_1 and θ_2 are set to 1 and 2. 500 simulations were performed for each scenario. Simulation results are shown in Table 7. The simulations results show that: 1. both the nonparametric and semi-parametric estimators are consistent; 2. variance parameter of semi-parametric estimator works better when sample size gets larger; 3. semiparametric estimator is more efficient than non-parametric method; 4. finally, the coverage rates become closer to the nominal 95% as sample size increase, which indicate asymptotically normality works.

3.5 DANISH FIRE INSURANCE DATA

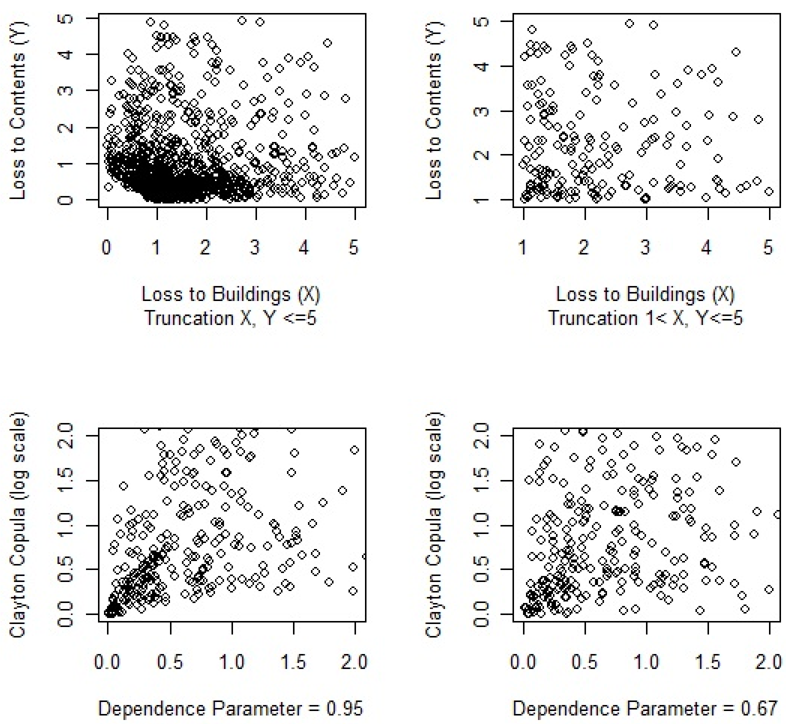
The Danish fire data consist of 2167 losses of over one million Danish Krone from the years 1980 to 1990 inclusive. The total losses are divided into three categories: damage to buildings, damage to contents (furniture and personal property) and damage to profits. Because the third component rarely has losses, we will only focus on the dependence between the loss of building and loss of content. We also apply our methods to the log losses as in Esmaeili and Klüppelberg [2010a]. After inflation adjustment to the 1985 value and splitting the total loss to three components, the losses become lower than one million Danish Krone. Esmaeili and Klüppelberg [2010a] suggested to use the data which are greater than one

Table 7: Simulation results for **bivariate CPP** with **Lévy copula**: mean of estimates (EST), mean of standard error (MSE), empirical standard error (ESE), coverage rate of 95% confidence interval (COV)

λ_1	λ_2	δ	Stat	Two-stage semi-parametric estimator	Nonparametric estimator (Kendall's τ)
			EST	1.01	1.00
200	160	1.00	MSE	0.068	0.094
			ESE	0.062	0.094
			COV	0.98	0.96
			EST	2.00	2.01
200	160	2.00	MSE	0.097	0.12
			ESE	0.094	0.12
			COV	0.96	0.96
			EST	1.00	1.00
400	320	1.00	MSE	0.048	0.066
			ESE	0.044	0.064
			COV	0.97	0.96
			EST	2.00	2.00
400	320	2.00	MSE	0.068	0.086
			ESE	0.068	0.085
			COV	0.96	0.96

million Danish Krone after inflation adjustment in both coordinates. We also propose to use this truncated data in order to get a better goodness of fit to the Clayton Lévy copula, which can be visually checked by the Figure 1. We compare the scatter plot of joint losses size and the targeted copula function to do such visual check. The top two panels of the figure are the scatter plots of joint non-zero losses with different truncation windows, while the bottom two are simulated ordinary Clayton copula with different dependence parameters. Visual inspection indicates that the data which are greater than one in both coordinates (right top) is much closer to the simulated Clayton copula with dependence parameter 0.67 (right bottom) than the original data (left top). The simulated Clayton copula with parameter 0.95 (left bottom) seems different from the truncated data.

Figure 1: Comparison between Danish Fire Data and Clayton Copula.



After truncation, we have totally 940 data points available (Figure 2), 298 of which are joint jumps. By simulation results, such amount of joint jumps are enough to guarantee the convergence of variance estimator.

Table 8 shows that the estimates by the non-parametric method and the two-stage semiparametric method are lower than that of the full-likelihood method, which is obtained from Esmaeili and Klüppelberg [2010a]. In their paper, Weibull distribution is assumed for the marginal losses in the full likelihood method. Histogram and QQ plot methods were performed to check the appropriateness of the Weibull assumption. Although the Weibull distribution fits the marginal losses very well, the estimates of marginal parameters by MLE are very different from those by the full likelihood method, which is probably due to unstable numerical methods used. Thus, the resulting dependence parameter estimate (0.953) is unreliable, because they are affected by the marginal estimates. Visual inspection also verifies this point. On the other hand, the proposed method in this thesis uses two-stage strategy, so that marginal and dependence parameters don't affect each other when performing numerical maximization procedures. We can also see that semiparametric estimator is more efficient than the nonparametric estimator while less efficient than the full likelihood method. The nonparametric estimate is also very close to the semiparametric two-stage estimate, which shows that the assumption of Clayton Lévy copula is approximately valid.

3.6 DISCUSSION

This thesis focuses only on the two-stage inference of the spectrally positive Lévy copula. There are several directions to extend current work:

- Develop a method to test the goodness of fit for the Lévy copula.
- Extend to general bivariate Lévy process with both positive and negative jumps. For the Lévy process with infinite variation, a small jumps truncation rule needs to be specified. This extension can be applied to stock price association analysis.
- Extend to the multivariate case with associative dependence structure. When the multivariate density function is not easy to derive by taking derivatives with respect to each component in the Lévy copula, a composite pairwise likelihood can be utilized. In this case, a two-stage estimator becomes more important because a full likelihood method in the multivariate case does not appear to be feasible.

Although the two-stage strategy leads to a less efficient estimator than the full likelihood, its simplicity is still attractive. As long as the sample size is large enough, the two-stage estimator is still a good choice.

Figure 2: Trajectory of Danish Fire Data

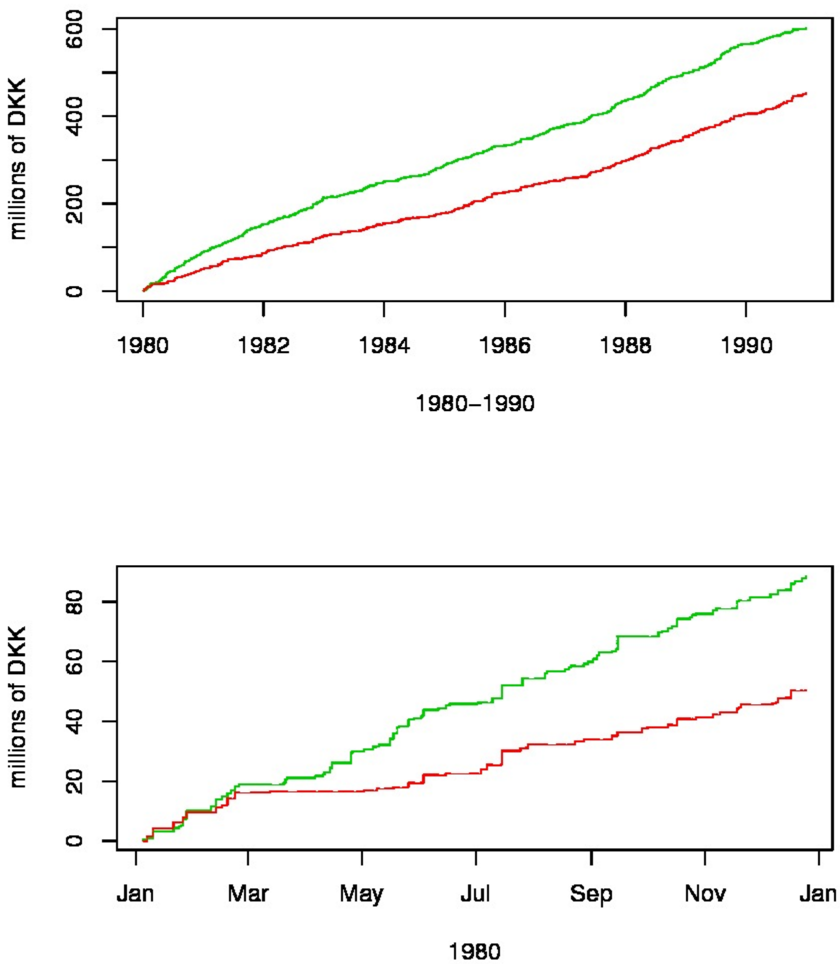


Table 8: Dependence between Loss to Building and Loss to Contents from Danish Fire Data

