# Conditional Copula-Graphic Estimator for Semi-Competing Risks Data

by

Shamsia Sobhan

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

Master of Science

Department of Statistics
University of Manitoba
Winnipeg

**Abstract**

In semi-competing risks data, interest lies in the estimation of the survival function of a non-terminal event time, which is subject to dependent censoring by a terminal event. This problem has been extensively studied in the literature, but mostly focusing on unconditional settings. In this thesis, we propose two versions of conditional copula-graphic estimator: one allows for covariate adjustment only in the marginal survival functions of non-terminal and terminal events, and the other allows for covariate adjustment in both the marginal survival functions and the dependence structure of non-terminal and terminal event times.

The proposed estimators are semiparametric. In both, the conditional copula is assumed to belong to a one-parameter Archimedean copula family, but the copula parameter is estimated parametrically in the first version and nonparametrically in the second one. Both versions employ Beran's estimator in the estimation of the conditional marginal survival functions. The performance of the conditional copula-graphic estimators is assessed using a simulation study and is compared to that of the unconditional copula-graphic estimator to investigate the cost of ignoring covariate effects.

Our findings suggest that, in the presence of covariates, the conditional copula-graphic estimators are more efficient and less biased than the unconditional

copula-graphic estimator. If interest centres on the estimation of the marginal survival function of the non-terminal event, both versions of the conditional copula-graphic estimator perform similarly. However, if the estimation of the conditional dependence structure is also of interest, the second version more accurately captures the underlying dependence structure. The performance of the conditional copula-graphic estimators deteriorates with the increase in the censoring rate of the non-terminal event. A real data example on breast cancer recurrence is provided to illustrate the proposed approach in comparison to the unconditional copula-graphic estimator.

# Acknowledgment Page

At first I express my gratitude to the merciful and almighty God for the blessing, invisible guidance and immeasurable grace for a successful completion of the program. It is a great pleasure in expressing endless gratefulness and sincere gratitude to my respected supervisor, Dr. Elif Fidan Acar. In spite of her busy schedule, I was fortunate enough to have her excellent supervision, continuous help and guidance, valuable suggestions and prudential advices throughout the period of my thesis. Her continuous encouragement and support boosted my confidence whenever I was stressed. Special thanks to my committee members Dr. Po Yang and Dr. Melody Ghahramani for their insightful comments. Thanks to my colleagues and friends for their support and encouragement. Special Thanks to my mother Hosne Ara Sobhan, my sisters and relatives for all the support and prayers.

# Dedication Page

This thesis is dedicated to my mother. Whatever I am today is the result of her hard work and sacrifices.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Survival analysis is a statistical technique where the outcome variable is the time until an event (e.g., death, disease recurrence) occurs. In survival analysis, complete data may not be observed as some subjects do not experience the event during the study period due to drop-out or death. In such cases, the event time is called *censored*. Depending on the type of events and their underlying censoring mechanisms, different approaches are taken in survival analysis (e.g., Lawless, 2011; Klein and Moeschberger, 2006; Crowder, 2012). This thesis focuses on *semi-competing risks*.

Semi-competing risks refer to a situation where a subject may experience a terminal event (e.g., death) before the occurrence of a non-terminal event (e.g., cancer recurrence) and where both events are subject to independent (administrative) censoring. Since the censoring of the non-terminal event by the terminal event is informative, the dependence between these two event

times needs to be taken into account when estimating the marginal survival function of the non-terminal event time.

The motivating example in this thesis comes from the German Breast Cancer Study (GBCS), where interest lies in the recurrence time of breast cancer. The original dataset consists of 720 patients with primary node positive breast cancer between July 1984 and December 1989 (Schmoor et al., 1996). The available subset consists of 686 observations with times to cancer recurrence and times to death, indicators of their censoring statuses and continuous covariates, such as age, tumour size, number of positive lymph nodes, progesterone receptor level and estrogen receptor level, and categorical covariates, such as menopausal status, tumour grade and hormone treatment. In the study, all patients had breast cancer history along with covariate information but were cancer-free prior to the study start. Recurrence times of breast cancer are likely to be associated with lifetimes as both events are subject to some common risk factors such as genetic background, immune system and patient's lifestyle. The recurrence time can be censored by either death or the end of study period (administrative censoring), and death may be censored as well by the end of study period. The administrative censoring is assumed to be independent of the two events. Since 171 patients were alive at the end of the study period, this dataset defines an example of semi-competing risks. If the terminal event were observed for all study subjects, the data structure would fall under dependent right-censoring.

In the case of dependent right-censored data the copula-graphic estimator (Zheng and Klein, 1995) is commonly used to estimate the marginal survival function of the event of interest. This estimator has a closed-form expression when the copula of the non-terminal and terminal event times is Archimedean (Rivest and Wells, 2001). For semi-competing risks data, where there is additional independent censoring, Lakhal et al. (2008) proposed a copula-graphic estimator using one parameter Archimedean copulas, which was extended to multi-parameter Archimedean copulas by Heuchenne et al. (2014) using a pseudo-maximum likelihood approach. The large sample properties of the copula-graphic estimator were studied in Laurent (2013) and (Rivest and Wells, 2001).

While semi-competing risks data have been extensively studied in the literature, only few works have addressed the incorporation of covariates in the modelling strategy. While most research efforts in this domain focused on regression analysis of marginal survival functions, some addressed potential covariate effects on the dependence structure for some special cases, such as discrete covariates (e.g., Hsieh et al., 2008; Chen, 2012) and time-varying dependence parameters (e.g., Peng and Fine, 2007; Hsieh and Huang, 2012). Peng and Fine (2007) proposed regression analysis based on a parametric copula model and developed nonparametric estimator using non-linear estimating equation. Hsieh and Huang (2012) specified the dependence structure of non-terminal and terminal events via a conditional copula models and the covariate effects

on margins via regression models. They proposed a conditional likelihood approach for estimating the covariate effects on the margins. In some other work, the dependence structure is specified using a frailty model (e.g., Ghosh, 2006; Xu et al., 2010). Ghosh (2006) proposed a gamma frailty model for analyzing semi-competing risks data conditional on covariates and showed in a data application of leukaemia that the strength of dependence between times to disease recurrence and death differs for different types of leukaemia. Xu et al. (2010) proposed an illness-death model with gamma frailty for covariate adjusted analysis of semi-competing risks data. They applied their method to Nasopharyngeal cancer data and found a significant effect of the tumour size and nodal status on the dependence structure of cancer relapse and death times. Some recent work also considered quantile regression (e.g., Li and Peng, 2015; Yang and Peng, 2016) for analyzing conditional semi-competing risks data.

In this thesis, we develop an extension of the copula-graphic estimator of Braekers and Veraverbeke (2005) for dependent right-censored data to the semi-competing risks setting. We propose two versions of conditional copula-graphic estimator that allows for covariate adjustment, (i) on the marginal survival functions and (ii) on both the margins and the dependence structure of non-terminal and terminal event times. The performances of the proposed estimators are evaluated in a simulation study under different rates of censoring and for different dependence structures. A comparison of conditional and unconditional copula-graphic estimators is also provided to investigate the cost

of ignoring covariate effects.

In the following sections, we provide an overview of copulas and survival analysis methods relevant to this thesis.

## 1.1 Background on Copulas

Copulas are joint distribution functions of random variables having standard uniform marginal distributions. Copula models have gained popularity in many applied fields such as finance (e.g., Uyttendaele and Mazo, 2016; Krupskii and Joe, 2015, 2013), actuarial science (e.g., Embrechts, 2009; Frees and Valdez, 1998) and health sciences (e.g., Geerdens et al., 2017; Li and Cheng, 2016; Heuchenne et al., 2014). They offer a strategy to study dependence relationships in multivariate data, define scale-free measures of dependence, and construct multivariate distributions.

Suppose $Y = (Y_1, Y_2)$ be a random vector with the joint distribution function $H$ and marginal distribution functions $F_i$ for $i = 1, 2$. Using Sklar's Theorem (Sklar, 1959), the joint distribution function can be written as

$$H(y_1, y_2) = \mathcal{C}(F_1(y_1), F_2(y_2)). \tag{1.1}$$

Here $\mathcal{C}$ is the copula of $Y_1$ and $Y_2$ linking their dependence.

A large number of parametric copula families has been proposed, with the most

popular families being elliptical copulas (e.g. Gaussian and Student t-copulas) and Archimedean copulas (e.g., Clayton, Gumbel and Frank copulas). In this research, Archimedean copulas are considered as they lead to a closed form expression for the copula-graphic estimator. In addition, Archimedean copulas describe very different dependence structures including lower and upper tail dependence.

Archimedean copulas are constructed explicitly via a generator function $\psi$ and take the form

$$C_\psi(u_1, u_2) = \psi^{[-1]}\{\psi(u_1) + \psi(u_2)\},$$

where $\psi : (0, 1] \to R^+$ is a continuous function satisfying $\psi(1) = 0$, $\psi'(t) < 0$ (strictly decreasing), $\psi''(t) > 0$ (convex), for all $t \in (0, 1]$ and $\psi^{[-1]}$ is the pseudo-inverse of $\psi$ such that $\psi^{-1} : (0, \infty] \to (0, 1]$ and

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

The Pearson's correlation coefficient is not an appropriate dependence measure for the Archimedean family of copulas. Even joint elliptical distributions, there are situations where linear correlation is not appropriate. If the model come from a heavy-tailed distribution linear correlation coefficient can not be defined because of infinite second moments (Embrechts et al., 2001). Kendall's tau is

commonly used to quantify the strength of dependence in copulas. Kendall's tau for the bivariate random vector $(y_1, y_2)^T$ is defined as

$$\tau(y_1, y_2) = \mathcal{P}\{(y_1 - \tilde{y_1})(y_2 - \tilde{y_2}) > 0\} - \mathcal{P}\{(y_1 - \tilde{y_1})(y_2 - \tilde{y_2}) < 0\},$$

where $(\tilde{y_1}, \tilde{y_2})^T$ is an independent copy of $(y_1, y_2)^T$. In other words Kendall's tau for $(y_1, y_2)^T$ is the probability of concordance minus the probability of discordance. For an Archimedean copula with parameter $\theta$, Kendall's tau can be written in terms of the generator function as

$$\tau = 4 \int_0^1 \frac{\psi_\theta(u)}{\psi'_\theta(u)} du + 1.$$

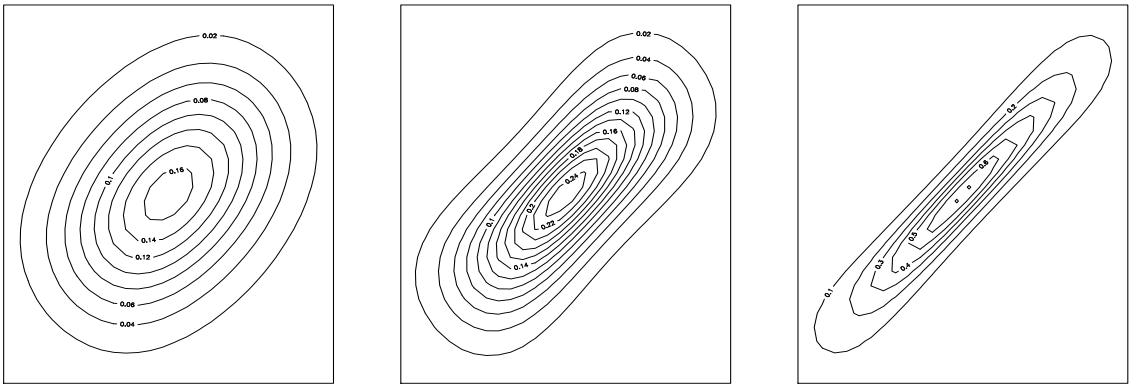
Below we describe the Archimedean copulas used in this thesis, along with their contour and the conversions between their parameter and Kendall's tau. We also provide the generator functions, as well as their inverses and derivatives, as these are need for likelihood construction in Chapter 2.

**Frank Copula:** The Frank copula is given by

$$C_\theta(u_1, u_2) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right\}, \qquad \theta \in \mathcal{R} \setminus \{0\}.$$

The Frank copula has no tail dependence. The contour plots of the Frank copula at different $\tau$ values are given in Figure (1.1) with standard normal margins.

Figure 1.1: Contour plots of the Frank copula when $\tau = 0.2$ (left panel), 0.5 (middle panel) and 0.8 (right panel).



The conversion of the parameter to Kendall's tau is given by

$$\tau = 1 + \frac{4}{\theta}[D_1(\theta) - 1],$$

where $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t)-1}dt$ is the Debye function.

The generator function of the Frank copula along with its first derivative, inverse and the first two derivatives of the inverse functions are given by

$$\psi_\theta(z) = -\ln\left\{\frac{e^{-\theta z}-1}{e^{-\theta}-1}\right\},$$

$$\psi'_\theta(z) = \frac{\theta e^{-\theta z}}{e^{-\theta z}-1},$$

$$\psi_\theta^{-1}(z) = -\frac{1}{\theta}\ln\{e^{-z}(e^{-\theta}-1)+1\},$$

$$\psi_\theta^{-1}(z)' = \frac{1}{\theta}\frac{e^{-z}(e^{-\theta}-1)}{e^{-z}(e^{-\theta}-1)+1},$$

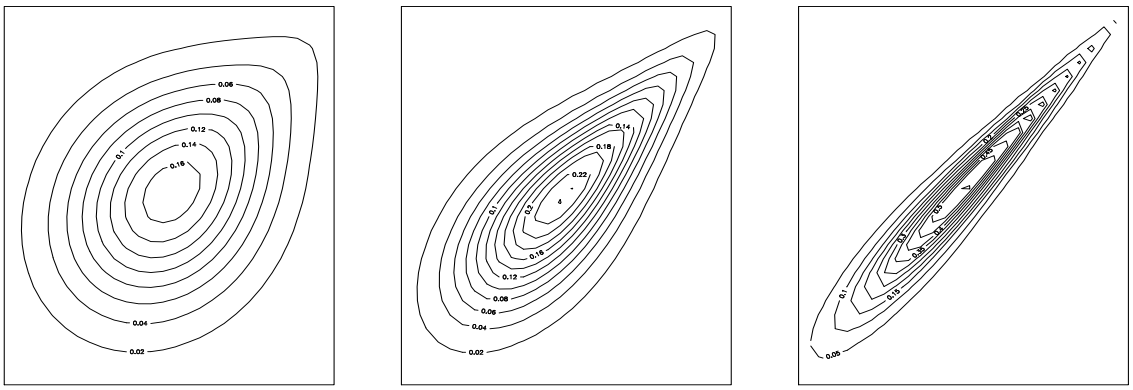
$$\psi_\theta^{-1}(z)'' = -\frac{1}{\theta}\frac{e^{-z}(e^{-\theta-1})}{\{e^{-z}(e^{-\theta}-1)+1\}^2}.$$

**Gumbel Copula:** The Gumbel copula is given by

$$C_\theta(u_1,u_2) = e^{-\{(-\ln u_1)^\theta+(-\ln u_2)^\theta\}^{\frac{1}{\theta}}}, \qquad \theta \in [1,\infty).$$

The Gumbel copula has upper tail dependence. The contour plots of Gumbel copula at different $\tau$ values are given in Figure (1.2) with standard normal margins.

Figure 1.2: Contour plots of the Gumbel copula with standard normal margins when $\tau = 0.2$ (left panel), 0.5 (middle panel) and 0.8 (right panel).



