White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/222417/

Version: Published Version

Article:
Dias, Alexandra, Ismail, Isaudin and Aihua, Zhang (2025) Copula-Based Risk Aggregation and the Significance of Reinsurance. Risks. 44. ISSN: 2227-9091

https://doi.org/10.20944/preprints202410.0743.v1

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

risks

MDPI

Article

Copula-Based Risk Aggregation and the Significance of Reinsurance

Alexandra Dias [1], Isaudin Ismail [2,3] and Aihua Zhang [4,5,*]

[1] School for Business and Society, University of York, York YO10 5ZF, UK; alexandra.dias@york.ac.uk
[2] Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh 84600, Johor, Malaysia; isaudin@uthm.edu.my
[3] Department of Decision Sciences, HEC Montréal, Montréal, QC H3T 2A7, Canada
[4] Department of Mathematical Sciences, College of Science, Mathematics and Technology, Wenzhou-Kean University, 88 Daxue Road, Ouhai, Wenzhou 325060, China
[5] Department of Mathematical Sciences, Dorothy and George Hennings College of Science, Mathematics and Technology, Kean University, 1000 Morris Ave, Union, NJ 07083, USA
* Correspondence: azhang@kean.edu

Abstract: Insurance companies need to calculate solvency capital requirements in order to ensure that they can meet their future obligations to policyholders and beneficiaries. The solvency capital requirement is a risk management tool essential for addressing extreme catastrophic events that result in a high number of possibly interdependent claims. This paper studies the problem of aggregating the risks coming from several insurance business lines and analyses the effect of reinsurance on the level of risk. Our starting point is to use a hierarchical risk aggregation method which was initially based on two-dimensional elliptical copulas. We then propose the use of copulas from the Archimedean family and a mixture of different copulas. Our results show that a mixture of copulas can provide a better fit to the data than an individual copula and consequently avoid over- or under-estimation of the capital requirement of an insurance company. We also investigate the significance of reinsurance in reducing the insurance company's business risk and its effect on diversification. The results show that reinsurance does not always reduce the level of risk, but can also reduce the effect of diversification for insurance companies with multiple business lines.

Keywords: copula; reinsurance; capital requirement; risk aggregation; value at risk; tail value at risk

check for updates

Academic Editor: Alexander Melnikov

Received: 9 October 2024
Revised: 4 February 2025
Accepted: 19 February 2025
Published: 26 February 2025

Citation: Dias, Alexandra, Isaudin Ismail, and Aihua Zhang. 2025. Copula-Based Risk Aggregation and the Significance of Reinsurance. *Risks* 13: 44. https://doi.org/10.3390/risks13030044

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Determining the level of capital required for business continuity is essential for insurance companies. Interest in the literature concerning the significance of capital requirements is increasing, as evidenced by the studies by Nguyen and Molinari (2011), Floreani (2013), Clemente et al. (2015), Becker et al. (2022), and, most recently, Li and Yin (2024). The capital, which is mostly referred to as a capital requirement or business capital, should support an insurance company in minimizing the risk of insolvency and fulfilling its obligations to its policyholders. When extreme events happen, such as floods, earthquakes, hurricanes, and other catastrophic events, the claims amount to be paid by an insurance company can be extremely high. However, part of the claims can be passed to reinsurance companies. An insurance company (cedent) can transfer some risks to another insurer (the reinsurer), exchanging part of its unexpected future losses by the payment of a fixed premium.

Typically, the cedent insurance company keeps most of the risk and when large amounts of claims occur, these can originate not just from one business line but involve

other products as well. In other words, some insurance business lines are dependent on each other, in the sense that an increase in the claims amount being filled in one business line is accompanied by a higher claims amount in other business lines too. In addition, these claims due to correlated catastrophic events could lead to defaults in insurance systems, as highlighted by Giuzio et al. (2019) and Torri et al. (2022). Therefore, it is essential that regulators monitor these issues. To the best of our knowledge, the interaction between two critical aspects of insurance—dependence among risks and partial risk cession—remains underexplored. This study addresses this gap by evaluating the impact of reinsurance on the composition and risk profile of an insurer's portfolio across multiple business lines. Leveraging flexible copula models and portfolio diversification metrics, we analyse empirical data from the Australian insurance market to quantify these effects. Our findings provide actionable insights for policymakers, enabling them to design more effective strategies for the reinsurance industry and enhance systemic stability. Hence, there is a need to properly model the aggregate risk of losses across a broad range of insurance products.

Aggregating the risk of losses for insurance companies is challenging. The most crucial aspect of the aggregation process is modelling the dependence structure between the risks of losses across different business lines. Examining linear correlations is a classic approach to model risk dependence but fails to incorporate all possible dependence structures. The use of the classic multivariate Gaussian model implicitly assumes symmetry between the multivariate tails and a linear relationship between variables. To investigate potential non-linearities, for instance, whether losses are more strongly related than gains in financial data, we use copula models, which can be seen as a generalisation of the classic multivariate Gaussian model. When selecting a copula model, if the multivariate Gaussian is the most appropriate distribution, then it will still be the chosen model. In the case of insurance claims, we must consider the possibility of asymmetries and non-linearities. Small claims might originate more independently from each other, while very large claims might have underlying common causes (e.g., weather-related events) and thus exhibit stronger dependence. When linearity and symmetry may not hold, copulas are the appropriate method to model the dependence structure. Copulas have received increasing interest from researchers and practitioners in recent years.

This aims of this paper are twofold . First, we focus on modelling the aggregation of risks from different business lines in insurance. Second, we explore the effect of reinsurance on the level of risks and how this relates to the dependence structure between different business lines. To model the multivariate structure of the insurance risks, we use a hierarchical risk aggregation method based on two-dimensional copulas. As an alternative to hierarchical copulas, one could use vine copula models, as in Joe (1996) and Czado (2019), but these may not perform as well as hierarchical copulas when some variables, more strongly related than others, form natural groups inherent to the specific application (Chaoubi et al. 2021). However, hierarchical copulas may outperform vine copulas when certain variables that are more strongly related than others form natural groups inherent to the specific application. If we ignore the particular application, both methods can be used to model multivariate distributions. These two procedures have been compared in Nanthakumar (2022) from a purely probabilistic point of view. In our study, fire and house insurance claims are expected to be strongly related, as are liability and CTP* (compulsory third party liability) insurance claims. This natural hierarchical dependence is confirmed in our data analysis, justifying our choice of the hierarchical copula methodology.

The hierarchical risk aggregation approach recently adopted by Côté and Genest (2015) was developed by Arbenz et al. (2012). The hierarchical aggregation procedure is based on rooted trees, which include branching and leaf nodes, and uses the elliptical copula family for each aggregation step. However, as highlighted by Embrechts et al. (2003), this copula

family has certain drawbacks, such as its inability to capture dependence structures that are not radially symmetric. Especially in the case of extreme events, the dependence of large losses from different business lines cannot be modelled by the elliptical copula family (see Nguyen and Molinari 2011).

In this paper, we address this issue by proposing a novel hierarchical framework that integrates copulas from the Archimedean family. Archimedean copulas offer inherent asymmetry and flexibility for modelling diverse dependence structures, particularly in tail dependencies. To further enhance adaptability, we incorporate mixtures and rotated variants of Archimedean copulas, selected through rigorous goodness of fit tests to match empirical data patterns. This dual approach of combining hierarchical aggregation with Archimedean-based copulas represents a key innovation, enabling robust risk modelling without relying on restrictive conditional assumptions.

For the empirical application, we use data from the Australian Prudential Regulation Authority (APRA). Modelling insurance claims for capital requirements using Australian data is relevant not only for Australian insurers but also for insurers based in other countries. While Solvency II establishes the capital requirements in Europe, APRA's GPS Prudential Standards similarly regulate insurers' capital requirements in Australia. Solvency II establishes the equivalence between the two supervisory frameworks, which is essential for European insurers operating in Australia. Hence, the methodology and results from this study are relevant for all countries with regulatory equivalence under Solvency II.

Tang and Valdez (2006) analyse APRA's data consisting of 19 semi-annual gross incurred claims and earned premiums from December 1992 to June 2002. We chose to use more recent data with a quarterly frequency in order to increase the sample size and improve the estimation of the risk aggregation model. As a result, a total of 28 observations, consisting of quarterly premiums earned and incurred claims—both gross and net of reinsurance—for five business lines, were selected for the period between September 2010 and June 2017. The quarterly incurred claims and premiums earned were then used to calculate loss ratios for the five different business lines. The risk aggregation model was selected based on the resulting loss ratios to measure the associated risks. The gross and net of reinsurance loss ratios were used to examine the change in the level of risk for each business line and for the aggregate risk.

Research on risk aggregation with copulas applied to insurance was pioneered by Wang (1998). This research introduces the concept of copulas and chooses a Gaussian copula as one of the useful tools in determining the risk aggregation of an insurance company by combining correlated loss distributions. More specifically, the aggregate loss distribution is determined by the combination of the effects of claim frequency and claim severity distribution. By contrast, Tang and Valdez (2006) use copula models to aggregate risks in order to determine economic capital as well as diversification benefits, focusing on the insurance industry. Using multiple insurance business lines' data, they analyse the importance of selecting an appropriate copula model to avoid underestimation or overestimation of capital required, which consequently may affect the level of capital for insurance products. Bürgi et al. (2008) highlight that modelling the dependence between risks is important as it is a form of rule for risk aggregation. Their research also considers various methods for modelling dependencies, which subsequently affect diversification benefits, and shows that overestimation of diversification may cause inaccurate computation of risk-based capital (RBC). Nguyen and Molinari (2011) use copulas to address the loopholes of Solvency II, such as linear correlations being used to measure the dependence structure of correlated risks. However, a linear correlation may not be suitable for modelling dependence structure and may not be able to capture all the information of a tail distribution. To overcome this problem, the authors propose a method of risk aggregation via copula to determine the

dependence structure between risks. Nevertheless, their focus is based on the perspective of Solvency II, rather than on risk aggregation modelling based on real data.

Modelling risk aggregation using a dimensional copula can be very challenging and requires more parameters to be estimated than traditional two-dimensional or bivariate copula models (Bürgi et al. 2008). With this in mind, we consider hierarchical aggregation as an alternative modelling technique based on two-dimensional copulas for high-dimensional copulas. This model, introduced by Arbenz et al. (2012), has the advantage of not requiring the specification of a copula for all business lines. Instead, a copula and the joint dependence between the aggregated sub-business lines are determined at each aggregation step. The aggregation model is represented by a rooted tree, which consists of branching nodes and leaves based on graph theory.

