

**Rank-based methods for modeling  
dependence between loss triangles**

M.-P. Côté, C. Genest,  
A. Abdallah

G-2015-137

December 2015

---

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2015.

The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2015.



# Rank-based methods for modeling dependence between loss triangles

**Marie-Pier Côté**<sup>a</sup>

**Christian Genest**<sup>a</sup>

**Anas Abdallah**<sup>b</sup>

<sup>a</sup> GERAD & Department of Mathematics and Statistics,  
McGill University, Montréal (Québec) Canada,  
H3A 0B9

<sup>b</sup> École d'actuariat, Université Laval, Québec (Québec)  
Canada, G1V 0A6

marie-pier.cote2@mail.mcgill.ca  
cgenest@math.mcgill.ca  
anas.abdallah.1@ulaval.ca

**December 2015**

**Les Cahiers du GERAD  
G–2015–137**

Copyright © 2015 GERAD

**Abstract:** In order to determine the risk capital for their aggregate portfolio, property and casualty insurance companies must fit a multivariate model to the loss triangle data relating to each of their lines of business. As an inadequate choice of dependence structure may have an undesirable effect on reserve estimation, a two-stage inference strategy is proposed in this paper to assist with model selection and validation. Generalized linear models are first fitted to the margins. Standardized residuals from these models are then linked through a copula selected and validated using rank-based methods. The approach is illustrated with data for six lines of business from a large Canadian insurance company for which two hierarchical dependence models are considered, i.e., a fully nested Archimedean copula structure and a copula-based risk aggregation model.

**Key Words:** Capital allocation, copula GLM, hierarchical modeling, nested Archimedean copulas, parametric bootstrap, rank-based estimation, risk aggregation, run-off triangles.

---

**Acknowledgments:** This work was supported by the Canada Research Chairs Program, scholarships and grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN/39476-2011), the Canadian Statistical Sciences Institute, the Fonds de recherche du Québec – Nature et technologies (2015-PR-183236), Mitacs, and the Cooperators General Insurance Company.

# 1 Introduction

In Canada, the Own Risk and Solvency Assessment (ORSA) guideline from the Office of the Superintendent of Financial Institutions (OSFI) requires that insurance companies set internal targets for risk capital that are tailored to their consolidated operations. In order to relate risk to capital and consider their operations as a whole, insurers are encouraged to develop internal models for the aggregation of dependent risks. Similar regulations exist in many countries worldwide.

To comply with regulatory standards, property and casualty insurance companies have to hold reserves and risk capital relating to losses that are incurred but not yet paid. For each line of business, payments relating to past claims are usually structured in a run-off triangle arranged to rows according to the accident years, and to columns according to the development periods, i.e., the years since the accident occurred. In order to determine a reserve, one must forecast the payments that these ongoing claims will induce in future years, i.e., one must extend each triangle to a rectangle by predicting the missing entries.

Several nonparametric approaches are available for developing claims in a run-off triangle, most notably the chain-ladder method. In order to account for the dependence between triangles, multivariate extensions of this technique have been proposed, e.g., by Braun (2004), Pröhl and Schmidt (2005), Schmidt (2006), Merz and Wüthrich (2008), and Zhang (2010). These techniques account for dependence in the computation of reserves and their prediction errors but they do not provide the predictive distribution needed to obtain risk measures such as Value-at-Risk (VaR) or Tail Value-at-Risk (TVaR). Their use in the determination of risk capital is therefore limited.

Parametric approaches leading to the distribution of unpaid losses have been considered, e.g., by Brehm (2002), Shi and Frees (2011), De Jong (2012), Shi et al. (2012), Merz et al. (2013) and Abdallah et al. (2015). Models investigated in these articles incorporate dependence between lines of business and/or within calendar years of a line of business through Gaussian, Archimedean or Hierarchical Archimedean copulas. In these papers, the total reserve estimate in the presence of dependence is not equal to the sum of the marginal reserves estimated assuming independence. This is a by-product of the joint estimation of the marginal and dependence parameters, which relies heavily on the choice of multivariate model for the run-off triangles. An inadequate choice of dependence structure may then have a large, undesirable effect on the estimation of the reserves. This is particularly alarming given that this choice is typically based on very few data points (e.g., 55 observations for 10 accident years and 10 development periods). Tools are thus needed for assessing the dependence between run-off triangles and selecting an appropriate model.

In this paper, we address this inferential issue within the context of a multivariate extension of the pairwise model of Shi and Frees (2011), where the dependence between corresponding cells of different run-off triangles is described by a copula. We propose to use an alternative two-stage inference strategy, in which generalized linear models (GLMs) are first fitted to the margins, thereby fixing the estimates of the reserves. In the second step, standardized residuals from those models are linked through a dependence structure estimated using rank-based methods. This general approach has a long history in the copula modeling literature; see, e.g., Genest and Favre (2007) or Genest and Nešlehová (2012) for reviews. When dealing with identically distributed data, rank-based methods are well-established tools for selecting, estimating and validating copulas. To our knowledge, however, these techniques have never been applied to run-off triangles.

To illustrate the proposed approach, we consider run-off triangles for six portfolios from a large Canadian property and casualty insurance company. These data are described in Section 2 and appended. In Section 2.1, GLMs with log-normal and Gamma distributions are fitted to the individual portfolios, and the properties of these two parametric families are exploited in Section 2.2 to define residuals that are suitable for a dependence analysis through ranks. Two different hierarchical approaches are then explored for modeling the dependence between the lines of business.

In Section 3, a nested Archimedean copula model is fitted, along the same lines as Abdallah et al. (2015). As this model imposes many constraints on the dependence structure and the choice of copulas, a more flexible approach considered by Arbenz et al. (2012) and Côté and Genest (2015) is implemented in Section 4. Risk capital calculations and allocations for the two models are compared in Section 5, and

Section 6 summarizes the pros and cons of the two approaches. Appendix A contains density calculations for the nested Archimedean copula model, and the data (up to a multiplicative factor for confidentiality purposes) are provided in Appendix B, along with parameter estimates of the marginal GLM models.

## 2 Data

The run-off triangle data considered in this paper are from a large Canadian property and casualty insurance company. They consist of the cumulative paid losses and net earned premiums for six lines of automobile and home insurance business. Tables 12–17 in Appendix B shows the paid losses for accident years 2003–12 inclusively for each of the six lines of business developed over at most ten years. To preserve confidentiality, all figures were multiplied by a constant  $c > 0$ . However, this is inconsequential because in order to account for the volume of business, the analysis focuses on the paid loss ratios, i.e., the payments divided by the net earned premiums.

Table 1 gives a descriptive summary of each line of business (LOB). There are five run-off triangles of personal and commercial auto lines with Accident Benefits and Bodily Injury coverages from three regions (Atlantic, Ontario and the West). Atlantic Canada consists of New Brunswick, Nova Scotia, Prince Edward Island and Newfoundland/Labrador; the West comprises Manitoba, Saskatchewan, Alberta, British Columbia, Northwest Territories, Yukon, and Nunavut. Given that Québec has a public plan for this section of auto insurance, business for that province is included only in the sixth triangle, which comprises the company’s country-wide Liability personal and commercial home insurance.

Table 1: Descriptive summary of six lines of business for a Canadian insurance company.

LOB	Region	Product	Coverage
1	Atlantic	Auto	Bodily Injury
2	Ontario	Auto	Bodily Injury
3	West	Auto	Bodily Injury
4	Ontario	Auto	Accident Benefits excluding Disability Income
5	Ontario	Auto	Accident Benefits: Disability Income only
6	Country-wide	Home	Liability

Bodily Injury (BI) coverage provides compensation to the insured if the latter is injured or killed through the fault of a motorist who has no insurance, or by an unidentified vehicle. The Accident Benefits (AB) coverage provides compensation, regardless of fault, if a driver, passenger, or pedestrian suffers injury or death in an automobile collision. Disability income is an insurance product that provides supplementary income when the accident results in a disability that prevents the insured from working at his/her regular employment. For this reason, AB disability income is considered separately from other AB. Finally, Liability insurance covers an insured for his/her legal liability for injuries or damage to others.

### 2.1 Marginal GLMs for incremental loss ratios

For LOB  $\ell \in \{1, \dots, 6\}$ , denote by  $Y_{ij}^{(\ell)}$  the incremental payment for the  $i$ th accident year and the  $j$ th development period, where  $i, j \in \{1, \dots, 10\}$ . Given that the earned premiums  $p_i^{(\ell)}$  vary with accident year  $i$  and line of business  $\ell$ , it is convenient to model the loss ratios, defined by

$$X_{ij}^{(\ell)} = Y_{ij}^{(\ell)} / p_i^{(\ell)}.$$

In Figure 1, loss ratios  $X_{ij}^{(\ell)}$  for  $i = 1, 2$ ,  $j = 1, \dots, 11 - i$  and  $\ell = 1, \dots, 6$  are shown. It is clear from the graph that the loss ratio depends on the development lag for every portfolio. By comparing the solid and dashed lines of the same color, one can also see that the accident year has an impact. In order to capture these patterns, we consider a regression model with two explanatory variables, i.e., accident year and development period. This is in line with the classical chain-ladder approach.