The conversion of the parameter to Kendall's tau is given by

$$\tau = 1 - \frac{1}{\theta}.$$

The generator function of the Gumbel copula along with its first derivative, inverse and the first two derivatives of the inverse functions are given by

$$\psi_\theta(z) = (-\ln z)^\theta,$$

$$\psi'_\theta(z) = -\frac{\theta}{z}(-\ln z)^{\theta-1},$$

$$\psi_\theta^{-1}(z) = e^{-z^{\frac{1}{\theta}}},$$

$$\psi_\theta^{-1}(z)' = -\frac{1}{\theta} z^{\frac{1}{\theta}-1} e^{-z^{\frac{1}{\theta}}},$$

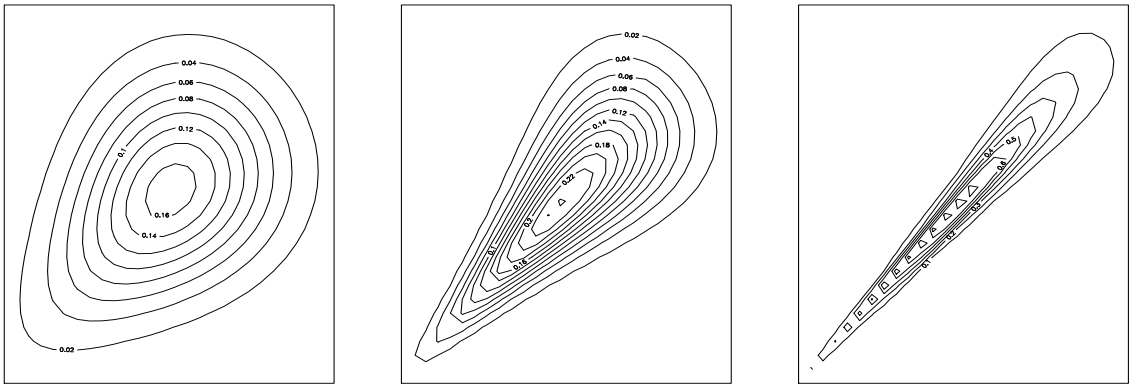
$$\psi_\theta^{-1}(z)'' = \frac{1}{\theta^2} z^{\frac{1}{\theta}-2} e^{-z^{\frac{1}{\theta}}} (z^{\frac{1}{\theta}} - 1 + \theta).$$

**Clayton Copula:** The Clayton copula is given by

$$C_\theta(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}, \qquad \theta \in (0, \infty).$$

The Clayton copula has lower tail dependence. The contour plots of Clayton copula at different $\tau$ values are given in Figure (1.3) with standard normal margins.

Figure 1.3: Contour plots of the Clayton copula with standard normal margins when $\tau = 0.2$ (left panel), 0.5 (middle panel) and 0.8 (right panel).



The conversion of the parameter to Kendall's tau is given by

$$\tau = \frac{\theta}{\theta + 2}.$$

The generator function of the Clayton copula along with its first derivative, inverse and the first two derivatives of the inverse functions are given by

$$\psi_\theta(z) = \frac{1}{\theta}(z^{-\theta} - 1),$$

$$\psi'_\theta(z) = -z^{-(\theta+1)},$$

$$\psi_\theta^{-1}(z) = (\theta z + 1)^{-\frac{1}{\theta}},$$

$$\psi_\theta^{-1}(z)' = -(\theta z + 1)^{-(\frac{1}{\theta}+1)},$$

$$\psi_\theta^{-1}(z)'' = (\theta + 1)(\theta z + 1)^{-(\frac{1}{\theta}+2)}.$$

## 1.2 Background on Survival Analysis

This section provides a brief overview of the survival analysis tools used in this thesis.

**Kaplan-Meier Estimator:** When the event time and censoring time are independent, the marginal survival function of the event time is often estimated using the Kaplan-Meier estimator. Let $Y$ be the event time, $Z$ be the censoring time. The observed random variables are $T = \min\{Y, Z\}$ and $\Delta = 1\{Y \leq Z\}$. Given that the observed data is $\{(T_i, \Delta_i), i = 1, \cdots, n\}$, the Kaplan-Meier estimator (Kaplan and Meier, 1958) is defined as

$$\hat{\Gamma}(t) = \prod_{i:T_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \tag{1.2}$$

where $n_i$ be the number of subjects at risk at $T_i$, $d_i$ be the number of events that happened at time $t_i$. Asymptotic properties of the Kaplan-Meier estimator are verified in many literature (e.g., Gonzalez-Manteiga and Cadarso-Suarez, 1994; Cai, 1998).

**Copula-Graphic Estimator for Dependent Censoring:** In case of dependent censoring, Zheng and Klein (1995) proposed a copula-graphic estimator where the dependence between the event time and censoring time is described by a known copula. This estimator is consistent and reduces to the Kaplan-Meier estimator when event and censoring times are independent (Rivest and Wells,

2001). Using previous notations when $Y \not\perp C$, the copula-graphic estimator of the marginal survival function is defined as

$$\hat{S}(t) = \psi_\theta^{-1}\left[-\sum_{T_i \le t, \Delta_i = 1} \psi_\theta(\hat{\pi}(T_i)) - \psi_\theta(\hat{\pi}(T_i) - 1/n)\right], \tag{1.3}$$

where $\hat{\pi}(t) = \sum_{i=1}^{n} 1\{T_i \ge t\}/n$, is the estimate of the survival function of $T = \min\{Y, Z\}$ with the censoring indicator $\Delta_1$.

**Copula-Graphic Estimator for Semi-competing Risks:** Consider a semi-competing risks setting with $Y_1, Y_2$, the non-terminal and terminal event times which are dependent on each other. Let $Z$ be the censoring time. The observed random variables are $T_1 = \min\{Y_1, Y_2, Z\}$, $T_2 = \min\{Y_2, Z\}$, $\Delta_1 = 1\{Y_1 \le Y_2, Y_1 \le Z\}$ and $\Delta_2 = 1\{Y_2 \le Z\}$. For analyzing such data, Lakhal et al. (2008) proposed an extension of copula-graphic estimator is given by

$$\widehat{S}_1(t) = \psi_\theta^{-1}\left(-\sum_{T_{1i} \le t, \Delta_i = 1} \left(\psi_\theta(\hat{\Gamma}(T_{1i}^-)) - \psi_\theta(\hat{\Gamma}(T_{1i}))\right)\right), \tag{1.4}$$

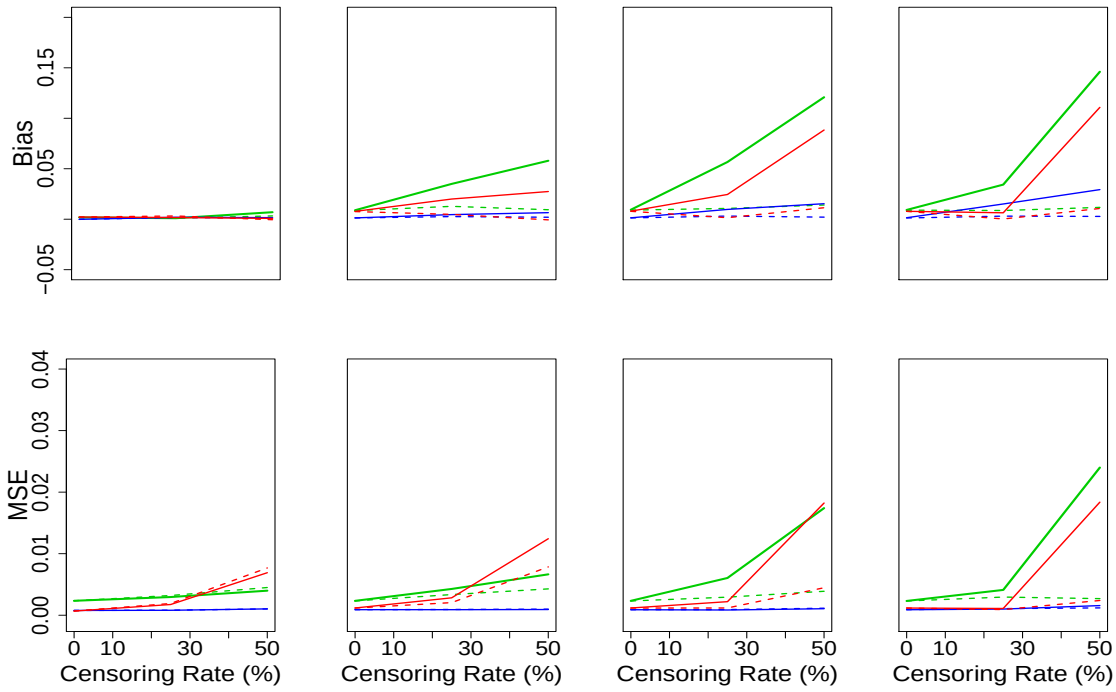
where $\hat{\Gamma}(.)$ is the survival function estimator of $T^* = \min\{Y_1, Y_2\}$, obtained using the Kaplan-Meier estimator with the censoring indicator $\Delta_3 = 1\{\min(Y_1, Y_2) < Z\} = \min(1, \Delta_1 + \Delta_2)$ and $T_1^-$ is the time just before $T_1$.

## 1.3 Dependent versus Independent Censoring: A Simulation Study

In this section, we perform a small-scale simulation study to compare the Kaplan-Meier estimator and the copula-graphic estimator at dependence levels $\tau = 0$, 0.2, 0.5 and 0.8, and censoring rates 0%, 25% and 50%. We generated samples of size, $n = 100$ under the Frank family (see details in Section 3.1) and obtained the Kaplan-Meier and copula graphic-estimators for M=100 Monte-Carlo replicates.

Figure (1.4) displays the bias and mean squared error of these estimators for $Q_{10}$, $Q_{50}$ and $Q_{90}$ across different censoring rates and at different dependence levels. When $\tau = 0$, the event time and censoring times are independent. Here the Kaplan-Meier (KM) estimator coincides with the copula-graphic estimator having equal bias and mean squared error (MSE) for all quantiles. As $\tau$ and/or censoring rate increase, the bias and the mean squared error of the Kaplan-Meier estimator deviate from those of the copula-graphic estimator for the $10^{\text{th}}$ and $50^{\text{th}}$ percentiles. Since bias $> 0$, the Kaplan-Meier estimator overestimates the survival probability for increasing $\tau$ and/or censoring rate. This simulation study indicates that ignoring the dependent censoring leads to biased and less efficient estimates.

Figure 1.4: Bias and Mean squared error of the Kaplan-Meier (solid line) and copula-graphic (dashed line) estimates, calculated over 100 Monte Carlo samples, at p = 0.1 (red), 0.5 (green), 0.9 (blue) under the Frank copula with $\tau = 0.0$, 0.2, 0.5 and 0.8 (left to right columns).




## 1.4 Thesis Outline

This thesis is organized as follows. Chapter 2 describes the model and the proposed estimation procedure for both the conditional marginal survival functions and the conditional copula. This chapter also outlines an iterative algorithm for conditional copula-graphic estimators. Chapter 3 presents the simulation results to evaluate the performance of the conditional copula-graphic

estimators in comparison to that of the unconditional version. Chapter 4 contains the analysis of German Breast Cancer data where we asses the effect of age on both the time to cancer relapse and the dependence between the relapse and death times. Chapter 5 summarizes our main findings and outlines future direction.

# Chapter 2

# Conditional Copula-Graphic Estimator

This chapter contains the methodological contributions of the thesis. After introducing the notation in Section 2.1, we describe the conditional copula model for the conditional joint survival function in Section 2.2. Two versions of the conditional-copula graphic estimator are introduced in Section 2.3 along with the estimation procedure. Finally, in Section 2.4, we outline an iterative algorithm for the estimation of the conditional copula parameter and of the conditional survival function.

## 2.1 Notation

Let us consider a semi-competing risks setting where $Y_1$ is the non-terminal event time, $Y_2$ is the terminal event time and $Z$ is the censoring time. The observed random variables are $T_1 = \min\{Y_1, Y_2, Z\}$, $T_2 = \min\{Y_2, Z\}$, $\Delta_1 = 1\{Y_1 \leq$

$Y_2, Y_1 \leq Z\}$ and $\Delta_2 = 1\{Y_2 \leq Z\}$. When $\Delta_1 = \Delta_2 = 1$, both $Y_1$ and $Y_2$ are observed, when $\Delta_1 = 1$, $Y_1$ and the minimum of $(Y_2, Z)$ are observed and when $\Delta_2 = 1$, $Y_2$ is observed. The situation where either non-terminal or terminal or both events occur is defined using the indicator $\Delta_3 = 1\{\min(Y_1, Y_2) < Z\} = \min(1, \Delta_1 + \Delta_2)$ for $T^* = \min\{Y_1, Y_2\}$. Suppose $X$ is a continuous covariate that affects both the marginal survival functions and the dependence structure of $Y_1$ and $Y_2$. The observed data is $\{(T_{1i}, T_{2i}, \Delta_{1i}, \Delta_{2i}, X_i), i = 1, \cdots, n\}$.

## 2.2 Model

For the setting described in Section 2.1, the conditional joint survival function of $Y_1$ and $Y_2$ given $X = x$ can be represented as

$$H_X(t_1, t_2 \mid x) = \mathcal{C}_X\{S_{1|X}(t_1 \mid x), S_{2|X}(t_2 \mid x) \mid x\},$$

where $S_{j|X}(t_j \mid x) = P(Y_j > t_j \mid X = x)$ is the conditional marginal survival function of $Y_j$ given $X = x$ for $j = 1, 2$, and $\mathcal{C}_X$ is the conditional copula. We assume that $\mathcal{C}_X$ belongs to the same Archimedean copula family with the generator function $\psi : (0, 1] \to R$ for each $x$ in the range of the covariate. Let us consider two forms of joint survival function.

**Constant conditional copula model:**

$$\mathcal{C}_X\{S_{1|X}(t_1 \mid x), S_{2|X}(t_2 \mid x) \mid x\} = \psi_\theta^{-1}\left(\psi_\theta(S_{1|X}(t_1 \mid x)) + \psi_\theta(S_{2|X}(t_2 \mid x))\right), \tag{2.1}$$

where $\theta$ is the copula parameter and $\psi : (0,1] \in \mathbb{R}^+$ is the Archimedean generator.

**Varying conditional copula model:**

$$\mathcal{C}_X\{S_{1|X}(t_1 \mid x), S_{2|X}(t_2 \mid x) \mid x\} = \psi^{-1}_{\theta(x)}\left(\psi_{\theta(x)}(S_{1|X}(t_1 \mid x)) + \psi_{\theta(x)}(S_{2|X}(t_2 \mid x))\right), \tag{2.2}$$

where $\theta(x)$ is the copula parameter and $\psi : (0,1] \in \mathbb{R}^+$ is the Archimedean generator. The generator functions as well as their inverses and derivatives for the commonly used Archimedean copulas can be found in Section 1.1.

## 2.3 Estimation of the Conditional Marginal Survival Functions

Given the observed data, $\{(T_{1i}, T_{2i}, \Delta_{1i}, \Delta_{2i}, X_i), i = 1, \ldots, n\}$ one needs to estimate $S_{1|X}(t_1 \mid x)$, $S_{2|X}(t_2 \mid x)$ and $\theta(\cdot)$ to fit the model in (2.1). As the terminal event time is independent of the censoring time, $S_{2|X}(\cdot \mid x)$ can be estimated using Beran's estimator (Beran, 1981). In survival analysis, event (recurrence of any disease) or censoring (death) times often accompanied by some characteristics (e.g., age, sex). For these type of situation, Beran (1981) proposed a nonparametric estimator on conditional margins, which takes the form

$$\widehat{S}_{2|X}(t|x) = \prod_{T_{2i}\leq t, \Delta_{2i}=1} \left(1 - \frac{w_{ni}(x,h)}{1-\sum_{j=1}^{i-1} w_{ni}(x,h)}\right), \tag{2.3}$$

with the weights $w_{ni}(x,h) = K_h(X_i - x)/\sum_{j=1}^{n} K_h(X_j - x)$, where $K_h(\cdot) = K(\cdot \mid h)/h$, with $K$ the kernel function and $h$ the bandwidth parameter. The choice of bandwidth $h$ plays an important role in Beran's estimator of marginal survival function. We select the bandwidth value that minimizes the leave-one-out cross validated criterion

$$\widehat{B}(h) = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{n} \Delta_{ji}^{*}(1\{T_{2i} \leq T_{2j}\} - \widehat{F}_{2|X}^{(-i)}(T_{2j} \mid X_i))^2, \tag{2.4}$$

where $\widehat{F}_{2|X}^{(-i)}(\cdot \mid X_i) = 1 - \widehat{S}_{2|X}^{(-i)}(\cdot \mid X_i)$ is the estimator of the conditional marginal distribution obtained from the Beran's estimator using the data $\{(T_{1i}, T_{2i}, \Delta_{1i}, \Delta_{2i}, X_i), i = 1, \ldots, n\}$ with $j = 1, \cdots, i-1, i+1, \cdots, n$. Here, following Geerdens et al. (2017), we use the indicator $\Delta_{ji}^{*} = 1$ for a useful pair of observed times and $\Delta_{ji}^{*} = 0$ for all other pairs. A pair of observed times, $(T_{2i}, T_{2j})$ is useful if one of the following holds: (i) $(\Delta_{2i}, \Delta_{2j}) = (1,1)$; (ii) $(\Delta_{2i}, \Delta_{2j}) = (1,0)$ and $T_{2i} \leq T_{2j}$ ; (iii) $(\Delta_{2i}, \Delta_{2j}) = (0,1)$ and $T_{2i} \geq T_{2j}$ ; or (iv) i = j with $i, j = 1, \cdots, n$.