In addition, we investigate the significance of reinsurance from a risk management perspective. Insurance companies are able to transfer risks to reinsurance, and, as a result, capital is saved from being allocated to these risks (Baur et al. 2004). Previous research by Cummins et al. (2008) proves that insurance companies purchase reinsurance for the benefits of reducing the loss ratio, measured by its volatility. Reinsurance also provides protection against catastrophes by limiting the liability on specific risks. The drawback of reinsurance is that insurers' cost of production is increased. Furthermore, reinsurance provides other benefits, such as capital relief as well a flexible financing. In our study, we use two measures of portfolio diversification, where the portfolio is defined as a collection of insurance business lines. The first measure, diversification, is derived from Shannon entropy (Shannon 1948), while the second was introduced in Choueifaty and Coignard (2008). While the first measure concentrates on the weights of the business lines, the second one focuses on the risk of the portfolio. In our empirical analysis of Australian insurance data, we find that, although the portfolio of business lines without reinsurance is more balanced and hence more diversified according to Shannon's measure, reinsurance reduces the risk of the (reinsured) portfolio, thereby increasing diversification according to Choueifaty and Coignard's measure.

The remainder of this paper is organised as follows: Section 2 discusses the methods for aggregating risk using a hierarchical copula aggregation model, copula simulations, and the determination of capital requirements. Section 3 contains the estimation of the hierarchical aggregation copula model and analysis of the results. In Section 4, we study the effects of reinsurance on the level of risk and diversification of the portfolio of different business lines. Section 5 concludes the paper.

2. Copula-Based Hierarchical Aggregation Model

In finance and insurance, popular models for problems involving a large number of random variables have been based on copula functions. Different copula models have been proposed. These include Archimedean and elliptical copula models (Genest and Nešlehová 2012), vine copula models (Aas et al. 2009; Kurowicka and Joe 2010), and hierarchical copula models (Mai and Scherer 2012; Côté and Genest 2015; Cossette et al. 2017).

While some of these models impose restrictive dependence structures, complicating inference, vine copulas—despite their growing popularity—require a critical assumption: each conditional bivariate copula must be independent of the conditioning variable, except through its marginal distributions (Haff et al. 2010). This assumption, while essential for accurate inference, limits their applicability in practice. In contrast, hierarchical copulas eliminate the need for conditional copulas entirely, avoiding such assumptions. Their straightforward structure requires only a single copula and corresponding joint dependence at each aggregation step, significantly simplifying estimation. In this study, we adopt

hierarchical copulas to strike an optimal balance between ease of estimation and model flexibility, offering a robust alternative to more complex approaches. This article adopts the hierarchical copula model with the goal of achieving a good compromise between ease of estimation and flexibility.

2.1. The Definition of Copula

Bivariate copulas are the main building block of hierarchical aggregation copula models. Here we only provide the basic definition in order to introduce the notation and we refer the reader to Nelsen (2006) and McNeil et al. (2015) for an introduction to copulas and the definition of specific copula families.

Given a d-dimensional random vector $(X_1, X_2, \ldots, X_d)'$,[1] according to Sklar (1959), there exists a function $C : [0,1]^d \to [0,1]$ such that

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)),$$

where $F_i(x) = \mathbb{P}(X_i \leq x_i)$ is the cumulative distribution function (cdf) of X_i for $i = 1, 2, \ldots, d$, and C is a copula function. In fact, a copula is a multivariate joint cdf with uniform margins. If the univariate cdf's F_i are continuous, then the copula function C is unique.

2.2. Hierarchical Aggregation Copula Models

Hierarchical copula models draw on results from graph theory on rooted trees (Diestel 2017). Following the notation used in Arbenz et al. (2012), a rooted tree τ is composed of leaf nodes and branching nodes where one of the branching nodes is the root. The subset of branching nodes is denoted by $\mathscr{B}(\tau)$, the subset of leaf nodes is denoted by $\mathscr{L}(\tau)$, and the root node by $\varnothing$. Naturally, $\mathscr{B}(\tau) \cup \mathscr{L}(\tau) = \tau$ and $\mathscr{B}(\tau) \cap \mathscr{L}(\tau) = \emptyset$. In order to use rooted trees to aggregate the losses of several business lines, we make the following assumptions:

- Each leaf node in the rooted tree is associated with the loss of business line i, represented by a random variable X_i.
- Each branching node is associated with the sum of the business lines mapped to that node's children.

In Figure 1, we illustrate the mapping to a rooted tree of three insurance loss random variables, X_m, X_f, and X_h, representing the business lines Motor, Fire, and Household, respectively. Each leaf node corresponds to a business line and each branching node corresponds to the sum of the variables associated with its children leaf nodes. As in Arbenz et al. (2012), for simplicity, we assume that each branching node has two children, although the results on rooted trees used in this paper are valid for branching nodes with any number of children (see Arbenz et al. 2012). By assuming that each branching node has only two children, we can simplify the construction and estimation of the model, as only bivariate copulas are involved. In order to define the aggregation model, we denote by $(X_i)'_{i \in T} = (X_1, X_2, \ldots, X_d)'$ the vector of random variables, where each X_i represents the loss of the business line i. The rooted tree aggregation model for the random vector $(X_i)_{i \in \tau}$ is determined by

- a rooted tree structure τ,
- univariate cdf's $F_i : \mathbb{R} \to [0,1]$ for all leaf nodes i in $\mathscr{L}(\tau)$, and
- bivariate copula functions $C_j : [0,1]^2 \to [0,1]$ for the two children of each branching node j in $\mathscr{B}(\tau)$.

We denote the tree aggregation model by $(\tau, (F_i)_{i\in\mathscr{L}(\tau)}, (C_j)_{j\in\mathscr{B}(\tau)})$. Using this modelling approach, we obtain the distribution of the root node, which represents the aggregate total loss

$$X_\varnothing = X_1 + X_2 + \ldots + X_d = \sum_{i\in\mathscr{L}(\tau)} X_i$$

based on the univariate cdf's F_i for the business lines associated with the leaf nodes and the bivariate copulas associated with the branching nodes.

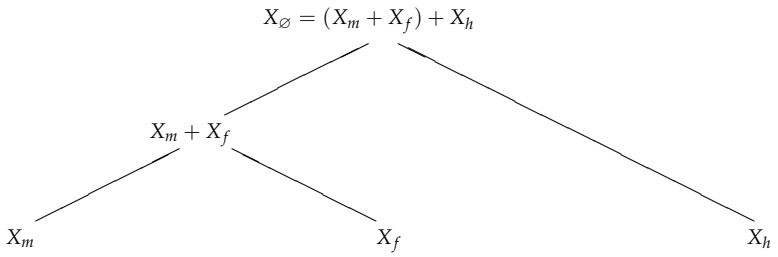

Figure 1. Illustration of a hierarchical loss aggregation copula model built by allocating each of the three individual business lines, represented by X_m, X_f, and X_h, to a leaf node of a rooted tree. The structure of the tree in this example is determined by the assumption that the pair (X_m, X_f) has the strongest dependence among the three possible pairs of individual business lines.

2.2.1. Existence and Uniqueness of a Joint Distribution

The existence and uniqueness of the joint distribution of the hierarchical aggregation copula model for the vector $(X_1, X_2, \ldots, X_d)'$ have been studied in Arbenz et al. (2012). Here we only summarise the conditions and the main results necessary in this paper. Given a rooted tree aggregation model $(\tau, (F_i)_{i\in\mathscr{L}(\tau)}, (C_j)_{j\in\mathscr{B}(\tau)})$ where each branching node $j \in \mathscr{B}(\tau)$ is the sum of its children, the random vector $(X_i)_{i\in\tau}$ is called *mildly tree-dependent*. A mildly tree-dependent random vector $(X_i)_{i\in\tau}$ is called *tree-dependent* if for each branching node $i \in \mathscr{B}(\tau)$, given X_i, its descendants $(X_j)_{j\in\mathscr{D}(i)}$, where $\mathscr{D}(i)$ is the set of descendent nodes, are conditionally independent of the remaining nodes $(X_j)_{j\in\tau\setminus\mathscr{D}(i)}$; that is,

$$(X_j)_{j\in\mathscr{D}(i)} \perp (X_j)_{j\in\tau\setminus\mathscr{D}(i)} \mid X_i \qquad \text{for all } i \in \mathscr{B}(\tau).$$

This conditional independence condition does not, however, imply that $(X_j)_{j\in\mathscr{D}(j)}$ is independent of $(X_j)_{j\in\tau\setminus\mathscr{D}(i)}$, because their dependence may come from X_i.

Theorem 1. *Given a rooted tree aggregation model* $\left(\tau, (F_i)_{i\in\mathscr{L}(\tau)}, (C_j)_{j\in\mathscr{B}(\tau)}\right)$*, a tree-dependent random vector exists and its joint distribution is unique.*

For the proof of this result, see Arbenz et al. (2012). For the example illustrated in Figure 1, the joint distribution of the hierarchical aggregation copula model for the vector $(X_m, X_f, X_h)'$ exists and is unique if and only if

$$(X_m, X_f) \perp (X_\varnothing, X_h) \mid X_m + X_f,$$

where $X_\varnothing = X_m + X_f + X_h$. This means that all the information in X_m and X_f that influences X_h is contained in $X_m + X_f$.

Under the above theorem, if all the univariate and copula distributions are absolutely continuous, then the joint density function is given by the following proposition shown in Côté and Genest (2015).