	Two-stage semi-parametric estimator	Nonparametric estimator (Kendall's τ)	Full Likelihood
$\hat{\delta}$(SD)	0.675(0.088)	0.546(0.112)	0.953(0.030)

APPENDIX A

CAUSE SPECIFIC ASSOCIATION MEASURE

A.1 ASYMPTOTIC PROPERTIES OF $\hat{\zeta}_{CP}$

Recall the extended association measure for mother-children data within region Ω_{qr}

$$\zeta_{CP}(\Omega_{qr}) = \frac{\int\int_{\Omega_{qr}} w(s,t)EN_{11}^*(ds,dt)EH^*(s,t)}{\int\int_{\Omega_{qr}} w(s,t)EN_{10}^*(ds,t)EN_{01}^*(s,dt)},$$

where N_{11}^*, N_{10}^*, N_{01}^* and H^* are the double or single event processes and at-risk process defined over all mother-child pairs; for instance, $N_{11}^*(s,t) = \sum_j I(T_j \leq s, \epsilon_j = 1, T_0 \leq t, \epsilon_0 = 1)$. We now consider asymptotic properties of the nonparametric plug-in estimator

$$\hat{\zeta}_{CP}(\Omega_{qr}) = \frac{\int\int_{\Omega_{qr}} \hat{w}(s,t)\mathbb{P}N_{11}^*(ds,dt)\mathbb{P}H^*(s,t)}{\int\int_{\Omega_{qr}} \hat{w}(s,t)\mathbb{P}N_{10}^*(ds,t)\mathbb{P}N_{01}^*(s,dt)},$$

given in (2.6).

Following the arguments in Cheng and Fine [2008], we establish consistency and asymptotic normality as below. For simplicity, in the sequel, we omit the integration region Ω_{qr}. Observe that

$$\hat{\zeta}_{CP} - \zeta_{CP} = \frac{A}{\{\int\int \hat{w}(s,t)\mathbb{P}_n N_{10}^*(ds,t)\mathbb{P}_n N_{01}^*(s,dt)\} \times \{\int\int w(s,t)EN_{10}^*(ds,t)EN_{01}^*(s,dt)\}},$$

where

$$
\begin{aligned}
A &= \int\int w(s,t)EN_{10}^*(ds,t)EN_{01}^*(s,dt)\int\int \hat{w}(s,t)\mathbb{P}_nN_{11}^*(ds,dt)\mathbb{P}_nH^*(s,t) \\
&- \int\int w(s,t)EN_{11}^*(ds,dt)EH^*(s,t)\int\int \hat{w}(s,t)\mathbb{P}_nN_{10}^*(ds,t)\mathbb{P}_nN_{01}^*(s,dt) \\
&= \int\int w(s,t)EN_{10}^*(ds,t)EN_{01}^*(s,dt)\Big\{\int\int \hat{w}(s,t)\mathbb{P}_nN_{11}^*(ds,dt)\mathbb{P}_nH^*(s,t) \\
&\quad -\int\int w(s,t)EN_{11}^*(ds,dt)EH^*(s,t)\Big\} \\
&- \int\int w(s,t)EN_{11}^*(ds,dt)EH^*(s,t)\Big\{\int\int \hat{w}(s,t)\mathbb{P}_nN_{10}^*(ds,t)\mathbb{P}_nN_{01}^*(s,dt) \\
&\quad -\int\int w(s,t)EN_{10}^*(ds,t)EN_{01}^*(s,dt)\Big\}.
\end{aligned}
$$

Consistency of $\hat{\zeta}_{CP}$: Since $N_{11}^*(s,t)$ and $H^*(s,t)$ are summations of bivariate indicator functions with $EM < \infty$, we have $\sqrt{n}\begin{pmatrix} \mathbb{P}_nN_{11}^* - EN_{11}^* \\ \mathbb{P}_nH^* - EH^* \end{pmatrix} \rightsquigarrow \begin{pmatrix} Z_{N_{11}^*} \\ Z_{H^*} \end{pmatrix}$, where $Z_{N_{11}^*}$ and Z_{H^*} are mean zero Gaussian processes, and $\rightsquigarrow$ denotes convergence in distribution. By the functional δ-method,

$$
\begin{aligned}
&\sqrt{n}\Big\{\int\int \mathbb{P}_nN_{11}^*(ds,dt)\mathbb{P}_nH^*(s,t) - \int\int EN_{11}^*(ds,dt)EH^*(s,t)\Big\} \\
&\xrightarrow{P} \int\int Z_{N_{11}^*}(ds,dt)EH^*(s,t) + \int\int EN_{11}^*(ds,dt)Z_{H^*}(s,t),
\end{aligned}
$$

where $\xrightarrow{P}$ denotes convergence in probability. Then,

$$
\begin{aligned}
&\left|\int\int \hat{w}\mathbb{P}_nN_{11}^*(ds,dt)\mathbb{P}_nH^*(s,t) - \int\int wEN_{11}^*(ds,dt)EH^*(s,t)\right| \\
\leq &\left|\int\int \mathbb{P}_nN_{11}^*(ds,dt)\mathbb{P}_nH^*(s,t) - \int\int EN_{11}^*(ds,dt)EH^*(s,t)\right| \\
&\quad +\|\hat{w} - w\|_\infty \cdot |\tau_{1,q+1} - \tau_{1,q}| \cdot |\tau_{2,r+1} - \tau_{2,r}| \xrightarrow{P} 0,
\end{aligned}
$$

as $n \to \infty$, where $\|\cdot\|_\infty$ denotes supremum over region Ω_{qr}. Similarly, we get

$$
\left|\int\int \hat{w}\mathbb{P}_nN_{10}^*(ds,t)\mathbb{P}_nN_{01}^*(s,dt) - \int\int wEN_{10}^*(ds,t)EN_{01}^*(s,dt)\right| \xrightarrow{P} 0,
$$

as $n \to \infty$. The consistency of $\hat{\zeta}_{CP}^{qr}$ is immediate.

Asymptotic normality of $\sqrt{n}(\hat{\zeta}_{CP} - \zeta_{CP})$: By some simple algebra,

$$
\begin{aligned}
\sqrt{n}A &= \int\int wEN^*_{10}(ds,t)EN^*_{01}(s,dt)\Big\{\int\int \sqrt{n}(\hat{w}-w)\mathbb{P}_nN^*_{11}(ds,dt)\mathbb{P}_nH^*(s,t) \\
&\quad + \int\int wEN^*_{11}(ds,dt)\sqrt{n}(\mathbb{P}_n - E)H^*(s,t) \\
&\quad + \int\int w\sqrt{n}(\mathbb{P}_n - E)N^*_{11}(ds,dt)\mathbb{P}_nH^*(s,t)\Big\} \\
&- \int\int wEN^*_{11}(ds,dt)EH^*(s,t)\Big\{\int\int \sqrt{n}(\hat{w}-w)\mathbb{P}_nN^*_{10}(ds,t)\mathbb{P}_nN^*_{01}(s,dt) \\
&\quad + \int\int wEN^*_{10}(ds,t)\sqrt{n}(\mathbb{P}_n - E)N^*_{01}(s,dt) \\
&\quad + \int\int w\sqrt{n}(\mathbb{P}_n - E)N^*_{10}(ds,t)\mathbb{P}_nN^*_{01}(s,dt)\Big\}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\int\int wEN^*_{10}(ds,t)EN^*_{01}(s,dt)\int\int \sqrt{n}(\hat{w}-w)\mathbb{P}_nN^*_{11}(ds,dt)\mathbb{P}_nH^*(s,t) \\
-&\int\int wEN^*_{11}(ds,dt)EH^*(s,t)\int\int \sqrt{n}(\hat{w}-w)\mathbb{P}_nN^*_{10}(ds,t)\mathbb{P}_nN^*_{01}(s,dt) \\
=&\int\int wEN^*_{10}(ds,t)EN^*_{01}(s,dt)\Big\{\int\int \sqrt{n}(\hat{w}-w)EN^*_{11}(ds,dt)EH^*(s,t) + o_P(1)\Big\} \\
-&\int\int wEN^*_{11}(ds,dt)EH^*(s,t)\Big\{\int\int \sqrt{n}(\hat{w}-w)EN^*_{10}(ds,t)EN^*_{01}(s,dt) + o_P(1)\Big\} \\
=&\ o_P(1),
\end{aligned}
$$

since we assume constant ζ in the integration region. It follows that under time-invariance, $\sqrt{n}(\hat{\zeta}_{CP} - \zeta_{CP}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n I_i + o_P(1)$, where the influence function

$$
\begin{aligned}
I_i &= \frac{1}{\int\int wEN^*_{10}(ds,t)EN^*_{01}(s,dt)}\Big[\int\int wEN^*_{11}(ds,dt)\{\sum_{j\le m_i} I(Y_{ij}\ge s, Y_{i0}\ge t) \\
&- E(m_i)P(Y_{i1}\ge s, Y_{i0}\ge t)\} + \int\int w\{\sum_{j\le m_i} I(Y_{ij}=s,\eta_{ij}=1,Y_{i0}=t,\eta_{i0}=1) \\
&- E(m_i)P(Y_{i1}=s,\eta_{i1}=1,Y_{i0}=t,\eta_{i0}=1)\}EH^*(s,t)\Big] \\
&- \frac{\int\int wEN^*_{11}(ds,dt)EH^*(s,t)}{\{\int\int wEN^*_{10}(ds,t)EN^*_{01}(s,dt)\}^2}\Big[\int\int wEN^*_{10}(ds,t)\{\sum_{j\le m_i} I(Y_{ij}\ge s, Y_{i0}=t,\eta_{i0}=1) \\
&- E(m_i)P(Y_{i1}\ge s, Y_{i0}=t,\eta_{i0}=1)\} + \int\int w\{\sum_{j\le m_i} I(Y_{ij}=s,\eta_{ij}=1,Y_{i0}\ge t) \\
&- E(m_i)P(Y_{i1}=s,\eta_{i1}=1,Y_{i0}\ge t)\}EN^*_{01}(s,dt)\Big].
\end{aligned}
$$