However, the estimation of $S_{1|X}(\cdot \mid x)$ is more involved due to dependent censoring. Suppose the conditional copula $C_X$ is known. Then, one can utilize

the dependence between $\widehat{S}_{1|X}(\cdot \mid x)$ and $\widehat{S}_{2|X}(\cdot \mid x)$ to estimate $\theta(\cdot)$. Hence, the conditional copula-graphic estimator takes a form similar to the copula-graphic estimator,

$$\widehat{S}_{1|X}(t|x) = \psi_\theta^{-1}\left(-\sum_{T_{1i}\leq t, \Delta_{1i}=1}\left(\psi_\theta(\hat{\Gamma}(T_{1i}^-|x)) - \psi_\theta(\hat{\Gamma}(T_{1i}|x))\right)\right), \tag{2.5}$$

where $\hat{\Gamma}(.)$ is the conditional survival function estimator of $T^* = \min\{Y_1, Y_2\}$, obtained using Beran's estimator in equation (2.3) with the censoring indicator $\Delta_3$. Bandwidth selection mechanism for the conditional copula graphic estimator is similar as the Beran's bandwidth selection given in equation (2.4). Since the conditional copula $C_X(\cdot, \cdot \mid x)$ is unknown in practice, one needs to fit the conditional copula model, which we discuss next.

## 2.4 Estimation of the Conditional Copula Parameter

Based on the two models introduced in Section 2.2 , we propose two versions of conditional copula-graphic estimator (CCGE).

**Case I:** Covariate only affects the marginal survival functions and the copula parameter is a constant.

**Case II:** Covariate affects both the margins and the dependent structure, copula parameter is a function of covariate.

### 2.4.1 Estimation of $\theta$

Given the estimates of the conditional marginal survival functions, the copula parameter $\theta$ is estimated maximizing the pseudo copula log-likelihood function, which is similar to the unconditional setting in Heuchenne et al. (2014). The pseudo copula log-likelihood function is given by

$$\sum_{i=1}^{n} \ell(\theta, \widehat{S}_{1|X}(t_{1i} \mid x), \widehat{S}_{2|X}(t_{2i} \mid x); \Delta_{1i}, \Delta_{2i}),$$

where $\ell(\cdot)$ is the pseudo copula log-likelihood function and $\widehat{S}_{1|X}(t_1 \mid x)$ is the conditional copula-graphic estimator referred to as CCGE. For one observation, Heuchenne et al. (2014) defines the likelihood as

$$L(\theta, \widehat{S}_{1|X}, \widehat{S}_{2|X} \mid \cdot) = [\psi'_\theta(\widehat{S}_{1|X})]^{\Delta_1}[\psi'_\theta(\widehat{S}_{2|X})]^{\Delta_2} \times (\psi_\theta^{-1})^{(\Delta_1+\Delta_2)} \left(\psi_\theta(\widehat{S}_{1|X}) + \psi_\theta(\widehat{S}_{2|X})\right). \tag{2.6}$$

### 2.4.2 Estimation of $\theta(X)$

Given the estimates of the conditional marginal survival functions, the copula parameter $\theta(x)$ is estimated using the approach in Geerdens et al. (2017) for bivariate right-censored data. The estimation is achieved by maximizing the local pseudo copula log-likelihood function

$$\sum_{i=1}^{n} \ell(\theta(X_i), \widehat{S}_{1|X}(t_{1i} \mid x), \widehat{S}_{2|X}(t_{2i} \mid x); \Delta_{1i}, \Delta_{2i})\ K_{h_C}(X_i - x), \tag{2.7}$$

where $\ell(\cdot)$ is the local pseudo copula log-likelihood function.

Due to the restricted parameter range of some copula families, $\theta(X)$ are re-parametrized as $\theta(x) = g^{-1}(\eta(x))$, where $\eta(.)$ is called the calibration function and $g^{-1} : R \to \Theta$ is a inverse-link function with $\Theta$ being the parameter space of a given copula family Acar et al. (2011). For different Archimedean copula families $\theta$ is re-parametrized as $\theta(x) = \eta(x)$ for the Frank copula, $\theta(x) = \exp(\eta) + 1$ for the Gumbel copula and $\theta(x) = \exp(\eta)$ for the Clayton copula.

In semi-competing risks setting, it is difficult to capture the effect of covariate on the dependence structure. In the literature, it is advised to employ nonparametric strategies for determining the impact of covariate on the dependence strength (Acar et al., 2011). Suppose $\gamma(.)$ is sufficiently smooth with $(p+1)^{\text{th}}$ derivative at the point $x$, then, for a given covariate value $X_i$, in a neighbourhood of $x$, $\gamma(X_i)$ can be approximated by Taylor expansion of order $p$ :

$$\gamma(X_i) \approx \gamma(x) + \gamma^{(1)}(x)(X_i - x) + \cdots + \frac{\gamma^{(p)}(x)}{p!}(X_i - x)^p.$$

Letting $p = 1$, i.e. ignoring the higher order derivatives, $\theta(X_i)$ can be approximated as,

$$\theta(X_i) = g^{-1}(\gamma_{0x} + \gamma_{1x}(X_i - x)).$$

The estimate of $\gamma_{0x}$ gives the copula parameter estimate at $x$, i.e. $\hat{\theta}(x) = \hat{\gamma}_{0x}$. In this case, the conditional copula-graphic estimator, $\widehat{S}_{1|X}(t_1 \mid x)$ is referred to as CCGE*. The bandwidth value, $h_C$ of equation (2.7) is determined by maximizing the leave-one-out cross-validated local log-likelihood function (Acar et al., 2011).

## 2.5 Iterative Algorithm

As discussed in Section 2.3, $\widehat{S}_{1|X}$ is needed to estimate the copula parameter, $\hat{\theta}(X)$ and $\hat{\theta}(X)$ is needed to estimate, $\widehat{S}_{1|X}$. Therefore we define an iterative algorithm for estimating copula parameter and marginal survival function of the non-terminal event simultaneously using equation (2.5) and equation (2.6). The steps of the iterative algorithm are provided in Algorithm 1.

**Algorithm 1:** Iterative algorithm for estimation of the conditional survival time and conditional dependence parameter in semi-competing risks.

**Step 1.** Given data, use leave-one-out cross validation technique for selecting bandwidth value and obtain $\widehat{S}_{2|X}(t_{2i}|x_i)$ using the Beran's estimator.

For $m = 1, 2, ..$

**Step 2.** Select bandwidth value using leave-one-out cross validation technique and given $\hat{\theta}^{(m-1)}(x_i)$, obtain $\widehat{S}^{(m)}_{1|X}(t_{1i}|x_i)$ for $i = 1, \ldots, n$ using the conditional copula-graphic estimator.

**Step 3.** Given $\{\widehat{S}^{(m)}_{1|X}(t_{1i}|x_i), \widehat{S}_{2|X}(t_{2i}|x_i); i = 1, \ldots, n\}$ obtain bandwidth value $h_C$ of the likelihood function and estimate $\hat{\theta}^{(m)}(x_i)$ for $i = 1, \ldots, n$ using local pseudo log-likelihood function.

**Step 4.** Bandwidth values can be fixed or updated in each iteration. Repeat step 2-3 until convergence in $\hat{\theta}(x_i)$ or $\widehat{S}_{1|X}(t_{1i}|x_i)$ is achieved.

# Chapter 3

# Simulation Study

This chapter presents the results of our simulation study, where we investigate the performances of the proposed conditional copula-graphic estimators in comparison to the unconditional version. We describe the simulation setting in Section 3.1, and the evaluation criterion in Section 3.2. The estimation results under various data generating models are provided in Sections 3.3 – 3.5. The copula-graphic estimators, including the iterative algorithm are implemented in **R** using own codes. For data generation and likelihood calculations, we used the **R**-copula (Hofert et al., 2014) and **R**-VineCopula (Schepsmeier et al., 2012) packages.

## 3.1 Simulation Setting

We compare the performance of our proposed conditional copula-graphic estimators (CCGE, CCGE*) with the unconditional copula-graphic estimator

(CGE). For the conditional survival function, we consider the exponential model,

$$S_{j|X}(Y_{ji}) = \exp(-\lambda_j Y_{ji} \exp(\beta_j X_i)),$$

where $\lambda_j$ is a constant and $\beta_j$ is the coefficient of covariate, for $j = 1, 2$. We consider the following models for the covariate effects.

- Model 1: $\beta_1 = \beta_2 = 0$ and $\tau = 0.5$
- Model 2: $\beta_1 = \beta_2 = 1$ and $\tau = 0.5$
- Model 3: $\beta_1 = \beta_2 = 1$ and $\tau = 2(x - 0.5)^2 + 0.3$ with $\tau \in (0.30, 0.80)$
- Model 4: $\beta_1 = \beta_2 = 1$ and $\tau = -2(x - 0.5)^2 + 0.8$ with $\tau \in (0.30, 0.80)$
- Model 5: $\beta_1 = \beta_2 = 1$ and $\tau = 3.5(x-0.5)^3 - 0.003$ with $\tau \in (-0.41, 0.43)$

These models describe the situations where (1) covariate has no effect on the marginal survival functions or dependence (Model 1), (2) covariate affects only the marginal survival functions (Model 2), and (3) covariate affects both the marginal survival functions and the dependence (Model 3, 4 and 5). The models are specified in terms of Kendall's tau from which the copula parameter, $\theta$ is obtained using the conversions in Section 1.1.

Under Models 1 - 4, we generated data $\{(U_{1i}, U_{2i} \mid X_i) : i = 1, 2, \ldots, n\}$ of size $n = 100$ from the Clayton, Frank and Gumbel families. We use the fixed design for the covariate $X$, with values equally spaced between 0 and 1. Under

Model 5, where $\tau(X)$ takes negative values, we consider only the Frank family. Then, we apply the inverse-cdf method to obtain the event times $Y_{1i} = S^{-1}_{1|X}(U_{1i})$ and $Y_{2i} = S^{-1}_{2|X}(U_{2i})$ from the copula data using the exponential model. We set $\lambda_1 = 1$ and determine $\lambda_2$ value so that the non-terminal event has no (approximately 0%), low (approximately 25%) and moderate (approximately 50%) censoring rate. The censoring variable $Z$ has Uniform(0, b) distribution with $P(Y_2 > Z) = 0.20$. The observed data are the minimum of the events time and censoring times with indicator variables as discussed in Section 2.1.

## 3.2 Evaluation Criteria for Estimators

After data generation, we estimated survival probabilities using CGE, CCGE, CCGE* and copula parameters $\theta$ or $\theta(X)$. Let $\hat{\alpha}$ stands for the estimator $\widehat{S}_1$ or $\hat{\tau}$. Under each setting, we replicate the experiment M=1000 times and evaluated estimation performance through the integrated squared bias ($\text{IBias}^2$), integrated mean squared Error (IMSE) and integrated variance (IVAR) given by

$$\text{IBias}^2(\hat{\alpha}) = \int_X [E[\hat{\alpha}(x)] - \alpha(x)]^2 dx,$$

$$\text{IVAR}(\hat{\alpha}) = \int_X E[\{\hat{\alpha}(x) - E[\hat{\alpha}(x)]\}^2] dx,$$

$$\text{IMSE}(\hat{\alpha}) = \int_X E[\{\hat{\alpha}(x) - \hat{\alpha}(x)\}^2] dx = \text{IBias}^2(\hat{\alpha}) + \text{IVAR}(\hat{\alpha}).$$

Under each model, we first display the estimation results of $\widehat{S}_{1|X}$ (or $\widehat{S}_1$) and the summarize the results of $\hat{\tau}(X)$ or $(\hat{\tau})$.

## 3.3 Simulation Results under Model 1

When there is no covariate effect, the copula-graphic estimator (CGE) is expected to fit best among the three estimators. The estimation results of $\widehat{S}_1$ under the Frank, Gumbel and Clayton families with $\tau = 0.5$ are summarized in Table 3.1 at three different quantiles p = (0.1, 0.5, 0.9).

From Table 3.1, we notice that CCGE and CCGE* estimates have very small integrated bias and comparatively small integrated mean square error. This indicates that incorporating covariate effect even when it is not significant does not have a significant negative impact on the estimation performance. We also notice that the estimation performance deteriorates at higher censoring rates of the non-terminal event. Similar conclusions are reached under three copula families and across three quantiles in this setting.

The results of the Kendall's tau estimates for the data generated under Model 1 are summarized in Table 3.2. When there is no covariate effect, all three estimators (CGE, CCGE and CCGE*) have very small integrated bias. We also notice that, for no and low censoring rates, the integrated bias is comparatively higher for Clayton family and for moderate censoring rate the integrated bias is comparatively higher for Gumbel family. The integrated variance and inte-

grated mean squared errors for CCGE* are the highest in all cases. Note that the copula parameter is estimated nonparametrically in this estimator, hence we expect more variability in CCGE* than CCGE. An increase in censoring rate reduces the efficiency of all three estimators as can be concluded from the increasing mean square errors.

A graphical representation of the Kendall's tau estimates of CCGE* at different censoring rates under each copula family with $\tau = 0.5$ is given in Figure (3.1). As can be seen, the Kendall's tau estimates coincide with the true parameter value; the pattern is similar for the three copula families.

The results above are for moderate dependence ($\tau = 0.5$). To address the impact of the strength of dependence, we focus on the Frank copula and consider $\tau = 0.2$, 0.5 and 0.8. The results of the conditional marginal survival function estimates are given in Table 3.3, and the summary of Kendall's tau estimates is provided in Table 3.4. From Table 3.3, we see that the estimation performance of the three estimators do not change with the $\tau$ values. While all three estimators produce low bias in the Kendall's tau estimates, the integrated mean squared errors and variances are smaller for larger $\tau$ values, as can be seen in Table 3.4.

