



**HAL**  
open science

# Semiparametric copula models applied to the decomposition of claim amounts

Sébastien Farkas, Olivier Lopez

► **To cite this version:**

Sébastien Farkas, Olivier Lopez. Semiparametric copula models applied to the decomposition of claim amounts. 2023. hal-04300493v2

**HAL Id: hal-04300493**

**<https://hal.science/hal-04300493v2>**

Preprint submitted on 14 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Semiparametric copula models applied to the decomposition of claim amounts.

Sébastien FARKAS<sup>1</sup>, Olivier LOPEZ<sup>2</sup>.

March 14, 2024

## Abstract

In this paper, we develop a conditional copula model to analyze the distribution of a claim that generates different types of costs and/or simultaneously impacts several guarantees. Our methodology is adapted to taking into account the particular structure of our data, since observations are subject to right-censoring. Right-censoring occurs since payment of a claim is not made instantaneously, and therefore unsettled claims only provide a partial information on the phenomenon that one wishes to model. The new methodology that we develop is supported by theoretical results that show the asymptotic normality of our estimators. A simulation study and a real data analysis illustrate the method.

**Key words:** Conditional copula; Right-censoring; Claim reserving; Insurance.

**Short title:** Semiparametric copulas for the decomposition of claim amounts.

<sup>1</sup> Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, LPSM, 4 place Jussieu, F-75005 Paris, France, <sup>2</sup> CREST Laboratory, CNRS, Groupe des Écoles Nationales d'Économie et Statistique, Ecole Polytechnique, Institut Polytechnique de Paris, 5 avenue Henry Le Chatelier 91120 PALAISEAU, France, E-mails: [sebastien.farkas@sorbonne-universite.fr](mailto:sebastien.farkas@sorbonne-universite.fr), [olivier.lopez@ensae.fr](mailto:olivier.lopez@ensae.fr)

# 1 Introduction

Analyzing and predicting the evolution of claims is a challenging aspect of risk management, especially in some branches where the volatility of the final amount may be important. The question of computing the appropriate amount of reserve is, of course, crucial, but such type of analysis also enable to take actions before the settlement of a claim, in order to reduce its impact once a difficulty has been identified. In this paper, we consider the particular situation where the final cost of a claim is decomposed between several categories of costs. For example, a medical malpractice claim (example that we consider in this paper) will lead to a direct compensation of the victim, but also legal fees or experts costs. In the field of corporate insurance, incidents like mechanical failure, fire, or even cyber attack, may cause direct damage to the type of equipment that is affected, but may also generate business interruption. This problematic is also present for many types of claims, for which, for example, several guarantees are simultaneously activated. In any case, a clear analysis of what generates the expanses for the company is required to improve the vision one has of the reserve, but also to develop product innovation via a sharp tailoring of what precise type of cost should be covered in case of claim.

With the increase of available data to analyze claims, many techniques have been proposed recently to perform an accurate evaluation of RBNS (Reported But Not Settled) claims. Traditional aggregated methods like chain-ladder, see Mack (1993), Merz et al. (2013) or Saluz et al. (2014), can be modified to incorporate additional informations. For example, Wüthrich (2016) and Wüthrich (2017) considered the introduction of covariates in the development factors, allowing to use machine learning techniques to increase the precision of the reserve computation. Alternatively, micro-level reserving methods directly consider the prediction of the evolution of a claim based on its characteristics. Such type of methods have been proposed for example by Norberg (1993), Norberg (1999), Antonio and Plat (2010), Antonio et al. (2016), or Pigeon et al. (2014) in a dynamic setting (the time phenomena being modeled by Poisson processes). Lopez et al. (2016) proposed regression tree techniques to predict the amount of a claim based on information available at its occurrence, while Sabban et al. (2022) used deep learning methods to deduce from insurance reports, the outcome of severe claims. A comparison between micro-level and macro-level methods can be found for example in Jin and Frees (2013).

In the present paper, we develop a methodology that is close to Lopez (2019). Our model to predict the outcome of a claim is based on incomplete data: settled claims

but also unsettled one. This phenomenon is related to right-censoring, which is classical in survival analysis, see Fleming and Harrington (2011) for example. The general idea is the following: settled claims are, in average, closed faster than the one that are still open at the extraction of the database. Consequently, calibrating a model based on the settled claims solely typically tends to lead to an under-evaluation of the reserve. Hence an adapted methodology should be developed to correct this bias. The second axis of our methodology is to deal with an outcome of a claim which is multivariate, since the cost is decomposed into several lines of businesses or types of expanses. Hence, it is natural to rely on copula theory (see for example Nelsen (2006)) to model the dependence structure of each component of the vector of losses.

An advantage of the copula approach is the possibility to use models of various types to describe the distribution of each margin. Since the distribution type of each category of expense may be quite different (typically some may be heavy-tailed, some may not), copula theory allows to perform the analysis of these marginal distributions separately, while the dependence structure is, in a second step, done through the fitting of a parametric copula function. This explains the popularity of such techniques, see Zhao and Zhou (2010), Bouyé et al. (2000) or Jaworski et al. (2010) for examples of applications. Due to the presence of covariates describing the circumstances and nature of the claim, the dependence structure may not be the same for all claims, leading to a conditional copula modeling, see Fermanian and Wegkamp (2004) or Veraverbeke et al. (2011). Our approach is then close to the semiparametric model developed by Abegaz et al. (2012), but with an adaptation to the particular structure of our data. Apart from the bias caused by censoring, a difficulty arises since, among the covariates that may have impact on the dependence structure, one of them is unavailable for open claims (namely, the time before settlement). Hence, when it comes to prediction, an evaluation of this time before settlement must be combined with the conditional copula model we develop.

The rest of the paper is organized as follows. In Section 2, we describe the general model we develop to analyse the joint distribution of the loss vector. Theoretical validity of this approach is provided in Section 3. A simulation study and a real data analysis demonstrate the practical feasibility of the method in Section 4.

## 2 A model for the decomposition of the claim cost

This section is devoted to the description of the model used to describe the cost of a claim, and to the techniques that can be used to calibrate its parameters. Section 2.1 describes the structure of our data, with the description of the right-censoring phenomenon, caused by the fact that some claims in the portfolio are still open at the time of the statistical analysis. The procedure to model and predict the open claims is then described in Sections 2.2 to 2.5. The different steps of the procedure are the following:

- Correction of the bias caused by the censoring via the introduction of weights that dispatch the partial information contained in the open claims among complete observations (Section 2.2);
- Marginal models: each claims generates different lines of payment (compensation to the policyholder, expert fees...) that are modeled separately (Section 2.3);
- A conditional copula model is used to model the dependence structure between these lines of payment (Section 2.4);
- Joining step 2 and 3 allows to obtain a distribution for the losses that is used to predict the amount of each open claim (Section 2.5).

### 2.1 Model and observations

In many situations, a single insurance claim can trigger several guarantees and generate varied additional expenses, like expert cost, legal fees and so on. In Section 4.2, we give an example in the case of medical malpractice claims, but many other fields may be affected by such a decomposition of the costs. The cost of a claim is decomposed into  $\mathbf{L} = (L^{(1)}, \dots, L^{(d)})$ , the total cost being  $L_{tot} = \sum_{k=1}^d L^{(k)}$ . It is expected that these partial costs may not be independent from each other, since related to the same claim. Moreover, their distributions may be quite different since they are not of the same nature (and not affected with the same limits). Hence, a joint modeling of the different components of the random vector  $\mathbf{L}$  may be delicate.

Copula analysis is a convenient way to deal with such difficulties. Sklar's Theorem, see Sklar (1959), is at the core of the copula approach, and states that

$$F(l_1, \dots, l_d) = \mathfrak{C}(F^{(1)}(l_1), \dots, F^{(d)}(l_d)),$$

where  $F(l_1, \dots, l_d) = \mathbb{P}(L^{(1)} \leq l_1, \dots, L^{(d)} \leq l_d)$ ,  $F^{(k)}(l) = \mathbb{P}(L^{(k)} \leq l)$ , and  $\mathfrak{C}$  is a copula function, that is a function from  $[0, 1]^d \rightarrow [0, 1]$  which is the distribution of a  $d$ -dimensional random vector whose margins are uniformly distributed over  $[0, 1]$ . The copula function  $\mathfrak{C}$  is unique when the margins are continuous (which will be our assumption throughout this paper), hence this object characterizes the dependence structure of the random vector  $\mathbf{L}$  (information on the marginal distributions are contained in the one-dimensional cumulative distribution functions  $F^{(k)}$ ).

In our case, covariates are present since one has information on the circumstances of the claim and on the characteristics of the policyholder. These covariates  $\mathbf{X} \in \mathbb{R}^p$  have an impact on the marginal distribution, but potentially also on the dependence structure. Moreover, the time required to settle the claim, that is the difference between the date of occurrence of the claim and the date of its settlement, denoted by  $T$  in the following, is expected to also have a strong correlation with the final cost: in average, claims that are rapidly settled tend to be associated with a smaller amount. However, unlike  $\mathbf{X}$ , the information on  $T$  is available only at the final settlement, and not at the opening of the claim. This is a right-censoring framework, see Fleming and Harrington (2011): instead of observing  $T$ , one observes  $Y = \inf(T, C)$  where  $C$  is a censoring time, and  $\delta = \mathbf{1}_{T \leq C}$ . If  $\delta = 1$  the claim is closed, and one observes  $Y = T$ . In the opposite situation, the claim is still open since  $Y = C$ ,  $C$  being the difference between the date of occurrence and the date at which one loses track on the claim (because data has been extracted before its closure, or because the claim is part of a retroceded portfolio). If  $\delta = 1$ , the final amounts  $\mathbf{L}$  are known, while it is not the case when  $\delta = 0$ .

The reason for not calibrating the model only on claims such that  $\delta_i = 1$  is explained in detail in Lopez (2019). Typically, a positive (and potentially strong) correlation is expected between  $T_i$  and the total loss  $L_{tot}$ , based on the heuristic that "the longer it takes for a claim to be solved, the higher it costs". Calibrating a model only on closed claims is likely to tend to underestimate the typical values taken by  $L_{tot}$ , since the population of closed claims is characterized by an overrepresentation of claims with small final amount. The methodology that we propose (see section 2.2) consists in correcting the bias caused by the incompleteness of some observations via a proper weighting scheme, which allows more importance to observations in the area of the distribution where information is missing.

Let  $F(l_1, \dots, l_d | \mathbf{x}, t) = \mathbb{P}(L^{(1)} \leq l_1, \dots, L^{(d)} \leq l_d | \mathbf{X} = \mathbf{x}, T = t)$ , and  $F^{(k)}(l | \mathbf{x}, t) = \mathbb{P}(L^{(k)} \leq l | \mathbf{X} = \mathbf{x}, T = t)$ . Then, the conditional copula of  $\mathbf{L}$  conditionally to  $\mathbf{X} = \mathbf{x}$  and

$T = t$  (see e.g. Veraverbeke et al. (2011)) is the copula function  $\mathfrak{C}^{(\mathbf{x}, t)}$  such that

$$F(l_1, \dots, l_d | \mathbf{x}, t) = \mathfrak{C}^{(\mathbf{x}, t)}(F^{(1)}(l_1 | \mathbf{x}, t), \dots, F^{(d)}(l_d | \mathbf{x}, t)).$$

Standard regression models can be used to estimate each of the marginal conditional distribution functions, see examples in Section 2.3 below. Our main purpose is to focus on the estimation of the dependence structure.

Introducing a parametric copula family  $\mathcal{C} = \{\mathfrak{C}_\theta : \theta \in \Theta\}$ , where  $\Theta \subset \mathbb{R}^m$  is finite dimensional, we assume that, for all possible values of  $x$  and  $t$ ,  $\mathfrak{C}^{(\mathbf{x}, t)} \in \mathcal{C}$ . Let  $\theta(\mathbf{x}, t)$  denote the function such that  $\mathfrak{C}^{(\mathbf{x}, t)} = \mathfrak{C}_{\theta(\mathbf{x}, t)}$ . Our aim is to retrieve this function, either using a parametric model or a semiparametric model, as described in Section 2.4 below.

To estimate this function, we rely on observations of a set of  $n$  past claims. We assume that  $(\mathbf{L}_i, \mathbf{X}_i, T_i, C_i)_{1 \leq i \leq n}$  are assumed to be i.i.d. replications of  $(\mathbf{L}, \mathbf{X}, T, C)$ . Then, the observations are i.i.d. replications  $(\mathbf{M}_i, \mathbf{X}_i, Y_i, \delta_i)_{1 \leq i \leq n}$  of  $(\mathbf{M}, \mathbf{X}, Y, \delta)$  where

$$\begin{cases} Y &= \inf(T, C), \\ \delta &= \mathbf{1}_{T \leq C}, \\ \mathbf{M} &= (M^{(1)}, \dots, M^{(d)}), \end{cases}$$

with  $M^{(k)} = L^{(k)}$  if  $\delta = 1$ , and  $M^{(k)} = 0$  otherwise.  $M^{(k)}$  represents partial payments done on the  $k$ -th line at the end of the observation of the claim. The aim of the procedure will then be able to predict the final amount of open claims,  $L_i^{(k)}$  for all  $k$  and all  $i$  such that  $\delta_i = 0$ . This will require to also predict  $T_i$  when  $\delta_i = 0$ , but only as a mean to ultimately use it to predict the amount.

Let us note that the present framework can be seen as a particular case of a more general model where each  $M^{(k)}$  would be  $M^{(k)} = \inf(L^{(k)}, D^{(k)})$  where  $D^{(k)}$  is a censoring variable with  $D^{(k)} > L^{(k)}$  only if  $\delta = 1$ .  $D^{(k)}$  could represent payments that already occurred on line  $k$ , but that are still incomplete. To simplify, we do not consider this more general model here, since a dependence structure on the  $D^{(k)}$  should also be assumed in this case. On the other hand, if information about these payments are available (and assuming that there is no information in the dependence structure between the variables  $D^{(k)}$ ), this information can be included in the covariates  $X$  instead.

Let us also note that we here only consider the database of open claims, that is that are already known to the insurer, that is "Reported But Not Settled Claims" (RBNS). Incurred But Not Reported claims are not in the scope of the analysis, and should be studied separately when it comes to constituting a reserve (we only focus in this paper on the part of the reserve related to RBNS claims).

## 2.2 Inverse probability of censoring weighting (IPCW)

Since a duration phenomenon is present in the data acquisition process (one must wait until the settlement of a claim to know its final state), censoring would introduce some bias if no correction is performed. The Inverse Probability of Censoring Weighting methodology (IPCW), see for example Van der Laan and Robins (2003), is a simple way to proceed. It consists of determining an appropriate weight to put on each observation to asymptotically cancel the bias caused by the censoring. Only the uncensored observations are affected with a non-zero weight (since these observations are complete), but the partial information contained in the censored one is used to determine these weights.

