ICA CIA BERLIN 2018

Prediction of the amount of a claim using a copula model in micro-level reserving

Olivier Lopez, Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, Paris

About the speaker

- **Olivier Lopez**
- Professor at Sorbonne Université, Paris
- Director of ISUP (Statistics and actuarial department at Sorbonne Université); Fully qualified member of Institut des Actuaires.

SORBONNE UNIVERSITÉ
CRÉATEURS DE FUTURS DEPUIS 1257

- **Sorbonne Université**
- Actuarial department (ISUP)
- Laboratoire de Probabilités, Statistique et Modélisation (LPSM)

Agenda

Introduction

Conditional copula modeling

Practical example

Appendix

Agenda

Introduction

Conditional copula modeling

Practical example

Appendix

Agenda

Introduction

Claim reserving in non-life insurance

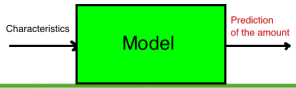
- After the occurrence of a claim, the amount M is not known before a random time T.
- T may be a few years.
- M and T are not independent.
- How to estimate M (and eventually T) to constitue a reserve?

Classical methods

- Aggregated approach: (Chain Ladder)
  - $C_{i,j}$ = cumulated amount of claims for year i, paid at year $i+j$.
  - temporal stability assumption on the rythm of payments.

Accident Year	Reported Losses ($000)						
	c(i,1)	c(i,2)	c(i,3)	c(i,4)	c(i,5)	c(i,6)	c(i,7)
i = 1	584	2,846	3,100	3,674	4,874	5,028	4,034
i = 2	658	3,214	3,888	4,777	5,539	5,925	
i = 3	1,847	4,121	6,662	7,186	7,844		
i = 4	451	2,847	3,361	3,565			
i = 5	1,421	4,879	5,433				
i = 6	325	3,542					
i = 7	1,942						

- Micro-level approach:
  - use the characteristics of a claim to predict its amount



Micro-level reserving

- Use the information one has on the claims (covariates $X \in \mathbb{R}^d$) to predict its amount.
- For applications of this idea in a chain-ladder approach, see Wüthrich, M. V. (2016) (*Machine learning in individual claims reserving*, http://dx.doi.org/10.2139/ssrn.2867897).
- In the micro-level approach, the problem is that we face censoring.

Observations

Bivariate censoring

Observations are n i.i.d. copies of

$$\begin{aligned} Y_i &= \inf(T_i, C_i), \\ \delta_i &= \mathbf{1}_{T_i \leq C_i}, \\ N_i &= \delta_i M_i, \\ X_i &\in \mathbb{R}^d. \end{aligned}$$

Interpretation of these quantities

- $T =$ time between occurrence of the claim and final paiement; $C =$ time between occurrence and end of observation for another cause than the settlement.
- Possible difference with classical right-censoring problems: C_i may be observed.
- Remark: the partial information one may have on the value of the claim can be present in the covariates X_i.

Agenda

ICA CIA BERLIN 2018

Conditional copula modeling

- Copula theory
- Maximum likelihood estimation under censoring
- Weighted log-likelihood
- Theoretical results

Sklar's Theorem

Let T, M be two random variables of c.d.f. F_T and F_M respectively. Let $F(t, m) = \mathbb{P}(T \leq t, M \leq m)$. Then

$$F(t, m) = \mathfrak{C}(F_T(t), F_M(m)),$$

where $\mathfrak{C}$ is a copula function (unique when T and M are continuous).

- A copula function is a c.d.f. on $[0, 1]^2$ with margins uniformly distributed on $[0, 1]$.

Conditional copula modeling

- Modeling of $T|X = x$:
  - Proportional hazard model, or Cox model;
  - Accelerated failure time model;
  - ...
- Modeling of $M|X = x$:
  - Generalized linear model;
  - Extreme value theory;
  - ...
- $\mathfrak{C}_x =$ the copula function for $(T, M)|X = x$.
- Assumption: $\mathfrak{C}_x =: \mathfrak{C}$ does not depend on x.

Parametric copula modeling

- Parametric assumption: $\mathfrak{C} \in \mathcal{C} = \{\mathfrak{C}_\theta; \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$.
- Let us denote θ_0 the "true" parameter, that is $\mathfrak{C} = \mathfrak{C}_{\theta_0}$.
- Associated copula density to $\mathfrak{C}_\theta$: $\mathfrak{c}_\theta$.
- Let $A_i = F_T(T_i|X_i)$ and $B_i = F_M(M_i|X_i)$, in the uncensored case, maximum likelihood estimation is

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log \mathfrak{c}_\theta(A_i, B_i).$$

Difference with the usual case

- Two classical bad ideas:
  - Wrong method number 1:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log \mathfrak{c}_\theta(F_T(Y_i|X_i), F_M(N_i|X_i)),$$

  - Wrong method number 2:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \log \mathfrak{c}_\theta(F_T(Y_i|X_i), F_M(N_i|X_i))^{\delta_i}.$$

- Uncensored and censored observations both bring information (of different nature) on the phenomenon.