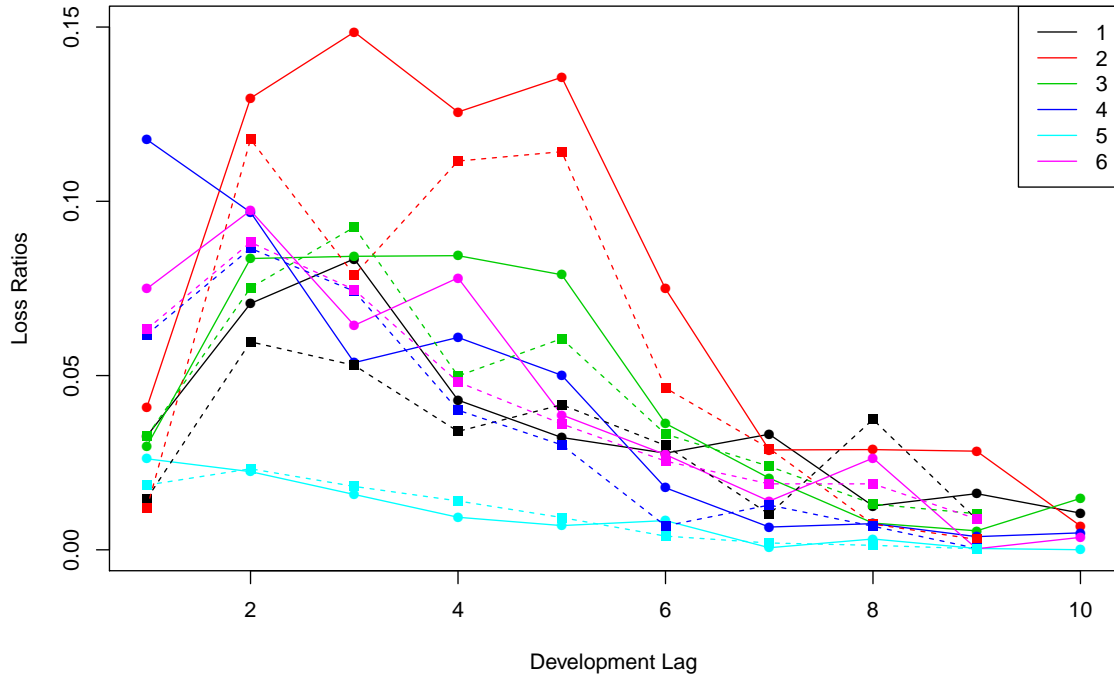


Figure 1: Loss ratios for years 2003 (solid line) and 2004 (dashed line) in function of the development lag for the six lines of business.

For LOB  $\ell \in \{1, \dots, 6\}$ , let  $\kappa_i^{(\ell)}$  be the effect of accident year  $i \in \{1, \dots, 10\}$  and  $\lambda_j^{(\ell)}$  be the effect of development period  $j \in \{1, \dots, 10\}$ . The systematic component for the  $\ell$ th line of business can then be written as

$$\eta_{ij}^{(\ell)} = \zeta^{(\ell)} + \kappa_i^{(\ell)} + \lambda_j^{(\ell)},$$

where  $\zeta^{(\ell)}$  is the intercept, and for parameter identification, we set  $\kappa_1^{(\ell)} = \lambda_1^{(\ell)} = 0$ . There is no interaction term in this model, i.e., it is assumed that the effect of a given development period does not vary by accident year. While this assumption is hard to check, it is required to ensure that all parameters can be estimated from the 55 observations available.

In their analysis of dependent loss triangles using copulas, Shi and Frees (2011) use the log-normal and Gamma distributions for incremental claims. Their justification applies here as well. Following these authors, we consider the link

$$\mu_{ij}^{(\ell)} = \eta_{ij}^{(\ell)}$$

for a log-normal distribution with mean  $\mu_{ij}^{(\ell)}$  and standard deviation  $\sigma^{(\ell)}$  on the log scale. For the Gamma distribution, however, we use the exponential link instead of the canonical inverse link in order to enforce positive means. When the Gamma distribution is selected, therefore, its scale and shape parameters are respectively denoted by  $\beta_{ij}^{(\ell)}$  and  $\alpha^{(\ell)}$ , and it is assumed that

$$\beta_{ij}^{(\ell)} = \exp(\eta_{ij}^{(\ell)})/\alpha^{(\ell)}.$$

Log-normal and Gamma distributions were fitted to all lines of business by the method of maximum likelihood. Table 2 shows the corresponding values of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). These criteria suggest the choice of the log-normal distribution for the first line of business and the Gamma distribution for all others. These choices of models are confirmed by the Kolmogorov–Smirnov goodness-of-fit test, whose  $p$ -values are also given in Table 2. No model is rejected at the 1% level. Q-Q plots (not shown) of standardized residuals (defined below) provide visual confirmation that the selected models are adequate, although the fit for LOB 6 is borderline.

Table 2: Fit statistics and goodness-of-fit test of marginals.

LOB	AIC		BIC		<i>p</i> -value of the Kolmogorov–Smirnov test
	Log-normal	Gamma	Log-normal	Gamma	
1	−294	−291	−254	−251	0.886
2	−266	−270	−226	−230	0.643
3	−323	−324	−283	−283	0.397
4	−272	−276	−232	−236	0.135
5	−441	−444	−401	−404	0.478
6	−259	−267	−219	−226	0.019

Parameter estimates of the fitted models are given in Appendix B along with their standard errors. Using these values, one can estimate the total reserve of the portfolio by

$$\sum_{\ell=1}^6 \sum_{i=2}^{10} \sum_{j=10-i+2}^{10} p_i^{(\ell)} \mathbb{E}(X_{ij}^{(\ell)}),$$

where  $\mathbb{E}(X_{ij}^{(\ell)})$  is the projected unpaid loss ratio, and  $p_i^{(\ell)}$  is the premiums earned in the corresponding accident year  $i$ . For  $\ell = 1$ , we have

$$\mathbb{E}(X_{ij}^{(1)}) = \exp\{\widehat{\mu}_{ij}^{(1)} + (\widehat{\sigma}^{(1)})^2/2\},$$

while for  $\ell > 1$ ,  $\mathbb{E}(X_{ij}^{(\ell)}) = \widehat{\beta}_{ij}^{(\ell)} \widehat{\alpha}^{(\ell)}$ . The estimated reserves of the six lines of business are given at the bottom of Table 18 in Appendix B, along with those derived from the chain-ladder method, which is the industry's benchmark. The two methods lead to similar results and total reserve estimates of \$438,088 and \$453,686, respectively.

## 2.2 Exploratory dependence analysis

One would expect intuitively that the AB, BI and Liability claim payments are associated, as these coverages all involve compensation for injuries or damage to the insured or to others. One may also wonder whether there exist interactions between portfolios across regions. In order to account for such dependencies between  $d \geq 2$  triangles, Shi and Frees (2011) propose to link the marginal GLMs through a copula. This approach involves expressing the joint distribution of the loss ratios in the form

$$\Pr(X_{ij}^{(1)} \leq x_{ij}^{(1)}, \dots, X_{ij}^{(d)} \leq x_{ij}^{(d)}) = C\{\Pr(X_{ij}^{(1)} \leq x_{ij}^{(1)}), \dots, \Pr(X_{ij}^{(d)} \leq x_{ij}^{(d)})\},$$

where  $C$  is a  $d$ -variate cumulative distribution function with uniform margins on  $(0, 1)$ .

In order to select a copula  $C$  that appropriately reflects the dependence in the data, it is best to rely on rank-based techniques as they allow to separate the effect of the marginals from the dependence structure (Genest and Favre, 2007; Genest and Nešlehová, 2012). To illustrate this point, consider first the graph displayed in the left panel of Figure 2, which shows a scatter plot of the pairs  $(X_{ij}^{(3)}, X_{ij}^{(6)})$  with  $i, j \in \{1, \dots, 10\}$  and  $j \leq i$ . This graph suggests a strong, positive dependence between BI in Western Canada and country-wide Liability; in particular, the Pearson correlation is 0.56. However, the pattern of points on this graph is induced by the systematic effects of the development lags and accident years. For example, the seven points in the lower left corner of the graph all correspond to development years 7–10. As these effects are already accounted for by the marginal GLMs, this graph is uninformative (not to say misleading) for the selection of  $C$ .

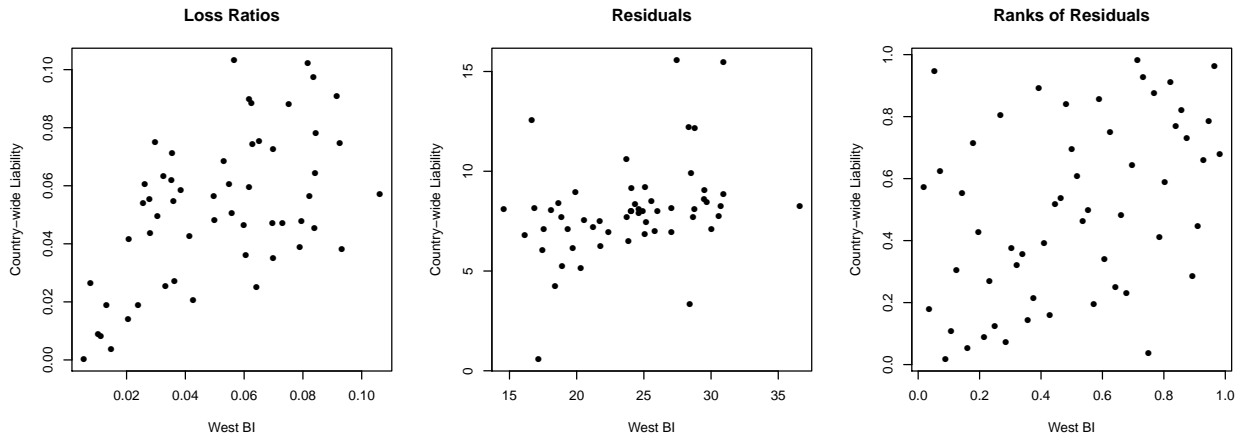


Figure 2: Loss ratios (left), residuals (middle) and standardized ranks of the latter (right) for LOBs 3 and 6.