By the asymptotic linearity, with $EM^2 < \infty$, we have the normality of $\sqrt{n}(\hat{\zeta}_{CP} - \zeta_{CP})$ via a standard central limit theorem.

A.2 ASYMPTOTIC PROPERTIES OF $\tilde{\theta}_{CP}^{L}$

Without confusion we simply write $\theta_{CP}(\Omega_{qr})$ for θ. Taking the first derivative of L_n in θ and multiplying by $\frac{\theta}{\binom{n}{2}}$, we have the following estimating equation

$$\begin{aligned}
\Psi_n(\theta) &= \frac{1}{\binom{n}{2}}\Bigg[\sum_{1\le i\le n}\sum_{d=1}^{m_{max}} I\{i \in D_d\}\frac{\phi_{i,d0}^{qr}\{a(Y_{id}, Y_{i0}; k, l) - 1\}}{a(Y_{id}, Y_{i0}; k, l) - 1 + \theta} \\
&\quad - \sum_{1\le i<j\le n}\sum_{d=1}^{m_{max}} I\{i, j \in D_d\}\frac{\psi_{ij,dd00}^{qr}\theta}{a(Y_{(idjd)}, Y_{(i0j0)}; k, l) - 1 + \theta}\Bigg] \\
&= \frac{1}{\binom{n}{2}}\Bigg[\sum_{1\le i<j\le n}\sum_{d=1}^{m_{max}} I\{i, j \in D_d\}\frac{\phi_{ij,dd00}^{qr}}{a(Y_{(idjd)}, Y_{(i0j0)}; k, l) - 1 + \theta} \\
&\quad - \sum_{1\le i<j\le n}\sum_{d=1}^{m_{max}} I\{i, j \in D_d\}\frac{\psi_{ij,dd00}^{qr}\theta}{a(Y_{(idjd)}, Y_{(i0j0)}; k, l) - 1 + \theta}\Bigg] = 0,
\end{aligned}$$

where $\phi_{ij,dd00}^{qr}$ is the concordant indicator for U statistics.

Define $X_i = (Y_{i0}, Y_{i1}, \dots, Y_{im_i}, \eta_{i0}, \eta_{i1}, \dots, \eta_{im_i}, m_i), i = 1, \dots, n$, and denote

$$h(X_i, X_j; \theta) = \sum_{d=1}^{m_{max}} I\{i, j \in D_d\}\left[\frac{\phi_{ij,dd00}^{qr} - \psi_{ij,dd00}^{qr}\theta}{a(Y_{(idjd)}, Y_{(i0j0)}; k, l) - 1 + \theta}\right]. \tag{A.1}$$

Then $\Psi_n(\theta)$ is a U statistic of order 2 with the kernel function $h(X_i, X_j; \theta)$.

Let $\Psi(\theta) = E\Psi_n(\theta)$ and θ_0 be the solution to $\Psi(\theta) = 0$. Define $\hat{\Psi}(\theta) = \frac{2}{n}\sum_{i=1}^{n} h_1(X_i; \theta)$, where $h_1(X_i; \theta) = Eh(X_i, X_j; \theta|X_i) - \Psi(\theta)$. Since $\left|\frac{\phi_{ij,dd00}^{qr} - \psi_{ij,dd00}^{qr}\theta}{a(Y_{(idjd)}, Y_{(i0j0)}; k,l) - 1 + \theta}\right| \le \frac{1+\theta}{\theta}$, for any $\theta > 0$, and m_{max} is finite for a family study, we have $Eh^2(X_i, X_j; \theta) < \infty$. By Theorem 12.3 of van der Vaart [1998], $\sqrt{n}(\Psi_n(\theta) - \Psi(\theta) - \hat{\Psi}(\theta)) \xrightarrow{P} 0$ and $\sqrt{n}(\Psi_n(\theta) - \Psi(\theta))$ converges to a mean zero random variable Z for any $\theta > 0$. Note that the map $\theta \to \Psi_n(\theta)$ is nonincreasing. By Lemma 5.10 of van der Vaart [1998], we have $\tilde{\theta}_{CP}^{L} \to \theta_0$ in probability, as $n \to \infty$.

Next, we will establish the asymptotic normality of the estimator by using a master theorem of Z estimators. For any $\theta_1, \theta_2 > 0$, note that $|h(X_i, X_j; \theta_1) - h(X_i, X_j; \theta_2)| = \left|\sum_{d=1}^{m_{max}} I\{i, j \in D_d\} \frac{\{\phi+\psi(a-1)\}(\theta_2-\theta_1)}{(a-1+\theta_1)(a-1+\theta_2)}\right| \leq \frac{m_{max}}{a(Y_{(idjd)}, Y_{(i0j0)}; k,l)-1}|\theta_1 - \theta_2|$. Therefore,

$$
\begin{aligned}
& \sqrt{n}(\Psi_n - \Psi)(\tilde{\theta}^L_{CP}) - \sqrt{n}(\Psi_n - \Psi)(\theta_0) \\
= \ & \sqrt{n}\{\hat{\Psi}(\tilde{\theta}^L_{CP}) - \hat{\Psi}(\theta_0)\} + o_P(1) \\
= \ & \sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} \{h_1(X_i; \tilde{\theta}^L_{CP}) - h_1(X_i; \theta_0)\} + o_P(1) \\
\leq \ & \sqrt{n}|\tilde{\theta}^L_{CP} - \theta_0| \cdot E\Big\{\frac{m_{max}}{a(Y_{(idjd)}, Y_{(i0j0)}; k,l) - 1}\Big\} + o_P(1) \\
= \ & o_P(1 + \sqrt{n}|\tilde{\theta}^L_{CP} - \theta_0|),
\end{aligned}
$$

since $a(Y_{(idjd)}, Y_{(i0j0)}; k, l)$ is the at-risk set of order n. Let $\dot{\Psi}(\theta_0) = E\{\frac{\partial h_1(X_i;\theta)}{\partial\theta}|_{\theta=\theta_0}\}$. $\tilde{\theta}^L_{CP}$ is chosen such that $|\Psi_n(\tilde{\theta}^L_{CP})| = o_P(n^{-1/2})$. By Theorem 13.4 of Kosorok [2007], we have $\sqrt{n}(\tilde{\theta}^L_{CP} - \theta_0) \to -\dot{\Psi}^{-1}_{\theta_0} Z$ in distribution. The bootstrap validity also follows.

A.3 DERIVATION OF CONDITIONAL PROBABILITIES $P(\Delta|\mathcal{A})$

We now give a detailed derivation for the conditional probabilities of $P(\Delta = 1|\mathcal{A})$ for bivariate mother-child data. Assume that a cause-specific mother-child pair $(Y_0, \eta_0, Y_1, \eta_1)$ comes from bivariate competing risks data $\{(Y_{i0}, \eta_{i0}, Y_{i1}, \eta_{i1}); i = 1, 2, ...n\}$ and consider the set $\mathcal{T}$ of points (s,t) such that

(a) a(s, t; k, l) $= \sharp\{i \colon Y_{i1} \geq s, Y_{i0} \geq t\} \geq 1$

(b) for some i, $Y_{i1} = s, \eta_{i1} = k, Y_{i0} \geq t$

(c) for some j, $Y_{j0} = t, \eta_{j0} = l, Y_{j1} \geq s$

Thus, $\mathcal{T}$ is a subset of the Cartesian product of $\{Y_{i1}\}$ and $\{Y_{j0}\}$. Then $\Delta(s, t; k, l) = 1$ if, for some h, $s = Y_{h1}$ and $t = Y_{h0}$, and $\Delta(s, t; k, l) = 0$ otherwise. The probability of $\Delta(s, t; k, l)$

given the risk set $\mathcal{A}(s,t;k,l)$ can be calculated thus:

$$\begin{aligned} P\Big(\Delta(s,t;k,l)=1|\mathcal{A}(s,t;k,l)\Big) &= a(s,t;k,l)P(Y_0=t,\eta_0=l)P(Y_1=s,\eta_1=k|Y_0=t,\eta_0=l), \\ P\Big(\Delta(s,t;k,l)=0|\mathcal{A}(s,t;k,l)\Big) &= a(s,t;k,l)P(Y_0=t,\eta_0=l) \\ &\quad \cdot (a(s,t;k,l)-1)\,P(Y_1=s,\eta_1=k|Y_0\geq t). \end{aligned} \tag{A.2}$$