Table 3.1: Mean, integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the $\widehat{S}_1(.)$ at different quantiles under Model 1.

| Copula Family | p | Censoring Rate | CGE | | | | CCGE | | | | CCGE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $E(\widehat{S}_1(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE |
| Frank | 0.1 | 0% | 0.102 | 0.000 | 0.089 | 0.089 | 0.102 | 0.000 | 0.131 | 0.131 | 0.102 | 0.000 | 0.131 | 0.131 |
| | | 25% | 0.102 | 0.000 | 0.118 | 0.118 | 0.103 | 0.001 | 0.209 | 0.210 | 0.102 | 0.000 | 0.168 | 0.168 |
| | | 50% | 0.099 | 0.000 | 0.310 | 0.310 | 0.084 | 0.026 | 0.691 | 0.717 | 0.084 | 0.025 | 0.723 | 0.747 |
| | 0.5 | 0% | 0.496 | 0.001 | 0.252 | 0.253 | 0.496 | 0.002 | 0.427 | 0.428 | 0.496 | 0.002 | 0.426 | 0.428 |
| | | 25% | 0.497 | 0.000 | 0.292 | 0.292 | 0.497 | 0.001 | 0.483 | 0.484 | 0.497 | 0.001 | 0.455 | 0.456 |
| | | 50% | 0.497 | 0.001 | 0.360 | 0.361 | 0.495 | 0.002 | 0.553 | 0.556 | 0.495 | 0.004 | 0.593 | 0.597 |
| | 0.9 | 0% | 0.900 | 0.000 | 0.085 | 0.085 | 0.900 | 0.000 | 0.130 | 0.130 | 0.900 | 0.000 | 0.130 | 0.130 |
| | | 25% | 0.900 | 0.000 | 0.090 | 0.090 | 0.900 | 0.000 | 0.130 | 0.130 | 0.900 | 0.000 | 0.145 | 0.145 |
| | | 50% | 0.900 | 0.000 | 0.102 | 0.102 | 0.900 | 0.000 | 0.157 | 0.157 | 0.898 | 0.000 | 0.164 | 0.165 |
| Gumbel | 0.1 | 0% | 0.102 | 0.000 | 0.088 | 0.089 | 0.102 | 0.000 | 0.130 | 0.131 | 0.102 | 0.000 | 0.130 | 0.131 |
| | | 25% | 0.102 | 0.000 | 0.128 | 0.129 | 0.102 | 0.001 | 0.218 | 0.219 | 0.102 | 0.000 | 0.175 | 0.175 |
| | | 50% | 0.100 | 0.000 | 0.347 | 0.347 | 0.084 | 0.024 | 0.719 | 0.743 | 0.085 | 0.022 | 0.749 | 0.771 |
| | 0.5 | 0% | 0.496 | 0.001 | 0.251 | 0.253 | 0.496 | 0.002 | 0.424 | 0.426 | 0.496 | 0.002 | 0.424 | 0.426 |
| | | 25% | 0.496 | 0.001 | 0.274 | 0.275 | 0.496 | 0.002 | 0.456 | 0.458 | 0.496 | 0.002 | 0.440 | 0.441 |
| | | 50% | 0.493 | 0.005 | 0.339 | 0.344 | 0.489 | 0.012 | 0.523 | 0.536 | 0.489 | 0.012 | 0.555 | 0.567 |
| | 0.9 | 0% | 0.900 | 0.000 | 0.085 | 0.085 | 0.900 | 0.000 | 0.129 | 0.130 | 0.900 | 0.000 | 0.129 | 0.130 |
| | | 25% | 0.900 | 0.000 | 0.083 | 0.083 | 0.900 | 0.000 | 0.132 | 0.132 | 0.900 | 0.000 | 0.130 | 0.131 |
| | | 50% | 0.900 | 0.000 | 0.086 | 0.086 | 0.899 | 0.000 | 0.137 | 0.137 | 0.899 | 0.000 | 0.143 | 0.143 |
| Clayton | 0.1 | 0% | 0.102 | 0.000 | 0.089 | 0.089 | 0.102 | 0.000 | 0.131 | 0.131 | 0.102 | 0.001 | 0.194 | 0.194 |
| | | 25% | 0.103 | 0.001 | 0.121 | 0.122 | 0.104 | 0.001 | 0.192 | 0.194 | 0.103 | 0.001 | 0.181 | 0.182 |
| | | 50% | 0.109 | 0.007 | 0.252 | 0.260 | 0.109 | 0.008 | 0.386 | 0.394 | 0.118 | 0.032 | 0.994 | 1.026 |
| | 0.5 | 0% | 0.496 | 0.001 | 0.252 | 0.254 | 0.496 | 0.002 | 0.426 | 0.427 | 0.497 | 0.001 | 0.445 | 0.446 |
| | | 25% | 0.500 | 0.000 | 0.333 | 0.333 | 0.500 | 0.000 | 0.520 | 0.521 | 0.498 | 0.001 | 0.481 | 0.482 |
| | | 50% | 0.498 | 0.000 | 0.446 | 0.446 | 0.497 | 0.001 | 0.626 | 0.627 | 0.500 | 0.001 | 0.796 | 0.797 |
| | 0.9 | 0% | 0.900 | 0.000 | 0.085 | 0.086 | 0.900 | 0.000 | 0.130 | 0.130 | 0.900 | 0.000 | 0.130 | 0.130 |
| | | 25% | 0.900 | 0.000 | 0.090 | 0.090 | 0.900 | 0.000 | 0.142 | 0.142 | 0.900 | 0.000 | 0.136 | 0.136 |
| | | 50% | 0.900 | 0.000 | 0.098 | 0.098 | 0.899 | 0.000 | 0.146 | 0.146 | 0.899 | 0.000 | 0.158 | 0.158 |

Table 3.2: Integrated squared bias ($IBias^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the Kendall's tau estimates under Model 1.

| Copula Family | Censoring Rate | CGE | | | CCGE | | | CCGE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE |
| Frank | 0% | 0.004 | 0.242 | 0.246 | 0.003 | 0.281 | 0.284 | 0.002 | 0.793 | 0.796 |
| | 25% | 0.000 | 0.309 | 0.310 | 0.000 | 0.373 | 0.373 | 0.004 | 0.957 | 0.962 |
| | 50% | 0.005 | 0.491 | 0.496 | 0.020 | 0.614 | 0.634 | 0.038 | 1.905 | 1.943 |
| Gumbel | 0% | 0.000 | 0.295 | 0.295 | 0.004 | 0.299 | 0.303 | 0.009 | 0.755 | 0.763 |
| | 25% | 0.003 | 0.339 | 0.342 | 0.019 | 0.371 | 0.390 | 0.010 | 0.870 | 0.880 |
| | 50% | 0.046 | 0.512 | 0.559 | 0.140 | 0.600 | 0.740 | 0.127 | 1.764 | 1.891 |
| Clayton | 0% | 0.031 | 0.300 | 0.331 | 0.036 | 0.381 | 0.417 | 0.024 | 0.890 | 0.914 |
| | 25% | 0.029 | 0.375 | 0.404 | 0.038 | 0.495 | 0.533 | 0.041 | 1.153 | 1.193 |
| | 50% | 0.001 | 0.518 | 0.519 | 0.001 | 0.665 | 0.665 | 0.023 | 2.355 | 2.378 |

Figure 3.1: Mean, $5^{\text{th}}$ and $95^{\text{th}}$ quantiles of Kendall's tau estimates with no (left column), low (middle column) and moderate (right column) censoring rates for Frank (top), Gumbel (middle) and Clayton (bottom) families. Dashed, dotted and solid line represent the mean Kendall's tau estimates, quantiles and the true Kendall's tau, respectively.

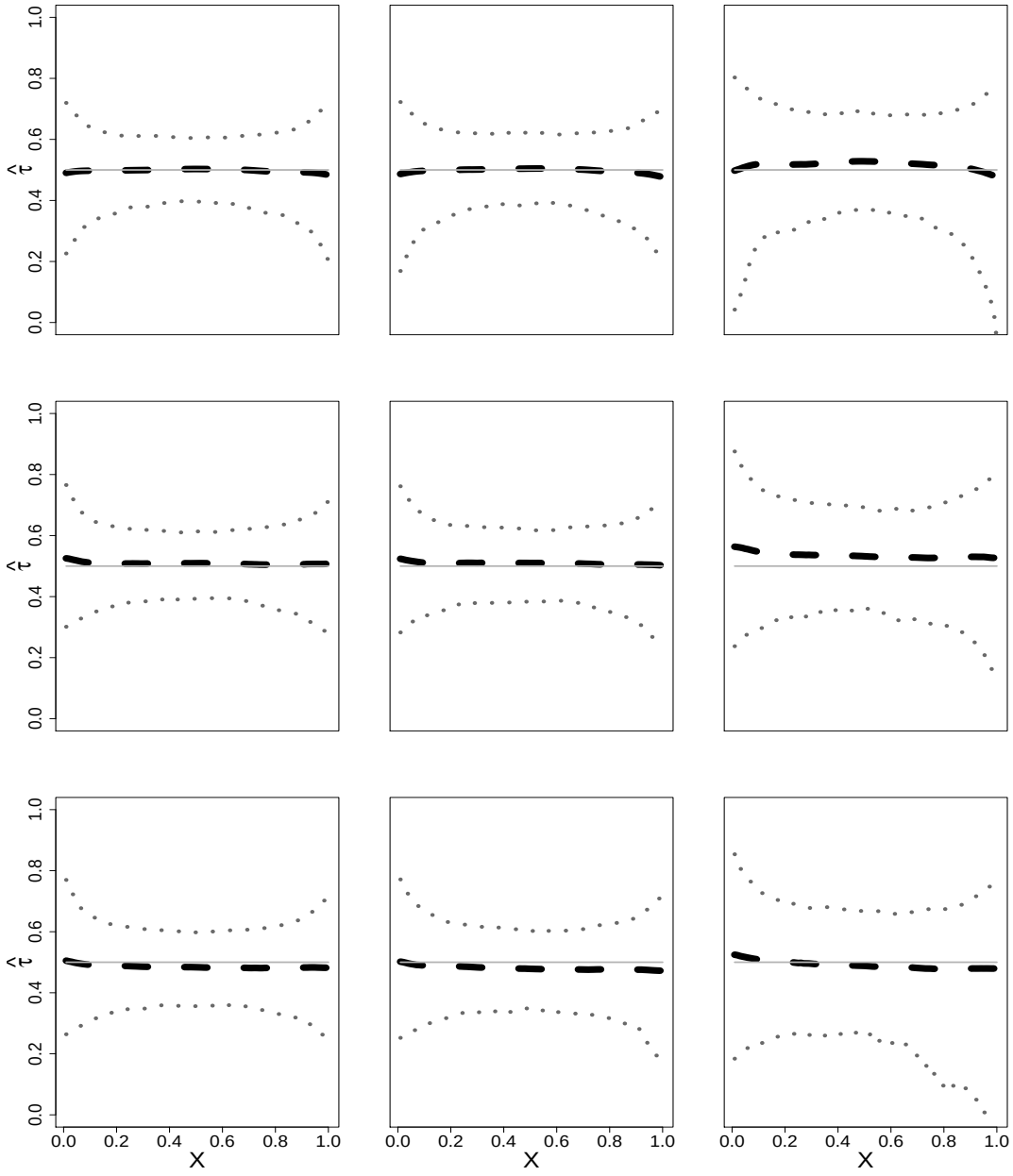


Table 3.3: Mean, integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the $\widehat{S}_1(.)$ at different quantiles with Frank copula and $\tau = 0.2$, 0.5 and 0.8 under Model 1.

| $\tau$ | p | Censoring Rate | CGE | | | | CCGE | | | | CCGE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $E(\widehat{S}_1(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE |
| 0.2 | 0.1 | 0% | 0.102 | 0.000 | 0.089 | 0.089 | 0.102 | 0.000 | 0.132 | 0.133 | 0.102 | 0.000 | 0.132 | 0.133 |
| | | 25% | 0.102 | 0.000 | 0.132 | 0.132 | 0.102 | 0.000 | 0.211 | 0.211 | 0.103 | 0.001 | 0.212 | 0.213 |
| | | 50% | 0.087 | 0.016 | 0.580 | 0.596 | 0.072 | 0.081 | 0.782 | 0.863 | 0.061 | 0.149 | 0.889 | 1.038 |
| | 0.5 | 0% | 0.496 | 0.001 | 0.253 | 0.254 | 0.496 | 0.002 | 0.428 | 0.430 | 0.496 | 0.002 | 0.428 | 0.430 |
| | | 25% | 0.496 | 0.001 | 0.292 | 0.294 | 0.496 | 0.002 | 0.486 | 0.488 | 0.496 | 0.001 | 0.481 | 0.482 |
| | | 50% | 0.496 | 0.001 | 0.430 | 0.431 | 0.494 | 0.004 | 0.639 | 0.643 | 0.492 | 0.006 | 0.757 | 0.763 |
| | 0.9 | 0% | 0.900 | 0.000 | 0.085 | 0.085 | 0.900 | 0.000 | 0.130 | 0.130 | 0.900 | 0.000 | 0.130 | 0.130 |
| | | 25% | 0.900 | 0.000 | 0.087 | 0.087 | 0.900 | 0.000 | 0.139 | 0.139 | 0.900 | 0.000 | 0.137 | 0.137 |
| | | 50% | 0.899 | 0.000 | 0.094 | 0.094 | 0.899 | 0.000 | 0.146 | 0.146 | 0.898 | 0.001 | 0.156 | 0.157 |
| 0.5 | 0.1 | 0% | 0.102 | 0.000 | 0.089 | 0.089 | 0.102 | 0.000 | 0.131 | 0.131 | 0.102 | 0.000 | 0.131 | 0.131 |
| | | 25% | 0.102 | 0.000 | 0.118 | 0.118 | 0.103 | 0.001 | 0.209 | 0.210 | 0.102 | 0.000 | 0.168 | 0.168 |
| | | 50% | 0.099 | 0.000 | 0.310 | 0.310 | 0.084 | 0.026 | 0.691 | 0.717 | 0.084 | 0.025 | 0.723 | 0.747 |
| | 0.5 | 0% | 0.496 | 0.001 | 0.252 | 0.253 | 0.496 | 0.002 | 0.427 | 0.428 | 0.496 | 0.002 | 0.426 | 0.428 |
| | | 25% | 0.497 | 0.000 | 0.292 | 0.292 | 0.497 | 0.001 | 0.483 | 0.484 | 0.497 | 0.001 | 0.455 | 0.456 |
| | | 50% | 0.497 | 0.001 | 0.360 | 0.361 | 0.495 | 0.002 | 0.553 | 0.556 | 0.495 | 0.004 | 0.593 | 0.597 |
| | 0.9 | 0% | 0.900 | 0.000 | 0.085 | 0.085 | 0.900 | 0.000 | 0.130 | 0.130 | 0.900 | 0.000 | 0.130 | 0.130 |
| | | 25% | 0.900 | 0.000 | 0.090 | 0.090 | 0.900 | 0.000 | 0.130 | 0.130 | 0.900 | 0.000 | 0.145 | 0.145 |
| | | 50% | 0.900 | 0.000 | 0.102 | 0.102 | 0.900 | 0.000 | 0.157 | 0.157 | 0.898 | 0.000 | 0.164 | 0.165 |
| 0.8 | 0.1 | 0% | 0.102 | 0.000 | 0.088 | 0.089 | 0.102 | 0.000 | 0.130 | 0.131 | 0.102 | 0.000 | 0.130 | 0.131 |
| | | 25% | 0.102 | 0.001 | 0.128 | 0.128 | 0.102 | 0.000 | 0.227 | 0.227 | 0.102 | 0.000 | 0.156 | 0.157 |
| | | 50% | 0.104 | 0.001 | 0.183 | 0.185 | 0.103 | 0.001 | 0.486 | 0.487 | 0.103 | 0.001 | 0.491 | 0.493 |
| | 0.5 | 0% | 0.496 | 0.001 | 0.251 | 0.253 | 0.496 | 0.001 | 0.425 | 0.426 | 0.496 | 0.002 | 0.424 | 0.426 |
| | | 25% | 0.497 | 0.001 | 0.273 | 0.273 | 0.497 | 0.001 | 0.455 | 0.456 | 0.496 | 0.002 | 0.425 | 0.427 |
| | | 50% | 0.498 | 0.001 | 0.301 | 0.301 | 0.498 | 0.000 | 0.465 | 0.466 | 0.499 | 0.000 | 0.479 | 0.479 |
| | 0.9 | 0% | 0.900 | 0.000 | 0.085 | 0.085 | 0.900 | 0.000 | 0.130 | 0.130 | 0.900 | 0.000 | 0.130 | 0.130 |
| | | 25% | 0.901 | 0.000 | 0.099 | 0.099 | 0.902 | 0.000 | 0.151 | 0.151 | 0.900 | 0.000 | 0.133 | 0.133 |
| | | 50% | 0.900 | 0.000 | 0.111 | 0.111 | 0.902 | 0.000 | 0.156 | 0.156 | 0.901 | 0.000 | 0.157 | 0.157 |

Table 3.4: Integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the Kendall's tau estimates with Frank copula and $\tau = 0.2$, 0.5 and 0.8 under Model 1.

| $\tau$ | Censoring Rate | CGE | | | CCGE | | | CCGE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\text{IBias}^2$ | IVAR | IMSE | $\text{IBias}^2$ | IVAR | IMSE | $\text{IBias}^2$ | IVAR | IMSE |
| 0.2 | 0% | 0.001 | 0.423 | 0.424 | 0.001 | 0.482 | 0.483 | 0.004 | 1.468 | 1.473 |
| | 25% | 0.000 | 0.569 | 0.569 | 0.000 | 0.638 | 0.638 | 0.004 | 1.937 | 1.942 |
| | 50% | 0.000 | 0.971 | 0.971 | 0.011 | 1.100 | 1.111 | 0.025 | 3.770 | 3.795 |
| 0.5 | 0% | 0.004 | 0.242 | 0.246 | 0.003 | 0.281 | 0.284 | 0.002 | 0.793 | 0.796 |
| | 25% | 0.000 | 0.309 | 0.310 | 0.000 | 0.373 | 0.373 | 0.004 | 0.957 | 0.962 |
| | 50% | 0.005 | 0.491 | 0.496 | 0.020 | 0.614 | 0.634 | 0.038 | 1.905 | 1.943 |
| 0.8 | 0% | 0.010 | 0.046 | 0.056 | 0.029 | 0.081 | 0.110 | 0.029 | 0.195 | 0.224 |
| | 25% | 0.003 | 0.058 | 0.061 | 0.014 | 0.103 | 0.116 | 0.026 | 0.203 | 0.229 |
| | 50% | 0.005 | 0.087 | 0.092 | 0.000 | 0.169 | 0.169 | 0.007 | 0.403 | 0.411 |

## 3.4 Simulation Results under Model 2

When there is a covariate effect on only margins, we expect the conditional copula-graphic estimator (CCGE) to have the best performance among the three estimators. The unconditional copula-graphic estimator (CGE) is expected to perform the worst in this setting because of ignoring covariate effect.

