



HAL
open science

A censored copula model for micro-level claim reserving

Olivier Lopez

► **To cite this version:**

Olivier Lopez. A censored copula model for micro-level claim reserving. Insurance: Mathematics and Economics, 2019, pp.1-14. 10.1016/j.insmatheco.2019.04.001 . hal-01706935

HAL Id: hal-01706935

<https://hal.science/hal-01706935>

Submitted on 12 Feb 2018

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.

A censored copula model for micro-level claim reserving

Olivier LOPEZ¹

12 février 2018

Résumé

In this paper, we consider the question of predicting the final amount of a claim and its distribution from micro-level data. A copula model is used to describe the dependence between the amount of a claim and its duration (that is the time between its occurrence and its closure). Due to the presence of censoring, we adapt classical methodologies using a weighting scheme that corrects the bias caused by this incompleteness in the data. Theoretical results and simulation support the validity of the procedure. A real case coming from medical malpractice claims is presented.

Key words : Claim reserving, copula models, Kaplan-Meier estimator, censoring.

Short title : Censored copula for claim reserving

¹ Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, LPSM,
4 place Jussieu, F-75005 Paris, France, E-mail : olivier.lopez@sorbonne-universite.fr

1 Introduction

Evaluating loss reserves relies on the prediction of the final amount of claims that already happened and for which the insurer will be committed to pay. In the present paper, we propose a new methodology to perform this analysis at a micro-level. Our aim is to use known characteristics of a claim in order to predict its amount and the time before its settlement. A key feature of our method is to model the dependence structure between the amount and the duration using a copula model. Indeed, claims that take a long time before being closed are more likely to be expensive.

The task is made more difficult by the presence of censoring. This classical problem in survival analysis is unavoidable when dealing with duration variables. Indeed, at a given date, two types of situations may happen for the current status of a claim : either it is closed, in which case its final amount is known ; or it is still open, in which case the claim is said to be censored. If one does not take censored claims into account, the danger is to build a model based on claims that are, in average, smaller than a typical one. This is due to the fact that, among the uncensored claims, there is an overrepresentation of small claims that take a short time to be closed. In this paper, we propose a way to correct this bias caused by the censoring by introducing an appropriate weighting scheme which compensates this deficit in large claims.

In the literature of claim reserving, aggregated methods are often proposed. Chain-Ladder type methods (see e.g. Mack (1993), Merz et al. (2013), Pigeon et al. (2014), Saluz et al. (2014)) are a way to respond to the fact that amounts of claims are not paid right at the time at occurrence. All of these methods are all based on a stability assumption that may be difficult to validate. They also tend to loose the information one may have on each individual claim.

Various techniques have been proposed in the literature to introduce information one may have on the specificity of each claim. A first way to proceed is to make development factors, in the Chain-Ladder approach, depend on covariates, as in Wuthrich (2017) or Wuthrich (2016). Micro-level approaches consider each claim separately and use its characteristic to predict its amount, see Norberg (1993) and Norberg (1999), and see Jin and Frees (2013) for a comparison between the macro and the micro approach. When looking at IBNR (Incurred But Not Reported) claims, time is an important factor, since the question is to estimate the delay before reporting the claim. Antonio and Plat (2010) propose to improve the evaluation of the IBNR reserves by using Poisson processes. Zhao and Zhou (2010) propose a copula approach to study the dependence between the delay

of reporting and the time of occurrence of the claim. In their work, they deal with a left-truncation phenomenon due to the absence in the datasets of some claims that are still not reported.

In the present paper, we study the importance of the development of claims through time, but focusing mostly on RBNS claims (Reported But Not Settled). Our main concern is to provide a model that allows us to understand the dependence between the time before settlement (what we will call the "lifetime" of the claim) and its amount. This is of particular importance for some guarantees for which the settlement of a claim can take a lot of time. For example, the dataset we consider in the real application below comes from medical malpractice claims. When evaluating RBNS claims, most existing approaches (see e.g. Jin (2013) or Ayuso and Santolino (2008)) study the lifetime and the amount of a claim separately, or use multi-state modeling as in Antonio et al. (2016) to model the development of the claims. Nevertheless some estimation bias may occur when dealing with guarantees with large time before settlement. This is due to the censoring phenomenon that we already mentioned : databases that are used to calibrate the distribution of the amount of the claims lack of claims with important lifetime (since only closed claims are used to evaluate the distribution).

Our approach aims to fill this gap, by showing how survival analysis and copula techniques may be used in this context. Our methodology may contribute to complement a chain-ladder-type approach, by improving the evaluation of RBNS claims, taking advantage of all the information available on them. Apart from improving the way the reserves are evaluated, the aim is also to furnish a data-driven approach that does not necessarily require the intervention of an expert, sent physically to evaluate the amount. Based of a sufficient amount of gathered information on the circumstances of the claims, the idea is to let the model do the prediction, and send an expert only when the prediction is too blurry (based on the evaluation of the prediction uncertainty which can also be obtained using our method).

The paper is organized as follows. In section 2, we present the censoring model that we consider to model the evolution of the claims, and explain our estimation and prediction procedure. Section 3 is devoted to the theoretical behavior of the methodology, completed by some simulation study. A real data case, corresponding to medical malpractice claims, is shown in details in section 4. Technical results are gathered in section 6.

2 Model and assumptions

We first present the model in section 2.1, and then focus on the estimation procedure, decomposed into the estimation of the margins in section 2.2, and then on the dependence structure in section 2.3. The methodology to forecast the open claims is described in section 2.4.

2.1 Observations and setting

We consider a set of i.i.d. observations $(Y_i, \delta_i, C_i, X_i, N_i)_{1 \leq i \leq m}$, where

$$\begin{cases} Y_i &= \inf(T_i, C_i), \\ \delta_i &= \mathbf{1}_{T_i \leq C_i}, \\ N_i &= \delta_i M_i. \end{cases}$$

In the micro-level reserving situation, M_i is the amount of claim number i , $X_i \in \mathbb{R}^d$ are its characteristics. T_i is the time before the claim is stabilized, that is when one finally knows its total amount. This variable is censored by C_i , the time between the opening of the claim and today. This means that, if $T_i > C_i$ (which corresponds to the case $\delta_i = 0$), neither T_i or M_i is observed. This model is similar to the one studied in Lopez et al. (2016), except that here, we assume that C_i is observed.

The reason for supposing that C_i is observed, compared to a more general censoring model, is that we consider that the only cause for censoring is the end of the observation period. As mentioned in section 2.5 below, the approach we propose is extended to more general situations in which C_i may not be observed. Nevertheless, we prefer to focus on this more simple case, in order to take advantage of the additional information we have on the censoring. This will also allow us to obtain better asymptotic results for the estimators we propose.

Let us also mention that, in the above setting, no partial information on the amount of the claim is required, although it can be used if present. Indeed, according to this model, $N_i = 0$ if $\delta_i = 0$. If some elements are given (expertise, partial payments...), this information can be contained in the covariates $X_i \in \mathbb{R}^d$.

In such censoring models, an identifiability assumption is required to retrieve the distribution of (T, M, X) from the data. Throughout this paper, we will use the following assumption.

Assumption 1 *Assume that (T, X, M) is independent of C .*

As a consequence of this assumptions, the covariates X should not be dependent on the time at which the claim opened. Otherwise, X would be dependent of C , since C is the difference between the date of today and the date at which the claim occurred. This means that X should not include a time factor that could have impact on M (like some time-dependent indicators, or the year of occurrence). Therefore, the values of M have to be considered as corrected from an inflation factor. If no indication on the inflation is present, we discuss in section 4.2 how to remove such an inflation effect.

2.2 Estimation of the margins

In this section, we discuss how to model the conditional distribution of $T|X = x$, and the distribution of $M|X = x$. These variables are expected to have very different types of distributions. T is a duration variable, while M is an amount that may be very volatile.

2.2.1 Estimation of the conditional distribution of T

Various ways to estimate a regression model involving a censored variable T have been discussed in the literature. Defining $\mu(t|x) = \lim_{dt \rightarrow 0^+} dt^{-1} \mathbb{P}(T \in [t, t + dt] | X = x)$ the conditional hazard rate function, Cox regression model assumes that $\mu(t|x) = \mu_0(t) \exp(\alpha_0^T x)$, where μ_0 is an unknown baseline function and α_0 a finite dimensional unknown parameter. This model has the advantage to be semiparametric, since it allows us not to specify the baseline μ_0 . Estimation of α_0 can be performed using pseudo-likelihood maximization (see Cox (1975)), while the cumulative hazard function $M_0(t) = \int_0^t \mu_0(s) ds$ can be estimated nonparametrically using the Breslow estimator (see Burr (1994)).

An alternative to Cox regression is the Accelerated Failure Time model (AFT), where one assumes that

$$\mu(t|x) = \alpha_0 \mu_0(\alpha_0 t), \quad (2.1)$$

where μ_0 is a known baseline function, and α_0 an unknown parameter. Estimation of α_0 can be performed using maximum likelihood estimation (see Wei (1992)). This last type of models will be used in section 4.

2.2.2 Estimation of the conditional distribution of M

A classical way to model the amount of a claim is to consider a Generalized Linear Model. We here explain how to extend the estimation of a GLM model estimation to the context of censoring. We consider a known transformation $\psi(M)$ of M and suppose

that it fits a GLM. This model, see Nelder and Baker (1972), consists of assuming that $\Psi(M)|X$ has a distribution in a fixed exponential family, and that $g(E[\Psi(M)|X]) = \beta_0^T X$, where g is a known link function. The reason for considering $\Psi(M)$ and not merely M is that M may be heavy-tailed in some cases. In this case, using an exponential family of distribution is unadapted. On the other hand, a GLM assumption may be reasonable for $\log M$, for instance.