Proposition 1. *Given a rooted tree aggregation model* $(\tau, (F_i)_{i\in\mathscr{L}(\tau)}, (C_j)_{j\in\mathscr{B}(\tau)})$ *with d leaf nodes associated with the vector* $\mathbf{X} = (X_1, X_2, \ldots, X_d)'$*, the joint density function of the vector* $\mathbf{X}$ *is given by*

$$f_{\mathbf{X}}(x_1,\ldots,x_d) = \prod_{j=1}^{d-1} c_j \left[F_{\mathscr{LD}(j1)}\left(\sum_{i\in\mathscr{LD}(j1)} x_i \right), F_{\mathscr{LD}(j2)}\left(\sum_{i\in\mathscr{LD}(j2)} x_i \right) \right] \prod_{i=1}^{d} f_i(x_i),$$

for all $x_1, \ldots, x_d \in \mathbb{R}$*, where* $\mathscr{LD}(ji)$ *represents the leaf nodes in the set of descendants of child node i of the branching node j,* $F_{\mathscr{LD}(jk)}$ *is the cdf of the sum of the leaf nodes in* $\mathscr{LD}(jk)$*,* $f_1, \ldots, f_d$ *are the univariate density functions of* $X_1, X_2, \ldots, X_d$ *respectively, and* c_j *is the copula density function of the children of* X_j *for* $j \in \mathscr{B}(\tau)$*.*

As an example, for the business lines represented by the random vector $(X_m, X_f, X_h)'$ associated with the rooted tree τ illustrated in Figure 1, the joint density function is given by

$$f_{\mathbf{X}}(x_m, x_f, x_h) = c_{m,f}\Big(F_m(x_m), F_f(x_f)\Big) c_{m+f,h}\Big(F_{m+f}(x_m + x_f), F_h(x_h)\Big) \cdot f_m(x_m) f_f(x_f) f_h(x_h),$$

for all (x_m, x_f, x_h), where F_i is the cdf of the univariate random variable X_i with the density function f_i, F_{m+f} is the cdf of $X_m + X_f$, $c_{m,f}$ is the copula density function of (X_m, X_f), and $c_{m+f,h}$ is the bivariate copula density function of $((X_m + X_f), X_h)$.

2.2.2. Simulation of Joint Distributions

Given the set of d business lines represented by the random variables $X_1, X_2, \ldots, X_d$, we determine the tree structure by iteratively aggregating the pair of variables with the strongest dependence. We use Kendall's tau to measure the dependence between pairs of random variables in the hierarchical aggregation procedure. For the motivation and justification for using Kendall's tau in this setting, see Côté and Genest (2015).

After defining the structure of the tree, we proceed with selecting the probability distribution for the random variable allocated to a leaf node and the copula family for the two children of each branching node in order to specify the hierarchical aggregation model. We use maximum likelihood estimation to estimate the parameters, and the goodness of fit methods outlined in Anderson and Darling (1954) and Genest et al. (2009) to select the best probability distributions.

The hierarchical aggregation model allows for estimating measures of risk based on the sum of the individual variables considered. We estimate these risk measures based on the simulation of observations from the aggregation model by generalising the algorithm introduced in Arbenz et al. (2012), which consists of a numerical approximation procedure where sample reordering induces the dependence structure, a technique that goes back to the work of Iman and Conover (1982).

We present below the algorithm for the case where all branching nodes have two children and the functional that produces the aggregation is a weighted sum of the branching nodes. Brechmann (2014) generalize the case where the aggregation functionals are Kendall functions.

2.2.3. Sample Reordering Numerical Approximation Algorithm

1. Define the number of simulations $N \in \mathbb{N}$.
2. Simulate N independent samples from the univariate random variables X_i $(i \in \mathscr{L}(\tau))$ associated with d leaf nodes: $X_i^k \sim F_i$ for $k = 1, \ldots, N$ and $i = 1, \ldots, d$, where F_i is the pre-determined univariate cdf for X_i.

3. Simulate N independent samples from the bivariate copula C_j ($j \in \mathscr{B}(\tau)$) associated with each of the $d-1$ branching nodes: $\mathbf{U}_j^k = (U_{j1}^k, U_{j2}^k) \sim C_j$ for $k = 1, \ldots, N$ and $j = 1, \ldots, d-1$.
4. Following a bottom-up approach, beginning at the branching nodes closer to the leaf nodes and ending at the root nodes, define the approximation for the cdf of each branching node $j \in \mathscr{B}(\tau)$ as

$$F_j^N(x) = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}\left\{ w_{j1}\, x_{j1}^{(r_{j1}^k)} + w_{j2}\, x_{j2}^{(r_{j2}^k)} \leq x \right\},$$

recursively, where $\mathbb{1}$ is the indicator function[2], x_{j1}^k and x_{j2}^k are (simulated) sample values of the random variables associated with the two children of the branching node j, w_{ji} is the weight given to variable X_{ji}, r_{ji}^k is the (component-wise) rank of u_{ji}^k, and $\{x_{ji}^{(1)}, x_{ji}^{(2)}, \ldots, x_{ji}^{(N)}\}$, $i = 1, 2$ are the ordered sample.

Once we have the estimate for the cdf of the total aggregate loss, we proceed with estimating the risk measures of interest.

2.3. Risk Estimation of the Aggregate Loss

After building the model for the aggregate loss, using the hierarchical copula model, we can estimate the risk of the aggregate loss. In the process of risk aggregation, the bivariate copula is used to link every two losses, such as X_1 and X_2. TVaR is then applied to estimate the risks of the aggregate losses. TVaR is used to fulfil the coherent risk measure property, as suggested by Acerbi et al. (2001), Acerbi and Tasche (2002), and Bargès et al. (2009). The TVaR of the loss represented by the random variable X at the confidence level α, for $\alpha \in (0, 1)$, is defined as

$$\mathrm{TVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_u(X)\, du,$$

where the VaR_α of the random loss X is given by

$$\mathrm{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}.$$

Conventionally, α typically takes the values 90%, 95%, or 99%. In order to estimate the TVaR, we use the following non-parametric estimator, of which a more detailed description can be found in Adam et al. (2008). Given n observations $\{x_1, x_2, \ldots, x_n\}$ of the variable X, the TVaR estimator is given by

$$\widehat{\mathrm{TVaR}}_\alpha = \frac{1}{n(1-\alpha)} \left(\sum_{i=1}^{\lfloor n(1-\alpha) \rfloor} x_{(n-i+1)} + (n(1-\alpha) - \lfloor n(1-\alpha) \rfloor) x_{(n-\lfloor n(1-\alpha) \rfloor)} \right), \quad (1)$$

where $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ is the ordered sample, and $\lfloor v \rfloor$ denotes the largest integer not greater than v. In our setting, we estimate the TVaR by applying Equation (1) to the N observations simulated by the sample reordering algorithm. Given its wide use by insurance regulators, we also report the VaR estimates for the three commonly used confidence levels below.

2.4. The Data

The data on general insurance are obtained from the Australian Prudential Regulation Authority (APRA) (https://www.apra.gov.au/, accessed on 21 February 2025), as also used by Tang and Valdez (2006). However, we use a more recent time period and quarterly data instead of annual data to increase the sample size. Australia has a large market share

in the insurance industry within developed countries. Based on the data published by the OECD (see OECD 2017), Australia's general insurance is above the 70th percentile in terms of total gross premiums in 2016. In September 2010, a change in the reporting format was introduced, so the definitions of some variables used are also modified. To avoid inconsistencies, we focus on the period from September 2010 to June 2017. We are interested in four variables: gross incurred claims (including movements in outstanding claims liability during the period); gross earned premium; net incurred claims (net of reinsurance recoveries revenue); and net earned premium (net of outwards reinsurance expense). We consider both the gross and the net variables as one of our goals to evaluate the effect of reinsurance on capital requirements. As in MunichRe (2010) and Cipra (2010), reinsurance is a mechanism used by an insurance company (the reinsured, cedent, or primary insurance company) to transfer all or part of its unforeseen or extraordinary losses under a policy or policies that it has issued to another insurance company (the reinsurer). To indemnify the reinsurer, a premium is paid to the reinsurer by the ceding company. We source data for five insurance business lines, namely, domestic motor vehicle (hereafter referred to as Motor), houseowners/households (House), fire and ISR[3] (Fire), liability, and compulsory third-party motor vehicle (CTP). According to the data collected from the APRA webpage, these five business lines made up more than 85% of the Australian general insurance market in terms of net earned premiums as at June 2017. In the process of cleaning the data, we removed the observations from two quarters where there were two negative observations of gross incurred claims. Our final dataset has 26 observations for each business line.

Loss Ratios

To quantify the insurance risk, we use loss ratio (LR)[4], defined as

$$LR_{i,t} = \frac{IC_{i,t}}{EP_{i,t}}.$$

The numerator $IC_{i,t}$ denotes the incurred claims corresponding to the earned premium $EP_{i,t}$ (the denominator) for business line i at time t, based on the accident year insurance company's accounting principal; see Taylor (1997) for details on the loss ratio variable. The loss ratio can be seen as a measure of claims standardised by the risk exposure (given by the earned premium). Using loss ratios can eliminate temporal effects of business growth and inflation, allowing for comparisons between business lines with different risk exposures. Individual loss ratios are added up to form the aggregated loss ratio for capital requirement estimation.

The aggregate loss ratio at time t, LR_t, can then be written as the weighted sum of the individual loss ratios of the d business lines as

$$\begin{aligned} LR_t &= \frac{IC_t}{EP_t} \\ &= \frac{\sum_{i=1}^{d} IC_{i,t}}{\sum_{i=1}^{d} EP_{i,t}} \\ &= \frac{\sum_{i=1}^{d} \left(\frac{IC_{i,t}}{EP_{i,t}} \times EP_{i,t} \right)}{\sum_{i=1}^{d} EP_{i,t}} \\ &= \sum_{i=1}^{d} LR_{i,t} \times \frac{EP_{i,t}}{\sum_{i=1}^{d} EP_{i,t}} \\ &= \sum_{i=1}^{d} LR_{i,t} \times w_{i,t}, \end{aligned} \tag{2}$$

where IC_t and EP_t are the incurred claims and earned premium aggregated across all business lines, and $w_{i,t}$ is the weight of earned premiums for business line i in period t. Below we will also examine gross loss ratios and net (after reinsurance) loss ratios compared to the total earned premiums across all business lines. The gross loss ratio is the ratio of gross claims to gross premiums while the net loss ratio is the ratio of net claims to net premiums.

The descriptive statistics for the five business lines' loss ratios are summarised in Table 1. The 'Aggregate loss' column contains the quantities for the aggregate loss ratio, which are calculated as in Equation (2). From Table 1, we observe that for all the business lines, the average loss ratios—both gross and net of reinsurance—are not statistically different. Although reinsurance is essentially a risk transfer (or sharing) tool, loss distributions tend to be positively skewed and hence we would expect the average loss ratio to reduce from gross to net of reinsurance. However, reinsurance seems to have no strong effect on the average loss ratio. We explore later in the paper how this may result from the interplay between the premium ceded to and claim recoveries from reinsurance. The standard deviation is higher for Fire. However, while it decreases for House, Motor, and especially Fire, it actually increases for CTP and Liability when reinsurance is taken into account. The values estimated for skewness show that the loss ratios for House and Fire do not have symmetric distributions. There is also significant excess kurtosis of the loss ratios for House and Fire, with both reducing with reinsurance. In terms of the aggregate loss ratio, reinsurance has a larger effect on the skewness and kurtosis than on the mean and standard deviation of the loss ratio. Most notably, reinsurance reduces the excess kurtosis of the aggregate loss ratio by 74%.

