UNIVERSITÉ LAVAL

Modélisation de la dépendance et apprentissage automatique dans le contexte du provisionnement individuel et de la solvabilité en actuariat IARD

Thèse

Ihsan Chaoubi

Doctorat en actuariat
Philosophiæ doctor (Ph. D.)

Québec, Canada

Modélisation de la dépendance et apprentissage automatique dans le contexte du provisionnement individuel et de la solvabilité en actuariat IARD

Thèse

Ihsan Chaoubi

Sous la direction de:

Étienne Marceau, directeur de recherche
Hélène Cossette, codirectrice de recherche

Résumé

Les compagnies d'assurance jouent un rôle important dans l'économie des pays en s'impliquant de façon notable dans les marchés boursiers, obligataires et immobiliers, d'où la nécessité de préserver leur solvabilité. Le cycle spécifique de production en assurance amène des défis particuliers aux actuaires et aux gestionnaires de risque dans l'accomplissement de leurs tâches. Dans cette thèse, on a pour but de développer des approches et des algorithmes susceptibles d'aider à résoudre certaines problématiques liées aux opérations de provisionnement et de solvabilité d'une compagnie d'assurance. Les notions préliminaires pour ces contributions sont présentées dans l'introduction de cette thèse.

Les modèles de provisionnement traditionnels sont fondés sur des informations agrégées. Ils ont connu un grand succès, comme en témoigne le nombre important d'articles et documents actuariels connexes. Cependant, en raison de la perte d'informations individuelles des sinistres, ces modèles représentent certaines limites pour fournir des estimations robustes et réalistes dans des contextes succeptibles d'évoluer. Dans ce sens, les modèles de réserve individuels représentent une alternative prometteuse. En s'inspirant des récentes recherches, on propose dans le Chapitre 1 un modèle de réserve individuel basé sur un réseau de neurones récurrent. Notre réseau a l'avantage d'être flexible pour plusieurs structures de base de données détaillés des sinistres et capable d'incorporer plusieurs informations statiques et dynamiques. À travers plusieurs études de cas avec des jeux de données simulés et réels, le réseau proposé est plus performant que le modèle agrégé chain-ladder.

La détermination des exigences de capital pour un portefeuille repose sur une bonne connaissance des distributions marginales ainsi que les structures de dépendance liants les risques individuels. Dans les Chapitres 2 et 3 on s'intéresse à la modélisation de la dépendance et à l'estimation des mesures de risque. Le Chapitre 2 présente une analyse tenant compte des structures de dépendance extrême. Pour un portefeuille à deux risques, on considère en particulier à la dépendance négative extrême (antimonotonocité) qui a été moins étudiée dans la littérature contrairement à la dépendance positive extrême (comonotonocité). On développe des expressions explicites pour des mesures de risque de la somme d'une paire de variables antimontones pour trois familles de distributions. Les expressions explicites obtenues sont très utiles notamment pour quantifier le bénéfice de diversification pour des risques antimonotones.

Face à une problématique avec plusieurs lignes d'affaires, plusieurs chercheurs et praticiens se sont intéressés à la modélisation en ayant recours à la théorie des copules au cours de la dernière décennie. Cette dernière fournit un outil flexible pour modéliser la structure de dépendance entre les variables aléatoires qui peuvent représenter, par exemple, des coûts de sinistres pour des contrats d'assurance. En s'inspirant des récentes recherches, dans le Chapitre 3, on définit une nouvelle famille de copules hiérarchiques. L'approche de construction proposée est basée sur une loi mélange exponentielle multivariée dont le vecteur commun est obtenu par une convolution descendante de variables aléatoires indépendantes. En se basant sur les mesures de corrélation des rangs, on propose un algorithme de détermination de la structure, tandis que l'estimation des paramètres est basée sur une vraisemblance composite. La flexibilité et l'utilité de cette famille de copules est démontrée à travers deux études de cas réelles.

Mots clés : antimonotonocité, copules archimédiennes, copules bivariées asymétriques, copules hiérarchiques, estimation de la vraisemblance composite, mesures de risque, réserve individuelle, réseaux de neurones récurrents, sinistres larges.

Abstract

Insurance companies play an essential role in the countries economy by monopolizing a large part of the stock, bond, and estate markets, which implies the necessity to preserve their solvency and sustainability. However, the particular production cycle of the insurance industry may involve typical problems for actuaries and risk managers. This thesis project aims to develop approaches and algorithms that can help solve some of the reserving and solvency operations problems. The preliminary concepts for these contributions are presented in the introduction of this thesis.

In current reserving practice, we use deterministic and stochastic aggregate methods. These traditional models based on aggregate information have been very successful, as evidenced by many related actuarial articles. However, due to the loss of individual claims information, these models represent some limitations in providing robust and realistic estimates, especially in variable settings. In this context, individual reserve models represent a promising alternative. Based on the recent researches, in Chapter 1, we propose an individual reserve model based on a recurrent neural network. Our network has the advantage of being flexible for several detailed claims datasets structures and incorporating several static and dynamic information. Furthermore, the proposed network outperforms the chain-ladder aggregate model through several case studies with simulated and real datasets.

Determining the capital requirements for a portfolio relies on a good knowledge of the marginal distributions and the dependency structures linking the individual risks. In Chapters 2 and 3, we focus on the dependence modeling component as well as on risk measures. Chapter 2 presents an analysis taking into account extreme dependence structures. For a two-risk portfolio, we are particularly interested in extreme negative dependence (antimonotonicity), which has been less studied in the literature than extreme positive dependence (comonotonicity). We develop explicit expressions for risk measures of the sum of a pair of antimonotonic variables for three families of distributions. The explicit expressions obtained are very useful, e.g., to quantify the diversification benefit for antimonotonic risks.

For a problem with several lines of business, over the last decade, several researchers and practitioners have been interested in modeling using copula theory. The latter provides a flexible tool for modeling the dependence structure between random variables that may represent, for

example, claims costs for insurance contracts. Inspired by some recent researches, in Chapter 3, we define a new family of hierarchical copulas. The proposed construction approach is based on a multivariate exponential mixture distribution whose common vector is obtained by a top-down convolution of independent random variables. A structure determination algorithm is proposed based on rank correlation measures, while the parameter estimation is based on a composite likelihood. The flexibility and usefulness of this family of copulas are demonstrated through two real case studies.

Keywords: Archimedean Copulas, Asymmetric Pair-Copulas, Composite Likelihood Estimation, Counter-monotonicity, Hierarchical Copulas, Individual Claim Reserving, Large claims, Recurrent Neural Networks, Risk Measures.

Table des matières

Liste des tableaux

Liste des illustrations

À mes très chers parents et mari,
cette thèse est aussi la vôtre.

Remerciements

De prime abord, je tiens à exprimer mes sincères remerciements à mes directeurs de recherche, Hélène Cossette et Étienne Marceau, professeurs titulaires à l'École d'actuariat de l'Université Laval. Ce fut vraiment un réel plaisir et une grande chance de travailler avec vous et surtout d'apprendre de vous tout au long de mon projet de thèse. Merci énormément Hélène pour vos conseils, encouragements, disponibilités, soutien et aide très appréciés. Merci beaucoup Étienne pour votre partage généreux de connaissances, directives pertinentes de recherche et encouragements. Vous êtes une mine d'information, je le constatais à chaque échange. Je vous remercie encore une fois d'avoir cru en mes compétences.

Je remercie vivement tous les co-auteurs des articles de cette thèse qui ont considérablement enrichi les travaux de recherche : Camille Besse, Prof. Marie-Pier Côté, Simon Pierre Gadoury et Prof. Christian Y. Robert. Des collaborations avec plusieurs échanges fructueux qui ont beaucoup contribué à ma formation.

J'adresse ma reconnaissance et mes remerciements aux membres du jury, Prof. Arthur Charpentier, Prof. Edward W. Frees et Prof. Montserrat Guillen pour avoir accepté d'évaluer ma thèse ainsi que le temps qu'ils y ont consacré.

Mes remerciements vont également à mes collègues de l'École d'actuariat de l'Université Laval, pour leur bonne humeur et soutien. Je remercie chaleureusement les membres du laboratoire ACT&RISK, en particulier Christopher Blier-Wong, avec qui il fut agréable d'étudier et de s'entraider. Je tiens à remercier, spécialement, le Doyen de la faculté des sciences et de génie M. André Zacarin, de m'avoir octroyé la chance d'être chargée de cours durant ma thèse.

Je souhaite également exprimer ma reconnaissance au Conseil de Recherche en Sciences Naturelles et en Génie du Canada, la Chaire de recherche en actuariat de l'Université Laval, Le Centre de recherche et développement coopérative, Intact Assurance et la Banque Nationale du Canada pour leur soutien financier tout au long de ma thèse. Cela m'a permis de me concentrer à temps plein sur ma recherche.

Je remercie du fond du cœur ma famille et mes amis qui ont cru en moi. Votre soutien, sous différentes formes, m'a permis de progresser dans cette phase d'apprenti-chercheuse.

Un remerciement spécial et chaleureux ira à ma chère mère. Tu as toujours exprimé ton amour sincère à travers tes messages quotidiens, ce qui m'a alimenté en énergie positive tout au long du doctorat. Mon cher papa, je sais que tu es aussi préoccupé par mon cheminement que moi, je te remercie énomément et saches que tu es le modèle que j'ai suivi dans mes études. Ma sœur Ikram, merci du fond du coeur pour ton soutien et tes conseils qui m'ont été d'un grand secours pour accomplir ce projet. Mon frère Ismaël, merci de m'avoir accompagné dans cette expérience à l'Université Laval, très loin de chez nous. Mes frères Abdelmoughit et Mustapha, je vous remercie pour votre encouragements et suivi de mes travaux de recherche. Mon cher Reda, tu as apporté de la joie, de l'espoir et du sourire dans les périodes les plus difficiles.

Mon cher El Mehdi, se marier en plein doctorat n'est pas si évident, mais on a cru en notre capacité à surmonter ce défi ensemble. Je te remercie infiniment pour tes encouragements, ta patience et tes ondes positives. À chaque moment de doute, tu étais là pour redresser un peu, inspirer et rééquilibrer. Merci d'avoir partager avec moi le poids de cette thèse.

Avant-propos

Dans cette thèse, on s'intéresse à plusieurs problématiques liées aux opérations des compagnies d'assurance et des gestionnaires de risques. Particulièrement, on y étudie l'évaluation des réserves au niveau granulaire, des modèles de dépendance bivariée et multivariée hiérarchiques. La thèse est composée d'une introduction et une conclusion générales et trois chapitres constitués intégralement de trois articles scientifiques dont je suis l'auteure principale.

Le Chapitre 1 est basé sur un article co-écrit avec Camille Besse, professionnel de recherche à l'Institut Intelligence et Données, Marie-Pier Côté, et ma directrice de thèse, Hélène Cossette, professeures à l'École d'actuariat de l'Université Laval. L'article s'intitule *Micro-level Reserving for General Insurance Claims using a Long Short-Term Memory Network* et soumis pour publication le 10 août 2021 à la revue *ASTIN Bulletin*. En nous basant sur une technique d'apprentissage automatique profond nommée réseau *Long-Short-Term-Memory*, nous introduisons un nouveau modèle de réserve individuel. Nous étudions également la prévision des réserves pour les sinistres extrêmes en utilisant une distribution de Pareto généralisée pour modéliser l'excès de paiements censurés.

Le Chapitre 2 est constitué de l'article intitulé *On sums of two counter-monotonic risks* co-écrit avec mes directeurs de thèse Hélène Cossette et Étienne Marceau, professeur à l'École d'actuariat, et Simon Pierre Gadoury, auxiliaire de recherche au laboratoire ACT&RISK. Cet article a été publié en mai 2020 dans la revue *Insurance : Mathematics and Economics*. Dans un contexte de risque bivarié, nous nous intéressons à la dépendance négative extrême (antimonotonocité). Pour la perte agrégée d'une paire de risques antimonotones, nous développons des expressions analytiques pour diverses mesures de risque. Cette contribution permet également de quantifier facilement le bénéfice de diversification.

Le troisième article, intitulé *Hierarchical copulas with Archimedean blocks and asymmetric between-block pairs* est présenté au Chapitre 3. Il est co-écrit avec Hélène Cossette, Étienne Marceau et Christian Y. Robert, professeur à l'Université Claude Bernard Lyon 1, et fut publié en février 2021 dans la revue *Computational Statistics & Data Analysis*. Dans cet article, nous définissons une nouvelle famille de copules hiérarchiques se caractérisant par des distributions non échangeables entre les blocs. Nous proposons un algorithme de détermination de la structure accompagné d'une procédure d'estimation des paramètres.

Article associé au Chapitre 1

Chaoubi, I., Besse, C., Cossette, H., & Côté, MP. (2021). *Micro-level Reserving for General Insurance Claims using a Long Short-Term Memory Network*. ASTIN Bulletin (soumis pour publication).

Article associé au Chapitre 2

Chaoubi, I., Cossette, H., Gadoury, S. P., & Marceau, E. (2020). *On sums of two counter-monotonic risks*. Insurance : Mathematics and Economics, 92, 47-60.

Article associé au Chapitre 3

Chaoubi, I., Cossette, H., Marceau, E., & Robert, C. Y. (2021). *Hierarchical copulas with Archimedean blocks and asymmetric between-block pairs*. Computational Statistics & Data Analysis, 154, 107071.