Estimation of the parameter β_0 can be done by

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n W_{i,n} \log f_{\beta}(\psi(N_i)|\beta^T X_i), \quad (2.2)$$

where $f_{\beta}(\cdot|\beta^T x)$ denotes the conditional density of $\psi(M)|X$ according the model with parameter β , and where

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

and where $\hat{S}_C(t) = n^{-1} \sum_{i=1}^n \mathbf{1}_{C_i \geq t}$, the empirical survival function of C , which is a consistent estimator of $S_C(t) = \mathbb{P}(C \geq t)$. The reason for using (2.3) is the fact that, under Assumption 1,

$$E \left[\frac{\delta \phi(Y, N, X)}{S_C(Y)} \right] = E [\phi(T, M, X)], \quad (2.4)$$

for all functions ϕ such that $E[|\phi(T, M, X)|] < \infty$, and such that $\phi(t, m, x) = 0$ if $S_C(t) = 0$. Such type of weighting are of common use to correct the bias caused by the censoring (see Van der Laan and Robins (2003), or Lopez et al. (2016) for some applications of these Inverse Probability of Censoring Weights (IPCW) techniques). If the variables $(C_i)_{1 \leq i \leq n}$ are not observed, the empirical distribution function \hat{S}_C can be replaced by the Kaplan-Meier estimator of Kaplan and Meier (1958).

Note that (2.2) is not a true maximum likelihood estimator, but it has the advantage of relying on the maximization of a criterion which is numerically of the same complexity as in the uncensored case. An alternative way would be to use the expression of the maximum likelihood estimator under censoring, see Fleming and Harrington (2011), which, is more delicate to handle numerically. Hence we prefer to consider this easier optimization problem.

This estimation procedure is $n^{1/2}$ -consistent under mild assumptions by a direct adaptation of Stute (1999), who considers the special case of a nonlinear regression model with a quadratic loss function. We give elements, in the Appendix, to justify this extension, see section 6.5.

2.3 Copula modeling

Clearly, in our micro-level reserving problem, the variable T and the variable M are dependent. Indeed, a claim with a long time T before being settled, is expected to be associated with a large value of M , since small claims are more likely to be rapidly settled. The advantage of considering the dependence between M and T from a copula prospective stands in the fact that these two variables are of different nature (an amount and a time). Hence, it would be difficult to model them jointly. Besides, the copula approach allows us to separate the univariate modeling as in section 2.2, and then study the dependence structure separately.

The basic idea behind copula models arises from Sklar's Theorem (see Sklar (1959)) which states that, for two variables A and B ,

$$F_{A,B}(a, b) = \mathfrak{C}(F_A(a), F_B(b)), \quad (2.5)$$

where $F_{A,B}(a, b) = \mathbb{P}(A \leq a, B \leq b)$, F_A (resp. F_B) denotes the cumulative distribution function of A (resp. B), and \mathfrak{C} is a copula function (that is a distribution function on $[0, 1]^2$ with uniform margins). Moreover this decomposition is unique if the variables are continuous, which will be the case in our application. Hence, (2.5) introduces a separation between the marginal behavior of A and B (entirely described by F_A and F_B) and their dependence structure, solely contained in \mathfrak{C} .

In view of (2.5), let \mathfrak{C}_x denote the conditional copula of (T, M) given $X = x$. This means that, if we consider the joint conditional function $F_{T,M|x}(t, m|x) = \mathbb{P}(T \leq t, M \leq m|X = x)$, it decomposes into

$$F_{T,M|x}(t, m|x) = \mathfrak{C}_x(F_{T|x}(t|x), F_{M|x}(m|x)),$$

where $F_{T|x}(t|x) = \mathbb{P}(T \leq t|X = x)$ and $F_{M|x}(t|x) = \mathbb{P}(M \leq t|X = x)$. In the following, we assume that the dependence structure of (T, M) does not depend on the value of X , that is there exists a copula function \mathfrak{C} such that, for all x , $\mathfrak{C}_x = \mathfrak{C}$. Moreover, we assume that \mathfrak{C} belongs to a parametric family of copulas $\mathcal{C} = \{\mathfrak{C}_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$.

We denote by $\mathfrak{c}_\theta(u, v)$ the copula density associated to \mathfrak{C}_θ , that is $\mathfrak{c}_\theta(u, v) = \partial_{u,v}^2 \mathfrak{C}_\theta(u, v)$. Let $U_i = F_{T|x}(T_i|X_i)$ and $V_i = F_{M|x}(M_i|X_i)$. If $(U_i, V_i)_{1 \leq i \leq n}$ were observed, the maximum likelihood estimator of θ_0 would be $\hat{\theta}^{ML} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log \mathfrak{c}_\theta(U_i, V_i)$. In our setting, U_i and V_i can not be computed for two reasons. The first reason, common in copula modeling, is that the conditional marginal distributions $F_{T|x}$ and $F_{M|x}$ are not exactly known. However, estimators for the margins can be deduced from section 2.2, and therefore

we have at our disposal two estimated conditional distribution functions $\hat{F}_{T|X}$ and $\hat{F}_{M|X}$ that could be plugged into the maximum likelihood procedure. Nevertheless, in our case, censoring introduces an additional complexity, since the quantities $U_i^* = \hat{F}_{T|X}(T_i|X_i)$ and $V_i^* = \hat{F}_{M|X}(M_i|X_i)$ can not be computed if $\delta_i = 0$ (in this case, T_i and M_i are unknown). Therefore, we define $\hat{U}_i = \hat{F}_{T|X}(Y_i|X_i)$ and $\hat{V}_i = \hat{F}_{M|X}(N_i|X_i)$, and our copula parameter estimator

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n W_{i,n} \log \mathbf{c}_{\theta}(\hat{U}_i, \hat{V}_i), \quad (2.6)$$

where $W_{i,n}$ is defined in (2.3). From (2.3), we see that $\hat{U}_i = U_i^*$ and $\hat{V}_i = V_i^*$ when $W_{i,n} \neq 0$. Moreover, introducing $W_{i,n}$ is a natural way to correct the presence of the censoring from (2.4).

Our estimator (2.6) is motivated by the same arguments as in (2.2). An alternative procedure would be to use the estimator of Shih and Louis (1995), which is based on the extension of maximum likelihood estimation to the censoring framework. However, the simplicity of (2.6) (in terms of definitions but also in terms of numerical optimization) explains why we turn to $\hat{\theta}$.

2.4 Prediction of the final amount of the claims

The previous approach allows us to model the distribution of $(T, M)|X$. The idea is then to use this model in order to predict the final amount of each claim. More precisely, consider an open claim (that is with $\delta_i = 0$). For this claim, we already know that the unobserved variable T_i is greater than Y_i . Hence, the distribution of (T, M) given the information we have (that is given $X = X_i$ and $T \geq Y_i$) has a density that can be estimated by

$$\hat{f}_{Y_i, X_i}(t, m) = \frac{\mathbf{c}_{\hat{\theta}}(\hat{F}_{T|X}(t|x), \hat{F}_{M|X}(m|x)) \hat{f}_{T|X}(t|x) \hat{f}_{M|X}(m|x) \mathbf{1}_{t \geq Y_i}}{\int_{u=Y_i}^{\infty} \int_{v=-\infty}^{\infty} \mathbf{c}_{\hat{\theta}}(\hat{F}_{T|X}(u|x), \hat{F}_{M|X}(v|x)) \hat{f}_{T|X}(u|x) \hat{f}_{M|X}(v|x) dudv}, \quad (2.7)$$

where $\hat{f}_{T|X}(\cdot|x)$ (resp. $\hat{f}_{M|X}(\cdot|x)$) denotes the estimated conditional density of $T|X = x$ (resp. of $M|X = x$) and $\hat{F}_{T|X}(\cdot|x)$ (resp. $\hat{F}_{M|X}(\cdot|x)$) denotes the estimated conditional distribution function, obtained in section 2.2.

Next, the prediction can either be performed analytically, for example by computing $\int_{m=0}^{+\infty} \int_{t=Y_i}^{\infty} m \hat{f}_{Y_i, X_i}(t, m) dt dm$ (or alternatively by looking at the median of this estimated distribution), or using simulation. This second approach seems more convenient due to the difficulty to find a closed form for the analytic form. Moreover, the simulation approach

permits us to obtain easily an evaluation of the distribution of the amount of the open claim. Indeed, for each claim for which $\delta_i = 0$, the idea is to simulate a large number of variables under the estimated density (2.7). One then can estimate the amount of claim i by \hat{M}_i which is the empirical mean of the simulated sample, but also have insights about the uncertainty of this estimation.

2.5 Discussion and extensions

In this paper, we have chosen to consider the case where the dependence structure of $(T, M)|X = x$ does not depend on x . This is motivated by the search for a relatively simple model. Another argument is the difficulty of estimating a conditional copula function without having at our disposal a sufficiently large amount of data. This issue becomes even worse when the dimension d of X becomes large. However, the estimator (2.6) can be modified in order to consider the more general case where \mathfrak{C}_x depends on x , by, for example, adding kernel weights as in Abegaz et al. (2012) in the uncensored case.

Another limit of this procedure is the identifiability assumption (Assumption 1) we have made on the censoring. A common alternative would be to assume that C is independent of (T, M) given X . This would allow C to depend on the covariates, X . In this case, (2.4) has to be replaced by

$$E \left[\frac{\delta\phi(Y, N, X)}{S_{C|X}(Y|X)} \right] = E [\phi(T, M, X)],$$

where $S_{C|X}(t|x) = \mathbb{P}(C \geq t|X = x)$. The weights $W_{i,n}$ must be adapted, by replacing the estimator of the survival function of C by an estimator of the conditional survival function $S_{C|X}(\cdot|x)$.

We also considered that C was observed. If it is not the case, the procedure stays the same, but with the empirical distribution function \hat{S}_C replaced by the Kaplan-Meier estimator of the distribution of C (see Kaplan and Meier (1958)).

3 Consistency of the copula estimator

This section investigates the theoretical behavior of the copula estimator. Section 3.1 gathers the required assumptions, which are basically the same as in the classical uncensored case, plus some additional ones on the tails of the distributions due to the censoring. The consistency results are obtained in section 3.2. The simulation study in section 3.3 illustrates the behavior of the technique for finite sample size.

3.1 List of assumptions

Two categories of assumptions are required. First, Assumptions 2 and 3 are close to Assumptions A.1 to A.5 present in Tsukahara (2005), who gives conditions for the consistency of the maximum likelihood estimator of a copula (when the margins are estimated separately). These assumptions are satisfied by a large number of copula families.

Before stating these two assumptions, let us introduce some notations.