Table 1. Summary statistics of the loss ratios for the period from September 2010 to June 2017.

	House	**Fire**	**Motor**	**CTP**	**Liability**	**Aggregate Loss**
Gross loss ratios						
Mean	0.5849	0.7820	0.7211	0.8172	0.7024	0.7005
Standard deviation	0.2981	0.8334	0.0682	0.3100	0.1566	0.1971
Skewness	2.6290	3.6449	0.9729	−0.7432	−0.2392	2.8759
Excess kurtosis	8.0694	13.819	0.0075	0.0036	0.0671	9.6254
Average weight, $\bar{w}_{i,t}$	0.25	0.14	0.33	0.11	0.18	1
Weight at June 2017, $w_{i,T}$	0.26	0.12	0.33	0.13	0.16	1
Net loss ratios						
Mean	0.6272	0.6549	0.7394	0.8051	0.6499	0.7018
Standard deviation	0.2105	0.2639	0.0454	0.3333	0.1907	0.1659
Skewness	2.0440	1.4870	0.3835	−0.8458	−0.6556	1.3425
Excess kurtosis	5.6319	2.2074	−0.9542	0.0960	1.5980	2.4629
Average weight, $\bar{w}_{i,t}$	0.22	0.10	0.36	0.13	0.18	1
Weight at June 2017, $w_{i,T}$	0.24	0.09	0.36	0.13	0.17	1

3. Estimation of the Hierarchical Aggregation Copula Model

In this section, we implement the estimation of the hierarchical copula model for the aggregate loss from the individual business lines as presented in Section 2.2.

3.1. Tree Structure of the Hierarchical Copula Model

The first element of the hierarchical copula model is the rooted tree τ associated with random variables X_i, representing the loss ratios for individual business lines. As explained in Section 2.2, to build the tree, we start by allocating the loss ratio of each business line to one leaf node and then aggregate the two random loss ratios, with the highest dependence measured by Kendall's tau. Table 2 shows the Kendall's tau estimates for the gross loss

ratios of each pair of business lines. At each stage, we aggregate the two loss ratio random variables with the strongest Kendall's tau estimate.

Table 2. Sequential aggregation of the gross loss ratios for the five business lines.

		Stage 1		
	House	Fire	Motor	CTP
Fire	**0.5262**	1	–	–
Motor	0.4338	0.2308	1	–
CTP	0.0154	−0.0523	−0.1815	1
Liability	0.0585	−0.1323	0.1446	0.3662
		Stage 2		
		House + Fire	Motor	CTP
Motor		0.3169	1	–
CTP		−0.0400	−0.1815	1
Liability		−0.0338	0.1446	**0.3662**
		Stage 3		
		House + Fire	Motor	
Motor		**0.3169**	1	
CTP+Liability		0.0154	−0.0523	

After allocating each business line to a leaf node, as in the bottom row of the tree depicted in Figure 2, we aggregate the two business lines with the strongest dependence. From Table 2, we observe that House and Fire have the largest Kendall's tau. Hence, at this first stage, we aggregate these two business lines. In Stage 2, the largest Kendall's tau observed is between CTP and Liability. We then aggregate CTP and Liability. In Stage 3, the strongest dependence is between Motor and Fire + House, leading to the aggregate between them. In the final stage, we aggregate together the two resulting loss ratios, namely Motor + Fire + House and Liability + CTP. This is illustrated in Figure 2.

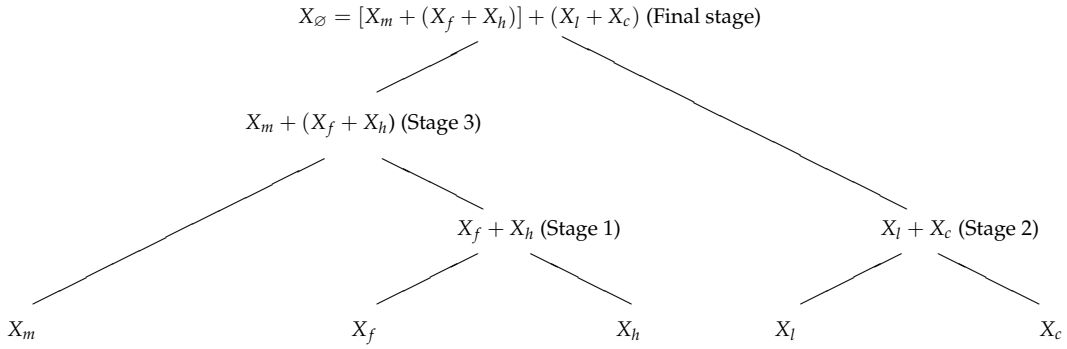

Figure 2. Hierarchical loss aggregation copula model for the gross (and net) loss ratio of the the five business lines, Motor, Fire, House, Liability and CTP, represented by X_m, X_f, X_h, X_l, and X_c, respectively. The structure of the tree is determined by iteratively aggregating the two nodes with the strongest dependence.

Table 3 contains the Kendall's tau values for the case of the net (after reinsurance) loss ratio for the five business lines. The more strongly dependent variables, at the different stages of construction of the tree, are the same as for the gross loss ratios case. As a consequence, the structure of the rooted tree for the net loss ratios hierarchical copula model is the same as for the gross loss ratios shown in Figure 2. In the last stage of the aggregation model, the Kendall's tau between the net loss ratios for House, Fire, Motor, CTP, and Liability together is −0.0892.

Table 3. Sequential aggregation of the net loss ratios for the five business lines.

	Stage 1			
	House	Fire	Motor	CTP
Fire	**0.5446**	1	–	–
Motor	0.4338	0.2492	1	–
CTP	−0.0154	−0.0031	−0.2369	1
Liability	0.0092	−0.0646	−0.0523	0.4954
	Stage 2			
	House + Fire		Motor	CTP
Motor	0.3969		1	–
CTP	−0.0400		−0.2369	1
Liability	−0.0523		−0.0523	**0.4954**
	Stage 3			
	House + Fire		Motor	
Motor	**0.3969**		1	
CTP + Liability	−0.0523		−0.2123	

3.2. Fitting the Univariate Probability Distributions

Next, we will fit the probability distributions for individual loss ratios, based on the maximum likelihood estimation method and Anderson and Darling's (A-D) goodness of fit test. As we are primarily interested in estimating the measures of risk, which are based on the tail of the distributions, it is important to use an appropriate test. It is known that the A-D test is more powerful and sensitive to the tails of the distribution (see Engmann and Cousineau 2011) than other alternative tests such as the commonly used Kolmogorov–Smirnov (Kolmogorov 1933; Smirnov 1948) goodness of fit test. Hence, we choose the distribution that produces the highest p-value according to the A-D test. For each business line, we fit the following families of distributions: lognormal, gamma, Weibull, log-logistic, Pareto, and Burr. The results for the distribution with the highest A-D test p-value and corresponding parameter estimates are listed in Table 4. The fitted distributions for the loss ratios are log-logistic, Burr, and Weibull distributions. These fitted distributions are also visualised in Figures 3 and 4 for gross and net loss ratios, respectively.

Table 4. Family of distributions selected for each business line's gross and net loss ratios. The parameter and corresponding standard error estimates are listed for each business line together with the Anderson and Darling (A-D) statistic and p-value. For the purpose of comparison, the table also has the estimates for the aggregate loss ratio with the weights fixed as at June 2017. * In the case of the Burr distribution, the value listed in the table as being the scale is in fact the estimate for the rate, which is 1/scale.

	House	**Fire**	**Motor**	**CTP**	**Liability**	**Aggregate Loss**
			Gross loss ratios			
Distribution	Log-logistic	Burr	Burr	Weibull	Burr	Burr
Shape 1	4.76266	0.19159	0.04799	3.00527	7.70166	0.3732
(s.e.)	(0.776)	(0.122)	(0.042)	(0.505)	(22.63)	(0.199)
Shape 2	–	8.11427	189.928	–	5.64960	15.8580
(s.e.)	–	(4.012)	(155.0)	–	(1.555)	(5.441)
Scale *	0.52243	3.04747	1.55319	0.90936	0.92955	1.70254
(s.e.)	(0.037)	(0.415)	(0.014)	(0.061)	(0.604)	(0.095)
A-D statistic	0.294	0.147	0.335	1.417	0.270	0.230
A-D p-value	0.942	0.998	0.909	0.197	0.958	0.979

Table 4. *Cont.*

	House	Fire	Motor	CTP	Liability	Aggregate Loss
			Net loss ratios			
Distribution	Log-logistic	Log-logistic	Log-logistic	Weibull	Weibull	Burr
Shape 1	6.37499	4.96750	27.9840	2.53352	3.87399	0.50244
(s.e.)	(1.031)	(0.801)	(4.469)	(0.439)	(0.599)	(0.269)
Shape 2	–	–	–	–	–	18.4406
(s.e.)	–	–	–	–	–	(5.898)
Scale *	0.59180	0.59840	0.73616	0.89199	0.71298	1.55857
(s.e.)	(0.031)	(0.041)	(0.009)	(0.071)	(0.037)	(0.073)
A-D statistic	0.246	0.455	0.371	1.962	0.602	0.197
A-D *p*-value	0.971	0.791	0.875	0.097	0.643	0.991