Introduction

Les actuaires et les gestionnaires de risques sont chargés d'assurer une bonne évaluation et une saine gestion quantitative des risques au sein d'un portefeuille d'une compagnie d'assurance. Dans ce projet de thèse, on développe des approches et des algorithmes susceptibles d'aider les actuaires face à certaines problématiques en relation avec les opérations de provisionnement, la solvabilité et l'allocation du capital en assurance IARD (Incendie, Accidents et Risques Divers).

Plus précisément, dans un premier volet nous nous intéressons à l'amélioration de la précision de l'estimation des provisions. Ces derniers constituent la partie majeure du passif d'une compagnie d'assurance. Une estimation juste permet de garder le montant nécessaire pour honorer les engagements futurs de la compagnie et placer le reste des fonds dans des actifs rentables. Dans ce contexte, nous proposons un modèle de réserve individuel basé sur un réseau de neurones récurrent qui permet de prédire le développement futur complet des sinistres encore ouverts à la date d'évaluation.

Le deuxième volet de notre contribution concerne la structure de dépendance des risques. Dans le cas où l'actuaire a des difficultés à identifier la dépendance exacte, une alternative est de considérer les structures de dépendance positive et négative extrême. Pour le cas bivarié, nous nous intéressons, particulièrement, à la dépendance négative extrême, peu étudiée dans la littérature, et nous développons des expressions explicites pour les mesures de risque populaires. Un autre thème dans le contexte de la modélisation de la dépendance porte sur les structures de dépendance hiérarchiques. Nous introduisons une nouvelle famille de copules hiérarchiques caractérisée par des structures non échangeables à l'extérieur des groupes. Nous proposons également des algorithmes d'identification de la hiérarchie et une méthode d'estimation des paramètres en illustrant leur performance à travers plusieurs études de cas.

Le présent chapitre est divisé en cinq sections où nous y introduisons le contexte général des problématiques étudiées dans cette thèse. La première section constitue un aperçu de la dynamique des compagnies d'assurance afin de mieux comprendre le lien entre les travaux élaborés dans les Chapitres 1 à 3. En effet, nous nous intéressons en particulier à deux opérations d'assurance à savoir le provisionnement et la solvabilité. Dans la Section 0.2, nous présentons une revue des modèles de provisionnement agrégés et individuels basés sur les réseaux de neurones.

Ces derniers représentent une technique d'apprentissage automatique que nous utilisons dans notre modèle de réserve individuel présenté dans le Chapitre 1. Ensuite, nous présentons des notions en relation avec la problématique de solvabilité et d'allocation de capital, et qui sont préliminaires pour les Chapitres 2 et 3. Ainsi, dans la Section 0.3 nous présentons les principaux résultats sur la modélisation de la dépendance, puis dans la Section 0.4 nous traitons le cas de la dépendance bivariée extrême. Dans la Section 0.5, nous nous intéressons aux modèles de dépendance hiérarchiques incluant une revue de la théorie des copules. Finalement, les principales contributions de cette thèse sont résumées dans la dernière section.

0.1 Dynamique de l'assurance

L'industrie de l'assurance se caractérise par un cycle de production inversé où une prime fixe est exigée en contrepartie de coûts incertains en cas de sinistre. Ainsi, les compagnies d'assurance adoptent un processus de fonctionnement adéquat dont les principales opérations sont présentées dans la Section 0.1.1. Les Sections 0.1.2 et 0.1.3 portent sur les deux opérations auxquelles on s'intéresse dans ce projet de thèse.

0.1.1 Opérations de compagnies d'assurance

En tant qu'intermédiaires financiers, les compagnies d'assurance jouent un rôle essentiel dans l'économie et accaparent un pourcentage important du produit intérieur brut de plusieurs pays. Elles sont le principal outil de gestion des risques pour les particuliers et les entreprises. Ce transfert de risque promeut les canaux de la productivité, des innovations technologiques, les taux d'épargne et permet aux assurés d'accepter des risques encore plus importants.

Un assureur accepte de couvrir un risque incertain et s'engage à payer les dommages en cas de survenance d'un sinistre en contrepartie d'une prime fixe au début du contrat. Ce cycle de production particulier oblige tout assureur à bien gérer et rentabiliser ses fonds de primes collectées pour honorer ses engagements futurs et préserver sa solvabilité. Par conséquent, les compagnies d'assurance sont alignées sur des processus d'assurance spécifiques que l'on peut agréger en cinq unités opérationnelles :

— **Souscription initiale et tarification** : À l'arrivée d'un nouvel assuré, l'analyste de l'assurance évalue le risque à couvrir et prend la décision de souscrire un contrat ou non. Ensuite, il faut attribuer une prime appropriée (ou taux) en tenant compte de l'exposition au risque, de la couverture proposée et de considérations commerciales. La règle de base consiste à déterminer le juste prix pour le bon risque et d'éviter la sélection contraire.

— **Renouvellement et tarification** : En général, les contrats d'assurance ont des durées d'exposition fixe (exemple ; 1 an) et parfois même très courtes (exemple : 6 mois). À la

fin de cette durée, l'assuré a le choix de renouveler son contrat ou de ne pas le faire, et l'assureur a également la possibilité de refuser la couverture du risque à nouveau ou bien d'ajuster la prime. En cas de renouvellement, l'analyste réévalue la prime selon les données collectées pendant la durée du contrat en tenant compte de l'objectif de fidéliser les clients rentables. Les informations détaillées concernant le renouvellement ou la résiliation sont envoyées à l'assuré, généralement, un mois avant l'échéance de son contrat.

— **Gestion des déclarations** : À la survenance d'un sinistre, l'assuré doit réclamer le plus tôt possible, en respectant les prescriptions prévues dans le contrat. Un agent intervient pour collecter les informations nécessaires à l'ouverture du dossier afin d'évaluer les responsabilités de l'assuré, vérifier si le sinistre est garanti par le contrat et qu'il n'y a surtout pas de fraude. Par la suite, il faut évaluer le montant d'indemnisation en tenant compte de la rétention si elle a lieu.

— **Provisionnement** : Au cours de l'évolution d'une demande d'indemnisation, il peut y avoir un ou plusieurs paiements jusqu'au moment où l'analyste considère que les obligations financières sont résolues et déclare la fermeture du dossier. Ce processus peut durer quelques jours ou même des années en contrepartie d'une prime fixe reçue au début du contrat. Périodiquement, la compagnie d'assurance utilise des outils analytiques pour fournir une estimation appropriée de ses obligations futures et quantifier l'incertitude. Cet exercice permet d'évaluer les fonds de réserve (provisions) nécessaires pour honorer les engagements futurs liés aux sinistres survenus (déclarés ou pas encore) à la date d'évaluation.

— **Solvabilité et allocation du capital** : Les primes collectées constituent la source principale de revenus d'une compagnie d'assurance. Elles sont placées sur le marché financier et immobilier, et contribuent à la constitution des fonds de réserve. La compagnie doit assurer un bon rendement de ses investissements et maintenir une liquidité supérieure à son passif. Pour ceci, elle doit quantifier le montant de capital requis pour préserver sa solvabilité, et décider des moyens de son affectation aux différentes activités et entre les multiples risques.

L'Illustration 0.1 schématise le processus de fonctionnement d'une compagnie d'assurance en indiquant l'enchainement des cinq opérations majeures. Dans cette thèse, nous proposons des contributions reliées au provisionnement, à la solvabilité et à l'allocation du capital. Dans les Sections 0.1.2 et 0.1.3 nous présentons une description plus détaillée de ces deux opérations ainsi que des notions primordiales pour les Chapitres 1 à 3.

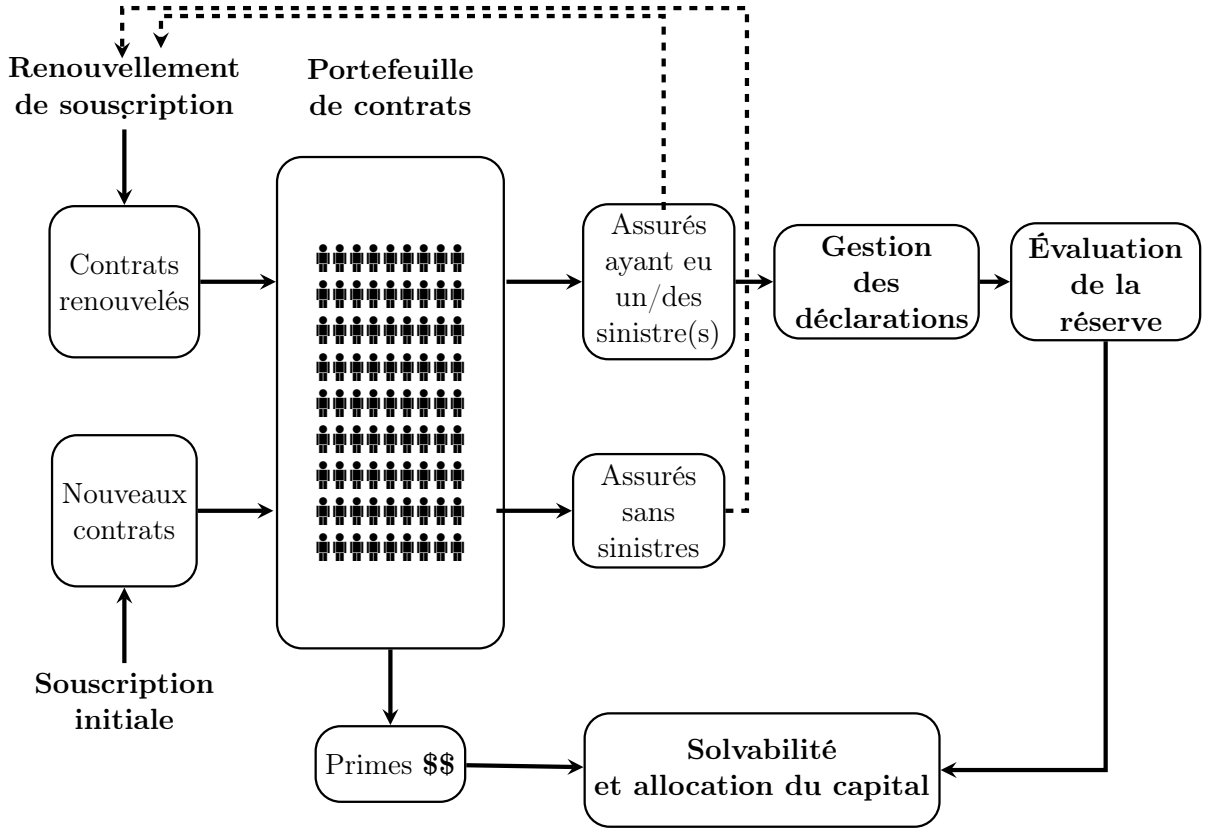


Illustration 0.1 – Opérations principales d'une compagnie d'assurance

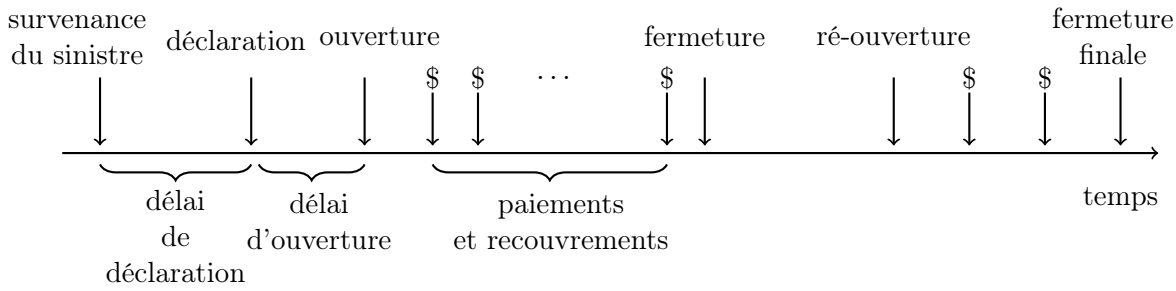


Illustration 0.2 – Exemple de développement d'un sinistre en assurance non-vie

0.1.2 Provisionnement

Au moment de la souscription d'un contrat d'assurance, la prime chargée à l'assuré est connue tandis que les coûts associés au contrat (indemnisation du sinistre) est une variable aléatoire (v.a.). Suite à la survenance d'un sinistre au cours de la période d'assurance, l'assuré le rapporte à l'assureur avec un certain délai de déclaration. Il est aussi possible d'avoir un délai quant à l'ouverture du dossier. À l'ouverture, les caractéristiques spécifiques sont connues, telles que la date de survenance, l'âge de l'assuré ainsi que la description du sinistre. Ces informations permettent d'élaborer une première estimation des coûts. Au cours du développement du

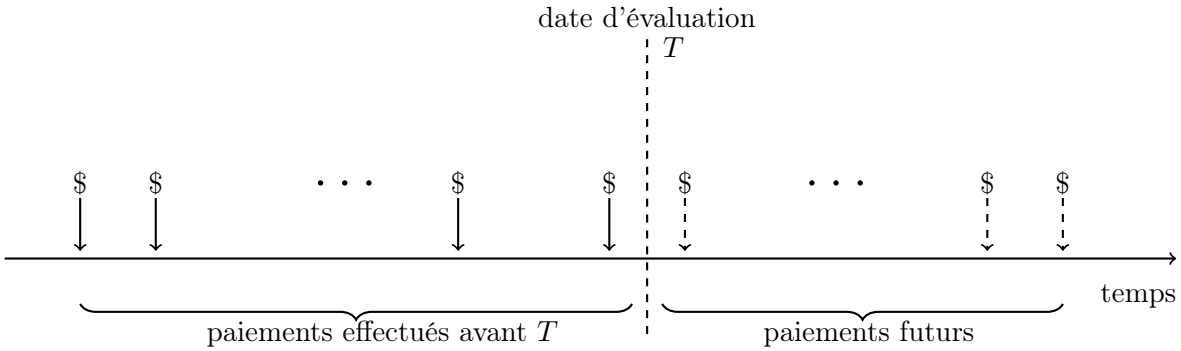


Illustration 0.3 – Exemple de développement d'un sinistre encore ouvert à la date d'évaluation

sinistre, plusieurs paiements peuvent être effectués à différents intervalles de temps avant la clôture du dossier. Le sinistre peut être réouvert puis fermé à nouveau. Tout au long de ce processus de règlement, on observe le développement des covariables du sinistre. Un exemple du développement est présenté à l'Illustration 0.2.

