Semi-parametric Estimation of the Regression Function for Right-Censored Data

Charles Fontaine* Taoufik Bouezmarni†
Université de Sherbrooke Université de Sherbrooke

October 22, 2014

Abstract

This article discusses the semi-parametric estimation of the regression function in situation of right-censored data. We present a new copula-based method to estimate the regression function. The key concept presented in this paper is to write the regression function in terms of copula and marginal distributions, and to substitute the marginal distribution of the response variable which is right-censored by an estimated distribution based on the product-limit estimator. We present the assumptions to be able to use our model and then, we derive asymptotic properties of our estimator and extend it to the multivariate case. We also provide a new method to avoid any problem of copula misspecification when the data come from an unknown distribution and, compare the performance of our estimator by simulations. Furthermore, we present two applications of the proposed technique: one to an heart transplant data set and another one to an insurance company losses and expenses for which our estimator gives good results.

Key words: *Semi-parametric estimation, Regression function, Censored data, Parametric copula models, Multidimensional covariates, Kaplan-Meier estimator.*

1 Introduction

Right-censoring on the response variable arises when the followed subjects in a study cannot provide us information on the goal of the study (e.g. voluntary retraction, moving, etc.) in

*corresponding autor
†T. Bouezmarni acknowledges financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC).

despite of our knowledge of the predictors for these subjects (e.g. age, time of follow, etc). Therefore, these failure time data lead researchers to infer these data. As a scientific study looks for a relation between these predictors and the response variable, let assume that there exists a dependence between these variables. Thus, given $X = (X_1, X_2, ..., X_d)^T$ the vector of d covariates for which we note the marginal distributions as $F(x) = (F_1(x_1), F_2(x_2), ..., F_d(x_d))$, and Y the response variable with a marginal distribution $F(y)$, it is obvious that the joint distribution of these variables $F(x, y)$ will be something else than the product of the marginal distributions and, using Sklar's theorem (Sklar, 1959), we can write that joint distribution in terms of copula.

In this paper, we will propose a new method to infer the regression function in case of right-censored data on the response variable. Our method is based on the the copula regression function estimator proposed by (Noh, El Ghouch, and Bouezmarni, 2013) for complete data for which we propose to estimate the marginal distribution $F(Y)$ at a value y by a well-known survival estimator to get an easy and robust implementation for any user who wants to use our method: the product-limit estimator (Kaplan-Meier estimator); denoted by $\Gamma_T(t)$.

In the literature, there exists a lot of techniques to evaluate that kind of right-censored regression function. At first, (Miller, 1976) and (Koul, Susarla, and Van Ryzin, 1981) propose estimators based on the standard linear model with some methodological weaknesses since that they both require a particular censoring pattern. Furthermore, the method proposed by (Buckley and James, 1979), which is also based on the the general linear model is not accurate since that it is an iterative sequence of estimators that can fail to converge. More recently, (Fan and Gijbels, 1994) proposed an estimator based on local linear regression. Since that (Noh, El Ghouch, and Bouezmarni, 2013) get better results for complete data than local linear methods, we propose to use that semi-parametric method.

Furthermore, to estimate the copula, as (Noh et al., 2013) did, since that the non-parametric approach requires to find a bandwidth parameter which can be very thoughtful to find, we use a semi-parametric approach (i.e. the copula is parametric but the margins are non-parametric). Therefore, it reduces the problem to the research of a copula parameter; which can be found from the Kendall tau inference on data.

2 Proposed estimator

Let Y, the response variable, a stochastic variable which has a distribution function F_Y and a density function f_Y; and $X = (X_1, ..., X_d)$, the predictive variable composed of d predictors, which has a distribution function $F_X(x) = (F_1(x_1), F_2(x_2), ..., F_d(x_d)), x = (x_1, ..., x_d)$ and a density function f_X. From the copula theory, it is known that the joint distribution $(Y, X)^T$ at any point $(y, x) \in \mathbb{R} \times \mathbb{R}^d$ is given by $C(F_0(y), F_X(x))$ where C is the copula distribution of (Y, X). Thus, for $(u_0, u_1, ..., u_d) \equiv (u_0, u) \in [0, 1] \times [0, 1]^d$ we have

$$C(u_0, u) = \mathbb{P}(U_0 \leq u_0, U_1 \leq u_1, ..., U_d \leq u_d)$$

where $U_0 = F_0(Y), U_d = F_d(X_d)$ for $j = 1, ..., d$. Hence, C is a distribution function with uniform margins on $[0, 1]$.

In the situation of complete data for all the variables, Noh et al. propose to estimate the regression function using the conditionality on variables. From the conditional density theory written in terms of densities of copulas (i.e. in terms of partial derivatives of the distribution of copula with respect to the margins) as:

$$f_{Y|X=x} = f_Y(y)\frac{c(F_0(y), F_X(x))}{\tilde{c}(F_X(x))}.$$

where $\tilde{c}$ is the copula density of X. Thus, they propose to rewrite the regression function, from its definition, as:

$$\begin{aligned} m(x) &= \mathbb{E}(Y|X = x) \\ &= \mathbb{E}(Yw(F_0(y), F_X(x))) \\ &= \frac{e(F_X(x))}{\tilde{c}(F_X(x))} \end{aligned} \tag{2.1}$$

where

$$w(u_0, u) = c(u_0, u)/\tilde{c}(u) \quad \text{and} \quad e(u) = \int_0^1 F_0^{-1}(u_0)c(u_0, u)du_0$$

with c and $\tilde{c}$ are the copula densities of (Y, X) and X, $c(u_0, u) = \frac{\partial^{d+1}C(u_0,u_1,u_2,...,u_d)}{\partial u_0 \partial u_1 \partial u_2 ... \partial u_d}$ and $\tilde{c}(u) = \frac{\partial^d C(1,u_1,u_2,...,u_d)}{\partial u_1 \partial u_2 ... \partial u_d}$; thus $w(u_0, u)$ is just the ratio of $c(u_0, u)$ over $\tilde{c}(u)$. Note that in the case where $d = 1$ or the case where the predictors are mutually independents, we have $\tilde{c} = 1$ and the regression function in 2.1 reduces to $e(F_X(x))$.

In this paper, the estimator that we propose adapts that result to right-censored data on the dependent variable. Let denote the observed vector of responses Z as $Z = \min(Y, C_0)$ where C_0 is a stochastic variable representing the right-censoring phenomena, and let denote the data completeness indicator such that $\delta = \mathbb{1}(Z \leq C_0)$. If, for example, a user observe only complete data, then $Z = Y$ and $\delta_j = 1$ for $j = 1, ..., n$ where n is the number of observations. From formula 2.1, we have to estimate the marginal distributions and the copula density function.

Our method uses the Kaplan-Meier estimator, Γ_n, of F_0 from which we deduce the Kaplan-Meier weights w, applied on the observed values Z_i to infer the marginal distribution F_0 :

$$\Gamma_n(z) = \begin{cases} 1 - \prod_{i:1 \leq i \leq n, Z_{(i)} \leq z} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & \text{if } z < Z_{(n)} \\ \\ 1 & \text{Otherwise.} \end{cases}$$

We also consider that all the explicative variables are uncensored then we use a rescaled empirical distribution function for marginal distribution of the predictors, F_j : for $j = 1, ..., d$:

$$\hat{F}_{j,n}(x_j) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x_j).$$

But, if the covariates are a mix of complete and censored random variables, we suggest to consider the empirical distribution for the marginal distributions of the uncensored variables and Kaplan-Meier estimator for the marginal distributions of the censored variables.