The core of such an approach is the following result. In the rest of the paper, we assume that  $(T, \mathbf{L}, \mathbf{X})$  is independent from the censoring mechanism  $C$ . Under this assumption, for any function  $\phi$  such that  $E[|\phi(T, \mathbf{L}, \mathbf{X})|] < \infty$ , and such that  $\phi(y, \mathbf{m}, \mathbf{x}) = 0$  if  $y$  is not in the support of the distribution of  $Y$ ,

$$E \left[ \frac{\delta \phi(Y, \mathbf{M}, \mathbf{X})}{S_C(Y)} \right] = E [\phi(T, \mathbf{L}, \mathbf{X})], \quad (2.1)$$

where  $S_C(y) = \mathbb{P}(C \geq y)$ . In the following, we assume to simplify that  $T$  and  $C$  have the same support, which guarantees that  $\inf\{t : \mathbb{P}(T \geq t) = 0\} = \inf\{t : \mathbb{P}(C \geq t) = 0\}$ , so that (2.1) holds for all function  $\phi$  with first order finite moment. In the general case, this would lead to consider truncated versions of functions whose support is not compact, introducing some bias which can not be removed without some additional parametric assumption (since one part of the distribution is not observed in this case).

Equation (2.1) implies that

$$\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(Y_i, \mathbf{M}_i, \mathbf{X}_i)}{S_C(Y_i)} \rightarrow_{n \rightarrow \infty} E[\phi(T, \mathbf{L}, \mathbf{X})], \text{ almost surely,} \quad (2.2)$$

from the Strong Law of Large Numbers. The left-hand side quantity has the advantage to be computable from observed quantities, while the limit is an expectation with respect to the variables we are actually interested in ( $T$ ,  $\mathbf{L}$  and  $\mathbf{X}$ ). Nevertheless, the function  $S_C$  is in general unknown and hard to model, due to the lack of visibility on the censoring process.

A Kaplan-Meier estimator (see Kaplan and Meier (1958)) can be used for  $S_C$ , that is, if  $T$  is continuous,

$$\hat{S}_C(t) = \prod_{Y_i \leq t} \left( 1 - \frac{\delta_i}{\sum_{j=1}^n Y_j \mathbf{1}_{Y_j \geq t}} \right),$$



a more general expression can be found in Gill (1983) that covers also the discrete case. Defining

$$W_{i,n} = \frac{1}{n} \frac{\delta_i}{\hat{S}^C(Y_i)}, \quad (2.3)$$

the IPCW approach consists in estimating any quantity of the type  $E[\phi(T, \mathbf{L}, \mathbf{X})]$  by

$$\sum_{i=1}^n W_{i,n} \phi(Y_i, \mathbf{M}_i, \mathbf{X}_i).$$

These computable weighted sums will replace every empirical mean that we would use in the case of complete (uncensored) data.

### 2.3 Regression model for the margins

Since the margins are expected to have heterogeneous behaviors (some lines may be very volatile while some others are not, some of them may have a low probability of activation...), the models that are used to study the distribution of each  $L^{(k)}$  may be of different types. Typically we distinguish between fully parametric models, and semi-parametric or non-parametric ones.

**Fully parametric models.** In a fully parametric model, one assumes that  $F^{(k)}(l|\mathbf{x}, t) = F_\beta(l|\mathbf{x}, t)$ , where  $\{F_\beta(\cdot|\mathbf{x}, t), \mathbf{x} \in \mathcal{X}, t \geq 0, \beta \in \mathcal{B}\}$  is a parametric set of distribution functions (here,  $\mathcal{X}$  is the support of the random vector  $\mathbf{X}$ ). Typically, this is the framework of the Generalized Linear Model (see Nelder and Baker (1972)). According to this model, the distribution of  $L^{(k)}|\mathbf{X} = \mathbf{x}, T = t$  has density  $f_{\beta(\mathbf{x}, t)}(l)$ , where  $\{f_\beta : \beta \in \mathcal{B}\}$  is the collection of densities from a given exponential-type distribution. Moreover, it is assumed that

$$g(E[L^{(k)}|\mathbf{X} = \mathbf{x}, T = t]) = h(\beta(\mathbf{x}, t)) = \alpha' \mathbf{z}, \quad (2.4)$$

where  $\mathbf{z} = (\mathbf{x}, t)$ ,  $g$  is a fixed monotonic function ( $h$  is deduced from  $g$ ) and  $\alpha$  a finite dimensional parameter (here, the  $'$  symbol denotes the transpose of a vector). From an estimator  $\hat{\alpha}$  of  $\alpha$ , one deduce an estimator of the conditional distribution function  $\hat{F}^{(k)}(l|\mathbf{x}, t) = \int_0^l f_{g^{-1}(\hat{\alpha}' \mathbf{z})}(u) du$ , where  $\mathbf{z} = (\mathbf{x}, t)'$ , simply plugging  $\hat{\alpha}$  in (2.4). Since  $L^{(k)}$  is subject to censoring, a specific adaptation of the log-likelihood estimation is required. Consistent estimators of  $\hat{\alpha}$  can be used, such as Stute (1999) (see also Lopez (2009)). The idea is essentially to modify traditional estimation criteria by using the weights  $W_{i,n}$  within the loss function. Stute (1999) relies on the optimization of a weighted quadratic loss, a weighted log-likelihood can be used instead.

**Semi-parametric or nonparametric models.** Since the covariates are smooth, a simple way to estimate the distribution nonparametrically consists in using kernel estimator for the conditional distribution function.

This leads to

$$\hat{F}^{(k)}(l|\mathbf{z}) = \sum_{i=1}^n W_{i,n} \frac{\tilde{K}\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h'}\right)}{\sum_{j=1}^n \tilde{K}\left(\frac{\mathbf{Z}_j - \mathbf{z}}{h'}\right)} \mathbf{1}_{M_i^{(k)} \leq l}, \quad (2.5)$$

for some kernel  $\tilde{K} \geq 0$ . A traditional choice is to take a product kernel, that is  $\tilde{K}(\mathbf{z}) = \prod_{j=1}^{p+1} \tilde{k}(z_j)$ , where  $\int k(z)dz = 1$  and  $\int zk(z)dz = 0$ . With some anticipation, let us note that the bandwidth  $h'$  should be different from the smoothing parameter  $h$  that will be used for the specific estimation of the association parameter (see section 2.4).

More elaborate models, coming from example from machine-learning field, consist in decomposing  $L^{(k)} = m^{(k)}(\mathbf{X}, T) + \varepsilon^{(k)}$ , where  $m$  is a function belonging to a potentially infinite dimension class, and  $\varepsilon$  is a residual. For example, in the case where  $m^{(k)}(\mathbf{x}) = E[L^{(k)}|\mathbf{X} = \mathbf{x}, T = t]$ ,  $E[\varepsilon^{(k)}|\mathbf{X} = \mathbf{x}, T = t] = 0$ , while in the case of median regression,  $m^{(k)}(\mathbf{x})$  is the conditional median and  $\varepsilon$  is such that  $\mathbb{P}(\varepsilon^{(k)} \geq 0|\mathbf{X} = \mathbf{x}, T = t) = 1/2$ . If one assumes that  $\varepsilon$  does not depend on  $\mathbf{X}$  or  $T$ ,  $F^{(k)}(l|\mathbf{x}, t) = F^{(\varepsilon^{(k)})}(l - m(\mathbf{x}, t))$ , where  $F^{(\varepsilon^{(k)})}$  is the cumulative distribution function of  $\varepsilon^{(k)}$ . Based on an estimator  $\hat{m}^{(k)}$  of  $m^{(k)}$ , one can define  $\hat{\varepsilon}_i^{(k)} = L_i^{(k)} - \hat{m}^{(k)}(\mathbf{X}_i, T_i)$ , and

$$\hat{F}^{(\varepsilon^{(k)})}(e) = \sum_{i=1}^n W_{i,n} \mathbf{1}_{\hat{\varepsilon}_i^{(k)} \leq e},$$

and  $\hat{F}^{(k)}(l|\mathbf{x}, t) = \hat{F}^{(\varepsilon^{(k)})}(l - \hat{m}^{(k)}(\mathbf{x}, t))$ . Various machine learning techniques have been proposed for censoring models, like regression trees (see Bou-Hamad et al. (2011)), random forests (see Ishwaran et al. (2008), Gerber et al. (2021)) or neural networks (see Sabban et al. (2022)).

## 2.4 Conditional copula model estimation

Let  $\mathbf{U}_i = (U_i^{(1)}, \dots, U_i^{(d)})$  where  $U_i^{(k)} = F^{(k)}(L_i^{(k)}|\mathbf{X}_i, T_i)$ . The variables  $U_i^{(k)}$  are uniformly distributed over  $[0, 1]$ . Moreover, the conditional density of  $\mathbf{U}_i$  conditionally to  $\mathbf{Z}_i = (\mathbf{X}_i, T_i)$  is  $\mathfrak{C}_{\theta(\mathbf{x}_i, T_i)}$ .

If we had the ability to observe  $(\mathbf{U}_i)_{1 \leq i \leq n}$ , one could rely on a localized version of the maximum likelihood estimator. Let us introduce a kernel function  $K$  (that is function from  $\mathbb{R}^p$  to  $\mathbb{R}$  such that  $\int K(\mathbf{z})d\mathbf{z} = 1$ , with  $\int \mathbf{z}K(\mathbf{z})d\mathbf{z} = 0$ ). To simplify, we will consider in

the following a product kernel, that is a function  $K(\mathbf{z}) = \prod_{j=1}^{p+1} k(z_j)$ . One could maximize

$$\frac{1}{h^{p+1}} \sum_{i=1}^n K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \log \mathbf{c}_\theta(\mathbf{U}_i),$$

where  $\mathbf{c}_\theta$  is the copula density associated with  $\mathfrak{C}_\theta$ , that is  $\mathbf{c}_\theta(\mathbf{u}) = \partial^d \mathfrak{C}_\theta(\mathbf{u}) / \partial u^{(1)} \dots \partial u^{(d)}$ . Here, we consider the same bandwidth  $h$  (which is of course tending to zero with  $n$ ) for each coordinate of  $\mathbf{z}$  to simplify the notations, but in full generality these bandwidths may be different.

In our case, the vectors  $(\mathbf{U}_i)_{1 \leq i \leq n}$  are not directly observed, but one can compute

$$\hat{U}_i^{(k)} = \hat{F}^{(k)}(M_i^{(k)} | \mathbf{X}_i, Y_i),$$

from the marginal distribution estimators of Section 2.3. Let us note that, because of the censoring, this quantity will not be close to  $U_i^{(k)}$  when  $\delta_i = 0$ , since, in this case,  $M_i^{(k)} \neq L_i^{(k)}$ . Nevertheless, we will require to compute this quantity only when  $\delta_i = 1$ . This is the consequence of the IPCW technique of Section 2.2 that is required to correct the bias of the censoring: this technique leads to the following estimator

$$\hat{\theta}(\mathbf{x}, t) = \arg \max_{\theta \in \Theta} \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \log \mathbf{c}_\theta(\hat{\mathbf{U}}_i), \quad (2.6)$$

where we recall that  $W_{i,n} = 0$  if  $\delta_i = 0$ . In this expression, the vectors  $\mathbf{Z}_i$  (which depend on  $T_i$ ) are fully observed only if  $\delta_i = 1$ , but, again, their computation is not required in case  $\delta_i = 0$  due to the nullity of the weight. The estimator (2.6) can be seen as an adaptation of the estimator of Abegaz et al. (2012).

## 2.5 Prediction of an open claim

Let us consider an open claim with characteristics  $\mathbf{x}$ . The claim is open since a duration  $y$ . The conditional copula model and the marginal regression models described above allows to determine the conditional distribution of  $\mathbf{L}$  conditionally to  $\mathbf{X} = \mathbf{x}$  and  $T = t$ . When it comes to predicting an open claim, they can not be used directly, since  $T$  is unknown (one only knows that  $T \geq y$ ). A possibility is to rely on a predictor of  $T$ , say  $\hat{T}$ . Let

$$\hat{m}(\mathbf{x}, t) = \int_{\mathbb{R}^d} \mathbf{l} \times \mathbf{c}_{\hat{\theta}(\mathbf{x}, t)}(\hat{F}^{(1)}(l^{(1)} | \mathbf{x}), \dots, \hat{F}^{(d)}(l^{(d)} | \mathbf{x})) d\hat{F}^{(1)}(l^{(1)} | \mathbf{x}, t) \dots d\hat{F}^{(d)}(l^{(d)} | \mathbf{x}, t),$$

where we used the notation  $\mathbf{l} = (l^{(1)}, \dots, l^{(d)})$ . This is an estimator of  $m(\mathbf{x}, t) = E[\mathbf{L} | \mathbf{X} = \mathbf{x}, T = t]$ . Based on  $\hat{T}$ , one can use  $\hat{m}(\mathbf{x}, \hat{T})$  to predict the final state of the claim. This

only gives the "central scenario". The fitted conditional copula density and the marginal models can be used more generally to simulate the conditional distribution of  $\mathbf{L}$ , and thus to get an analysis on the volatility of this prediction.

The crucial question is hence to determine a proper prediction method  $\hat{T}$ . This can be based on a standard regression model on the censored variable  $T$ . An accelerated failure-time model (see Wei (1992)) can for example be fitted to obtain an estimator of  $F^T(t|\mathbf{x}) = \mathbb{P}(T \leq t|\mathbf{X} = \mathbf{x})$ , or a Cox model (see Cox (1975)). Other semiparametric models, like Stute (1999) or Lopez (2009) can also be used. Machine learning methods, like for example survival forests (see Ishwaran et al. (2008)) are also available. In each case, one obtains an estimator  $\hat{F}^T(t|\mathbf{x})$  of  $F^T$ , from which a predictor of  $T$  can be obtained.

For an open claim with characteristics  $\mathbf{x}$ , open since  $y$ , the idea is to estimate  $E[T|\mathbf{X} = \mathbf{x}, T \geq y]$ . This leads to

$$\hat{T} = \frac{\int_y^\infty t d\hat{F}^T(t|\mathbf{x})}{\hat{S}^T(y|\mathbf{x})}, \quad (2.7)$$

where  $\hat{S}^T(y|\mathbf{x}) = 1 - \hat{F}^T(y|\mathbf{x})$ . The quality of the prediction will of course rely on the regression model on  $T$ .