To get insight into the dependence structure, it is more relevant to consider the residuals from the GLM models. For LOB 1, (standardized) residuals of the log-normal regression model can be defined, for all  $i, j \in \{1, \dots, 10\}$  and  $j \leq i$ , as

$$\epsilon_{ij}^{(1)} = \{\ln(X_{ij}^{(1)}) - \hat{\mu}_{ij}^{(1)}\} / \hat{\sigma}^{(1)},$$

while for LOB  $\ell \in \{2, \dots, 6\}$ , the fact that Gamma regression models were used leads to set

$$\epsilon_{ij}^{(\ell)} = X_{ij}^{(\ell)} / \hat{\beta}_{ij}^{(\ell)}.$$

In this fashion, the vectors  $(\epsilon_{ij}^{(1)}, \dots, \epsilon_{ij}^{(6)})$  with  $i, j \in \{1, \dots, 10\}$  and  $j \leq i$  form a pseudo-random sample from a distribution with copula  $C$  and margins approximately  $\mathcal{N}(0, 1)$  for  $\ell = 1$  and  $\mathcal{G}(\hat{\alpha}^\ell, 1)$ , for  $\ell = 2, \dots, 6$ . The middle panel of Figure 2 shows a scatter plot of the pairs  $(\epsilon_{ij}^{(3)}, \epsilon_{ij}^{(6)})$ . This graph suggests a form of positive dependence (Pearson's correlation is 0.34), but the message is blurred by the effect of the Gamma marginals. As the goal is to select the copula  $C$ , which does not depend on the margins, it is preferable to plot the pairs of normalized ranks, as in the right panel of Figure 2. For arbitrary  $i, j \in \{1, \dots, 10\}$  and  $j \leq i$ , the standardized rank of residual  $\epsilon_{ij}^{(\ell)}$  is defined by

$$R_{ij}^{(\ell)} = \frac{1}{55 + 1} \sum_{i^*=1}^{10} \sum_{j^*=1}^{11-i^*} \mathbf{1}(\epsilon_{i^*j^*}^{(\ell)} \leq \epsilon_{ij}^{(\ell)}),$$

where, in general,  $\mathbf{1}(A)$  is the indicator function of the set  $A$  and the division by 56 rather than 55 is to ensure that all standardized ranks are strictly comprised between 0 and 1.

Let  $C_n$  be the empirical distribution function of the vectors  $(R_{ij}^{(1)}, \dots, R_{ij}^{(d)})$ , with  $i, j \in \{1, \dots, 10\}$  and  $j \leq i$ . It can be shown, under suitable conditions on the underlying copula  $C$ , that  $C_n$  is a consistent estimator thereof. Accordingly, the vectors of standardized ranks, which form the support of  $C_n$ , are a reliable tool for copula selection, fitting and validation. For example, the right panel of Figure 2 shows the pairs of standardized ranks associated with the residuals from the West BI and the country-wide Liability coverages. One can see from this graph that there is a residual dependence between these two portfolios. In particular, the correlation between these pairs is 0.40; this rank-based correlation is a consistent estimate of Spearman's  $\rho$ . Alternative copula-based measures of association between two variables are Kendall's  $\tau$  and the Van der Waerden's coefficient  $\Upsilon$ . As  $\rho = \tau = \Upsilon = 0$  when the variables are independent, one can test the hypothesis  $\mathcal{H}_0 : C_{k\ell}(u, v) \equiv uv$  using the rank-based estimates of these coefficients; see, e.g., Hollander et al. (2014). For the pair  $(\epsilon^{(3)}, \epsilon^{(6)})$ , these estimates and the  $p$ -values of the corresponding tests are shown in Table 3; the null hypothesis of independence can be rejected at the 1% level in all cases.

Given the presence of dependence, the challenge is then to select a copula that best reflects the association between the variables. Many parametric families of copulas are available; see, e.g., McNeil et al. (2015) or

Table 3: Nonparametric tests of independence.

Kendall's test		Spearman's test		Van der Waerden test	
$\hat{\tau}$	$p$ -value	$\hat{\rho}$	$p$ -value	$\hat{Y}$	$p$ -value
0.29	0.0021	0.40	0.0023	18.27	0.0055

Nelsen (2006) for the definition and properties of the Clayton, Frank, Plackett and  $t$  copula families used subsequently. Given a class  $\mathcal{C} = \{C_\theta : \theta \in \Theta\}$  of  $d$ -dimensional copulas, a rank-based estimate of the dependence parameter  $\theta$  can be obtained from loss-triangle data by maximizing the pseudo log-likelihood

$$\mathcal{L}(\theta) = \sum_{i=1}^{10} \sum_{j=1}^{11-i} \ln\{c_\theta(R_{ij}^{(1)}, \dots, R_{ij}^{(d)})\},$$

where  $c_\theta$  is the density of  $C_\theta$ . The consistency and asymptotic normality of the resulting estimator  $\hat{\theta}$  was established by Genest et al. (1995) under broad regularity conditions. The adequacy of the class  $\mathcal{C}$  can then be tested using the Cramér-von-Mises statistic defined by

$$S_n = \int_{[0,1]^d} \{C_n(u_1, \dots, u_d) - C_{\hat{\theta}}(u_1, \dots, u_d)\}^2 du_1 \cdots du_d.$$

The  $p$ -value of a test of the hypothesis  $\mathcal{H}_0 : C \in \mathcal{C}$  based on the statistic  $S_n$  can be computed via a parametric bootstrap procedure described in Genest et al. (2009). Both the estimation and the goodness-of-fit procedures are available in the R package `copula`. For illustration, Table 4 shows the parameter estimates, standard deviation and the  $p$ -value of the goodness-of-fit test for four copula families fitted to the pair of residuals  $(\epsilon^{(3)}, \epsilon^{(6)})$  from the West BI and country-wide Liability triangles. This suggests that the Clayton copula would be a poor choice for these data; given the small sample size, however, it does not seem possible to discriminate between the other three copula families on the basis of  $S_n$ .

Table 4: Parameter estimates and goodness-of-fit test  $p$ -value.

Copula	Parameter	Standard Deviation	$p$ -value
Clayton	0.584	0.194	0.0804
Frank	2.804	0.836	0.7557
Plackett	3.777	1.426	0.7747
$t(2)$	0.375	0.155	0.2323

This model selection, fitting and validation procedure is standard and straightforward to implement in two dimensions. However, the canonical  $d$ -variate generalizations of bivariate copulas typically lack flexibility: either they are exchangeable and/or their lower-dimensional margins are all from the same type. With six lines of business, these assumptions may be too restrictive. As one can see in Figure 3, different pairs of residuals have different types of association; in particular, there is positive dependence within Ontario LOBs, and the BI coverages for Ontario and the West are negatively associated. This is confirmed by the empirical values of Kendall's  $\tau$ , which are displayed in Table 5 for all pairs of variables. In this table, values shown in bold are significantly different from 0 at the 5% level.

The fact that many variables exhibit positive dependence is due in part to exogenous common factors such as inflation and interest rates. Furthermore, strategic decisions can impact several portfolios, e.g., the acceleration of payments on all lines of the liability insurance sector could induce some dependence between West BI and country-wide Liability. At a more basic level, the positive association between Ontario AB and BI can be explained by the fact that the same accident will often arise in both coverages. Finally, jurisprudence can play a role. For example, reforms were engaged in the Atlantic region to control BI costs; this may explain why LOB 1 is seemingly independent from all other lines of business.

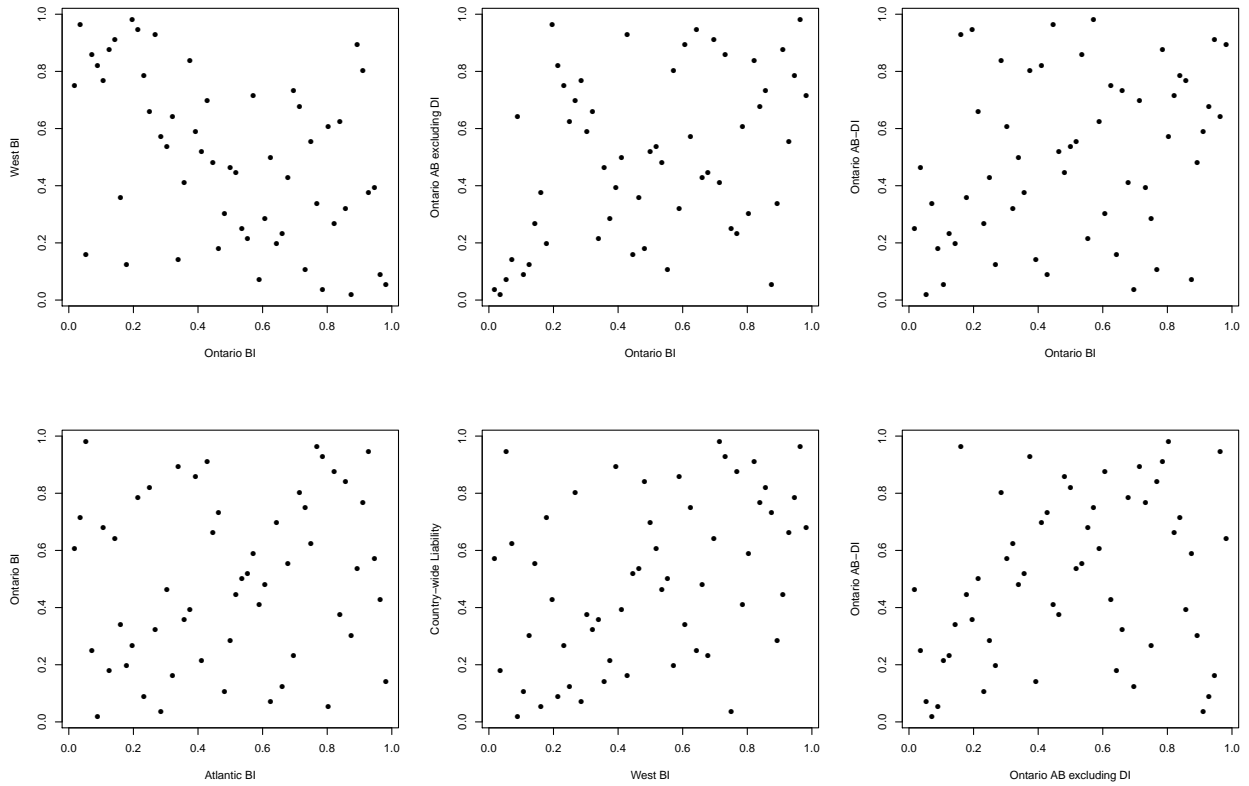


Figure 3: Scatter plot of residuals between different LOBs.

Table 5: Empirical values of Kendall’s  $\tau$  for all pairs in the portfolio.