We recall that a function $r : (0, 1) \rightarrow (0, \infty)$ is called u -shaped if r is symmetric about $1/2$ and increasing on $(0, 1/2]$. For such a u -shaped function r , and for $0 < \beta < 1$, define

$$r_\beta(t) = r(\beta t)\mathbf{1}_{0 < t \leq 1/2} + r(1 - \beta(1 - t))\mathbf{1}_{1/2 < t \leq 1}.$$

A reproducing u -shaped function is a u -shaped function such that $r_\beta \leq M_\beta r$ for all $\beta > 0$ in a neighborhood of 0, with M_β a finite constant.

Let \mathcal{R} denote the set of reproducing u -shaped functions in the following, and \mathcal{Q} denote the set of continuous functions on $[0, 1]$ which are positive on $(0, 1)$, symmetric about $1/2$, increasing on $[0, 1/2]$ and such that $\int_0^1 q(t)^{-2} dt < \infty$.

Assumption 2 gives conditions on the first order derivatives (with respect to θ) of the log-likelihood.

Assumption 2 Let $\phi_\theta(u, v) = \nabla_\theta \log c_\theta(u, v)$, $\phi_{\theta_0}^{(1)} = \partial_u \phi_{\theta_0}(u, v)$, and $\phi_{\theta_0}^{(2)}(u, v) = \partial_v \phi_{\theta_0}(u, v)$. Assume that there exist functions $r_j \in \mathcal{R}$, $\tilde{r}_j \in \mathcal{R}$, and $q_j \in \mathcal{Q}$, for $j = 1, 2$ such that, for all (u, v, θ) ,

$$\begin{aligned} \|\phi_\theta(u, v)\| &\leq r_1(u)r_2(v), \\ \|\phi_{\theta_0}^{(1)}(u, v)\| &\leq \tilde{r}_1(u)r_2(v), \\ \|\phi_{\theta_0}^{(2)}(u, v)\| &\leq r_1(u)\tilde{r}_2(v), \end{aligned}$$

and such that

$$\int \frac{\left\{ \prod_{i=1,2} r_i(F_T(t|x), F_M(m|x)) \right\}^2}{S_C(t)} \times f_T(t|x) f_M(m|x) d\mathfrak{C}(F_T(t|x), F_M(m|x)) < \infty, \quad (3.1)$$

$$\int \frac{\tilde{r}_1(F_T(t|x), F_M(m|x)) q_1(F_T(t|x)) r_2(F_T(t|x), F_M(m|x))}{S_C(t)} \times f_T(t|x) f_M(m|x) d\mathfrak{C}(F_T(t|x), F_M(m|x)) < \infty, \quad (3.2)$$

$$\int \frac{r_1(F_T(t|x), F_M(m|x)) \tilde{r}_2(F_T(t|x), F_M(m|x)) q_2(F_M(m|x))}{S_C(t)} \times f_T(t|x) f_M(m|x) d\mathfrak{C}(F_T(t|x), F_M(m|x)) < \infty. \quad (3.3)$$

The conditions on the second order derivatives with respect to θ are presented in Assumption 3.

Assumption 3 *There exists a neighborhood of θ_0 such that $\phi_\theta(u, v)$ is continuously differentiable, and such that the Hessian matrix*

$$\Sigma_\theta(u, v) = \nabla_\theta^2 \log \mathfrak{c}_\theta(u, v),$$

satisfies

$$\|\Sigma_\theta(u, v)\| \leq h_1(u) h_2(v),$$

with $h_i \in \mathcal{R}$ for $i = 1, 2$ and

$$\int \frac{f_T(t|x) f_M(m|x)}{S_C(t)} \left\{ \prod_{i=1,2} h_i(F_T(t|x), F_M(m|x)) \right\} d\mathfrak{C}(F_T(t|x), F_M(m|x)) < \infty.$$

Moreover, assume that $\Sigma = \int \Sigma_{\theta_0}(u, v) d\mathfrak{C}(u, v)$ is nonsingular.

The main difference with the conditions in Tsukahara (2005) is the presence of the function S_C at the denominator, which is caused by censoring. Of course, compared to an uncensored situation, these conditions may not hold if censoring is too heavy (that is if $S_C(t)$ decreases too fast to zero when t tends to infinity).

The second type of assumptions lists requirements on the estimation of the margins. Indeed, in our setting, we wish to have enough flexibility to model the margins, so that the asymptotic properties of our estimators may be obtained for a large number of situations.

Assumption 4 ensures that the margins are estimated with a sufficiently fast convergence rate.

Assumption 4 Assume that

$$\sup_{i=1,\dots,n} \frac{|\hat{U}_i - U_i|}{q_1(U_i)} + \frac{|\hat{V}_i - V_i|}{q_2(V_i)} = O_P(n^{-1/2}), \quad (3.4)$$

where q_j for $j = 1, 2$ are defined in Assumption 2. Moreover, assume that

$$\sup_{i=1,\dots,n} \frac{U_i}{\hat{U}_i} + \frac{1 - U_i}{1 - \hat{U}_i} + \frac{V_i}{\hat{V}_i} + \frac{1 - V_i}{1 - \hat{V}_i} = O_P(1). \quad (3.5)$$

Additionally, one requires an asymptotic representation for the estimated margins is required.

Assumption 5 The margins are estimated so that an asymptotic representation of the following type holds,

$$\hat{U}_i - U_i = \frac{1}{n} \sum_{j=1}^n \lambda_{Y_i, X_i}^U(Y_j, \delta_j, c, X_j) + R_{1,n}, \quad (3.6)$$

$$\hat{V}_i - V_i = \frac{1}{n} \sum_{j=1}^n \lambda_{N_i, X_i}^V(N_j, Y_j, \delta_j, c, X_j) + R_{2,n}, \quad (3.7)$$

where $\sup_{1 \leq i \leq n} |R_{i,n}| = o_P(n^{-1/2})$.

The next assumption imposes moment conditions on the terms in these asymptotic representations (in order to apply a Central Limit Theorem to the terms that appear in the asymptotic study of the estimator).

Assumption 6 Using the notations of Assumption 5,

$$\Lambda_U(y, d, x) = E \left[\phi_{\theta_0}^{(1)}(U_1, V_1) \lambda_{T_1, X_1}^U(Y_2, \delta_2, X_2) | Y_2 = y, \delta_2 = d, X_2 = x \right],$$

$$\Lambda_V(m, y, d, x) = E \left[\phi_{\theta_0}^{(2)}(U_1, V_1) \lambda_{M_1, X_1}^V(N_2, Y_2, \delta_2, X_2) | N_2 = m, Y_2 = y, \delta_2 = d, X_2 = x \right].$$

Assume that

$$E[\Lambda_U(Y_i, \delta_i, X_i)^2] + E[\Lambda_V(M_i, Y_i, \delta_i, X_i)^2] < \infty.$$

Assumptions 4 to 6 can be easily checked for the estimators we consider in section 2.2, see section 6.5.

3.2 Asymptotic results

We now state our main result, which consists of an asymptotic representation of the estimated association parameter $\hat{\theta}$. As a by-product, the asymptotic normality is obtained.

Theorem 3.1 *Under Assumptions 2 to 6,*

$$\hat{\theta} - \theta_0 = \Sigma^{-1} \left(\frac{1}{n} \sum_{i=1}^n \Psi(Y_i, N_i, \delta_i, C_i, X_i) \right) + o_P(n^{-1/2}),$$

where

$$\Psi(y, m, d, x) = -\Gamma_{\phi_{\theta_0}(F_{T|X}(\cdot|\cdot), F_{M|X}(\cdot|\cdot))}(y, n, d, c, x) - \Lambda_U(y, d, c, x) - \Lambda_V(m, y, d, c, x),$$

where

$$\Gamma_{\phi}(y, m, d, c, x) = \frac{d\phi(y, m, d, c, x)}{S_C(y)} + E \left[\frac{\phi(T, M, X)[S_C(T) - \mathbf{1}_{C \geq T}]}{S_C(T)} \mid C = c \right].$$

Hence,

$$n^{1/2}(\hat{\theta} - \theta_0) \implies \mathcal{N}(0, \Sigma^{-1} E [\Psi(Y, M, \delta, C, X)^2] \Sigma^{-1}).$$

The function Ψ is composed of three terms. The first one ($\Gamma_{\phi_{\theta_0}}$) corresponds to the asymptotic representation we would have if the conditional margins $F_{T|X}$ and $F_{M|X}$ were exactly known. It is itself composed of two terms : the first one, $d\phi_{\theta_0}(y, m, d, c, x)[S_C(y)]^{-1}$, is the term one would have if the distribution of the censoring S_C were exactly one, while the second one is present due to the estimation of S_C by its empirical counterpart. The terms Λ_U and Λ_V come from the estimation of the conditional margins, and are dependent on the way this estimation is performed.

The asymptotic variance can naturally be estimated by replacing each expectation and unknown quantity by its empirical counterpart (and θ_0 by $\hat{\theta}$).

3.3 Simulation study

To assess the finite sample validity of our method, we consider the following simulation settings. We simulate $X_i = (X_{i,1}, X_{i,2}, X_{i,3})^T$ where the $X_{i,j}$ are uniform over $[0, 1]$ and mutually independent. The variable T_i is simulated according to a log-normal distribution conditionally to the variables X_i . More precisely, $\log T_i = \alpha_{0,0} + (\alpha_{0,1}, \dots, \alpha_{0,3})X_i + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. For the amount M_i , we consider a GLM model for $\log M_i$ with exponential distribution and using the identity as a link function, that is $E[\log M_i | X_i] =$

$\beta_{0,0} + (\beta_{0,1}, \dots, \beta_{0,3})X_i$, and $\log M_i|X_i$ is exponentially distributed. This produces an heavy tail conditional distribution for M_i , since the conditional survival function of $M_i|X_i$ has a polynomial rate of decay. The censoring C_i is simulated independently under a Weibull distribution of parameters k and λ , that is of density $k\lambda^{-k}c^{k-1} \exp(-c/\lambda)^k \mathbf{1}_{c \geq 0}$. The values for the respective parameters involved in the simulation of the margins and of the censoring are summarized in Table 1. We consider different values for the parameters of the censoring in order to investigate the evolution of the performance when the strength of the censoring (i.e. the average percentage of censored observations in the sample) increases.