**Remark 2.1** *It is important to use a predictor  $\hat{T}$  of the form (2.7), that is an estimator of  $E[T|\mathbf{X} = \mathbf{x}, T \geq y]$ , instead of a more simple estimator of  $E[T|\mathbf{X} = \mathbf{x}]$ : this alternative method does not take into account all the available information on  $T$ , and could lead to predictions for which  $\hat{T} < y$ .*

### 3 Asymptotic consistency

To study the theoretical behavior of the method, we first describe in Section 3.1 the assumptions required to obtain consistency of the estimation of the dependence structure. Our asymptotic results are gathered in Section 3.2.

#### 3.1 List of assumptions and discussion

We distinguish between three types of assumptions required to obtain the theoretical results: on the copula family, on the estimation of the margins, and regularity assumptions related to kernel estimation. The assumptions are of the same type as the one used by Omelka et al. (2021) and Tsukahara (2005) to study the behavior of semiparametric estimators of copulas (that is the copula structure is parametric, but the margins are estimated through a nonparametric estimator, namely the empirical distribution function),

but with small adaptations required by the context of censoring. In our case, we add assumptions on how the margins are estimated, since we want to be able to consider various types of models for the margins.

**Assumptions on the copula family.**

Before stating the assumptions, we need to introduce some notations. We use the bracketing number (see, for example, Chapter 19 in Van der Vaart (2000)) to define the richness of a class of functions. A bracket  $[\mathbf{u}, \mathbf{l}]$ , where  $\mathbf{u}$  and  $\mathbf{l}$  are functions such that  $\mathbf{u} \leq \mathbf{l}$ , is the set of functions  $\mathbf{f}$  such that  $\mathbf{u} \leq \mathbf{f} \leq \mathbf{l}$ . A  $\varepsilon$ -bracket is such that  $E[(\mathbf{u}(\mathbf{U}) - \mathbf{l}(\mathbf{U}))^2] \leq \varepsilon^2$ .  $N_{[]}(\varepsilon, \mathcal{F})$  denotes the number of  $\varepsilon$ -brackets required to cover a class of functions  $\mathcal{F}$ . How fast  $N_{[]}(\varepsilon, \mathcal{F})$  explodes when  $\varepsilon$  tends to zero is an indication about the complexity of the class of functions  $\mathcal{F}$ .

**Assumption 1** Let  $\mathcal{F}_1 = \{\mathbf{u} \rightarrow \log \mathbf{c}_\theta(\mathbf{u}) : \theta \in \Theta\}$ , and assume that there exists an envelope function  $\Phi$  such that  $|\log \mathbf{c}_\theta(\mathbf{u})| \leq \Phi(\mathbf{u})$  for all  $\theta$  and  $\mathbf{u}$  with

$$E \left[ \frac{\Phi(\mathbf{U})}{S^C(T)^\iota} \right] < \infty, \tag{3.1}$$

for some  $\iota > 0$ . Assume that  $N_{[]}(\varepsilon, \mathcal{F}) \leq A\varepsilon^{-\kappa}$ , for some  $A$  and  $\kappa > 0$ .

Moreover, assume that there exists  $\Phi_1$  such that,  $\forall \phi \in \mathcal{F}_1$ ,  $\|\nabla_{\mathbf{u}}\phi(\mathbf{u}) - \nabla_{\mathbf{u}}\phi(\mathbf{u}')\| \leq \Phi_1(\mathbf{u})$ , where  $\Phi_1$  is continuous and  $\nabla_{\mathbf{u}}f$  denotes the gradient vector of a function  $f$  with respects to its arguments, with  $E[\Phi_1(\mathbf{U})] < \infty$ .

This assumption is relatively easy to fulfill for a classical parametric copula family (like Gaussian copula, standard Archimedean copulas...). Typically, a polynomial bound  $A\varepsilon^{-\kappa}$  for the covering number is obtained when the class  $\mathcal{F}$  is regular enough (typically, when this class is Lipschitz with respect to the parameter  $\theta$ , see Example 19.7 in Van der Vaart (2000)).

Next, we need to dominate the class of copula functions and some of their derivatives, with an assumption which is close to the one present in Tsukahara (2005) and in Omelka et al. (2021). The only difference stands in the presence of censoring in our case, which strengthens the assumptions. Right-censoring induces potentially erratic behavior when studying the right-tail of the distribution. Typically, this explains the introduction of  $S^C$  (the survival function of the censoring variable) and a function  $\mathfrak{K}$  in the last two moment conditions of Assumption 2, defined as

$$\mathfrak{K}(t) = \left[ - \int_{\infty}^t \frac{dS^C(s)}{S^C(s)^2 S^T(s)} \right]^{-1},$$

where  $S^T$  is the survival function of  $T$ . If these two functions decreased too fast (compared to the tail of the distribution of  $T$ ), the proper convergence rate can not be achieved. Truncation of the highest observations is required, leading to some bias that can not be cancelled even asymptotically. This type of conditions is classical in presence of censoring, see for example Gill (1983).

**Assumption 2** *Introduce*

$$\mathcal{F}_2 = \{\mathbf{u} \rightarrow \nabla_{\theta} \log c_{\theta}(\mathbf{u}) : \theta \in \Theta\}.$$

Let

$$\forall u \in [0, 1], m(u) = u(1 - u),$$

and assume that there exists some  $0 \leq \beta < 1/2$ ,  $0 < \alpha < 1/2$ , and some positive constant  $A_2$  such that

$$\forall \phi \in \mathcal{F}_2, \forall k \in \{1, \dots, d\}, |\dot{\phi}^{(j)}(\mathbf{u})| m(u^{(k)})^{\alpha} \leq \sum_{k=1}^d \frac{A_2}{m(u^{(k)})^{\beta}},$$

where  $\dot{\phi}^{(j)}$  denotes the derivative with respect to the  $j$ -th component of  $\mathbf{u}$ .

This assumption holds for standard classes of copula, see for example Remark 3 in Omelka et al. (2021). It essentially controls the explosion of the score function close to the limits of  $[0, 1]^d$ . The fact that  $\beta < 1/2$  allows  $E[m(U^{(k)})^{-\beta}]$  to be finite.

**Assumptions on the estimation of the margins.**

The margins estimation should satisfy the following assumption.

**Assumption 3** *Assume that*

$$\sup_{i=1, \dots, n: \delta_i=1} \sup_{k=1, \dots, d} \left| \frac{U_i^{(k)}}{\hat{U}_i^{(k)}} + \frac{1 - U_i^{(k)}}{1 - \hat{U}_i^{(k)}} \right| = O_P(1), \quad (3.2)$$

$$\sup_{i=1, \dots, n: \delta_i=1} \sup_{k=1, \dots, d} \left| \frac{U_i^{(k)} - \hat{U}_i^{(k)}}{[U_i^{(k)}(1 - U_i^{(k)})]^{\alpha}} \right| = O_P(\varepsilon_n), \quad (3.3)$$

for some sequence  $\varepsilon_n$  tending to zero, and where  $\alpha$  is defined in Assumption 2.

This conditions would be fulfilled if one would consider the empirical distribution function to estimate the margins, or with parametric models. These condition also hold for parametric estimation of the margins, as long as the model is regular enough. If a

parametric estimator is used for the margins, then one can expect  $\varepsilon_n = n^{-1/2}$ . In absence of covariates, if the margins are estimated by the empirical distribution function, (3.3) holds easily from the uniform convergence of weighted empirical processes, see Example 19.12 in Van der Vaart (2000). The most delicate case concerns the estimator (2.5), where the rate is slower. In section 6.5, we show that a proper choice of the bandwidth allows to obtain a sufficiently fast convergence.

**Assumptions on the kernel and on the regularity of the copula regression model.**

Let us recall that this last set of assumptions is only required if the simplifying assumption does not hold.

**Assumption 4** *The kernel function  $K : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ , has the following properties:*

$$\begin{aligned} \int K(\mathbf{u})d\mathbf{u} &= 1, \\ \int \mathbf{u}K(\mathbf{u})d\mathbf{u} &= 0, \\ \int \|\mathbf{u}\|_2^2 |K(\mathbf{u})|d\mathbf{u} &< \infty. \end{aligned}$$

Additionally, we need a regularity assumption on the conditional distribution of  $\mathbf{U}$  given  $\mathbf{Z}$ .

**Assumption 5** *Let  $f_{\mathbf{U},\mathbf{Z}}(\mathbf{u}, \mathbf{z})$  denote the joint density of  $(\mathbf{U}, \mathbf{Z})$  computed at point  $(\mathbf{u}, \mathbf{z})$ . Assume that, for all  $\mathbf{z} \in \mathcal{Z}$ ,*

$$\forall \mathbf{v}, \left| \mathbf{v}' \nabla_{\mathbf{z}}^2 f_{\mathbf{U},\mathbf{Z}}(\mathbf{u}, \mathbf{z}) \mathbf{v} \right| \leq \|\mathbf{v}\|_2^2 \mathfrak{s}(\mathbf{u}), \quad (3.4)$$

with

$$\int \varphi_{\theta(\mathbf{z})}^{(j,k)}(\mathbf{u}) \mathfrak{s}(\mathbf{u}) d\mathbf{u} < \infty. \quad (3.5)$$

### 3.2 Asymptotic behavior of the estimator of the association parameter

We now state our main theoretical result on the asymptotic behavior of the estimates of the association parameter. We obtain an asymptotic representation of the estimator which makes it asymptotically equivalent to the estimator we would compute if we knew exactly the distribution of the margins and of the censoring mechanism. From this, asymptotic normality is easily derived.

**Theorem 3.1** Let  $\theta_h^*(\mathbf{z}) = \arg \max_{\theta} E \left[ K \left( \frac{\mathbf{Z}-\mathbf{z}}{h} \right) \log \mathbf{c}_{\theta}(\mathbf{U}) \right]$ , and

$$\begin{aligned}\varphi_{\theta(\mathbf{z})}(\mathbf{u}) &= \nabla_{\theta} \log \mathbf{c}_{\theta}(\mathbf{u}), \\ \varphi_{\theta(\mathbf{z})}^{(j,k)}(\mathbf{u}) &= \partial_{j,k}^2 \log \mathbf{c}_{\theta}(\mathbf{u}),\end{aligned}$$

where  $\partial_{j,k}^2$  denotes the second order derivatives with respect to  $\theta_j$  and  $\theta_k$ . Assume that, for some  $\iota > 0$ ,

$$E \left[ \frac{\varphi_{\theta(\mathbf{z})}(\mathbf{U})}{\mathfrak{K}(T)^{1/2+\iota}} \right] < \infty.$$

Under Assumptions 1 to 3, and under Assumption 4 and 5, and if

$$n\varepsilon_n^2 h^{p+1} \rightarrow 0, \quad (3.6)$$

we have

$$\hat{\theta}(\mathbf{z}) - \theta_h^*(\mathbf{z}) = -\Sigma(\mathbf{z})^{-1} \left\{ \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n}^* K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \varphi_{\theta(\mathbf{z})}(\mathbf{U}_i) \right\} + o_P(n^{-1/2} h^{(p+1)/2}), \quad (3.7)$$

with  $\Sigma(\mathbf{z}) = (\sigma^{(j,k)}(\mathbf{z}))_{j,k}$ , is supposed to be invertible, with

$$\sigma^{(j,k)}(\mathbf{z}) = E \left[ \varphi_{\theta(\mathbf{z})}^{(j,k)}(\mathbf{U}_i) | \mathbf{Z} = \mathbf{z} \right].$$

Hence,

$$n^{1/2} h^{(p+1)/2} \{ \hat{\theta}(\mathbf{z}) - \theta_h^*(\mathbf{z}) \} \implies \mathcal{N}(\mathbf{0}, \Sigma(\mathbf{z})^{-1} S(\mathbf{z}) S(\mathbf{z})' \Sigma(\mathbf{z})^{-1}), \quad (3.8)$$

where  $S(\mathbf{z}) = E \left[ \varphi_{\theta(\mathbf{z})}(\mathbf{U}) | \mathbf{Z} = \mathbf{z} \right] f_{\mathbf{Z}}(\mathbf{z})$ , and the transpose of a matrix  $A$  is denoted by  $A'$ .

Let us note that (3.6) typically imposes that the margins are estimated at a rate which is faster than the expected rate for the conditional copula parameter. It can be easily shown, from the proof of Theorem 3.1, that if this condition does not hold, the rate becomes  $\varepsilon_n$ .

Finally, let us note that (3.7) only concerns the stochastic part of the error, and does not include the bias term, that is the difference between  $\theta_j^*(\mathbf{z})$  and  $\theta(\mathbf{z})$ . The bias term is covered by the following Proposition 3.2.

**Proposition 3.2** Under Assumptions 2, 4 and 5,

$$\sup_{\mathbf{z}} \|\theta_h^*(z) - \theta(\mathbf{z})\| = O(h^2).$$

All proofs are postponed to the appendix section 6.



## 4 Empirical evidence

In this section, we provide empirical evidence on the behavior of the procedure. This is done through a simulation study in section 4.1, and a real data application in section 4.2.

### 4.1 Simulation setting

In this section, we investigate the practical behavior of the procedure for finite sample size. More precisely, we look at the impact of the following components of the model:

- the dependence structure, in terms of copula family, but also on the strength of the dependence;
- the strength of the censoring (that is the average proportion of censored observations in the sample);
- the smoothing parameter  $h > 0$ .

We consider a two-dimensional covariate  $\mathbf{X}$  with independent margins uniformly distributed over  $[0, 1]$ . The conditional distribution of the duration  $T$  is lognormal, with  $E[\log T | \mathbf{X} = \mathbf{x}] = \beta_0 + \sum_{j=1}^2 \beta_j X^{(j)}$  and  $Var(\log T | \mathbf{X} = \mathbf{x}) = 1$ . The censoring variable  $C$  follows a Weibull distribution with shape parameter  $1/2$ , that is with survival function

$$\mathbb{P}(C \geq t) = \exp(-\{t/c\}^{1/2}),$$

where the parameter  $c$  is used to change the average proportion of censored observations in the sample (30%, 40%, and 50%).

We consider the case of a two-dimensional loss vector  $\mathbf{L}$ , where the margins are conditionally log-Gamma distributed (that is  $\log L^{(1)}$  and  $\log L^{(2)}$  are Gamma distributed) with rate 2 and shape parameter ( $s_j$  for  $L^j$ ) defined as

$$s_j(\mathbf{x}, t) = \alpha_{0,j} + \sum_{k=1}^2 \alpha_{k,j} X^{(k)} + \alpha_{3,j} \bar{T},$$

where  $\bar{T}$  is a centered standardized version of  $T$ .