Le règlement d'un sinistre n'est pas forcément effectué pendant la période de validité du contrat et d'encaissement de la prime. Le développement du sinistre peut durer plusieurs années. Pour assurer sa solvabilité et honorer ses engagements à l'égard des assurés, la compagnie d'assurance doit estimer avec précision les paiements à effectuer pour les sinistres survenus et constituer des provisions. L'exercice de provisionnement doit être fait périodiquement. À l'Illustration 0.3, on présente un exemple de paiements à prédire pour un sinistre encore ouvert à la date d'évaluation T. À noter que le nombre de paiements et la date de fermeture sont des données également inconnues à T.

Le nombre de paiements et la durée de développement diffèrent d'une demande d'indemnisation à une autre, ce qui conduit à des données de sinistres non structurées et difficiles à stocker en temps continu et de manière ordonnée. Pour simplifier l'évaluation des provisions, les actuaires agrègent les informations des sinistres hétérogènes pour appliquer des méthodes de provisionnement dites "macro". Généralement, les informations sont agrégées par périodes d'occurrence et de développement sous forme d'un triangle. À l'Illustration 0.4, on présente un triangle de développement agrégé où $C_{i,j}$ peut représenter, par exemple, les paiements pour sinistres survenus à la date i et cumulatifs jusqu'à la période de développement j, pour $i,\ j \in \{1,\ldots,n\}$. La partie supérieure en bleu correspond aux données observées avant la date d'évaluation, tandis que la partie inférieure est à prédire. Il en résulte une perte de caractéristiques individuelles qui peut limiter la précision des prévisions, en particulier dans le cas où la composition du portefeuille ou les attributs des sinistres évoluent au fil du temps. On distingue deux types de modèles agrégés. Le premier est déterministe, tel que le célèbre modèle chain-ladder, tandis que le deuxième est stochastique, où les paiements futurs sont considérés comme des variables aléatoires. Pour un aperçu des approches agrégées de provisionnement,

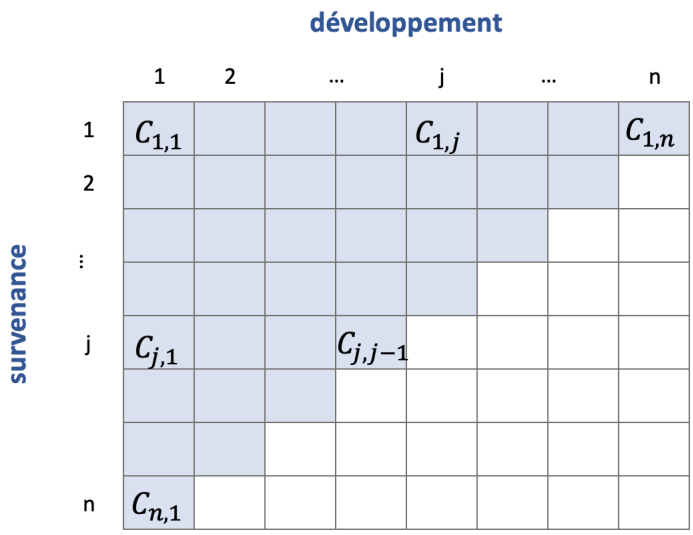


Illustration 0.4 – Exemple de triangle de développement pour sinistres agrégés par période d'occurrence et de développement

voir par exemple Wüthrich and Merz (2008).

Récemment, des chercheurs ont proposé des approches de provisionnement au niveau micro qui utilisent des informations individuelles des sinistres pour améliorer la précision des prévisions. Elles permettent d'obtenir le développement prévisionnel d'un portefeuille sur la base de l'évolution individuelle des sinistres. Toutefois, elles posent de nouveaux défis tels que le traitement des données avec deux types de variables : statiques et dynamiques. Les variables statiques, telles que l'âge du réclamant, sont constantes tout au long du développement du sinistre, tandis que les variables dynamiques peuvent évoluer dans le temps. Un exemple de variables dynamiques est le paiement incrémental en temps continu ou discret. Pour une comparaison entre divers modèles de réserve agrégés et individuels, voir par exemple Charpentier and Pigeon (2016).

Les modèles de réserve individuels, en temps continu ou discret, peuvent être classifiés en deux grandes catégories. La première regroupe les modèles paramétriques basés sur des distributions probabilistes pour les variables aléatoires à prédire. Ces modèles considèrent des formes structurelles fixes et peuvent potentiellement être surparamétrés. La deuxième catégorie regroupe les modèles semi-paramétriques et non paramétriques où l'apprentissage se fait principalement à partir des données. Dans cette catégorie on retrouve, par exemple, le modèle XGBoost, Extra Trees et les réseaux de neurones. Ces derniers nécessitent peu d'hypothèses, intègrent des tendances non linéaires complexes et ont des performances prédictives élevées. Dans la Section 0.2.3, on présente un aperçu des modèles individuels basés sur des réseaux de neurones.

L'Illustration 0.5 schématise la classification des modèles de provisionnement discutés précé-

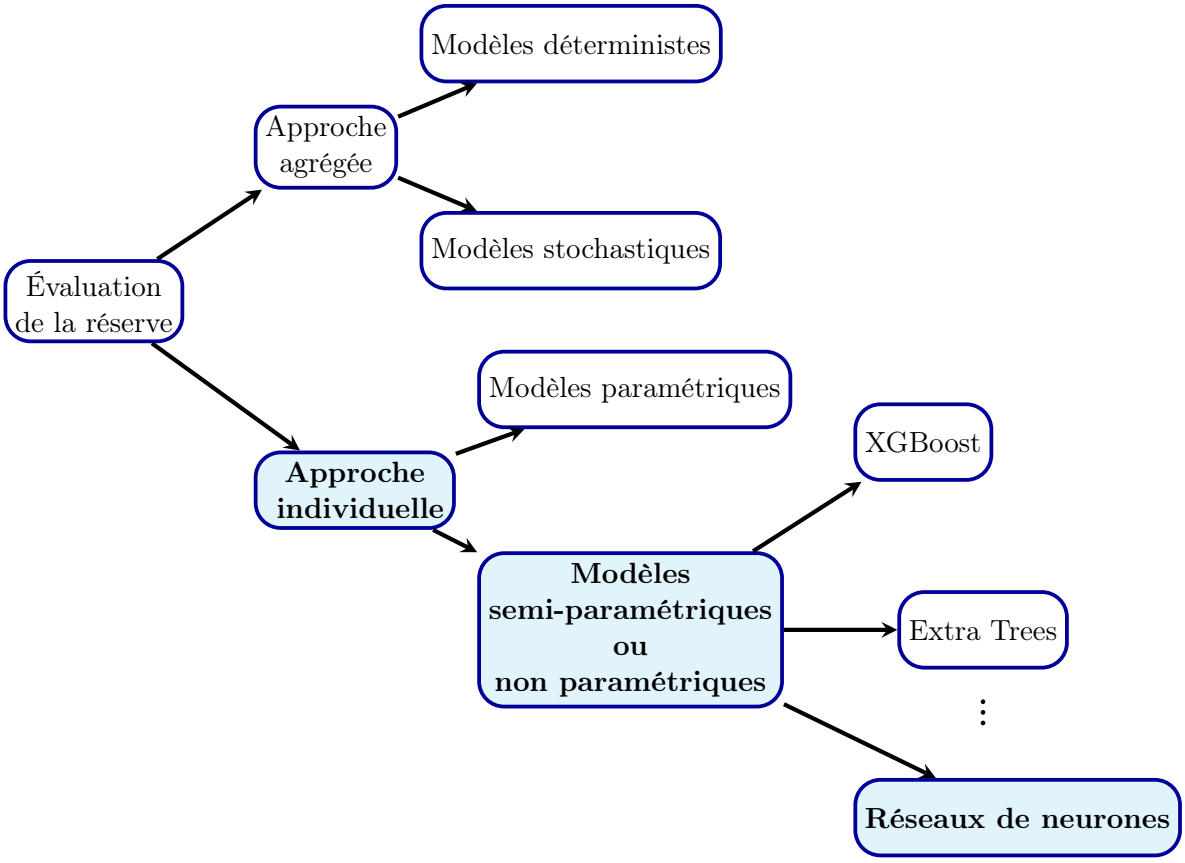


Illustration 0.5 – Classification des approches de provisionnement

demment. On identifie en bleu la catégorie du modèle que nous proposons dans le Chapitre 1, et qui est basé sur un réseau de neurones. Ce dernier est défini dans la Section 0.2.1.

0.1.3 Solvabilité et allocation du capital

Comme toute entreprise, une compagnie d'assurance doit assumer une bonne identification et gestion de ses risques afin de préserver sa croissance et solvabilité dynamique. Cette dernière fait référence à sa capacité d'honorer ses obligations qui sont en majorité les coûts des sinistres à payer. C'est une question importante qui concerne non seulement les actionnaires de la compagnie et les assurés, mais aussi les autorités de régulation. La solvabilité fait partie du domaine émergeant nommé la gestion quantitative des risques (voir par exemple Embrechts and Hofert (2014)) qui se base sur la modélisation mathématique des risques. Étant donné que l'assurance est fondée sur le principe de mutualisation, en plus de l'analyse individuelle des risques, il est nécessaire de comprendre le comportement du portefeuille global. Souvent les risques sont influencés par un ou plusieurs facteurs communs, par exemple d'ordre économique, géographique, technologique ou social. Ainsi, il est moins réaliste de supposer l'indépendance entre les risques et il est plus prudent de modéliser adéquatement la dépendance au sein du portefeuille. Dans la Section 0.3, nous présentons les notions de base et des approches générales

pour modélisation de la dépendance.

L'évaluation du capital économique est nécessaire pour la solvabilité de la compagnie. Elle est souvent fondée sur des mesures de risque appliquées à la somme des risques individuels du portefeuille. Ces mesures sont un outil fondamental dans la gestion et la quantification des risques. En plus de bien évaluer les exigences en capital, il est également important de connaître la méthode rentable d'allocation de ce capital pour l'ensemble des risques individuels du portefeuille. Pour ceci, on peut utiliser des mesures de risque pour calculer la contribution de chaque risque à la perte globale. On peut également déduire le gain de diversification au sein d'un portefeuille défini comme étant le capital épargné en raison de la combinaison de plusieurs risques au lieu de les traiter séparément. La Définition 1 propose une définition générale d'une mesure de risque (voir par exemple Denuit et al. (2006)).

Définition 1 *Soit un risque représenté par la v.a. X. La fonction $\rho : X \to \mathbb{R}^+$ est une mesure de risque reliant le risque X à un nombre réel non négatif $\rho(X)$, éventuellement infini, représentant une provision sécuritaire qui rend le risque acceptable.*

De nombreuses mesures de risque ont été proposées en assurance et en finance. Il est essentiel de comprendre quel aspect incertain du risque on veut quantifier pour choisir la mesure de risque adéquate. Selon les écoles de pensée, il est convenu que toute mesure de risque doit satisfaire certaines propriétés souhaitables. Nous nous intéressons particulièrement aux mesures de risque dites cohérentes vérifiant les conditions énoncées dans la Proposition 1 (voir par exemple Artzner et al. (1999)).

Proposition 1 *Une mesure de risque ρ est dite cohérente si elle possède les propriétés suivantes :*

1. *Homogénéité positive : Pour tout risque X et toute constante $a \in \mathbb{R}^+$, on a*

$$\rho(aX) = a\rho(X).$$

2. *Invariance par rapport à la translation : Pour tout risque X et toute constante $a \in \mathbb{R}^+$, on a*

$$\rho(X + a) = \rho(X) + a.$$

3. *Monotonocité : Soit deux risques X et Y. Alors, on a*

$$\Pr(X \leq Y) = 1 \Rightarrow \rho(X) \leq \rho(Y).$$

4. *Sous-additivité : Soit deux risques X et Y. Alors, on a*

$$\rho(X + Y) \leq \rho(X) + \rho(Y).$$

La Propriété 1.1 signifie que toute modification d'unité monétaire apportée à X conduit à un changement du même ordre pour la mesure de risque. Rappelons que cette dernière permet d'évaluer le capital sécuritaire pour faire face à une incertitude. Par conséquent, il est évident que toute augmentation du risque X d'un montant déterministe a devrait entraîner la même augmentation du capital tel que formuler dans la Propriété 1.2. Selon la Propriété 1.3, plus le risque est élevé, plus le capital requis est grand. La Propriété1.4 signifie que le risque global peut être réduit par la diversification. Dans le cas d'égalité, on parle d'additivité des risques.