To estimate the copula density c, we suppose a parametric family , $c(\cdot, \cdot, \theta)$ and then one estimate the parameters vector $\theta \in \mathbb{R}^p$. Thus, $c(\cdot, \cdot, \theta)$ is estimated by $c(\cdot, \cdot, \hat{\theta}_n)$. Therefore, we propose the semi-parametric regression estimator :

$$\begin{aligned} \hat{m}(x) &= \int y c(\Gamma_n(z), \hat{F}_{1,n}(x_1), ..., \hat{F}_{d,n}(x_d), \hat{\theta}_n) d\Gamma_n(z) \\ &= \sum_{i=1}^{n} Z_i w_i c(\Gamma_n(Z_i), \hat{F}_{1,n}(x_1), ..., \hat{F}_{d,n}(x_d), \hat{\theta}_n). \end{aligned}$$

where $\omega_1 = \Gamma_n(Z_{(1)})$ and $\omega_i = \Gamma_n(Z_{(i)}) - \Gamma_n(Z_{(i-1)})$ for $i = 2, ..., n$

Remarks 2.1. *We can do the following assumption for the case where we use any other estimator than the empirical cumulative distribution function or the Kaplan-Meier estimator to evaluate F_0 and $F_{\mathbf{X}}$ respectively.*

- *Let $\tilde{F}(x_i)$ denote the estimator used for the covariates and $\tilde{\Gamma}_n(y)$ the one for the response variable. Therefore, the estimators have to satisfy that, for any point x where we want to estimate the regression function such that $x \in \mathbb{R}^d$, the estimators are such that:*
  - $\tilde{F}(x_i) = \hat{F}_{i,n}(x) + o_p(n^{-1/2})$ *for* $i = 1, ..., d$
  - $\tilde{\Gamma}_n(y) = \Gamma_n(y) + o_p(n^{-1/2})$.

3 Theoretical framework

In this section, we provide two asymptotic results. The first establishes the uniform weak consistency of the proposed estimator and the second gives his i.i.d. representation. From the last results, we deduce the asymptotic normality of the estimator. We calculate his asymptotic variance and suggest a practical estimator of the asymptotic variance. In the next section, we present the assumptions and some notations needed for the main results.

3.1 Assumptions on copula and its parameters

Let us provide some preliminary assumptions to use our main result in a good theoretical framework. Moreover, we have to make assumptions on the copula parameter and for that, we will generalize some results in the case of bidimensional covariates for multidimensional covariates. Thus, $\hat{\theta}_n$ has to satisfy the following:

Assumption A:

$$\hat{\theta}_n - \theta_0 = \frac{1}{n}\sum_{i=1}^{n} \zeta_i + o_p(n^{-1/2})$$

where ζ_i are random i.i.d. variables with zero mean and finite variance.

In this paper, we consider the estimation of θ found by (Shih and Louis, 1995). Let remember that in their article, they proposed a two-stage parametric maximum likelihood estimation method in a censored data scheme; which is a special case of the IFM (inference function for margins) estimation method developed in (Joe and Xu, 1996). The validity of this method holds since that the limiting distribution of $\hat{\theta}_n$ is $\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(0, G^{-1}(\theta_0))$ where G is the Godambe information matrix. Therefore, they defined, when $d = 1$ and for two censored random variables a result which is easily extendable to $d > 1$, ζ_i such that

$$\zeta_i = W(\theta_0, F_0(Y_i), F_1(X_{1i})) + I_{0i} + I_{1i}$$

for which

$$W(\theta_0, F_0(Y_i), F_1(X_{1i})) = \frac{\partial l(\theta_0, F_0(Y_i), F_1(X_{1i}))}{\partial \theta_0}$$

where l is the log of the likelihood of θ_0 and where, for $i = 1, ..., n$, I_{0i}, and I_{1i} are defined by the following:

$$I_{0i} \equiv I_0(Y_i, \delta_{0i}, \theta_0) = \int_A \frac{\partial^2 l(\theta_0, F_0(y), F_1(x_1))}{\partial \theta_0 \partial F_0(y)} I_0^0(Y_i, \delta_{0i})(y) dH_{\theta_0}(y, x_1)$$

and

$$I_{1i} \equiv I_1(X_{1i}, \delta_{1i}, \theta_0) = \int_A \frac{\partial^2 l(\theta_0, F_0(y), F_1(x_1))}{\partial \theta_0 \partial F_1(x_1)} I_1^0(X_{1i}, \delta_{1i})(y) dH_{\theta_0}(y, x_1)$$

for which H_{θ_0} is the joint distribution of $(Y_i, \delta_{0i}), (X_{1i}, \delta_{1i})$ where δ is the censoring indicator vector which will be equal to one for all the explicative variables, i.e. $\delta_1 = 1$, and $A = [0, t_{00}] \times [0, t_{01}]$ with $t_{00} = \sup\{t : \mathbb{P}(Y > t, C_0 > 0) > 0\}$, and $t_{01} = \sup\{t : \mathbb{P}(X_1 > t, C_1 > 0) > 0\}$. Furthermore, $I_0^0(Y, \delta_{0i})(y)$ is defined as follows:

$$I_0^0(Y, \delta_{0i})(y) = -(1 - F_0(y)) \left[\int_0^y \frac{1}{\mathbb{P}(Y \geq u, C_0 \geq u)} dN_{0i}(u) - \int_0^y \frac{\mathbb{1}(Y \geq u)}{\mathbb{P}(Y \geq u, C_0 \geq u)} d\Lambda_0(u) \right]$$

where C is the distribution function of the censoring times, $N_{0i}(u) = \mathbb{1}(Y_i \leq u, \delta_{0i} = 1)$ and $\Lambda_0 = -\log F_0(Y)$. We can do the same for the covariate X_1. We can extend that result to a distribution function of d covariates, by a similar approach, and show that

$$\zeta_i = W(\theta_0, F_0(Y_i), F_1(X_{1i})) + \sum_{j=0}^{d} I_{ji}$$

where I_{ji} is defined in a the similar way as in the case with one covariate.

To make assumptions on the copula, we have to use the following notation from Noh et al.:

- $e(\mathbf{u}) = \mathbb{E}(Yc(F_0(Y), \mathbf{u})) = \int_0^1 F_0^{-1}(u_0)c(u_0, \mathbf{u})du_0$, $\mathbf{u} = (u_1, ..., u_d)$.
- $\partial_j c = \frac{\partial c}{\partial u_j}$ for $j = 0, 1, ..., d$ and $\dot{\mathbf{c}} = \left(\frac{\partial c}{\partial \theta_1}, ..., \frac{\partial c}{\partial \theta_p}\right)^T$.