We then consider three families of copula functions to describe its dependence structure, namely Clayton, Frank and Gumbel families whose definitions are recalled in Table 1. These three Archimedean families of copula (see Nelsen (2006) for the definition of the Archimedean class) present some interesting features for this comparison:

Family	Copula function	$\Theta$	$\lambda_L$	$\lambda_U$
Clayton	$\mathfrak{C}_\theta(u, v) = [\max\{u^{-\theta} + v^{-\theta} - 1; 0\}]^{-1/\theta}$	$[-1; \infty) - \{0\}$	$2^{-1/\theta}$	0
Frank	$\mathfrak{C}_\theta(u, v) = -\frac{1}{\theta} \log \left[ 1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1} \right]$	$\mathbb{R} - \{0\}$	0	0
Gumbel	$\mathfrak{C}_\theta(u, v) = \exp \left[ - \left( (-\log u)^\theta + (-\log v)^\theta \right)^{1/\theta} \right]$	$[1, \infty)$	0	$2 - 2^{1/\theta}$

Table 1: Copula families used in the simulation setting.

Clayton family presents lower tail dependence, that is, for  $(U, V)$  with Clayton c.d.f.,  $\lambda_L = \lim_{u \rightarrow 0} \mathbb{P}(U \leq u | V \leq u) > 0$ ; Gumbel has upper tail dependence, that is for, for  $(U, V)$  with Gumbel c.d.f.,  $\lambda_U = \lim_{u \rightarrow 1} \mathbb{P}(U \geq u | V \geq u) > 0$ ; while Frank's family is such that  $\lambda_U = \lambda_L = 0$ .

Regarding the target function  $\theta(\mathbf{x}, t)$ , we want to make things comparable between the different families, therefore we need to take into account the fact that the meaning of the association parameter is not the same from one family to another. Therefore, we consider a common value for the conditional Kendall's tau coefficient, rather than for the association parameter directly. This dependence measure is defined as

$$\begin{aligned} \tau(\mathbf{x}, t) &= \mathbb{P} \left( [L_1^{(1)} - L_2^{(1)}][L_1^{(2)} - L_2^{(2)}] > 0 | \mathbf{X} = \mathbf{x}, T = t \right) \\ &\quad - \mathbb{P} \left( [L_1^{(1)} - L_2^{(1)}][L_1^{(2)} - L_2^{(2)}] < 0 | \mathbf{X} = \mathbf{x}, T = t \right). \end{aligned}$$

We set

$$\text{logit}\{\tau(\mathbf{x}, t)\} = \gamma_0 + \sum_{j=1}^2 \gamma_j X^{(j)} + \gamma_3 \bar{T}.$$

Then we compute, for each family of copula the corresponding  $\theta(\mathbf{x}, t)$  since a 1-1 correspondence between these two quantities exist in the families we consider. The values of the parameters used to simulate the sample are gathered in Table 2.

For a simulated sample, we estimate  $\hat{\theta}(\mathbf{x}, t)$  for different value of the smoothing parameter, and, for each estimator, we measure the error based on the following criteria:

1. an empirical measure linked to the ability to estimate the copula function, that is

$$E_c = \left\{ \sum_{i=1}^n W_{i,n} \left( \mathfrak{C}_{\hat{\theta}(\mathbf{x}, t)}(\hat{\mathbf{U}}_i) - \mathfrak{C}_{\theta(\mathbf{X}_i, Y_i)}(\hat{\mathbf{U}}_i) \right)^2 \right\}^{1/2};$$

2. a way to measure the difference between the corresponding Kendall's tau, in order to get something which can be comparable from one family to another,

$$E_\tau = \left\{ \sum_{i=1}^n W_{i,n} \left( \hat{\tau}(\mathbf{X}_i, Y_i) - \tau(\mathbf{X}_i, Y_i) \right)^2 \right\}^{1/2}.$$

Parameter	Value	Parameter	Value
$\alpha_{0,1}$	0.2	$\alpha_{0,2}$	0.2
$\alpha_{1,1}$	0.3	$\alpha_{1,2}$	0.3
$\alpha_{2,1}$	0.1	$\alpha_{2,2}$	0.1
$\alpha_{3,1}$	0.6	$\alpha_{3,2}$	0.8
$\beta_0$	0.03	$\gamma_0$	-3
$\beta_1$	0.04	$\gamma_1$	-1
$\beta$	-0.03	$\gamma_2$	1.5
		$\gamma_3$	2

Table 2: Value of the parameters used in the simulation setting.

On the other hand, the final aim is to estimate the amount of reserve. For each claim that is open (that is with  $\delta_i = 0$ ), we develop the technique of section 2.5. Let  $\hat{L}_i^{(j)}$  denote the prediction of the  $j$ -th component of the loss for observation  $i$ . Next, we define  $\hat{R}^{(j)} = \sum_{i=1}^n (1 - \delta_i) \hat{L}_i^{(j)}$ , that is the estimation of the reserve for the  $j$ -th component of the loss, and  $R^{(j)} = \sum_{i=1}^n (1 - \delta_i) L_i^{(j)}$  is the true value.

To assess the quality of the reserve estimation, we compute

$$E_R = \frac{\hat{R}^{(1)} + \hat{R}^{(2)} - R^{(1)} - R^{(2)}}{R^{(1)} + R^{(2)}}.$$

Regarding smoothing, we consider a Gaussian kernel and a grid of bandwidth  $h \in \{1, 1.1, \dots, 2\}$ . The average errors are reported in Table 3 to 5. These average errors show a relative stability for the considered set of bandwidths. The percentage of censoring generally diminishes the performance to estimate Kendall's tau coefficient and the copula function. The results are quite similar from one copula family to another in this example. The impact of the copula family appears more clearly when it comes to considering the reserve. Under Clayton and Frank copula models, the absence of upper tail dependence makes the estimation of the reserve less volatile compared to Gumbel. Let us note that an increase of the percentage of censoring here makes the error  $E_R$  decrease. This is essentially caused by the fact that the number of claims to evaluate becomes higher when censoring increases. Hence, the individual errors are more absorbed by the more important number of claims to evaluate. In Figure 1 and 2, we display some boxplots for a selected number of configurations. The error on copula estimation is shown in Figure 1 while Figure 2 shows the error for the reserve. For the sake of brevity, boxplots for Kendall's tau coefficient are not shown, since they give a similar picture as the one provided for

the error  $E_c$ . We can see from these boxplots that the dispersion of the error over the simulated sample stays relatively stable.

Bandwidth	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
Frank 30%, n=500	0.0081	0.0081	0.0082	0.0082	0.0082	0.0082	0.0082	0.0082	0.0082	0.0082	0.0083
Frank 40%, n=500	0.0088	0.0088	0.0088	0.0089	0.0089	0.0089	0.0089	0.0089	0.0089	0.0089	0.0090
Frank 50%, n=500	0.0084	0.0084	0.0085	0.0085	0.0085	0.0086	0.0086	0.0086	0.0086	0.0086	0.0086
Clayton 30%, n=500	0.0076	0.0077	0.0077	0.0077	0.0078	0.0078	0.0078	0.0078	0.0078	0.0078	0.0079
Clayton 40%, n=500	0.0078	0.0078	0.0078	0.0079	0.0079	0.0079	0.0079	0.0079	0.0079	0.0079	0.0079
Clayton 50%, n=500	0.0082	0.0083	0.0083	0.0083	0.0083	0.0084	0.0084	0.0084	0.0084	0.0084	0.0084
Gumbel 30%, n=500	0.0082	0.0082	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0083	0.0084
Gumbel 40%, n=500	0.0078	0.0079	0.0079	0.0080	0.0080	0.0080	0.0080	0.0080	0.0081	0.0081	0.0081
Gumbel 50%, n=500	0.0081	0.0081	0.0081	0.0082	0.0082	0.0082	0.0082	0.0082	0.0083	0.0083	0.0083
Frank 30%, n=1000	0.0067	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.0069	0.0069	0.0069	0.0069
Frank 40%, n=1000	0.0070	0.0070	0.0071	0.0071	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072
Frank 50%, n=1000	0.0071	0.0072	0.0072	0.0072	0.0073	0.0073	0.0073	0.0073	0.0073	0.0074	0.0074
Clayton 30%, n=1000	0.0066	0.0067	0.0067	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068
Clayton 40%, n=1000	0.0069	0.0070	0.0071	0.0071	0.0071	0.0072	0.0072	0.0072	0.0072	0.0072	0.0072
Clayton 50%, n=1000	0.0069	0.0070	0.0070	0.0071	0.0071	0.0071	0.0071	0.0071	0.0071	0.0071	0.0072
Gumbel 30%, n=1000	0.0066	0.0067	0.0067	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.0068	0.0069
Gumbel 40%, n=1000	0.0067	0.0068	0.0068	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0070
Gumbel 50%, n=1000	0.0072	0.0072	0.0072	0.0072	0.0073	0.0073	0.0073	0.0073	0.0074	0.0074	0.0074

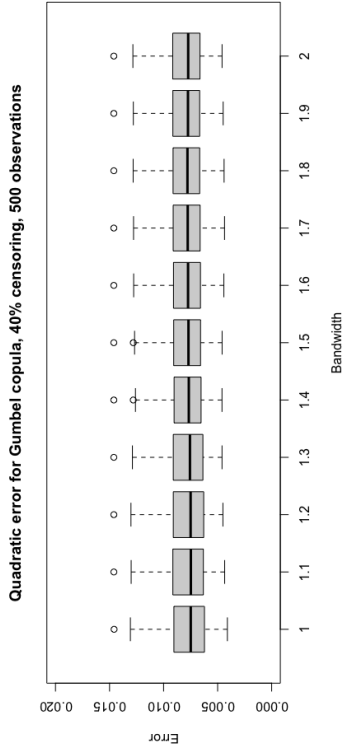
Table 3: Average error  $E_c$  over the 100 simulations, for different values of bandwidth, copula families, and proportion of censoring.

Bandwidth	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
Frank 30%, n=500	0.0540	0.0542	0.0545	0.0547	0.0548	0.0548	0.0548	0.0548	0.0549	0.0550	0.0550
Frank 40%, n=500	0.0576	0.0579	0.0581	0.0583	0.0584	0.0586	0.0587	0.0588	0.0588	0.0588	0.0589
Frank 50%, n=500	0.0552	0.0555	0.0557	0.0560	0.0562	0.0563	0.0565	0.0566	0.0567	0.0568	0.0569
Clayton 30%, n=500	0.0513	0.0516	0.0519	0.0521	0.0523	0.0523	0.0524	0.0525	0.0526	0.0527	0.0529
Clayton 40%, n=500	0.0524	0.0527	0.0529	0.0531	0.0533	0.0534	0.0534	0.0534	0.0535	0.0535	0.0536
Clayton 50%, n=500	0.0556	0.0559	0.0561	0.0563	0.0564	0.0565	0.0566	0.0568	0.0569	0.0571	0.0571
Gumbel 30%, n=500	0.0549	0.0553	0.0555	0.0557	0.0557	0.0557	0.0557	0.0557	0.0558	0.0560	0.0561
Gumbel 40%, n=500	0.0533	0.0536	0.0539	0.0542	0.0544	0.0546	0.0547	0.0547	0.0548	0.0549	0.0549
Gumbel 50%, n=500	0.0542	0.0545	0.0548	0.0550	0.0551	0.0552	0.0553	0.0554	0.0555	0.0556	0.0557
Frank 30%, n=1000	0.0447	0.0450	0.0452	0.0454	0.0455	0.0455	0.0455	0.0456	0.0456	0.0458	0.0458
Frank 40%, n=1000	0.0459	0.0464	0.0468	0.0472	0.0475	0.0476	0.0477	0.0478	0.0478	0.0478	0.0479
Frank 50%, n=1000	0.0469	0.0473	0.0477	0.0479	0.0480	0.0482	0.0483	0.0484	0.0486	0.0487	0.0488
Clayton 30%, n=1000	0.0446	0.0450	0.0452	0.0454	0.0455	0.0455	0.0456	0.0456	0.0457	0.0458	0.0460
Clayton 40%, n=1000	0.0464	0.0468	0.0472	0.0475	0.0478	0.0480	0.0481	0.0482	0.0482	0.0483	0.0483
Clayton 50%, n=1000	0.0469	0.0473	0.0475	0.0477	0.0479	0.0480	0.0481	0.0482	0.0483	0.0485	0.0485
Gumbel 30%, n=1000	0.0451	0.0455	0.0459	0.0460	0.0460	0.0460	0.0461	0.0462	0.0463	0.0465	0.0467
Gumbel 40%, n=1000	0.0454	0.0457	0.0460	0.0463	0.0466	0.0467	0.0468	0.0469	0.0469	0.0469	0.0470
Gumbel 50%, n=1000	0.0483	0.0487	0.0488	0.0490	0.0491	0.0493	0.0494	0.0496	0.0498	0.0498	0.0499

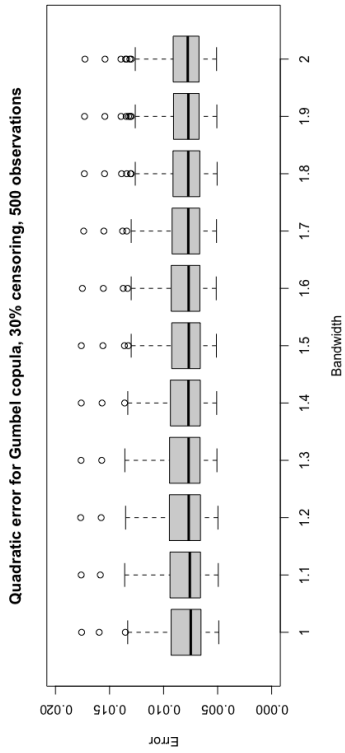
Table 4: Average error  $E_\tau$  over the 100 simulations, for different values of bandwidth, copula families, and proportion of censoring.