Il existe d'autres propriétés auxquelles peut satisfaire une mesure de risque. Il n'y a pas une "meilleure" mesure de risque mais, il suffit de choisir l'une des mesures de risque qui vérifient les propriétés désirées selon le domaine d'application. Dans la littérature actuarielle, il est admis que certaines propriétés soient souhaitables. Premièrement, une marge de risque non excessive ; le capital mis de côté pour faire face à un risque X ne doit pas excéder la valeur de la perte maximale. Deuxièmement, une marge de sécurité positive ; le capital minimum doit excéder les coûts espérés de X afin d'éviter la ruine. Troisièmement, une marge de risque justifiée ; le capital associé à un portefeuille avec coûts déterministes égale à une constante a, ne doit pas être différent de a. Pour plus de détails, voir par exemple Denuit et al. (2006) et Marceau (2013).

Au cours de la dernière décennie, la mesure "Value-at-Risk" (VaR) est devenue la mesure de risque de référence pour les compagnies d'assurance et les institutions financières due à son interprétation simple. Pour un niveau de confiance $\kappa \in (0,1)$, $VaR_\kappa(X)$ correspond au κ^e quantile de la distribution du risque X à garder en provision pour assurer la solvabilité. Par conséquent, cette mesure ne donne aucun renseignement sur la distribution de X au-delà de la valeur $VaR_\kappa(X)$. C'est une mesure non cohérente qui ne satisfait pas la Propriété 1.4 celle-ci fort désirable en pratique. Une alternative populaire est la mesure de risque cohérente "Tail Value-at-Risk" (TVaR).

Définition 2 *Soit X une v.a. définie sur $\mathbb{R}$ avec fonction de répartition F_X et fonction quantile*

$$F_X^{-1}(u) = \inf\{x,\ F_X(x) \geq u\}, u \in (0,1).$$

En supposant que $E[X] < \infty$, la $TVaR$ de X de niveau de confiance $\kappa \in (0,1)$ est définie par

$$TVaR_\kappa(X) = \frac{1}{1-\kappa}\int_\kappa^1 VaR_u(X)\,\mathrm{d}u, \tag{1}$$

où $VaR_\kappa(X) = F_X^{-1}(\kappa)$.

En se basant sur l'équation (1), la $TVaR_\kappa(X)$ représente l'espérance de la perte sachant que cette dernière prend des valeurs supérieures à $VaR_\kappa(X)$. Autrement dit, la mesure de risque $TVaR_\kappa(X)$ est une moyenne des $VaR_p(X)$, pour $p \in (\kappa,1)$. Elle permet donc de donner une

indication sur la queue droite de la distribution du risque. Cette mesure satisfait les trois propriétés désirables en actuariat, contrairement à la $VaR_\kappa(X)$ qui vérifie uniquement deux propriétés, à savoir une marge de risque justifiée et non excessive.

0.2 Réseau de neurones et provisions

Les réseaux de neurones (RNs) sont une classe de techniques d'apprentissage automatique. Ils sont très flexibles pour modéliser des données structurées et non structurées, ce qui leur a permis de gagner en popularité dans plusieurs domaines d'application. De nombreuses contributions récentes ont montré comment les RNs peuvent améliorer la précision de la prédiction des provisions individuelles. Dans cette section, nous présentons en premier les RNs, puis un aperçu de leur utilisation dans des modèles de réserve agrégés et individuels.

0.2.1 Réseau de neurones

McCulloch and Pitts (1943) a introduit les RNs, en s'inspirant de la structure et le fonctionnement du cerveau humain. Les RNs ont l'avantage de nécessiter peu d'hypothèses, d'intégrer des tendances non linéaires complexes et d'avoir des performances prédictives élevées. En revanche, ils sont parfois délicats à calibrer, ont un grand nombre de paramètres, une faible capacité d'interprétation et nécessitent un grand nombre d'observations. L'Illustration 0.6 représente un simple exemple de RN avec plusieurs variables $\{x_i,\ i = 1,\dots,p\}$ entièrement connectées à une couche cachée pour obtenir deux variables réponses $\{y_1,y_2\}$. Les nœuds du réseau sont connectés par des poids (paramètres) qui sont optimisés en tenant compte d'une fonction de perte. Pour un compte rendu détaillé sur le fonctionnement des RNs, voir par exemple Goodfellow et al. (2016).

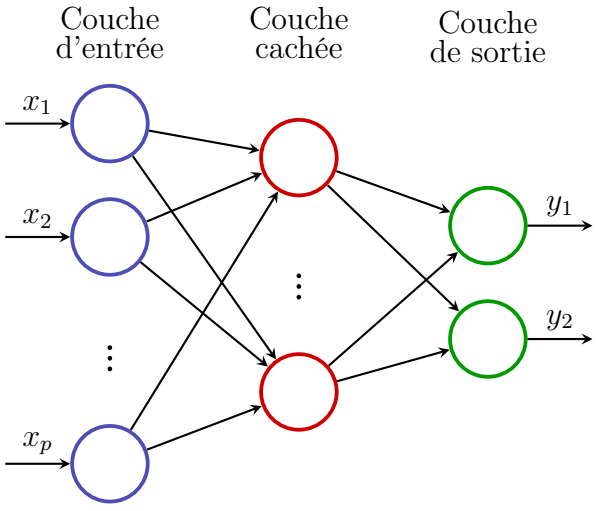


Illustration 0.6 – Exemple de RN avec plusieurs entrées et sorties

Les réseaux de neurones récurrents (RNR) représentent une classe populaire de RNs introduite par Hopfield (1982). Au cours des dernières décennies, ils ont prouvé leur potentiel dans divers domaines tels que la classification, la reconnaissance vocale, la modélisation du langage et le guidage de robots. Leur architecture les rend plus adaptés aux données séquentielles. En effet, comme représenté dans l'Illustration 0.7, un RNR est un enchaînement de plusieurs copies du même réseau de neurones (module). À chaque étape, le réseau reçoit un vecteur de variables $\boldsymbol{X}_t$, pour $t = 1, \ldots, n$, et génère l'état caché $\boldsymbol{h}_t$. La structure des RNRs permet de considérer les informations pertinentes et tenir compte des dépendances temporelles. Cependant, en raison du gradient multiplicatif qui peut croître de manière exponentielle, il est difficile de capturer les dépendances à long terme avec un RNR simple.

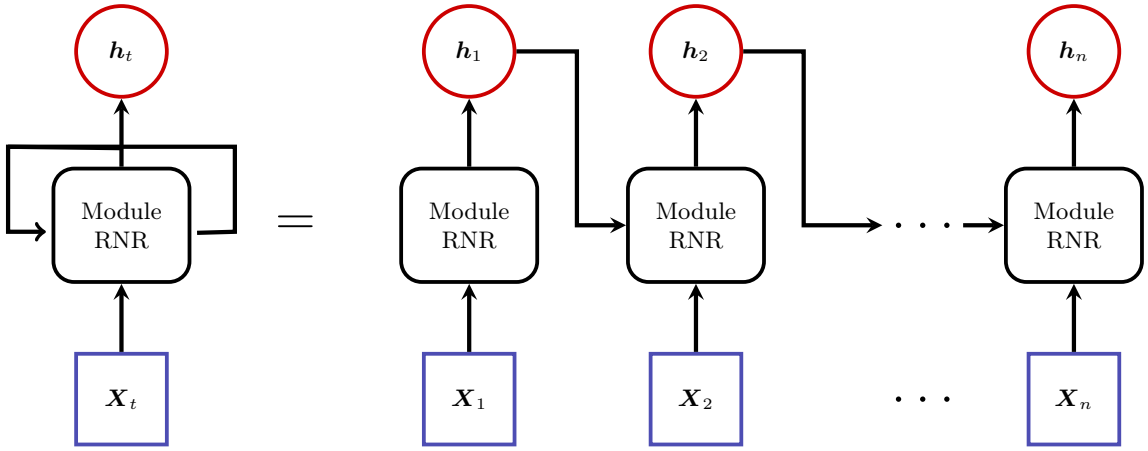


Illustration 0.7 – Structure d'un RNR avec variables d'entrée dynamiques

Pour remédier au problème du gradient, Hochreiter and Schmidhuber (1997) a introduit les réseaux récurrents à mémoire court et long terme (*Long Short-Term Memory* – LSTM) en tant que classe spécifique des RNRs. Ainsi, un LSTM peut être également représenté par la structure 0.7, cependant, le module répétitif a une architecture complexe avec de multiples couches interagissant de manière particulière. Le module est également caractérisé par une cellule mémoire permettant le passage des informations sans problème de disparition de gradient. Les réseaux LSTMs ont montré des résultats compétitifs en comparaison à d'autres architectures traditionnelles de RNRs. Voir, par exemple, Graves and Schmidhuber (2005), Sundermeyer et al. (2012), Graves (2013) et Weninger et al. (2015).

Durant les dernières années, des chercheurs et praticiens se sont intéressés davantage à l'utilisation des RNs pour résoudre des problématiques actuarielles. Particulièrement, ils ont développé plusieurs modèles de réserve basés sur différentes classes et structures de RNs. Dans les Sections 0.2.2 et 0.2.3 nous présentons, respectivement, une revue de ces modèles de réserve agrégés et individuels. En s'inspirant de la performance de ces modèles, dans le Chapitre 1 nous proposons un nouveau modèle de réserve individuel utilisant une structure de réseau LSTM qui est plus adaptée aux données de sinistre temporelles.

0.2.2 Réseau de neurones et provisions agrégées

Les méthodes de réserve agrégées habituelles consistent à compléter la partie inférieure du triangle de l'Illustration 0.4, en estimant les facteurs de développement afin de prédire les paiements futurs agrégés par période. Certains modèles agrégés classiques ont connu des extensions en utilisant des techniques d'apprentissage automatique notamment des RNs.

Wüthrich (2018) propose une extension de la méthode agrégée *chain-ladder* en modélisant les facteurs de développement de la cadence des paiements avec un RN utilisant les informations des sinistres. Ce réseau permet de tenir compte de l'hétérogénéité des observations. En se basant sur des triangles de développement pour différentes lignes d'affaires, Kuo (2019) développe un RN profond (à plusieurs couches cachées) permettant une modélisation conjointe du nombre de sinistres survenus ainsi que les paiements futurs. Malgré les résultats encourageants, les triangles de données agrégées n'exploitent pas pleinement le potentiel des RNs. Ces derniers peuvent potentiellement mieux fonctionner avec de grandes bases de données cachant des structures complexes. Pour améliorer la performance des RNs dans un contexte de réserve agrégée, il est possible d'y intégrer des modèles classiques de réserve paramétriques. Ceci correspond à l'implémentation du modèle classique en premier pour utiliser les résultats obtenus comme point de départ de l'apprentissage du réseau. Ainsi, au cours du processus d'entraînement, les poids du RN vont s'ajuster plus rapidement pour minimiser les erreurs de prédiction. Cette approche garantit la stabilité et accélère l'apprentissage tel que montré dans Gabrielli et al. (2020) et Gabrielli (2020b). En effet, Gabrielli et al. (2020) initialise un RN avec un modèle de régression Poisson surdispersé à classification croisée. Le réseau proposé permet d'améliorer le processus d'estimation des paramètres de la régression en recherchant une structure de modèle au-delà de celle spécifiée et ceci en utilisant des triangles de sinistres de différentes lignes d'affaires. Cette idée est étendue à la prédiction conjointe du nombre de sinistres et paiements futurs dans Gabrielli (2020b). Une autre approche combinant des modèles de régression plus complexes et des RNs est introduite par Lindholm et al. (2020). Cette méthode a l'avantage de produire des réserves séparées pour les sinistres encore ouverts et ceux survenus mais non encore déclarés.

0.2.3 Réseau de neurones et provisions individuelles

Avec l'avènement du Big Data et l'évolution des techniques de collecte et stockage des données individuelles des réclamations, il devient de plus en plus intéressant d'appliquer des méthodes de réserve individuelle (réserve granulaire). Les modèles individuels paramétriques sont difficiles à estimer avec un très grand nombre d'observations non structurées. Récemment, plusieurs chercheurs et praticiens se sont intéressés à l'utilisation de divers RNs pour la prédiction des réserves individuelles. Dans la littérature, on identifie des RNs simples et d'autres avec des architectures complexes nécessitant des connaissances approfondies. Par conséquent, les actuaires doivent développer leur compréhension des RNs pour implémenter et mettre en

pratique ces modèles.

Gabrielli (2020a) a structuré un RN effectuant simultanément deux tâches pour estimer les paiements futurs. La tâche de classification permet de prédire le type de paiement, tandis que la régression prédit le montant. Au lieu d'utiliser l'historique exact de la réclamation comme entrée du réseau, les covariables individuelles observées sont remplacées par des indicateurs. Sur la base de l'historique complet des sinistres, Delong and Wüthrich (2020) a utilisé des RNs pour prédire le développement conjoint des sinistres encourus et des paiements individuels. La tâche de provisionnement est décomposée en six étapes, et un RN est formé pour chaque tâche. Une autre façon d'utiliser l'évolution passée des sinistres est de recourir à des RNRs. Pour plusieurs tâches séquentielles de traitement de l'information, Kuo (2020) a récemment proposé une approche de provisionnement individuelle avec un modèle de prévision de paiements sur plusieurs périodes en utilisant des réseaux LSTMs avec densité de mélange bayésien. Ce modèle est constitué d'un LSTM encodeur pour les flux financiers passés et les séquences d'état des sinistres, et un LSTM décodeur pour générer une distribution des pertes payées. Cependant, il n'améliore pas la précision des prédictions par rapport au modèle classique chain-ladder.