Then we can derive the probability distribution of $\Delta(s,t;k,l)$ by noting that

$$P\Big(\Delta(s,t;k,l)=1|\mathcal{A}(s,t;k,l)\Big)+P\Big(\Delta(s,t;k,l)=1|\mathcal{A}(s,t;k,l)\Big)=1$$

and the definition of cause-specific association

$$\theta_{CS}(s,t,k,l)=\frac{\lambda_{1,k}(s|Y_0=t,\eta_0=l)}{\lambda_{1,k}(Y_1|Y0>t)}.$$

Next, we give a detailed derivation for the conditional probabilities of $P(\Delta=1|\mathcal{A})$ for sibship data. We will use $\Delta,\mathcal{A},a,a_1,a_2$ for simplicity. For any valid pair (s,t), we define two observable random variables B_s and B_t, where $B_s=i$ if s is from the first child of some pair in dataset a,where $i=1,2$, and $B_t=j$ if t is from the second child of some pair in dataset j, where $j=1,2$. It is easy to prove that $P(B_s=i,B_t=j)=\frac{1}{4},i,j=1,2$ by the exchangeability assumption. We also know $B_s=B_t$ if $\Delta=1$, then we have

$$\begin{aligned} P(\Delta=1|\mathcal{A}) &= \sum_{B_s,B_t} P(\Delta=1,B_s,B_t|\mathcal{A}) \\ &= \sum_{B_s,B_t} P(\Delta=1|B_s,B_t,\mathcal{A})\cdot P(B_s,B_t) \\ &= \frac{1}{4}P(\Delta=1|B_s=1,B_t=1,\mathcal{A})+\frac{1}{4}P(\Delta=1|B_s=1,B_t=2,\mathcal{A}) \\ &\quad +\frac{1}{4}P(\Delta=1|B_s=2,B_t=1,\mathcal{A})+\frac{1}{4}P(\Delta=1|B_s=2,B_t=2,\mathcal{A}) \\ &= \frac{1}{4}a_1\cdot P(Y_1=s,\eta_1=k)P(Y_2=t,\eta_2=l|Y_1=s,\eta_1=k) \\ &\quad +\frac{1}{4}a_2\cdot P(Y_2=s,\eta_2=k)P(Y_1=t,\eta_1=l|Y_2=s,\eta_2=k) \\ &= \frac{1}{4}a\cdot P(Y_1=s,\eta_1=k)P(Y_2=t,\eta_2=l|Y_1=s,\eta_1=k), \end{aligned} \tag{A.3}$$

where the exchangeability assumption is applied in the last equality. When $\Delta = 0$, s and t must from two different pairs, we have

$$\begin{aligned}
P(\Delta = 0|\mathcal{A}) &= \sum_{B_s, B_t} P(\Delta = 0, B_s, B_t|\mathcal{A}) \\
&= \sum_{B_s, B_t} P(\Delta = 0|B_s, B_t, \mathcal{A}) \cdot P(B_s, B_t) \\
&= \frac{1}{4}P(\Delta = 0|B_s = 1, B_t = 1, \mathcal{A}) + \frac{1}{2}P(\Delta = 0|B_s = 1, B_t = 2, \mathcal{A}) \\
&\quad + \frac{1}{4}P(\Delta = 0|B_s = 2, B_t = 1, \mathcal{A}) + \frac{1}{2}P(\Delta = 0|B_s = 2, B_t = 2, \mathcal{A}) \\
&= \frac{1}{4} \cdot a_1 P(Y_1 = s, \eta_1 = k) \cdot (a_1 - 1)P(Y_2 = t, \eta_2 = l|Y_1 \geq s) \\
&\quad + \frac{1}{4} \cdot a_1 P(Y_1 = s, \eta_1 = k) \cdot a_2 P(Y_1 = t, \eta_2 = l|Y_2 \geq s) \\
&\quad + \frac{1}{4} \cdot a_2 P(Y_2 = s, \eta_2 = k) \cdot a_1 P(Y_2 = t, \eta_2 = l|Y_1 \geq s) \\
&\quad + \frac{1}{4} \cdot a_2 P(Y_2 = s, \eta_2 = k) \cdot (a_2 - 1)P(Y_1 = t, \eta_1 = l|Y_2 \geq s) \\
&= \frac{1}{4}a \cdot P(Y_1 = s, \eta_1 = k) \cdot (a - 1) \cdot P(Y_2 = t, \eta_2 = l|Y_1 \geq s), \qquad \text{(A.4)}
\end{aligned}$$

where exchangeability assumption is applied again in the last equality. Then the same conditional distribution of $\Delta(s, t; k, l)$ as in the child-mother case can be shown by using ($A.3$) and ($A.4$), except that the calculation of $a(s, t; k, l)$ is not only based on the original dataset 1 but also on the switched dataset 2.

A.4 FORMULA FOR CORRECTED U

We now define concordant and discordant pairs for tied sibship observations as follows:

$$\begin{aligned}
\tilde{\phi}^{qr}_{ij,acbd} &= I\{Y_{ia} = Y_{jc}, Y_{ib} > Y_{jd}, (Y_{ia}, Y_{jd}) \in \Omega_{qr}, \eta_{ia} \neq 1, \eta_{jc} = 1, \eta_{jd} = 1\} \\
&+ I\{Y_{ia} = Y_{jc}, Y_{ib} < Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \eta_{ia} = 1, \eta_{jc} \neq 1, \eta_{ib} = 1\} \\
&+ I\{Y_{ia} > Y_{jc}, Y_{ib} = Y_{jd}, (Y_{jc}, Y_{ib}) \in \Omega_{qr}, \eta_{jc} = 1, \eta_{ib} \neq 1, \eta_{jd} = 1\} \\
&+ I\{Y_{ia} < Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \eta_{ia} = 1, \eta_{ib} = 1, \eta_{jd} \neq 1\} \\
&+ 0.5 \times \Big\{ I\{Y_{ia} \neq Y_{jc}, Y_{ib} = Y_{jd}, (Y_{(iajc)}, Y_{ib}) \in \Omega_{qr}, \eta_{(iajc)} = 1, \eta_{ib} = \eta_{jd} = 1\} \\
&\quad I\{Y_{ia} = Y_{jc}, Y_{ib} \neq Y_{jd}, (Y_{ia}, Y_{(ibjd)}) \in \Omega_{qr}, \eta_{ia} = \eta_{jc} = 1, \eta_{(ibjd)} = 1\} \\
&\quad I\{Y_{ia} = Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \\
&\qquad \eta_{ia} = \eta_{jc} = 1, \{\eta_{ib} = 1, \eta_{jd} \neq 1\} \text{ or } \{\eta_{ib} \neq 1, \eta_{jd} = 1\}\} \\
&\quad I\{Y_{ia} = Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \\
&\qquad \{\eta_{ia} = 1, \eta_{jc} \neq 1\} \text{ or } \{\eta_{ia} \neq 1, \eta_{jc} = 1\}, \eta_{ib} = \eta_{jd} = 1\} \Big\} \\
&+ 0.5 \times I\{Y_{ia} = Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \eta_{ia} = \eta_{jc} = \eta_{ib} = \eta_{jd} = 1\},
\end{aligned}$$

$$\begin{aligned}
\tilde{\psi}^{qr}_{ij,acbd} &= I\{Y_{ia} = Y_{jc}, Y_{ib} > Y_{jd}, (Y_{ia}, Y_{jd}) \in \Omega_{qr}, \eta_{ia} = 1, \eta_{jc} \neq 1, \eta_{jd} = 1\} \\
&+ I\{Y_{ia} = Y_{jc}, Y_{ib} < Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \eta_{ia} \neq 1, \eta_{jc} = 1, \eta_{ib} = 1\} \\
&+ I\{Y_{ia} > Y_{jc}, Y_{ib} = Y_{jd}, (Y_{jc}, Y_{ib}) \in \Omega_{qr}, \eta_{jc} = 1, \eta_{ib} = 1, \eta_{jd} \neq 1\} \\
&+ I\{Y_{ia} < Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \eta_{ia} = 1, \eta_{ib} \neq 1, \eta_{jd} = 1\} \\
&+ 0.5 \times \Big\{ I\{Y_{ia} \neq Y_{jc}, Y_{ib} = Y_{jd}, (Y_{(iajc)}, Y_{ib}) \in \Omega_{qr}, \eta_{(iajc)} = 1, \eta_{ib} = \eta_{jd} = 1\} \\
&\quad I\{Y_{ia} = Y_{jc}, Y_{ib} \neq Y_{jd}, (Y_{ia}, Y_{(ibjd)}) \in \Omega_{qr}, \eta_{ia} = \eta_{jc} = 1, \eta_{(ibjd)} = 1\} \\
&\quad I\{Y_{ia} = Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \\
&\qquad \eta_{ia} = \eta_{jc} = 1, \{\eta_{ib} = 1, \eta_{jd} \neq 1\} \text{ or } \{\eta_{ib} \neq 1, \eta_{jd} = 1\}\} \\
&\quad I\{Y_{ia} = Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \\
&\qquad \{\eta_{ia} = 1, \eta_{jc} \neq 1\} \text{ or } \{\eta_{ia} \neq 1, \eta_{jc} = 1\}, \eta_{ib} = \eta_{jd} = 1\} \Big\} \\
&+ 0.5 \times I\{Y_{ia} = Y_{jc}, Y_{ib} = Y_{jd}, (Y_{ia}, Y_{ib}) \in \Omega_{qr}, \eta_{ia} = \eta_{jc} = \eta_{ib} = \eta_{jd} = 1\}.
\end{aligned}$$