	$\epsilon^{(1)}$	$\epsilon^{(2)}$	$\epsilon^{(3)}$	$\epsilon^{(4)}$	$\epsilon^{(5)}$	$\epsilon^{(6)}$
$\epsilon^{(1)}$	1.000	0.115	0.024	-0.061	0.014	0.076
$\epsilon^{(2)}$	0.115	1.000	<b>-0.331</b>	<b>0.244</b>	<b>0.209</b>	-0.090
$\epsilon^{(3)}$	0.024	<b>-0.331</b>	1.000	0.040	-0.079	<b>0.285</b>
$\epsilon^{(4)}$	-0.061	<b>0.244</b>	0.040	1.000	<b>0.200</b>	0.030
$\epsilon^{(5)}$	0.014	<b>0.209</b>	-0.079	<b>0.200</b>	1.000	0.046
$\epsilon^{(6)}$	0.076	-0.090	<b>0.285</b>	0.030	0.046	1.000

### 3 Nested Archimedean copula model

Nesting Archimedean copulas is a popular way to construct non-exchangeable multivariate dependence models. This approach, originally proposed by Joe (1997), was further investigated, e.g., by Embrechts et al. (2003), Whelan (2004) and Savu and Tiede (2010). In the reserving literature, Abdallah et al. (2015) exploited nested Archimedean copulas to model the dependence between two run-off triangles. In what follows, this approach is extended to higher dimensions using a specific structure called fully nested Archimedean copulas.

Following Genest and MacKay (1986) or Nelsen (2006), a bivariate copula is said to be Archimedean with generator  $\varphi_1 : [0, 1] \rightarrow [0, \infty]$  if it can be expressed, for all  $(u_1, u_2) \in (0, 1)^2$ , in the form

$$C_1(u_1, u_2) = \varphi_1^{-1}\{\varphi_1(u_1) + \varphi_1(u_2)\},$$

where  $\varphi_1$  is convex, decreasing and such that  $\varphi_1(1) = 0$ . More generally, a  $(d + 1)$ -variate copula  $C_d$  is said to be a fully nested Archimedean copula with generators  $\varphi_1, \dots, \varphi_d$  if it is defined recursively for all

$(u_1, \dots, u_{d+1}) \in (0, 1)^{d+1}$ , by

$$\begin{aligned} C_2(u_1, u_2, u_3) &= \varphi_2^{-1}[\varphi_2(u_3) + \varphi_2\{C_1(u_1, u_2)\}], \\ \vdots &= \vdots \\ C_d(u_1, \dots, u_{d+1}) &= \varphi_d^{-1}[\varphi_d(u_{d+1}) + \varphi_d\{C_{d-1}(u_1, \dots, u_d)\}]. \end{aligned}$$

As shown by McNeil (2008), sufficient compatibility conditions for  $C_d$  to be a copula are:

- (i)  $\varphi_1^{-1}, \dots, \varphi_d^{-1}$  are completely monotone, i.e., Laplace transforms;
- (ii)  $\varphi_{k+1} \circ \varphi_k^{-1}$  has completely monotone derivatives for all  $k \in \{1, \dots, d-1\}$ .

This model is such that if  $(U_1, \dots, U_{d+1})$  is distributed as  $C_d$ , the copula linking variables  $U_j$  and  $U_k$  is Archimedean with generator  $\varphi_{k-1}$  for all  $j < k$ . In addition, because of condition (ii), one must have

$$\tau(U_k, U_\ell) \leq \tau(U_i, U_j), \quad i < j < \ell, \quad k < \ell. \quad (1)$$

Algorithms for generating data from  $C_d$  were given by McNeil (2008) and Hofert (2011). Hofert and Mächler (2011) also wrote the R package `nacopula` (now merged into `copula`) that can be used to simulate from fully nested Archimedean copulas in any dimension.

Figure 4 depicts the fully nested Archimedean structure used to model the dependence between the residuals of the six lines of business. In this structure, copula  $C_1$  links the two components of the Ontario AB coverage. Their dependence with Ontario BI coverage is then incorporated at level 2. The West BI and the country-wide Liability coverages are then included at levels 3 and 4, respectively. Anti-ranks (i.e., the ranks of the negative residuals) had to be used at levels 3 and 4, because of the constraints imposed by (1) and the fact that the residuals for LOB 3 are negatively associated with LOB 2 and positively associated with LOB 6. Finally, the Atlantic BI coverage was included at the last step given its apparent lack of dependence with the other lines of business. This overall structure is in accordance with ratemaking practices, as the rating is typically performed on a territorial basis. One may thus expect the dependence between lines of business to be larger when they are from the same region than when they are not.

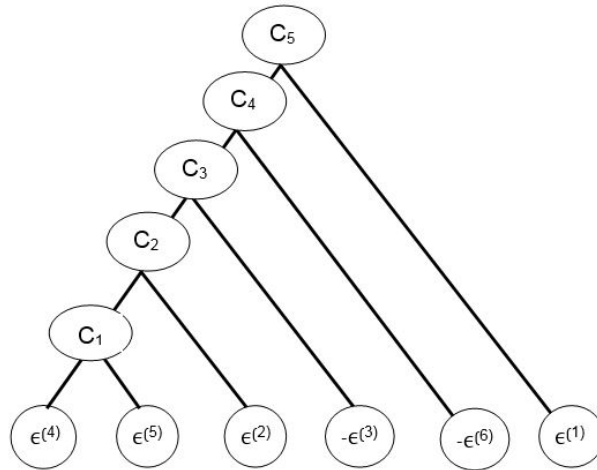


Figure 4: Tree structure for the fully nested Archimedean copula model.

In what follows, it is assumed that for each  $k \in \{1, \dots, 5\}$  and all  $t \in (0, 1)$ ,

$$\varphi_k(t) = -\ln \left( \frac{e^{-t\theta_k} - 1}{e^{-\theta_k} - 1} \right)$$

for some  $\theta_k \in \mathbb{R}$ . In other words, the nested copulas are taken to be from the Frank family, which spans all degrees of dependence between  $-1$  and  $1$ , as measured by Kendall's tau. A rank-based estimate  $\hat{\theta}$  of

the vector  $\theta = (\theta_1, \dots, \theta_5)$  characterizing the dependence structure is then obtained by maximizing the pseudo-likelihood function

$$\mathcal{L}(\theta) = \sum_{i=1}^{10} \sum_{j=1}^{11-i} \ln\{c(R_{ij}^{(4)}, R_{ij}^{(5)}, R_{ij}^{(2)}, 1 - R_{ij}^{(3)}, 1 - R_{ij}^{(6)}, R_{ij}^{(1)}; \theta)\},$$

where  $c$  is the density of the fully nested Archimedean copula, which is obtained by differentiating the copula. As shown in Appendix A, the evaluation of this density is straightforward but computationally intensive in high dimensions. Therefore, due to evidence that residuals for LOB 1 are independent from residuals for other LOBs, the last level in the hierarchy is independence and  $\theta_5$  is set to be 0.

The maximization of the pseudo-likelihood for the model with four levels leads to the parameter estimate  $\hat{\theta} = (2.693, 2.354, 1.7817, 0.867)$ . However, a 95% confidence interval for  $\theta_4$  based on 1000 bootstrap replicates is  $[0, 1.92)$ . As a value of  $\theta = 0$  corresponds to independence in the Frank copula family, the dependence is significant only in the first three levels of the hierarchy. The parameters of the reduced model with  $\theta_4 = \theta_5 = 0$  were estimated once again by the maximum pseudo-likelihood method. This led to  $\hat{\theta} = (2.577, 2.233, 1.776)$ , whose components are all significantly different from 0.

Figure 5 shows the approximate distribution of  $\hat{\theta}_3$  (left),  $\hat{\theta}_2$  (middle), and  $\hat{\theta}_1$  (right) based on 10,000 bootstrap replicates. In that figure, the dashed blue lines represent 95% confidence intervals for the parameters, none of which includes 0. There are hints in the figure that the distribution of the estimators (especially  $\hat{\theta}_1$ ) may not be Normal. This is likely due to the constraint  $\theta_3 \leq \theta_2 \leq \theta_1$ . In the bottom row of Figure 5, one can observe that parameters on the boundary of their domain are relatively frequent:  $\hat{\theta}_1 = \hat{\theta}_2$  in 14.3% of the replicates,  $\hat{\theta}_3 = \hat{\theta}_2$  in 9.9% of the replicates, and  $\hat{\theta}_1 = \hat{\theta}_2 = \hat{\theta}_3$  in 4.8% of the replicates.

To check for model adequacy, a random sample of size 500 from the fitted model was generated. A test of the hypothesis that the underlying copula of this sample is the same as that of the original data was then carried out using the rank-based procedure of Rémillard and Scaillet (2009). The test statistic was

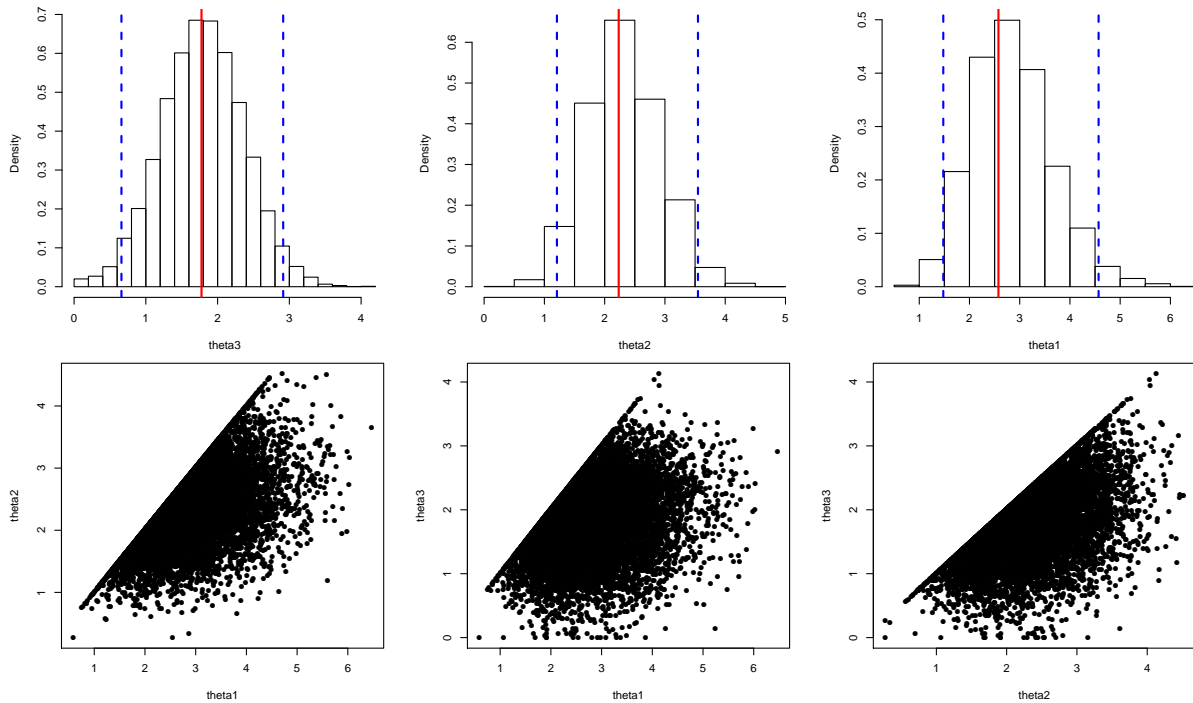


Figure 5: Fully nested Archimedean copula model: histograms of bootstrap parameters with 95% confidence interval (top row) and scatter plots of bootstrap replications (bottom row).

computed with the R package `TwoCop` and led to an approximate  $p$ -value of 31%, suggesting that the fit is not inadequate.