This leads us to the following assumptions:

Assumption B:

Let g be a function of $\dot{\mathbf{c}}$ or $\partial_j c$, $j = 0, 1, ..., d$ and $x \in \mathbb{R}^d$ be a given point of interest such that $F(\mathbf{x}) \in (0, 1)^d$. Thus:

- $(\mathbf{u}, \theta) \to g_{u_0}(\mathbf{u}, \theta) \equiv g(u_0, \mathbf{u}; \theta)$is continuous at $(F(\mathbf{x}), \theta_0)$ uniformly in $u_0 \in [0, 1]$
- $u_0 \to g(u_0, F(\mathbf{x}); \theta_0)$ is continuous in [0,1].

3.2 Main results

We start by considering the case of only one covariate, X_1. We prove, in Proposition 3.1, the uniform convergence in probabilities of the proposed estimator, provide his i.i.d. representation, in Theorem 3.1, and then, in Proposition 3.2, give the asymptotic normality of $\hat{m}$.

Proposition 3.1. *Under assumptions A and B we have*

$$\sup_{x_1} (|\hat{m}(x_1) - m(x_1)|) \quad \overset{P}{\longrightarrow} \quad 0.$$

The proof of Proposition 3.1 is shown in the appendix.

Theorem 3.1. *Under assumptions A and B, $\hat{m}$ admits the following i.i.d. representation:*

$$\hat{m}(x_1)(F(x_1)) = m(x_1) + \frac{1}{n}\sum_{i=1}^{n} \eta_i(F(x_1)) + o_p(n^{-1/2}) \tag{3.2}$$

where, $\eta_i = \sum_{j=1}^4 \eta_{i,j}(F(x_1))$, *with*

$$\eta_{i,1}(F(x_1)) = \int_0^{\tau} y\xi(Z_i, \delta_i, y)\partial c_0(F_0(y), F_1(x_1), \theta_0)dF_0(y),$$

$$\eta_{i,2}(F(x_1)) = (I(X_{1i} \leq x_1) - F_1(x_1)) \int_0^{\tau} y\partial c_1(F_0(y), F_1(x_1), \theta_0)dF_0(y),$$

$$\eta_{i,3}(F(x_1)) = \int_0^\tau y\zeta_i \dot{\boldsymbol{c}}(F_0(y), F_1(x_1), \theta_0) dF_0(y),$$

and

$$\eta_{i,4}(F(x_1)) = \int_0^\tau \xi(Z_i, \delta_i, y) d(yc(F_0(y), F_1(x_1), \theta_0)).$$

Proposition 3.2. *Under assumptions A, B and C, the asymptotic normality of $\hat{m}(x)$ is such that:*

$$\sqrt{n}\,(\hat{m}(x_1) - m(x_1)) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = \mathbb{E}(\eta_1^2)$.

3.3 Extensions to the multivariate case

Now, for multivariate case, $(d \geq 2)$, from 2.1the regression function is given by

$$m(\boldsymbol{x}) = \frac{e(\boldsymbol{F}(\boldsymbol{x})}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x})}. \tag{3.3}$$

We suppose a parametric model for the numerator, $e(\boldsymbol{F}(\boldsymbol{x}); \boldsymbol{\theta}_0)$, and the estimation can be done as in the single covariate case by

$$\hat{e}(\hat{\boldsymbol{F}}(\boldsymbol{x})) = \sum_{i=1}^{n} Z_i w_i c(\Gamma_n(Z_i), \hat{F}_{1,n}(x_1), ..., \hat{F}_{d,n}(x_d), \hat{\theta}_n)$$

where $\hat{\boldsymbol{F}}(\boldsymbol{x}) = (\hat{F}_1(x_1), \ldots, \hat{F}_d(x_d))$.

Following the proof of Theorem 3.1, one can easily check that

$$\hat{e}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) - e(\boldsymbol{F}(\boldsymbol{x})) = n^{-1} \sum_{i=1}^{n} \varphi_i(\boldsymbol{F}(\boldsymbol{x})) + o_p(n^{-1/2}), \tag{3.4}$$

where $\varphi_i = \sum_{j=1}^4 \varphi_{i,j}$ with

$$\varphi_{i,1}(\boldsymbol{F}(\boldsymbol{x})) = \int_0^\tau y\xi(Z_i, \delta_i, y) \partial c_0(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0) dF_0(y),$$

$$\varphi_{i,2}(\boldsymbol{F}(\boldsymbol{x})) = \sum_{j=1}^{d}(I(X_{ji} \leq x_1) - F_1(x_1)) \int_0^\tau y\partial c_j(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0)dF_0(y),$$

$$\varphi_{i,3}(\boldsymbol{F}(\boldsymbol{x})) = \int_0^\tau y\zeta_i^T \dot{\mathbf{c}}(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0)dF_0(y),$$

and

$$\varphi_{i,4}(\boldsymbol{F}(\boldsymbol{x})) = \int_0^\tau \xi(Z_i, \delta_i, y)d(yc(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0)).$$

Using the fact that $c_{\boldsymbol{X}}(\boldsymbol{u}) = \mathbb{E}[c(F_0(Y), \boldsymbol{u})]$, we propose to estimate $c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))$ by

$$\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) = \sum_{i=1}^{n} w_i c(\Gamma_n(Z_i), \hat{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}}).$$

Thus, our estimator of $m(\boldsymbol{x})$ is given by

$$\hat{m}(\boldsymbol{x}) = \frac{\hat{e}(\hat{\boldsymbol{F}}(\boldsymbol{x}))}{\hat{c}_{\boldsymbol{X}}(\hat{\boldsymbol{F}}(\boldsymbol{x}))} = \sum_{i=1}^{n} Y_i \frac{w_i c(\hat{F}_0(Y_i), \hat{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}})}{\sum_{i=1}^{n} w_i c(\hat{F}_0(Y_i), \hat{\boldsymbol{F}}(\boldsymbol{x}); \hat{\boldsymbol{\theta}})}.$$

The asymptotic representation of $\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x}))$ follows by using similar arguments as in the proof of Theorem 3.1. In fact, one can easily check that

$$\hat{c}_{\boldsymbol{X}}(\tilde{\boldsymbol{F}}(\boldsymbol{x})) - c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x})) = n^{-1}\sum_{i=1}^{n} \phi_i + o_p(n^{-1/2}), \tag{3.5}$$

where $\phi_i(\boldsymbol{F}(\boldsymbol{x})) = \sum_{j=1}^{4} \phi_{i,j}(\boldsymbol{F}(\boldsymbol{x}))$ with

$$\phi_{i,1}(\boldsymbol{F}(\boldsymbol{x})) = \int_0^\tau y\xi(Z_i, \delta_i, y)\partial c_0(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0)dF_0(y),$$

$$\phi_{i,2}(\boldsymbol{F}(\boldsymbol{x})) = \sum_{j=1}^{d}(I(X_{ji} \leq x_1) - F_1(x_1)) \int_0^\tau y\partial c_j(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0)dF_0(y),$$

$$\phi_{i,3}(\boldsymbol{F}(\boldsymbol{x})) = \int_0^\tau y\zeta_i^T \dot{\mathbf{c}}(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0)dF_0(y),$$

and

$$\phi_{i,4}(\boldsymbol{F}(\boldsymbol{x})) = \int_0^\tau \xi(Z_i, \delta_i, y) d(yc(F_0(y), \hat{\boldsymbol{F}}(\boldsymbol{x}), \theta_0)).$$

Combining (3.4) with (3.5) leads to our main result.