MLE under bivariate censoring (Shih and Louis, 1993)

- The log-likelihood is

$$\sum_{i=1}^{n} \delta_i \log \mathfrak{c}_\theta(F_T(Y_i|X_i), F_M(N_i|X_i)) + (1-\delta_i) \log \mathfrak{S}_\theta(F_T(Y_i|X_i), F_M(N_i|X_i)),$$

  where $\mathfrak{S}_\theta$ is the survival function associated to $\mathfrak{C}_\theta$.
- Practical implementation can be difficult.

Alternative idea

- The parameter θ_0 is defined as

$$\theta_0 = \arg\max_{\theta\in\Theta} E\left[\log \mathfrak{c}_\theta(F_T(T_i|X_i), F_M(M_i|X_i))\right] = \arg\max_{\theta\in\Theta} M(\theta).$$

- Idea: find an estimator $M_n(\theta)$ of $M(\theta)$, and define

$$\hat{\theta} = \arg\max_{\theta\in\Theta} M_n(\theta).$$

- We search an estimator M_n of the following type :

$$M_n(\theta) = \sum_{i=1}^{n} W_{i,n} \log \mathfrak{c}_\theta(F_T(Y_i|X_i), F_M(N_i|X_i)).$$

The weights 1/2

- The key is to find appropriate weights that correct the presence of the censoring, that is such that

$$\sum_{i=1}^{n} W_{i,n}\phi(Y_i, N_i, X_i) \rightarrow_{n\rightarrow\infty} E[\phi(T, M, X)] \; a.s.$$

- If $(T, X) \perp C$,

$$W^*_{i,n} = \frac{1}{n}\frac{\delta_i}{1 - G(Y_i-)},$$

would be ok, where $G(t) = \mathbb{P}(C \leq t)$, but G is unknown.

The weights 2/2

- G can be replaced by its Kaplan-Meier estimator...
- ... or by the empirical distribution function of C if $(C_i)_{1\leq i\leq n}$ are observed, that is

$$\hat{G}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{C_i \leq t}.$$

- Let

$$W_{i,n}^{*} = \frac{1}{n}\frac{\delta_i}{1-\hat{G}(Y_i-)}.$$

Final estimator

- Let $\hat{A}_i' = \hat{F}_T(Y_i|X_i)$, and $\hat{B}_i' = \hat{F}_M(N_i|X_i)$.
- With at hand two estimators of the conditional distribution functions of the margins, say $\hat{F}_T(\cdot|x)$ and $\hat{F}_M(\cdot|x)$,

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{1-\hat{G}(Y_i-)} \log \mathfrak{c}_\theta(\hat{A}_i', \hat{B}_i').$$

Theoretical properties

Asymptotic representation of the pseudo-MLE

Under some assumptions,

$$(\hat{\theta} - \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \phi_{1,G}(A_i', B_i', \delta_i) + \phi_{2,G}(Y_i, \delta_i) + \phi_3(A_i', B_i', \delta_i) \right\} + o_P(n^{-1/2}),$$

where $A_i' = F_T(Y_i|X_i)$ and $B_i' = F_M(N_i|X_i)$, where each function ϕ has 0 expectation.

Interpretation of this decomposition

1. $\phi_{1,G}$ corresponds to the asymptotic behavior of the method when the margins and G are known.
2. $\phi_{2,G}$ is due to the fact that G is unknown.
3. ϕ_3 is due to the estimation of the margins.
4. Cov represents the covariance between these three terms.
5. Hence the impact of each of these terms is retrieved in the asymptotic variance, since

$$n^{1/2}(\hat{\theta} - \theta_0) \Longrightarrow \mathcal{N}(0, Var(\phi_{1,G} + \phi_{2,G} + \phi_3 + Cov)).$$

Assumptions

- Assumptions on the copula family: the copula family should be regular enough, and the copula density should not explode to much near the frontier of the unit square.
- See Tsukahara (2005) in *Canadian Journal of Statistics.*

- Assumptions on the margins: the margins should be estimated at $n^{-1/2}-$rate.

Assumptions on the censoring

- Let $\phi_\theta(u,v) = \nabla_\theta \log \mathfrak{c}_\theta(u,v)$.
- Following the assumptions of Tsukahara (2005), we require that

$$|\phi_\theta(u,v)| \leq r_1(u) r_2(v),$$

  where r_1 and r_2 are reproducing $U-$shaped functions (similar assumption on the second variable).
- The strongest assumption (caused by the censoring) that we have is

$$\int \frac{r_1(F_T(t|x))^2 r_2(F_M(m|x))^2}{S_C(t)} d\mathbb{P}(t,m,x) < \infty.$$

Agenda

ICA CIA BERLIN 2018

Practical example

- The dataset
- Model for the margins
- Copula estimation
- Prediction error

Short description of the dataset

- Medical malpractice insurance: 4724 claims observed during 27 years.
- We remove the information from the last 10 years (censoring).
- Inflation "correction" on the amount.
- The average "lifetime" of the claim is around 7 years.
- 1146 censored claims that we wish to predict.

