Are Frequency and Severity independent in health insurance?

Santiago Mosquera - Daza[1] Norman Maldonado[2]

11.02.2021

[1]Corresponding author: santiago.mosquedaza@gmail.com

[2]Research center on health economics and social protection (PROESA) - Icesi University, Santiago de Cali, Colombia: normanmva@gmail.com

Content

Content

Content

Context ...

System [1–7]

- Health System
  - IPS (IHSP)
  - **EPS** (HMO)
    - SR
    - **CR**
  - Population
  - Other institutions

- IHSP: Institutional Health Service Providers.
- HMO: Health Maintenance Organization.

Contributory Regime - CR [1, 8–12]

VHI (Commercial Premiums)

- VHI (CRM) [13]
  - Compound Processes
  - **Separated Processes**

Separated Processes [13–15]

- Claim.
- Exposure.
- Pure Premium.
- **Frequency**.
- **Severity**.

Content

Motivation

> **Factors**: Exposure $\rightarrow$ Frequency $\leftrightarrow$ Severity $\leftarrow$ Exposure.

Motivation

> Frequency - Severity Independence (FSI) assumption:

- It is "*frequently adopted in the non-life insurance technique*" [13] and it is an *a priori* condition [16, 17].
- It provides some advantages: Statistical inference, less computationally-intensive processes, ... [17–20].
- It is restrictive and invalid [17, 21, 22]
- It could negatively impact on:
  - Total loss (underestimation [9, 22] or overestimation [21]).
  - Commercial premiums (*Rate making*) [16].
  - Risk control and risk management [17, 23].
- The correct modelling is relevant in public health: Systemic risk and insurance issues [23].

Goal

Objective

We aim to provide evidence of the relationship between the claim frequency and claim severity by testing the independence assumption hypothesis using the Multivariate Approach, which would also derives the dependence structure.

Content

Content

Model

$$S = \begin{cases} 0 & N = 0 \\ \sum_{i=1}^{N} X_i & N > 0 \end{cases} \quad (1)$$

- $N \in \mathbb{Z}^* \mid \mathbb{Z}^* = \{0\} \cup \mathbb{Z}^+$
- $Y_j \in \mathbb{R}^* | \mathbb{R}^* = \{0\} \cup \mathbb{R}^+$
- $X_j = \Pi(Y_j)$ tal que $X_j \leq Y_j$.

Univariate Approach

FSM-IUA

E$[S] = \mathbb{E}[X_1]\,\mathbb{E}[N]$

[13, 14, 24]

FSM-CUA

$f(N,X) =$
$f(N) \times f(X|N)$

[16, 17, 19, 20]

Multivariate Approach
FSM-MA

$f(X,N) = h(X,N) \leftarrow H(X,N)$ [21, 22, 25]

Content

Test

FSM - CUA

$$f\left(N, X; \Theta\right) = f(N; \theta_1) \times f(X \mid N; \theta_2)$$
$$\theta_2 = \{\theta^{X|N}, \theta_X\}$$

- $\hat{\theta}_{X|N} = 0$
- $\hat{\theta}_{X|N} > 0$
- $\hat{\theta}_{X|N} < 0$

FSM - MA

$$f\left(N, X; \Theta\right) = H\left(N, X; \gamma\right)$$

- $\hat{\gamma} = 0$
- Association measures: $\delta\left(\cdot\right)$

Literature

Non-health insurance

- Collision insurance: small $\hat{\theta}_{X|N} < 0$ [17].
- Liability insurance: $\hat{\theta}_{X|N} > 0$ [20].
- Car insurance:
  - $\hat{\theta}_{X|N} > 0, \quad \hat{\delta} > 0, \quad \hat{\gamma} \neq 0$ [16].
  - $\hat{\delta} > 0, \quad \hat{\gamma} \neq 0$ - Underestimation [22].
  - $\hat{\delta} < 0, \quad \hat{\gamma} \neq 0$ - Overestimation [21].

Health insurance

- Prediction models for expenditure [26]:
  - Inpatient: $\hat{\theta}_{X|N} = 0$.
  - Outpatient: $\hat{\theta}_{X|N} > 0$.
- Bivariate distributions of drug claims and other health claims: $\hat{\gamma} \neq 0$ [27].
- Health insurance and health care demand (married couples) [28].
- CRM with dependencies between Frequency and Severity with different hospital settings [29].