Theorem 3.1. *Under Assumption C, if $\tilde{F}$ satisfies Assumption A and $\hat{\boldsymbol{\theta}}$ satisfies Assumption B, then we have*

$$\hat{m}(\boldsymbol{x}) - m(\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} \frac{1}{c_{\boldsymbol{X}}(\boldsymbol{F}(\boldsymbol{x}))} \left[\varphi_i(\boldsymbol{F}(\boldsymbol{x})) - m(\boldsymbol{x})\phi_i(\boldsymbol{F}(\boldsymbol{x}))\right] + o_p(n^{-1/2}).$$

4 Choice of copula

When the copula family is given and specified, the proposed estimator gives results as close of the true regression function as the results shown in this paper. In other case, we say that the copula parameter is misspecified and it can leads us to poor results. Because, in practice, the copula parameter is often unknown, we propose here a technique that works only in the case of parametric copulas to avoid the problems of misspecification.

Let consider $c_1, c_2, ..., c_k$ k parametric copula models. For example, c_1 can be a Gaussian copula, c_2 a Gumbel copula and so on. Furthermore, let consider the regression estimator

$$\hat{m}^{c_j}(x) = \sum_{i=1}^{n} y_i c_j(\Gamma_n(y_i), \hat{F}_{i,n}(x_i), \hat{\theta}) w(y)$$

for which j lies between 1 and k. Therefore, we propose to compute a difference function T for the k possible copulas for a given situation such that

$$T^{c_j} = \sum_{i=1}^{n} (y_i - \hat{m}^{c_j}(x_i))^2 w(y_i),\ \ j = 1, 2, ..., k.$$

Thus, the copula model which give us $\min(T^{c_1}, T^{c_2}, ..., T^{c_k})$ is the one to choose to get a consistent estimator of the regression function.

5 Simulations

In this section, we show that our copula selection criteria holds when censoring arises, and then we show that the performance of our estimator on some DGPs.

5.1 Copula selection criteria

Let consider the simulation scheme presented in table 1. We generated 200 data with 5 different copulas. For each copula, we used the same Kendall tau, which is 0.80, to generate data. For each copula, we generate three level of censoring: 0 %, 25% and 32% and replicate the simulation 500 times to find the estimator that minimize the value of T^{c_j}. The censoring was uniformly distributed.

As it can be seen in table 1, at a censoring level of 0 and 25%, the criteria to avoid misspecification holds; except for the Student copula, which can be explained by the similarity between the Gaussian and the Student copula. Therefore, that misspecification is not really a problem here. For a censoring level of 32%, the same issue happens for the Student copula. Furthermore, for the Frank copula, the copula that minimize the lower value of T^c is given by the Clayton copula and the Frank copula get the second position. The censoring on data probably explain this misspecification. Therefore, we assume the validity of our criteria to avoid misspecification until a censoring level around 30%.

5.2 Performance of the proposed estimator

In this section, we will perform 2 data generating process with different censoring levels. The censoring will be generated using an uniform distribution and will be around 30 percents. The DGPs chosen for this paper are the following:

- **DGP 1:** $(F_Y(Y), F_X(X))$ follows a Gaussian copula with parameter $\rho = 0.8$; $Y \sim (N)(\mu_Y = 0, \sigma_Y^2 = 1)$

  1. The resulting regression function is $m(x) = 0.8\Phi + 1(F_X(x))$ where Φ is the cdf on a $\mathcal{N}(0, 1)$ distribution
  2. X is generated from $\mathcal{N}(0, 1)$.

Estimator \ Generator		Gaussian	Clayton	Gumbel	Frank	Student
Gaussian	$c = 0\%$	**1.671**	1.630	1.703	1.459	1.776
		$\sigma = 0.285$	$\sigma = 0.283$	$\sigma = 0.301$	$\sigma = 0.145$	$\sigma = 0.347$
	$c = 25\%$	**5.449**	5.474	5.924	5.345	**5.485**
		$\sigma = 0.864$	$\sigma = 1.028$	$\sigma = 0.797$	$\sigma = 0.971$	$\sigma = 0.891$
	$c = 32\%$	**5.585**	5.755	6.170	6.748	**5.710**
		$\sigma = 0.944$	$\sigma = 1.030$	$\sigma = 0.932$	$\sigma = 0.824$	$\sigma = 0.946$
Clayton	$c = 0\%$	1.767	**1.593**	1.820	1.436	1.877
		$\sigma = 0.290$	$\sigma = 0.278$	$\sigma = 0.320$	$\sigma = 0.165$	$\sigma = 0.359$
	$c = 25\%$	5.827	**5.114**	5.560	5.865	5.888
		$\sigma = 0.735$	$\sigma = 0.841$	$\sigma = 0.671$	$\sigma = 0.719$	$\sigma = 0.757$
	$c = 32\%$	6.065	**5.318**	6.910	6.820	6.177
		$\sigma = 0.767$	$\sigma = 0.829$	$\sigma = 0.776$	$\sigma = 0.727$	$\sigma = 0.817$
Gumbel	$c = 0\%$	1.677	1.969	**1.702**	1.494	1.776
		$\sigma = 0.278$	$\sigma = 0.299$	$\sigma = 0.291$	$\sigma = 0.150$	$\sigma = 0.332$
	$c = 25\%$	6.586	6.312	**5.549**	6.074	6.618
		$\sigma = 1.255$	$\sigma = 1.401$	$\sigma = 1.218$	$\sigma = 1.690$	$\sigma = 1.266$
	$c = 32\%$	6.805	6.541	**5.710**	7.479	6.894
		$\sigma = 1.414$	$\sigma = 1.379$	$\sigma = 1.394$	$\sigma = 1.147$	$\sigma = 1.329$
Frank	$c = 0\%$	1.685	1.594	1.718	**1.445**	1.795
		$\sigma = 0.288$	$\sigma = 0.283$	$\sigma = 0.307$	$\sigma = 0.142$	$\sigma = 0.352$
	$c = 25\%$	5.722	5.636	5.813	**5.262**	6.263
		$\sigma = 0.908$	$\sigma = 1.049$	$\sigma = 0.840$	$\sigma = 0.940$	$\sigma = 0.939$
	$c = 32\%$	5.953	5.928	6.051	**6.711**	6.526
		$\sigma = 0.969$	$\sigma = 1.039$	$\sigma = 0.967$	$\sigma = 0.876$	$\sigma = 1.008$
Student	$c = 0\%$	1.676	1.864	1.705	1.493	**1.776**
		$\sigma = 0.288$	$\sigma = 0.186$	$\sigma = 0.302$	$\sigma = 0.1.541$	$\sigma = 0.348$
	$c = 25\%$	6.228	5.865	6.299	5.698	5.785
		$\sigma = 0.904$	$\sigma = 1.070$	$\sigma = 0.836$	$\sigma = 1.017$	$\sigma = 0.924$
	$c = 32\%$	6.429	6.096	6.522	7.001	6.080
		$\sigma = 0.0.993$	$\sigma = 1.046$	$\sigma = 0.971$	$\sigma = 0.943$	$\sigma = 0.954$

Table 1: *Evaluation of the value of T^c for different copulas at a Kendall tau of 0.75. Values are in percentage.*