Bandwidth	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
Frank 30%, n=500	-0.0025	-0.0030	-0.0040	-0.0031	-0.0035	-0.0022	-0.0032	-0.0032	-0.0037	-0.0028	-0.0020
Frank 40%, n=500	-0.0240	-0.0220	-0.0214	-0.0224	-0.0171	-0.0196	-0.0192	-0.0210	-0.0187	-0.0191	-0.0165
Frank 50%, n=500	-0.0137	-0.0132	-0.0131	-0.0130	-0.0129	-0.0134	-0.0125	-0.0129	-0.0118	-0.0127	-0.0118
Clayton 30%, n=500	-0.0109	-0.0123	-0.0139	-0.0130	-0.0147	-0.0115	-0.0111	-0.0148	-0.0122	-0.0118	-0.0172
Clayton 40%, n=500	-0.0038	-0.0047	-0.0044	-0.0030	-0.0041	-0.0045	-0.0040	-0.0031	-0.0040	-0.0037	-0.0052
Clayton 50%, n=500	-0.0034	-0.0056	-0.0050	-0.0047	-0.0044	-0.0036	-0.0041	-0.0047	-0.0050	-0.0038	-0.0032
Gumbel 30%, n=500	-0.0048	-0.0038	-0.0028	-0.0022	-0.0029	-0.0026	-0.0033	-0.0032	-0.0046	-0.0050	-0.0030
Gumbel 40%, n=500	-0.4867	-0.4222	-0.4895	-0.5568	-0.5090	-0.4131	-0.8108	-0.4890	-0.4632	-0.8127	-0.8134
Gumbel 50%, n=500	-0.1425	-0.0111	-0.0193	-0.0144	-0.0175	-0.0184	-0.0140	-0.0455	-0.0151	-0.0174	-0.0154
Frank 30%, n=1000	-0.0007	-0.0014	-0.0004	-0.0008	-0.0006	-0.0014	-0.0013	-0.0017	-0.0010	-0.0015	-0.0003
Frank 40%, n=1000	-0.0132	-0.0115	-0.0138	-0.0136	-0.0119	-0.0141	-0.0131	-0.0130	-0.0131	-0.0126	-0.0134
Frank 50%, n=1000	-0.0107	-0.0158	-0.0139	-0.0131	-0.0139	-0.0114	-0.0175	-0.0289	-0.0160	-0.0102	-0.0113
Clayton 30%, n=1000	-0.0031	-0.0024	-0.0017	-0.0009	-0.0047	-0.0023	-0.0022	-0.0017	-0.0023	-0.0032	-0.0035
Clayton 40%, n=1000	-0.0082	-0.0080	-0.0058	-0.0551	-0.0570	-0.0085	-0.0182	-0.0210	-0.0358	-0.0150	-0.0157
Clayton 50%, n=1000	-0.0082	-0.0105	-0.0085	-0.0143	-0.0439	-0.0089	-0.0068	-0.0054	-0.0224	-0.0081	-0.0117
Gumbel 30%, n=1000	-0.2684	-0.0409	-0.1381	-0.0808	-0.0345	-0.1033	-0.0645	-0.0337	-0.0384	-0.1303	-0.0381
Gumbel 40%, n=1000	-0.1059	-0.0630	-0.0601	-0.0590	-0.0132	-0.0494	-0.1588	-0.0412	-0.0422	-0.0332	-0.0314
Gumbel 50%, n=1000	-0.0174	-0.0102	-0.0159	-0.0215	-0.0620	-0.0567	-0.0140	-0.0098	-0.0124	-0.0088	-0.0313

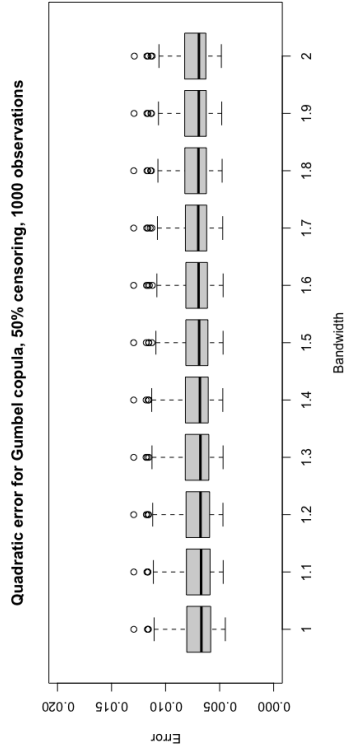
Table 5: Average error  $E_R$  over the 100 simulations, for different values of bandwidth, copula families, and proportion of censoring.



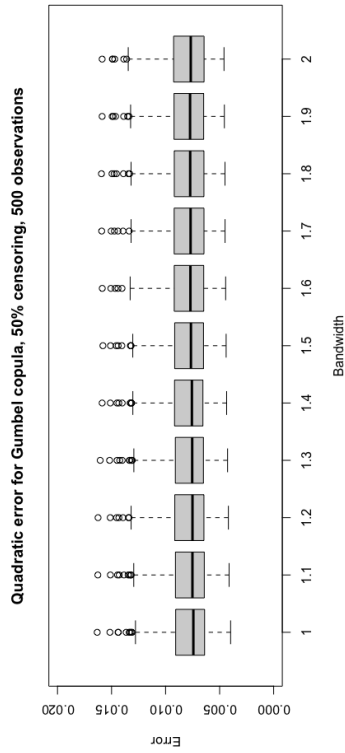
(a) 30% censoring and 500 observations.



(b) 40% censoring and 500 observations.



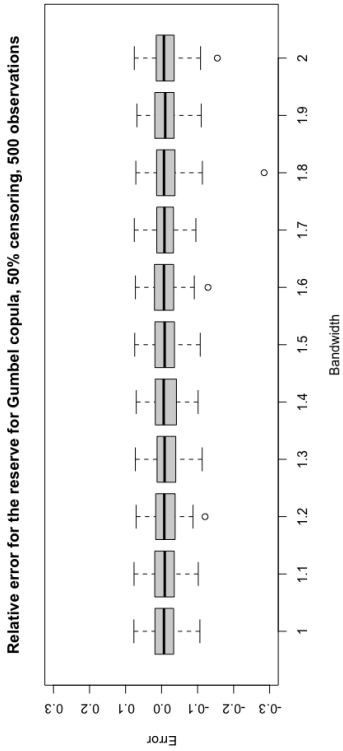
(c) 50% censoring and 500 observations.



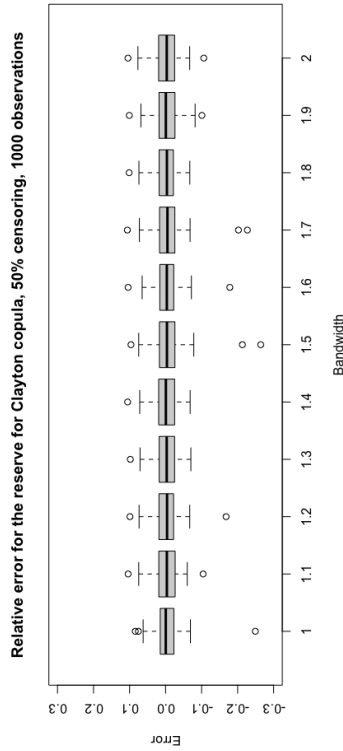
(d) 50% censoring and 1000 observations.

Figure 1: Error  $E_c$  for Gumbel copula, for different intensity of censoring, size of data sample, and values of the bandwidth.

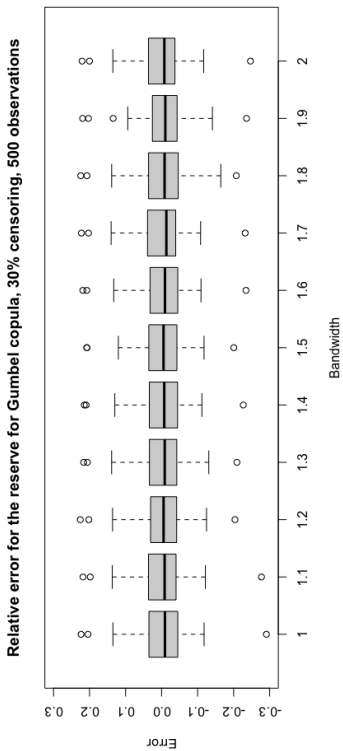




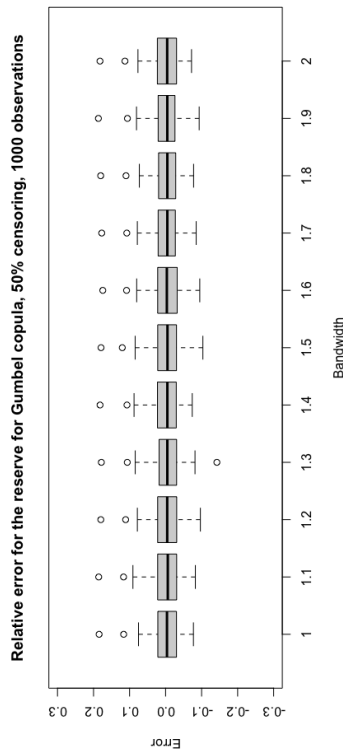
(a) Gumbel, 30% censoring and 500 observations.



(b) Gumbel, 50% censoring and 500 observations.



(c) Gumbel, 50% censoring and 1000 observations.



(d) Clayton, 50% censoring and 1000 observations.

Figure 2: Error  $E_R$  in different configurations, and for different values of the bandwidth.

	$X_1$	$X_2$	$X_3$	$X_4$	$T$	$L^{(1)}$	$L^{(2)}$
Mean	176	43	56 882	9 944	744	163 397	38 379
$q_{0.25}$	1	38	5 000	115	418	25 000	2 000
$q_{0.5}$	6	43	14 875	2 500	682	60 000	12 843
$q_{0.75}$	133	49	42 375	10 000	1009	175 500	41 000

Table 6: Descriptive statistics on closed claims. The notation  $q_\alpha$  refers to the  $\alpha$ -th empirical quantile.

## 4.2 Real data example

For this real data analysis, we consider a dataset from the Texas Department of Insurance on medical malpractice claims. This dataset gathers 40 868 injury claims, closed between 2007 and 2012. This dataset comes from the annual reports of the Texas Department of Insurance (<https://www.tdi.texas.gov/>). Additionally, the dataset for the claims that were closed in 2007 is also available at <https://doi.org/10.7910/DVN/2ZX2XS>. We consider the cost of each claim  $L_{(tot,i)}$  to be composed of the indemnity for the injured party,  $L_i^{(1)}$ , and the expenses linked to justice expenses,  $L_i^{(2)}$ . For simplicity, we focus on claims with impacts on both margins, resulting to 20 634 claims. For a global approach, the following models may be plugged to a mixture model, in order to incorporate models of claims with univariate costs.

As already mentioned, the database is composed of closed claims. To put ourselves in a situation where we need to predict open claims, we look at the status of a claim at 31/12/2010. Claims that occurred after this date are out of the scope of our analysis, and claims that occurred before 31/12/2010 with a settlement between 2010 and 2012 are considered as censored. This leaves 15 717 closed claims and 4 917 open claims, with the aim to assess the costs of the last category.

We rely on 4 covariates to predict open claims costs : the delay between the injury and the insurance report  $X^{(1)}$ , the age of the injured party  $X^{(2)}$ , the initial estimation of the indemnity  $X^{(3)}$ , and the initial estimation of the expenses  $X^{(4)}$ . Descriptive statistics of the dataset are given in Tables 7 and 6. As expected, the closing delay  $T$  is globally inferior on closed claims than on open claims. The distributions of covariates  $X^{(1)}$  and  $X^{(2)}$  are quite similar whereas the first estimations, as the final costs, are higher for closed claims than open ones. In addition, the Kendall's tau between  $L^{(1)}$  and  $L^{(2)}$  is higher on closed claims (0.40) than for open claims (0.30).

	$X_1$	$X_2$	$X_3$	$X_4$	$Y$	$T$	$L^{(1)}$	$L^{(2)}$
Mean	169	44	68 297	11 880	598	933	221 684	53 486
$q_{0.25}$	1	40	6 000	103	254	621	50 000	7 000
$q_{0.5}$	6	43	16 000	5 000	511	857	100 000	22 698
$q_{0.75}$	138	49	50 000	12 993	844	1167	247 500	56 000

Table 7: Descriptive statistics on open claims. The informations of  $T$ ,  $L^{(1)}$  and  $L^{(2)}$  are assumed to be unknown on the 31/12/2010.

Regarding the regression models for the margins, the covariates are standardized (that is, for  $j = 1, \dots, 4$  are replaced by  $\tilde{X}_j = (X_j - \bar{X}_j)/\hat{\sigma}_{X_j}$ , where  $\bar{X}_i$  is the empirical mean of covariate  $i$  among the whole database, and  $\hat{\sigma}_{X_j}$  the corresponding empirical standard deviation) to avoid scale effects in fitting the coefficients of the models. The variable  $T$  is also normalized ( $\tilde{T}$  denotes the normalized variable), but the empirical mean and standard deviation are taken among the closed claims. The normalized vector  $\mathbf{Z}$  (resp.  $\mathbf{X}$ ) is denoted as  $\tilde{\mathbf{Z}}$  (resp.  $\tilde{\mathbf{X}}$ ).

For the distribution of the margins, we consider:

- for  $L^{(1)}$  a Generalized Linear model with Gamma distribution and logarithmic link function, that is the density of  $L^{(1)}|\tilde{\mathbf{Z}} = \mathbf{z}$  is

$$f_{\kappa,\lambda}(l) = \Gamma(\kappa)^{-1} l^{\kappa-1} \mu^{-\kappa} \exp(-l/\mu),$$

with  $\kappa = \kappa_1$  and

$$E \left[ \log L^{(1)} | \tilde{\mathbf{Z}} = \mathbf{z} \right] = \alpha_1 + \sum_{k=1}^4 \alpha_k x^{(k)} + \alpha_5 t;$$

- for  $L^{(2)}$  a Weibull regression model, that is  $\log L^{(2)}|\tilde{\mathbf{Z}} = \mathbf{z}$  is distributed according to the density

$$f_{\kappa,\lambda}(l) = \kappa l^{\kappa-1} \lambda(\mathbf{z})^{-\kappa} \exp(-(l/\lambda(\mathbf{z}))^\kappa), \quad (4.1)$$

with  $\kappa = \kappa_2$ , and

$$\lambda_L(\mathbf{z}) = \frac{1}{\Gamma(1 + 1/\kappa_2)} \left( \alpha_{0,2} + \sum_{k=1}^4 \alpha_{k,2} x^{(k)} + \alpha_{4,2} t \right).$$

To predict the final value of open claims, we also need a model on  $T|\tilde{\mathbf{X}}$ . We consider that  $T|\tilde{\mathbf{X}} = \mathbf{x}$  is Weibull distributed (density defined in (4.1)) with  $\kappa = \kappa_T$  fixed, and

$$\lambda_T(\mathbf{x}) = \frac{1}{\Gamma(1 + 1/\kappa_T)} \left( \alpha_0^T + \sum_{k=1}^4 \alpha_k^T x^{(k)} \right).$$

$L^{(1)}$	$\alpha_{0,1}$	3.88	$L^{(2)}$	$\alpha_{0,2}$	3.65	$T$	$\alpha_0^T$	6.91
	$\alpha_{1,1}$	-0.96		$\alpha_{1,2}$	0.34		$\alpha_1^T$	0.06
	$\alpha_{2,1}$	-0.04		$\alpha_{2,2}$	-0.01		$\alpha_2^T$	-0.01
	$\alpha_{3,1}$	0.20		$\alpha_{3,2}$	0.19		$\alpha_3^T$	-0.01
	$\alpha_{4,1}$	0.11		$\alpha_{4,2}$	0.82		$\alpha_4^T$	0.01
	$\alpha_{5,1}$	0.13		$\alpha_{5,2}$	0			
	$\kappa_1$	0.2		$\kappa_2$	1.25		$\kappa_T$	0.59

Table 8: Estimated parameters for the estimation of the marginal distributions.