Dans le Chapitre 1, nous développons un nouveau modèle de provisionnement individuel basé sur un LSTM et applicable à toute base de données détaillées des sinistres à long développement. Notre réseau effectue deux tâches afin de prédire les paiements futurs : une classification pour déterminer la probabilité d'avoir un paiement non nul et une régression pour évaluer le montant incrémental sachant qu'il est non nul. Ce réseau multitâche permet une régularisation plus robuste et d'améliorer les performances en raison du partage des caractéristiques individuelles entre les deux tâches.

0.3 Modélisation de la dépendance

La modélisation de la dépendance est la description probabiliste de la relation entre deux ou plusieurs variables aléatoires. Elle constitue l'un des sujets les plus étudiés par les chercheurs en mathématique et statistique dans le but de mieux comprendre le monde ; identifier les causes des maladies, évaluer le changement climatique, prévoir les cycles économiques ou se prémunir contre les crises financières. En assurance, pour évaluer adéquatement le capital requis pour couvrir la perte totale d'un portefeuille, il est important de comprendre la structure de dépendance définissant la relation entre les différents risques individuels qui le composent. Ces risques sont liés entre eux, notamment par des facteurs communs comme la géographie, le climat, le contexte économique, etc. Dans cette section, on présente les notions préliminaires pour la modélisation de la dépendance ainsi que les mesures d'association et de concordance.

0.3.1 Classes de Fréchet

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un vecteur de n variables aléatoires avec fonction de répartition conjointe

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n), \tag{2}$$

avec $(x_1, \ldots, x_n) \in \mathbb{R}^n$. La fonction de survie associée à $\mathbf{X}$ est définie par

$$\overline{F}_{\mathbf{X}}(x_1, \ldots, x_n) = \Pr(X_1 > x_1, \ldots, X_n > x_n).$$

En se basant sur (2), on peut déduire la fonction de répartition marginale de la v.a. X_i, pour $i = 1, \ldots, n$, avec

$$F_i(x_i) = \Pr(X_i \leq x_i) = \lim_{x_j \to \infty,\ i \neq j \in \{1,\ldots,n\}} F_{\mathbf{X}}(x_1, \ldots, x_n).$$

On peut construire différentes structures de dépendance entre les composantes du vecteur $\mathbf{X}$. L'ensemble des fonctions de répartition multivariées $F_{\mathbf{X}}$ possibles constitue la classe de Fréchet, notée $\Gamma(F_1, \ldots, F_n)$. On connaît au moins une notion de dépendance qui appartient à $\Gamma(F_1, \ldots, F_n)$; c'est l'indépendance où

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = F_1(x_1) \times \cdots \times F_n(x_n). \tag{3}$$

Ainsi, $\Gamma(F_1, \ldots, F_n)$ est un ensemble non vide. La popularité de cette classe émane de ses bornes (inférieure et supérieure) introduites par Fréchet (1951) et dont les propriétés sont étudiées par plusieurs chercheurs. Toute fonction de répartition conjointe $F_{\mathbf{X}} \in \Gamma(F_1, \ldots, F_n)$ est comprise entre deux fonctions à n dimensions définies en termes des marginales $F_1, \ldots, F_n$. La borne supérieure de Fréchet est définie par $M_n(x_1, \ldots, x_n) = \min\{F_1(x_1), \ldots, F_n(x_n)\}$ tandis que la borne inférieure de Fréchet correspond à $W_n(x_1, \ldots, x_n) = \max\{\sum_{i=1}^{n} F_i(x_i) - n + 1; 0\}$. Pour $F_{\mathbf{X}} \in \Gamma(F_1, \ldots, F_n)$, on a

$$W_n(x_1, \ldots, x_n) \leq F_{\mathbf{X}}(x_1, \ldots, x_n) \leq M_n(x_1, \ldots, x_n), \tag{4}$$

La borne supérieure de Fréchet $M_n \in \Gamma(F_1, \ldots, F_n)$ est la fonction de répartition pour un vecteur avec dépendance positive extrême (comonotonocité, voir Section 0.4.1). Ce résultat est vrai : quelle que soit la dimension $n \geq 2$. En contrepartie, pour $n = 2$, la borne inférieure de Fréchet est la fonction de répartition pour la structure de dépendance bivariée négative extrême (antimonotonocité, voir Section 0.4.2), avec $W_2 \in \Gamma(F_1, F_2)$. Pour $n \geq 3$, W_3 n'est pas forcément une fonction de répartition multivariée (voir les contre-exemples fournis par Tchen (1980) et l'exemple 3.1 de Joe (1997)).

0.3.2 Mesures d'association et de concordance.

En outre de l'identification de la structure de dépendance entre des v.a., il est également important de mesurer l'association entre celles-ci. Scarsini (1984) a défini certaines propriétés

souhaitables pour une mesure d'association entre deux v.a. Les mesures satisfaisant ces propriétés sont dites concordantes (voir Définition 3). L'idée intuitive du concept de concordance est la suivante : deux v.a. X_1 et X_2 sont concordantes si les grandes valeurs de X_1 vont de pair avec de grandes valeurs de X_2 et réciproquement pour les petites valeurs.

Définition 3 *Soit X_1 et X_2 deux v.a. La fonction r mesurant l'association entre X_1 et X_2 est dite concordante si et seulement si elle satisfait les propriétés suivantes :*

P1 *Symétrie : $r(X_1,X_2) = r(X_2,X_1)$;*

P2 *Normalisation : $-1 \leq r(X_1,X_2) \leq 1$;*

P3 *Si (X_1,X_2) est une paire comonotone : $r(X_1,X_2) = 1$.*

P4 *Si (X_1,X_2) est une paire antimonotone : $r(X_1,X_2) = -1$;*

P5 *Pour $f : \mathbb{R} \to \mathbb{R}$ strictement monotone :*

$$r(f(X_1),X_2) = \begin{cases} r(X_1,X_2), & \text{si } f \text{ est croissante,} \\ -r(X_1,X_2), & \text{si } f \text{ est décroissante.} \end{cases}$$

Le coefficient populaire de corrélation de Pearson est une mesure de dépendance linéaire seulement et ne permet pas de mesurer l'association non linéaire entre deux v.a. Ce coefficient satisfait uniquement les Propriétés **P1** et **P2** et n'est donc pas une mesure de concordance. Il existe d'autres mesures satisfaisant les Propriétés **P1**-**P5**, dont le tau de Kendall et le rhô de Spearman dans le cas de v.a. continues. Ces deux mesures sont des mesures de corrélation de rang dont la valeur ne dépend pas des marginales (voir par exemple Denuit and Dhaene (2003) et Denuit and Lambert (2005)). Ainsi, ces mesures permettent d'évaluer la dépendance entre les variables sans tenir compte de leur comportement individuel.

Définition 4 *Soit (X_1,X_2) une paire de v.a. continues et (X_1',X_2') une paire indépendante et identiquement distribuée à (X_1,X_2), telles que $(X_1,X_2) =_d (X_1',X_2')$. Le tau de Kendall mesurant la corrélation de rang entre X_1 et X_2 est défini par :*

$$\begin{aligned} \tau(X_1,X_2) &= \Pr(\textit{concordance}) - \Pr(\textit{discordance}) \\ &= \Pr\left((X_1 - X_1')(X_2 - X_2') > 0\right) - \Pr\left((X_1 - X_1')(X_2 - X_2') < 0\right) \end{aligned}$$

Selon la Définition 4, le tau de Kendall correspond à la probabilité de concordance moins la probabilité de discordance. Dans un échantillon d'observations, la mesure non paramétrique du tau de Kendall est basée sur le nombre de concordances et de discordances. La concordance se produit lorsque les paires d'observations varient dans le même sens, et la discordance se produit lorsque des paires d'observations varient différemment.

Définition 5 *Soit* $\mathbf{X} = (X_1,X_2)$ *une paire de v.a. continues avec fonction de répartition conjointe* $F_{\mathbf{X}} \in \Gamma(F_1,F_2)$. *On définit le rhô de Spearman comme suit :*

$$\begin{aligned}\rho_s(X_1,X_2) =& Cor(F_1(X_1),F_2(X_2)) \\ =& \frac{E\left[F_1(X_1)F_2(X_2)\right] - 1/4}{1/12} \\ =& 12 \int\int F_1(x_1)F_2(x_2)dF_{\mathbf{X}}(x_1,x_2) - 3,\end{aligned}$$

où $Cor(\cdot,\cdot)$ *désigne le coefficient de corrélation de Pearson et* $F_1(X_1) =_d F_2(X_2) \sim Unif(0,1)$.

Le rhô de Spearman pour une paire de v.a. continues X_1 et X_2 correspond au coefficient de corrélation de Pearson pour les rangs $F_1(X_1)$ et $F_2(X_2)$.

Les mesures de concordance tau de Kendall et rhô de Spearman prennent souvent des valeurs assez différentes, mais toujours appartenant à l'intervalle $[-1,1]$ (Propriété **P2**) et vérifiant l'inégalité suivante (voir Denuit et al. (2006)) :

$$\begin{aligned}\frac{3\tau - 1}{2} \leq \rho_s \leq \frac{1 + 2\tau - \tau^2}{2}, \ \tau \geq 0, \\ \frac{\tau^2 + 2\tau - 1}{2} \leq \rho_s \leq \frac{1 + 3\tau}{2}, \ \tau \leq 0,\end{aligned}$$

avec $\tau = \tau(X_1,X_2)$ et $\rho_s = \rho_s(X_1,X_2)$.

0.4 Dépendances bivariées extrêmes

Pour un vecteur aléatoire $\mathbf{X}$, la modélisation de la structure de dépendance peut conduire à des défis mathématiques et statistiques. Récemment, plusieurs travaux de recherche se sont intéressés aux concepts de dépendance extrême ainsi qu'à leur application dans plusieurs domaines. Ces concepts permettent de quantifier des bornes pour les mesures de risque selon les scénarios extrémums. Dans cette section, on présente les structures de dépendance bivariée positive et négative extrême.

0.4.1 Comonotonocité

Lors d'un évènement naturel extrême tel qu'un tremblement de terre, un tsunami, une famine ou une épidémie, les personnes, propriétés et facteurs économiques se trouvant dans la même zone graphique seront affectés de façon similaire. Dans un tel contexte, on parle de risques avec dépendance positive extrême. Ce cas particulier de relation de dépendance est nommé la comonotonicité.

Les composantes d'un vecteur aléatoire sont dites comonotones si elles se comportent de la même manière telle que des valeurs élevées pour l'une des variables impliquent des valeurs

élevées pour toutes les autres et réciproquement. Pour une revue de littérature sur les développements du concept de comonotonicité, voir, par exemple, Deelstra et al. (2011) qui présente également plusieurs exemples d'application en finance et assurance. Dans ce qui suit, on considère uniquement un contexte de risque bivarié pour présenter les propriétés importantes des variables comonotones et on note $\mathbf{X}^+ = (X_1^+, X_2^+)$ une paire de v.a. comonotones. Tous les résultats peuvent être facilement généralisés pour plus de deux v.a. comonotones.

Définition 6 *Une paire de v.a.* $\mathbf{X}^+ = \left(X_1^+, X_2^+\right)$ *est comonotone si et seulement s'il existe une v.a.* Z *et des fonctions non décroissantes* ϕ_1 *et* ϕ_2 *telles que*

$$\mathbf{X}^+ =_d \left(\phi_1(Z), \phi_2(Z)\right).$$

Selon la Définition 6, les composantes de $\mathbf{X}^+$ s'expriment sous forme de fonctions croissantes d'un facteur de risque commun. Par exemple, dans un marché financier, tous les actifs peuvent être influencés par un choc économique unique tel que le cas de la crise financière mondiale de 2007-2008. Dans une catastrophe naturelle, l'intensité peut être vue comme un facteur de risque commun pour toutes les personnes impliquées. Plus l'ampleur de la catastrophe est élevée, plus les dommages sont importants.

Proposition 2 *Soit* $U \sim Unif(0,1)$ *et* $\mathbf{X} = (X_1, X_2)$ *un vecteur aléatoire dont les composantes sont définies par*

$$X_i = F_i^{-1}(U), \tag{5}$$

pour $i = 1,2$. *Alors, par la Définition 6, les composantes de* $\mathbf{X}$ *sont dites comonotones.*

On déduit également que si les variables aléatoires X_1^+ et X_2^+ sont comonotones et identiquement distribuées, tels que $X_1^+ \sim X_2^+ \sim X$, cela revient à dire qu'elles prennent toutes la même valeur. La relation en (5) est importante pour établir la formule de la fonction de répartition conjointe du vecteur $\mathbf{X}^+$, l'algorithme de simulation ainsi que l'agrégation de risques comonotones comme présenté ci-dessous.

Proposition 3 *Le vecteur de v.a.* $\mathbf{X} = (X_1, X_2)$ *a des composantes comonotones si et seulement si sa fonction de répartition conjointe est la borne supérieure de Fréchet.*

Preuve. Pour le vecteur de v.a. $\mathbf{X}^+$ défini en (5), on a

$$\begin{aligned}
F_{\mathbf{X}^+}(x_1,x_2) &= \Pr\left(X_1^+ \leq x_1, X_2^+ \leq x_2\right) \\
&= \Pr\left(F_1^{-1}(U) \leq x_1, F_2^{-1}(U) \leq x_2\right) \\
&= \Pr\left(U \leq F_1(x_1), U \leq F_2(x_2)\right) \\
&= \Pr\left(U \leq \min\left(F_1(x_1); F_2(x_2)\right)\right) \\
&= M_2(x_1,x_2).
\end{aligned}$$

■

Selon la Proposition 3, on a $F_{\mathbf{X}^+} = M_2 \in \Gamma(F_1,F_2)$. En se basant sur la représentation stochastique (5), on peut déduire la procédure de simulation d'une paire de variables aléatoires comonotones présentée dans l'Algorithme 1.

