Individual Claims Reserving with dependent censored data

Marie Michaelides

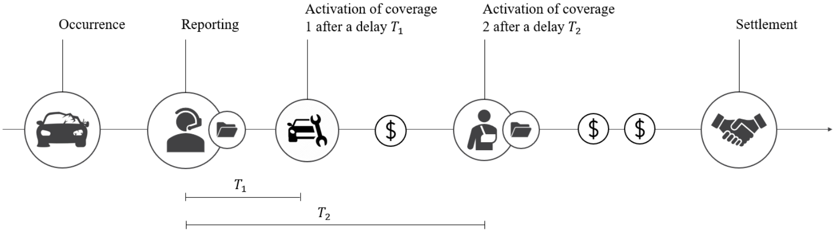
michaelides.marie@uqam.ca

Joint work with Mathieu Pigeon (UQAM) and Hélène Cossette (ULaval)

June 2023

Introduction

- Consider a policy providing two insurance coverages.
- $\boldsymbol{T} = (T_1, T_2)$: activation delays for both coverages.

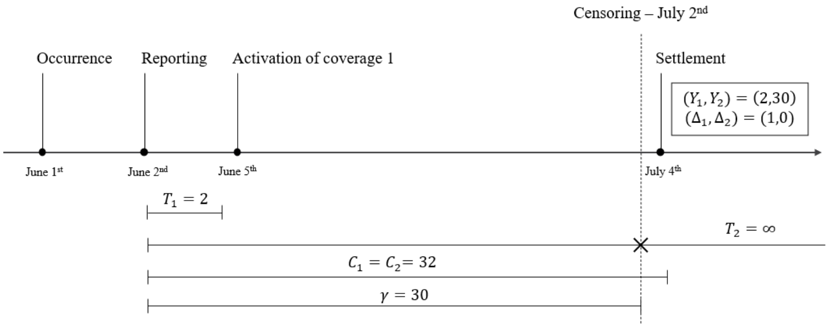


If a claim activates coverage 1, will it also activate coverage 2? If so, how long after?

$\Longrightarrow$ We work with **censored dependent** data.

Notation

- $\boldsymbol{T} = (T_1, T_2)$: vector of true survival times (*activation delays*)
- $\boldsymbol{C} = (C_1, C_2)$ with $C_1 = C_2$: vector of censoring variables (*settlement delays*)
- γ_i for $i = 1, 2$: limits imposed on the survival times
- $\boldsymbol{Y} = (Y_1, Y_2)$ with $Y_i = \min(T_i, C_i, \gamma_i)$ for $i = 1, 2$: observed times
- $\boldsymbol{\Delta} = (\mathbb{1}[Y_1 = T_1], \mathbb{1}[Y_2 = T_2])$: censoring indicators



Graphical comparison of Archimedean copulas

- Archimedean copula: $C_\alpha(u_1, u_2) = \phi_\alpha^{-1}(\phi_\alpha(u_1) + \phi_\alpha(u_2))$

- $Z = C_\alpha(U_1, U_2)$ follows **Kendall**'s distribution:
  $K(z) = z - \frac{\phi(z)}{\phi^{(1)}(z)} = z - \lambda(z)$ for $0 < z \leq 1$

- Knowing $\hat{K}(z)$, a non-parametric estimator of $K(z)$, estimate Kendall's tau and retrieve the copula parameters

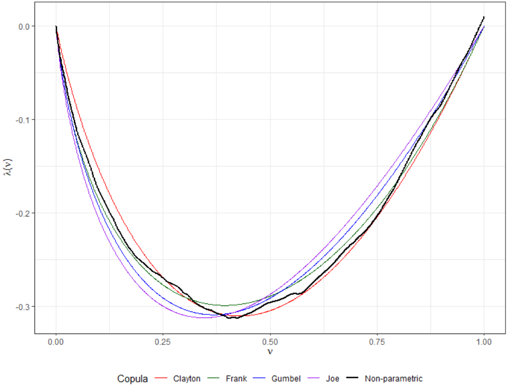
$$\hat{\tau} = 3 - 4\int_0^1 \hat{K}(z)dz \implies \hat{\alpha} = g^{-1}(\hat{\tau})$$

- Compare the plot of $\hat{\lambda}(z) = z - \hat{K}(z)$ to those of $\hat{\lambda}_{\hat{\alpha}}(z)$ to find the most appropriate copula.