TABLE 1 – Parameters used in the simulation of the margins

Parameter name	Parameter value	Proportion of censored observations
$\alpha_{0,0}$	0	
$\alpha_{0,1}$	0.5	
$\alpha_{0,2}$	0.2	
$\alpha_{0,3}$	-0.1	
$\beta_{0,0}$	0.1	
$\beta_{0,1}$	0.03	
$\beta_{0,2}$	0.05	
$\beta_{0,3}$	-0.02	
(k, λ)	(2.7,3)	30%
(k, λ)	(0.5,3)	50%
(k, λ)	(0.15,3)	60%

We consider three copula families to describe the dependence between T and M , that is Frank's, Gumbel's and Clayton's copula families. The corresponding values of the copula functions are summarized in Table 2. Through the use of these three different families, we expect to observe the behavior of the technique in situations where the dependence structure is quite different. Clayton's family allows us to incorporate a left tail dependence, while Gumbel's family induces a right tail dependence, and Frank's family does not have any dependence in the tails. For each copula family, we consider two values of the association parameter. To facilitate the comparison between the families, these two cases correspond to values of Kendall's τ coefficient equal to 0.5 and 0.1. We recall that Kendall's τ coefficient associated to a copula function \mathfrak{C} is a dependence measure that can be expressed as $\tau = 4 \int \int \mathfrak{C}(a, b) d\mathfrak{C}(a, b) - 1$, see Nelsen (2006).

In each simulation setting, we simulate B samples using the distributions that have

TABLE 2 – Copula families, and corresponding values of Kendall’s τ coefficient

Copula family	Copula function	Condition on θ	Kendall’s τ
Clayton	$\frac{1}{(a^{-\theta} + b^{-\theta} + 1)^{1/\theta}}$	$\theta > 0$	$\frac{\theta}{\theta + 2}$
Frank	$-\frac{1}{\theta} \log \left(1 + \frac{(\exp(-\theta a) - 1)(\exp(-\theta b) - 1)}{\exp(-\theta) - 1} \right)$	$\theta \neq 0$	$1 + 4 \frac{\theta^{-1} \int_0^\theta \frac{t^k dt}{\exp(t) - 1} - 1}{\theta}$
Gumbel	$\exp \left(- \left[-(\log a)^\theta + (-\log b)^\theta \right]^{1/\theta} \right)$	$\theta \geq 1$	$1 - \theta^{-1}$

been previously described. For the b -th simulated sample, we compute the estimators of the parameters corresponding to the margins and to the dependence structure according to the procedures of sections 2.2 and 2.3. Then, we predict each of the censored claims using the technique described in section 2.4. That is, for each simulated sample $(Y_i^b, N_i^b, C_i^b, \delta_i^b, X_i^b)_{1 \leq i \leq n}$, we compute, for each censored claim with $\delta_i^b = 0$, the corresponding estimated density (2.7) and simulate 1000 replications under this distribution. Then, the prediction \hat{M}_i^b of the amount $M_{i,b}$ is taken as the mean-value of this simulated sample.

Let $R^b = \sum_{i=1}^n (1 - \delta_i) M_i^b$, that is the reserve that one should constitute if the amounts of all claims were known, and $\hat{R}^b = \sum_{i=1}^n (1 - \delta_i) \hat{M}_i^b$ denote the predicted one. Tables 3 and 4 below quantify different measurement errors, defining

$$\begin{aligned}
 E_1 &= \frac{1}{B} \sum_{b=1}^B |R^b - \hat{R}^b|, \\
 E_{1,r} &= \frac{1}{B} \sum_{b=1}^B \frac{|R^b - \hat{R}^b|}{R^b}, \\
 E_2 &= \frac{1}{B} \sum_{b=1}^B \frac{1}{\sum_{i=1}^n (1 - \delta_i)} \sum_{i=1}^n (1 - \delta_i) |\hat{M}_i^b - M_i^b|.
 \end{aligned}$$

The column $E[(\hat{\theta} - \theta_0)^2]^{1/2}$ presents the empirical estimation of this quantity, that is $\{B^{-1} \sum_{b=1}^B (\hat{\theta}^b - \theta_0)^2\}^{1/2}$ where $\hat{\theta}^b$ is the estimator of the copula parameter for the simulation b .

Errors E_1 and $E_{1,r}$ focus on the global error, that is the difference (or relative difference for $E_{1,r}$) between R^b and \hat{R}^b . Error E_2 focuses on an individual version of the error (claim by claim).

For each model, the relative error $E_{1,r}$ seems relatively stable in the different configurations, with no major difference between the copula families. E_1 tends to decrease when the proportion of censoring increases. This is due to the fact that the number of claim

TABLE 3 – Error measurement for Kendall's τ coefficient equal to 0.1, $n = 5000$

Proportion of censoring	Copula family	$E[(\hat{\theta} - \theta_0)^2]^{1/2}$	E_1	$E_{1,r}$	E_2
30 %	Clayton	0.04118	10.33	0.00595	0.1205
	Frank	0.1314	15.38	0.01221	0.11980
	Gumbel	0.01847	22.20	0.01258	0.1046
50 %	Clayton	0.04579	10.512	0.00363	0.1209
	Frank	0.1544	11.39	0.01215	0.1223
	Gumbel	0.0238	3.450	0.000846	0.1092
60 %	Clayton	0.1009	12.41	0.00325	0.1267
	Frank	0.3027	10.81	0.01209	0.1328
	Gumbel	0.0575	3.171	0.0009402	0.1202

TABLE 4 – Error measurement for Kendall's τ coefficient equal to 0.5, $n = 5000$

Proportion of censoring	Copula family	$E[(\hat{\theta} - \theta_0)^2]^{1/2}$	E_1	$E_{1,r}$	E_2
30 %	Clayton	0.1091	30.59	0.0166	0.1205
	Frank	0.1869	71.236	0.03102	0.1209
	Gumbel	0.05371	86.05	0.0455	0.1051
50 %	Clayton	0.1233	12.74	0.0043	0.1254
	Frank	0.1787	13.09	0.0003760	0.1276
	Gumbel	0.1040	14.64	0.00023	0.1105
60 %	Clayton	0.3113	12.83	0.0034	0.1415
	Frank	0.2386	11.92	0.003494	0.1474
	Gumbel	0.2207	1.6411	0.00055	0.1249

to predict increases : since our estimators are asymptotically unbiased, overestimation of some claims is compensated by under estimation of others. For the individual prediction, we see that the error E_2 (as for the estimation of the copula parameter) tends to increase with the proportion of censoring : with less information, the model has a poorer performance and it leads to a negative impact on the prediction of each amount.

4 Real data example

4.1 Description of the data

We consider a dataset containing $n = 4724$ observations coming from medical malpractice claims, covering a period of 17 years. Among these claims, 3578 are uncensored and have their final amount known. For the 1146 censored ones, we have at our disposal an additional base giving us the final amount. The estimators and the prediction models do not use the information used in this control database, which will only be used to measure the accuracy of the predictions we develop.

A preliminary treatment has been done on the database. First, the unit of the amounts has been changed in order to preserve confidentiality (multiplication by a constant). Second, some inflation factor has been removed. We explain this procedure in section 4.2 below.

As we can see in Table 5 below, the typical values taken by T when the claim is censored are globally higher than when it is not. Apparently, there is no significant difference between the values of M when $\delta = 0$ and when $\delta = 1$, but as we shall see in the following, a positive correlation between T and M exists, making the information on T crucial in order to predict M for the censored claims.

TABLE 5 – Descriptive statistics on M and T .

Variable	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
M	0.3938	1.8570	2.6820	2.7920	3.5810	8.8180
T	0.05706	2.7860	3.7800	3.937	5.2010	16.7300
M ($\delta = 0$)	0.3938	1.7910	2.6170	2.7400	3.5160	7.8710
M ($\delta = 1$)	0.394	1.881	2.718	2.814	3.599	8.818
T ($\delta = 0$)	0.1819	3.240	4.010	4.2050	5.4130	16.1500
T ($\delta = 1$)	0.05706	2.59000	3.67100	4.00600	4.99700	16.7300

On each claim, some characteristics related to the context of the claim are present, denoted X_1 to X_3 . In Table 6 below, we present a few descriptive statistics about the quantitative variables X_1 and X_2 . X_3 is a binary variable (equal to 1 for 29% of the observations, to 0 for the other 71%).

TABLE 6 – Description of the continuous covariates X_1 and X_2 .

Variable	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
X_1	0	0.2188	0.3750	0.4134	0.5312	1
X_2	0	0.5714	0.7143	0.6391	0.7143	1

4.2 Removing inflation

Various methodologies may be used to correct the effect of inflation on the costs. Among them, using external data sources may be useful. However, this does not seem accurate in our case of medical malpractice data, since the inflation of the amounts does not follow the inflation of the common consuming products. Let M'_i denote the amount of a claim before removing the part due to inflation. We assume that

$$\log M'_i = \mu d_i + \log M_i, \quad (4.1)$$

where d_i is the date at which the i -th claim occurred, and μ is a drift coefficient that we estimate using our database. We assume that $(M_i)_{1 \leq i \leq n}$ are i.i.d. and independent from $(d_i)_{1 \leq i \leq n}$. The dates d_i take their value in $\{0, \dots, D\}$ (where $D + 1$ years are observed).

We then proceed in the following way :

- compute $m_{i,j}$, defined as the average of the claims that occurred on the fiscal year d_i , and which are settled after j years (for (i, j) such that $d_i + j \leq D$), and let $n_{i,j}$ denote the number of such claims ;
- then, under (4.1), $\log m_{i,j} \approx \mu d_i + \lambda_j$, where $\lambda_j = E[\log M_i | T_i = j]$. For each j , we compute $\hat{\mu}_j$ the weighted least-square estimator of the slope μ based on the points $(\log m_{i,j}, d_i)_{i:d_i+j \leq D}$, using the weights $n_{i,j}$, that is we compute

$$(\hat{\lambda}_j, \hat{\mu}_j) = \arg \min_{a,b} \sum_{i:d_i+j \leq D} n_{i,j} (\log m_{i,j} - a - b d_i)^2.$$

- we compute our final estimator of μ ,

$$\hat{\mu} = \frac{\sum_j n_j^{-1/2} \hat{\mu}_j}{\sum_j n_j^{-1/2}},$$

where $n_j = \sum_i n_{i,j}$.