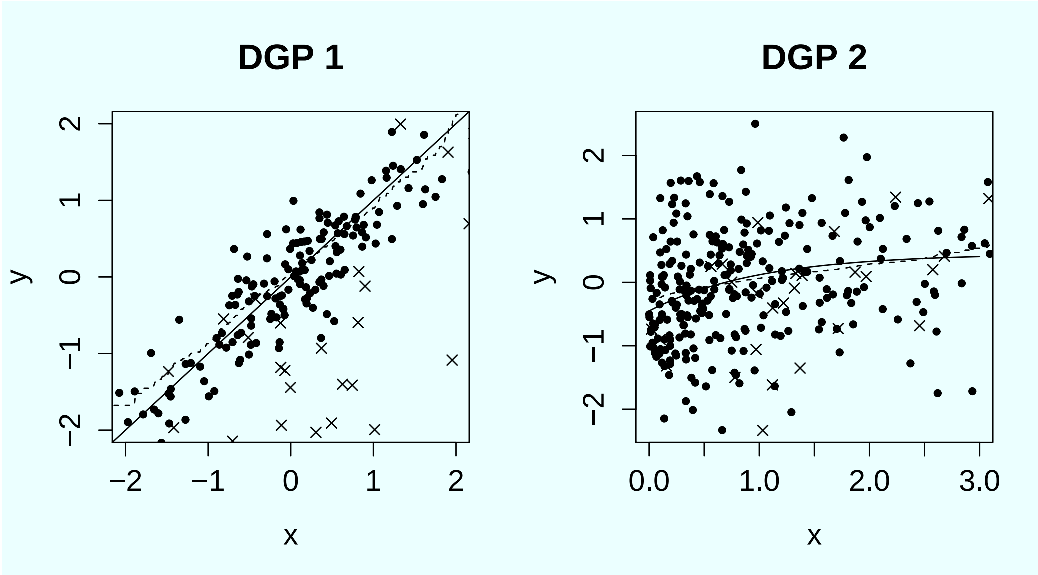

Figure 1: *Scatterplot of both DGPs with the copula-based regression function (dashed line) with 300 data and a censoring level of 25 percents. Points represent complete data, cross represent censored data; a solid line represent the theoretical regression function and a dashed line represent the proposed estimator.*

- **DGP 2:** $(F_Y(Y), F_X(X))$ follows a FGM copula with parameter $\theta = 0.8$

  1. The resulting regression function is $m(x) = (-\frac{0.8}{\sqrt{\pi}}) + 2(\frac{0.8}{\sqrt{\pi}})\sigma_Y F_X(x)$
  2. $Y \sim \mathcal{N}(0,1)$, $X \sim exp(1)$

At Figure 1, one can see an example of both DGP with their censored data represented by a cross. For each DGP, we performed 300 replications of the simulation process for 100, 200 and 500 observations. Then, we collected the average of the mean squared error (MSE) and the average of the standard deviation (SD) over the 300 replications as it is shown on table 2. One should note that the computed error is based on the difference between the theoretical value of the regression function and the regression fit found by the copula-based method. Therefore, in some case, the theoretical value is not the true regression function on a specific data set.

		n=100	n=250	n=500
DGP 1	MSE	0.1148	0.1127	0.1096
	$Bias^2$	0.0151	0.0141	0.0135
	SD	0.0997	0.0986	0.0961
DGP 2	MSE	0.1284	0.1281	0.1267
	$Bias^2$	0.0187	0.188	0.0176
	SD	0.1097	0.1093	0.1091

Table 2: *Evaluation of the average mean squared error and standard deviation for 300 replications of 2 DGPs at 3 levels of data*

6 Real data study

6.1 Heart transplant data

From the real data set provided by (Miller and Halpern, 1982), we will visually compare our estimator to the their one. The Stanford heart transplant data are extracted from a program of heart transplant which began in October 1967. The patients had to be selected to take part of that program and therefore receive a transplant. Between the selection of the patient and the transplant, some people die which leads their survival time to the value 0. The cut-off date of the study was February 1980, and at this moment, the data about the heart transplant of 184 patients were collected. The variables of interest in that study are the survival time (in days), the survival status (1 if dead, 0 if alive; which is equivalent to 1 if it is a complete data and 0 if it is a censored one), the age at the time of the first transplant and the mismatch score.

To take track of exactly the same data than in Miller and Halpern, we keep only data about patients who survived at least 10 days; and those for who the mismatch score is not missing. Therefore, we get a sample of 152 data with a censorship of 36,18 percents. The better regression function estimator to infer the regression in that data set presenter by these autors is the Buckley and James quadratic regression estimator for which the parameters are shown in table 3.

As if can be seen on Figure 2, it is visually evident that both estimators give us a quadratic regression form. However, for the Buckley and James regression estimator, the user has to specify that he desires that kind of estimation in despite of the copula-based estimator for which we don't have that kind of parameters of precise. Even if both estimator use the Kaplan-

Estimator	Intercept		Age		Age^2	
	$\hat{\alpha}$	$SD(\hat{\alpha})$	$\hat{\beta_1}$	$SD(\hat{\beta_1})$	$\hat{\beta_2}$	$SD(\hat{\beta_2})$
Buckley & James	1.35	0.71	0.107	0.037	-0.0017	0.0005

Table 3: *Coefficients for the Buckley and James regression method*

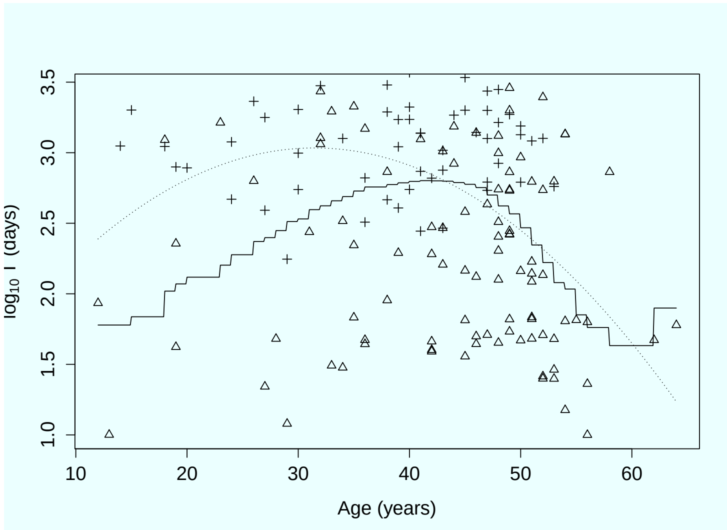

Figure 2: *Scatterplot of survival times versus age at transplant of 152 heart transplant patients. Dashed line represents the Buckley and James quadratic regression estimator and solid line represents the copula-based regression estimator. A cross represent a censored data.*

Meier estimator to infer the marginal distribution of the censored covariate, the copula-based estimator provide a better adjustment to data since that it is not simply a smoothed parabola over all the data range, but it consider local variations on the data dispersion. Note that for this data set, our copula selection criteria suggested to use an ellipsoidal copula; the gaussian one.