The estimated parameters, given in Table 8, allow to obtain bivariate pseudo observations of  $(L^{(1)}, L^{(2)})$  illustrated in Figure 3.

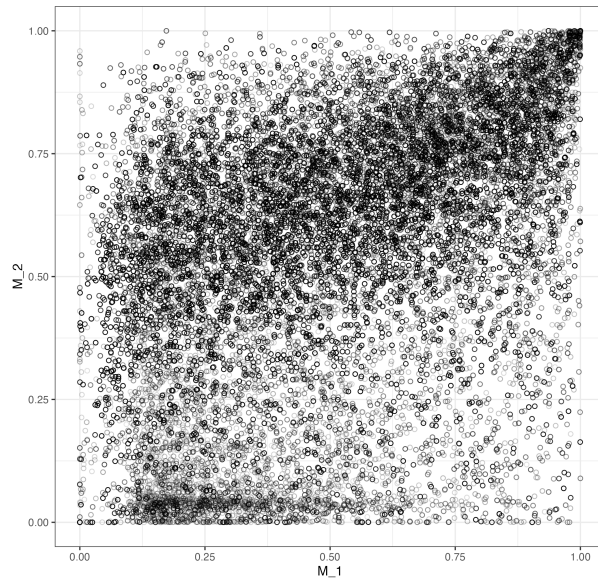


Figure 3: Scatter plot of pseudo observations  $(U^{(1)}, U^{(2)})$ . The pseudo observations with high weight appear darker.

We consider a Gaussian kernel. To select the bandwidth in our kernel estimator, we rely on 3-fold cross validation techniques, using a weighted log-likelihood as a performance measure. That is, for an estimator  $\hat{\theta}(\mathbf{z})$  in a given copula family, we compute

$$\frac{\sum_{i \in \mathcal{I}} W_{i,n} \log c_{\hat{\theta}(\mathbf{z}_i)}(\hat{U}_i)}{\sum_{i \in \mathcal{I}} W_{i,n}},$$

as a measure on the fold identified to the set of indexes  $\mathcal{I}$ . The results, for different bandwidths, are shown in Table 9.

	h=0.05	h=0.1	h=0.15	h=0.2
Clayton	0.0235	0.0242	0.0143	0.0087
Frank	0.0506	0.0732	0.0118	0.0103
Gumbel	0.0008	0.0013	0.0011	0.0096

Table 9: 3-fold cross validation performances, in terms of log-likelihood, depending on the parameter  $h$  of the Kernel  $K$ . For each copula family, the highest performance is in bold, and lead to parameters used in Section 4.2 for semi-parametric copula estimation.

In the following, we therefore keep the value  $h = 0.1$  for the smoothing parameter, which gives the best results for every copula family that we considered. We observe that, according to this criterion, Frank’s copula provides the best fit.

Then, based on the conditional copula model and the prediction  $\hat{T}$  for each open claim, we simulate 1000 replications of the estimated conditional distribution of open claims. Gathering the predictions for all claims gives a distribution for the reserve to constitute, that can be compared to the true value of these claims, which is approximately 1,340 M USD. Figure 4 shows the histograms of these three simulated distributions. We see that, for the aggregated amounts, the three models gives relatively similar results. Table 10 also provides the mean and median of these distributions. All models tend to overestimate the amount of reserve, with a slight advantage to Frank’s copula model (a focus on this model is given in Figure 5). One can compare this approach with the one that does not take into account the dependence structure between the two lines. In this approach, these two lines are modeled separately according to the previous marginal models that we mention. The average predicted amount is then 1,375 M USD. One can observe that taking into account the dependence structure significantly contributes to make the predicted amount close to the target.

Copula	Mean	Median
Clayton	1,341.9	1,341.7
Frank	1,341.8	1,341.1
Gumbel	1,341.9	1,341.6

Table 10: Mean and median of the simulated distribution according to the three conditional copula model we considered. (Amounts in M USD)

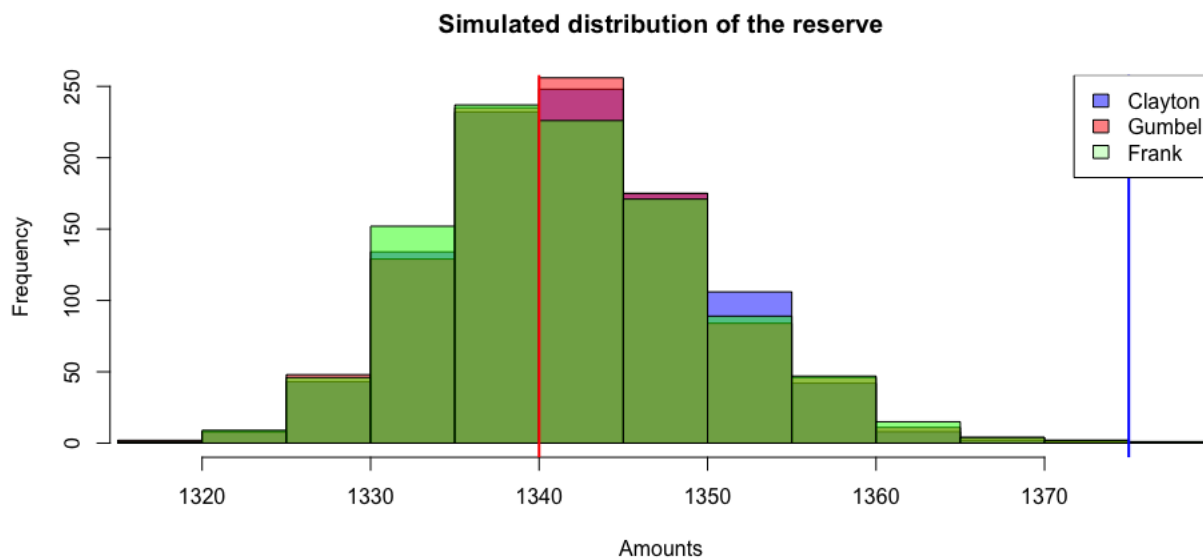


Figure 4: Aggregated simulated amount of open claims (M USD), adding the amounts for the two lines of business. The red line shows the true value (1,340). The blue line shows the amount predicted without considering the dependence structure between the two lines (1,375).

## 5 Conclusion

In this paper, we considered the case of analyzing the cost of claims in the case where this cost is decomposed in several lines of payment. In the framework we develop, the interaction between these lines of payment is modeled using copulas. Moreover, we take into account the link between the "lifetime of the claim" (time before settlement) and its final cost, with the correction of the bias caused by censoring. This decomposition can be useful to better understand what drives the severity of a claim, and potentially to improve risk management by winning on some lines (like legal fees or expert costs) that can be more easily controlled. Let us also mention another possible application to the proper design of insurance products, through the example of cyber insurance. Due to the variety of situations in cyber risk (data theft, business interruption, legal dispute after a data breach, corporal prejudice...), an important question is to define which will be covered or not by an insurance product. Romanosky et al. (2019) show some analysis of some products available on the US market, showing that their nature may be strongly differ. Carefully analyzing the different components of the loss is essential to better understand

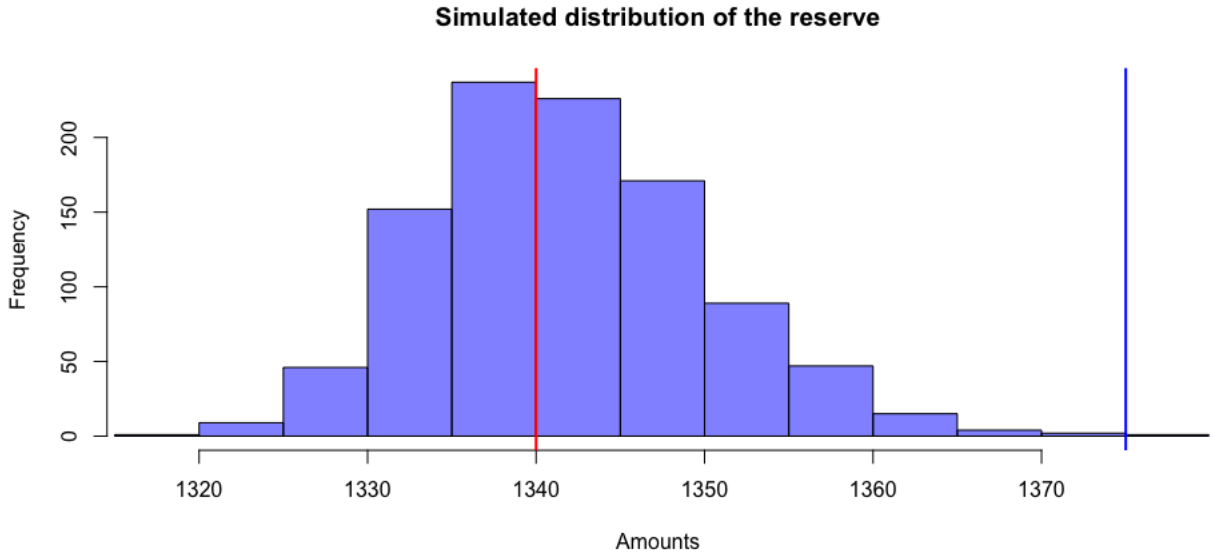


Figure 5: Aggregated simulated amount of open claims (M USD) using a Frank conditional copula model.

which (and at which level) some types of losses can be covered or not.

## 6 Appendix

The Appendix section is organized as follows. In Section 6.1, we show the convergence of the bias term of our kernel estimator. Preliminary results to show Theorem 3.1 are gathered in Section 6.2 (convergence results) and 6.3 (rate of convergence). The proof of Theorem 3.1 is then given in Section 6.4. Section 6.5 show the convergence rate for the margins in the most delicate case where the estimator that is used is the nonparametric estimator (2.5).

### 6.1 Proof of Proposition 3.2

Let

$$M(\theta, \mathbf{z}) = \frac{1}{h^{p+1}} E \left[ \log \mathbf{c}_\theta(\mathbf{U}) K \left( \frac{\mathbf{Z} - \mathbf{z}}{h} \right) \right].$$

We have  $\theta_h^*(\mathbf{z}) = \arg \max_\theta M(\theta, \mathbf{z})$ . Since  $\nabla_\theta M(\theta(\mathbf{z}), \mathbf{z}) = 0$ , a second order Taylor expansion yields

$$\theta_h^*(\mathbf{z}) - \theta(\mathbf{z}) = -\nabla_\theta^2 M(\tilde{\theta}, \mathbf{z})^{-1} \nabla_\theta M(\theta(\mathbf{z}), \mathbf{z}), \quad (6.1)$$

where each component of  $\tilde{\theta}$  is between the corresponding components of  $\theta_h^*(\mathbf{z})$  and  $\theta(\mathbf{z})$ .

From Lemma 6.6,

$$\nabla_{\theta} M(\theta(\mathbf{z}), \mathbf{z}) = O(h^2).$$

On the other hand, from dominated convergence,

$$\nabla_{\tilde{\theta}}^2 M(\tilde{\theta}, \mathbf{z}) = \nabla_{\theta}^2 M(\theta(\mathbf{z}), \mathbf{z}) + o(1),$$

and, from Lemma 6.6,

$$\nabla_{\theta}^2 M(\theta(\mathbf{z}), \mathbf{z}) = \Sigma(\mathbf{z}).$$

## 6.2 Technical results for convergence of the estimator

**Proposition 6.1** *Assume that  $\sup_{i=1, \dots, n} \|\hat{\mathbf{U}}_i - \mathbf{U}_i\| = o_P(1)$ . Consider a class of functions  $\mathcal{F}$  on  $[0, 1]^d$  such that*

$$\forall \phi \in \mathcal{F}, \forall \mathbf{u} \in [0, 1]^d, \|\nabla_{\mathbf{u}} \phi(\mathbf{u})\| \leq \Phi(\mathbf{u}), \quad (6.2)$$

where  $\Phi$  is continuous on  $(0, 1)^d$ , and  $\nabla_{\mathbf{u}}$  denotes the gradient vector with respect to the arguments of the function. Moreover, assume that

$$E[\Phi(\mathbf{U})] < \infty. \quad (6.3)$$

Then,

$$\sup_{\phi \in \mathcal{F}} \left| \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \{\phi(\mathbf{U}_i) - \phi(\hat{\mathbf{U}}_i)\} \right| = o_P(1).$$

**Proof.** We have

$$\begin{aligned} \left| \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \{\phi(\mathbf{U}_i) - \phi(\hat{\mathbf{U}}_i)\} \right| &\leq \sup_{i=1, \dots, n} \|\mathbf{U}_i - \hat{\mathbf{U}}_i\| \\ &\quad \times \left| \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} \Phi(\tilde{\mathbf{U}}_i) |K|\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \right|, \end{aligned}$$

where each component of  $\tilde{\mathbf{U}}_i$  is between the corresponding component of  $\mathbf{U}_i$  and  $\hat{\mathbf{U}}_i$ . Since  $\sup_{i=1, \dots, n} \|\hat{\mathbf{U}}_i - \mathbf{U}_i\| = o_P(1)$ ,  $\tilde{\mathbf{U}}_i$  can be replaced by  $\mathbf{U}_i$  in the right-hand side of the last equation. Hence, the proof will be complete if we show that

$$\left| \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} \Phi(\mathbf{U}_i) |K|\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \right| = O_P(1).$$



This can be done by observing that

$$\left| \frac{W_{i,n}}{W_{i,n}^*} \right| \leq \sup_{t \leq Y_{(n)}} \frac{S^C(t)}{\hat{S}^C(t)} = O_P(1),$$

where  $Y_{(n)}$  denotes the largest observation of the sample  $(Y_1, \dots, Y_n)$  (from Lemma 2.6 in Gill (1983)). This implies that

$$\left| \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} \Phi(\mathbf{U}_i) |K| \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \right| \leq O_P(1) \times \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n}^* \Phi(\mathbf{U}_i) |K| \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right),$$

and the second term on the right hand side has finite expectation. ■

**Proposition 6.2** Consider a class of functions  $\mathcal{F}$  on  $[0, 1]^d$  such that

$$\forall \phi \in \mathcal{F}, \forall \mathbf{u} \in [0, 1]^d, |\phi(\mathbf{u})| \leq \Phi(\mathbf{u}),$$

with

$$E \left[ \frac{\Phi(\mathbf{U})}{S^C(T)^\iota} \right] < \infty.$$

Then,

$$\sup_{\phi \in \mathcal{F}} \left| \frac{1}{h^{p+1}} \sum_{i=1}^n [W_{i,n} - W_{i,n}^*] K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \phi(\mathbf{U}_i) \right| = o_P(1).$$