Finally, for each i such that $\delta_i = 1$, compute $\log \hat{N}_i = \log M'_i - \hat{\mu} d_i$, the amount of the claim, once the inflation effect has been removed. In the following, for the sake of simplicity and of clarity, we continue to use the notation N_i instead of \hat{N}_i , since our main task is to illustrate the copula methodology and not to focus on the inflation effect.

Remark 4.1 *To materialize the inflation of the cost, we chose to consider the impact of the origin year of the claim d_i , while it would be probably more natural to use the year at which the claim is settled, f_i . However, this would create a supplementary difficulty, since the year of the settlement is not known in advance. The procedure described in the following permits us to predict the lifetime of the claim, but to achieve this aim, we first need to remove the inflation. Due to this constraint, modeling inflation through (4.1) seems a reasonable compromise.*

4.3 Estimation of the margins

To model the conditional distribution of $T|X = (X^{(1)}, X^{(2)}, X^{(3)})$, we consider a standard linear model for $\log T$ with Gaussian residuals, that is

$$\log T = \alpha_{0,0} + \sum_{j=1}^3 \alpha_{0,j} X^{(j)} + \varepsilon,$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. A Gamma regression assumption is made for the conditional distribution of the logarithm of $\log(1 + M)$ given X , that is $\log(1 + M)|X \sim \Gamma(r(X), \lambda)$, where $\Gamma(r, \lambda)$ denotes the distribution with density

$$f_{r,\lambda}(m) = \frac{m^{r-1}}{\Gamma(r)\lambda^r} \exp(-m/\lambda) \mathbf{1}_{m \geq 0}.$$

A parametric form is given for $r(X)$, that is

$$r(X) = \beta_{0,0} + \beta_{0,1}X^{(1)} + \beta_{0,2}X^{(1)2} + \beta_{0,3}X^{(1)3} + \beta_{0,4}X^{(3)}.$$

The estimated parameters are given in Table 7.

4.4 Copula estimation

We consider three different copula models to estimate the dependence structure, from Clayton, Frank and Gumbel families. The corresponding copula functions are given in Table 2. Estimation of the association parameter is performed using the procedure of section 2.3 in each model, and the results are given in Table 8.

These three copula families lead to quite different dependence structures. A key issue is then to be able to determine which of them is more adapted to describe the distribution of data. To this purpose, we compare the resulting estimators with a nonparametric estimator of the copula function defined by Gribkova and Lopez (2015). The idea behind

TABLE 7 – Estimated parameters for the margins

Parameter	Estimated value
$\alpha_{0,0}$	1.5268
$\alpha_{0,1}$	-0.2755
$\alpha_{0,2}$	-0.1478
$\alpha_{0,3}$	0.0515
σ	0.5970
$\beta_{0,0}$	0.5780
$\beta_{0,1}$	2.1310
$\beta_{0,2}$	-4.4447
$\beta_{0,3}$	2.6880
$\beta_{0,4}$	0.1431
λ	3.8660

TABLE 8 – Estimated association parameters for the different copula models

Copula family	Association parameter
Clayton	0.3343
Frank	2.3276
Gumbel	1.3479

this nonparametric estimator is to generalize the empirical copula estimator of Deheuvels (1979) to a censoring framework. The estimator is defined as

$$\hat{\mathfrak{C}}(a, b) = \hat{F}(\hat{F}_1^{-1}(a), \hat{F}_2^{-1}(b)),$$

where $\hat{F}(t, m) = \sum_{i=1}^n W_{i,n} \mathbf{1}_{Y_i \leq t, N_i \leq m}$ is a nonparametric estimator of the joint distribution function $F(y, m)$, $\hat{F}_1(t) = \hat{F}(t, \infty)$, and $\hat{F}_2(m) = \hat{F}(\infty, m)$. The difference between $\hat{\mathfrak{C}}$ and $\mathfrak{C}_{\hat{\theta}}$ can be visualized in Figure 1.

We then compute, for each copula family, $d(\hat{\mathfrak{C}}, \mathfrak{C}_{\hat{\theta}})$, where d denotes some distance between the copula functions. We consider two distances d_1 and d_2 as criterion,

$$d_1(\hat{\mathfrak{C}}, \mathfrak{C}_{\hat{\theta}}) = \|\hat{\mathfrak{C}} - \mathfrak{C}_{\hat{\theta}}\|_{\infty},$$

$$d_2(\hat{\mathfrak{C}}, \mathfrak{C}_{\hat{\theta}}) = \left\{ \int (\mathfrak{C}(a, b) - \mathfrak{C}_{\hat{\theta}}(a, b))^2 d\hat{\mathfrak{C}}(a, b) \right\}^{1/2}.$$

The distance d_1 corresponds to a Kolmogorov-Smirnov type criterion, while d_2 is close to

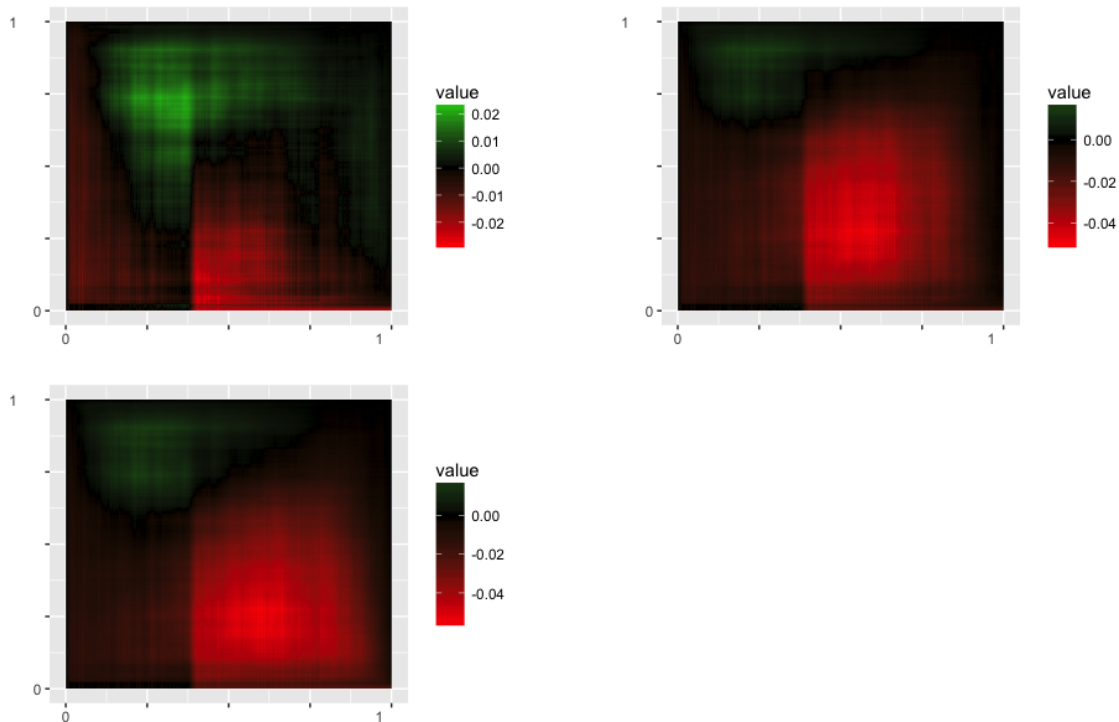


FIGURE 1 – Difference $\hat{\mathfrak{C}}(a, b) - \mathfrak{C}_{\hat{\theta}}(a, b)$. From left to right and top to down : Clayton's copula, Frank's copula, Gumbel's copula.

a Cramer-von Mises approach. The values for the different copula models are summarized in Table 9.

TABLE 9 – Distance d_1 and d_2 for the three copula models.

Copula family	$d_1(\hat{\mathfrak{C}}, \mathfrak{C}_{\hat{\theta}})$	$d_2(\hat{\mathfrak{C}}, \mathfrak{C}_{\hat{\theta}})$
Clayton	0.02827	0.00784
Frank	0.05109	0.02038
Gumbel	0.05695	0.02347

Clayton's copula model leads to the smallest distance according to both criteria d_1 and d_2 . Although this is only an heuristic argument in view of choosing the proper model, empirical results on the prediction of the reserve, as we shall see in section 4.5, also advocate for this model.

4.5 Estimation of the reserve

With at hand the model we obtained for the distribution of (T, M) , we are now able to use the procedure described in section 2.4 in order to estimate the conditional distribution of the open claims (with $\delta_i = 0$). From these individual simulations, we can deduce an estimation of the distribution of the reserve $R = \sum_{i=1}^n (1 - \delta_i) M_i$. Indeed, for each claim with $\delta_i = 0$, we simulate $(T_i^b, M_i^b)_{b=1, \dots, B}$ according to our estimated version of (2.7). We then compute $R^b = \sum_{i=1}^n (1 - \delta_i) M_i^b$.

Figures 2 present histograms of this reserve for the three copula models that we consider, for $B = 10000$. For each copula model, we also compute an estimation of $E[R]$ (that is $B^{-1} \sum_{b=1}^B R^b$), which can be seen as our central scenario. This prediction is compared to the real value of R , using our second database which contains the true realizations of the claims. This real value is $R = 3436.51$. The results are presented in Table 10, with 95% confidence intervals deduced from the estimated distribution of R . To look at the quality of the individual prediction, we also compute the bias term $E_0 = \{\sum_{i=1}^n (1 - \delta_i) [\hat{M}_i - M_i]\} (\sum_{i=1}^n (1 - \delta_i))^{-1}$, the error E_1 as the median of the values $|\hat{M}_i - M_i|$ for i such that $\delta_i = 0$, and $E_2 = \{\sum_{i=1}^n (1 - \delta_i) |\hat{M}_i - M_i|\} (\sum_{i=1}^n (1 - \delta_i))^{-1}$.