Content

Content

Definition

- F - CA: Frequency (N_i)- Claim Amount (X_i) [13, 17].
- F - ACA: Frequency (N_i) - Average Claim Amount (X_i/N_i) [10, 12, 22].
- Exposure: Sum of affiliates weighted by their period of membership.

Method

Independence test

Independence $\Rightarrow (\delta = 0)$

$(\delta \neq 0) \Rightarrow$ Dependence

$$H_o : \delta = 0$$

Dependence Structure

Measures of association [30–36]

- Pearson: r_P.
  - Linearity.
  - Normality.
- Spearman: ρ
- Kendall: τ

Copula [31, 33, 34, 37–44]

1. Copula specification.
   - Parametric estimation: IFME (*Inference Functions for Margins Estimator*).
   - Non-parametric estimation: MPLE (*Maximum Pseudo-Likelihood Estimator*).
2. Measures of association with transformation of the copula parameter:
   - Kendall: τ
   - Blomqvist: β
   - Lower Tail: λ_L
   - Upper Tail: λ_U

Copula

Theorem (Sklar's theorem)

$$H\left(x,y\right)=C\left(F(x),G(y)\right)$$

where H is a joint distribution function with margins F and G for random variables X and Y, respectively, the copula C is unique if F and G are continuous. Although the claim frequency marginal is discrete, which means that C is uniquely determined on Ran$F \times$ Ran$G$ [33, Theorem 2.3.3].

However, the copula representation and its non-uniqueness do not "(...) inhibit empirical applications" [40]

Copula Estimation and Association

IFME

1. Theoretical distributions [41]
   - Severity: *Gamma, Log-Normal, Exponential, Log-Logistic (Fisk), Pareto and Gumbel*.
   - Frequency: *Geometric, Poisson and Negative Binomial*.
2. MLE.

PMLE [38]

1. Pseudo-observations.
2. MLE.

Fréchet-Hoeffding bounds [31, 33, 43]

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u})$$

- $C(\mathbf{u}) \to M(\mathbf{u})$: Comonotonic $\equiv \delta(N, X) = 1$.
- $C(\mathbf{u}) \to W(\mathbf{u})$: Countermonotonic $\equiv \delta(N, X) = -1$.

Copulas [33, 38]

- Gaussian, Clayton, Frank, Gumbel, Joe, BB6 and BB8.
- Selection by AIC and Kendall's method (GoF: White and Kendall) [38].

Content

Data

> **Data:** HMO - VHI (EPS-PAC): 35,744 affiliates.

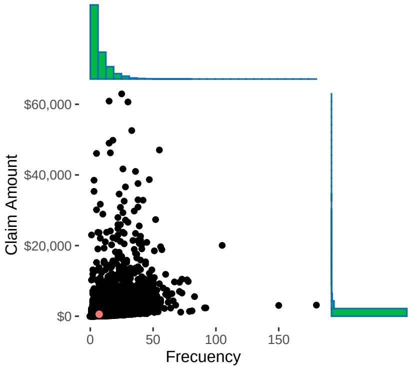
Table: Descriptive statistics

Variables	All	Female	Male	Over 60 y/o	Under 60 y/o
Claim Amount	$ 19,788,384.7	$ 12,281,185.0	$ 7,507,199.8	$ 5,600,321.0	$ 14,188,063.8
Number of Claims	254,155	166,883	87,272	54,523	199,632
Affiliates	35,744	20,523	15,221	4,724	31,020
Exposure	32,275	18,562	13,713	4,536	27,739
Severity	$ 77.9	$ 73.6	$ 86	$ 102.7	$ 71.1
Claim Amount per Person	$ 553.6	$ 598.4	$ 493.2	$ 1,185.5	$ 457.4
Frequency	7.9	9.0	6.4	12.0	7.2
Pure Premium	$ 613.1	$ 661.6	$ 547.5	$ 1,234.6	$ 511.5