Algorithme 1 : Simulation des réalisations de $\mathbf{X}^+$

1. On simule une réalisation $U^{(j)}$ de la v.a. $U \sim Unif\,(0,1)$.
2. On calcule $X_1^{+(j)} = F_1^{-1}\left(U^{(j)}\right)$, $X_2^{+(j)} = F_2^{-1}\left(U^{(j)}\right)$.

Pour une paire de variables aléatoires comonotones, on note le risque total par S^+. Selon l'Éq. (5), on a

$$S^+ = \sum_{i=1}^{2} F_i^{-1}(U) = \varphi^+(U), \tag{6}$$

où $U \sim Unif(0,1)$ et $\varphi^+(u)$ est une fonction croissante pour $u \in (0,1)$. En se basant sur l'expression (6), on peut déduire une expression explicite et simple pour les mesures de risque VaR et $TVaR$ présentée dans la Proposition 4. L'idée clé de la démonstration de cette proposition est la propriété de la VaR qui stipule que pour toute fonction croissante et continue φ de X, on a $VaR_\kappa(\varphi(X)) = \varphi\,(VaR_\kappa(X))$, pour $\kappa \in (0.1)$.

Proposition 4 ***(Additivité des VaR et des TVaR).*** *Soit $\mathbf{X}^+ = (X_1^+,X_2^+)$ le vecteur de v.a. comonotones et $S^+ = X_1^+ + X_2^+$. Pour $\kappa \in (0.1)$, on a*

$$VaR_\kappa\left(S^+\right) = \sum_{i=1}^{2} VaR_\kappa\left(X_i\right). \tag{7}$$

et

$$TVaR_\kappa\left(S^+\right) = \sum_{i=1}^{2} TVaR_\kappa\left(X_i\right). \tag{8}$$

En théorie du risque et en finance, on s'intéresse souvent à la somme S des risques individuels d'un portefeuille. Dans le cas de difficulté d'identification de la distribution de S, il est très utile de connaitre les bornes inférieures et supérieures pour diverses quantifications du risque. Soit un vecteur de v.a. $\mathbf{X} = (X_1,X_2)$ où $F_{\mathbf{X}} \in \Gamma\,(F_1,F_2)$. Soit $F_{\mathbf{X}^+} \in \Gamma\,(F_1,F_2)$ où

$$F_{\mathbf{X}}\,(x_1,x_2) \leq F_{\mathbf{X}^+}\,(x_1,x_2) = M_2\,(x_1,x_2).$$

On définit $S = \sum_{i=1}^{2} X_i$ et $S^+ = \sum_{i=1}^{2} X_i^+$. En supposant que $E[X_i^2] < \infty$, on a $E\,[S] = E\,[S^+]$. Pour $\kappa \in (0,1)$, on a

$$TVaR_\kappa\,(S) \quad \leq \quad TVaR_\kappa\left(S^+\right). \tag{9}$$

Cependant, la relation $VaR_\kappa(S) \leq VaR_\kappa(S^+)$ n'est pas vérifiée pour tout $\kappa \in (0,1)$ (voir par exemple Denuit et al. (2006) et Deelstra et al. (2011)). La relation de comonotonocité représente la dépendance la plus risquée au sens de la TVaR la plus élevée et fort utilisée pour la modélisation des catastrophes.

0.4.2 Antimonotonocité

L'antimonotonocité correspond à la relation de dépendance négative extrême. Contrairement à la comonotonocité, le concept d'antimonocité est défini uniquement dans un contexte bivarié. Les composantes d'une paire de v.a., notée $\mathbf{X}^- = (X_1^-,X_2^-)$, sont dites antimonotones si elles se comportent de manière opposée telle que des valeurs élevées pour la première impliquent des valeurs faibles pour la seconde et vice versa. Ce cas particulier de relation est souhaitable pour une mutualisation de risques profitable au sein d'un portefeuille dans le but de minimiser le risque total. L'antimonotonocité, importante autant que la comonotonocité pour résoudre de nombreux problèmes d'optimisation, a été moins étudiée dans la littérature principalement en raison de son extension difficile aux dimensions supérieures à 2.

Définition 7 *Une paire de v.a.* $\mathbf{X}^- = \left(X_1^-,X_2^-\right)$ *est antimonotone si et seulement s'il existe une v.a.* Z*, une fonction croissante* ϕ_1 *et une fonction décroissante* ϕ_2 *telles que*

$$\mathbf{X}^- =_d \left(\phi_1(Z),\phi_2(Z)\right).$$

Dans la Définition 7, la monotonocité opposée des fonctions ϕ_1 et ϕ_2 traduit la condition du comportement opposé des v.a. antimonotones. En analogie avec la Proposition 2, la Proposition 5 présente une représentation stochastique très utile pour un couple de v.a. antimonotones.

Proposition 5 *Soit* $U \sim Unif(0,1)$ *et* $\mathbf{X} = (X_1,X_2)$ *une paire aléatoire dont les composantes sont définies par*

$$X_1 = F_1^{-1}(U) \text{ et } X_2 = F_2^{-1}(1-U). \tag{10}$$

Alors, $\mathbf{X}$ *est antimonotone en vertu de la Définition 7.*

Proposition 6 *Le couple de v.a.* $\mathbf{X} = (X_1,X_2)$ *a des composantes antimonotones si et seulement si sa fonction de répartition conjointe est la borne inférieure de Fréchet* W_2*.*

Preuve. Pour le couple de v.a. $\mathbf{X}^-$ défini en (10), on a

$$\begin{aligned} F_{\mathbf{X}^-}(x_1,x_2) &= \Pr(X_1 \leq x_1, X_2 \leq x_2) \\ &= \Pr\left(F_1^{-1}(U) \leq x_1, F_2^{-1}(1-U) \leq x_2\right) \\ &= \Pr(U \leq F_1(x_1), 1-U \leq F_2(x_2)) \\ &= \Pr(U \leq F_1(x_1), U > 1-F_2(x_2)) \end{aligned}$$

qui correspond à $F_1(x_1)+F_2(x_2)-1$, si $F_1(x_1) \geq 1-F_2(x_2)$, ou à 0, si $F_1(x_1) < 1-F_2(x_2)$. Alors, on déduit

$$\begin{aligned} \Pr(U \leq F_1(x_1), 1-U \leq F_2(x_2)) &= \max(F_1(x_1)+F_2(x_2)-1;0) \\ &= W_2(x_1,x_2). \end{aligned}$$

■

Ainsi, on a $F_{\mathbf{X}^-} = W_2 \in \Gamma(F_1,F_2)$. Rappelons que W_n n'est pas toujours une fonction de répartition pour $n > 2$. En se basant sur la représentation (10), on peut facilement déduire la procédure de simulation de la paire (X_1^-,X_2^-) comme décrit dans l'Algorithme 2.

Algorithme 2 : Simulation des réalisations de $\mathbf{X}^-$

1. On simule une réalisation $U^{(j)}$ de la v.a. $U \sim Unif(0,1)$.
2. On calcule $X_1^{-(j)} = F_1^{-1}\left(U^{(j)}\right)$, $X_2^{-(j)} = F_2^{-1}\left(1-U^{(j)}\right)$.

On note la somme d'une paire antimonotone par S^-. En utilisant la représentation (10), on obtient

$$S^- = F_1^{-1}(U) + F_2^{-1}(1-U) = \varphi^-(U). \tag{11}$$

Contrairement à la monotonocité croissante de φ^+ en fonction de u (voir Eq. (6)), pour S^- la forme de $\varphi^-(u)$, $u \in (0,1)$, est influencée par le choix des marginales F_1 et F_2. Ainsi, on ne peut pas utiliser la propriété de la VaR pour déduire une décomposition de $VaR_\kappa(S^-)$ et $TVaR_\kappa(S^-)$ analogue à (7) et (8). Dans le Chapitre 2, nous proposons des expressions explicites pour ces deux mesures pour S^- pour plusieurs exemples de distributions bivariées strictement continues. Cette contribution permet de déduire des expressions explicites (fermées) pour la prime stop-loss ainsi que divers indices de diversification.

De manière analogue à l'inégalité (9), pour toute structure de dépendance du vecteur aléatoire (X_1,X_2), la $TVaR$ de la somme $S = X_1 + X_2$ est bornée à gauche par sa contrepartie antimonotone. En se basant sur la mesure $TVaR$, l'antimonotonocité est la dépendance la moins risquée. Pour $\kappa \in (0,1)$, on a

$$TVaR_\kappa\left(S^-\right) \leq TVaR_\kappa(S) \leq TVaR_\kappa\left(S^+\right). \tag{12}$$

Grâce à notre contribution dans le Chapitre 2, pour plusieurs classes de distribution un gestionnaire peut évaluer un intervalle pour la mesure $TVaR$ en utilisant l'inégalité (12) et cadrer le coût du risque en absence d'information sur la distribution de la perte globale S.

0.5 Modèles de dépendance multivariés

La majorité des problématiques de modélisation impliquent plusieurs facteurs aléatoires, notamment en actuariat. Les modèles de dépendance multivariés sont donc nécessaires. Les

copules constituent une option de modélisation très populaire étant donné leur flexibilité qui permet d'intégrer divers degrés et formes de dépendance entre les variables. Dans cette section, on présente une revue de la théorie des copules et ses différentes familles. Ces notions sont essentielles pour le Chapitre 3.

0.5.1 Copules

Au cours des dernières années, l'intérêt pour la théorie des copules a connu une croissance importante. Une simple recherche sur Google portant sur le terme « copula » produisait 0.6M résultats en 2005, et 1.13M en 2007. En 2021, cette même requête engendre presque 9M de résultats. Les copules sont de plus en plus utilisées par les chercheurs et praticiens dans divers domaines tels que l'actuariat (voir par exemple Czado et al. (2012), Abdallah et al. (2015) et Côté et al. (2016)), la finance (voir par exemple Brechmann et al. (2012), Dissmann et al. (2013) and Shi et al. (2017)), biostatistique, etc.

La notion de copule a été introduite par Sklar (1959). Une copule C de dimension n est la fonction de répartition multivariée sur $[0,1]^n$ d'un vecteur de v.a. $\mathbf{U} = (U_1, \ldots, U_n)$ dont toutes les composantes U_i $(i = 1, \ldots, n)$ obéissent à la loi uniforme (0,1). On a

$$C(u_1, \ldots, u_n) = \Pr(U_1 \leq u_1, \ldots, U_n \leq u_n), \qquad (u_1, \ldots, u_n) \in [0,1]^n.$$

Définition 8 *Une copule $C(u_1, \ldots, u_n)$ est une application $[0,1]^n \to [0,1]$ ayant les propriétés suivantes (celles d'une fonction de répartition conjointe) :*

- *$C(u_1, \ldots, u_n)$ est non décroissante sur $[0,1]^n$;*
- *$C(u_1, \ldots, u_n)$ est continue à droite sur $[0,1]^n$;*
- *Si au moins une composantes u_i est égale à 0, alors $\lim_{u_i \longrightarrow 0} C(u_1, \ldots, u_n) = 0$, pour $i = 1, \ldots, n$;*
- *Si toutes les composantes, sauf u_i, sont égales à 1, alors $\lim_{u_j \longrightarrow 1, j \neq i} C(u_1, \ldots, u_n) = u_i$;*
- *Inégalité de l'hyperrectangle : pour tout $a_1 \leq b_1$, ..., $a_n \leq b_n$, on a*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1 + \cdots + i_n} C(u_{1,i_1}, \ldots, u_{n,i_n}) \geq 0,$$

avec $u_{j,1} = a_j$ et $u_{j,2} = b_j$ pour $j \in \{1, 2, \ldots, n\}$.

Dans la Définition 8, la propriété de l'inégalité de l'hyperrectangle correspond à :

$$\begin{aligned} \sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1 + \cdots + i_n} C(u_{1,i_1}, \ldots, u_{n,i_n}) &= \Pr\left(\bigcap_{i=1}^{n} \{U_i \in (a_i, b_i]\} \right) \\ &= \Pr(\mathbf{U} \in (a_1, b_1] \times \cdots \times (a_n, b_n]) \geq 0, \end{aligned}$$

ce qui signifie que C est une copule à la condition qu'elle place une masse positive sur tout hyperrectangle $(a_1,b_1] \times \cdots \times (a_n,b_n]$ inclus dans $[0,1]^n$.

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un vecteur aléatoire. Dans le cas d'indépendance entre les composantes de $\mathbf{X}$, la fonction de répartition conjointe $F_{\mathbf{X}}(x_1, \ldots, x_n)$ est égale au produit des fonctions de répartition marginales $F_i(x_i)$, pour $i = 1, \ldots, n$ (voir Eq. (3)). Lorsque les variables aléatoires $X_1, \ldots, X_n$ sont dépendantes, la fonction de répartition conjointe $F_{\mathbf{X}}(x_1, \ldots, x_n)$ est différente du produit des marginales et il est souhaitable de pouvoir la décrire en fonction des marginales $F_i(x_i)$, $i = 1, \ldots, n$, pour voir clairement la différence entre $F_{\mathbf{X}}(x_1, \ldots, x_n)$ et le cas indépendant. Ceci est possible grâce au Théorème de Sklar qui constitue le fondement de la théorie des copules (voir par exemple Joe (1997) and Nelsen (2007)). Ce théorème comporte deux volets. Le premier volet permet l'identification d'une copule C à partir d'une fonction de répartition d'une loi multivariée, et le deuxième volet permet la construction d'une loi multivariée à partir d'une copule C et de marginales.