As an additional informal check, random samples of size 55 were drawn from the fitted 6-dimensional copula and compared visually to the empirical copula by looking at rank plots of selected pairs. Figure 6 shows one result from such a comparison of pairs (LOB 2, LOB  $\ell$ ) with  $\ell \in \{3, 4, 5\}$  and (LOB 3, LOB 4). The rank plots derived from the residuals are in the top row, and those corresponding to the random sample are in the bottom row. The positive dependence between Ontario risks seems to be accurately captured by the model. Although the negative association between LOB 2 and 3 is taken into account, one can see in the second column of Figure 6 that negative dependence is induced between LOB 3 and 4. This is an artifact of the dependence structure, which assumes from the start that the pairs  $(-3, \ell)$ , with  $\ell \in \{2, 4, 5\}$  have the same degree of association. Table 5 suggests that this is not the case. This issue could have been avoided by grouping LOB 2 and 3 earlier in the structure, but at the expense of the overall fit of the model. A more flexible modeling approach is presented below.

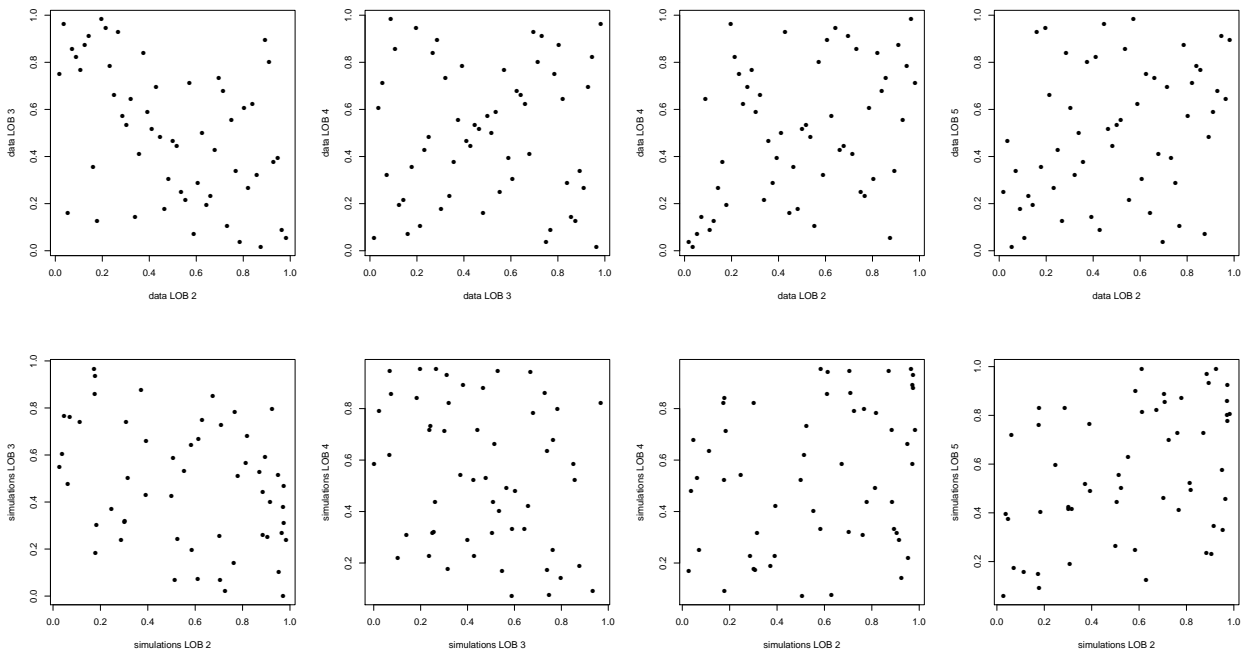


Figure 6: Fully nested Archimedean copula model adequacy checking: ranks of pairs of residuals (top row) and pairs of simulations from the model (bottom row).

## 4 Copula-based risk aggregation model

In this section, a hierarchical approach to loss triangle modeling is considered. It appears to have been originally proposed by Swiss reinsurance practitioners (Bürgi et al., 2008; SCOR, 2008) but was formalized by Arbenz et al. (2012). Estimation and validation procedures for this class of models are described in Côté (2014) and Côté and Genest (2015), where rank-based clustering techniques are also proposed for selecting an appropriate structure.

The model is defined using a tree comprising  $d - 1$  nodes, each of which has two branches. An example of such a structure is shown in the left panel of Figure 7. At each node, a copula describes the dependence between the two components which are then summed and viewed as a single risk in higher levels of the hierarchy. For example,  $C_{4,5}$  denotes the copula linking  $\epsilon^{(4)}$  and  $\epsilon^{(5)}$  and  $S_{4,5} = \epsilon^{(4)} + \epsilon^{(5)}$ , while  $C_{2,\dots,6}$  is the copula linking aggregated risks  $S_{2,3,6}$  and  $S_{4,5}$ .

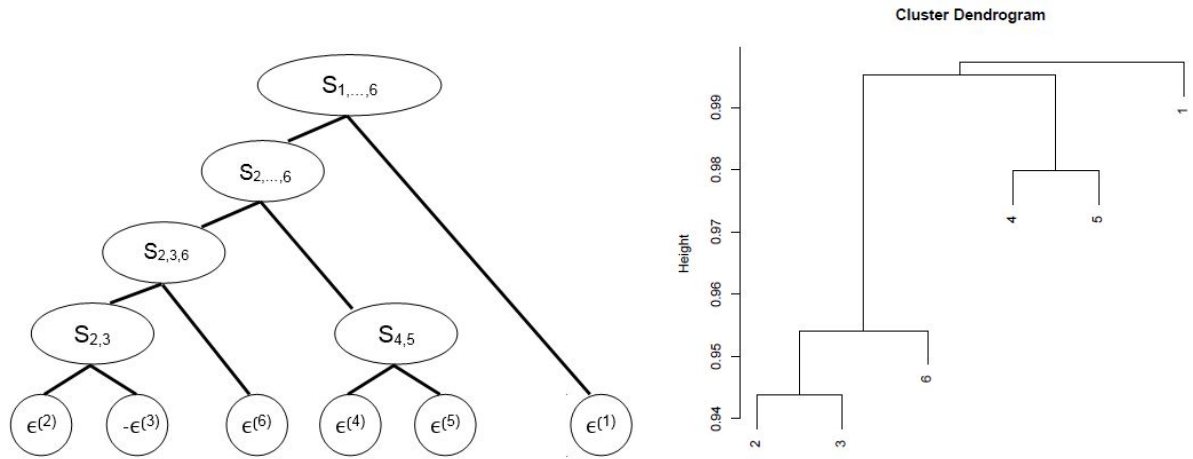


Figure 7: Illustration of the tree structure and dendrogram for the copula-based aggregation model.

A joint distribution for the  $d$  variables is then defined in terms of  $d - 1$  bivariate copulas and  $d$  marginal distributions under a conditional independence assumption. This assumption, which is reasonable in the present context, states that conditional on a sum at a given node, the descendants of that node are independent of the non-descendants. For additional details, see Arbenz et al. (2012) and Côté and Genest (2015).

This modeling strategy is simple to implement, as it builds on tools already available for bivariate copula selection, inference and validation. Furthermore, each of the  $d - 1$  copulas in the model can be chosen freely, thereby providing great flexibility in the dependence structure. Moreover, hierarchical clustering techniques can be adapted to obtain an appropriate tree structure.

As explained in Côté and Genest (2015), it is appealing to model first the risks that are the most dependent in some sense. In this paper, the distance based on Kendall’s tau,

$$\Delta(\epsilon^{(\ell)}, \epsilon^{(k)}) = \sqrt{1 - \tau^2(\epsilon^{(\ell)}, \epsilon^{(k)})},$$

is maximized at each step to obtain the dendrogram displayed in the right panel of Figure 7. Risks 2 and 3 are grouped in the first step. Given that they are negatively associated, it was deemed preferable to work with  $-\epsilon^{(3)}$  as was done in the previous section.

Before selecting appropriate copulas for each aggregation step, Kendall and van der Waerden tests of independence were performed to see if the dependence is significant. The resulting  $p$ -values are shown in Table 6, where one can see that independence is rejected for the first four aggregation steps, but not at the last one. This is not surprising as the preliminary analysis of the data already suggested that the Atlantic BI line of business is not related to the others. Unlike the nested Archimedean copula model, the risk aggregation model captures the existing dependence between West BI and country-wide Liability lines, and includes the latter in the dependence analysis.

Table 6: Results of tests of independence at each aggregation step.

Variables		$\tau$	$p$ -value	
			Van den Waerden test	Kendall test
$\epsilon^{(2)}$	$-\epsilon^{(3)}$	0.331	0.0004	0.0004
$S_{2,3}$	$\epsilon^{(6)}$	0.300	0.0020	0.0012
$\epsilon^{(4)}$	$\epsilon^{(5)}$	0.200	0.0541	0.0311
$S_{2,3,6}$	$S_{4,5}$	0.098	0.0406	0.2925
$S_{2,\dots,6}$	$\epsilon^{(1)}$	0.075	0.3401	0.4204

Given that the independence hypothesis cannot be rejected at the last node, there are only four copulas to be fitted, namely  $C_{2,3}$ ,  $C_{2,3,6}$ ,  $C_{4,5}$  and  $C_{2,\dots,6}$ . Based on rank plots, tests of extremeness from Ben Ghorbal et al. (2009) and goodness-of-fit tests based on the Cramér-von-Mises distance  $S_n$ , parametric families of bivariate copulas were selected and fitted by maximum pseudo-likelihood. The final choices are summarized in Table 7.