**Proof.** Write

$$|W_{i,n} - W_{i,n}^*| = |\hat{S}^C(Y_i) - S^C(Y_i)|^\iota \times \left[ \frac{|\hat{S}^C(Y_i) - S^C(Y_i)|}{\hat{S}^C(Y_i)} \right]^{1-\iota} \times \left[ \frac{S^C(Y_i)}{\hat{S}^C(Y_i)} \right]^\iota \times \frac{\delta_i}{n S^C(Y_i)^{1+\iota}}.$$

From Stute and Wang (1993), the first term on the right hand side tends to zero, while, from (Gill (1983)), the next two terms are  $O_P(1)$ . The result then follows from the fact that

$$E \left[ \frac{\delta}{S^C(Y)^{1+\iota}} \Phi(\mathbf{U}) |K| \left( \frac{\mathbf{Z} - \mathbf{z}}{h} \right) \right] = E \left[ \frac{1}{S^C(T)^\iota} \Phi(\mathbf{U}) |K| \left( \frac{\mathbf{Z} - \mathbf{z}}{h} \right) \right].$$

■

### 6.3 Technical results with rates of convergence

**Proposition 6.3** Under Assumption 2 and 3, let  $\phi \in \mathcal{F}_2$  where  $\mathcal{F}_2$  is defined in Assumption 2. Then

$$\left| \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \{ \phi(\mathbf{U}_i) - \phi(\hat{\mathbf{U}}_i) \} \right| = O_P(\varepsilon_n).$$

**Proof.** We have

$$\left| \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \{ \phi(\mathbf{U}_i) - \phi(\hat{\mathbf{U}}_i) \} \right| \leq$$

$$\sup_{i=1, \dots, n, k=1, \dots, n} \left| \frac{U_i^{(k)} - \hat{U}_i^{(k)}}{m(U_i^{(k)})^\alpha} \right| \times \left| \frac{1}{h^{p+1}} \sum_{i=1}^n \sum_{k=1}^d W_{i,n} m(U_i^{(k)})^\alpha \dot{\phi}^{(k)}(\tilde{U}_i^{(k)}) |K| \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \right|,$$

where  $\tilde{U}_i^{(k)}$  is between  $U_i^{(k)}$  and  $\hat{U}_i^{(k)}$ . From Assumption 3, we get the proper rate, we then need to show that the last term in the previous equation is  $O_P(1)$ . To obtain this, first replace  $W_{i,n}$  by  $W_{i,n}^*$  by observing that, as in the proof of Proposition 6.1, the ratio of these to weight is uniformly bounded by an  $O_P(1)$  term. Moreover, define, for  $L \geq 0$ ,

$$E_{L,n} = \bigcap_{k=1}^d \left\{ \inf_{1 \leq i \leq n} \inf \left( \frac{\hat{U}_i^{(k)}}{U_i^{(k)}}, \frac{1 - \hat{U}_i^{(k)}}{1 - U_i^{(k)}} \right) \geq \frac{1}{L} \right\}.$$

On the event  $E_{L,n}$ , from Assumption 2,

$$m(U_i^{(k)})^\alpha \dot{\phi}^{(k)}(\tilde{U}_i^{(k)}) \leq \frac{m(U_i^{(k)})^\alpha}{m(\tilde{U}_i^{(k)})^\alpha} \times \left( \sum_{k=1}^d \frac{L^\beta A_2}{m(U_i^{(k)})^\beta} \right)$$

$$\leq L^{\alpha+\beta} \sum_{k=1}^d \frac{A_2}{m(U_i^{(k)})^\beta}.$$

The result then follows from the fact that

$$nE \left[ W_{i,n}^* \sum_{k=1}^d \frac{1}{m(U_i^{(k)})} \right] = E \left[ \frac{\delta_i}{S^C(Y_i)} \sum_{k=1}^d \frac{1}{m(U_i^{(k)})^\beta} \right] = E \left[ \sum_{k=1}^d \frac{1}{m(U_i^{(k)})^\beta} \right] < \infty,$$

and that  $\lim_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(E_{L,n}) = 0$  from Assumption 3. ■

**Proposition 6.4** *Let  $\phi$  denote a function such that  $E \left[ \frac{|\phi(\mathbf{U}_i)|}{\mathfrak{K}(T_i)^{1/2+\iota}} \right] < \infty$ . Then, for all  $\phi \in \mathcal{F}_2$ ,*

$$\left| \frac{1}{h^{p+1}} \sum_{i=1}^n (W_{i,n} - W_{i,n}^*) K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \phi(\mathbf{U}_i) \right| = O_P(n^{-1/2}).$$

**Proof.** Decompose

$$|W_{i,n} - W_{i,n}^*| = W_{i,n}^* \frac{|\hat{S}_C(Y_i) - S^C(Y_i)|}{\hat{S}_C(Y_i)}$$

$$= \frac{W_{i,n}^*}{\mathfrak{K}(Y_i)^{1/2+\iota}} \frac{\mathfrak{K}(Y_i)^{1/2+\iota} |\hat{S}_C(Y_i) - S^C(Y_i)|}{\hat{S}_C(Y_i)}$$

$$\leq \frac{W_{i,n}^*}{\mathfrak{K}(Y_i)^{1/2+\iota}} \times \sup_{t \leq Y_{(n)}} \left\{ \mathfrak{K}(t)^{1/2+\iota} \frac{|\hat{S}_C(t) - S^C(t)|}{S^C(t)} \right\} \times \sup_{t \leq Y_{(n)}} \left| \frac{S^C(t)}{\hat{S}_C(t)} \right|.$$

From Theorem 2.1 in Gill (1983), we get

$$\sup_{t \leq Y_{(n)}} \left\{ \mathfrak{K}(t)^{1/2+\iota} \frac{|\hat{S}^C(t) - S^C(t)|}{S^C(t)} \right\} = O_P(n^{-1/2}),$$

and, from Lemma 2.6 in Gill (1983),

$$\sup_{t \leq Y_{(n)}} \left| \frac{S^C(t)}{\hat{S}^C(t)} \right| = O_P(1).$$

This shows that

$$\begin{aligned} & \left| \frac{1}{h^{p+1}} \sum_{i=1}^n (W_{i,n} - W_{i,n}^*) K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \phi(\mathbf{U}_i) \right| \\ & \leq O_P(n^{-1/2}) \times \left( \frac{1}{h^{p+1}} \sum_{i=1}^n \frac{W_{i,n}^*}{\mathfrak{K}(Y_i)^{1/2+\iota}} |K| \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) |\phi(\mathbf{U}_i)| \right). \end{aligned}$$

The term in parenthesis on the right hand side has finite bounded expectation since

$$nE \left[ \frac{W_{i,n}^*}{\mathfrak{K}(Y_i)^{1/2+\iota}} |\phi(\mathbf{U}_i)| \right] = E \left[ \frac{\delta_i |\phi(\mathbf{U}_i)|}{S^C(Y_i) \mathfrak{K}(Y_i)^{1/2+\iota}} \right] = E \left[ \frac{|\phi(\mathbf{U}_i)|}{\mathfrak{K}(T_i)^{1/2+\iota}} \right].$$

■

## 6.4 Proof of Theorem 3.1

Let us define

$$\begin{aligned} \hat{M}_n(\theta, \mathbf{z}) &= \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \log \mathfrak{c}_\theta(\hat{\mathbf{U}}_i), \\ \tilde{M}_n(\theta, \mathbf{z}) &= \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \log \mathfrak{c}_\theta(\mathbf{U}_i), \\ M_n^*(\theta, \mathbf{z}) &= \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n}^* K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \log \mathfrak{c}_\theta(\mathbf{U}_i), \\ M(\theta, \mathbf{z}) &= E \left[ \log \mathfrak{c}_\theta(\mathbf{U}) K \left( \frac{\mathbf{Z} - \mathbf{z}}{h} \right) \right]. \end{aligned}$$

The first step of the proof is to show the consistency of  $\hat{\theta}(\mathbf{z})$ , obtained in Proposition 6.5 below.

**Proposition 6.5** *Under the assumptions of Theorem 3.1,*

$$\hat{\theta}(\mathbf{z}) - \theta_h^*(\mathbf{z}) = o_P(1).$$

**Proof.** We shall get  $\hat{\theta}(\mathbf{z}) - \theta_h^*(\mathbf{z}) = o_P(1)$ , from the fact that

$$\sup_{\theta \in \Theta} |\hat{M}_n(\theta, \mathbf{z}) - M(\theta, \mathbf{z})| = o_P(1). \quad (6.4)$$

From the triangular inequality,

$$|\hat{M}_n(\theta, \mathbf{z}) - M(\theta, \mathbf{z})| \leq |\hat{M}_n(\theta, \mathbf{z}) - \tilde{M}_n(\theta, \mathbf{z})| + |\tilde{M}_n(\theta, \mathbf{z}) - M_n^*(\theta, \mathbf{z})| + |M_n^*(\theta, \mathbf{z}) - M(\theta, \mathbf{z})|.$$

Then, (6.4) will be obtained by showing that each of these three terms tends to zero in probability.

The uniform convergence of the first one is a consequence of Proposition 6.1. Conditions (6.2) and (6.3) correspond to the last part of Assumption 1, and the convergence of the pseudo-observations comes from Assumption 3.

To obtain  $\sup_{\theta \in \Theta} |M_n^*(\theta, \mathbf{z}) - \tilde{M}_n^*(\theta, \mathbf{z})| = o_P(1)$ , we rely on Proposition 6.2, whose assumptions hold thanks to Assumption 1.

Finally,  $\sup_{\theta \in \Theta} |M(\theta, \mathbf{z}) - \tilde{M}_n^*(\theta, \mathbf{z})| = o_P(1)$  is a consequence from Theorem 4 in Einmahl et al. (2005), where we use Assumption 1. ■

To get the convergence rate, we use the fact that, by definition,  $\nabla_{\theta} M_n(\hat{\theta}(\mathbf{z})) = 0$ . Moreover, from a Taylor expansion,

$$\nabla_{\theta} \hat{M}_n(\hat{\theta}(\mathbf{z}), \mathbf{z}) = \nabla_{\theta} \hat{M}_n(\theta_h^*(\mathbf{z}), \mathbf{z}) + \nabla_{\theta}^2 \hat{M}_n(\tilde{\theta}(\mathbf{z})) (\hat{\theta}(\mathbf{z}) - \theta_h^*(\mathbf{z})),$$

where each component of  $\tilde{\theta}(\mathbf{z})$  is between  $\hat{\theta}(\mathbf{z})$  and  $\theta_h^*(\mathbf{z})$ . Note that  $\tilde{\theta}(\mathbf{z}) - \theta_h^*(\mathbf{z}) = o_P(1)$  from the consistency of  $\hat{\theta}(\mathbf{z})$  obtained in Proposition 6.5.

We can decompose

$$\begin{aligned} \nabla_{\theta} \hat{M}_n(\theta_h^*(\mathbf{z}), \mathbf{z}) &= \nabla_{\theta} \{ \hat{M}_n(\theta_h^*(\mathbf{z}), \mathbf{z}) - \tilde{M}_n(\theta_h^*(\mathbf{z}), \mathbf{z}) \} + \nabla_{\theta} \{ \tilde{M}_n(\theta_h^*(\mathbf{z}), \mathbf{z}) - M_n^*(\theta_h^*(\mathbf{z}), \mathbf{z}) \} \\ &\quad + \nabla_{\theta} \{ M_n^*(\theta_h^*(\mathbf{z}), \mathbf{z}) - M(\theta_h^*(\mathbf{z}), \mathbf{z}) \}. \end{aligned}$$

The first term is  $O_P(\varepsilon_n)$  from Proposition 6.3. Proposition 6.4 shows that the second term is  $O_P(n^{-1/2}) = o_P(n^{-1/2} h^{-(p+1)/2})$ . The last term is

$$\frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n}^* K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \varphi_{\theta(\mathbf{z})}(\mathbf{U}_i),$$

since  $\nabla M(\theta_h^*(\mathbf{z}), \mathbf{z}) = 0$ .

On the other hand,  $\nabla_{\theta}^2 \hat{M}_n(\tilde{\theta}(\mathbf{z}))$  is a matrix whose coefficient  $(j, k)$  is

$$\sigma_n^{(j,k)} = \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K \left( \frac{\mathbf{Z}_i - \mathbf{z}}{h} \right) \varphi_{\tilde{\theta}(\mathbf{z})}^{(j,k)}(\hat{\mathbf{U}}_i).$$

The consistency of  $\tilde{\theta}(\mathbf{z})$  and the continuity of  $\varphi_{\theta}^{(j,k)}$  lead to

$$\sigma_n^{(j,k)} = \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n} K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \varphi_{\tilde{\theta}(\mathbf{z})}^{(j,k)}(\mathbf{U}_i) + o_P(1).$$

Then we apply Proposition 6.2 to show that

$$\sigma_n^{(j,k)} = \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n}^* K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \varphi_{\theta(\mathbf{z})}^{(j,k)}(\mathbf{U}_i) + o_P(1).$$

Then, using the continuity of the map  $\theta \rightarrow \varphi_{\theta}$  at point  $\theta(\mathbf{z})$ , we get

$$\sigma_n^{(j,k)} = \frac{1}{h^{p+1}} \sum_{i=1}^n W_{i,n}^* K\left(\frac{\mathbf{Z}_i - \mathbf{z}}{h}\right) \varphi_{\theta(\mathbf{z})}^{(j,k)}(\mathbf{U}_i) + o_P(1).$$

We then can apply Lemma 6.6 to show that

$$\sigma_n^{(j,k)} = \sigma^{(j,k)} + o_P(1),$$

where  $\sigma^{(j,k)}$  is the  $(j, k)$ -coefficient of the matrix  $\Sigma(\mathbf{z})$ . This shows that

$$\nabla_{\theta}^2 \hat{M}_n(\tilde{\theta}(\mathbf{z})) = \Sigma(\mathbf{z}),$$

and shows (3.7), from which the asymptotic normality follows.