Théorème 1 ***Théorème de Sklar***

Soit $F_{\mathbf{X}} \in \Gamma(F_1, \ldots, F_n)$ une fonction de répartition multivariée ayant des fonctions de répartition marginales F_1, ..., F_n. Alors, il existe une copule C telle que pour tout $(x_1, \ldots, x_n) \in \mathbb{R}^n$

$$F_{\mathbf{X}}(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)). \tag{13}$$

Si $F_1, \ldots, F_n$ sont continues, alors C est unique. Sinon, C est uniquement déterminée sur les rangs des marginales $RanF_1 \times \cdots \times RanF_n$. Inversement, si C est une copule et $F_1, \ldots, F_n$ sont des fonctions de répartition, alors la fonction définie par (13) est une fonction de répartition multivariée avec les fonctions de répartition marginales $F_1, \ldots, F_n$.

Soit $F_{\mathbf{X}} \in \Gamma(F_1, \ldots, F_n)$, dont les marginales $F_1, \ldots, F_n$ sont continues. Alors, la copule C associée à $F_{\mathbf{X}}$ selon (13) est donnée par

$$C(u_1, \ldots, u_n) = F_{\mathbf{X}}\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right).$$

Cette approche de construction des copules est dite méthode par inversion. Si la copule C est continue, sa fonction de densité est

$$c(u_1, \ldots, u_n) = \frac{\partial^n}{\partial u_1 \ldots \partial u_n} C(u_1, \ldots, u_n).$$

En se basant sur la seconde partie du Théorème de Sklar, une copule permet de combiner différentes lois marginales afin de créer une loi multivariée. Cette approche élargit la gamme des modèles de dépendance disponibles et permet une modélisation très flexible des risques. En effet, la copule C décrit uniquement la relation de dépendance entre les v.a. $X_1, \ldots, X_n$ alors

que les marginales décrivent le comportement de chacune des variables. Cet effet de dissociation constitue une caractéristique forte des copules et l'une des raisons de leur popularité.

Soit $\mathbf{U} = (U_1, \ldots, U_n)$ un vecteur de v.a. uniformes sur (0,1) avec fonction de répartition $F_{\mathbf{U}} = C$. La copule pour le vecteur $(1 - U_1, \ldots, 1 - U_n)$ est appelée copule de survie et notée $\hat{C}$. À partir de cette dernière et des fonctions de survie marginales $\overline{F}_i$, $i = 1, \ldots, n$, il est possible de construire une fonction de survie multivariée telle que

$$\overline{F}_X(x_1, \ldots, x_n) = \hat{C}\left(\overline{F}_1(x_1), \ldots, \overline{F}_n(x_n)\right).$$

Inversement, si la fonction de survie conjointe $\overline{F}_{\mathbf{X}}$ est continue, avec fonctions de survie marginales $\overline{F}_i$, $i = 1, \ldots, n$ et ses inverses $\overline{F}_i^{-1}$ continues à gauche, on déduit la copule $\hat{C}$

$$\hat{C}(u_1, \ldots, u_n) = \overline{F}_{\mathbf{X}}\left(\overline{F}_1^{-1}(u_1), \ldots, \overline{F}_n^{-1}(u_n)\right), \tag{14}$$

pour $(u_1, \ldots, u_n) \in (0,1)^n$ (voir par exemple Joe, 1997). La copule de survie $\hat{C}$ est une copule et vérifie les propriétés de la Définition 8. L'Eq. (14) est utilisée pour la construction d'une nouvelle famille de copules dans le Chapitre 3.

Les copules étant des fonctions de répartition multivariées, elles appartiennent à la classe de Fréchet avec marginales $Unif(0,1)$ et les bornes de Fréchet (4) s'appliquent également. Pour une copule C de dimension n, on a

$$W_n(u_1, \ldots, u_n) \leq C(u_1, \ldots, u_n) \leq M_n(u_1, \ldots, u_n),$$

avec $W_n(u_1, \ldots, u_n) = \max\{u_1 + \cdots + u_n - n + 1; 0\}$ et $M_n(x_1, \ldots, x_n) = \min\{u_1, \ldots, u_n\}$. Pour tout $n \geq 2$, la borne supérieure M_n est appelée la copule borne supérieure de Fréchet, notée $M_n = C^+$, et correspond à la fonction de répartition d'un vecteur $\mathbf{U}^+$ dont les n composantes sont comonotones et suivent la loi $Unif(0,1)$. Pour $n = 2$ seulement, la borne inférieure W_2 correspond à la copule borne inférieure de Fréchet pour la paire antimonotone (U_1^-, U_2^-) et on note $W_2 = C^-$.

Les copules bornes de Fréchet constituent des copules de base en plus de la copule indépendance définie par $C^I(u_1, ..., u_n) = u_1 \times \cdots \times u_n$. Cette dernière est la fonction de répartition de v.a. indépendantes et uniformes sur (0,1). Selon l'Eq (3), la copule C^I est le produit des marginales de loi $Unif(0,1)$. L'Illustration 0.8 présente les nuages de points pour des simulations d'une paire (U_1, U_2) avec fonction de répartition, soit C^I, C^+ ou C^-. La copule C^+ correspond à une unité de masse répartie sur la diagonale principale $u_1 = u_2$, tandis que le support de la copule C^- est la diagonale secondaire $u_1 = 1 - u_2$.

Étant donné que les mesures de concordance tau de Kendall et rhô de Spearman sont des mesures de corrélation de rang, il est possible de les exprimer uniquement en fonction de la copule associée à la fonction de répartition de la paire de v.a. (X_1, X_2). En se basant sur la

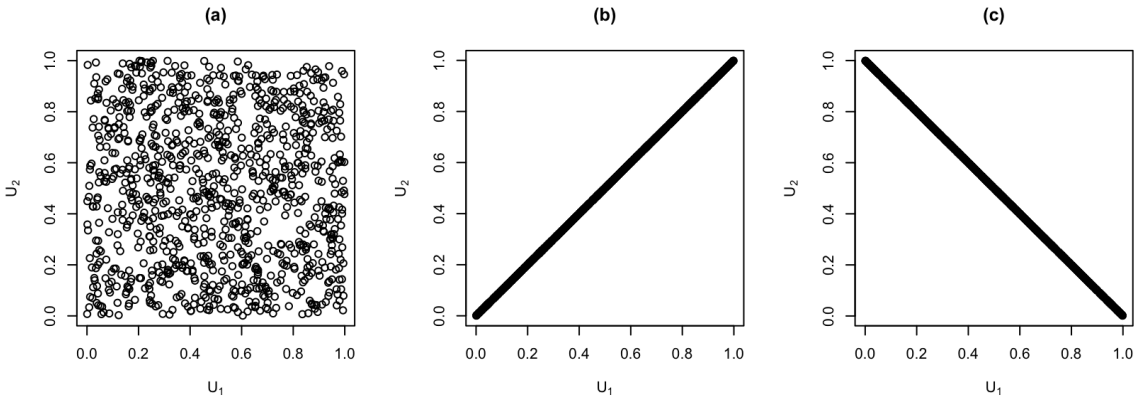


Illustration 0.8 – Les graphiques (a), (b) et (c) reproduisent 1000 réalisations de (U_1,U_2) provenant de la copule indépendance, bornes supérieure et inférieure de Fréchet

Définition 5 et en posant $U_1 = F_1(X_1)$ et $U_2 = F_2(X_2)$, on obtient

$$\begin{aligned}\rho_s(X_1,X_2) &= 12\int_0^1\int_0^1 u_1u_2\mathrm{d}C(u_1,u_2) - 3\\ &= 12E[U_1U_2] - 3.\end{aligned}$$

Pour le tau de Kendall, en développant l'expression fournie dans la Définition 4, on obtient

$$\begin{aligned}\tau(X_1,X_2) &= 2\Pr(\text{concordance}) - 1\\ &= 2\left(\Pr\left(X_1 > X_1',X_2 > X_2'\right) + \Pr\left(X_1 \le X_1',X_2 \le X_2'\right)\right) - 1\\ &= 4\int_0^1\int_0^1 C(u_1,u_2)\,\mathrm{d}C(u_1,u_2) - 1\\ &= 4E[C(U_1,U_2)] - 1.\end{aligned}$$

En plus de la mesure d'association pour une paire $(U_1,U_2) \sim C$, il est également intéressant de pouvoir comparer les forces de dépendance entre différentes copules. La notion d'ordre de concordance permet de comparer des distributions multivariées appartenant à la même classe de Fréchet, c.-à-d. avec le même ensemble de marginales univariées, afin d'identifier laquelle des distributions multivariées représente le plus de dépendance.

Définition 9 *Soit C_1 and C_2 deux copules de dimension n avec, respectivement, tau de Kendall τ_1, τ_2, et rhô de Spearman $\rho_s^{(1)}$, $\rho_s^{(2)}$. C_1 est dite moins concordante que C_2, noté $C_1 \prec_c C_2$, si*

$$C_1(\mathbf{u}) \le C_2(\mathbf{u}) \quad \textit{and} \quad \overline{C}_1(\mathbf{u}) \le \overline{C}_2(\mathbf{u}),$$

avec $\mathbf{u} \in [0,1]^n$ et $\overline{C}_i$ la fonction de survie conjointe associée à C_i, pour $i = 1,2$. Alors, cela implique que $\tau_1 \le \tau_2$ et $\rho_s^{(1)} \le \rho_s^{(2)}$.

Dans le cas bivarié, on a

$$\begin{aligned} \overline{C}_1(u_1,u_2) &\leq \overline{C}_2(u_1,u_2) \\ 1-u_1-u_2+C_1(u_1,u_2) &\leq 1-u_1-u_2+C_2(u_1,u_2) \\ C_1(u_1,u_2) &\leq C_2(u_1,u_2). \end{aligned}$$

Il en résulte qu'une seule des inégalités de la Définition 9 doit être vérifiée pour établir l'ordre de concordance entre les copules bivariées. Par sa relation directe avec les mesures de corrélation de rang, l'ordre de concordance permet d'établir un ordre de dominance basé sur la force de dépendance entre les copules (voir Joe, 1997). Les mesures de corrélation de rang ainsi que l'ordre de concordance sont utilisés dans la procédure d'identification de la structure hiérarchique de la famille de copules introduite dans le Chapitre 3.

0.5.2 Copules archimédiennes

Les copules archimédiennes constituent une importante classe de copules. Elles se caractérisent par leur simple construction, l'extension facile en grande dimension et la capacité de capturer diverses structures de dépendance. Une copule C de dimension n est dite archimédienne si elle s'écrit sous la forme

$$C(u_1,\ldots,u_n) = \psi\left(\psi^{-1}(u_1)+\cdots+\psi^{-1}(u_n)\right), \tag{15}$$

pour $(u_1,\ldots,u_n) \in [0,1]^n$ où la fonction $\psi : [0,\infty) \to [0,1]$ est appelée le générateur de la copule et satisfait les propriétés suivantes :

- $\psi(0) = 1$ et $\lim_{x\to\infty} \psi(x) = 0$;
- ψ est continue et strictement décroissante sur $[0,\infty)$;
- $\psi^{-1} : [0,1] \to [0,\infty)$ est la fonction inverse avec $\psi^{-1}(x) = \inf\{u,\ \psi(u) \leq x\}$.

En se référant à Kimberling (1974), l'expression (15) est une copule en toute dimension $n \geq 2$ si et seulement si la fonction ψ est complètement monotone, telle que pour $k \in \mathbb{N}$ et $\forall x \in [0,\infty)$

$$(-1)^k \frac{\mathrm{d}^k \psi(x)}{\mathrm{d}x^k} \geq 0.$$

Théorème 2 ***Théorème de Bernstein***

Une fonction $\psi : [0,\infty) \to [0,1]$ est dite complètement monotone si et seulement si elle est la TLS d'une v.a. strictement positive Θ, tel que $\psi = \mathcal{L}_\Theta$.

Selon le Théorème de Bernstein (voir Feller, 1971), ψ est un générateur de copule archimédienne si et seulement si elle est la TLS d'une v.a. strictement positive Θ et on peut réécrire l'expression (15) comme suit :

$$C(u_1,\ldots,u_n) = \mathcal{L}_\Theta\left(\mathcal{L}_\Theta^{-1}(u_1)+\cdots+\mathcal{L}_\Theta^{-1}(u_n)\right), \tag{16}$$

où $\mathcal{L}_{\Theta}^{-1}$ est l'inverse de la TLS de Θ.