The estimation result of the conditional and unconditional copula-graphic estimators calculated at different quantiles are given in Table 3.5. In this setting, both CCGE and CCGE* estimates are similar with low integrated bias and integrated mean squared error. This suggests that considering covariate effect on dependence, when it is not necessary, does not deteriorate the estimation of survival function. However, when unconditional copula-graphic estimator (CGE) is used, estimates have considerably high integrated bias and integrated mean squared error. Overall, the variances and mean squared error are highest at the $50^{\text{th}}$ quantile and at higher censoring rates.

The Kendall's tau estimates of the three estimators for data generated under Model 2 are summarized in Table 3.6. While all three estimators produce low bias, CCGE* yields comparatively higher variance and mean squared error in this setting. Under the Gumbel family, all three estimators produce slightly higher bias for moderate censoring case than other cases. When censoring rate is higher, estimates have more variability under all families.

Figure (3.2) represent the Kendall's tau estimates at different censoring rates under the three copula families with $\tau = 0.5$ when the covariate affects only the margins, but not the dependence structure. From Figure (3.2) we see that the Kendall's tau estimates coincide with the true value. The Frank family has comparatively wider confidence intervals for $\tau$ in comparison to other families and the confidence intervals get wider with the increase in censoring rates.

To see the effect of the strength of dependence under this setting, we consider $\tau = 0.2$, 0.5 and 0.8 under Frank family as before. From Table 3.7 we found that the value of Kendall's tau does not seem to affect the estimation performance of the three estimators of the survival function. The Kendall's tau estimates for the data generated under Model 2 are summarized in Table 3.8. As Kendall's tau increases, integrated mean squared error of the estimated Kendall's tau decreases for all three estimators.

Table 3.5: Mean, integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the $\widehat{S}_1(.)$ at different quantiles under Model 2.

| Copula Family | p | Censoring Rate | CGE | | | | CCGE | | | | CCGE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $E(\widehat{S}_1(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE |
| Frank | 0.1 | 0% | 0.124 | 0.467 | 0.096 | 0.564 | 0.108 | 0.053 | 0.202 | 0.255 | 0.108 | 0.053 | 0.202 | 0.255 |
| | | 25% | 0.125 | 0.473 | 0.130 | 0.603 | 0.110 | 0.065 | 0.258 | 0.323 | 0.109 | 0.059 | 0.232 | 0.291 |
| | | 50% | 0.125 | 0.538 | 0.288 | 0.826 | 0.109 | 0.150 | 0.473 | 0.623 | 0.109 | 0.164 | 0.503 | 0.667 |
| | 0.5 | 0% | 0.490 | 0.924 | 0.234 | 1.158 | 0.495 | 0.099 | 0.580 | 0.678 | 0.495 | 0.098 | 0.580 | 0.679 |
| | | 25% | 0.490 | 0.912 | 0.269 | 1.180 | 0.496 | 0.111 | 0.632 | 0.742 | 0.496 | 0.104 | 0.611 | 0.716 |
| | | 50% | 0.491 | 0.899 | 0.347 | 1.246 | 0.494 | 0.152 | 0.708 | 0.860 | 0.494 | 0.162 | 0.738 | 0.900 |
| | 0.9 | 0% | 0.893 | 0.091 | 0.088 | 0.180 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.185 |
| | | 25% | 0.893 | 0.091 | 0.098 | 0.190 | 0.898 | 0.008 | 0.194 | 0.202 | 0.898 | 0.008 | 0.189 | 0.196 |
| | | 50% | 0.893 | 0.093 | 0.109 | 0.202 | 0.897 | 0.011 | 0.203 | 0.214 | 0.896 | 0.013 | 0.209 | 0.222 |
| Gumbel | 0.1 | 0% | 0.123 | 0.473 | 0.100 | 0.572 | 0.108 | 0.053 | 0.201 | 0.254 | 0.108 | 0.053 | 0.201 | 0.254 |
| | | 25% | 0.124 | 0.479 | 0.139 | 0.619 | 0.110 | 0.064 | 0.273 | 0.337 | 0.109 | 0.058 | 0.240 | 0.297 |
| | | 50% | 0.122 | 0.513 | 0.301 | 0.814 | 0.107 | 0.126 | 0.518 | 0.644 | 0.108 | 0.131 | 0.546 | 0.677 |
| | 0.5 | 0% | 0.493 | 0.927 | 0.240 | 1.167 | 0.495 | 0.098 | 0.580 | 0.678 | 0.495 | 0.098 | 0.580 | 0.678 |
| | | 25% | 0.493 | 0.915 | 0.263 | 1.178 | 0.494 | 0.110 | 0.604 | 0.714 | 0.495 | 0.106 | 0.600 | 0.707 |
| | | 50% | 0.491 | 0.906 | 0.337 | 1.243 | 0.488 | 0.154 | 0.664 | 0.818 | 0.489 | 0.163 | 0.718 | 0.882 |
| | 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.184 |
| | | 25% | 0.894 | 0.086 | 0.093 | 0.179 | 0.898 | 0.008 | 0.178 | 0.186 | 0.898 | 0.008 | 0.178 | 0.187 |
| | | 50% | 0.893 | 0.092 | 0.104 | 0.197 | 0.897 | 0.111 | 0.180 | 0.191 | 0.897 | 0.012 | 0.191 | 0.203 |
| Clayton | 0.1 | 0% | 0.123 | 0.473 | 0.100 | 0.572 | 0.108 | 0.053 | 0.202 | 0.255 | 0.108 | 0.053 | 0.202 | 0.255 |
| | | 25% | 0.123 | 0.474 | 0.125 | 0.600 | 0.110 | 0.069 | 0.250 | 0.319 | 0.109 | 0.063 | 0.230 | 0.293 |
| | | 50% | 0.126 | 0.505 | 0.245 | 0.750 | 0.118 | 0.122 | 0.379 | 0.501 | 0.126 | 0.174 | 0.715 | 0.889 |
| | 0.5 | 0% | 0.493 | 0.927 | 0.241 | 1.168 | 0.495 | 0.099 | 0.579 | 0.678 | 0.495 | 0.098 | 0.579 | 0.677 |
| | | 25% | 0.495 | 0.925 | 0.297 | 1.222 | 0.498 | 0.113 | 0.685 | 0.798 | 0.497 | 0.100 | 0.632 | 0.733 |
| | | 50% | 0.497 | 0.909 | 0.399 | 1.308 | 0.498 | 0.155 | 0.794 | 0.950 | 0.501 | 0.148 | 0.896 | 1.045 |
| | 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.184 |
| | | 25% | 0.894 | 0.084 | 0.097 | 0.181 | 0.898 | 0.008 | 0.187 | 0.196 | 0.898 | 0.008 | 0.182 | 0.190 |
| | | 50% | 0.894 | 0.087 | 0.108 | 0.195 | 0.898 | 0.011 | 0.189 | 0.205 | 0.898 | 0.013 | 0.195 | 0.208 |

Table 3.6: Integrated squared bias ($IBias^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the Kendall's tau estimates under Model 2.

| Copula Family | Censoring Rate | CGE | | | CCGE | | | CCGE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE |
| Frank | 0% | 0.005 | 0.237 | 0.242 | 0.003 | 0.259 | 0.262 | 0.006 | 0.741 | 0.747 |
| | 25% | 0.015 | 0.305 | 0.320 | 0.000 | 0.323 | 0.324 | 0.005 | 0.895 | 0.901 |
| | 50% | 0.040 | 0.471 | 0.511 | 0.030 | 0.536 | 0.566 | 0.055 | 1.773 | 1.828 |
| Gumbel | 0% | 0.026 | 0.284 | 0.309 | 0.013 | 0.293 | 0.306 | 0.016 | 0.827 | 0.843 |
| | 25% | 0.045 | 0.329 | 0.374 | 0.036 | 0.353 | 0.389 | 0.021 | 0.878 | 0.899 |
| | 50% | 0.139 | 0.506 | 0.645 | 0.183 | 0.550 | 0.733 | 0.196 | 1.628 | 1.824 |
| Clayton | 0% | 0.006 | 0.285 | 0.291 | 0.056 | 0.375 | 0.431 | 0.080 | 0.893 | 0.973 |
| | 25% | 0.002 | 0.370 | 0.372 | 0.056 | 0.423 | 0.478 | 0.085 | 1.080 | 1.165 |
| | 50% | 0.008 | 0.502 | 0.510 | 0.006 | 0.595 | 0.601 | 0.017 | 2.169 | 2.186 |

Figure 3.2: Mean, $5^{\text{th}}$ and $95^{\text{th}}$ quantiles of Kendall's tau estimates with no (left column), low (middle column) and moderate (right column) censoring rates for Frank (top), Gumbel (middle) and Clayton (bottom) families. Dashed, dotted and solid line represent the mean Kendall's tau estimates, quantiles and the true Kendall's tau, respectively.

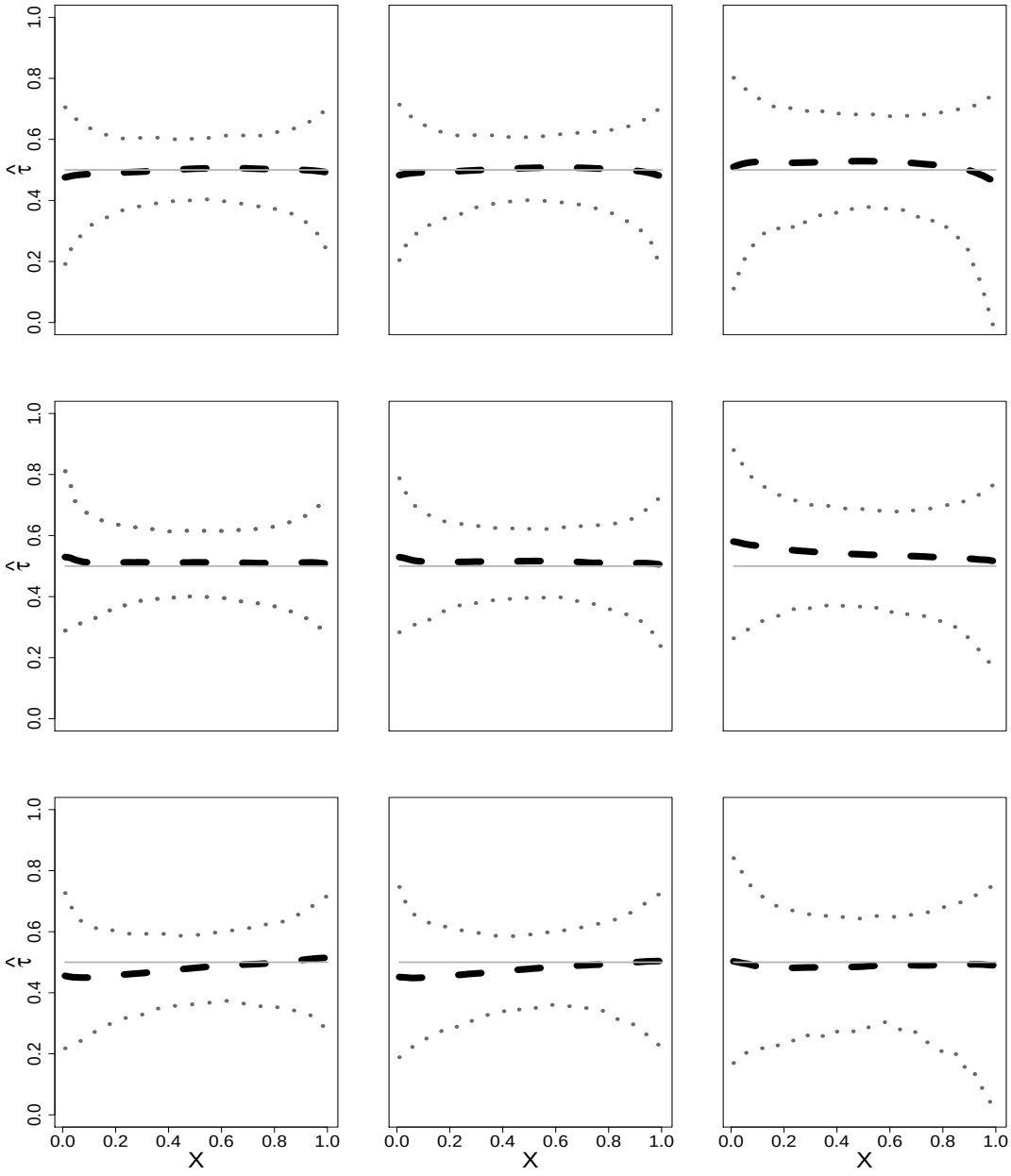


Table 3.7: Mean, integrated squared bias (IBias$^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the $\widehat{S}_1(.)$ at different quantiles with Frank copula and $\tau = 0.2$, 0.5 and 0.8 under Model 2.