Similarly we define concordant and discordant pairs for the reversed tied observations under exchangeability $\tilde{\phi}^{qr}_{ij,adbc}$ and $\tilde{\psi}^{qr}_{ij,adbc}$. Then we have the modified U statistic for the sibship association

$$\hat{\theta}^{MU}_{CC}(\Omega_{qr};k,l) = \frac{\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \sum_{1\leq a<b\leq m_i} \sum_{1\leq c<d\leq m_j} (\phi^{qr}_{ij,acbd} + \tilde{\phi}^{qr}_{ij,acbd} + \phi^{qr}_{ij,adbc} + \tilde{\phi}^{qr}_{ij,adbc})}{\binom{n}{2}^{-1} \sum_{1\leq i<j\leq n} \sum_{1\leq a<b\leq m_i} \sum_{1\leq c<d\leq m_j} (\psi^{qr}_{ij,acbd} + \tilde{\psi}^{qr}_{ij,acbd} + \psi^{qr}_{ij,adbc} + \tilde{\psi}^{qr}_{ij,adbc})}. \tag{A.5}$$

APPENDIX B

RECURRENT FAILURE TIMES WITH MARKS

B.1 CONSISTENCY AND ASYMPTOTIC NORMALITY

B.1.1 Consistency

We begin with the following assumptions:

A0 The conditional distribution of tail integrals $f(u,v|\delta) = \frac{\partial^2 \mathcal{C}_\delta(u,v)}{\partial u \partial v}$ or the truncated version $f(u,v|\delta) = \frac{1}{\mathcal{C}_\delta(\lambda_1,\lambda_2)} \frac{\partial^2 \mathcal{C}_\delta(u,v)}{\partial u \partial v}$ is distinct for different values of δ (identifiability).

A1 The density functions $f(u,v|\delta)$ indexed by δ have common support, so that the support $\mathcal{A} = \{(u,v) : f(u,v|\delta) > 0\}$ is independent of δ.

A2 The observations $(X_i^{\|}, Y_i^{\|})$,$i = 1, \ldots, n$ are iid with distribution function $C(\bar{F}_1(x), \bar{F}_2(y))$, where $\bar{F}_1(\cdot)$ and $\bar{F}_2(\cdot)$ are marginal survival functions, and $C_\delta(u,v)$,$0 \le u \le 1, 0 \le v \le 1$ is the survival copula of the joint jump sizes, which is determined by the Lévy copula $\mathcal{C}_\delta(u,v)$,$0 \le u \le 1, 0 \le v \le 1$.

A3 The parameter space Ω of δ contains an open set ω of which the true parameter value δ_0 is an interior point.

A4 For every $(u,v) \in \mathcal{A}$, the density $f(u,v|\delta)$ is differentiable with respect to δ, u, v and the first and second order of derivatives or partial derivatives in δ, u, v are all continuous and bounded by some functions of u and v, which have finite expectations with probability 1. The mixed third order derivatives are also continuous.

A5 The integral $\int f(u,v|\delta)dudv$ can be twice differentiated under the integral sign, which ensures $E[\partial \log f(u,v|\delta)/\partial\delta] = 0$ and $E[-\partial^2 \log f(u,v|\delta)/\partial\delta] = E[\partial \log f(u,v|\delta)/\partial\delta]^2$.

Proof. of Lemma 3.3.4 Analogous to the proof of theorem 6.3.2 of Lehmann [1998], the inequality is equivalent to

$$\frac{1}{n}\sum_i \log f(\delta, \hat{\lambda}_1 \hat{F}_1(x_i^{\|}), \hat{\lambda}_2 \hat{F}_2(y_i^{\|})) < \frac{1}{n}\sum_i \log f(\delta_0, \hat{\lambda}_1 \hat{F}_1(x_i^{\|}), \hat{\lambda}_2 \hat{F}_2(y_i^{\|})).$$

Similar to the proof of Theorem 2 in Shih and Louis [1995], note that $\log f(u,v|\delta)$ is continuous in u and v, thus $|\log f(\delta, \hat{\lambda}_1 \hat{\bar{F}}_1(X), \hat{\lambda}_2 \hat{\bar{F}}_2(Y)) - \log f(\delta, \lambda_{10}\bar{F}_{10}(X), \lambda_{20}, \bar{F}_{20}(Y))|$ converges to 0 in probability. Thus $\frac{1}{n}\sum_i \log f(\delta, \hat{\lambda}_1, \hat{\bar{F}}_1(x_i), \hat{\lambda}_2, \hat{\bar{F}}_2(y_i))$ is asymptotically equivalent to $\frac{1}{n}\sum_i \log f(\delta, \lambda_{10}F_{10}(x_i), \lambda_{20}F_{20}(y_i)))$, which by law of large numbers converges to its expectation $E\Big[\log f(\delta, \lambda_{10}\bar{F}_{10}(x), \lambda_{20}\bar{F}_{20}(x))\Big]$.

Similarly, $\frac{1}{n}\sum_i \log f(\delta_0, \hat{\lambda}_1 \hat{\bar{F}}_1(x_i), \hat{\lambda}_2 \hat{\bar{F}}_2(y_i)) \xrightarrow{P} E\Big[\log f(\delta_0, \lambda_{10}\bar{F}_{10}(x), \lambda_{20}\bar{F}_{20}(x))\Big]$.

Next, use ϕ_0 as the true marginal parameters or distribution functions. Note that $-\log$ is strictly convex, by Jensen's inequality:

$$E_{\delta_0,\phi_0}\left[\log f_i(\delta,\phi_0)/f_i(\delta_0,\phi_0)\right] < \log E_{\delta_0,\phi_0}[f_i(\delta,\phi_0)/f_i(\delta_0,\phi_0)] = 0.$$

This completes the proof. □

B.1.2 Asymptotic Normality

In the following, S will be used to represent the survival function instead of $\bar{F}$.

Proof. of Theorem 3.3.6

Expand $l_\delta(\delta, \hat{\lambda}_1, \hat{S}_1, \hat{\lambda}_2, \hat{S}_2)$ around δ_0 and evaluate at $\hat{\delta}$ to get:

$$\sqrt{n}(\hat{\delta} - \delta_0) \approx \frac{l_\delta(\delta_0, \hat{\lambda}_1, \hat{S}_1, \hat{\lambda}_2, \hat{S}_2)/\sqrt{n}}{\left[-\sum_{j=1}^{n} l_{\delta\delta}\{\delta_0, \hat{\lambda}_1, \hat{S}_1(x_j), \hat{\lambda}_2, \hat{S}_2(y_j)\} \Big/ n\right]}.$$

In the denominator, because $\hat{\lambda}_i$ and $\hat{S}_i$ are consistent estimators and $l_{\delta\delta}(\cdot)$ is continuous, we know $l_{\delta\delta}(\delta_0, \hat{\lambda}_1, \hat{S}_1(x_j), \hat{\lambda}_2, \hat{S}_2(y_j))$ converges to $l_{\delta\delta}(\delta_0, \lambda_{10}, S_{10}(x_j), \lambda_{20}, S_{20}(y_j))$, so

$-\sum_{j=1}^{n} l^j_{\delta\delta}(\delta_0, \hat\lambda_1, \hat S_1, \hat\lambda_2, \hat S_2)$ is asymptotically equivalent to $-\sum_{j=1}^{n} l^j_{\delta\delta}(\delta_0, \lambda_{10}, S_{10}, \lambda_{20}, S_{20})$, which converges to I_δ. The superscript j indicates a function of the j-th pair of the joint jumps.