6.2 Canadian insurer losses ans expense data set

The data provided in Frees and Valdez which are from a large Canadian insurer indemnity claims randomly chosen comprise 1500 liability claims. We wish to thank the Society of Actuaries, through the courtesy of Edward (Jed) Frees and Emiliano Valdez, for allowing use of the data in this paper. The data are constituted of a variable which is an indemnity payment (LOSS, X) and an allocated loss adjustment expense ($ALAE$, Y). In facts, the allocated loss adjustment expense is the amount which is specifically attributable to the settlement of

	ALAE	Loss	Loss (Uncensored)	Loss (Censored)
Number	1 500	1 500	1 466	34
Mean	12 588	41 208	37 110	217 491
Median	5 471	12 000	11 048	100 000
Standard Error	28 146	102 748	92 513	258 205
Minimum	15	10	10	5 000
Maximum	501 863	2173595	2 173 595	1 000 000

Table 4: *Statistics of Canadian insurer losses data set*

individual claims. Our goal is to fit the regression function across these data in despite of the censoring.

Even if there are only 34 censored data over the 1500 data, one cannot ignore them since that the statistics over the censored losses data are pretty much different than the ones over the complete data as the table 4 shows.

It can be seen that on that data set, censoring is on the covariate in despite of the response variable. Note that in this way, the regression function can be found by a similar pattern using Kaplan-Meier to infer the marginal distribution of X in despite of Y. Furthermore, one should note that regression function was computed on the log-scaled data to simplify the computation process. Using the copula selection criteria presented in the previous section to avoid any copula misspecification, the selected copula was the Frank's one since that she gives about a 10 percents better adjustment than the Gaussian one. Therefore, one can see at Figure 3

7 Conclusion

Appendix

We start this section with some notations. Let

$$g(t) = \int_0^t (1 - H(s))^{-2} \, dH^1(s),$$

where H is the distribution of Z,$H^1(t) := P(X_i \leq t, \delta_i = 1) = \int_0^t (1 - G_0(s)) dF_0(s)$ denote the sub-distribution of the uncensored observations and G_0 is the distribution of C_0 . For positive

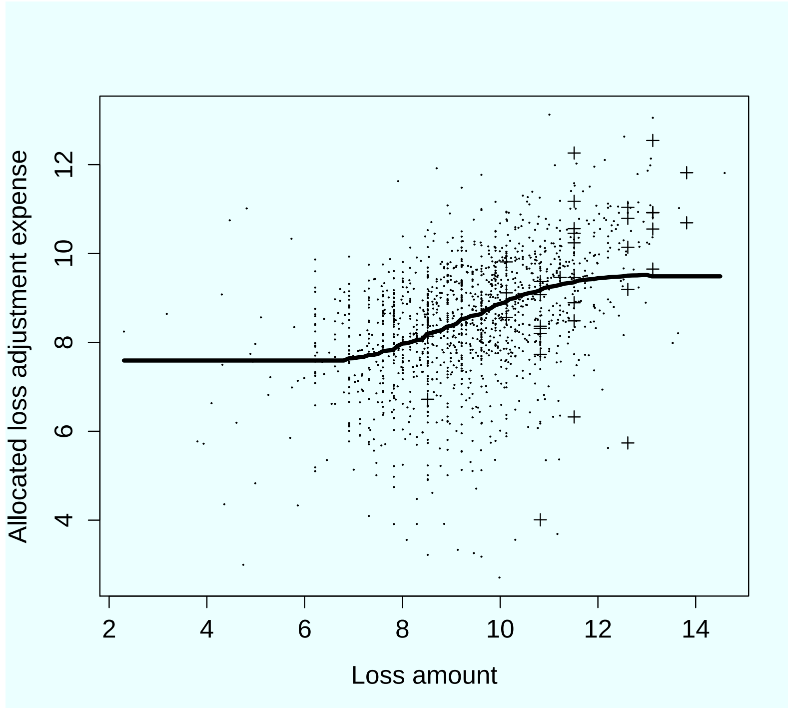

Figure 3: *Copula-based Regression function of the losses and adjusted losses of an insurance company. Points represent the complete data and cross represent the censored one.*

real numbers z and x, and $\delta = 0$ or 1, let

$$\xi(z, \delta, t) = -g(z \wedge t) + (1 - H)^{-1} I(z \leq t, \delta = 1).$$

We also set $\xi_i(x) = \xi(X_i, \delta_i, x)$. First note that

$$\mathbb{E}(\xi_i(x)) = 0 \text{ and } Cov(\xi_i(t), \xi_i(s)) = g(s \wedge t).$$

The following lemma is a du to (Lo, Mack, and Wang, 1989) expresses the KM estimator as an i.i.d. mean process with a remainder of negligible order.

Lemma 7.1. *(Lo et al. (1989))*

For $t \leq T$

$$\Gamma_n(y) - F_0(y) = n^{-1} \bar{F}(y) \sum_{i=1}^{n} \xi_i(y) + r_n(y),$$

where ξ *are*

$$\sup_{0\leq x\leq T}|r_n(y)| = O(logn/n) \quad a.s.,$$

and for any $\alpha \geq 1$,

$$\sup_{0\leq y\leq T}\mathbb{E}|r_n(y)|^\alpha = O([logn/n]^\alpha).$$

We start by recalling the result from Lemma 7.1 such that

$$\begin{aligned}\Gamma_n(y) &= F_0(y) + n^{-1}\sum_{i=1}^n \xi(Z_i,\delta_i,y) + r_n(y)\\ &= F_0(y) + \kappa(y) + r_n(y)\end{aligned}$$

where

$$\sup_{0\leq y\leq Y}|r_n(y)| = O(n^{-3/4}(\log \mathrm{n})^{3/4}),$$

$\xi(z,\delta,t) = (1-F(t))[g(z\wedge t) + \frac{1}{1-H(z)}\mathbb{1}(z\leq t,\delta=1)]$ where $H = F\cdot G$, G is the censoring distribution, $g(x) = \int_0^x [1-H(s)]^{-2}d\bar{H}_1(s)$, $\bar{H}_1(x) = 1-\mathbb{P}(Z_1\leq x,\delta_1=1)$ and $\kappa(y) = n^{-1}\sum_{i=1}^n \xi(Z_i,\delta_i,y)$. The next lemma provide an interesting decomposition of the regression estimation function at a point x_1. This expression will be useful to prove Proposition 3.1 and Theorem 3.1.