| $\tau$ | p | Censoring Rate | CGE | | | | CCGE | | | | CCGE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $E(\widehat{S}_1(\cdot))$ | IBias$^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | IBias$^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | IBias$^2$ | IVAR | IMSE |
| 0.2 | 0.1 | 0% | 0.124 | 0.468 | 0.097 | 0.565 | 0.108 | 0.054 | 0.203 | 0.257 | 0.109 | 0.060 | 0.200 | 0.260 |
| | | 25% | 0.127 | 0.483 | 0.188 | 0.671 | 0.110 | 0.065 | 0.284 | 0.349 | 0.111 | 0.068 | 0.290 | 0.357 |
| | | 50% | 0.120 | 0.594 | 0.461 | 1.055 | 0.106 | 0.186 | 0.661 | 0.848 | 0.102 | 0.190 | 0.761 | 0.951 |
| | 0.5 | 0% | 0.490 | 0.925 | 0.234 | 1.159 | 0.495 | 0.099 | 0.580 | 0.679 | 0.494 | 0.110 | 0.578 | 0.688 |
| | | 25% | 0.489 | 0.912 | 0.298 | 1.210 | 0.495 | 0.104 | 0.656 | 0.760 | 0.495 | 0.106 | 0.656 | 0.762 |
| | | 50% | 0.490 | 0.884 | 0.409 | 1.293 | 0.493 | 0.141 | 0.799 | 0.940 | 0.497 | 0.105 | 0.818 | 0.923 |
| | 0.9 | 0% | 0.893 | 0.091 | 0.088 | 0.180 | 0.898 | 0.008 | 0.176 | 0.184 | 0.898 | 0.008 | 0.178 | 0.187 |
| | | 25% | 0.893 | 0.093 | 0.096 | 0.189 | 0.898 | 0.008 | 0.184 | 0.192 | 0.898 | 0.008 | 0.186 | 0.194 |
| | | 50% | 0.892 | 0.095 | 0.104 | 0.199 | 0.897 | 0.010 | 0.194 | 0.204 | 0.897 | 0.009 | 0.200 | 0.209 |
| 0.5 | 0.1 | 0% | 0.124 | 0.467 | 0.096 | 0.564 | 0.108 | 0.053 | 0.202 | 0.255 | 0.108 | 0.053 | 0.202 | 0.255 |
| | | 25% | 0.125 | 0.473 | 0.130 | 0.603 | 0.110 | 0.065 | 0.258 | 0.323 | 0.109 | 0.059 | 0.232 | 0.291 |
| | | 50% | 0.125 | 0.538 | 0.288 | 0.826 | 0.109 | 0.150 | 0.473 | 0.623 | 0.109 | 0.164 | 0.503 | 0.667 |
| | 0.5 | 0% | 0.490 | 0.924 | 0.234 | 1.158 | 0.495 | 0.099 | 0.580 | 0.678 | 0.495 | 0.098 | 0.580 | 0.679 |
| | | 25% | 0.490 | 0.912 | 0.269 | 1.180 | 0.496 | 0.111 | 0.632 | 0.742 | 0.496 | 0.104 | 0.611 | 0.716 |
| | | 50% | 0.491 | 0.899 | 0.347 | 1.246 | 0.494 | 0.152 | 0.708 | 0.860 | 0.494 | 0.162 | 0.738 | 0.900 |
| | 0.9 | 0% | 0.893 | 0.091 | 0.088 | 0.180 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.185 |
| | | 25% | 0.893 | 0.091 | 0.098 | 0.190 | 0.898 | 0.008 | 0.194 | 0.202 | 0.898 | 0.008 | 0.189 | 0.196 |
| | | 50% | 0.893 | 0.093 | 0.109 | 0.202 | 0.897 | 0.011 | 0.203 | 0.214 | 0.896 | 0.013 | 0.209 | 0.222 |
| 0.8 | 0.1 | 0% | 0.124 | 0.468 | 0.097 | 0.564 | 0.108 | 0.052 | 0.203 | 0.255 | 0.108 | 0.052 | 0.202 | 0.255 |
| | | 25% | 0.124 | 0.465 | 0.116 | 0.581 | 0.109 | 0.061 | 0.272 | 0.333 | 0.108 | 0.054 | 0.221 | 0.275 |
| | | 50% | 0.126 | 0.492 | 0.191 | 0.682 | 0.113 | 0.107 | 0.356 | 0.463 | 0.113 | 0.106 | 0.362 | 0.468 |
| | 0.5 | 0% | 0.490 | 0.924 | 0.233 | 1.157 | 0.495 | 0.098 | 0.582 | 0.679 | 0.495 | 0.098 | 0.582 | 0.679 |
| | | 25% | 0.491 | 0.921 | 0.246 | 1.167 | 0.496 | 0.106 | 0.614 | 0.720 | 0.495 | 0.102 | 0.588 | 0.690 |
| | | 50% | 0.492 | 0.918 | 0.276 | 1.194 | 0.498 | 0.147 | 0.612 | 0.759 | 0.498 | 0.141 | 0.619 | 0.761 |
| | 0.9 | 0% | 0.893 | 0.091 | 0.088 | 0.179 | 0.898 | 0.008 | 0.178 | 0.185 | 0.898 | 0.008 | 0.178 | 0.185 |
| | | 25% | 0.894 | 0.091 | 0.097 | 0.188 | 0.900 | 0.007 | 0.207 | 0.215 | 0.899 | 0.007 | 0.186 | 0.194 |
| | | 50% | 0.894 | 0.089 | 0.112 | 0.200 | 0.900 | 0.009 | 0.208 | 0.217 | 0.900 | 0.008 | 0.207 | 0.215 |

Table 3.8: Integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the Kendall's tau estimates with Frank copula and $\tau = 0.2$, 0.5 and 0.8 under Model 2.

| $\tau$ | Censoring Rate | CGE | | | CCGE | | | CCGE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\text{IBias}^2$ | IVAR | IMSE | $\text{IBias}^2$ | IVAR | IMSE | $\text{IBias}^2$ | IVAR | IMSE |
| 0.2 | 0% | 0.059 | 0.421 | 0.481 | 0.000 | 0.480 | 0.480 | 0.001 | 1.430 | 1.431 |
| | 25% | 0.079 | 0.627 | 0.706 | 0.005 | 0.604 | 0.609 | 0.014 | 1.852 | 1.866 |
| | 50% | 0.080 | 0.930 | 1.010 | 0.034 | 1.069 | 1.103 | 0.086 | 3.483 | 3.570 |
| 0.5 | 0% | 0.005 | 0.237 | 0.242 | 0.003 | 0.259 | 0.262 | 0.006 | 0.741 | 0.747 |
| | 25% | 0.015 | 0.305 | 0.320 | 0.000 | 0.323 | 0.324 | 0.005 | 0.895 | 0.901 |
| | 50% | 0.040 | 0.471 | 0.511 | 0.030 | 0.536 | 0.566 | 0.055 | 1.773 | 1.828 |
| 0.8 | 0% | 0.007 | 0.047 | 0.053 | 0.037 | 0.071 | 0.108 | 0.044 | 0.186 | 0.230 |
| | 25% | 0.004 | 0.049 | 0.053 | 0.016 | 0.089 | 0.105 | 0.040 | 0.195 | 0.235 |
| | 50% | 0.003 | 0.077 | 0.080 | 0.002 | 0.159 | 0.161 | 0.013 | 0.356 | 0.368 |

## 3.5 Simulation Results under Model 3 – 5

When there is a covariate effect on both margins and dependence, we expect the second version of conditional copula-graphic estimator (CCGE*) to perform the best among the three estimators.

Table 3.9, Table 3.10 and Table 3.11 provide the summary of survival function estimates of these estimators under Model 3 (convex), Model 4 (concave) and Model 5 (zero centred), respectively. The results in Table 3.9, Table 3.10 and Table 3.11 are very similar and can be summarized by the following key points. When the covariate affects both margins and dependence, CGE produces high integrated bias and integrated mean squared error in the survival function estimates in comparison to CCGE and CCGE*. On other hand, CCGE and CCGE* have comparatively higher variance than CGE in this setting. For Model 3 and Model 4, we do not see much difference in the performance of the estimators under the three copula families. An increase in the censoring rate results in an increase in variance and mean squared error for each family. We also observe that at $10^{\text{th}}$ quantile, all estimators overestimate the survival probability, whereas at $50^{\text{th}}$ and $90^{\text{th}}$ quantiles the survival probabilities are underestimated.

The Kendall's tau estimates for the data generated under Model 3, Model 4 and Model 5 are summarized in Table 3.12, Table 3.13 and Table 3.14, respectively. From Table 3.12, Table 3.13 and Table 3.14 we see that both CGE and CCGE produce high integrated bias under all copula families because of ignoring the

covariance effect on dependence.

The graphical representations of the Kendall's tau estimates at different censoring rates under the three copula families with $\tau = 0.5$ under Model 3, Model 4 and Model 5 are given in Figure (3.3), Figure (3.4) and Figure (3.5), respectively. From Figure (3.3), Figure (3.4) and Figure (3.5) we see that the Kendall's tau estimates coincide with the truth in all cases, and have wider confidence intervals at higher censoring rates for all three families under each model.

Table 3.9: Mean, integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the $\widehat{S}_1(.)$ at different quantiles under Model 3.

| Copula Family | p | Censoring Rate | CGE | | | | CCGE | | | | CCGE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $E(\widehat{S}_1(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE |
| Frank | 0.1 | 0% | 0.124 | 0.467 | 0.096 | 0.563 | 0.108 | 0.053 | 0.202 | 0.255 | 0.108 | 0.052 | 0.202 | 0.254 |
| | | 25% | 0.123 | 0.445 | 0.132 | 0.577 | 0.107 | 0.054 | 0.259 | 0.313 | 0.108 | 0.051 | 0.231 | 0.282 |
| | | 50% | 0.126 | 0.515 | 0.300 | 0.815 | 0.106 | 0.117 | 0.475 | 0.592 | 0.107 | 0.083 | 0.432 | 0.515 |
| | 0.5 | 0% | 0.490 | 0.925 | 0.234 | 1.159 | 0.495 | 0.098 | 0.579 | 0.677 | 0.495 | 0.098 | 0.579 | 0.677 |
| | | 25% | 0.485 | 0.938 | 0.271 | 1.209 | 0.490 | 0.115 | 0.640 | 0.755 | 0.491 | 0.107 | 0.620 | 0.727 |
| | | 50% | 0.487 | 0.920 | 0.348 | 1.269 | 0.488 | 0.172 | 0.697 | 0.869 | 0.483 | 0.175 | 0.715 | 0.891 |
| | 0.9 | 0% | 0.893 | 0.091 | 0.088 | 0.180 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.184 |
| | | 25% | 0.893 | 0.092 | 0.098 | 0.190 | 0.898 | 0.009 | 0.193 | 0.202 | 0.897 | 0.009 | 0.192 | 0.201 |
| | | 50% | 0.894 | 0.091 | 0.108 | 0.199 | 0.897 | 0.013 | 0.197 | 0.211 | 0.895 | 0.016 | 0.215 | 0.231 |
| Gumbel | 0.1 | 0% | 0.123 | 0.473 | 0.100 | 0.572 | 0.108 | 0.053 | 0.201 | 0.254 | 0.108 | 0.053 | 0.201 | 0.254 |
| | | 25% | 0.121 | 0.449 | 0.135 | 0.584 | 0.107 | 0.052 | 0.277 | 0.330 | 0.107 | 0.051 | 0.240 | 0.291 |
| | | 50% | 0.122 | 0.502 | 0.314 | 0.816 | 0.105 | 0.103 | 0.526 | 0.630 | 0.104 | 0.091 | 0.540 | 0.630 |
| | 0.5 | 0% | 0.493 | 0.927 | 0.240 | 1.167 | 0.495 | 0.098 | 0.580 | 0.678 | 0.495 | 0.098 | 0.580 | 0.678 |
| | | 25% | 0.488 | 0.929 | 0.266 | 1.196 | 0.489 | 0.118 | 0.614 | 0.732 | 0.492 | 0.111 | 0.603 | 0.715 |
| | | 50% | 0.489 | 0.912 | 0.341 | 1.252 | 0.484 | 0.166 | 0.684 | 0.851 | 0.482 | 0.188 | 0.753 | 0.941 |
| | 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.184 |
| | | 25% | 0.893 | 0.090 | 0.094 | 0.185 | 0.897 | 0.010 | 0.181 | 0.191 | 0.897 | 0.010 | 0.182 | 0.192 |
| | | 50% | 0.892 | 0.095 | 0.105 | 0.200 | 0.895 | 0.014 | 0.181 | 0.196 | 0.894 | 0.019 | 0.200 | 0.219 |
| Clayton | 0.1 | 0% | 0.123 | 0.472 | 0.100 | 0.572 | 0.108 | 0.052 | 0.202 | 0.254 | 0.109 | 0.055 | 0.263 | 0.318 |
| | | 25% | 0.121 | 0.447 | 0.125 | 0.572 | 0.108 | 0.057 | 0.255 | 0.311 | 0.109 | 0.054 | 0.228 | 0.282 |
| | | 50% | 0.122 | 0.465 | 0.240 | 0.705 | 0.113 | 0.092 | 0.455 | 0.547 | 0.114 | 0.078 | 0.468 | 0.547 |
| | 0.5 | 0% | 0.493 | 0.928 | 0.242 | 1.169 | 0.495 | 0.098 | 0.581 | 0.679 | 0.495 | 0.099 | 0.598 | 0.697 |
| | | 25% | 0.492 | 0.954 | 0.302 | 1.256 | 0.493 | 0.117 | 0.704 | 0.821 | 0.494 | 0.102 | 0.650 | 0.752 |
| | | 50% | 0.495 | 0.945 | 0.409 | 1.353 | 0.494 | 0.176 | 0.818 | 0.995 | 0.488 | 0.148 | 0.827 | 0.975 |
| | 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.185 |
| | | 25% | 0.895 | 0.083 | 0.097 | 0.180 | 0.898 | 0.009 | 0.187 | 0.196 | 0.898 | 0.009 | 0.186 | 0.195 |
| | | 50% | 0.896 | 0.083 | 0.105 | 0.188 | 0.898 | 0.013 | 0.183 | 0.196 | 0.896 | 0.014 | 0.194 | 0.209 |

Table 3.10: Mean, integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the $\widehat{S}_1(.)$ at different quantiles under Model 4.

| Copula Family | p | Censoring Rate | CGE | | | | CCGE | | | | CCGE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $E(\widehat{S}_1(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE |
| Frank | 0.1 | 0% | 0.123 | 0.472 | 0.100 | 0.572 | 0.108 | 0.053 | 0.203 | 0.256 | 0.108 | 0.053 | 0.203 | 0.256 |
| | | 25% | 0.121 | 0.452 | 0.120 | 0.572 | 0.107 | 0.056 | 0.237 | 0.293 | 0.108 | 0.059 | 0.227 | 0.285 |
| | | 50% | 0.121 | 0.505 | 0.220 | 0.726 | 0.111 | 0.142 | 0.404 | 0.546 | 0.114 | 0.199 | 0.521 | 0.720 |
| | 0.5 | 0% | 0.493 | 0.927 | 0.240 | 1.167 | 0.495 | 0.098 | 0.581 | 0.679 | 0.495 | 0.098 | 0.581 | 0.679 |
| | | 25% | 0.488 | 0.937 | 0.263 | 1.199 | 0.491 | 0.117 | 0.613 | 0.730 | 0.494 | 0.112 | 0.602 | 0.714 |
| | | 50% | 0.489 | 0.928 | 0.311 | 1.239 | 0.494 | 0.155 | 0.682 | 0.837 | 0.499 | 0.160 | 0.767 | 0.927 |
| | 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.177 | 0.185 | 0.898 | 0.008 | 0.177 | 0.185 |
| | | 25% | 0.894 | 0.087 | 0.104 | 0.191 | 0.900 | 0.007 | 0.194 | 0.201 | 0.899 | 0.007 | 0.188 | 0.195 |
| | | 50% | 0.894 | 0.089 | 0.123 | 0.212 | 0.900 | 0.009 | 0.211 | 0.219 | 0.899 | 0.008 | 0.221 | 0.228 |
| Gumbel | 0.1 | 0% | 0.123 | 0.473 | 0.100 | 0.572 | 0.108 | 0.052 | 0.202 | 0.254 | 0.108 | 0.052 | 0.202 | 0.254 |
| | | 25% | 0.121 | 0.458 | 0.126 | 0.584 | 0.107 | 0.059 | 0.245 | 0.304 | 0.108 | 0.059 | 0.229 | 0.288 |
| | | 50% | 0.122 | 0.499 | 0.253 | 0.752 | 0.112 | 0.121 | 0.443 | 0.564 | 0.114 | 0.161 | 0.505 | 0.666 |
| | 0.5 | 0% | 0.493 | 0.927 | 0.240 | 1.168 | 0.495 | 0.098 | 0.580 | 0.677 | 0.495 | 0.098 | 0.580 | 0.678 |
| | | 25% | 0.489 | 0.928 | 0.257 | 1.185 | 0.491 | 0.114 | 0.591 | 0.705 | 0.494 | 0.111 | 0.594 | 0.706 |
| | | 50% | 0.490 | 0.903 | 0.309 | 1.212 | 0.491 | 0.144 | 0.671 | 0.816 | 0.495 | 0.150 | 0.729 | 0.879 |
| | 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.177 | 0.184 | 0.898 | 0.008 | 0.177 | 0.184 |
| | | 25% | 0.893 | 0.090 | 0.094 | 0.183 | 0.897 | 0.009 | 0.176 | 0.186 | 0.898 | 0.009 | 0.177 | 0.186 |
| | | 50% | 0.892 | 0.095 | 0.103 | 0.197 | 0.899 | 0.009 | 0.188 | 0.196 | 0.900 | 0.010 | 0.195 | 0.204 |
| Clayton | 0.1 | 0% | 0.122 | 0.472 | 0.100 | 0.572 | 0.108 | 0.053 | 0.203 | 0.255 | 0.108 | 0.053 | 0.203 | 0.255 |
| | | 25% | 0.120 | 0.448 | 0.119 | 0.567 | 0.108 | 0.060 | 0.242 | 0.301 | 0.108 | 0.060 | 0.232 | 0.292 |
| | | 50% | 0.119 | 0.460 | 0.183 | 0.644 | 0.113 | 0.113 | 0.368 | 0.481 | 0.119 | 0.164 | 0.498 | 0.662 |
| | 0.5 | 0% | 0.493 | 0.927 | 0.240 | 1.168 | 0.495 | 0.098 | 0.579 | 0.677 | 0.495 | 0.098 | 0.579 | 0.677 |
| | | 25% | 0.492 | 0.961 | 0.281 | 1.242 | 0.495 | 0.123 | 0.656 | 0.779 | 0.495 | 0.112 | 0.628 | 0.740 |
| | | 50% | 0.496 | 0.960 | 0.361 | 1.321 | 0.503 | 0.174 | 0.816 | 0.990 | 0.508 | 0.158 | 0.887 | 1.045 |
| | 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.177 | 0.185 | 0.898 | 0.008 | 0.178 | 0.185 |
| | | 25% | 0.895 | 0.082 | 0.100 | 0.182 | 0.900 | 0.007 | 0.188 | 0.195 | 0.899 | 0.007 | 0.186 | 0.192 |
| | | 50% | 0.895 | 0.082 | 0.112 | 0.194 | 0.901 | 0.008 | 0.196 | 0.204 | 0.901 | 0.007 | 0.202 | 0.209 |