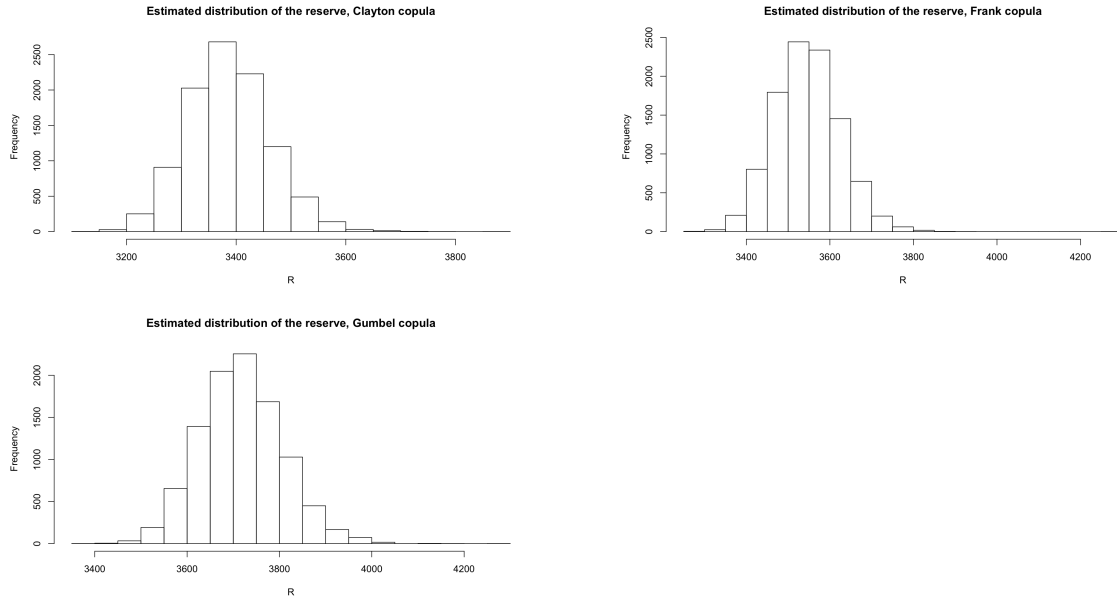


FIGURE 2 – Distribution of the reserve obtain by simulations. From left to right and top to down : Clayton’s copula, Frank’s copula, Gumbel’s copula.

As we can see, Frank’s and Gumbel’s models lead to an overestimation (3.19% of the real value for Frank and a much higher value, 8.18%, for Gumbel). This tendency of

TABLE 10 – Estimation of the reserve and confidence intervals.

Copula family	$E[R]$	95% confidence interval	E_0	E_1	E_2
Clayton	3386.03	[3246.43,3539.08]	-0.0440	0.8541	0.9297
Frank	3546.30	[3401.20,3704.53]	0.09580	0.8722	0.9251
Gumbel	3717.49	[3552.77,3901.84]	0.2451	0.9087	0.9944

obtaining a larger estimator of the reserve by the Gumbel model was expected, due to the right-tail dependence of this copula family. We can also observe that this is caused by a positive bias (0.2451) when estimating each claim amount individually. Here, the true value of the reserve does not belong to the confidence interval. On the other hand, the true value of R belongs to Frank’s copula interval, although close to its lower frontier. Clayton’s model produces the closest prediction of the reserve, with a smaller negative bias. When it comes to looking at the quality of individual prediction, Clayton’s model and Frank’s model give similar performances in terms of errors E_2 , with a small advantage on error E_1 for Clayton. This comforts the fact that Clayton’s model is a reasonable choice among these three models, as proposed at the end of section 4.4.

5 Conclusion and extensions

In this paper, we have presented a new methodology to predict the final amount of an open claim. The procedure is based on using the dependence structure between the amount of the claim and its "lifetime". Using a copula model is a convenient way to understand this dependence, since it allows us to model separately each margin, which is easier than studying them jointly. We have proposed a way to use this model to evaluate the distribution of the reserve corresponding to these open claims. The procedure can be a valuable tool to estimate the amount of each open claim and identify the ones that require more attention (for example by sending an expert to evaluate them). In the practical illustration we presented, no information was available on preliminary evaluations of the claims (by experts) or on partial payments or reevaluations that may occur during the development period. These additional information may be incorporated in the procedure by considering it as covariates, without changing much the structure of the procedure.

Moreover, in the model we have presented, we implicitly have considered that all claims were reported instantly to the insurer. In practice, this is of course not the case. A possibility to correct this issue is to incorporate left-truncation in our procedure. Indeed,

defining τ_i as the delay in reporting the i -th claim, data are composed of claims for which $Y_i \geq \tau_i$. Neglecting the left-truncation phenomenon leads to an overestimation of the typical values taken by T (and, due to the positive correlation between T and M , to a more pessimistic vision of the amount). Adaptation of our technique to left-truncation is straightforward : our estimator of S_C can be replaced by the estimator of Tsai et al. (1987), which changes the definition of the weights we use (see Sellero et al. (2005)).

6 Appendix

This section is devoted to the technical arguments required to obtain the results of section 3.2. The proof of Theorem 3.1 (see section 6.4) relies on several preliminary results. Section 6.1 furnishes general results on weighted empirical means (where the weights are designed, accordingly to our procedure, to compensate for censoring). Section 6.2 shows how to deal with pseudo-observations. Finally, section 6.3 presents consistency results and asymptotic representation that are a cornerstone of the proof of Theorem 3.1.

In the following proofs, we often refer to two classical properties of the empirical distribution function, that we list below. Recall that,

$$\sup_{t \leq Y_{(n)}} |\hat{S}_C(t) - S_C(t)| = O_P(n^{-1/2}), \quad (6.1)$$

where $Y_{(n)} = \max_{i=1, \dots, n} Y_i$, from Donsker's Theorem, see for example van der Vaart (1998). On the other hand, we also have

$$\sup_{t \leq Y_{(n)}} S_C(t) \hat{S}_C(t)^{-1} = O_P(1), \quad (6.2)$$

from Shorack and Wellner (2009).

6.1 Preliminary results on empirical weighted sums

We here give an uniform law of large numbers result for weighted sums (Proposition 6.1).

Proposition 6.1 *Let $\mathcal{F} = \{(t, m, x) \rightarrow f_\tau(t, m, x) : \tau \in \mathcal{T}\}$, where \mathcal{T} is a compact subset of \mathbb{R}^k , a class of R^m -valued functions such that, for some $\tau_0 \in \mathcal{T}$,*

$$\|f_\tau(t, m, x) - f_{\tau'}(t, m, x)\| \leq \|\tau - \tau'\| \phi(t, m, x), \quad (6.3)$$

with

$$E \left[\frac{|f_{\tau_0}(T, M, X)|}{S_C(T)} \right] < \infty, \quad (6.4)$$

$$E \left[\frac{\phi(T, M, X)}{S_C(T)} \right] < \infty. \quad (6.5)$$

Then, assuming that, for all $\tau \in \mathcal{T}$, $f_\tau(t, m, x) = 0$ if $S_C(t) = 0$,

$$\sup_{\tau \in \mathcal{T}} \left| \sum_{i=1}^n W_{i,n} f_\tau(Y_i, N_i, X_i) - E[f_\tau(T, M, X)] \right| = o_P(1).$$

Proof. Let $Z_n(\mathcal{F}) = \sup_{\phi \in \mathcal{F}} |\sum_{i=1}^n W_{i,n} \phi(Y_i, N_i, X_i) - E[\phi(T, M, X)]|$, $W_{i,n}^* = \delta_i S_C(Y_i)^{-1} n^{-1}$, and let D denote the diameter of \mathcal{T} . Decompose $Z_n(\mathcal{F}) \leq Z_n^*(\mathcal{F}) + R_n(\mathcal{F})$, where

$$Z_n^*(\mathcal{F}) = \sup_{\phi \in \mathcal{F}} \left| \sum_{i=1}^n W_{i,n}^* \phi(Y_i, N_i, X_i) - E[\phi(T, M, X)] \right|,$$

$$R_n(\mathcal{F}) = \sup_{\phi \in \mathcal{F}} \left| \sum_{i=1}^n \{W_{i,n}^* - W_{i,n}\} \phi(Y_i, N_i, X_i) \right|.$$

From (2.4), we can rewrite $Z_n^*(\mathcal{F}) = \sup_{g \in \mathcal{G}} |n^{-1} \sum_{i=1}^n g(\delta_i, Y_i, N_i, X_i) - E[g(\delta, Y, N, X)]|$, where $g \in \mathcal{G} = [\delta S_C(Y)^{-1}] \times \mathcal{F}$. From Theorem 2.10.6 in van der Vaart and Wellner (1996), we get that $Z_n^*(\mathcal{F}) = o_P(1)$.

On the other hand,

$$R_n(\mathcal{F}) \leq \sup_{t \in \mathbb{R}} |\hat{S}_C(t) - S_C(t)| \times \sup_{t \leq Y_{(n)}} \left| \frac{S_C(t)}{\hat{S}_C(t)} \right| \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \Phi(Y_i, N_i, X_i)}{S_C(Y_i)^2},$$

where $\Phi(t, m, x) = f_{\tau_0}(t, m, x) + D\phi(t, m, x)$. Next, use (6.1) and (6.2). From (2.4) and (6.4) and (6.5), we get $R_n(\mathcal{F}) = o_P(1)$ (in fact, we even get $R_n(\mathcal{F}) = O_P(n^{-1/2})$). ■

The following result furnishes an asymptotic representation of weighted sums. Similar type of results can be found in Stute (1995) or Stute (1996) when S_C is a Kaplan-Meier estimator, but here with weaker conditions due to the fact that $(C_i)_{1 \leq i \leq n}$ are observed (which is not the case in a classical right-censoring scheme).

Theorem 6.2 *Let ϕ denote a function such that*

$$E \left[\frac{\phi(T, M, X)^2}{S_C(T)} \right] < \infty. \quad (6.6)$$

Then

$$\sum_{i=1}^n W_{i,n} \phi(Y_i, N_i, X_i) = \frac{1}{n} \sum_{i=1}^n \Gamma_\phi(Y_i, N_i, \delta_i, C_i, X_i) + o_P(n^{-1/2}),$$

where Γ_ϕ is defined in Theorem 3.1.