Finally we consider the numerator $l_\delta(\delta_0, \hat\lambda_1, \hat S_1, \hat\lambda_2, \hat S_2)/\sqrt{n}$. Assume that the empirical joint distribution of jump sizes is $H_n(x, y)$ while the true distribution is H_{δ_0}; then we decompose it as:

$$
\begin{aligned}
n^{-1/2} l_\delta(\delta_0, \hat\lambda_1, \hat S_1, \hat\lambda_2, \hat S_2) &= \sqrt{n} \int_{R^2_+} W_\delta(\delta_0, \hat\lambda_1, \hat S_1(x), \hat\lambda_2, \hat S_2(y)) dH_n(x, y) \\
&= \sqrt{n} \int_{R^2_+} W_\delta(\delta_0, \hat\lambda_1, \hat S_1(x), \hat\lambda_2, \hat S_2(y)) dH_{\delta_0}(x, y) \\
&+ \sqrt{n} \int_{R^2_+} W_\delta(\delta_0, \hat\lambda_1, \hat S_1(x), \hat\lambda_2, \hat S_2(y))(dH_n - dH_{\delta_0})(x, y) \\
&= R_n(\delta_0, \hat\lambda_1, \hat S_1, \hat\lambda_2, \hat S_2) + Z_n(\delta_0, \hat\lambda_1, \hat S_1, \hat\lambda_2, \hat S_2),
\end{aligned}
$$

where $W_\delta(x, y) = \partial \log f(\delta, \phi)/\partial\delta$. We further decompose Z_n thus:

$$
\begin{aligned}
Z_n(\delta_0, \hat\lambda_1, \hat S_1, \hat\lambda_2, \hat S_2) &= \sqrt{n} \int_{R^2_+} W_\delta(\delta_0, \lambda_{10}, S_{10}(x), \lambda_{20}, S_{20}(y)) d(H_n - H_{\delta_0})(x, y) \\
&+ \sqrt{n} \int_{R^2_+} \Big[W_\delta(\delta_0, \hat\lambda_1, \hat S_1(x), \hat\lambda_2, \hat S_2(y)) \\
&- W_\delta(\delta_0, \lambda_{10}, S_{10}(x), \lambda_{20}, S_{20}(y)) \Big] d(H_n - H_{\delta_0})(x, y).
\end{aligned}
$$

The second term converges to 0 because all the marginal estimates are consistent and $\sqrt{n}(H_n - H_{\delta_0}) \to O_p(1)$. By **A4**, W_δ is continuous and bounded; thus by dominated convergence theorem, we can interchange the limit and integral, which makes the second term 0. The first term is just a sum of i.i.d. random variables of mean 0 and variance $I_{\delta\delta}$. By the central limit theorem, it converges to a normal distribution with mean 0 and variance $I_{\delta\delta}$.

To derive the asymptotic properties of R_n, we need apply the von Mises expansion to R_n around $\lambda_{10}, S_{10}(x), \lambda_{20}, S_{20}(y)$:

$$
\begin{aligned}
R_n(\delta_0, \hat{\lambda}_1, \hat{S}_1, \hat{\lambda}_2, \hat{S}_2) \simeq\ & R_n(\delta_0, \lambda_{10}, S_{10}, \lambda_{20}, S_{20}) \\
& + \sqrt{n} \int_{R_+^2} \frac{\partial}{\partial \lambda_1} W_\delta \cdot dH_{\delta_0}(x,y)(\hat{\lambda}_1 - \lambda_{10}) \\
& + \sqrt{n} \int_{R_+^2} \frac{\partial}{\partial \lambda_2} W_\delta dH_{\delta_0}(x,y) \cdot (\hat{\lambda}_2 - \lambda_{20}) \\
& + \sqrt{n} \int_{R_+} I_{c_1}(x) d(\hat{S}_1 - S_{10})(x) + \sqrt{n} \int_{R_+} I_{c_2}(y) d(\hat{S}_2 - S_{20})(y),
\end{aligned}
$$

where I_{c_1} is found by differentiating $R_n(\delta_0, \hat{\lambda}_1, (1-\epsilon_1)S_{10} + \epsilon_1 \hat{S}_1, \hat{\lambda}_2, (1-\epsilon_2)S_{20} + \epsilon_2 \hat{S}_2)$ with respect to ϵ_1 and ϵ_2 and evaluating at $\epsilon_1 = \epsilon_2 = 0$:

$$
\begin{aligned}
I_{c_1}(x) &= \int_0^x \int_{R_+} \frac{\partial}{\partial S_1} W_\delta(\delta_0, \lambda_{10}, S_{10}(u), \lambda_{20}, S_{20}(y)) h_{\delta_0}(u,y) dy du, \\
I_{c_2}(y) &= \int_0^y \int_{R_+} \frac{\partial}{\partial S_2} W_\delta(\delta_0, \lambda_{10}, S_{10}(x), \lambda_{20}, S_{20}(u)) h_{\delta_0}(x,u) dx du.
\end{aligned}
$$

Note that $R_n(\delta_0, \lambda_{10}, S_{10}, \lambda_{20}, S_{20}) = 0$, $\int_{R_+^2} \frac{\partial}{\partial \lambda_i} W_\delta \cdot dH_{\delta_0}(x,y) = I_{\delta \lambda_i}, i = 1,2$ and $\hat{\lambda}_i - \lambda_{i0} = [\hat{\lambda}_i - \hat{\lambda}^{\|} - (\lambda_{i0} - \lambda^{\|})] + [\hat{\lambda}^{\|} - \lambda^{\|}]$, where $\sqrt{n - n_i}[\hat{\lambda}_i - \hat{\lambda}^{\|} - (\lambda_{i0} - \lambda^{\|})]$ is asymptotically equivalent to $\sum_{j=1}^{n_i - n} I^{-1}_{\lambda_i - \lambda^{\|}} (\frac{1}{\lambda_i - \lambda^{\|}} - T^{\perp}_{ij}) / \sqrt{n_i - n}$ and $\sqrt{n}[\hat{\lambda}^{\|} - \lambda^{\|}]$ is asymptotically equivalent to $\sum_{j=1}^{n} I^{-1}_{\lambda^{\|}} (\frac{1}{\lambda^{\|}} - T^{\|}_j) / \sqrt{n}$ for i=1,2, where $T^{\perp}_{ij}$ is the j-th interarrival time of single jumps of component i and $T^{\|}_j$ is the j-th interarrival time of the joint jumps. Also note that $n_i/n = \lambda_{i0}/\lambda^{\|}$. Then we can asymptotically decompose R_n into three independent parts, corresponding to single jumps of component 1, single jumps of component 2 and joint jumps respectively, thus $R_n = \frac{1}{\sqrt{n_1 - n}} \sum_{j=1}^{n_1 - n} \Phi_{1j} + \frac{1}{\sqrt{n_2 - n}} \sum_{j=1}^{n_2 - n} \Phi_{2j} + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Phi^{\|}_j$

$$
\begin{aligned}
\Phi_{1j} &= \sqrt{\frac{\lambda^{\|}}{\lambda_1 - \lambda^{\|}}} I_{\delta \lambda_1} I^{-1}_{\lambda_1} \left(\frac{1}{\lambda_1 - \lambda^{\|}} - T^{\perp}_{1j} \right) - \frac{\sqrt{(\lambda_1 - \lambda^{\|})\lambda^{\|}}}{\lambda_1} \left(I_{c_1}(x^{\perp}_j) - E(I_{c_1}(x) \right), \\
\Phi_{2j} &= \sqrt{\frac{\lambda^{\|}}{\lambda_2 - \lambda^{\|}}} I_{\delta \lambda_2} I^{-1}_{\lambda_2} \left(\frac{1}{\lambda_2 - \lambda^{\|}} - T^{\perp}_{2j} \right) - \frac{\sqrt{(\lambda_2 - \lambda^{\|})\lambda^{\|}}}{\lambda_2} \left(I_{c_2}(y^{\perp}_j) - E(I_{c_2}(y) \right), \\
\Phi^{\|}_j &= (I_{\delta \lambda_1} + I_{\delta \lambda_2}) I^{-1}_{\lambda^{\|}} \left(\frac{1}{\lambda^{\|}} - T^{\|}_j \right) - \frac{\lambda^{\|}}{\lambda_1} \left(I_{c_1}(x^{\|}_j) - E(I_{c_1}(x)) \right) - \frac{\lambda^{\|}}{\lambda_2} \left(I_{c_2}(y^{\|}_j) - E(I_{c_2}(y)) \right).
\end{aligned}
$$

Thus, R_n is asymptotically normal with mean 0 and variance: $\tau = E[(\Phi_{1j})^2] + E[(\Phi_{2j})^2] + E[(\Phi_j^{\|})^2]$.

Then by the method in Appendix 1 of [Shih and Louis, 1995], it can be shown that Z_n and R_n are asymptotically uncorrelated, thus $\sqrt{n}(\hat{\delta} - \delta_0)$ is asymptotically normal with mean 0 and variance $\frac{I_{\delta\delta}+\tau}{I_{\delta\delta}^2}$ □

Variance Estimator Note that Fisher information can be obtained by its empirical counterpart. To estimate $I_{c_1}(x), I_{c_2}(y)$, we use the following formula:

$$I_{c_1}(x) = E\left[\frac{\partial}{\partial S_1} W_\delta(\delta_0, \lambda_{10}, S_{10}(u), \lambda_{20}, S_{20}(v)) \cdot \mathbf{1}(u < x)\right].$$

Then it is can by estimated by

$$\frac{1}{n}\sum_{j=1}^{n}\left[\frac{\partial}{\partial S_1} W_\delta(\hat{\delta}, \hat{\lambda}_1, \hat{S}_1(x_j), \hat{\lambda}_2, \hat{S}_2(y_j)) \cdot \mathbf{1}(x_j < x)\right].$$

$I_{c_2}(y)$ can be similarly approximated. Then the variance can be estimated by the observed variance of Φ_i.