Simulation study

Setup:

- $n = 500$ observations
- Clayton copula with $\tau = 0.25$
- Exponential(1) marginals
- $\pm 20\%$ censoring



Results validation

- **Omnibus procedure**: comparison of $\hat{\alpha}$ with $\hat{\alpha}^* = \arg\max L(u_1, u_2, \delta_1, \delta_2; \alpha)$

Copula	$\hat{\alpha}$	$\hat{\alpha}^*$
Clayton	0.723	0.767
Frank	2.540	2.156
Gumbel	1.361	1.190
Joe	1.650	1.127

- **Wang (2010)**'s goodness-of-fit test for censored data

Table 1: Percentage of rejection of the null hypothesis for different copulas.

True copula	τ	Copula under H_0 Clayton p_W	Gumbel p_W
Clayton	0.2	0.30	0.26
	0.4	0.20	0.94
Gumbel	0.2	0.02	0.90
	0.4	0.03	0.72

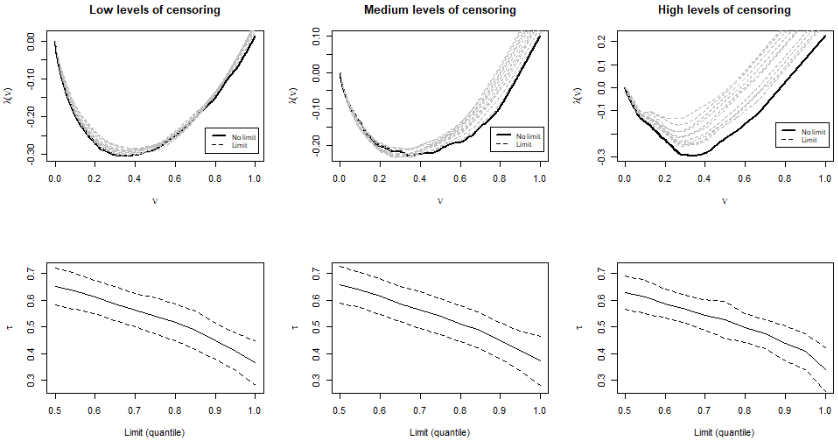
- **L^2-norm** test: $S(\hat{\alpha}) = \int_0^1 \left(\hat{K}(z) - K_{\hat{\alpha}}(z)\right)^2 dK_{\hat{\alpha}}(z)$

Table 2: Percentage of rejection of the null hypothesis for different copulas

True copula	τ	Copula under H_0			
		Clayton	Frank	Gumbel	Joe
		$p_{S(\hat{\alpha})}$	$p_{S(\hat{\alpha})}$	$p_{S(\hat{\alpha})}$	$p_{S(\hat{\alpha})}$
Frank	0.2	1.00	0.20	0.80	1.00
	0.4	1.00	0.36	0.64	1.00
	0.6	1.00	0.34	0.70	0.96
Gumbel	0.2	1.00	0.88	0.46	0.66
	0.4	1.00	0.96	0.14	0.90
	0.6	1.00	0.98	0.02	1.00
Joe	0.2	1.00	1.00	0.82	0.18
	0.4	1.00	1.00	0.84	0.16
	0.6	1.00	1.00	0.84	0.16

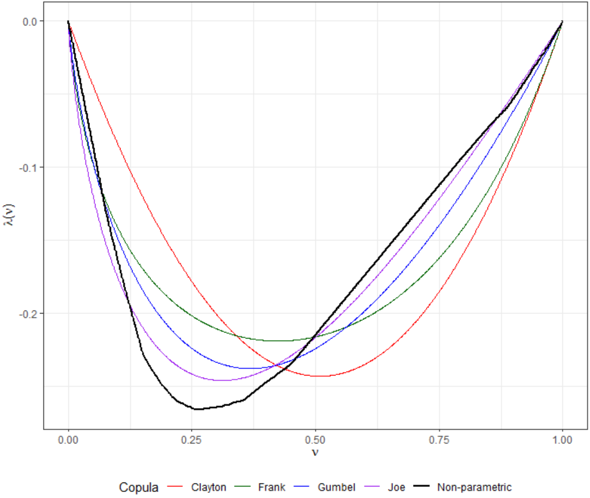
Impact of decreasing γ

Sample of $n = 500$ observations from a Clayton Copula with $\alpha = 0.7$ and margins such that $X_1 \sim \text{LogNormal}(\mu_1 = 8, \sigma_1 = 1)$ and $X_2 \sim \text{LogNormal}(\mu_2 = 7, \sigma_2 = 3)$.