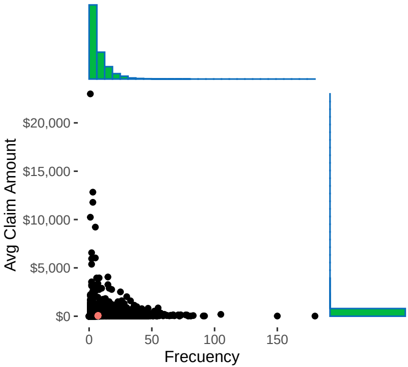
Source: Authors

Features

Figure: Marginal Plots of F-CA and F-ACA



(a) F-CA



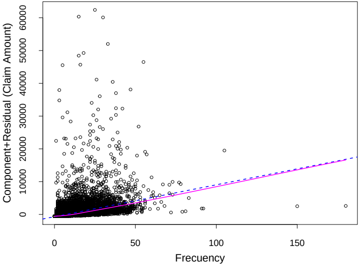
(b) F-ACA

Source: Authors

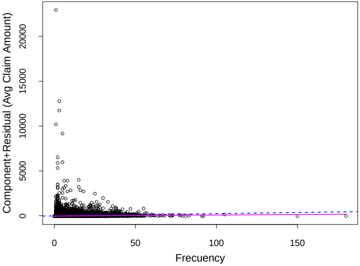
Content

Content

Gaussian?



(c) F-CA



(d) F-ACA

Figure: Partial residuals plots of F-CA and F-ACA

Gaussian?

Table: Shapiro-Wilk test

Group	**Statistic**	**p - value**
F-CA	0.2184	< 0.01
F-ACA	0.234	< 0.01

Source: Authors

Measures of Association

Table: Measures of association

Group	Method	Outliers Yes	Outliers No
F-CA	$r_{\mathbf{P}}$	0.4482***	0.6733***
	$\rho_{\mathbf{S}}$	0.9102***	0.9099***
	$\tau_{\mathbf{K}}$	0.7732***	0.7726***
F-ACA	$r_{\mathbf{P}}$	0.0934***	0.1615***
	$\rho_{\mathbf{S}}$	0.6707***	0.665***
	$\tau_{\mathbf{K}}$	0.5115***	0.5059***

Source: Authors

Content

IFME

Table: Fitted theoretical distributions by IFME

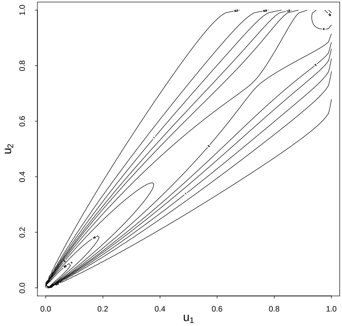
Variable	Distribution	Parameters	Estimate	Std.Error	p-value
Frequency	Negative Binomial	θ_{NB}	0.803	0.007	< 0.01
		μ_{NB}	6.915	0.043	< 0.01
Claim Amount	Log-normal	μ_{LN}	5.302	0.009	< 0.01
		σ_{LN}	1.539	0.006	< 0.01
Average Claim Amount	Log-logistic	β_{Ll}	1.946	0.010	< 0.01
		α_{Ll}	34.856	0.184	< 0.01

Source: Authors

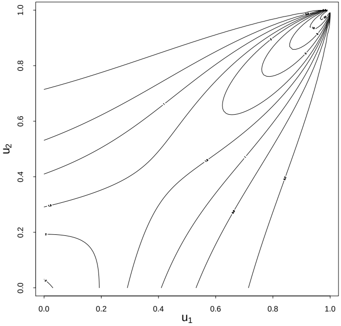
Copula

- Parametric estimation: IFME
  - Frequency: Negative Binomial.
  - Severity 1: Log-Normal.
  - Severity 2: Log-logistic (Fisk).
- Non-parametric estimation: MPLE:
  - Pseudo-observations.
  - Graphics of F-CA Non-parametric.

Survival BB6 (θ_1=5.764, θ_2=1.37)



Joe (θ=3.216)



Copula and Association

Table: Association measures by Copula

Class	Group	Method	Copula	Parameter 1	Paramater 2	τ_K	β_B	λ_L	λ_U
Parametric	F-CA	AIC	Survival BB8	5.214	1	0.56	0.606	0	0
		Kendall	Joe	2.672		0.475	0.486	0	0.704
	F-ACA	AIC	Survival BB8	6	0.438	0.314	0.35	0	0
		Kendall	Joe	1.443		0.2	0.191	0	0.383
Non-parametric	F-CA	AIC	Survival BB6	5.764	1.37	0.791	0.819	0.908	0
		Kendall	Joe	3.216		0.543	0.561	0	0.759
	F-ACA	AIC	Survival Joe	3.867		0.603	0.629	0.804	0
		Kendall	Joe	1.613		0.255	0.248	0	0.463

Source: Authors

Content

Limitations

- Copula: Gaussian, Clayton, Gumbel, Frank, Joe, BB8, and BB6.
- No Zero-Inflated Models.
- The VHI covers different hospital settings.
- The VHI is a “Plan de Atención Complementaria": Dependence factors $\leftarrow$ e.g. Hospital Network.