B.2 OVERVIEW OF LÉVY PROCESS

B.2.1 Overview

It was the French probabilist Paul Lévy who first studied what is now known as Lévy processes in the 1930s. The properties were further investigated in the 1930s and 1940s by Paul Lévy himself, the Russian mathematician A.N.Khintchine and the Japanese mathematician K.Itô. Recently, the theoretical development and applications in finance, insurance and other fields have led to a revival of interest in Lévy processes. These novel applications of Lévy processes are due to their great flexibility to satisfy more and more demanding modeling needs.

A Lévy process is the continuous analogue of random walk, which is the simplest example of stochastic processes in discrete time. Random walk can be represented as sum of independently identically distributed random variable, while the number of jumps is finite up to a specific time. In contrast, Lévy process may have an infinite number of jumps within a finite time interval and the increments of the stochastic process are stationary and independent, which can be continuously approximated by the random walk.

Furthermore, the stationary distribution of the increment is required to be infinitely divisible, such that the increment of the Lévy process at different scales follows the same type of distribution. It can be shown that there is a one-to-one mapping between infinitely divisible distributions and Lévy processes. This property connects Lévy process with many commonly used distributions, which are infinitely divisible, e.g. normal, Poisson, student t, gamma, α stable, log normal or negative binomial distributions. However, the uniform and binomial distributions are not, because they are distributions of non-degenerate bounded random variables. The one-to-one mapping between them provides great versatility in practical modeling, e.g., substituting the normal errors in time series models with errors from a infinitely divisible distribution, or substituting the driving force Brownian motion in stochastic models with a Lévy process. Thus, Lévy processes are natural models of noise that can be used to build stochastic integrals and to drive stochastic differential equations, e.g. in Barndorff-Nielsen and Shephard [2001], an Ornstein-Uhlenbeck process driven by positive Lévy process was proposed to model the volatility of stock price.

There are many important examples of Lévy processes, such as Brownian motion, Poisson processes, stable processes and subordinator (increasing Lévy processes). Convolution of Brownian motion and any of the above pure jump processes (e.g. compound Poisson processes or any infinite jump processes) can describe a stochastic path, which consists of a continuous part interlaced with jump discontinuities of random size at random times. For example, CPP can be used to model the aggregate insurance losses, which uses jump times to describe the claim times and jump sizes for the size of claims. For finance data, Lévy processes can naturally accommodate the characteristics of stock returns, which evolve continuously but may jump randomly resulting from abrupt incoming information.

Lévy processes have relatively easy structures compared to other class of processes, such as semi-martingales, Feller processes and self-similar processes. Semi-martingale can be decomposed into the sum of a local martingale and a càdlàg adapted process with finite variance, while for a Lévy process, by Lévy Itô decomposition, the local martingale part is the sum of a Brownian motion and compensated small jumps, while the sum of drift term and large jumps has finite variance. Moreover, a self-similar process is just a direct generalization of α stable Lévy processes, whose key property is that it remains unchanged under time aggregation, see Mandelbrot [1963].

The Lévy Itô decomposition simplified the interpretation of Lévy process, by decomposing it into a continuous part (Brownian motion with drift), infinite small jumps and finite big jumps. In addition, Lévy's work using Fourier analysis can also help explain the structure of general Lévy processes; for example, there is the famous Lévy Khintchine formula, which states that the characteristic function of Lévy processes L_t are determined by the Lévy triplets (ν, A, γ), where A is the covariance matrix of a Brownian motion, γ is the drift term and ν is called the Lévy measure, which describes the mean number of jumps for a certain range of jump size within a unit time interval. The continuous part described by A and γ has already been well studied, while current interest focuses on the pure jump part described by ν. Thus, the Lévy Khintchine formula provides an easy way to characterize the Lévy processes, by studying the specific form of the Lévy measure ν.

B.2.2 Lévy Process

Definition A càdlàg stochastic process $(X_t)_{t\geq 0}$ on $\{\Omega, \mathcal{F}, \mathbb{P}\}$ with values in $\mathbb{R}^d$ such that $X_0 = 0$ is called a Lévy Process if it possesses the following properties:

1 . Independent increments

2 . Stationary increments

3 . Stochastic continuity: $\forall \varepsilon > 0, lim_{h\to 0}\mathbb{P}(|X_{t+h} - X_t| \leq \varepsilon)$

The definition implies that a Lévy process is infinitely divisible, means that X_t can be represented as sum of arbitrary number of i.i.d. increments. Many distributions are infinitely divisible, as mentioned in the overview section, which leads to very flexible models based on Lévy processes.

Infinite divisibility also implies that the characteristic function of a Lévy process X_t is

$$\Phi_t(z) \equiv \Phi_{X_t}(z) \equiv E[e^{iz\cdot X_t}] = e^{t\phi(z)}, z \in \mathbb{R}^d,$$

which can be easily derived by representing X_t as the sum of i.i.d. increments within a unit time interval.

The *Lévy measure* is used to study the discontinuous property of Lévy process:

Definition Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$. The measure ν on $\mathbb{R}^d$ defined by:

$$\nu(A) = E[\sharp\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\}], \quad A \in \mathcal{B}\{\mathbb{R}^d\}$$

is called Lévy measure of X_t : $\nu(A)$ is the expected number, per unit time, of jumps whose size belongs to A.

Theorem B.2.1. Lévy-Itô decomposition: *Let$(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ and ν its Lévy measure.*

- *ν is a Random measure on $\mathbb{R}^d \backslash \{\mathbf{0}\}$ and satisfies*

$$\int_{|x|\leq 1} |x|^2 \nu(dx) < \infty \quad \int_{|x|\geq 1} \nu(dx) < \infty \tag{B.1}$$

- *The jump measure of X, denoted by J_X, is a Poisson random measure on $[0,\infty) \times \mathbb{R}^d$ with intensity measure $\nu(dx)dt$.*
- *There exits a vector γ and a d-dimensional Brownian motion $(B_t)_{t\geq 0}$ with covariance matrix A such that*

$$\begin{aligned}
X_t &= \gamma t + B_t + X_t^l + \lim_{\epsilon \downarrow 0} \tilde{X}_t^\epsilon \\
X_t^l &= \int_{|x|\geq 1, s\in[0,t]} x J_X(ds \times dx) \\
\tilde{X}_t^\epsilon &= \int_{\epsilon \leq |x| < 1, s\in[0,t]} x\{J_X(ds \times dx) - \nu(dx)ds\} \\
&\equiv \int_{\epsilon \leq |x| < 1, s\in[0,t]} x \tilde{J}_X(ds \times dx)
\end{aligned}$$

All the three terms are independent and the convergence in the last term is almost sure and uniform in t on $[0,T]$

Theorem B.2.2. Lévy-Khintchine representation*: Let X_t be a Lévy process on $\mathbb{R}^d$ with characteristic triplet* (ν, A, γ)*. Then*

$$E[e^{iz.X_t}] = e^{t\phi(z)}, z \in \mathbb{R}^d$$
$$\textit{with } \ \phi(z) = -\frac{1}{2}z \cdot Az + i\gamma \cdot z + \int_{\mathbb{R}^d} (e^{iz \cdot x} - 1 - iz \cdot x1_{|x| \leq 1})\nu(dx).$$

Note that a Lévy process is completely determined by the characteristic triplet, in which A is the covariance structure of the Brownian part, ν describes the jump behavior of the pure jump part, $\nu(A)$ is the expected number, per unit time, of jumps whose size belongs to A. γ depends on the choice of truncation function, which is $g(x) = 1_{|x| \leq 1}$ in the above case.

Whether or not to truncate small jumps depends on the behavior of jumps, which is totally determined by $\nu(x)$. Indeed, a Lévy process may explode as jump size approaches 0. This is why it is necessary to put constraints on $\nu(x)$ such that:

$$\int_{|x| \leq 1} |x|^2 \nu(dx) < \infty,$$

which is used to control the variance of small jump part in the Lévy Itô decomposition.

For details, see Cont and Tankov [2004] and Sato [1999].

B.2.3 Difficulty in Dependence Modeling by a Lévy Process

The dependence structure of the jump part of Lévy process is more difficult to model than that of other continuous processes. Thus, multivariate applications of stochastic processes have been dominated by Brownian motion, whose dependence is easy to quantify and the simulation is easy as well.

A simple method to solve this problem is Brownian subordination, which replaces the time of Brownian motion by an increasing Lévy processes, which is called subordinator, resulting in another Lévy process. The multivariate variance gamma, normal inverse Gaussian and generalized hyperbolic Lévy processes all fall into this category [Cont and Tankov, 2004]. Dependence structure is introduced by the covariance structure of the underlying Brownian motion. The advantages of this model are its simplicity and analytical tractability. But it can only handle limited range of dependence. In particular, independence of Brownian subordinator does not exist. Furthermore, all the marginal components must be of the same type and there is no quantitative measure of dependence.