Application to claims reserving

- Automobile insurance dataset with 2 insurance coverages
- Limit: $\gamma_1 = \gamma_2 = 730$ days



Thank you!

Selected references

Akritas, M. (1994), Nearest neighbor estimation of a bivariate distribution under random censoring, *Annals of Statistics*, 22, 1299-1327.

Akritas, M.G., Van Keilegom, I. (2003), Estimation of Bivariate and Marginal Distributions with Censored Data, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65:2, 457-471.

Beran, R. (1981), Nonparametric Regression with Randomly Censored Survival Data, *Berkeley, CA: University of California*.

Denuit, M., Purcaru, O., Van Keilegom, I. (2006), Bivariate Archimedean copula modelling for loss-ALAE data in non-life insurance, *Journal of Actuarial Practice*, 5, 1993-2006.

Genest, C., Rivest, L. (1993), Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of the American Statistical Association*, 88, 1034-1043.

Wang, A. (2010), Goodness-of-fit tests for Archimedean copula models, *Statistica Sinica*, 20, 441-453.

Wang, W., Wells, M. T. (2000), Model selection and semi-parametric inference for bivariate censored data, *Journal of the American Statistical Association*, 95, 62-72.

Appendix

Statistical model

Step 1: Extension of **Beran** (1981)'s estimator

$$\hat{F}_{1|2}(y_1|y_2) = 1 - \prod_{Y_{i1} \leq y_1, \Delta_{i1}=1} \left(1 - \frac{W_{ni2}(y_2; h_n)}{\sum_{j=1}^{n} W_{nj2}(y_2; h_n)\, \mathbb{1}_{Y_{j1} \geq Y_{i1}}}\right)$$

Step 2: Joint distribution $\hat{F}(\boldsymbol{y})$

$$\hat{F}(\boldsymbol{y}) = w(\boldsymbol{y}) \int_0^{y_2} \hat{F}_{1|2}(y_1|z_2) d\tilde{F}_2(z_2) + (1 - w(\boldsymbol{y})) \int_0^{y_1} \hat{F}_{2|1}(y_2|z_1) d\tilde{F}_1(z_1)$$

with $\tilde{F}_j(z_j)$ the marginal estimators of **Kaplan-Meier** (1958).

Step 3: Wang and Wells (2000)'s estimator for $\hat{K}(t)$

$$\hat{K}(t) = \int_0^{\infty} \int_0^{\infty} \mathbb{1}[\hat{F}(\boldsymbol{y}) \leq t] d\hat{F}(\boldsymbol{y})$$

Step 4: Comparison with selected copula models using $\hat{\lambda}(.)$

Popular Archimedean copulas

Table 1: Most commonly used bivariate Archimedean copulas. $\tilde{u} = -\ln u$ and $\bar{u} = 1 - u$.

Copula	$C_\alpha(u_1, u_2)$	$\phi_\alpha(t)$	τ
Clayton	$(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha}$	$t^{-\alpha} - 1$	$\alpha/(\alpha+2)$
Frank	$-\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha u_1}-1)(e^{-\alpha u_2}-1)}{e^{-\alpha}-1}\right)$	$-\ln\left(\frac{e^{-\alpha t}-1}{e^{-\alpha}-1}\right)$	$1 + \frac{4}{\alpha}\left(\int_0^\alpha \frac{\xi}{\alpha(e^\xi-1)}d\xi - 1\right)$
Gumbel-Hougaard	$\exp\left(-[(\tilde{u}_1)^\alpha + (\tilde{u}_2)^\alpha]^{1/\alpha}\right)$	$(-\ln t)^\alpha$	$1 - 1/\alpha$
Joe	$1 - (\bar{u}_1^\alpha + \bar{u}_2^\alpha - \bar{u}_1^\alpha \bar{u}_2^\alpha)^{1/\alpha}$	$-\ln(1-(1-t)^\alpha)$	-

Results validation

Omnibus procedure

- Likelihood function:

$$L(u_1, u_2, \delta_1, \delta_2; \alpha) = \prod_{i=1}^{n} c(u_{1i}, u_{2i}; \alpha)^{\delta_{1i}\delta_{2i}} + \left(\frac{\partial C(u_{1i}, u_{2i}; \alpha)}{\partial u_1}\right)^{\delta_{1i}(1-\delta_{2i})} + \left(\frac{\partial C(u_{1i}, u_{2i}; \alpha)}{\partial u_2}\right)^{(1-\delta_{1i})\delta_{2i}} + C(u_{1i}, u_{2i}; \alpha)^{(1-\delta_{1i})(1-\delta_{2i})}.$$

- Optimal dependence parameter value to compare to that found using $\hat{K}(z)$:

$$\hat{\alpha}^* = \arg\max L(u_1, u_2, \delta_1, \delta_2; \alpha)$$

Likelihood comparison test

Let $L(u_1, u_2, \delta_1, \delta_2; \alpha_i^*)$ be the likelihood function associated to copula model i for $i = 1, ..., M$ with α_i^* the dependence parameter estimated via the omnibus procedure.