Content

Conclusion

What did we get?

- Rejection of the H_o : FSI.
- No linearity - No bivariate normal.
- IFME underestimates the association.
- Positive association.
- Tail dependence.
- The structure: Joe Copula is remarkable.

What's next?

> Following [22], if the association is positive, it could be an underestimation of the total loss in car insurance.

- **Warning**: Does the FSI lead to a underestimation of total loss in health insurance?
  - Premium reserves.
  - VHI.
  - Compulsory Health Insurance and Social Health Insurance.

Content

Copula Regression

- **Conditional copulas** [34, 38]: *They capture the dependence among the components of* $\mathbf{X}$ *conditionally on* $\mathbf{Z}$ [38].
- If C has a density (mixed derivate of order d): $c \Rightarrow$ from the CDF $H_Z(\mathbf{X}) = C_z\left(F_{Z,1}(x_1), \ldots, F_{Z,d}(x_d)\right)$, the PDF is:

$$h_z = c_{\varphi(z)}\left(F_{1,\varphi_1(z)}(x_1), \ldots, F_{d,\varphi_d(z)}(x_d)\right)\prod_{j=1}^{d} f_{j,\varphi_j(z)}(x_j)$$

- Conditional log-likelihood function of the model is:

$$L_n(\beta, \beta_{\mathbf{1}}, \ldots, \beta_{\mathbf{d}}) = \sum_{i=1}^{n} \log c_{\phi}(\mathbf{Z}_i)\left(F_{1,\phi_1(\mathbf{Z}_i)}\left(X_{i1}\right), \ldots, F_{d,\phi_d(\mathbf{Z}_i)}\left(X_{id}\right)\right) +$$
$$\sum_{j=1}^{d}\sum_{i=1}^{n} \log f_{j,\phi_j(\mathbf{Z}_i)(X_{ij})}$$

Copula Regression: Traditional case

- $d = 2$ y $n = 35,744^*$

$$L_n(\beta, \beta_{\mathbf{1}}, \beta_{\mathbf{2}}) = \overbrace{\sum_{i=1}^{n} \log c_{\phi}(\mathbf{Z}_i)\left(F_{1,\phi_1(\mathbf{Z}_i)}\left(X_{i1}\right), F_{2,\phi_2(\mathbf{Z}_i)}\left(X_{i2}\right)\right)}^{JoeCopula} + \underbrace{\sum_{i=1}^{n} \log f_{1,\phi_1(\mathbf{Z}_i)(X_{i1})}}_{Poisson} + \underbrace{\sum_{i=1}^{n} \log f_{2,\phi_2(\mathbf{Z}_i)(X_{i2})}}_{Gamma}$$

- Tweedie: *Compound Poisson-gamma process* [11, 14].
- FSM - IUA: Frequency (Poisson) - Severity (Gamma).
- FSM - MA: Frequency (Poisson) - Severity (Gamma).

Example

Table: Estimated Total Loss

Pure Premium	Value	Std. Dev.	Lower Conf. Int	Upper Conf. Int	P-value
Tweedie	667.49	276.79	661.07	673.90	< 0.001
Ind. Poisson – Gamma	631.49	324.56	623.97	639.02	< 0.001
Ind. NB – Gamma	631.40	313.54	624.13	638.67	< 0.001
Cop. Poisson – Gamma	631.28	324.78	623.75	638.81	< 0.001
Cop. NB – Gamma	650.50*	318.65	643.12	657.89	< 0.001

Source: Authors

$$\frac{\hat{E}_{\text{independence}}(PP)}{\hat{E}_{\text{Joe}}(PP)} = 0.97^{**}$$

* Greater than real (613.1).

** 0.934 with Clayton for car insurance [22] - **Preliminary Findings**

- Marginal misspecification $\Rightarrow$ Fisk: Pure Premium (Ind)= USD 598.41.

¡Muchas gracias!

Referencias I

[1] R. Guerrero et al. "Sistema de salud de Colombia". In: *Salud Pública de México* 53.2 (2011), pp. 144–155.