Table 3.11: Mean, integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the $\widehat{S}_1(.)$ at different quantiles under Model 5 (Frank copula).

| p | Censoring Rate | CGE | | | | CCGE | | | | CCGE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $E(\widehat{S}_1(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE | $E(\widehat{S}_{1\|X}(\cdot))$ | $\text{IBias}^2$ | IVAR | IMSE |
| 0.1 | 0% | 0.122 | 0.472 | 0.101 | 0.573 | 0.108 | 0.056 | 0.203 | 0.259 | 0.108 | 0.054 | 0.203 | 0.258 |
| | 25% | 0.111 | 0.460 | 0.245 | 0.705 | 0.102 | 0.151 | 0.428 | 0.579 | 0.105 | 0.072 | 0.355 | 0.427 |
| | 50% | 0.093 | 0.589 | 0.557 | 1.146 | 0.086 | 0.375 | 0.775 | 1.150 | 0.079 | 0.259 | 0.831 | 1.091 |
| 0.5 | 0% | 0.493 | 0.928 | 0.241 | 1.169 | 0.495 | 0.102 | 0.578 | 0.680 | 0.495 | 0.100 | 0.579 | 0.679 |
| | 25% | 0.492 | 0.967 | 0.316 | 1.282 | 0.490 | 0.246 | 0.665 | 0.911 | 0.493 | 0.126 | 0.629 | 0.755 |
| | 50% | 0.494 | 0.972 | 0.459 | 1.432 | 0.489 | 0.392 | 0.813 | 1.205 | 0.490 | 0.102 | 0.958 | 1.060 |
| 0.9 | 0% | 0.894 | 0.085 | 0.091 | 0.176 | 0.898 | 0.008 | 0.176 | 0.184 | 0.898 | 0.008 | 0.176 | 0.184 |
| | 25% | 0.895 | 0.083 | 0.095 | 0.178 | 0.898 | 0.013 | 0.175 | 0.187 | 0.897 | 0.012 | 0.172 | 0.184 |
| | 50% | 0.895 | 0.082 | 0.099 | 0.181 | 0.898 | 0.016 | 0.174 | 0.190 | 0.897 | 0.012 | 0.187 | 0.199 |

Table 3.12: Integrated squared bias ($\text{IBias}^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the Kendall's tau estimates under Model 3.

| Copula Family | Censoring Rate | CGE | | | CCGE | | | CCGE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\text{IBias}^2$ | IVAR | IMSE | $\text{IBias}^2$ | IVAR | IMSE | $\text{IBias}^2$ | IVAR | IMSE |
| Frank | 0% | 2.224 | 0.276 | 2.500 | 2.238 | 0.299 | 2.538 | 0.281 | 0.914 | 1.195 |
| | 25% | 2.247 | 0.340 | 2.587 | 2.228 | 0.377 | 2.604 | 0.332 | 1.080 | 1.412 |
| | 50% | 2.256 | 0.565 | 2.821 | 2.253 | 0.619 | 2.872 | 0.580 | 1.907 | 2.487 |
| Gumbel | 0% | 2.233 | 0.328 | 2.561 | 2.225 | 0.327 | 2.552 | 0.262 | 1.216 | 1.477 |
| | 25% | 2.292 | 0.376 | 2.668 | 2.268 | 0.398 | 2.666 | 0.294 | 1.629 | 1.922 |
| | 50% | 2.358 | 0.597 | 2.955 | 2.397 | 0.636 | 3.034 | 0.789 | 1.858 | 2.647 |
| Clayton | 0% | 2.268 | 0.321 | 2.589 | 2.348 | 0.413 | 2.761 | 0.272 | 1.394 | 1.665 |
| | 25% | 2.234 | 0.404 | 2.638 | 2.283 | 0.531 | 2.814 | 0.340 | 1.328 | 1.667 |
| | 50% | 2.225 | 0.560 | 2.786 | 2.230 | 0.708 | 2.939 | 0.562 | 1.969 | 2.531 |

Table 3.13: Integrated squared bias ($IBias^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the Kendall's tau estimates under Model 4.

| Copula Family | Censoring Rate | CGE | | | CCGE | | | CCGE* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE |
| Frank | 0% | 2.225 | 0.156 | 2.381 | 2.245 | 0.183 | 2.428 | 0.081 | 1.023 | 1.194 |
| | 25% | 2.240 | 0.165 | 2.406 | 2.225 | 0.181 | 2.406 | 0.087 | 1.226 | 1.313 |
| | 50% | 2.271 | 0.293 | 2.565 | 2.251 | 0.325 | 2.576 | 0.107 | 2.366 | 2.473 |
| Gumbel | 0% | 2.225 | 0.194 | 2.419 | 2.225 | 0.218 | 2.443 | 0.205 | 0.825 | 1.030 |
| | 25% | 2.252 | 0.190 | 2.443 | 2.244 | 0.206 | 2.449 | 0.211 | 0.939 | 1.150 |
| | 50% | 2.329 | 0.324 | 2.653 | 2.335 | 0.344 | 2.679 | 0.406 | 1.644 | 2.050 |
| Clayton | 0% | 2.303 | 0.242 | 2.544 | 2.415 | 0.317 | 2.733 | 0.421 | 1.195 | 1.616 |
| | 25% | 2.248 | 0.265 | 2.513 | 2.316 | 0.349 | 2.664 | 0.454 | 1.436 | 1.890 |
| | 50% | 2.224 | 0.372 | 2.596 | 2.253 | 0.485 | 2.738 | 0.612 | 2.450 | 3.062 |

Table 3.14: Integrated squared bias ($IBias^2$), integrated variance (IVAR) and integrated mean square error (IMSE) (multiplied by 100) of the Kendall's tau estimates under Model 5 (Frank copula).

| Censoring Rate | CGE | | | CCGE | | | CCGE* | | |
|---|---|---|---|---|---|---|---|---|---|
| | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE | $IBias^2$ | IVAR | IMSE |
| 0% | 2.841 | 0.475 | 3.316 | 2.740 | 0.531 | 3.271 | 0.266 | 1.602 | 1.868 |
| 25% | 2.902 | 0.759 | 3.661 | 2.785 | 0.856 | 3.642 | 0.340 | 2.061 | 2.401 |
| 50% | 2.863 | 1.064 | 3.927 | 2.802 | 1.254 | 4.056 | 0.410 | 3.570 | 3.980 |

Figure 3.3: Mean, $5^{\text{th}}$ and $95^{\text{th}}$ quantiles of Kendall's tau estimates with no (left column), low (middle column) and moderate (right column) censoring rates for Frank (top), Gumbel (middle) and Clayton (bottom) families under Model 3. Dashed, dotted and solid line represent the mean Kendall's tau estimates, quantiles and the true Kendall's tau, respectively.

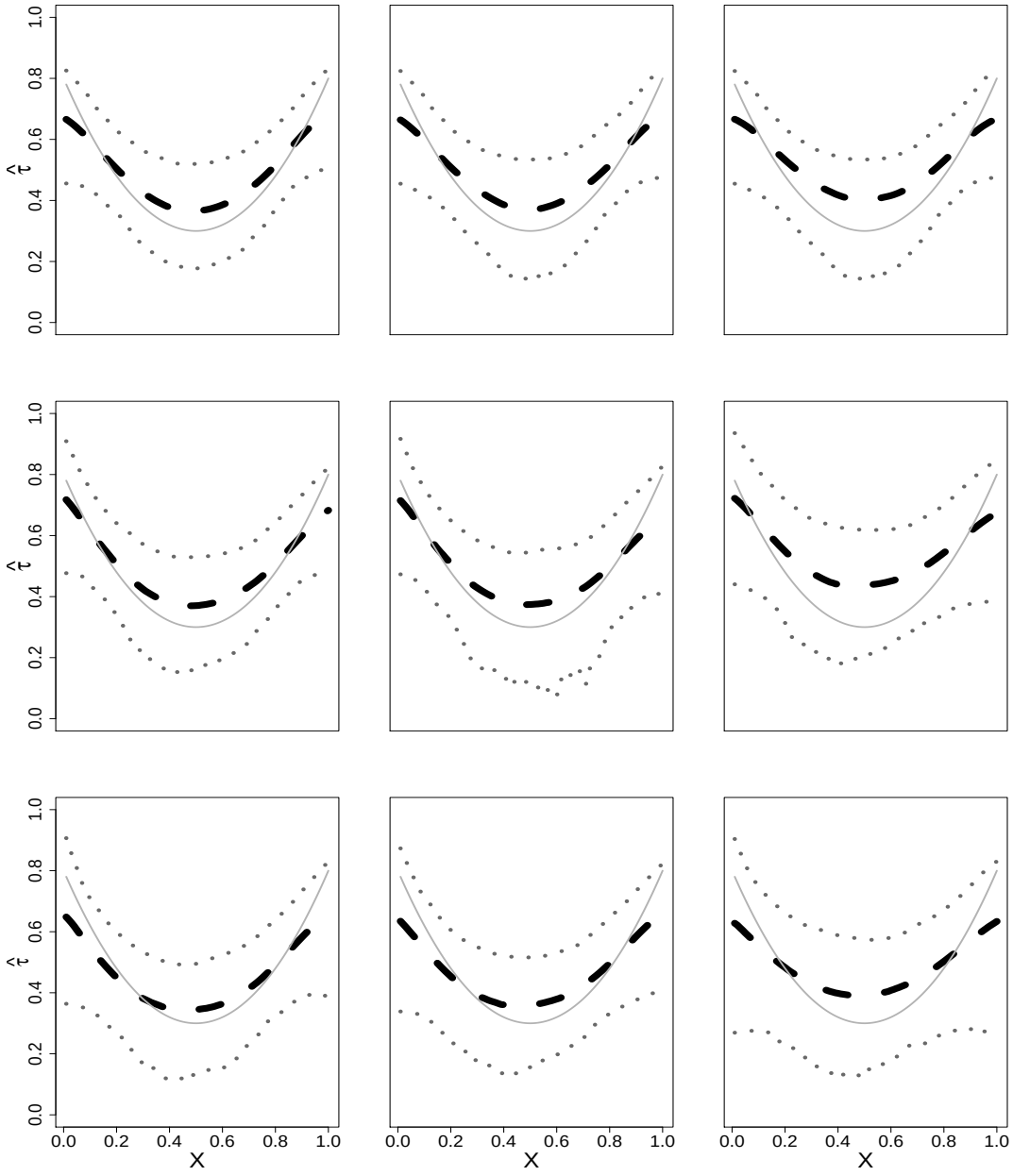


Figure 3.4: Mean, $5^{\text{th}}$ and $95^{\text{th}}$ quantiles of Kendall's tau estimates with no (left column), low (middle column) and moderate (right column) censoring rates for Frank (top), Gumbel (middle) and Clayton (bottom) families under Model 4. Dashed, dotted and solid line represent the mean Kendall's tau estimates, quantiles and the true Kendall's tau, respectively.

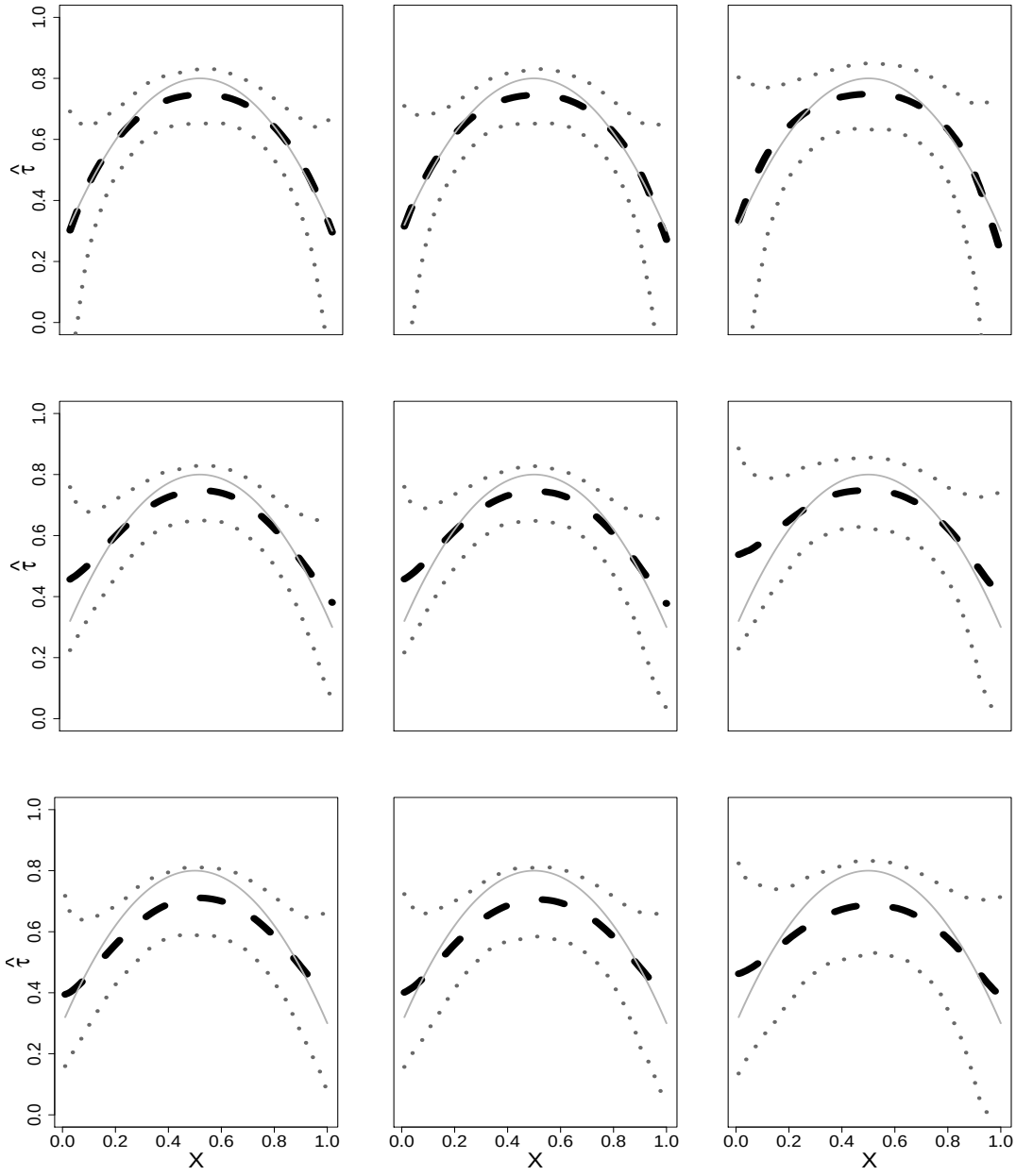
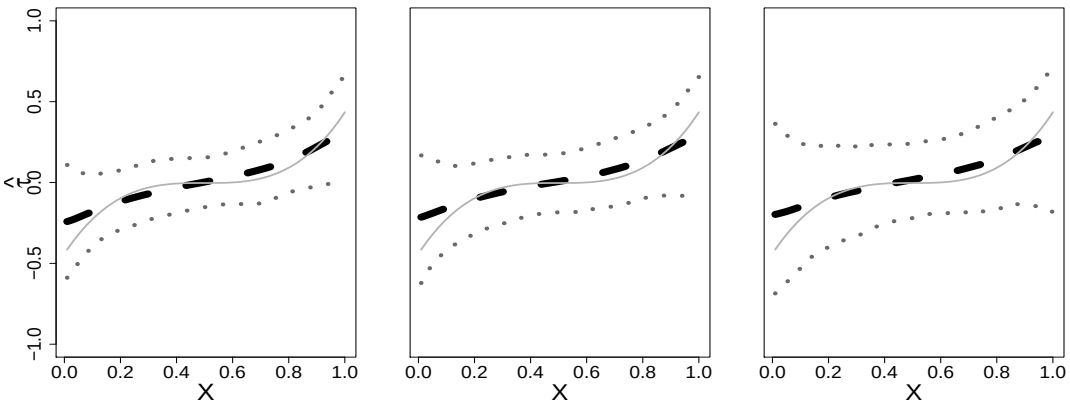


Figure 3.5: Mean, $5^{\text{th}}$ and $95^{\text{th}}$ quantiles of Kendall's tau estimates with no (left column), low (middle column) and moderate (right column) censoring rates for Frank family under Model 5. Dashed, dotted and solid line represent the mean Kendall's tau estimates, quantiles and the true Kendall's tau, respectively.