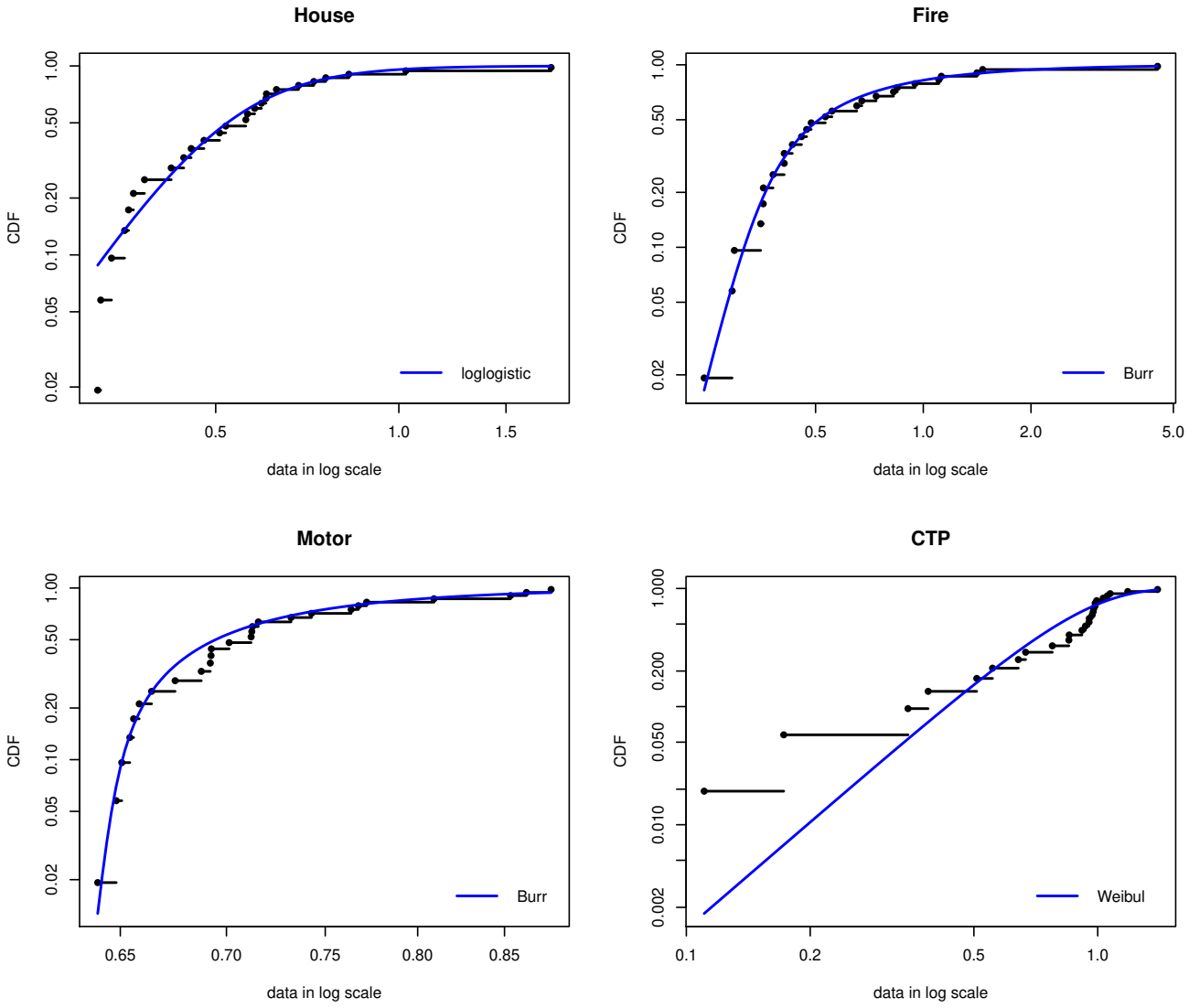

Figure 3. *Cont.*

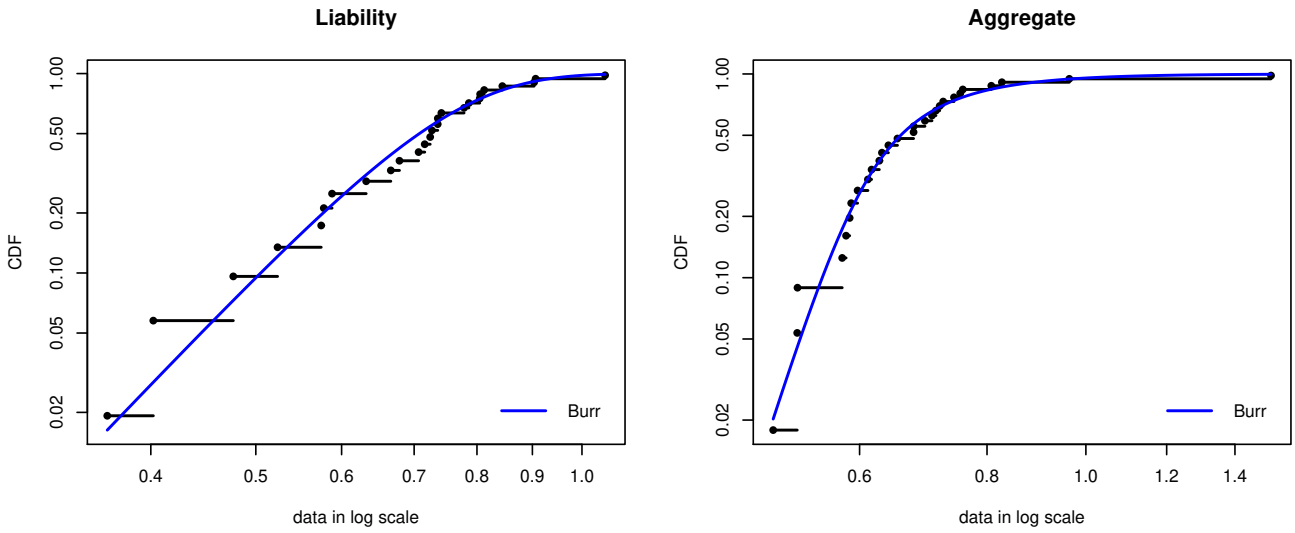

Figure 3. Fitted probability distributions (in blue) vs. observed cumulative distribution functions (CDF) for the gross loss ratios.

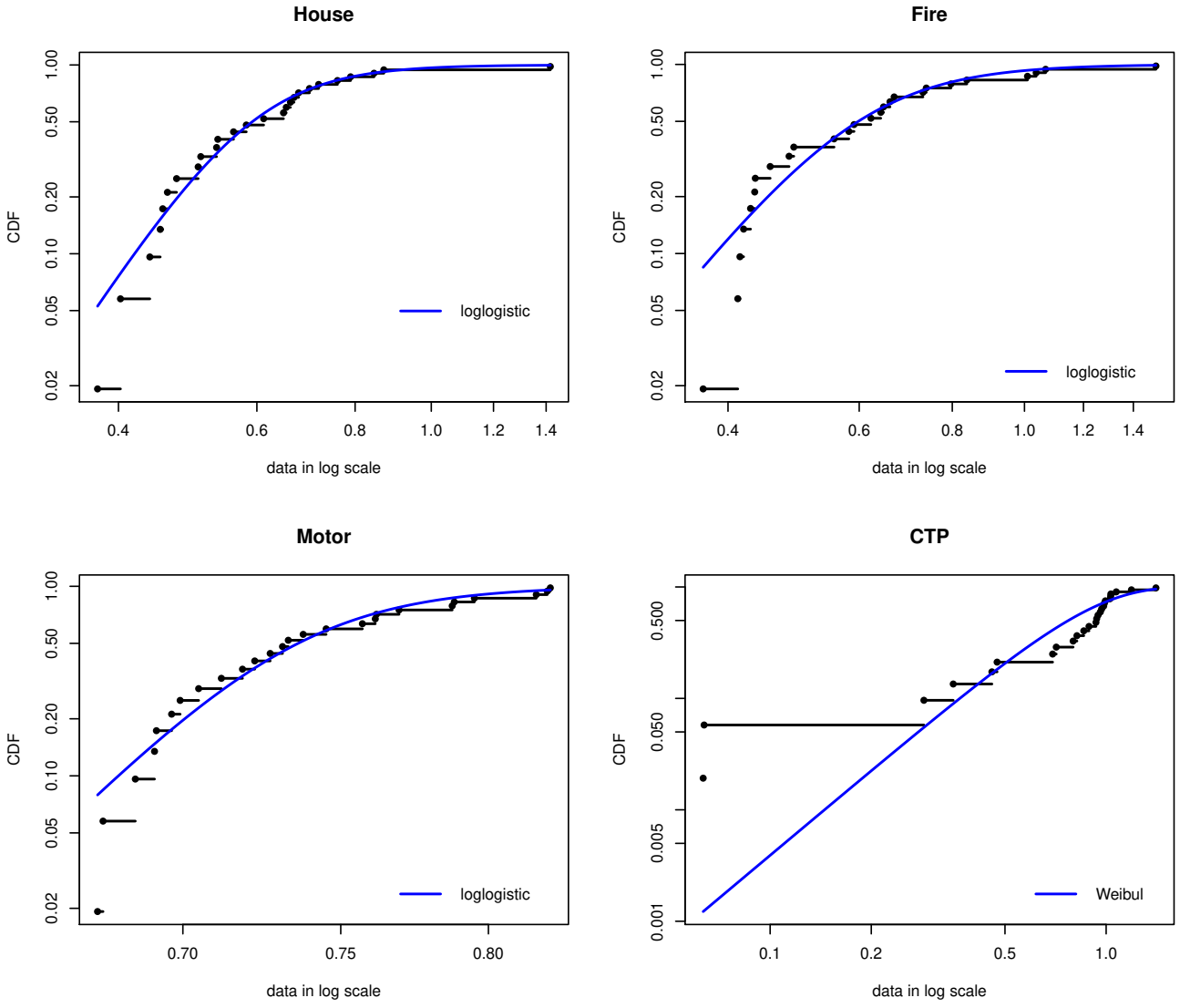

Figure 4. *Cont*.