[2] R. Castaño and A. Zambrano. "Biased selection within the social health insurance market in Colombia". In: *Health Policy* 79.2-3 (2006), pp. 313–324. DOI: 10.1016/j.healthpol.2006.01.010.

[3] R. Rodriguez-Monguió and A. Infante. "Universal health care for Colombians 10 years after Law 100: Challenges and opportunities". In: *Health Policy* 68.2 (2004), pp. 129–142. DOI: 10.1016/j.healthpol.2003.10.004.

[4] J. L. Londoño and J. Frenk. "Structured pluralism: Towards an innovative model for health system reform in Latin America". In: *Health Policy* 41.1 (1997), pp. 1–36. DOI: 10.1016/S0168-8510(97)00010-9.

[5] N. Paredes Cubillos. "¿Es necesario repensar el cálculo de la UPC?" In: *Conexxión2* (2015), pp. 12–17.

[6] Congreso de la República de Colombia. *Ley 100 de 1993*. 1993.

[7] Congreso de la República de Colombia. *Ley 1122 de 2007*. 2007.

[8] Congreso de la República de Colombia. *Ley 1438 del 2011*. 2011.

[9] D. Cummins and L. Wiltbank. "Estimating the Total Claims Distribution Using Multivariate Frequency and Severity Distributions". In: *The Journal of Risk and Insurance* 50.3 (1983), p. 377. DOI: 10.2307/252434.

[10] D. Cummins. "Statistical and Financial Models of Insurance Pricing and the Insurance Firm". In: *The Journal of Risk and Insurance* 58.2 (1991), p. 261. DOI: 10.2307/253237.

[11] B. Jorgensen. "Exponential Dispersion Models". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 49.2 (1987), pp. 127–162. DOI: https://doi.org/10.1111/j.2517-6161.1987.tb01685.x.

[12] P. Shi and J. Guszcza. "Frameworks for General Insurance Ratemaking: Beyond the Generalized Linear Model". In: *Predictive Modeling Applications in Actuarial Science*. Ed. by E. W. Frees, G. Meyers, and R. A. Derrig. Vol. 2. International Series on Actuarial Science. Cambridge University Press, 2016, pp. 100–125. DOI: 10.1017/CBO9781139342681.005.

Referencias II

[13] E. Pitacco. *Health Insurance*. Trieste: Springer, 2014. DOI: 10.1007/978-3-319-12235-9.

[14] M. Goldburd, A. Khare, and D. Tevet. *Generalized linear models for insurance data*. 5. 2008, pp. 1–196. DOI: 10.1017/CBO9780511755408.

[15] G. Werner et al. *Basic Ratemaking*. 5th ed. Casualty Actuarial Society, 2016, pp. 1–423.

[16] P. Shi, X. Feng, and A. Ivantsova. “Dependent frequency-severity modeling of insurance claims”. In: *Insurance: Mathematics and Economics* 64 (2015), pp. 417–428. DOI: 10.1016/j.insmatheco.2015.07.006.

[17] J. Garrido, C. Genest, and J. Schulz. “Generalized linear models for dependent frequency and severity of insurance claims”. In: *Insurance: Mathematics and Economics* 70 (2016), pp. 205–215. DOI: 10.1016/j.insmatheco.2016.06.006.

[18] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss models: from data to decisions, 3rd ed.* Wiley Online Library, 2012. DOI: 10.1002/9780470391341.

[19] W. Lee, S. C. Park, and J. Y. Ahn. “Investigating dependence between frequency and severity via simple generalized linear models”. In: *Journal of the Korean Statistical Society* 48.1 (2018), pp. 13–28. DOI: 10.1016/j.jkss.2018.07.003.

[20] G. Y. Lee and P. Shi. “A dependent frequency–severity approach to modeling longitudinal insurance claims”. In: *Insurance: Mathematics and Economics* 87 (2019), pp. 115–129. DOI: 10.1016/j.insmatheco.2019.04.004.

[21] C. Czado et al. “A mixed copula model for insurance claims and claim sizes”. In: *Scandinavian Actuarial Journal* 4 (2012), pp. 278–305. DOI: 10.1080/03461238.2010.546147.

[22] N. Krämer et al. “Total loss estimation using copula-based regression models”. In: *Insurance: Mathematics and Economics* 53.3 (2013), pp. 829–839. DOI: 10.1016/j.insmatheco.2013.09.003. arXiv: 1209.5356.