Another dependence modeling method is to construct the dependence structure of jump sizes of CPP. This method argues that the jump risks of CPP come from several sources. For example, jumps of stocks may be induced by global risk, sector risk or idiosyncratic risk. This method is practical only when there are a few sources, but otherwise leads to an intractable model. Indeed, a lot of parameters need to be specified, such as intensity and jump size distribution of single jumps, intensity and jump size distribution of the joint jumps and dependence structure (e.g. copula function) of the joint jump sizes. Since the intensity of the marginal component is determined by the marginal intensity and dependence structure together, this method does not separate the dependence structure and margins, which is not a desirable property when dependence is of primary interest while margins are nuisance parameters. To separate them, copula for the whole processes is the third choice.

An ordinary copula method allows us to separate the dependence structure of a random vector from its univariate margins. It not only characterizes the dependence structure, but also provides a parametric representation of the multivariate distribution. The common copula method for random variables has been well investigated, but little effort has been made to study the dependence structure of dynamic processes. The Lévy copula is introduced in Tankov [2003] and Kallsen and Tankov [2006] to deal with the problems of modeling dependence structure of the Lévy process by ordinary copula, which will be discussed below. Lévy copula directly connects the 1-d Lévy measure by Lévy copula function and assures that the resulting high dimensional process is still a Lévy process.

B.3 COPULA

The ordinary copula is a popular method in constructing dependence between random variables [Nelsen, 1999, Joe, 1997]. The statistical law of a two-dimensional vector (X, Y) is usually described by its cumulative distribution function:

$$F(x, y) = P(X < x, Y < y) \tag{B.2}$$

with marginal laws:

$$F_1(x) = F(x, \infty), \quad F_2(y) = F(\infty, y). \tag{B.3}$$

The volume of a rectangle $B = [x_1, x_2] \times [y_1, y_2] \subset S_1 \times S_2$, where S_1 and S_2 are two possible infinite closed interval in $\bar{R} = \mathbb{R} \cup \{\infty\} \cup \{-\infty\}$, is defined based on $F(x, y)$:

$$V_F(B) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1). \tag{B.4}$$

Two other concepts are required before a copula can be defined:

- F is **2-increasing** if for every rectangle B in its domain, $V_F(B) > 0$,
- F is **grounded** if for every $x \in S_1$, $F(x, \min S_2) = 0$ and for every $y \in S_2$, $F(\min S_1, y) = 0$.

Definition A two-dimensional copula is a function C with domain $[0,1]^2$ such that

- C is grounded and 2-increasing
- C has margins $C_k, k = 1, 2$ which satisfy $C_k(u) = u$ for all u in $[0, 1]$

A two-dimensional copula is then a joint distribution function of two uniform random variables in $[0, 1]$. The famous Sklar' Theorem states the relationships between copula and joint distribution function:

Theorem B.3.1. *Let F be a two-dimensional distribution function with margins F_1, F_2. Then there exists a two-dimensional copula C such that for all $x \in \bar{R}^2$*

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{B.5}$$

If F_1, F_2 are continuous then C is unique, otherwise C is uniquely determined on $RanF_1 \times RanF_2$, where $RanF$ is range of F. Conversely, if C is a copula and F_1, F_2 are distribution functions, then the function F defined above is a two-dimensional distribution function with margins F_1, F_2.

It is natural to think about using an ordinary copula to connect the marginal distributions of a 2-d Lévy process (X_t, Y_t), leading to a time-varying copula process C_t, which, however, has the following drawbacks:

- The copula C_t may depend on t, therefore C_s for $s \neq t$ is generally not able to be calculated from C_t.
- It is unclear which copula will generate a two-d infinitely divisible law, for given one-d law of X_t and Y_t
- The law of components of multivariate Lévy process is generally given by its Lévy measure in its characteristic function. Thus probability density is hard to compute.

BIBLIOGRAPHY

E. W. Andersen. Composite likelihood and two-stage estimation in family studies. *Biometrika*, 5(1):15–30, 2004.

K. Bandeen-Roche and K. Liang. Modeling multivariate failure time associations in the presence of a competing risk. *Biometrika*, 89:299–314, 2002.

K. Bandeen-Roche and J. Ning. Nonparametric estimation of bivariate failure time associations in the presence of a competing risk. *Biometrika*, 95:221–232, 2008.

O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, 63(2):167–241, 2001.

I. V. Basawa and B. L. S. Prakasa Rao. *Statistical Inference for Stochastic Process*. Academic Press, London, 1980.

N. Bäuerle and R. Grübel. Multivariate counting processes: copulas and beyond. *ASTIN Bulletin*, 35(2):379–408, 2005.

B. Buckmann and G. Rudolf. Decompounding: An estimation problem for Poisson random sums. *The Annals of Statistics*, 31(4):1054–1074, 2003.

V. Chavez-Demoulin, P. Embrechts, and J. Nešlehvová. Quantitative models for operational risk: extremes, dependence and aggregation. Technical report, Dep. of Mathematics, ETH Zürich, 2005.

Y. Cheng and J. P. Fine. Nonparametric estimation of cause-specific cross hazard ratio with bivariate competing risks data. *Biometrika*, 95:233–240, 2008.

Y. Cheng, J. P. Fine, and M. R. Kosorok. Nonparametric analysis of bivariate competing risks data. *Journal of the American Statistical Association*, 102:1407–1416, 2007.

Y. Cheng, J. P. Fine, and M. R. Kosorok. Nonparametric association analysis of exchangeable clustered competing risks data. *Biometrics*, 65(2):385–393, 2009.

Y. Cheng, J. P. Fine, and K. Bandeen-Roche. Association analyses of clustered competing risks data via cross hazard ratio. *Biostatistics*, 11(1):82–92, Jan 2010.

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*, 65:141–152, 1978.

R. Cont and P. Tankov. *Financial modeling with jump processes*. CRC Press, New York, 2004.

H. Esmaeili and C. Klüppelberg. Parameter estimation of a bivariate compound Poisson process. *Insurance: Mathematics and Economics*, 47(2):224–233, 2010a.

H. Esmaeili and C. Klüppelberg. Parameter estimation of a bivariate stable Lévy process, 2010b.

H. Esmaeili and Claudia Klüppelberg. Two-step estimation of a multivariate Lévy process, 2010c.

J. P. Fine and H. Jiang. On association in a copula with time transformations. *Biometrika*, 87:559–571, 2000.

C. Genest, K. Ghoudi, and L.P. Rivest. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552, 1995.

P. Hougaard. *Analysis of Multivariate Survival Data*. Springer, New York, 2000.

T. Hu, B. Nan X. Lin, and J. M. Robins. Time-dependent cross ratio estimation for bivariate failure times. *Biometrika*, 98(2):341–354, 2011.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall Ltd, Boca Raton, FL, 1997.

H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419, 2005.

J. Kallsen and P. Tankov. Characterization of dependence of multidimensional Lévy processes using Lévy copula. *Journal of Multivariate Analysis*, 97:1551–1572, 2006.

M. R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer-Verlag, New York, 2007.

E. L. Lehmann. *Theory of Point Estimation 2nd*. Springer, 1998. ISBN 0-387-98052-6.

B. G. Lindsay. Composite likelihood method. *Contemporary Mathematics*, 80, 1988.

F. Lindskog and A. McNeil. Common Poisson shock models: Applications to insurance and credit risk modeling. *ASTIN Bulletin*, 33(2):209–238, 2003.

B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:392–419, 1963.

A. J. McNeil. Estimating the tails of loss severity distribution using extreme value theory. *ASTIN Bulletin*, 27:117–132, 1997.

R. B. Nelsen. *An Introduction to Copulas.* Springer-Verlag Inc, New York, 1999.

D. Oakes. A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B: Methodological*, 44:414–422, 1982.

D. Oakes. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84:487–493, 1989.

D. Pfeifer and J. Nešlehvová. Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin*, 34(2):333–360, 2004.

R. C. Pruitt. Identifiability of bivariate survival curves from censored data. *Journal of the American Statistical Association*, 88:573–579, 1993.

J. Rosinski. Tempering stable processes. *Stochastic processes and their applications*, 117(6): 677–707, 2007.

K. Sato. *Lévy processes and infinitely divisible distributions.* Cambridge University Press, Cambridge, 1999.

M. Schicks. Nonparametric estimation for multivariate compound Poisson processes, 2009.

J. H. Shih and P. S. Albert. Modeling familial association of ages at onset of disease in the presence of competing risk. *Biometrics*, 66:1012–1023, 2010.

J. H. Shih and T. A. Louis. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:1384–1399, 1995.

P. Tankov. Dependence structure of spectrally positive multi-dimensional Lévy processes. http://www.math.jussieu.fr/~tankov, 2003.

P. Tankov. Simulation and option pricing in Lévy copula. http://www.math.jussieu.fr/~tankov, 2006.

A. Tsiatis. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences*, 72:20–22, 1975.

A. W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, 1998.