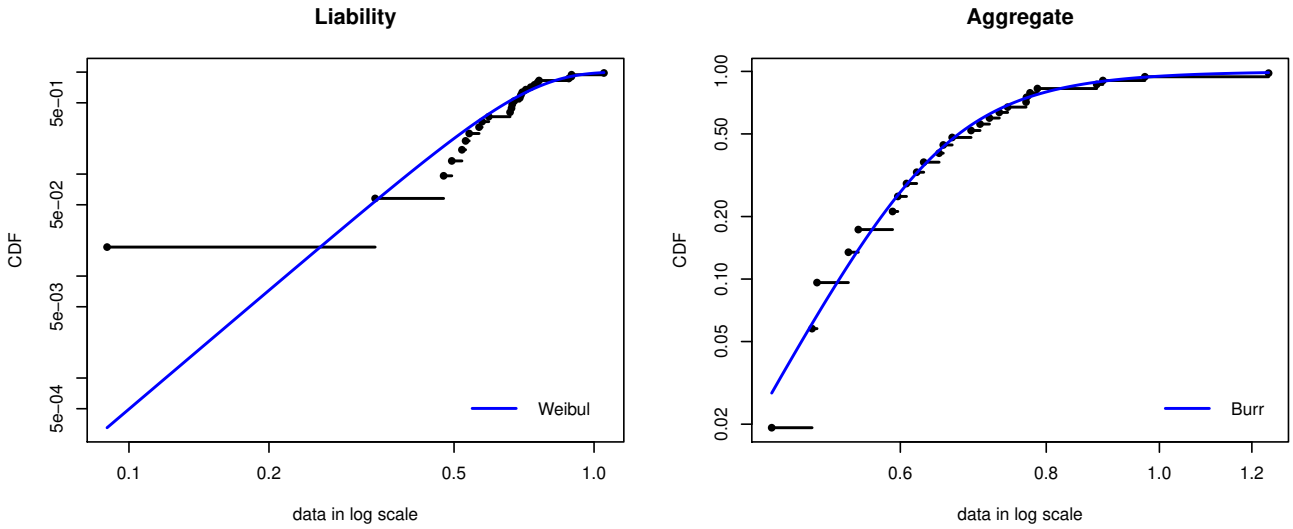

Figure 4. Fitted probability distributions (in blue) vs. observed cumulative distribution functions (CDF) for the net loss ratios.

3.3. Determining Joint Distribution Through Copulas

Having determined the best fit univariate distribution for each business loss ratio, we can estimate the joint distributions for each pair of loss ratios at each branching node in Figure 2 by coupling the corresponding univariate probability distributions. We consider the following commonly used copulas: Gaussian, Student-t, Frank, Clayton, Gumbel, and the mixtures of Clayton and Gumbel copulas and their corresponding survival copulas. The Gaussian copula is appropriate for modelling data with an elliptical shape but does not allow for tail dependence between the business lines. The Student-t copula, although also elliptical, allows for tail dependence. The Frank copula is suitable for symmetric data and can perform better than the Gaussian copula in certain cases. The Clayton and the Gumbel copulas are used when data show asymmetric tail dependence (lower or upper, respectively). Mixtures of Clayton and Gumbel copulas and their corresponding survival copulas can potentially perform better when the data exhibit tail dependence in both tails, even if asymmetric.

In Table 5, we calculate and report the non-parametric estimates of the upper and lower tail dependence coefficients (see Sibuya 1960; Schmid and Schmidt 2007) for each pair of loss ratios associated with the children of each branching node. As the risk of extreme events is one of the main concerns when it comes to capital requirements, it is important to pay particular attention to the tails of the copula distributions in the modelling process. Table 5 summarizes the results of the selected copulas for the four branching nodes for both gross and net loss ratios.

To select between copula models, we use the goodness of fit test statistic S_n from (Genest et al. 2009). In the cases of the mixture models, to determine the weight of each copula family, we fit a model with weight α for the first copula and $1-\alpha$ for the second copula, for a sequence of values α varying in $[0,1]$. We then choose the estimate $\hat{\alpha}$ for the mixture that maximises the goodness of fit statistic. For the gross loss ratios, the first node is Fire and House ($X_f + X_h$), as shown in Figure 2. From Table 5, we can see that both the lower (λ_L) and upper (λ_U) tail coefficient estimates are different from zero. The copula with the highest p-value (using the goodness of fit test statistic S_n from is a mixture of 40% Clayton and 60% survival Clayton copulas. As the Clayton copula allows for tail dependence, and the estimates for the upper and lower tail dependence are slightly different, the mixture model seems to be a reasonable choice. The p-value of the S_n goodness of fit test, the parameters, and the standard errors estimates are also listed in Table 5. For the CTP and

Liability business lines, the best copula is a mixture of 25% Clayton and 75% survival Clayton copulas. The same copula mixture is again the best for Motor and Fire plus House, but with only 10% weight on the Clayton component of the mixture. The estimates for the tail coefficients for the two root node children, Motor plus Fire plus House and CTP plus Liability, are zero. Indeed, the best copula, according to the goodness of fit test, is the Gaussian copula, which has no tail dependence. This choice makes sense, as the estimates for the upper and lower tail dependence are zero.

Table 5. Upper (λ_U) and lower (λ_L) tail coefficient non-parametric estimates for the pairs of children of each branching node of the copula hierarchical model tree. The best-fitting copula, corresponding goodness of fit test p-value, and parameter estimates (with standard errors in parenthesis) are also listed. For the mixture copulas, θ_1 is the parameter estimate of the first component of the mixture and θ_2 corresponds to the second component of the mixture. For the last pair of net loss ratios, $(X_m + X_f + X_h, X_c + X_l)$, λ_L measures the tail coefficient in the second quadrant of the sample space and λ_U measures the tail coefficient in the fourth quadrant.

	λ_L	λ_U	**Copula**	p**-Value**	$\hat{\theta}_1$ **(s.e.)**	$\hat{\theta}_2$ **(s.e.)**
Gross loss ratios						
(X_h, X_f)	0.5218	0.5694	0.4 Clayton + 0.6 SurvClayton	0.4640	4.886 (4.161)	2.148 (1.966)
(X_c, X_l)	0.1496	0.2742	0.25 Clayton + 0.75 SurvClayton	0.5410	1.022 (3.194)	1.482 (1.596)
$(X_m, X_f + X_h)$	0.2772	0.4383	0.1 Clayton + 0.9 SurvClayton	0.8986	1.160 (5.796)	1.029 (0.548)
$(X_m + X_f + X_h, X_c + X_l)$	0.0000	0.0000	Gaussian	0.9815	0.013036 (0.285)	
Net loss ratios						
(X_h, X_f)	0.5390	0.5401	0.6 Gumbel + 0.4 SurvGumbel	0.7298	2.126 (1.265)	2.801 (2.083)
(X_c, X_l)	0.2772	0.1070	Student-t	0.5549	0.7376 (0.115)	1.2910 (0.593)
$(X_m, X_f + X_h)$	0.3977	0.4038	0.7 SurvGumbel + 0.3 SurvClayton	0.7607	1.750 (0.954)	1.047 (2.884)
$(X_m + X_f + X_h, X_c + X_l)$	0.0143	0.1531	90° Rotated Gumbel	0.5569	1.0865 (0.186)	

For the net loss ratios, the best copula for the House and Fire pair is 60% Gumbel plus 40% survival Gumbel. The resulting copula has both upper and lower tail dependence, which is in line with the non-parametric estimates. The CTP and Liability pair is best modelled by a Student-t copula, which also allows for both upper and lower tail dependence. A mixture of 70% survival Gumbel with 30% survival Clayton has the higher p-value for Motor and Fire plus House. The estimate for the Kendall's tau for the Motor plus Fire plus House and the CTP plus Liability pairs is close to zero but negative. Hence, we flip the Motor plus Fire plus House variable after transforming it into the zero-one interval. By flipping, we mean subtracting the variable from one. The best copula for the resulting pair is then a Gumbel copula, which allows for tail dependence between low values of Motor plus House plus Fire and high values of CTP and Liability.

3.4. Simulation of the Aggregate Loss Ratios

In order to estimate VaR and TVaR from the hierarchical copula aggregation model, we can now simulate observations of aggregate loss ratios using the model constructed in the previous sections. We implement the sample reordering algorithm from Section 2.2.2 for the gross and net loss ratios using $N = 1000$. Using the estimator from Equation (1), we

estimate the TVaR for each business line's gross and net loss ratios for the confidence levels of 90%, 95%, and 99%. The VaR estimate for a given confidence level is the corresponding empirical quantile. The results are presented in Table 6, and their analysis follows in the next sections.

Table 6. VaR and TVaR estimates for the five business lines. The values in square brackets are 95% confidence intervals. The column labelled 'Weighted Sum of Risk Measures' corresponds to the weighted sum of the risk measures (VaR or TVaR) from each business line with weights as at June 2017. The column labelled 'Risk measure of Aggregate Loss' has the values obtained from the hierarchical aggregation copula model with weights for each business line as at June 2017.

	House	**Fire**	**Motor**	**CTP**	**Liability**	**Weighted Sum of Risk Measures**	**Risk Measure of Aggregate Loss, LR_t**
				Gross loss ratios			
90% VaR	0.8284	1.4422	0.8283	1.1991	0.8915	0.9603	0.8806
	[0.800,0.856]	[1.301,1.60]	[0.814,0.843]	[1.172,1.225]	[0.879,0.903]	[0.940,0.981]	[0.859,0.902]
95% VaR	0.9693	2.2516	0.8931	1.3088	0.9417	1.1377	1.0184
	[0.925,1.024]	[1.959,2.593]	[0.872,0.916]	[1.278,1.341]	[0.926,0.957]	[1.099,1.182]	[0.979,1.064]
99% VaR	1.365	6.2017	1.0602	1.5049	1.0346	1.8101	1.5937
	[1.227,1.534]	[4.463,8.642]	[1.008,1.122]	[1.451,1.56]	[1.008,1.061]	[1.603,2.095]	[1.385,1.891]
90% TVaR	1.063	4.1271	0.9299	1.3413	0.9576	1.4060	1.2644
	[1.007,1.128]	[2.873,6.202]	[0.906,0.957]	[1.313,1.37]	[0.944,0.972]	[1.256,1.652]	[1.118,1.518]
95% TVaR	1.2353	6.4776	1.0026	1.4322	1.0003	1.7755	1.5895
	[1.144,1.341]	[4.096,10.578]	[0.966,1.042]	[1.397,1.466]	[0.983,1.019]	[1.488,2.276]	[1.304,2.094]
99% TVaR	1.7244	18.4861	1.1898	1.6042	1.0836	3.4412	3.1437
	[1.459,2.074]	[8.037,37.647]	[1.1,1.299]	[1.54,1.669]	[1.051,1.118]	[2.178,5.761]	[1.897,5.461]
				Net loss ratios			
90% VaR	0.835	0.9313	0.7961	1.2386	0.8843	0.8821	0.801
	[0.813,0.857]	[0.9,0.965]	[0.791,0.801]	[1.207,1.273]	[0.869,0.899]	[0.874,0.89]	[0.792,0.81]
95% VaR	0.9383	1.0821	0.8177	1.3737	0.9462	0.9563	0.844
	[0.904,0.973]	[1.033,1.134]	[0.811,0.825]	[1.334,1.414]	[0.927,0.965]	[0.945,0.967]	[0.832,0.857]
99% VaR	1.2087	1.4985	0.8662	1.6234	1.0549	1.1271	0.9443
	[1.124,1.311]	[1.366,1.668]	[0.851,0.883]	[1.56,1.693]	[1.026,1.083]	[1.101,1.156]	[0.916,0.976]
90% TVaR	0.9987	1.1803	0.8273	1.4158	0.9638	0.9916	0.8651
	[0.959,1.04]	[1.121,1.246]	[0.821,0.835]	[1.379,1.453]	[0.948,0.981]	[0.979,1.004]	[0.853,0.878]
95% TVaR	1.1164	1.3622	0.8487	1.5303	1.0144	1.0674	0.9098
	[1.055,1.183]	[1.266,1.468]	[0.839,0.859]	[1.483,1.579]	[0.995,1.034]	[1.049,1.087]	[0.891,0.93]
99% TVaR	1.4311	1.8778	0.8983	1.7503	1.1077	1.2516	1.021
	[1.27,1.636]	[1.605,2.216]	[0.876,0.924]	[1.664,1.837]	[1.074,1.146]	[1.201,1.312]	[0.976,1.075]