**Lemma 6.6** *Let  $\phi$  be such that  $E[|\phi(\mathbf{U})|] < \infty$ , and assume that the density of  $(\mathbf{U}, \mathbf{Z})$ ,  $f_{\mathbf{U}, \mathbf{Z}}$  is continuous. Then,*

$$\frac{1}{h^{p+1}} E \left[ \frac{\delta\phi(\mathbf{U})}{S^C(Y)} K\left(\frac{\mathbf{Z} - \mathbf{z}}{h}\right) \right] = E[\phi(\mathbf{U}) | \mathbf{Z} = \mathbf{z}] f_{\mathbf{Z}}(\mathbf{z}) + o(1). \quad (6.5)$$

Moreover, under (3.4) and assuming that

$$\int \phi(\mathbf{u}) \mathfrak{s}(\mathbf{u}) d\mathbf{u} < \infty, \quad (6.6)$$

we have

$$\frac{1}{h^{p+1}} E \left[ \frac{\delta\phi(\mathbf{U})}{S^C(Y)} K\left(\frac{\mathbf{Z} - \mathbf{z}}{h}\right) \right] - E[\phi(\mathbf{U}) | \mathbf{Z} = \mathbf{z}] f_{\mathbf{Z}}(\mathbf{z}) = O(h^2). \quad (6.7)$$

**Proof.** We have

$$\frac{1}{h^{p+1}} E \left[ \frac{\delta\phi(\mathbf{U})}{S^C(Y)} K\left(\frac{\mathbf{Z} - \mathbf{z}}{h}\right) \right] = \frac{1}{h^{p+1}} E \left[ \phi(\mathbf{U}) K\left(\frac{\mathbf{Z} - \mathbf{z}}{h}\right) \right].$$

Through a change of variables,

$$\frac{1}{h^{p+1}} E \left[ \phi(\mathbf{U}) K\left(\frac{\mathbf{Z} - \mathbf{z}}{h}\right) \right] = \int K(\mathbf{v}) \phi(t, \mathbf{u}) f_{\mathbf{U}, \mathbf{Z}}(\mathbf{u}, \mathbf{z} + h\mathbf{v}) d\mathbf{u} d\mathbf{v}. \quad (6.8)$$

Since  $h$  tends to 0 and  $f_{\mathbf{U}, \mathbf{Z}}(t, \mathbf{u}, \mathbf{z} + h\mathbf{v})$  is continuous, then (6.5) follows from dominated convergence and the fact that  $\int K(\mathbf{v})d\mathbf{v} = 1$ .

To prove (6.7), apply a second order Taylor expansion to (6.8). We get, from (3.4) and the fact that  $\int \mathbf{v}K(\mathbf{v})d\mathbf{v} = 0$ ,

$$\left| \frac{1}{h^{p+1}} E \left[ \phi(T, \mathbf{U}) |K| \left( \frac{\mathbf{Z} - \mathbf{z}}{h} \right) \right] - \right| \leq h^2 \int \|\mathbf{v}\|_2^2 K(\mathbf{v}) \phi(\mathbf{u}) \mathfrak{s}(\mathbf{u}) d\mathbf{u} d\mathbf{v}.$$

The result then follows from (6.6). ■

## 6.5 Rate of convergence of the margins for estimator (2.5)

In this section, we show that Assumption 3 holds for the kernel estimator (2.5). This is in fact a consequence of Theorem 4 in Einmahl et al. (2005). We show the result under three additional assumptions on the model:

1. we have

$$\sup_{\mathbf{z}, y} \left| \frac{F^{(j)}(l)}{F^{(j)}(l|\mathbf{z})} + \frac{1 - F^{(j)}(l)}{1 - F^{(j)}(l|\mathbf{z})} \right| \leq \mathbf{a},$$

for some finite constant  $\mathbf{a}$ , where  $F^{(j)}(l) = \mathbb{P}(L^{(j)} \geq l)$ ;

2. the kernel function  $\tilde{K}$  is a continuous and bounded function, symmetric around 0, such that  $\int u^2 \tilde{K}(u) du < \infty$ , the density  $\mathbf{z} \mapsto f_{\mathbf{Z}}(\mathbf{z})$  and  $\mathbf{z} \mapsto F^{(j)}(l|\mathbf{z})$  are twice continuously differentiable with respect to  $\mathbf{z}$ , with uniformly bounded derivatives up to order 2;

3. we assume that, for some  $\iota > 0$  and  $0 < a < 1/2$ ,

$$E \left[ \frac{1}{S^C(Y_i)^{1+\iota}} \left( \frac{1}{[F^{(j)}(L_i^{(j)})]^a} \right)^{2+\iota} \right] < \infty,$$

and

$$E [\mathfrak{K}(T_i)^{1/2+\iota}] < \infty.$$

The first assumption is a way to consider that there is some kind of uniform domination of the behavior of the conditional distributions when  $\mathbf{x}$  changes. The second assumption is classical in kernel regression, and will help to control the bias term involved in smoothing techniques. Finally, the third assumption is required to deal with the censoring.

Introducing the kernel estimator of the density of  $\mathbf{Z}$ ,

$$\hat{f}_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{\tilde{h}^d} \sum_{i=1}^n W_{i,n} \tilde{K} \left( \frac{\mathbf{Z}_i - \mathbf{z}}{\tilde{h}} \right),$$

we can write, for  $t \leq 1/2$ ,

$$\hat{f}_{\mathbf{Z}}(\mathbf{z}) \frac{\hat{F}^{(j)}(l|\mathbf{z})}{[F^{(j)}(l|\mathbf{z})(1 - F^{(j)}(l|\mathbf{z}))]^a} = \frac{1}{n\tilde{h}^p} \sum_{i=1}^n \frac{W_{i,n}}{W_{i,n}^*} \tilde{K} \left( \frac{\mathbf{Z}_i - \mathbf{Z}}{\tilde{h}} \right) f_t(M_i^{(j)}),$$

where

$$\begin{aligned} f_t(l) &= \frac{\delta_i \mathbf{1}_{l \leq t}}{S^C(Y_i) [F^{(j)}(t|\mathbf{x})(1 - F^{(j)}(t|\mathbf{x}))]^a} \\ &\leq \frac{1}{[F^{(j)}(y|\mathbf{x})]^a [1 - F^{(j)}(1/2|\mathbf{x})]^a} \\ &\leq \frac{\alpha^a}{[F^{(j)}(y)]^a [1 - F^{(j)}(1/2|\mathbf{x})]^a}. \end{aligned}$$

Since

$$E \left[ \left( \frac{\delta_i}{S^C(Y_i) [F^{(j)}(M_i^{(j)})]^a} \right)^p \right] = E \left[ \frac{1}{S^C(Y_i)^{1+\iota}} \left( \frac{1}{[F^{(j)}(L_i^{(j)})]^a} \right)^{2+\iota} \right] < \infty,$$

for some  $\iota > 0$  and  $a < 1/2$ , and since the covering number of the class of functions  $f_t$  is controlled (see Example 19.12 in Van der Vaart (2000)), Theorem 4 in Einmahl et al. (2005) applies, showing that

$$\sup_{t \leq 1/2, x} \left| \frac{1}{n\tilde{h}^d} \sum_{i=1}^n \tilde{K} \left( \frac{\mathbf{Z}_i - \mathbf{z}}{\tilde{h}} \right) f_t(M_i^{(j)}) - E \left[ \tilde{K} \left( \frac{\mathbf{Z}_i - \mathbf{z}}{\tilde{h}} \right) f_t(M_i^{(j)}) \right] \right| = O_P \left( [\log n]^{1/2} n^{-1/2} \tilde{h}^{-d/2} \right).$$

Then, from a Taylor expansion and the second assumption of this section, we get

$$\begin{aligned} E \left[ f_t(M_i^{(j)}) \tilde{K} \left( \frac{\mathbf{Z}_i - \mathbf{z}}{\tilde{h}} \right) \right] &= E \left[ f_t(M_i^{(j)}) | \mathbf{Z}_i = \mathbf{z} \right] f_{\mathbf{Z}}(\mathbf{z}) + O(\tilde{h}^2) \\ &= \frac{F^{(j)}(t|\mathbf{x})}{[F^{(j)}(t|\mathbf{x})(1 - F^{(j)}(t|\mathbf{x}))]^a} f_{\mathbf{Z}}(\mathbf{z}) + O(\tilde{h}^2). \end{aligned}$$

Let us note that the  $\tilde{h}^2$  rate for the bias term can be improved if one uses a degenerate kernel with a sufficiently high number of moments equal to zero. Moreover, it can easily be shown that

$$\begin{aligned} \frac{1}{n\tilde{h}^p} \sum_{i=1}^n \frac{W_{i,n}}{W_{i,n}^*} \tilde{K} \left( \frac{\mathbf{Z}_i - \mathbf{Z}}{\tilde{h}} \right) f_t(M_i^{(j)}) &= \frac{1}{n\tilde{h}^{p+1}} \sum_{i=1}^n \tilde{K} \left( \frac{\mathbf{Z}_i - \mathbf{Z}}{\tilde{h}} \right) f_t(M_i^{(j)}) \\ &\quad + O_P(\tilde{h}^2 + [\log n]^{1/2} n^{-1/2} \tilde{h}^{-p+1/2}), \end{aligned}$$

provided that  $n\tilde{h}^4 = o(1)$  and  $n\tilde{h}^{p+1} \rightarrow \infty$ , using the same way to bound  $W_{i,n}/W_{i,n}^*$  as we did in the proof of Proposition 6.4 (using the third assumption of this section).

We can study  $\hat{f}_Z$  in the same way, to obtain the same convergence rate. This leads to

$$\sup_{t \leq 1/2, \mathbf{x}} \left| \frac{\hat{F}^{(j)}(t|\mathbf{x}) - F^{(j)}(t|\mathbf{x})}{[F^{(j)}(t|\mathbf{x})(1 - F^{(j)}(t|\mathbf{x}))]^\alpha} \right| = O_P(\tilde{h}^2 + [\log n]^{1/2} n^{-1/2} \tilde{h}^{-d/2}).$$

Studying the supremum for  $t > 1/2$  can be done in the same way, by studying  $1 - F^{(j)}$  instead of  $F^{(j)}$ .

## References

- Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *Journal of Multivariate Analysis*, 110:43 – 73. Special Issue on Copula Modeling and Dependence.
- Antonio, K., Godecharle, E., and Van Oirbeek, R. (2016). A multi-state approach and flexible payment distributions for micro-level reserving in general insurance. <http://dx.doi.org/10.2139/ssrn.2777467>.
- Antonio, K. and Plat, R. (2010). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014.
- Bou-Hamad, I., Larocque, D., Ben-Ameur, H., et al. (2011). A review of survival trees. *Statistics Surveys*, 5:44–71.
- Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2000). Copulas for finance—a reading guide and some applications. *Available at SSRN 1032533*.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.
- Einmahl, U., Mason, D. M., et al. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, 33(3):1380–1403.
- Fermanian, J.-D. and Wegkamp, M. (2004). Time dependent copulas. *Preprint INSEE, Paris, France*.
- Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*, volume 169. John Wiley & Sons.
- Gerber, G., Faou, Y. L., Lopez, O., and Trupin, M. (2021). The impact of churn on client value in health insurance, evaluation using a random forest under various censoring mechanisms. *Journal of the American Statistical Association*, 116(536):2053–2064.



- Gill, R. (1983). Large sample behaviour of the product-limit estimator on the whole line. *The annals of statistics*, 11(1):49–58.
- Ishwaran, H., Kogalur, U. B., Blackstone, E. H., Lauer, M. S., et al. (2008). Random survival forests. *The annals of applied statistics*, 2(3):841–860.
- Jaworski, P., Durante, F., Hardle, W. K., and Rychlik, T. (2010). *Copula theory and its applications*, volume 198. Springer.
- Jin, X. and Frees, E. W. J. (2013). Comparing micro and macro level loss reserving models. *Preprint*.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481.
- Lopez, O. (2009). Single-index regression models with right-censored responses. *Journal of Statistical Planning and Inference*, 139(3):1082–1097.
- Lopez, O. (2019). A censored copula model for micro-level claim reserving. *Insurance: Mathematics and Economics*, 87:1 – 14.
- Lopez, O., Milhaud, X., and Thérond, P.-E. (2016). Tree-based censored regression with applications in insurance. *Electron. J. Statist.*, 10(2):2685–2716.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *Astin bulletin*, 23(2):213–225.
- Merz, M., Wüthrich, M. V., and Hashorva, E. (2013). Dependence modelling in multivariate claims run-off triangles. *Annals of Actuarial Science*, 7(1):3–25.
- Nelder, J. A. and Baker, R. J. (1972). *Generalized linear models*. Wiley Online Library.
- Nelsen, R. B. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition.
- Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance 1. *ASTIN Bulletin: The Journal of the IAA*, 23(1):95–115.
- Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin*, 29(1):5–25.

- Omelka, M., Hudecová, Š., and Neumeyer, N. (2021). Maximum pseudo-likelihood estimation based on estimated residuals in copula semiparametric models. *Scandinavian Journal of Statistics*, 48(4):1433–1473.
- Pigeon, M., Antonio, K., and Denuit, M. (2014). Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics*, 58:121 – 131.
- Romanosky, S., Ablon, L., Kuehn, A., and Jones, T. (2019). Content analysis of cyber insurance policies: How do carriers price cyber risk? *Journal of Cybersecurity*, 5(1):tyz002.
- Sabban, I. C., Lopez, O., and Mercuzot, Y. (2022). Automatic analysis of insurance reports through deep neural networks to identify severe claims. *Annals of Actuarial Science*, 16(1):42–67.
- Saluz, A., Bühlmann, H., Gisler, A., and Moriconi, F. (2014). Bornhuetter-Ferguson reserving method with repricing. <https://ssrn.com/abstract=2697167>.
- Sklar, M. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.
- Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica*, pages 1089–1102.
- Stute, W. and Wang, J.-L. (1993). The strong law under random censorship. *The Annals of statistics*, pages 1591–1607.
- Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canadian Journal of Statistics*, 33(3):357–375.
- Van der Laan, M. J. and Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Springer Science & Business Media.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.
- Veraverbeke, N., Omelka, M., and Gijbels, I. (2011). Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, 38(4):766–780.
- Wei, L.-J. (1992). The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in medicine*, 11(14-15):1871–1879.

Wüthrich, M. V. (2016). Machine learning in individual claims reserving.  
<http://dx.doi.org/10.2139/ssrn.2867897>.

Wüthrich, M. V. (2017). Neural networks applied to chain-ladder reserving.  
<http://dx.doi.org/10.2139/ssrn.2966126>.

Zhao, X. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2):290 – 299.