Proof. Decompose

$$\begin{aligned} \sum_{i=1}^n W_{i,n} \phi(Y_i, N_i, X_i) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \phi(Y_i, N_i, X_i)}{S_C(Y_i)} + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i [S_C(Y_i) - \hat{S}_C(Y_i)] \phi(Y_i, N_i, X_i)}{S_C(Y_i)^2} \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{\delta_i [S_C(Y_i) - \hat{S}_C(Y_i)]^2 \phi(Y_i, N_i, X_i)}{S_C(Y_i)^2 \hat{S}_C(Y_i)}. \end{aligned}$$

The third term is $O_P(n^{-1})$ using (6.1), (6.2) and (6.6). The second term can be rewritten as

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_i [S_C(Y_i) - \mathbf{1}_{C_j \geq Y_i}] \phi(Y_i, N_i, X_i)}{S_C(Y_i)^2},$$

which is

$$\frac{1}{n} \sum_{j=1}^n E \left[\frac{\phi(T, M, X) [S_C(T) - \mathbf{1}_{C \geq T}] |C = C_j}{S_C(T)} \right] + o_P(n^{-1/2}),$$

the remainder term being a degenerated second order U -statistic. ■

6.2 Lemma on pseudo-observations

Lemma 6.3 *Let χ denote a positive function, and \mathcal{F} a class of continuous functions such that, for all $\phi \in \mathcal{F}$,*

$$|\phi(u, v)| \leq \mathbf{r}_1(u) \mathbf{r}_2(v),$$

where \mathbf{r}_1 and \mathbf{r}_2 are two functions in \mathcal{R} , such that

$$\sup_x \int \chi(t, m, x) f_T(t|x) f_M(m|x) \left\{ \prod_{i=1,2} \mathbf{r}_i(F_T(t|x), F_U(u|x)) \right\} \mathbf{c}(F_T(t|x), F_M(m|x)) dt dm < \infty. \quad (6.7)$$

Moreover, assume that $(\hat{U}_i)_{1 \leq i \leq n}$ and $(\hat{V}_i)_{1 \leq i \leq n}$ are such that

$$\sup_{i=1, \dots, n} \{ |\hat{U}_i - U_i| + |\hat{V}_i - V_i| \} = o_P(1),$$

and

$$\sup_{i=1, \dots, n} \left| \frac{U_i}{\hat{U}_i} + \frac{1 - U_i}{1 - \hat{U}_i} \right| + \left| \frac{V_i}{\hat{V}_i} + \frac{1 - V_i}{1 - \hat{V}_i} \right| = O_P(1). \quad (6.8)$$

Then,

$$\sup_{\phi \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \chi(Y_i, N_i, X_i)}{S_C(Y_i)} |\phi(\hat{U}_i, \hat{V}_i)| = O_P(1), \quad (6.9)$$

$$\sup_{\phi \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \chi(Y_i, N_i, X_i)}{S_C(Y_i)} |\phi(\hat{U}_i, \hat{V}_i) - \phi(U_i, V_i)| = o_P(1). \quad (6.10)$$

Proof. For $M > 0$, define the event

$$E_{M,n} = \left\{ \inf_{1 \leq i \leq n} \inf \left(\frac{\hat{U}_i}{U_i}, \frac{1 - \hat{U}_i}{1 - U_i} \right) \geq \frac{1}{M} \right\} \cap \left\{ \inf_{1 \leq i \leq n} \inf \left(\frac{\hat{V}_i}{V_i}, \frac{1 - \hat{V}_i}{1 - V_i} \right) \geq \frac{1}{M} \right\}.$$

On $E_{M,n}$, the left hand side in (6.9) is bounded by

$$\frac{M^2}{n} \sum_{i=1}^n \frac{\delta_i \chi(Y_i, N_i, X_i)}{S_C(Y_i)} \mathbf{r}_1(U_i) \mathbf{r}_2(V_i),$$

which as finite expectation from (6.7). Since $\lim_{M \rightarrow \infty} \bar{\lim}_n \mathbb{P}(E_{M,n}) \rightarrow 0$ from (6.8), we get (6.9).

Similarly for (6.10), the left hand-side tends to 0 on $E_{M,n}$ (the expectation tends to zero from Lebesgue's dominated convergence theorem). Once again, one uses that $\lim_{M \rightarrow \infty} \bar{\lim}_n \mathbb{P}(E_{M,n}) \rightarrow 0$ to conclude. ■

6.3 Auxiliary results

Proposition 6.4 is a preliminary result that ensures consistency of the estimator of the association parameter, while Proposition 6.5 furnishes an asymptotic representation of a term which appears in the proof of Theorem 3.1.

Proposition 6.4 *Under Assumptions 2 and 3,*

$$\hat{\theta} - \theta_0 = o_P(1).$$

Proof. For the sake of simplicity, we consider the case $\theta \in \mathbb{R}$. The general case can be obtained by looking at each component separately. Let

$$M_n(\theta) = \sum_{i=1}^n W_{i,n} \log \left\{ \frac{\mathbf{c}_\theta(\hat{U}_i, \hat{V}_i)}{\mathbf{c}_{\theta_0}(\hat{U}_i, \hat{V}_i)} \right\},$$

and $M(\theta) = E[\log \mathbf{c}_\theta(U_i, V_i)] - E[\log \mathbf{c}_{\theta_0}(U_i, V_i)]$. The result of Proposition 6.4 will be obtained if we show that $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| = o_P(1)$. From a first order Taylor expansion, for some $\tilde{\theta}$ between θ and θ_0 , we have

$$M_n(\theta) - M(\theta) = (\theta - \theta_0) \left\{ \sum_{i=1}^n W_{i,n} \phi_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i) - E[\phi_{\tilde{\theta}}(U_i, V_i)] \right\}.$$

Decompose

$$\begin{aligned} \sum_{i=1}^n W_{i,n} \phi_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i) - E[\phi_{\tilde{\theta}}(U_i, V_i)] &= \left\{ \sum_{i=1}^n W_{i,n}^* \phi_{\tilde{\theta}}(U_i, V_i) - E[\phi_{\tilde{\theta}}(U_i, V_i)] \right\} \\ &+ \left\{ \sum_{i=1}^n \{W_{i,n} - W_{i,n}^*\} \phi_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i) \right\} \\ &+ \left\{ \sum_{i=1}^n W_{i,n}^* \{ \phi_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i) - \phi_{\tilde{\theta}}(U_i, V_i) \} \right\} \end{aligned} \quad (6.11)$$

From Proposition 6.1, the first bracket converges towards 0 uniformly in θ . Indeed, the family of functions $\mathcal{F} = \{(t, m, x) \rightarrow \phi_\theta(F_T(t|x), F_M(m|x)) : \theta \in \Theta\}$ satisfies the assumptions of Proposition 6.1, when Assumptions 2 and 3 hold. Indeed,

$$\|\phi_\tau(t, m, x) - \phi_{\tau'}(t, m, x)\| \leq h_1(F_T(t|x)) h_2(F_M(m|x)) \|\tau - \tau'\|,$$

and the integrability conditions ensure that (6.4) and 6.5 hold.

The third bracket in (6.11) is bounded by

$$\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{S_C(Y_i)} \sup_{\theta \in \Theta} |\phi_\theta(\hat{U}_i, \hat{V}_i) - \phi_\theta(U_i, V_i)|,$$

which converges to 0 from Lemma 6.3. The second bracket in (6.11) is bounded by

$$\sup_{t \leq Y_{(n)}} |\hat{S}_C(t) - S_C(t)| \times \sup_{t \leq Y_{(n)}} S_C(t) \hat{S}_C(t)^{-1} \times \frac{1}{n} \sum_{i=1}^n \frac{\delta_i r_1(\hat{U}_i) r_2(\hat{V}_i)}{S_C(Y_i)}.$$

Lemma 6.3, (6.1) and (6.2) show that this term tends to 0. ■

Proposition 6.5 *Under Assumptions 2 to 6,*

$$\begin{aligned} \sum_{i=1}^n W_{i,n} \phi_{\theta_0}(\hat{U}_i, \hat{V}_i) &= \frac{1}{n} \sum_{i=1}^n \Gamma_{\phi_{\theta_0}(F_T(\cdot|\cdot), F_M(\cdot|\cdot))}(Y_i, N_i, \delta_i, C_i, X_i) \\ &+ \frac{1}{n} \sum_{i=1}^n \{ \Lambda_U(Y_i, \delta_i, X_i) + \Lambda_V(N_i, Y_i, \delta_i, X_i) \} + o_P(n^{-1/2}), \end{aligned}$$

where $\Gamma_{\phi_{\theta_0}(F_T(\cdot|\cdot), F_M(\cdot|\cdot))}$ is defined in Theorem 3.1.

Proof. From a first order Taylor expansion,

$$\begin{aligned} \sum_{i=1}^n W_{i,n} \phi_{\theta_0}(\hat{U}_i, \hat{V}_i) &= \sum_{i=1}^n W_{i,n} \phi_{\theta_0}(U_i, V_i) + \sum_{i=1}^n W_{i,n} [\hat{U}_i - U_i] \phi_{\theta_0}^{(1)}(\tilde{U}_i, \hat{V}_i) \\ &\quad + \sum_{i=1}^n W_{i,n} [\hat{V}_i - V_i] \phi_{\theta_0}^{(2)}(\hat{U}_i, \tilde{V}_i), \end{aligned}$$

where \tilde{U}_i (resp. \tilde{V}_i) is between U_i and \hat{U}_i (resp. V_i and \hat{V}_i). The asymptotic expansion of the first term is covered by Theorem 6.2. We only study the first of the two other terms, since both are similar. We have

$$\begin{aligned} \sum_{i=1}^n W_{i,n} [\hat{U}_i - U_i] \phi_{\theta_0}^{(1)}(\tilde{U}_i, \hat{V}_i) &= \sum_{i=1}^n W_{i,n}^* [\hat{U}_i - U_i] \phi_{\theta_0}^{(1)}(\tilde{U}_i, \hat{V}_i) \\ &\quad + \sum_{i=1}^n [W_{i,n} - W_{i,n}^*] [\hat{U}_i - U_i] \phi_{\theta_0}^{(1)}(\tilde{U}_i, \hat{V}_i) \\ &=: A_{1,n} + B_{1,n}. \end{aligned}$$

To show that the term $B_{1,n}$ is negligible, observe that

$$|B_{1,n}| \leq \sup_{1 \leq i \leq n} \frac{|\hat{U}_i - U_i|}{q_1(U_i)} \times \sup_{t \leq Y_{(n)}} |\hat{S}_C(t) - S_C(t)| \times \sup_{t \leq Y_{(n)}} \frac{S_C(t)}{\hat{S}_C(t)} \times \left[\frac{1}{n} \sum_{i=1}^n \frac{\delta_i |\phi_{\theta_0}^{(1)}(\tilde{U}_i, \hat{V}_i)| q_1(U_i)}{S_C(Y_i)^2} \right]. \quad (6.12)$$

Since Assumption 4 ensures that $\sup_{1 \leq i \leq n} |\hat{U}_i - U_i| q_1(U_i)^{-1} = o_P(1)$, and using (6.1)-(6.2), we get $|B_{1,n}| = o_P(n^{-1/2})$ as long as the bracket term in (6.12) is $O_P(1)$, which is obtained using Lemma 6.3.