Table 7: Copula family and parameter estimates.

Step	Copula	Parameter	SD	Kendall's $\tau$	$p$ -value GoF test
$C_{2,3}$	Plackett	5.349	2.021	0.36	0.523
$C_{2,3,6}$	Frank	2.864	0.986	0.29	0.714
$C_{4,5}$	Clayton	0.548	0.215	0.22	0.147
$C_{2,\dots,6}$	$t_2$	0.162	0.180	0.10	0.358

The model validation technique described in Côté and Genest (2015) was used. It relies on a simulation algorithm proposed by Arbenz et al. (2012) and validated by Mainik (2015). Based on a random sample of size 500 from the model, the test of Rémillard and Scaillet (2009) led to an approximate  $p$ -value of 52%. Therefore, the null hypothesis that both samples are coming from the same copula cannot be rejected. This suggests that the selected hierarchical model is appropriate, and that the conditional independence assumption is reasonable. A visual check of the latter assumption confirms this finding.

Looking at Figure 8, one can see that the pitfalls of the nested Archimedean copula model have been addressed: there is no negative dependence between LOBs 3 and 4, and the model induces positive dependence between LOBs 3 and 6. However, the extent of the association between Ontario AB and BI risks is not represented as fully in the aggregation model as it was in the nested Archimedean copula model. Over all, the risk aggregation model provides a faithful description of the data.

Note that if desired, a modification of the tree structure would make it possible to account for the dependence between LOB 2 and the pair (LOB 4, LOB 5). In that case, however, the negative dependence between LOBs 2 and 3 would be masked.

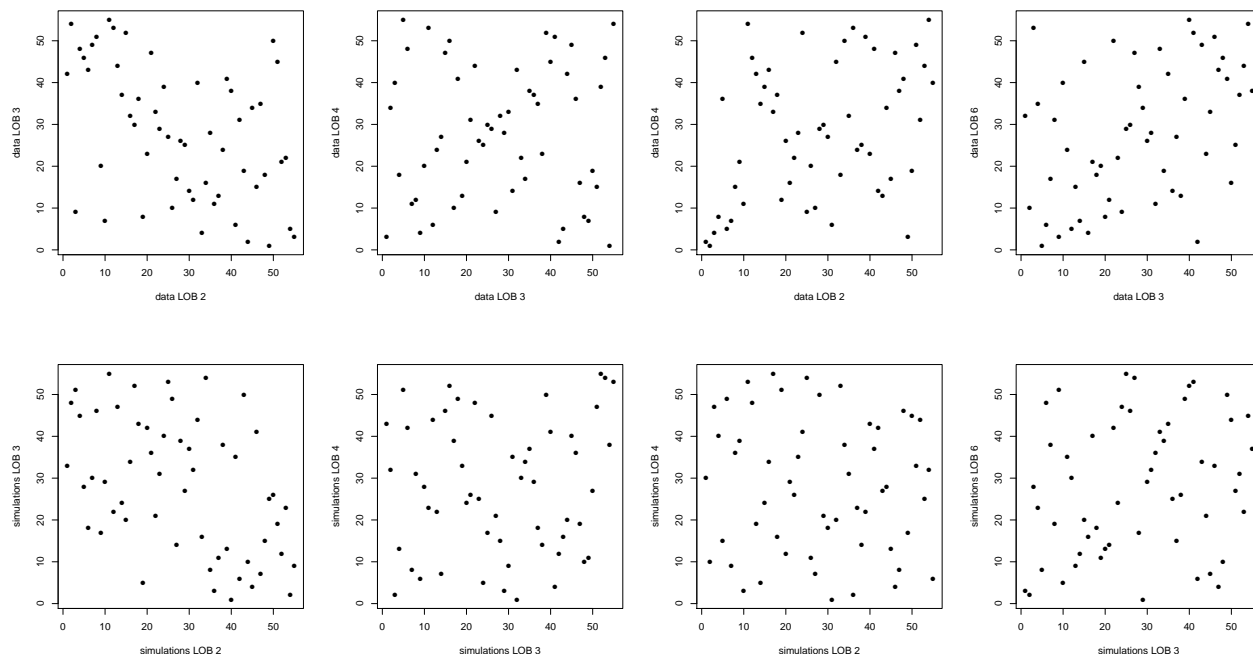


Figure 8: Copula-based risk aggregation model adequacy checking: ranks of pairs of residuals (top row) and pairs of simulations from the model (bottom row).

## 5 Predictive distribution and risk capital

The goal of loss triangle modeling is to forecast the unpaid loss by completing the triangle into a rectangle. Insurance companies are interested in the expected unpaid loss — the reserve — but also in its standard deviation, and other risk measures defined in terms of a risk tolerance  $\kappa \in (0, 1)$  such as the Value-at-Risk (VaR) and the Tail Value-at-Risk (TVaR). In principle, these various measures could all be computed for the nested Archimedean copula model (Model I) and the risk aggregation model (Model II), given that they both specify a distribution for the total unpaid claims. As these distributions cannot be obtained explicitly through a convolution, however, all risk measures must be estimated by simulation.

To obtain one realization of the total unpaid loss, one can proceed as follows:

1. Simulate 45 observations from the dependence model.
2. Transform these observations into loss ratios  $X_{ij}^{(\ell)}$  for LOB  $\ell = 1, \dots, 6$ , development years  $j = 2, \dots, 10$  and accident years  $i = 12 - j, \dots, 10$  by using appropriate inverse probability transforms.
3. For LOB  $\ell = 1, \dots, 6$ , compute the simulated unpaid loss

$$X^{(\ell)} = \sum_{i=2}^{10} \sum_{j=12-i}^{10} p_i^{(\ell)} X_{ij}^{(\ell)}$$

as well as the total unpaid loss  $S = X^{(1)} + \dots + X^{(6)}$ .

Consistent estimates of the risk measures can be derived easily from  $n$  independent copies of the unpaid loss  $S_1, \dots, S_n$ . Let  $F_n$  be the corresponding empirical distribution function. Then

$$\widehat{\text{VaR}}_{\kappa}(S) = \inf\{s \in \mathbb{R} | F_n(s) \geq \kappa\} = s_{\kappa}$$

and

$$\widehat{\text{TVaR}}_{\kappa}(S) = \frac{1}{1 - \kappa} \left[ \frac{1}{n} \sum_{j=1}^n S_j \mathbf{1}(S_j > s_{\kappa}) + s_{\kappa} \{F_n(s_{\kappa}) - \kappa\} \right].$$

Table 8 shows risk measures for the total unpaid loss based on 500,000 simulations for Models I and II. Given the GLMs fitted to the marginals, one would expect an average total unpaid loss of \$438,088; the small discrepancy between this value and the approximations is due to simulation. The risk measures are all smaller for Model I than for Model II. This is a little surprising because Model II takes into account the negative dependence between LOBs 2 and 3, which means there should be more risk diversification in that case than with Model I. Nevertheless, Model II is more conservative than Model I in the sense that it does not assume LOB 6 is independent from the others. In addition, Model II is based in part on Plackett and  $t(2)$  copulas, which exhibit tail dependence, whereas members of Frank's copula family in Model I do not.

Table 8: Risk measures for 500,000 simulations.

Model	Average	St. Dev.	VaR <sub>95%</sub>	VaR <sub>99%</sub>	TVaR <sub>99%</sub>
I	\$438,115	\$13,706	\$460,938	\$470,750	\$475,697
II	\$438,101	\$13,808	\$461,179	\$471,486	\$476,763

Insurance companies also have to determine capital allocations, i.e., the share of the risk capital to be allocated to each LOB. This capital allocation helps to point out most and least profitable sectors of activities in a company. Capital allocation principles have first been introduced by Tasche (1999), and Bargès et al. (2009) provide a review on that topic. Here, TVaR-based capital allocations are used. If

$$X^{(\ell)} = \sum_{i=2}^{10} \sum_{j=12-i}^{10} p_i^{(\ell)} X_{ij}^{(\ell)}$$

is the unpaid loss for LOB  $\ell$ , the capital allocated to that LOB is

$$\text{TVaR}_\kappa(X^{(\ell)}; S) = \frac{E[X^{(\ell)} \mathbf{1}\{S > \text{VaR}_\kappa(S)\}] + \beta_\kappa E[X^{(\ell)} \mathbf{1}\{S = \text{VaR}_\kappa(S)\}]}{1 - \kappa},$$

where  $\beta_\kappa = [F_S\{\text{VaR}_\kappa(S)\} - \kappa] / \Pr\{S = \text{VaR}_\kappa(S)\}$  if the denominator is strictly positive and 0 otherwise. This quantity can be estimated by

$$\widehat{\text{TVaR}}_\kappa(X^{(\ell)}; S) = \frac{1}{n(1 - \kappa)} \left\{ \sum_{j=1}^n X_j^{(\ell)} \mathbf{1}(S_j > s_\kappa) + \frac{F_n(s_\kappa) - \kappa}{\frac{1}{n} \sum_{\ell=1}^n \mathbf{1}(S_\ell = s_\kappa)} \sum_{j=1}^n X_j^{(\ell)} \mathbf{1}(S_j = s_\kappa) \right\},$$

where  $X_1^{(\ell)}, \dots, X_n^{(\ell)}$  are the  $n$  realizations of  $X^{(\ell)}$  corresponding to the realizations  $S_1, \dots, S_n$ .

In Table 9, TVaR-based capital allocations are shown for both models as well as for the ‘‘Silo’’ method, which is widespread in industry (Ajne, 1994). It is clear that the Silo method overestimates the total capital required as it implicitly assumes that risks are comonotonic, thereby preventing any form of diversification. The results for Models I and II are similar. While the capital allocations for LOBs 4 and 5 are higher in Model II than in Model I, they are lower for LOBs 2 and 3, outlining the additional risk diversification that is possible in the presence of negative dependence.