Lemma 7.2. *let* $\hat{m}$ *the regression estimator in 2.2. Under Assumption B we have*

$$\hat{m}(x_1) = m(x) + \sum_{j=1}^4 I_{n,j} + o_p(n^{-1/2}) \tag{7.6}$$

where

$$I_{n,1} = \int_0^\tau y(\Gamma_n(y) - F_0(y))\partial c_0(F_0(y),F_1(x_1),\theta_0)dF_0(y),$$

$$I_{n,2} = (\hat{F}_{1,n}(x_1) - F_1(x_1))\int_0^\tau y\partial c_1(F_0(y),F_1(x_1),\theta_0)dF_0(y),$$

$$I_{n,3} = \int_0^\tau y(\hat{\theta}_n - \theta_0)^T\dot{\mathbf{c}}(F_0(y),F_1(x_1),\theta_0)dF_0(y),$$

and

$$I_{n,4} = \int_0^\tau yc(F_0(y), F_1(x_1), \theta_0) d\kappa(y).$$

Proof of Lemma 7.6 Using the decomposition above of Lo and Singh (1985), we have

$$\begin{aligned}
\hat{m}(x_1) &= \int_0^\tau yc(\Gamma_n(y), \hat{F}_{1,n}(x_1), \hat{\theta}_n) d\Gamma_n(y) \\
&= \int_0^\tau yc(\Gamma_n(y), \hat{F}_{1,n}(x_1), \hat{\theta}_n) dF_0(y) + \int_0^\tau yc(\Gamma_n(y), \hat{F}_{1,n}(x_1), \hat{\theta}_n) d\kappa(y) \\
&+ \int_{-\infty}^{\infty} yc(\Gamma_n(y), \hat{F}_{1,n}(x_1), \hat{\theta}_n) dr_n(y) \\
&= I + II + III
\end{aligned}$$

First of all, let study the part (I) of that estimator. Using his first-order Taylor expansion around $(F_0(y), F_1(x_1), \theta_0)$, we have

$$I = \int_0^\tau yc(F_0(y), F_1(x_1), \theta_0) dF_0(y) + V_1 + V_2 + V_3$$

where

$$\begin{aligned}
V_1 &= \int_0^\tau y(\Gamma_n(y) - F_0(y)) \partial c_0(\xi_1, \xi_2, \xi_3) dF_0(y) \\
V_2 &= (\hat{F}_{1,n}(x_1) - F_1(x_1)) \int_0^\tau y \partial c_1(\xi_1, \xi_2, \xi_3) dF_0(y) \\
V_3 &= \int_0^\tau y(\hat{\theta}_n - \theta_0)^T \dot{\mathbf{c}}(\xi_1, \xi_2, \xi_3) dF_0(y)
\end{aligned}$$

with

$$\begin{aligned}
\xi_1 &= F_0(y) + t(\Gamma_n(y) - F_0(y)) \\
\xi_2 &= F_1(x_1) + t(\hat{F}_{1,n}(x_1) - F_1(x_1) \\
\xi_3 &= \theta_0 + t(\hat{\theta}_n - \theta_0)^T
\end{aligned}$$

where t belongs between 0 and 1; and $\partial c_0, \partial c_1, \dot{\mathbf{c}}$ are respectively the partial derivatives matrices of the copula distribution with respect to u_0, u_1 and θ_0. If we study V_2, it can be rewritten such that

$$
\begin{aligned}
V_2 &= (\hat{F}_{1,n}(x_1) - F_1(x_1) \int_0^\tau y\partial c_1(\xi_1, \xi_2, \xi_3) dF_0(y) \\
&= (\hat{F}_{1,n}(x_1) - F_1(x_1)) \int_0^\tau y\partial c_1(\xi_1, \xi_2, \xi_3) dF_0(y) \\
&= (\hat{F}_{1,n}(x_1) - F_1(x_1)) \int_0^\tau y\partial c_1(F_0(y), F_1(x_1), \theta_0) dF_0(y) \\
&+ (\hat{F}_{1,n}(x_1) - F_1(x_1)) \int_0^\tau y[\partial c_1(\xi_1, \xi_2, \xi_3) - \partial c_1(F_0(y), F_1(x_1), \theta_0)] dF_0(y)
\end{aligned}
$$

for which we assume that $[\partial c_1(\xi_1, \xi_2, \xi_3) - \partial c_1(F_0(y), F_1(x_1), \theta_0)] = o_p(1)$. Moreover, $sup_{x_1}|\hat{F}_{1,n}(x_1) - F_1(x_1)| = O_p(n^{-1/2})$ by Donsker's theorem. Consequently, we get

$$
\begin{aligned}
V_2 &= (\hat{F}_{1,n}(x_1) - F_1(x_1)) \int_0^\tau y\partial c_1(F_0(y), F_1(x_1), \theta_0) dF_0(y) + o_p(n^{-1/2}) \\
&\equiv I_{n,2} + o_p(n^{-1/2}).
\end{aligned}
$$

After that, let's look at the term V_3, which is

$$
\begin{aligned}
V_3 &= \int_0^\tau y(\hat{\theta}_n - \theta_0)^T \dot{\mathbf{c}}(\xi_1, \xi_2, \xi_3) dF_0(y) \\
&= \int_0^\tau y(\hat{\theta}_n - \theta_0)^T \dot{\mathbf{c}}(F_0(y), F_1(x_1), \theta_0) dF_0(y) \\
&+ \int_0^\tau y(\hat{\theta}_n - \theta_0)^T [\dot{\mathbf{c}}(\xi_1, \xi_2, \xi_3) - \dot{\mathbf{c}}(F_0(y), F_1(x_1), \theta_0)] dF_0(y)
\end{aligned}
$$

for which we assume that $[\dot{\mathbf{c}}(\xi_1, \xi_2, \xi_3) - \dot{\mathbf{c}}(F_0(y), F_1(x_1), \theta_0)] = o_p(1)$. From the work of (Shih and Louis, 1995), we know that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges to a normal with mean 0 and a certain variance $\sup |\hat{\theta} - \theta_0| = O_p(n^{-1/2})$. Therefore, we get as result for V_2

$$
\begin{aligned}
V_3 &= (\hat{\theta}_n - \theta_0)^T \int_0^\tau y\dot{\mathbf{c}}(F_0(y), F_1(x_1), \theta_0) dF_0(y) + o_p(n^{-1/2}) \\
&= I_{n,3} + o_p(n^{-1/2})
\end{aligned}
$$

For the case of V_1, it can be rewritten such that

$$\begin{aligned}
V_1 &= \int_0^\tau y(\Gamma_n(y) - F_0(y))\partial c_0(\xi_1, \xi_2, \xi_3) dF_0(y) \\
&= \int_0^\tau y(\Gamma_n(y) - F_0(y))\partial c_0(F_0(y), F_1(x_1), \theta_0) dF_0(y) \\
&+ \int_0^\tau y(\Gamma_n(y) - F_0(y))[\partial c_0(\xi_1, \xi_2, \xi_3) - \partial c_0(F_0(y), F_1(x_1), \theta_0)] dF_0(y)
\end{aligned}$$

In this case, we will assume that $[\partial c_0(\xi_1, \xi_2, \xi_3) - \partial c_0(F_0(y), F_1(x_1), \theta_0)] = o_p(1)$. From the work of (Lo and Singh, 1985), we know that $sup_y|\Gamma_n(y) - F_0(y)| = O_p(n^{-1/2}(log\ n)^{1/2})$. Moreover, from the condition C, for any constant $K \geq 1$, we have the majoration $|\partial c_0(\xi_1, \xi_2, \xi_3) - \partial c_0(F_0(y), F_1(x_1), \theta_0)| \leq K|\xi_1 - F_0(y)| \leq K|\Gamma_n(y) - F_0(y)| = O_p(n^{-1/2}(log\ n)^{1/2})$. Therefore,

$$\begin{aligned}
V_1 &= \int_0^\tau y(\Gamma_n(y) - F_0(y))\partial c_0(F_0(y), F_1(x_1), \theta_0) dF_0(y) + O_p(n^{-1}(log\ n)) \\
&= I_{n,1} + o_p(n^{-1/2})
\end{aligned}$$