## 3.6 Summary

Our key findings of the simulation study are: (1) Even if there is no covariate effect, CCGE and the CCGE* can be used for estimating the survival probabilities or dependence parameter. (2) When there is covariate effect on margins, CGE produces biased survival estimates. If the interest is in the dependence parameter, all three estimators can be used in this scenario. (3) When there is covariate effect on both margin and dependence, CGE again yields biased survival estimates. If the interest is in survival estimates, both the CCGE and CCGE* have similar performance in this setting. However, if the interest is in the dependence parameter, both CGE and CCGE produces very high bias as a result of ignoring covariate effect on dependence. (4) Overall the efficiency of estimators deteriorates with an increase in the censoring rates. (5) The performance of estimators are usually similar under different Archimedean families and different strength of dependence.

# Chapter 4

# Data Example

In this chapter, we introduce German Breast Cancer Data, estimate the dependence parameter and the survival probability with the unconditional and conditional copula-graphic estimators and briefly discuss our findings.

German Breast Cancer data is retrieved from the University of Massachusetts, Department of Statistics website where the data is collected from a clinical trial. The study is conducted between July 1984 and December 1989. There are 16 variables on $n = 686$ subjects in the dataset. The variables are described in Table 4.1.

This dataset was analyzed in Wey et al. (2015), Hess and Levin (2014), Ambler et al. (2002), which found age at diagnosis and progesterone receptor level to be significant covariates affecting cancer recurrence. We, therefore, consider these two covariates in our analysis.

Table 4.1: List of Variables in German Breast Cancer Data.

| Variable Name | Description | Code |
|---|---|---|
| id | Study ID | 1 - 686 |
| diagdate | Date of Diagnosis | dd-mm-yyyy |
| recdate | Date of Recurrence | dd-mm-yyyy |
| deathdate | Date of Death | dd-mm-yyyy |
| age | Age at Diagnosis | Years |
| menopause | Menopausal Status | 1 = Yes, 2 = No |
| hormone | Hormone Therapy | 1 = Yes, 2 = No |
| size | Tumor Size | mm |
| grade | Tumor Grade | 1 - 3 |
| nodes | Number of Nodes Involved | 1 - 51 |
| prog_recp | Number of Progesterone Receptors | 1 - 2380 |
| estrg_recp | Number of Estrogen Receptors | 1 - 1144 |
| rectime | Time to Recurrence | Days |
| censrec | Recurrence Censoring | 0 = Censored, 1 = Recurrence |
| survtime | Time to Death | Days |
| censdead | Death Censoring | 0 = Censored, 1 = Death |

We first employed the unconditional copula-graphic estimator to analyze the data under three Archimedean copula families. The dependence parameter estimates are given in Table 4.2.

Table 4.2: Kendall's tau estimates and likelihood for different copula families.

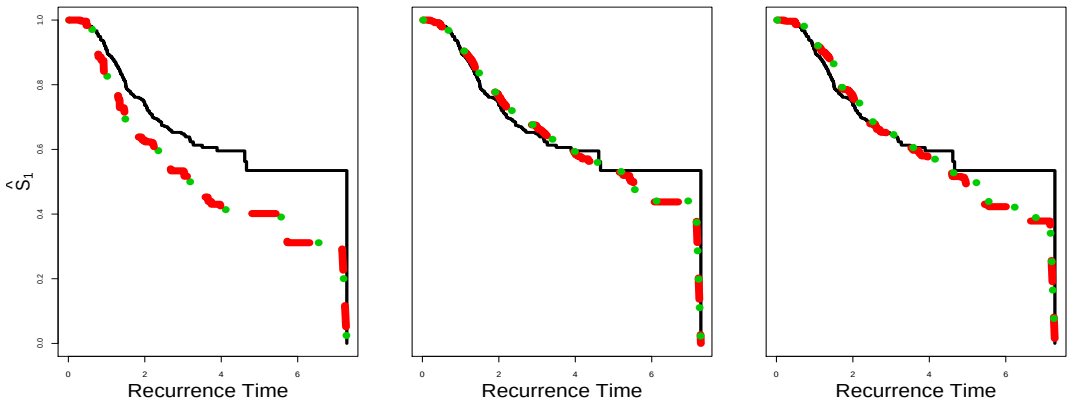
| Family | $\hat{\tau}$ | Log-likelihood |
|---|---|---|
| Frank | 0.735 | -93.143 |
| Gumbel | 0.589 | -93.207 |
| Clayton | 0.815 | -104.515 |

Since the Frank copula yielded the highest log-likelihood value, we used this family in our investigations of the covariate effects.

For the effect of age at diagnosis on the cancer recurrence time, we employed the conditional copula-graphic estimators and compared the results with those of

the unconditional copula-graphic estimator. The conditional survival function estimates are displayed in Figure (4.1) for three age values, 30, 50 and 70 at the time of diagnosis.

Figure 4.1: CGE (black), CCGE (red) and CCGE* (green) plots against the recurrence time (years) at three age groups: 30 years (left), 50 years (middle) and 70 years quantile (right).



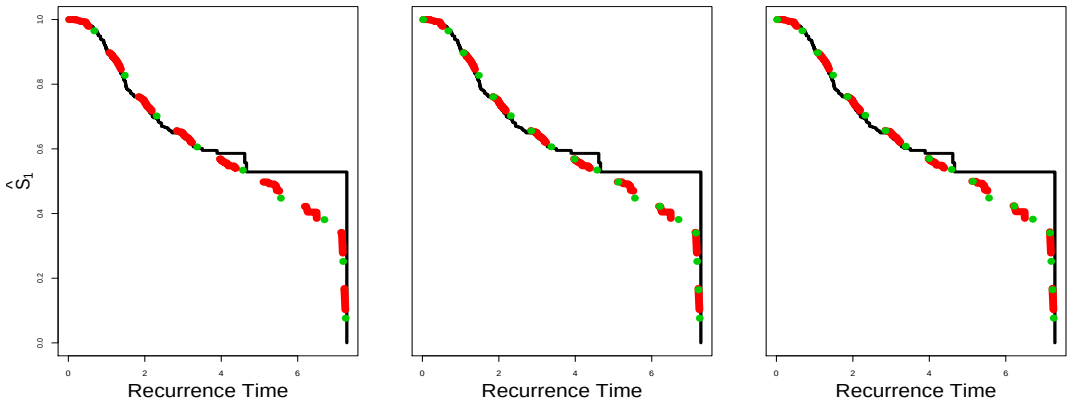

From Figure (4.1) we see that the survival plots of CCGE and CCGE* are coinciding for different age groups. We also find that CGE overestimates the survival probability for people with age 30 in comparison to CCGE or CCGE*. For people with age 50 and 70, CGE overestimates the survival probability after 5 years of study. Given the difference between the unconditional and conditional graphic estimators, we can conclude that the age at diagnosis affects the cancer recurrence time, with lower probability of recurrence at younger ages.

For exploring the effect of progesterone receptor level on the cancer recurrence time, we consider the first, second and third quantile values of progesterone receptor level to display the conditional copula-graphic estimators graphically

in Figure (4.2).

From Figure (4.2), we see that the survival plots at the three quantiles of the progesterone receptor level are all very similar. While the three estimators coincide in their survival probability estimates till the first four years, CGE overestimates the survival probability after four years. Since the discrepancy between the unconditional and conditional copula-graphic estimates is smaller for the progesterone receptor level, we conclude that this covariate is not as important as the age at diagnosis.

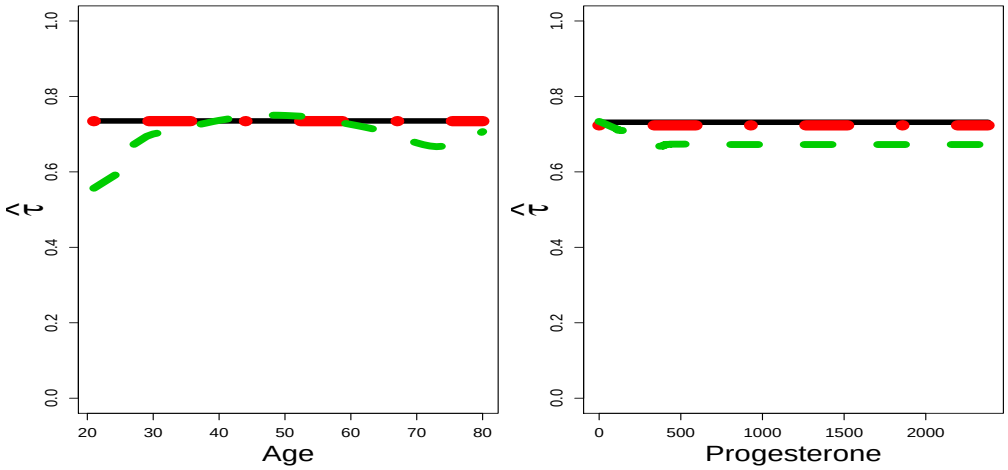
Figure 4.2: CGE(black), CCGE (red) and CCGE* (green) plots against the recurrence time (years) at three Cancer progesterone receptor groups: $1^{st}$ quantile (left), $2^{nd}$ (middle) and $3^{rd}$ quantile (right).




To see whether the age at diagnosis and progesterone receptor level affect the strength of dependence between the cancer recurrence time and the time to death, we compare the Kendall's tau estimates of the three estimators in Figure (4.3). While CGE and CCGE yield similar constant Kendall's tau

estimates, there is a slight variation in the Kendall's tau estimates of CCGE* across values of each covariate.

Figure 4.3: Kendall's tau estimates of CGE (black), CCGE (red) and CCGE* (green) at different age at diagnosis (left) and progesterone receptor levels (right).




From Figure (4.3) we see that CGE and CCGE overestimate the dependence for age groups below 30 and above 65 and underestimate the dependence for progesterone receptor level between 5000 and 1500. Nevertheless, the differences are small suggesting that the dependence structure does not drastically change with these covariates.

The data application suggests that both age at diagnosis and progesterone receptor level have effects on survival after four years of study period. These covariates have negligible effects on the dependence structure, except for certain groups of people.

# Chapter 5

# Conclusion

In situations of dependent censoring, the copula-graphic estimator is commonly used to estimate marginal survival function by accounting for the dependence structure. In this thesis, we proposed two extended versions of the copula-graphic estimator: CCGE and CCGE* to further for covariate effects. The performance of our proposed estimators were investigated in a simulation study and compared to that of the unconditional CGE. We considered different copula families and dependent parameter models in our evaluations.

First, we considered the unconditional setting and concluded that our proposed estimators (CCGE, CCGE*) perform well, though with a loss in efficiency, when there is no covariate effect. Next, we considered the setting where the covariate affects only the margins. In this setting, both conditional copula-graphic estimators (CCGE, CCGE*) showed better performance than the unconditional version (CGE). Finally, we considered the setting where the covariate

affects both the margins and the dependence. We noticed if interest is the effect of covariate on the marginal survival function of the non-terminal event, both versions of the conditional copula-graphic estimator (CCGE, CCGE*) performed better in comparison to CGE. When interest is the effect of covariate on dependence, CCGE* accurately captures the underlying effect, while CCGE and CGE may fail to do so, in cases where the dependence parameter changes with the covariate.

We also demonstrated the performance of our proposed estimators using a real data example. We considered the age at diagnosis and the progesterone receptor level as covariates with potential effects on the cancer recurrence time and the lifetime of cancer patients. We found that failing to account for the age at diagnosis would yield overestimation of the survival probability, especially for younger patients. Ignoring the progesterone receptor level may also result in overestimation of the survival probability for some groups of patients.

The proposed conditional copula-graphic estimators in this thesis cannot accommodate more than one covariate, due to the additional complexity in accounting for these covariates in Beran's estimator and in the local likelihood estimation of the conditional copula parameter. Future research is needed to extend the proposed estimators to settings involving two or more covariates.

# Bibliography

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics*, 67(2):445–453.

Ambler, G., Brady, A. R., and Royston, P. (2002). Simplifying a prognostic model: a simulation study based on clinical data. *Statistics in Medicine*, 21(24):3803–3822.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Technical Report, University of California, Berkeley.

Braekers, R. and Veraverbeke, N. (2005). A copula-graphic estimator for the conditional survival function under dependent censoring. *Canadian Journal of Statistics*, 33(3):429–447.

Cai, Z. (1998). Asymptotic properties of kaplan-meier estimator for censored dependent data. *Statistics & Probability Letters*, 37(4):381–389.

Chen, Y.-H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime Data Analysis*, 18(1):36–57.

Crowder, M. J. (2012). *Multivariate survival analysis and competing risks*. Chapman and Hall/CRC.

Embrechts, P. (2009). Copulas: A personal view. *Journal of Risk and Insurance*, 76(3):639–650.

Embrechts, P., Lindskog, F., and McNeil, A. (2001). Modelling dependence with copulas. *Rapport Technique, Département de Mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2(1):1–25.

Geerdens, C., Acar, E. F., and Janssen, P. (2017). Conditional copula models for right-censored clustered event time data. *Biostatistics*.

Ghosh, D. (2006). Semiparametric inferences for association with semi-competing risks data. *Statistics in Medicine*, 25(12):2059–2070.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized kaplan-meier estimator with some applications. *Communications in Statistics-Theory and Methods*, 4(1):65–78.

Hess, K. R. and Levin, V. A. (2014). Getting more out of survival data by using the hazard function. *Clinical Cancer Research*.

Heuchenne, C., Laurent, S., Legrand, C., and Keilegom, I. V. (2014). Likelihood-based inference for semi-competing risks. *Communications in Statistics-Simulation and Computation*, 43(5):1112–1132.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2014). Copula: multivariate dependence with copulas. *R Package Version 0.999-9, URL http://CRAN. R-project. org/package= copula*.

Hsieh, J.-J. and Huang, Y.-T. (2012). Regression analysis based on conditional likelihood approach under semi-competing risks data. *Lifetime Data Analysis*, 18(3):302–320.

Hsieh, J.-J., Wang, W., and Adam Ding, A. (2008). Regression analysis based on semicompeting risks data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):3–20.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Klein, J. P. and Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Krupskii, P. and Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101.

Krupskii, P. and Joe, H. (2015). Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis*, 138:53–73.

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

Laurent, S. (2013). Estimating the survival functions in a censored semi-competing risks model. *Sankhya A*, 75(2):231–252.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.

Li, R. and Cheng, Y. (2016). Flexible association modelling and prediction with semi-competing risks data. *Canadian Journal of Statistics*, 44(3):361–374.

Li, R. and Peng, L. (2015). Quantile regression adjusting for dependent censoring from semicompeting risks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):107–130.

Peng, L. and Fine, J. P. (2007). Regression modeling of semicompeting risks data. *Biometrics*, 63(1):96–108.

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

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T., et al. (2012). Vinecopula: statistical inference of vine copulas. *R Package Version*, 1.

Schmoor, C., Olschewski, M., and Schumacher, M. (1996). Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies. *Statistics in Medicine*, 15(3):263–271.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications de Institut de Statistique de Université de Paris*, 8:229–231.

Uyttendaele, N. and Mazo, G. (2016). Extending one-factor copulas. *arXiv preprint arXiv:1612.02848*.

Wey, A., Connett, J., and Rudser, K. (2015). Combining parametric, semi-parametric, and non-parametric survival models with stacked survival models. *Biostatistics*, 16(3):537–549.

Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*, 66(3):716–725.

Yang, J. and Peng, L. (2016). A new flexible dependence measure for semi-competing risks. *Biometrics*, 72(3):770–779.

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