Table 9: Risk capital allocation for 500,000 simulations.

Model	TVaR <sub>99%</sub> - based capital allocations						Total
	LOB 1	LOB 2	LOB 3	LOB 4	LOB 5	LOB 6	
Silo	\$42,510	\$157,764	\$87,141	\$90,237	\$22,027	\$118,807	\$518,485
I	\$37,006	\$151,247	\$82,578	\$74,320	\$18,639	\$111,907	\$475,697
II	\$36,891	\$147,418	\$79,719	\$81,928	\$19,285	\$111,521	\$476,763

The risk measures in Tables 8 and 9 could be used to set internal capital targets, but they do not incorporate parameter uncertainty, as the model is assumed to be correct. A way to quantify estimation error and to tackle potential model over-fitting is to use a parametric bootstrap; see, e.g., Taylor and McGuire (2007), Kirschner et al. (2008) and Shi and Frees (2011). The parametric bootstrap procedure is as follows:

1. Simulate 55 observations from the dependence model, and transform them into observations of the loss ratios for the top triangle, accident year  $i = 1, \dots, 10$  and development year  $j = 1, \dots, 11 - i$  using the inverse marginal distributions.
2. Fit the marginal GLMs (log-normal for LOB 1 and Gamma for LOBs 2–6).
3. Compute the residuals from the GLMs.
4. Fit the copula model to the ranks of the residuals obtained.
5. From this new model, simulate the total unpaid loss using the steps described for simulation. The aggregate value is the simulated total unpaid loss.
6. Make  $N = 10,000$  repetitions of steps 1–5 to obtain the approximate distribution of the unpaid loss.

It is generally known (Andrews, 2000) that the parametric bootstrap is not valid when a parameter is on or close to the boundary of the parameter space. In the nested Archimedean copula model, the constraints on the dependence parameters, and the fact that  $\theta_2$  is close to  $\theta_1$  is likely to invalidate the bootstrap procedure. Therefore, results for Model I should be interpreted with caution.

Tables 10 and 11 show risk measures and capital allocations obtained with 10,000 bootstrap simulations, while Figure 9 shows the predictive distribution obtained for Model I (left) and Model II (right). The risks measures in Table 10 are similar for both models and are much higher than those reported in Table 8; this highlights the importance of incorporating parameter uncertainty. The observations made on capital allocations obtained by simulation for LOBs 2, 3, 4 and 5 apply here as well. In the light of these results, the

insurer might thus consider increasing the volume of LOB 3 to take better advantage of risk diversification. Finally, observe that Model II allocates less capital to LOB 6 than Model I in reflexion of the fact that LOB 6 is related to LOBs 2 and 3 in Model II.

Table 10: Risk measures for 10,000 bootstrap simulations.

Model	Average	St. Dev.	VaR <sub>95%</sub>	VaR <sub>99%</sub>	TVaR <sub>99%</sub>
I	\$443,041	\$31,291	\$496,780	\$521,293	\$539,205
II	\$442,957	\$31,038	\$496,470	\$522,417	\$535,536

Table 11: Risk capital allocation for 10,000 bootstrap simulations.

Model	TVaR <sub>99%</sub> - based capital allocations						Total
	LOB 1	LOB 2	LOB 3	LOB 4	LOB 5	LOB 6	
Silo	\$60,740	\$189,466	\$103,465	\$111,946	\$26,637	\$157,345	\$649,599
I	\$40,519	\$167,492	\$90,228	\$75,015	\$18,565	\$147,386	\$539,205
II	\$41,919	\$158,306	\$83,978	\$88,665	\$20,858	\$141,810	\$535,536

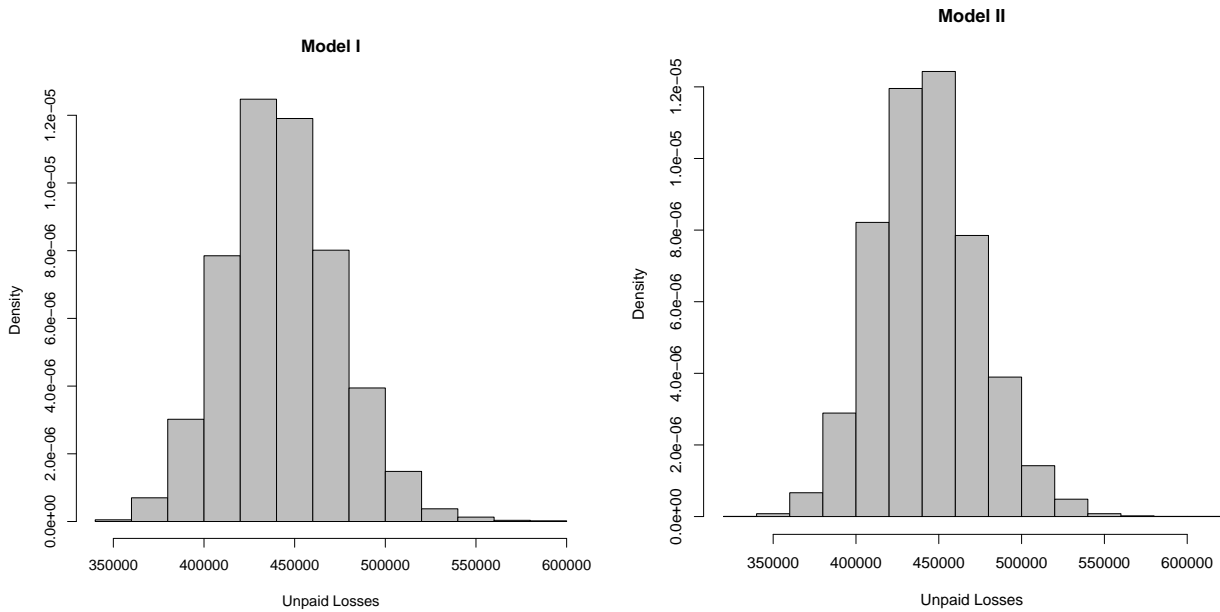


Figure 9: Predictive distributions based on 10,000 bootstrap replicates.

## 6 Summary and discussion

In this paper, rank-based procedures were introduced for the selection, estimation and validation of dependence structures for run-off triangles of property and casualty insurance claim data. The approach was illustrated using data for six lines of business from a large Canadian insurance company. Two hierarchical approaches were considered for modeling the pairwise dependence between different lines of business, i.e., fully nested Archimedean copulas and a copula-based risk aggregation model.

As simple and convenient as the nested Archimedean copula model may seem, its implementation raises more issues than one might anticipate initially. The success of this approach hinges on the choice of hierarchy and Archimedean generators at each of its levels. In principle, different Archimedean generators could be used throughout the structure, but the compatibility conditions to ensure the construction is valid are not always easy to verify. As there is no selection technique for generators, practitioners typically assume that

they are all from the same parametric family  $\varphi_\theta$ . In the latter case, conditions for the validity of the nested copula typically boil down to the constraint  $\theta_1 \geq \dots \geq \theta_d$ ; see, e.g., Hofert (2010).

As illustrated in the present paper, the use of the same generator throughout a fully nested Archimedean copula model has strong implications on the dependence structure. In particular, each variable is linked by the same bivariate copula to any variable appearing in a lower level of the hierarchy and, therefore, shares the same dependence characteristics with all of them in terms of symmetry, tail dependence, etc. In addition, the compatibility conditions imply Eq. (1) which are not only restrictive, but also raise issues in the parametric bootstrap. Indeed, when the data in a bootstrap sample would lead to unconstrained estimates  $\hat{\theta}_1, \dots, \hat{\theta}_d$  for which  $\hat{\theta}_1 \geq \dots \geq \hat{\theta}_d$  does not hold, one or more of the constrained parameter estimates end up being equal to 0. When this happens repeatedly, the dependence between the LOBs is underestimated. Thus, it is still unclear that this model can be used in a parametric bootstrap procedure to obtain the predictive distribution of unpaid losses, due to the optimization problem that is not standard.

Working with the risk aggregation model allows one to avoid most of these issues. The tree structure can be determined using hierarchical clustering, the copulas can be chosen freely at each aggregation step and standard tools for bivariate copula selection, estimation and validation are available. Furthermore, the application of the parametric bootstrap to this context is standard, as there are no constraints on the parameters. Overall, the model provides greater flexibility and the dependence structure can be considerably more complex than what can be achieved with the nested Archimedean approach. However, the conditional independence assumption must be satisfied (at least approximately) and formal tools for checking this assumption remain to be developed. Another minor irritant is the fact that simulation from this model relies on the Iman–Conover reordering algorithm, which while efficient is not yet included in standard software; in contrast, sampling from the fully nested Archimedean copula is easily done with the R package `copula`.

Perhaps the most significant limitation of the rank-based approach to risk aggregation modeling described here is that it can only be applied to data or residuals that are (at least approximately) identically distributed. Another requirement for this approach to make sense is that the sums that are linked by the copulas have the same number of components. This means that the risk aggregation model cannot be extended easily to include calendar year dependence, as Abdallah et al. (2015) did using nested Archimedean copulas. Unfortunately, this approach is not amenable to estimation and validation procedures based on ranks, as there is then only one observation for each copula in the model.

## A Nested Archimedean copula density

The 3-dimensional fully nested Archimedean copula is defined, for all  $u, v, w \in (0, 1)$ , by

$$C(u, v, w) = C_{\theta_2}\{w, C_{\theta_1}(u, v)\},$$

where  $\theta_1 \geq \theta_2 \geq 0$ . To ease notation, let  $C_\theta^{(i,j)}(u, v) = \partial^{i+j} C_\theta(u, v) / \partial u^i \partial v^j$  for  $i, j \in \{0, 1, 2\}$ . The density of the nested Archimedean copula can be derived easily using the chain rule, viz.

$$\begin{aligned} c(u, v, w) &= \frac{\partial^3}{\partial u \partial v \partial w} C_{\theta_2}\{w, C_{\theta_1}(u, v)\} = \frac{\partial^2}{\partial u \partial v} C_{\theta_2}^{(1,0)}\{w, C_{\theta_1}(u, v)\} \\ &= \frac{\partial}{\partial u} \left[ C_{\theta_2}^{(1,1)}\{w, C_{\theta_1}(u, v)\} C_{\theta_1}^{(0,1)}(u, v) \right] \\ &= C_{\theta_2}^{(1,2)}\{w, C_{\theta_1}(u, v)\} C_{\theta_1}^{(1,0)}(u, v) C_{\theta_1}^{(0,1)}(u, v) + C_{\theta_2}^{(1,1)}\{w, C_{\theta_1}(u, v)\} C_{\theta_1}^{(1,1)}(u, v). \end{aligned}$$

This expression is explicit, though it involves partial derivatives. In the case of the Frank copula, the expressions required are the copula

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left\{ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right\},$$

its density

$$C_\theta^{(1,1)}(u, v) = c_\theta(u, v) = \frac{-\theta e^{-\theta(u+v)}(e^{-\theta} - 1)}{\{(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1)\}^2},$$





Table 18: Parameter and reserve estimations.