[23] G. E. Rejda and M. M. J. *Principles of Risk Management and Insurance*. 12th ed. Pearson, 2014.

[24] H. Farhadi. *Aggregate Loss Models*. 2016.

Referencias III

[25] A. R. U. Kholifah, D. Lestari, and S. Devila. "Premium calculation using marginal generalized linear model combined with copula". In: *AIP Conference Proceedings* 2168.November (2019). DOI: 10.1063/1.5132462.

[26] F. Edward W. Frees PhD, J. G. PhD, and F. Marjorie A. Rosenberg PhD. "Predicting the Frequency and Amount of Health Care Expenditures". In: *North American Actuarial Journal* 15.3 (2011), pp. 377–392. DOI: 10.1080/10920277.2011.10597626.

[27] J. F. Carriere. "Testing independence in bivariate distributions of claim frequencies and severities". In: *Insurance: Mathematics and Economics* 21.1 (1997), pp. 81–89. DOI: 10.1016/S0167-6687(97)88952-4.

[28] D. M. Zimmer and P. K. Trivedi. "Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand". In: *Journal of Business and Economic Statistics* 24.1 (2006), pp. 63–76. DOI: 10.1198/073500105000000153.

[29] V. Erhardt and C. Czado. "Modeling dependent yearly claim totals including zero claims in private health insurance". In: *Scandinavian Actuarial Journal* 2 (2012), pp. 106–129. DOI: 10.1080/03461238.2010.489762.

[30] N. Wermuth and D. R. Cox. "Statistical Dependence and Independence". In: *Wiley StatsRef: Statistics Reference Online* 1980 (2014), pp. 4260–4264. DOI: 10.1002/9781118445112.stat05968.

[31] P. Embrechts, A. J. McNeil, and D. Straumann. "Correlation and Dependence in Risk Management: Properties and Pitfalls". In: (2002). Ed. by M. A. H. Dempster, pp. 176–223. DOI: 10.1017/CBO9780511615337.008.

[32] P. Kumar. "Copulas: Distribution Functions and Simulation". In: *International Encyclopedia of Statistical Science*. Ed. by M. Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 309–312. DOI: 10.1007/978-3-642-04898-2_191.

[33] R. Nelsen. *An Introduction to Copulas*. Springer, 2006, pp. 1–656. DOI: https://doi.org/10.1007/0-387-28678-0.

[34] H. Joe. *Multivariate Models and Dependence Concepts*. 1st ed. Chapman and Hall/CRC, 1997. DOI: https://doi.org/10.1201/9780367803896.

Referencias IV

[35] L. Puka. "Kendall's Tau". In: *International Encyclopedia of Statistical Science*. Ed. by M. Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 713–715. DOI: 10.1007/978-3-642-04898-2_324.

[36] R. Modarres. "Measures of Dependence". In: *International Encyclopedia of Statistical Science*. Ed. by M. Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 804–805. DOI: 10.1007/978-3-642-04898-2_359.

[37] B. Rémillard. "Tests of Independence". In: *International Encyclopedia of Statistical Science*. Ed. by M. Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1598–1601. DOI: 10.1007/978-3-642-04898-2_592.

[38] M. Hofert et al. *Elements of Copula Modeling with R*. Use R! Springer International Publishing, 2018, p. 267. DOI: 10.1007/978-3-319-89635-9.

[39] E. W. Frees and E. A. Valdez. "Understanding Relationships Using Copulas". In: *North American Actuarial Journal* 2.1 (1998), pp. 1–25. DOI: 10.1080/10920277.1998.10595667.

[40] P. K. Trivedi and D. M. Zimmer. *Copula modeling: An introduction for practitioners*. Vol. 1. 1. 2005, pp. 1–111. DOI: 10.1561/0800000005.

[41] Y.-K. Tse. "Part I: Loss models". In: *Nonlife Actuarial Models: Theory, Methods and Evaluation*. International Series on Actuarial Science. Cambridge University Press, 2009, pp. 1–112. DOI: 10.1017/CBO9780511812156.003.

[42] T. Nagler et al. *Statistical Inference of Vine Copulas*. 2019.

[43] S. Heilpern. "Multivariate Measures of Dependence based on Copulas". In: *Mathematical Economics* 10.17 (2014), pp. 17–32.

[44] L. G. Díaz Monroy and M. A. Morales Rivera. *Estadística multivariada: inferencia y métodos*. 2012.