Secondly, let's take the part (II) of the main equation; the integral

$$\int_0^\tau yc(\Gamma_n(y), \hat{F}_{1,n}(x_1), \hat{\theta}_n) d\kappa(y).$$

By a first-order Taylor expansion around $(F_0(y), F_1(x_1), \theta_0)$, we get

$$II = \int_0^\tau yc(F_0(y), F_1(x_1), \theta_0) d\kappa(y) + W_1 + W_2 + W_3$$

where

$$\begin{aligned}
W_1 &= \int_0^\tau y(\Gamma_n(y) - F_0(y))\partial c_0(\xi_1, \xi_2, \xi_3) d\kappa(y) \\
W_2 &= (\hat{F}_{1,n}(x_1) - F_1(x_1)) \int_0^\tau y\partial c_1(\xi_1, \xi_2, \xi_3) d\kappa(y) \\
W_3 &= (\hat{\theta}_n - \theta_0)^T \int_0^\tau y\dot{\mathbf{c}}(\xi_1, \xi_2, \xi_3) d\kappa(y).
\end{aligned}$$

If we rearrange the main part of (II), we get:

$$\begin{aligned}
\int_0^{\tau} yc(F_0(y), F_1(x_1), \theta_0) d\kappa(y) &= [yc(F_0(y), F_1(x_1), \theta_0)\xi(Z_i, \delta_i, y)\xi(Z_i, \delta_i, y)]_0^{\tau} \\
&- \int_0^{\tau} \kappa(y) d[yc(F_0(y), F_1(x_1), \theta_0)] \\
&= -\int_0^{\tau} \kappa(y) d[yc(F_0(y), F_1(x_1), \theta_0)] \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left(\int_0^{\tau} \xi(Z_i, \delta_i, y) d[yc(F_0(y), F_1(x_1), \theta_0)]\right).
\end{aligned}$$

For the part W_1 of (II), we have

$$\begin{aligned}
W_1 &= \int_0^{\tau} y(\Gamma_n(y) - F_0(y)) \partial c_0(\xi_1, \xi_2, \xi_3) d\kappa(y) \\
&= \int_0^{\tau} y(\Gamma_n(y) - F_0(y)) \partial c_0(F_0(y), F_1(x_1), \theta_0) d\kappa(y) \\
&+ \int_0^{\tau} y(\Gamma_n(y) - F_0(y))[\partial c_0(\xi_1, \xi_2, \xi_3) - \partial c_0(F_0(y), F_1(x_1), \theta_0)] d\kappa(y).
\end{aligned}$$

Since that we know that $\sup_{0\leq y\leq Y} |\Gamma_n(y) - F_0(y)| = O_p(n^{-1/2}(\log\ n)^{1/2})$, one can admit that $\sup_{0\leq y\leq Y} |\kappa(y)| = O_p(n^{-1/2}(\log\ n)^{1/2})$. Therefore,

$$\begin{aligned}
\left|\int_0^{\tau} y(\Gamma_n(y) - F_0(y)) d\kappa(y)\right| &\leq \tau \sup_{0\leq y\leq Y} |\Gamma_n(y) - F_0(y)| \cdot \left|\int_0^{\tau} d\kappa(y)\right| \\
&= o_p(n^{-1/2})
\end{aligned}$$

and, if we multiply by $\sqrt{n}$, we get:

$$\sqrt{n}\left|\int_0^{\tau} y(\Gamma_n(y) - F_0(y)) d\kappa(y)\right| = O_p(n^{-1/2}(log\ n))$$

which converges to 0. Using similar argument, we can show that $W_2 = o_p(n^{1/2})$ and $W_2 = o_p(n^{1/2})$.

Thirdly, if we look at the part (III) of the main equation, one can easily show that it is a negligeable term. In fact, we have

$$\begin{aligned}\int_{-\infty}^{\infty} yc(\Gamma_n(y), F_1(x_1), \hat{\theta}_n) dr_n(y) &= ||c||_\infty \left| \int_0^\tau dr_n(y) \right| \\ &= O_p(n^{-3/4}(log\ n)^{3/4}) \\ &= o_p(n^{-1/2}).\end{aligned}$$

Proof of Proposition 3.1 First, since $|\Gamma_n(y) - F_0(y)|$ and $\sup_{x_1} |F_{1,n}(x_1) - F_1(x_1)|$ and from Assumption B, the two terms, $I_{n,1}$ and $I_{n,2}$ in Lemma 7.6, converge to zero uniformly. Second, from Assumptions A and B, we have $I_n, 3$ in Lemma 7.6, converges to zero uniformly. Lastly, because $\kappa = n^{-1} \sum_{i=1}^n \xi(Z_i, \delta_i, y)$ and $\xi(Z_i, \delta_i, y))$ are i.i.d random variables with mean zero and finite variance, the term $I_{n,4}$ in Lemma 7.6 goes to zero uniformly. Which completes the proof.

Proof of Theorem 3.1 From Assumption A

$$\hat{\theta}_n - \theta_0 = \frac{1}{n} \sum_{i=1}^n \zeta_i + o_p(n^{-1/2})$$

where ζ_i are random i.i.d. variables with zero mean and finite variance. also, from Lo and Singh (1985) we have

$$\Gamma_n(y) = F_0(y) + n^{-1} \sum_{i=1}^n \xi(Z_i, \delta_i, y) + r_n(y)$$

We use again Lemma 7.6 we can conclude the proof.

Proof of Proposition 3.2 Proposition 3.2 can be deduced from Theorem 3.1.

References

Buckley, J. and I. James (1979). Linear regression with censored data. *Biometrika 66*, 429–436.

Fan, J. and I. Gijbels (1994). Censored-regression: local linear approximations and their applications. *Journal of the American Statistical Association 89*, 560–570.

Frees, E. and E. Valdez (1998). Annuity Valuation with Dependant Mortality. *North American Actuarial Journal 2*, 1–25.

Joe, H. and J. Xu (1996). The estimation method of inference functions for margins for multivariate models. *Technical Report, Department of Statistics, University of British Columbia 166*.

Koul, H., V. Susarla, and J. Van Ryzin (1981). Regression analysis with randomly right censored data. *Annals of Statistics 9*, 1276–1288.

Lo, S., Y. Mack, and J. Wang (1989). Density and hazard rate estimation for censored data via strong reptresentation of the kaplan-meier estimator. *Probability Theory and Relatd Fields 80*, 473–473.

Lo, S. and K. Singh (1985). The Product-Limit Estimator and the Bootstrap: Some Asymptotic Representations. *Probability Theory and Related Fields 71*, 455–465.

Miller, R. (1976). Least squares regression with censored data. *Biometrika 63*, 449–464.

Miller, R. and J. Halpern (1982). Regression with Censored Data. *Biometrika 69*, 521–531.

Noh, H., A. El Ghouch, and T. Bouezmarni (2013). Copula-Based Regression Estimation and Inference. *Journal of the American Statistical Association 108*, 676–688.

Shih, J. and T. Louis (1995). Inferences on the Association Parameter in Copula Models for Bivariate Survival Data. *Biometrics 51*, 1384–1399.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris 8*, 229–231.