Models for the margins

- $T|X = x$: accelerated failure time model with log-normal distribution, that is, defining

$$\lambda(t|x) = \lim_{dt \to 0^+} \mathbb{P}(T \in [t, t+dt] | T \geq t, X = x),$$

  we assume that

$$\lambda(t|x) = \exp(\beta_0^T x)\lambda_0(\exp(\beta_0^T x)t),$$

  where λ_0 is the hazard rate corresponding to a log-normal variable.

- $\log M$: Gamma GLM.

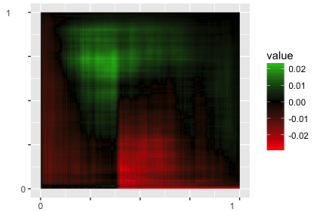
Copula models

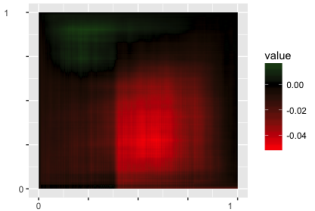
Model	$\mathfrak{C}_\theta(u, v)$	$\hat{\theta}$
Clayton	$\max((u^{-\theta} + v^{-\theta} - 1), 0)^{-1/\theta}$	0.3410
Frank	$-\theta^{-1} \log\left(1 + \frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{(\exp(-\theta)-1)}\right)$	2.3399
Gumbel	$\exp[-\{(-\log u)^\theta + (-\log v)^\theta\}]$	1.3505

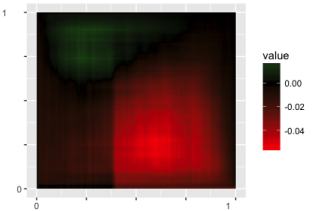
Table: Estimated association parameters for the different copula models

Comparison of the estimated copula and the empirical copula

- A nonparametric estimator of the copula is available in this framework (see Gribkova, Lopez, 2015).
- Difference between the empirical and the fitted copula : (from left to right: Clayton, Frank, Gumbel)







Prediction of open claims

- Consider a claim with:
  - $Y_i = y$,
  - $X_i = x$,
  - $\delta_i = 0$.
- The model can be used to compute an estimation of the distribution of $(T, M)|T \geq y, X = x$, with density $f_{y,x}(t, m)$.
- Predict (T_i, M_i), for example, with

$$(\hat{T}_i, \hat{M}_i) = \arg\max_{t,m} f_{y,x}(t, m).$$

- Simulate B claims according to this distribution (for each claim) to get a vision of the global charge associated with the claims.

Prediction error

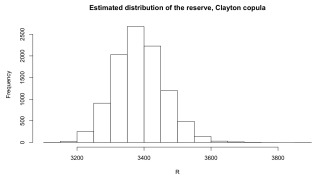
- The true value for the reserve is $R = 3436.51$, we look at the estimation provided by the different models.
- $B = 10000$ simulations, individual error $E = \{\sum_{i=1}^{n}(1-\delta_i)|\hat{M}_i - M_i|\}(\sum_{i=1}^{n}(1-\delta_i))^{-1}$.

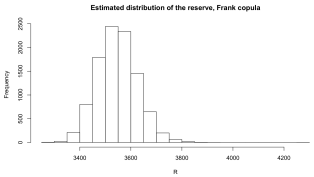
Table: Estimation of the reserve and confidence intervals.

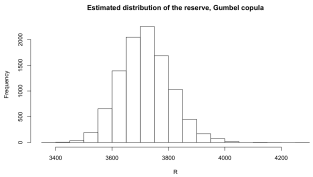
Copula family	$E[R]$	95% confidence interval	E
Clayton	3386.03	[3246.43,3539.08]	0.9297
Frank	3546.30	[3401.20,3704.53]	0.9251
Gumbel	3717.49	[3552.77,3901.84]	0.9944

Distribution of the reserve

- $B = 10000$ simulations.







Conclusion

- Model choice? (Possible solution: comparison with the empirical copula estimator, bootstrap procedure to compute the $p-$values)
- Assumption $\theta(x) = \theta_0$? (Difficulty: the amount of data required to estimate with sufficient precision this function $\theta(x)$)
- Left-truncation: possibility to modify the weighting scheme to introduce left-truncation, see for example: 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.

Thank you very much for your attention!

Contact details:
Olivier Lopez

address: Sorbonne Université, ISUP, Laboratoire LPSM
4 place Jussieu 75005 Paris, France
phone: +33 (0)144277056
mail: olivier.lopez@sorbonne-universite.fr
web: https://sites.google.com/site/olivierlopezcrest/

References

- Preprint: O. Lopez (2018), *A censored copula model for micro-level claim reserving.*
- Available online: https://hal.archives-ouvertes.fr/hal-01706935