Analysis of the Results

From Table 6, we can see that Fire has the largest VaR and TVaR among the five business lines for the gross loss ratios, followed by CTP, except for the 99% TVaR, where House has the second largest. When we consider reinsurance, by analysing the net loss ratios, CTP has the largest risk measure values while Fire has the second largest, except for the 99% TVaR, where Fire still has the largest value. Nevertheless, the 99% TVaR for Fire shows a staggering reduction after reinsurance. Overall, Motor has the lowest values of risk measures in terms of both gross and net loss ratios, implying that it is the least risky business line. The VaR and TVaR 95% confidence intervals for gross and net losses overlap in the cases of House, CTP, and Liability. For Fire, Motor, and (copula) aggregate losses, the confidence intervals for gross and net losses do not overlap. We conclude that reinsurance

is effectively reducing the level of risk only for Fire and Motor, and that this reduction is strong enough to carry over to the (copula) aggregate loss. The effect of reinsurance in changing the risk level for House, CTP, and Liability is much less pronounced. We come back to this point later in this article. It is worthwhile to recall here that the average loss is also not significantly different both with and without reinsurance.

Comparing the two right columns of Table 6, we can see that the weighted sum of the risk measures, VaR and TVaR, is larger than the value obtained using the hierarchical aggregation copula model. This is true both for VaR and TVaR at all the probability levels considered and for gross and net loss ratios. The risk measures obtained using the hierarchical copula model incorporate the dependence between the different business lines while the weighted sum of VaR and TVaR does not. Hence, the result obtained is clear evidence that there is a risk reduction effect in the tails when combining the five business lines. This reduction in risk by pooling different business lines (risks) corresponds to the well-known notion of diversification in financial portfolio selection and asset allocation.

4. The Effect of Reinsurance

Our goal now is to explore the effect of reinsurance on the diversification of the portfolio composed of the different business lines. Conceptually, we draw some parallel here between a portfolio of financial assets and the set of business lines. When addressing diversification in terms of portfolio selection, we can think of two aspects: First, diversification is affected by the weights of each component of the portfolio. Second, diversification can also be affected by the sources of risk and the interaction between the different business lines. We address these two cases separately below.

4.1. Reinsurance and Weighted Premiums Diversification

Here we evaluate the effect of reinsurance on diversification due to changing the proportion of underwritten premiums (weights) for the different business lines. Insurance companies cede risk to the reinsurer in different proportions for the different business lines. As a result, the weights of each business line—corresponding to the proportion of underwritten premiums—in the insurers portfolio before and after considering reinsurance are different. For the data analysed in this paper, the weights as at June 2017 are reported in Table 1; we can see that reinsurance reduces the proportion of the business lines of Fire and House, and increases the weight of Motor.

A measure of diversification which concentrates on the weights of each portfolio component is derived from the concept of Shannon's entropy, introduced in Shannon (1948) for information theory. Within a financial portfolio setting, Shannon's entropy measures diversification as

$$H(w) = -\sum_{i=1}^{N} w_i \ln w_i,$$

where $\sum_{i=1}^{N} w_i = 1$, $w_i \geq 0$ and N is the number of portfolio components. According to this measure, equal weights correspond to the highest diversification. The background idea is that equal weights correspond to maximum information. We refer to DeMiguel et al. (2009) for a study on the (superior) out-of-sample performance of an equally weighted financial portfolio.

The values obtained for the Shannon's entropy measure for the insurance data are listed in Table 7. We find that the diversification of the portfolio considering reinsurance is lower than the diversification of the portfolio without reinsurance. The change in the weights between the business lines is largely due to the higher cession rate on Fire. We can see that there is a link here between a higher cession rate (mainly) on the Fire business line through reinsurance and a reduction in the diversification of the portfolio. The value

for the Shannon's entropy of an equally weighted portfolio is 1.61. Therefore, the equally weighted portfolio is 6% more diversified than the portfolio without reinsurance and 8% more diversified than the portfolio after considering reinsurance.

Table 7. Shannon's entropy measure of diversification for the insurance portfolio of the five business lines using the weights as at June 2017.

	House	Fire	Motor	CTP	Liability	Shannon's Entropy
Gross loss ratio weights	0.26	0.12	0.33	0.13	0.16	**1.52**
Net loss ratio weights	0.24	0.09	0.36	0.13	0.17	**1.49**

4.2. Reinsurance and Source of Risk Diversification

One goal of diversification is to reduce the risk in the portfolio by taking advantage of the relation between the different components. One way of measuring portfolio diversification while taking the sources of risk into account is by calculating the diversification ratio (DR) from Choueifaty and Coignard (2008). This measure uses both the weights and the risk of each component of the portfolio, producing a weighted average of the components' risk. The expression for the diversification ratio is given by

$$DR = \frac{\sum_{i=1}^{N} w_i \lambda_i}{\lambda_P},$$

where $\sum_{i=1}^{N} w_i = 1$, $w_i \geq 0$, λ_i is the risk of component i, and N is the number of portfolio components. In the denominator of the diversification ratio, λ_P is the portfolio risk, and hence the relation between the different components of the portfolio is taken into account by the diversification ratio. Using standard deviation, with VaR and TVaR as measures of risk, we obtain the diversification ratio values. The VaR and TVaR risk measures for the weighted sum of business lines' loss ratios are the ones obtained by the hierarchical aggregation copula model. The weights are fixed and based on the premiums as at June 2017. The results are reported in Table 8.

Table 8. Weighted sum of risk measures for the five business lines compared with the risk measure of the weighted sum of business lines obtained by the hierarchical copula model. The last column of the table gives the values for the diversification ratio from Choueifaty and Coignard (2008).

	Weighted Sum of Risk Measures	Risk Measure of Aggregate Loss, LR_t	DR
Gross loss ratios			
σ	0.2654	0.1956	1.35
90% VaR	0.9603	0.8806	1.09
95% VaR	1.1377	1.0184	1.12
99% VaR	1.8101	1.5937	1.14
90% TVaR	1.4060	1.2644	1.11
95% TVaR	1.7755	1.5895	1.12
99% TVaR	3.4412	3.1437	1.09
Net loss ratios			
σ	0.1752	0.1040	1.68
90% VaR	0.8821	0.8010	1.10
95% VaR	0.9563	0.8440	1.13
99% VaR	1.1271	0.9443	1.19
90% TVaR	0.9916	0.8651	1.15
95% TVaR	1.0674	0.9098	1.17
99% TVaR	1.2516	1.0210	1.23

The results show that the diversification ratio increases with reinsurance most strikingly when we use standard deviation as measure of risk, where the diversification ratio increases by 24%. This indicates that reinsurance increases diversification for the smaller, more frequent claims to a larger extent than for the larger, less frequent claims. For the VaR and TVaR, the increase in diversification is modest, at around 3%, with the exception of TVaR, at the 99% level, to which reinsurance implies a 12% increase in the diversification ratio. This indicates that, from a multivariate or portfolio point of view, by reducing the high upper tail dependence of some of the loss ratios across the different business lines (the estimates for λ_U in Table 5 do not contradict this assertion), reinsurance increases the diversification ratio. Considering the results from Shannon's entropy measure, we conclude that, although reinsurance reduces the (weights') diversification, this is compensated by the reduction in risk producing higher diversification ratios. As far as we know, these effects of reinsurance on the multivariate overall portfolio of business lines have not been previously found in the literature.

5. Conclusions

It is important for every insurance company to determine and maintain the right amount of capital to maintain a solvency margin against the risk of not being able to cover the its liabilities. This calls for adequate methods of aggregating all risks and the use of appropriate risk measures to determine the capital requirement. In this article, we use a hierarchical aggregation copula model to address the dependence structure of the different insurance business lines. We use several copula families to model the aggregated loss, with particular emphasis on capturing the tail dependence. We consider a range of copulas: asymmetric, symmetric, with and without tail dependence, Gaussian, Student-t, Archimedean, Clayton, Gumbel, and Frank. Selecting the best copula families for the hierarchical aggregation model is crucial, as it influences the estimated level of risk and consequently avoids over- or underestimation of the capital required. We must stress that this study looks for empirical evidence of the effect of reinsurance on a portfolio of insurance business lines. This means that the methodology used is very much tailored for the particular case and dataset used, and the following conclusions are constrained by our specific application. Our goal is to show, based on the data used, that there are important issues to be considered and show a possible roadmap for other cases.

A very important tool for risk management is reinsurance. Insurance companies diversify away part of their underwriting risk to reinsurance companies. In this paper, we use the case of the Australian insurance market to investigate the effect and relevance of reinsurance on the risk of individual business lines, and particularly on the aggregate risk. These effects are measured in this paper by considering both gross and net loss ratios, where gross loss ratios are used to measure the insurance risk without considering the reinsurance business, while net loss ratios are used to determine the insurance risk while taking reinsurance into account. For most business lines, especially Fire and Motor, reinsurance reduces the risk. However, it can also increase the risk, even when measured by the standard deviation, as we can see in Tables 1 and 6 for the CTP and Liability business lines.