Copula 1 provides the closest fit to the data among M copulas under H_0:

$$H_0 : \min_{k=2,\dots,M} \hat{L}(u_1, u_2, \delta_1, \delta_2; \hat{\alpha}_k^*) - \hat{L}(u_1, u_2, \delta_1, \delta_2; \hat{\alpha}_1^*) > 0$$

$$H_1 : \min_{k=2,\dots,M} \hat{L}(u_1, u_2, \delta_1, \delta_2; \hat{\alpha}_k^*) - \hat{L}(u_1, u_2, \delta_1, \delta_2; \hat{\alpha}_1^*) \leq 0$$

L^2-norm

$$S(\hat{\alpha}) = \int_0^1 \left(\hat{K}(z) - K_{\hat{\alpha}}(z)\right)^2 dK_{\hat{\alpha}}(z).$$

Riemann sum approximate:

$$\hat{S}(\hat{\alpha}) = \sum_{i=1}^{n} \left(\hat{K}(z_{(i)}) - K_{\hat{\alpha}}(z_{(i)})\right)^2 (z_{(i)} - z_{(i-1)}),$$

with $z_0 = 0$ and $z_{(1)} \leq ... \leq z_{(n)}$ are the ordered values of $\{z_j = \hat{F}(\boldsymbol{y_j}), j = 1, ..., n\}$.

Copula 1 provides the closest fit to the data among M copulas under H_0:

$$H_0 : \min_{k=2,\ldots,M} \hat{S}(\hat{\alpha}_k) - \hat{S}(\hat{\alpha}_1) > 0$$

$$H_1 : \min_{k=2,\ldots,M} \hat{S}(\hat{\alpha}_k) - \hat{S}(\hat{\alpha}_1) \leq 0$$

Wang (2010)'s goodness-of-fit test

Let $C_\alpha(u,v) = \phi_\alpha^{-1}\left[\phi_\alpha(u) + \phi_\alpha(v)\right]$ be an Archimedean copula. Then,

$$U = \frac{\phi(u)}{\phi[C(u,v)]}, \quad V = C(u,v)$$

are shown to be independent by Genest and Rivest (1993), leading to the test

$$H_0 : \rho = 0 \quad \text{vs} \quad H_1 : \rho \neq 0.$$

Let

$$r_n = \frac{\sum_{i=1}^n (\hat{U}_i - \bar{\hat{U}})(\hat{V}_i - \bar{\hat{V}})}{\sqrt{\sum_{i=1}^n (\hat{U}_i - \bar{\hat{U}})^2 \sum_{i=1}^n (\hat{V}_i - \bar{\hat{V}})^2}}$$

The test statistic is defined as

$$Z_n = \frac{1}{2}\log\left[\frac{1 + r_n}{1 - r_n}\right]$$

and we have that $\sqrt{n}Z_n \to N(0,1)$ in distribution. This test can be extended to censored data by adjusting the distributions of $\hat{U}$ and $\hat{V}$.

Application to claims reserving

Notation:

- $\boldsymbol{T} = (T_1, T_2)$: vector of activation delays (in days)
- $\boldsymbol{C} = (C_1, C_2)$ with $C_1 = C_2$: vector of censoring variables corresponding to the claim settlement delay (in days)
- γ_i for $i = 1, 2$: limits imposed on the activation delays **= 730 days**
- $\boldsymbol{Y} = (Y_1, Y_2)$ with $Y_i = \min(T_i, C_i, \gamma_i)$ for $i = 1, 2$: observed delays
- $\boldsymbol{\Delta} = (\mathbb{1}[Y_1 = T_1], \mathbb{1}[Y_2 = T_2])$: censoring indicators

Table 3: Percentage of claims with different activation delays (1 period = 6 months).

	Activation delays				
Coverage	No delay	1 period	2 periods	3 periods	$\geq$ 4 periods
Accident Benefits	93.84	5.73	0.29	0.08	0.06
Bodily Injury	85.86	9.86	1.46	1.13	1.69