On the other hand, $A_{1,n}$ can be decomposed into $A_{1,n} = A_{2,n} + B_{2,n}$, where

$$\begin{aligned} A_{2,n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{S_C(Y_i)} [\hat{U}_i - U_i] \phi_{\theta_0}^{(1)}(U_i, V_i), \\ B_{2,n} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{S_C(Y_i)} [\hat{U}_i - U_i] \left\{ \phi_{\theta_0}^{(1)}(U_i, V_i) - \phi_{\theta_0}^{(1)}(\tilde{U}_i, \hat{V}_i) \right\}. \end{aligned}$$

We have $|B_{2,n}| \leq \sup_{i=1, \dots, n} |\hat{U}_i - U_i| q_1(U_i)^{-1} \times n^{-1} \sum_{i=1}^n \delta_i q_1(U_i) S_C(Y_i)^{-1} \{ \phi_{\theta_0}^{(1)}(U_i, V_i) - \phi_{\theta_0}^{(1)}(\tilde{U}_i, \hat{V}_i) \}$. Assumption 4 (equation (3.4)) and Lemma 6.3 show that $|B_{2,n}| = o_P(n^{-1/2})$.

Next, using the decomposition (3.6) of $\hat{U}_i - U_i$,

$$A_{2,n} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left\{ \frac{\delta_i \phi_{\theta_0}^{(1)}(U_i, V_i) \lambda_{(Y_i, X_i)}^U(Y_j, X_j)}{S_C(Y_i)} - \Lambda_U(Y_j, X_j) \right\} + \frac{1}{n} \sum_{j=1}^n \Lambda_U(Y_j, X_j) + o_P(n^{-1/2}).$$

The first term can be decomposed as the sum of a degenerated second order U -statistic

$$\frac{1}{n^2} \sum_{i \neq j} \left\{ \frac{\delta_i \phi_{\theta_0}^{(1)}(U_i, V_i) \lambda_{(Y_i, X_i)}^U(Y_j, X_j)}{S_C(Y_i)} - \Lambda_U(Y_j, X_j) \right\},$$

which is $O_P(n^{-1})$ from Nolan and Pollard (1987), and $n^{-2} \sum_{i=1}^n \{\delta_i \phi_{\theta_0}^{(1)}(U_i, V_i) \lambda_{(Y_i, X_i)}^U(Y_i, X_i) S_C(Y_i)^{-1} - \Lambda_U(Y_i, X_i)\} = O_P(n^{-1})$. ■

6.4 Proof of Theorem 3.1

For the sake of simplicity, we consider the case where $\Theta \subset \mathbb{R}$. The general case can be obtained studying component by component.

From a second order Taylor expansion,

$$\left\{ \sum_{i=1}^n W_{i,n} \Sigma_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i) \right\} (\hat{\theta} - \theta_0) = - \sum_{i=1}^n W_{i,n} \phi_{\theta_0}(\hat{U}_i, \hat{V}_i), \quad (6.13)$$

where $\tilde{\theta}$ is between θ_0 and $\hat{\theta}$. Hence, from Proposition 6.4, $\tilde{\theta} - \theta_0 = o_P(1)$. An asymptotic expansion of the right-hand side of (6.13) is obtained from Proposition 6.5.

Next, decompose

$$\sum_{i=1}^n W_{i,n} \Sigma_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{S_C(Y_i)} \Sigma_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i) + \sum_{i=1}^n (W_{i,n} - W_{i,n}^*) \Sigma_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i).$$

From Lemma 6.3, the first term is $\Sigma + o_P(1)$, while the second can be bounded by

$$\sup_{t \leq Y_{(n)}} |\hat{S}_C(t) - S_C(t)| \times \sup_{t \leq Y_{(n)}} \frac{S_C(t)}{\hat{S}_C(t)} \times \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{S_C(Y_i)^2} \Sigma_{\tilde{\theta}}(\hat{U}_i, \hat{V}_i).$$

Using again Lemma 6.3 and (6.1)-(6.2), we get that this term is $O_P(n^{-1/2})$.

To summarize, we have

$$(\Sigma + o_P(1))(\hat{\theta} - \theta_0) = \frac{1}{n} \sum_{i=1}^n \Psi(Y_i, N_i, \delta_i, C_i, X_i) + o_P(n^{-1/2}),$$

and the result follows from the fact that $\Sigma > 0$.

6.5 Results for the estimation of the margins

In this section, we give some arguments to justify that estimators of the type (2.2) are $n^{1/2}$ -consistent (under some appropriate conditions) and admit an asymptotic representation similar to the one of Assumption 5. The result of Proposition 6.6 below, is similar to the results of Stute (1999).

Proposition 6.6 Let β_0 be the unique maximizer (over a set $\mathcal{B} \subset \mathbb{R}^7$) of $M_m(\beta) = E[l(\beta; T, M, X)]$, and $\hat{\beta} = \arg \max_{\beta \in \mathcal{B}} \sum_{i=1}^n W_{i,n} l(\beta; Y_i, N_i, X_i)$. Let $\lambda(\beta; t, m, x) = \nabla_{\beta} l(\beta; t, m, x)$ the vector of partial derivatives with respect to β , and $H(\beta; t, m, x) = \nabla_{\beta}^2 l(\beta; t, m, x)$. Assume that

$$\|H(\beta; t, m, x) - H(\beta'; t, m, x)\| \leq \|\beta - \beta'\| \phi(t, m, x),$$

for some function ϕ such that $E[\phi(T, M, X)S_C(T)^{-1}] < \infty$, and $E[H(\beta_0; T, M, X)S_C(T)^{-1}] < \infty$. Moreover, assume that $E[\lambda(\beta_0; T, M, X)^2 S_C(T)^{-1}] < \infty$.

Then,

$$\begin{aligned} \hat{\beta} - \beta_0 &= E[H(\beta_0; T, M, X)]^{-1} \times \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \lambda(\beta_0; T_i, N_i, X_i)}{S_C(Y_i)} \right. \\ &\quad \left. + E \left[\frac{\phi(T, M, X)[S_C(T) - \mathbf{1}_{C \geq T}]}{S_C(T)} \middle| C = C_i \right] \right). \end{aligned}$$

The proof of Proposition 6.6 is similar to the one of Theorem 3.1, but simpler since there is no pseudo-observations involved. It is a combination of a Taylor expansion with Proposition 6.1 and Theorem 6.2. From such a representation of $\hat{\beta}$, it is easy to come to a similar asymptotic representation for the pseudo-observations. Indeed, in the case of the GLM, $\hat{F}_{M|X}(N_i) = F_{\hat{\beta}}(\psi(N_i)|\hat{\beta}^T X_i)$ (where $F_{\beta}(\cdot|\beta^T x)$ is the cumulative distribution function associated to $f_{\beta}(\cdot|\beta^T x)$, using the notations of section 2.2.2). A linearization of $F_{\hat{\beta}}(\psi(N_i)|\hat{\beta}^T X_i) - F_{\beta_0}(\psi(N_i)|\beta_0^T X_i)$ using a Taylor expansion allows to obtain the representation.

Références

- 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.

- Ayuso, M. and Santolino, M. (2008). Prediction of individual automobile rbns claim reserves in the context of solvency ii. *IREA–Working Papers, 2008, IR08/006*.
- Burr, D. (1994). On inconsistency of breslow’s estimator as an estimator of the hazard rate in the cox model. *Biometrics*, 50(4) :1142–1145.
- Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2) :269–276.
- Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d’indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65(6) :274–292.
- Fleming, T. R. and Harrington, D. P. (2011). *Counting processes and survival analysis*, volume 169. John Wiley & Sons.
- Gribkova, S. and Lopez, O. (2015). Non-parametric copula estimation under bivariate censoring. *Scandinavian Journal of Statistics*, 42(4) :925–946. 10.1111/sjos.12144.
- Jin, X. (2013). Micro-level loss reserving models with applications in workers compensation insurance. *University of Wisconsin, Empirical paper*.
- 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. *J. Amer. Statist. Assoc.*, 53 :457–481.
- 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., Wuthrich, 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.
- Nolan, D. and Pollard, D. (1987). u -processes : Rates of convergence. *Ann. Statist.*, 15(2) :780–799.

- Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance1. *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.
- Pigeon, M., Antonio, K., and Denuit, M. (2014). Individual loss reserving using paid–incurred data. *Insurance : Mathematics and Economics*, 58 :121 – 131.
- Saluz, A., Bühlmann, H., Gisler, A., and Moriconi, F. (2014). Bornhuetter-ferguson reserving method with repricing. <https://ssrn.com/abstract=2697167>.
- Sellero, C. S., Manteiga, W. G., and Van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scandinavian Journal of Statistics*, 32(4) :563–581.
- Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4) :1384–1399.
- Shorack, G. and Wellner, J. (2009). *Empirical Processes with Applications to Statistics*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).
- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8 :229–231.
- Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.*, 23(2) :422–439.
- Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.*, 23(4) :461–471.
- Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica*, pages 1089–1102.
- Tsai, W.-Y., Jewell, N. P., and Wang, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, 74(4) :883–886.
- 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. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes with applications to statistics*. Springer Series in Statistics. Springer-Verlag, New York.
- 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.
- Wuthrich, M. V. (2016). Machine learning in individual claims reserving. <http://dx.doi.org/10.2139/ssrn.2867897>.
- Wuthrich, 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.