Another aspect of reinsurance pertains to diversification. By leveraging the interdependencies among various business lines, reinsurance enhances the diversification ratio, which takes into account both the weights and the source of risk. Conversely, it diminishes Shannon's entropy diversification, which solely focuses on weights. Consequently, in the context of the Australian insurance market, we deduce that reinsurance can decrease the aggregate risk's sensitivity to alterations in the proportions of different business lines. Therefore, if the objective is to manage risk by adjusting the proportion of underwriting among business

lines, reinsurance may alleviate the decrease in aggregate risk. Hence, a comprehensive risk management strategy must encompass the three facets of weights, interdependencies among business lines, and reinsurance cession rates to effectively mitigate the aggregate risk in the insurance portfolio when the primary insurer transfers risk through reinsurance.

Author Contributions: Conceptualization, A.D., I.I. and A.Z.; methodology, A.D., I.I. and A.Z.; software, A.D. and I.I.; validation, A.D., I.I. and A.Z.; formal analysis, A.D., I.I. and A.Z.; investigation, A.D. and I.I.; data curation, I.I.; writing—original draft preparation, A.D., I.I. and A.Z.; writing—review and editing, A.D., I.I. and A.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Zhang, A. acknowledges financial support from Wenzhou-Kean University 2024 Internal (Faculty/Staff) Start-Up Research Grant [ISRG2024001] and Wenzhou-Kean University 2025 International Collaborative Research Program [ICRPSP2025002].

Data Availability Statement: The data supporting the analyses presented in the paper can be found from the Australian Prudential Regulation Authority (APRA) (https://www.apra.gov.au/).

Conflicts of Interest: The authors report there are no competing interests to declare.

Notes

[1] The symbol $'$ denotes the transpose of vector.

[2]
$$\mathbb{1}\{s_j \leq x\} = \begin{cases} 1, & \text{if } s_i \leq x \\ 0, & \text{otherwise.} \end{cases}$$

[3] ISR stands for industrial special risk.

[4] To simplify notations, we will use X for the LR, unless otherwise stated.

References

Aas, Kjersti, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44: 182–98. [CrossRef]

Acerbi, Carlo, and Dirk Tasche. 2002. On the coherence of expected shortfall. *Journal of Banking & Finance* 26: 1487–503. [CrossRef]

Acerbi, Carlo, Claudio Nordio, and Carlo Sirtori. 2001. Expected shortfall as a tool for financial risk management. *arXiv* arXiv:condmat/0102304.

Adam, Alexandre, Mohamed Houkari, and Jean-Paul Laurent. 2008. Spectral risk measures and portfolio selection. *Journal of Banking & Finance* 32: 1870–82. [CrossRef]

Anderson, Theodore W., and Donald A. Darling. 1954. A test of goodness of fit. *Journal of the American Statistical Association* 49: 765–69. [CrossRef]

Arbenz, Philipp, Christoph Hummel, and Georg Mainik. 2012. Copula based hierarchical risk aggregation through sample reordering. *Insurance: Mathematics and Economics* 51: 122–33. [CrossRef]

Bargès, Mathieu, Hélène Cossette, and Etienne Marceau. 2009. Tvar-based capital allocation with copulas. *Insurance: Mathematics and Economics* 45: 348–61. [CrossRef]

Baur, Patrizia, Antoinette Breutel-O'Donoghue, and Thomas Hess. 2004. *Understanding Reinsurance: How Reinsurers Create Value and Manage Risk*. Zürich: Swiss Reinsurance Company.

Becker, Bo, Marcus M. Opp, and Farzad Saidi. 2022. Regulatory forbearance in the us insurance industry: The effects of removing capital requirements for an asset class. *The Review of Financial Studies* 35: 5438–82. [CrossRef]

Brechmann, Eike Christian. 2014. Hierarchical Kendall copulas: Properties and inference. *The Canadian Journal of Statistics* 42: 78–108. [CrossRef]

Bürgi, Roland, Michel M. Dacorogna, and Roger Iles. 2008. Risk aggregation, dependence structure and diversification benefit, Stress Testing for Financial Institutions. Available online: https://ssrn.com/abstract=1468526 (accessed on 21 February 2025).

Chaoubi, Ihsan, Hélène Cossette, Etienne Marceau, and Christian Y. Robert. 2021. Hierarchical copulas with archimedean blocks and asymmetric between-block pairs. *Computational Statistics & Data Analysis* 154: 107071. [CrossRef]

Choueifaty, Yves, and Yves L. Coignard. 2008. Toward maximum diversification. *Journal of Portfolio Management* 35: 4051.

Cipra, Tomas. 2010. *Financial and Insurance Formulas*. Zeitschrift: Physica.

Clemente, Gian Paolo, Nino Savelli, and Diego Zappa. 2015. The impact of reinsurance strategies on capital requirements for premium risk in insurance. *Risks* 3: 164–82. [CrossRef]

Cossette, Hélène, Simon-Pierre Gadoury, Étienne Marceau, and Itre Mtalai. 2017. Hierarchical archimedean copulas through multivariate compound distributions. *Insurance: Mathematics and Economics* 76: 1–13. [CrossRef]

Côté, M.-P., and Christian Genest. 2015. A copula-based risk aggregation model. *The Canadian Journal of Statistics* 43: 60–81. [CrossRef]

Cummins, J. David, Georges Dionne, Robert Gagné, and Abdelhakim Nouira. 2008. The Costs and Benefits of Reinsurance. Available online: https://ssrn.com/abstract=1142954 (accessed on 21 February 2025).

Czado, Claudia. 2019. Analyzing Dependent Data with Vine Copulas. In *Lecture Notes in Statistics/Springer*. New York: Springer.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies* 22: 1915–53. [CrossRef]

Diestel, Reinhard. 2017. *Graph Theory*, 5th ed. Heidelberg: Springer.

Embrechts, Paul, Filip Lindskog, and Alexander McNeil. 2003. Modelling dependence with copulas and application to risk management. In *Handbook of Heavy Tailed Distributions in Finance*. Edited by Svetlozar T. Rachev. Amsterdam: Elsevier, chp. 8. pp. 329–384.

Engmann, Sonja, and Denis Cousineau. 2011. Comparing distributions: The two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. *Journal of Applied Quantitative Methods* 6: 1–17.

Floreani, Alberto. 2013. Risk measures and capital requirements: A critique of the solvency ii approach. *The Geneva Papers on Risk and Insurance Issues and Practice* 38: 189–212. [CrossRef]

Genest, Christian, and Johanna Nešlehová. 2012. Copulas and copula models. In *Encyclopedia of Environmetrics*, 2nd ed. Edited by Abdel H. El-Shaarawi and Walter W. Piegorsch. Chichester: Wiley, vol. 2, pp. 541–53.

Genest, Christian, Bruno Rémillard, and David Beaudoin. 2009. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44: 199–214. [CrossRef]

Giuzio, Margherita, Dejan Krušec, Anouk Levels, Ana Sofia Melo, Katri Mikkonen, and Petya Radulova. 2019. Climate change and financial stability. *Financial Stability Review* 1. Available online: https://api.semanticscholar.org/CorpusID:213584241 (accessed on 21 February 2025).

Haff, Ingrid Hobæk, Kjersti Aas, and Arnoldo Frigessi. 2010. On the simplified pair-copula construction—Simply useful or too simplistic? *Journal of Multivariate Analysis* 101: 1296–310. [CrossRef]

Iman, Ronald, and William-Jay Conover. 1982. A distribution-free approach to including rank correlation among input variables. *Communications in Statistics—Simulation and Computation* 11: 311–34. [CrossRef]

Joe, Harry. 1996. Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series* 28: 120–41.

Kolmogorov, Andrey. 1933. Sulla determinazione empirica di una lgge di distribuzione. *Giorn Dell'inst Ital Degli Att* 4: 83–91.

Kurowicka, Dorota, and Harry Joe. 2010. *Dependence Modeling. Vine Copula Handbook*. Singapore: World Scientific.

Li, Pingyun, and Chuancun Yin. 2024. The tail mean–variance optimal capital allocation under the extended skew-elliptical distribution. *Journal of Computational and Applied Mathematics* 448: 115965. [CrossRef]

Mai, Jan-Frederik, and Matthias Scherer. 2012. H-extendible copulas. *Journal of Multivariate Analysis* 110: 151–60. [CrossRef]

McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts. 2015. *Quantitative Risk Management: Concepts, Techniques and Tools*, 2nd ed. Princeton: Princeton University Press.

MunichRe. 2010. *A Basic Guide to Facultative and Treaty Reinsurance*. Princeton: Munich Reinsurance America.

Nanthakumar, Ampalavanar. 2022. A comparison of vine and hierarchical copulas as discriminants. *International Journal of Statistics and Probability* 11: 13–24.

Nelsen, Roger B. 2006. *An Introduction to Copulas*, 2nd ed. New York: Springer.

Nguyen, Tristan, and Robert Danilo Molinari. 2011. Risk aggregation by using copulas in internal models. *Journal of Mathematical Finance* 1: 50. [CrossRef]

OECD. 2017. International comparisons. In *OECD Insurance Statistics 2016*. Paris: OECD Publishing, pp. 45–79.

Schmid, Friedrich, and Rafael Schmidt. 2007. Multivariate conditional versions of spearman's rho and related measures of tail dependence. *Journal of Multivariate Analysis* 98: 1123–40. [CrossRef]

Shannon, Claude Elwood. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27: 379–423. [CrossRef]

Sibuya, Masaaki. 1960. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics* 11: 195–210. [CrossRef]

Sklar, Abe. 1959. Fonctions de répartition à n dimensions et leurs marges. *Annales de l'ISUP* 8: 229–31.

Smirnov, Nickolay. 1948. Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics* 19: 279–81. [CrossRef]

Tang, Andrew, and Emiliano A. Valdez. 2006. Economic capital and the aggregation of risks using copulas. Paper presented at 28th International Congress of Actuaries, Paris, France, October 28; Available online: http://www.ica2006.com/282.html (accessed on 21 February 2025).

Taylor, J. M. 1997. *Claim Reserving Manual*. 1997 rev. ed. London: Faculty and Institute of Actuaries.

Torri, Gabriele, Davide Radi, and Hana Dvořáčková. 2022. Catastrophic and systemic risk in the non-life insurance sector: A micro-structural contagion approach. *Finance Research Letters* 47: 102718. [CrossRef]

Wang, Shaun. 1998. Aggregation of correlated risk portfolios: Models and algorithms. In *Proceedings of the Casualty Actuarial Society*. Arlington: Citeseer, vol. 85, pp. 848–939.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.