LOB $\ell$		1	2	3	4	5	6
GLM		Log-normal	Gamma	Gamma	Gamma	Gamma	Gamma
$\zeta^{(\ell)}$		-4.031 (0.157)	-3.628 (0.148)	-3.501 (0.098)	-2.365 (0.173)	-4.064 (0.148)	-2.872 (0.167)
Accident Year	2	-0.226 (0.153)	-0.750 (0.151)	0.053 (0.097)	-0.413 (0.174)	-0.121 (0.151)	0.101 (0.177)
	3	0.022 (0.161)	-0.729 (0.160)	-0.156 (0.100)	-0.196 (0.183)	0.171 (0.161)	0.163 (0.177)
	4	-0.028 (0.168)	-0.651 (0.168)	0.239 (0.105)	-0.112 (0.190)	0.129 (0.168)	-0.136 (0.184)
	5	-0.112 (0.177)	-0.741 (0.174)	0.137 (0.110)	-0.095 (0.199)	0.092 (0.173)	-0.024 (0.191)
	6	-0.183 (0.189)	-0.574 (0.185)	0.120 (0.117)	-0.001 (0.210)	0.396 (0.187)	0.095 (0.203)
	7	0.170 (0.205)	-0.574 (0.200)	0.003 (0.127)	0.197 (0.227)	0.254 (0.200)	0.069 (0.219)
	8	0.032 (0.228)	-0.658 (0.220)	-0.160 (0.141)	-0.012 (0.253)	0.055 (0.222)	-0.017 (0.246)
	9	0.131 (0.268)	-1.147 (0.255)	0.169 (0.167)	-0.628 (0.295)	-0.259 (0.260)	0.131 (0.289)
	10	0.261 (0.362)	-1.625 (0.340)	0.175 (0.226)	-0.754 (0.393)	-0.676 (0.348)	-0.032 (0.390)
	Dev. Lag	2	1.311 (0.154)	2.061 (0.145)	0.815 (0.096)	0.450 (0.167)	0.419 (0.149)
3		1.438 (0.161)	2.065 (0.151)	0.817 (0.101)	-0.055 (0.175)	0.114 (0.155)	0.076 (0.174)
4		1.150 (0.168)	2.018 (0.158)	0.849 (0.106)	-0.507 (0.183)	-0.358 (0.163)	-0.095 (0.182)
5		0.874 (0.177)	1.818 (0.166)	0.717 (0.112)	-0.759 (0.193)	-0.582 (0.173)	-0.406 (0.192)
6		0.636 (0.189)	1.297 (0.176)	0.283 (0.120)	-1.580 (0.207)	-1.154 (0.182)	-0.481 (0.206)
7		0.392 (0.205)	0.773 (0.193)	-0.115 (0.129)	-1.899 (0.223)	-1.870 (0.201)	-0.757 (0.226)
8		0.137 (0.228)	-0.493 (0.216)	-1.001 (0.143)	-2.670 (0.250)	-2.103 (0.219)	-1.215 (0.248)
9		-0.291 (0.268)	-0.429 (0.255)	-1.375 (0.169)	-3.762 (0.298)	-3.849 (0.257)	-2.612 (0.304)
10		-0.522 (0.362)	-1.358 (0.340)	-0.715 (0.226)	-2.960 (0.393)	-6.248 (0.348)	-2.764 (0.390)
sd or scale			0.326 (0.031)	10.700 (2.009)	24.046 (4.554)	8.038 (1.502)	10.078 (1.891)
Reserve		36,063	132,919	78,665	73,220	18,290	98,931
C-L Reserve		35,411	146,794	76,500	75,551	18,726	100,704

## References

- Abdallah A, Boucher JP, Cossette H (2015) Modeling dependence between loss triangles with hierarchical Archimedean copulas. *ASTIN Bulletin* 45:577-599
- Ajne B (1994) Additivity of chain-ladder projections. *ASTIN Bulletin* 24(2):311-318
- Andrews DW (2000) Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68:399-405
- Arbenz P, Hummel C, Mainik G (2012) Copula based hierarchical risk aggregation through sample reordering. *Insurance: Mathematics and Economics* 51:122-133
- Bargès M, Cossette H, Marceau E (2009) TVaR-based capital allocation with copulas. *Insurance : Mathematics and Economics* 45:348-361
- Ben Ghorbal N, Genest C, Nešlehová J (2009) On the Ghoudi, Khoudraji, and Rivest test for extreme value dependence. *The Canadian Journal of Statistics* 37:534-552
- Braun C (2004) The prediction error of the chain ladder method applied to correlated run-off triangles. *ASTIN Bulletin* 34:399-434
- Brehm P (2002) Correlation and the aggregation of unpaid loss distributions. *Casualty Actuarial Society Forum* (Fall) 2:1-23
- Bürgi R, Dacorogna MM, Iles R (2008) Risk aggregation, dependence structure and diversification benefit. *Stress Testing for Financial Institutions*, available at [www.ssrn.com/abstract=1468526](http://www.ssrn.com/abstract=1468526)
- Côté MP (2014) Copula-based risk aggregation modelling. Master's thesis, McGill University, Montréal, Québec, Canada
- Côté MP, Genest C (2015) A copula-based risk aggregation model. *Canadian Journal of Statistics* 43:60-81
- De Jong P (2012) Modeling dependence between loss triangles. *North American Actuarial Journal* 16:74-86
- Embrechts P, Lindskog F, McNeil AJ (2003) Modelling dependence with copulas and applications to risk management. In: Rachev S (ed) *Handbook of Heavy Tailed Distributions in Finance*, Elsevier/North-Holland, Amsterdam
- Genest C, Favre AC (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering* 12:347-368
- Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82:543-552

- Genest C, MacKay RJ (1986) Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *The Canadian Journal of Statistics* 14:145–159
- Genest C, Nešlehová J (2012) Copulas and copula models. In: El-Shaarawi AH, Piegorsch WW (eds) *Encyclopedia of Environmetrics*, 2nd edn, Wiley, Chichester
- Genest C, Rémillard B, Beaudoin D (2009) Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44:199–213
- Hofert M (2010) Construction and sampling of nested Archimedean copulas. In: Jaworski P, Durante F, Härdle WK, Rychlik T (eds) *Copula Theory and its Applications*, Lecture Notes in Statistics, vol 198, Springer, Berlin, pp 147–160
- Hofert M (2011) Efficiently sampling nested Archimedean copulas. *Computational Statistics and Data Analysis* 55:57–70
- Hofert M, Mächler M (2011) Nested Archimedean copulas meet R: The `nacopula` package. *Journal of Statistical Software* 39:1–20
- Hollander M, Wolfe DA, Chicken E (2014) *Nonparametric Statistical Methods*, 3rd edn. Wiley, Hoboken, NJ
- Joe H (1997) *Multivariate Models and Dependence Concepts*. Chapman & Hall, London
- Kirschner G, Kerley C, Isaacs B (2008) Two approaches to calculating correlated reserve indications across multiple lines of business. *Variance* 2(1):15–38
- Mainik G (2015) Risk aggregation with empirical margins: Latin hypercubes, empirical copulas, and convergence of sum distributions. *Journal of Multivariate Analysis* 141:197–216
- McNeil AJ (2008) Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation* 78:567–581
- McNeil AJ, Frey R, Embrechts P (2015) *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd edn. Princeton University Press, Princeton, NJ
- Merz M, Wüthrich M (2008) Prediction error of the multivariate chain ladder reserving method. *North American Actuarial Journal* 12:175–197
- Merz M, Wüthrich M, Hashorva E (2013) Dependence modelling in multivariate claims run-off triangles. *Annals of Actuarial Science* 7(1):3–25
- Nelsen RB (2006) *An Introduction to Copulas*. Springer, Berlin
- Pröhl C, Schmidt K (2005) Multivariate chain-ladder. *ASTIN Colloquium 2005*, ETH Zürich, Switzerland
- Rémillard B, Scaillet O (2009) Testing for equality between two copulas. *Journal of Multivariate Analysis* 100:377–386
- Savu C, Trede M (2010) Hierarchies of Archimedean copulas. *Quantitative Finance* 10:295–304
- Schmidt K (2006) Optimal and additive loss reserving for dependent lines of business. *Casualty Actuarial Society Forum* (fall):319–351
- SCOR (2008) From principle-based risk management to solvency requirements. Technical report, SCOR, Switzerland, available at: [www.scor.com/images/stories/pdf/scorpapers/sstbook\\_second\\_edition\\_final.pdf](http://www.scor.com/images/stories/pdf/scorpapers/sstbook_second_edition_final.pdf)
- Shi P, Basu S, Meyers G (2012) A Bayesian log-normal model for multivariate loss reserving. *North American Actuarial Journal* 16:29–51
- Shi P, Frees E (2011) Dependent loss reserving using copulas. *ASTIN Bulletin* 41:449–486
- Tasche D (1999) Risk contributions and performance measurement. Working paper, Technische Universität München, Germany
- Taylor G, McGuire G (2007) A synchronous bootstrap to account for dependencies between lines of business in the estimation of loss reserve prediction error. *North American Actuarial Journal* 11(3):70–88
- Whelan N (2004) Sampling from Archimedean copulas. *Quantitative Finance* 4:339–352
- Zhang Y (2010) A general multivariate chain ladder model. *Insurance: Mathematics and Economics* 46:588–599