Les copules archimédiennes sont des exemples de distributions multivariées obtenues à partir de lois mélanges selon l'approche de Marshall and Olkin (1988). Soit $\mathbf{Y} = (Y_1, \ldots, Y_n)$ un vecteur aléatoire et Θ une v.a. strictement positive avec fonction de répartition F_{Θ} et TLS $\mathcal{L}_{\Theta}$. La v.a. Θ peut être discrète ou continue. Sachant la v.a. mélange (facteur commun) Θ, on suppose que les composantes du vecteur aléatoire $\mathbf{Y}$ sont conditionnellement indépendantes et identiquement distribuées, telles que $(Y_i|\Theta = \theta) \sim Exp(\theta)$, pour $i = 1, \ldots, n$. Par conséquent, $\mathbf{Y}$ a une distribution exponentielle multivariée mixte. La fonction de survie de la v.a. Y_i est

$$\overline{F}_{Y_i}(x_i) = E\left[\overline{F}_{Y_i|\Theta}(x_i)\right] = E\left[\mathrm{e}^{-\Theta x_i}\right] = \mathcal{L}_{\Theta}(x_i), \qquad i = 1, \ldots, n,$$

et la fonction de survie conjointe de $\mathbf{Y}$ est

$$\begin{aligned}
\overline{F}_{\mathbf{Y}}(x_1, \ldots, x_n) &= \int_0^{\infty} \overline{F}_{\mathbf{Y}|\Theta=\theta}(x_1, \ldots, x_n)\, \mathrm{d}F_{\Theta}(\theta) \\
&= \int_0^{\infty} \overline{F}_{Y_1|\Theta=\theta}(x_1) \times \cdots \times \overline{F}_{Y_n|\Theta=\theta}(x_n)\, \mathrm{d}F_{\Theta}(\theta) \\
&= \int_0^{\infty} \mathrm{e}^{-\theta x_1} \times \cdots \times \mathrm{e}^{-\theta x_n} \mathrm{d}F_{\Theta}(\theta) \\
&= \mathcal{L}_{\Theta}(x_1 + \cdots + x_n).
\end{aligned}$$

En appliquant le Théorème de Sklar et en se référant à l'expression (14), la copule associée à la fonction de survie $\overline{F}_{\mathbf{Y}}$ est

$$\begin{aligned}
C(u_1, \ldots, u_n) &= \overline{F}_{\mathbf{Y}}\left(\overline{F}_{Y_1}^{-1}(u_1), \ldots, \overline{F}_{Y_n}^{-1}(u_n)\right) \\
&= \mathcal{L}_{\Theta}\left(\mathcal{L}_{\Theta}^{-1}(u_1) + \cdots + \mathcal{L}_{\Theta}^{-1}(u_n)\right),
\end{aligned}$$

pour $(u_1, \ldots, u_n) \in (0,1)^n$.

Cette approche stochastique de construction des copules archimédiennes est la base de l'algorithme populaire de simulation de Marshall & Olkin (voir par exemple Marshall and Olkin, 1988; Hofert, 2008). Le point de départ de cet algorithme est la simulation d'une réalisation de la v.a. Θ. Dans le cas où la distribution de cette v.a. n'est pas connue, il est possible d'utiliser une procédure d'inversion numérique de sa TLS pour générer une simulation (voir par exemple Hofert, 2008). Il existe également d'autres approches de simulation, telles que la méthode conditionnelle basée sur l'inverse de la copule conditionnelle (voir par exemple Hofert, 2010).

Parmi les copules archimédiennes les plus utilisées dans la littérature et ayant une expression explicite, on retrouve les copules Ali-Mikhail-Haq, Frank, Joe, Gumbel et Clayton. Le Tableau 4.1 dans Nelsen (2007) liste les expressions et les générateurs de ces copules. Plusieurs chercheurs ont proposé des packages sur R avec des fonctions relatives aux calculs de ces copules (voir par exemple Hofert et al., 2014). L'Illustration 0.9 présente le nuage de points

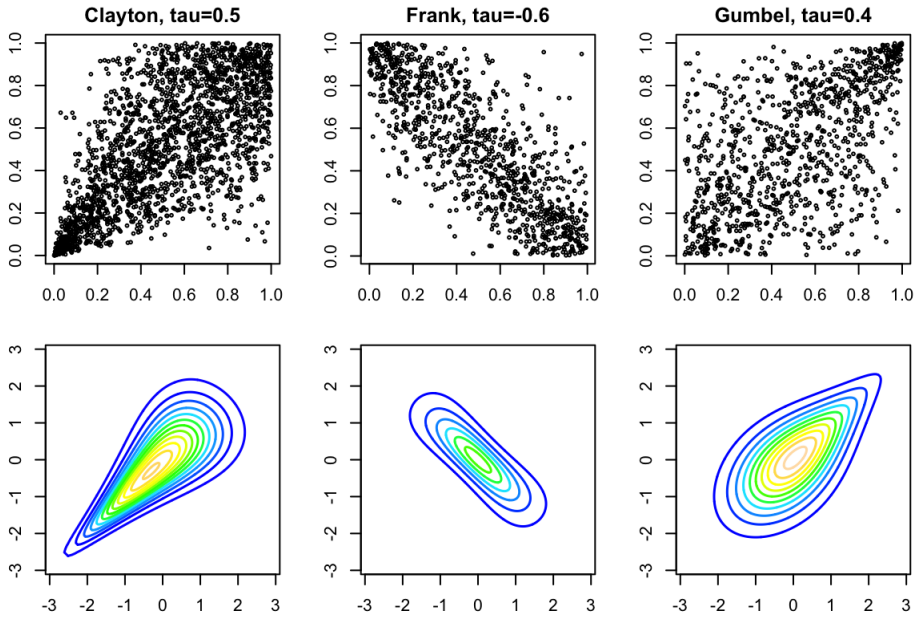


Illustration 0.9 – Ligne du haut : nuages de points de 2000 simulations, pour les copules Clayton, Frank et Gumbel, et ligne du bas : contour de densité normalisée

ainsi que le contour de densité pour des échantillons simulés de trois exemples de copules archimédiennes dont les formes de dépendance sont différentes.

Selon l'expression générale des copules archimédiennes (15), elles sont échangeables, c.-à-d. pour une permutation arbitraire σ de $\{1,\ldots,n\}$, on a $C(u_1,\ldots,u_n) = C(u_{\sigma_1},\ldots,u_{\sigma_n})$. À l'aide l'Illustration 0.9, on peut constater la symétrie des copules illustrées où $C(u_1,u_2) = C(u_2,u_1)$. Malgré leurs nombreuses propriétés, cette caractéristique de symétrie constitue un inconvénient et rend ces copules peu utiles dans des contextes plus pratiques. En effet, ces copules impliquent que la structure de dépendance entre toutes les paires est la même. Ceci n'est pas toujours le cas dans les portefeuilles d'assurance constitués de risques hétérogènes.

0.5.3 Copules non échangeables

En utilisant différentes techniques de construction, plusieurs copules non échangeables ont été proposées dans la littérature. McNeil and Nešlehová (2010) a introduit les copules de Liouville construites à partir de distributions de survie multivariées de Liouville. Les copules bivariées Archimax, proposées par Capéraà et al. (2000), peuvent également créer une dépendance asymétrique. Elles comprennent à la fois les copules de valeurs extrêmes ainsi que les copules archimédiennes comme cas particulier. Une extension multivariée des copules Archimax a été présentée par Mesiar and Jagr (2013). Nous référons Genest and Nešlehová (2013) pour une revue sur la construction de structures de dépendance asymétriques bivariées, les mesures d'asymétrie ainsi que les tests de symétrie.

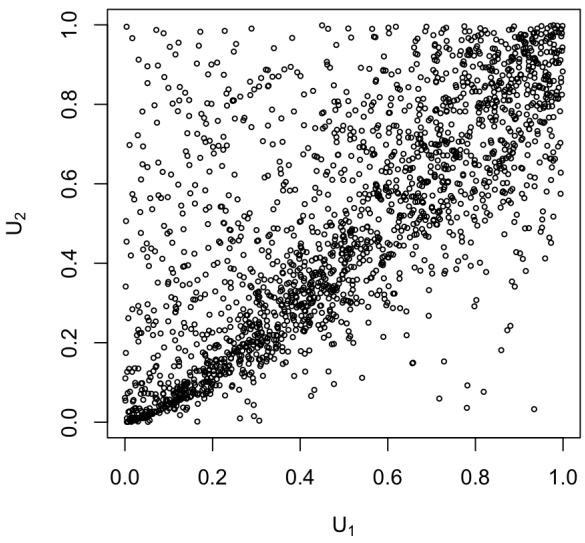


Illustration 0.10 – Nuage de points de 2000 simulations de la paire (U_1,U_2) distribuée selon la copule asymétrique $C_{0.95,0.7}$ avec $C_1(u_1,u_2) = u_1u_2$ et C_2 la copule Clayton avec paramètre $\alpha = 8$

Certaines copules non échangeables ne sont pas facilement utilisées en pratique puisqu'elles peuvent nécessiter des fonctions supplémentaires pour caractériser leurs dépendances complexes. Par exemple, pour une copule Archimax on a besoin de la fonction de dépendance de Pickands pour sa construction. Cette fonction nécessite parfois des dérivations statistiques complexes. Parmi les techniques de construction les plus populaires et les plus pratiques, on retrouve la méthode de produits proposée par Khoudraji (1995). Cette approche de génération de copules asymétriques bivariées est basée sur le biais de produits de copules (non nécessairement archimédiennes). Soit deux copules bivariées C_1 et C_2. Pour $\delta_1,\delta_2 \in [0,1]$, on a

$$C_{\delta_1,\delta_2}(u_1,u_2) = C_1\left(u_1^{1-\delta_1},u_2^{1-\delta_2}\right) C_2\left(u_1^{\delta_1},u_2^{\delta_2}\right),$$

pour $u_1,u_2 \in [0,1]$. Si $C_1(u_1,u_2) = u_1u_2$ est la copule d'indépendance, on obtient

$$C_{\delta_1,\delta_2}(u_1,u_2) = u_1^{1-\delta_1}u_2^{1-\delta_2}C_2\left(u_1^{\delta_1},u_2^{\delta_2}\right). \tag{17}$$

La copule C_{δ_1,δ_2} est asymétrique si $\delta_1 \neq \delta_2$. Si $\delta_1 = \delta_2 = 1$, C_{δ_1,δ_2} correspond à la copule C_2, et si $\delta_1 = \delta_2 = 0$, C_{δ_1,δ_2} est tout simplement la copule d'indépendance. Une extension multivariée de cette procédure multiplicative de Khoudraji (1995) a été proposée par Liebscher (2008).

L'Illustration 0.10 présente 2000 simulations pour une paire de v.a. distribuées selon la copule asymétrique (17), tel que $C_2(u_1,u_2) = (u_1^{-1/\alpha} + u_2^{-1/\alpha} - 1)^{-\alpha}$ est la copule archimédienne de Clayton de paramètre $\alpha = 8$. Le nuage de points occupe la partie au-dessus de la diagonale principale avec une dispersion aléatoire des points reflétant l'impact de la copule d'indépendance C_1. On observe également une forte dépendance aux valeurs extrêmes de la queue à

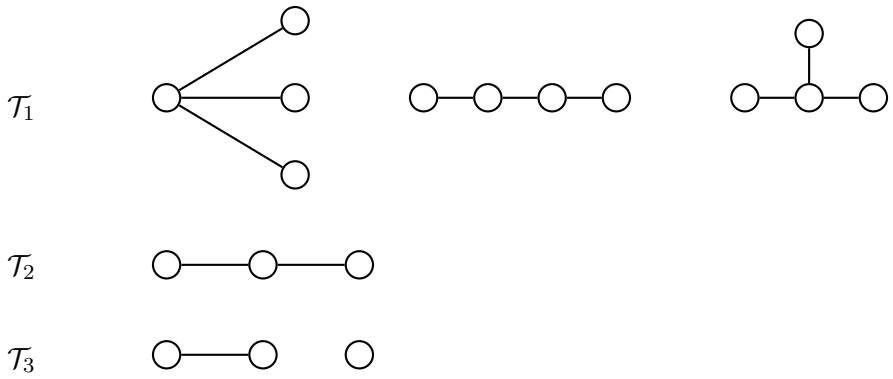


Illustration 0.11 – Exemples des dispositions possibles pour les arbres de 1er ($\mathcal{T}_1$), 2e ($\mathcal{T}_2$) et 3e ($\mathcal{T}_3$) niveaux pour la représentation d'une copule vine de dimension 4

gauche et une faible dépendance aux valeurs extrêmes de la queue à droite comme la forme d'une copule de Clayton. Dans le Chapitre 3, on introduit une nouvelle famille de copules hiérarchiques en s'intéressant à des dépendances non échangeables inter-blocs. Dans notre structure, nous observons particulièrement des produits de copules construites selon l'approche de Khoudraji (1995).

0.5.4 Copules multivariées hiérarchiques

Les familles de copules discutées dans les Sections 0.5.2 et 0.5.3 sont dites des copules à un niveau. Ces structures de dépendance ne peuvent pas être utilisées lors de la modélisation d'un portefeuille composé de plusieurs sous-groupes de risques non homogènes. Il existe d'autres types de copules multivariées capables de modéliser des dépendances plus complexes et qui gagnent de plus en plus d'intérêt dans les récentes recherches en assurance et finance.

Les copules vine, initialement introduites par Joe (1996), sont construites en se basant sur une décomposition en cascade d'une densité multivariée en un produit de densités conditionnelles de copules bivariées. Elles peuvent être représentées comme une séquence d'arbres, où à chaque nœud, une copule bivariée modélise la dépendance conditionnelle. L'Illustration 0.11 présente des exemples de différentes dispositions d'arbres pour une copule vine de dimension 4. Ces copules sont flexibles et peuvent tenir compte des asymétries et de la dépendance de queue. Cependant, elles nécessitent l'estimation de plusieurs paramètres et sont moins adaptées lorsque les données révèlent des structures de dépendance hiérarchiques. Ils existent plusieurs types de copules vine dont les plus populaires en pratique sont C-vine et D-vine. Pour plus de détails sur les copules vine, voir par exemple, Joe and Kurowicka (2011) and Czado (2019).

Les copules archimédiennes hiérarchiques représentent une extension des copules archimédiennes. Joe (1997) a proposé une première approche de construction de ces copules nommée copules archimédiennes imbriquées, obtenues par l'imbrication de copules archimédiennes les unes dans les autres. Considérons une copule archimédienne imbriquée de dimension 4,

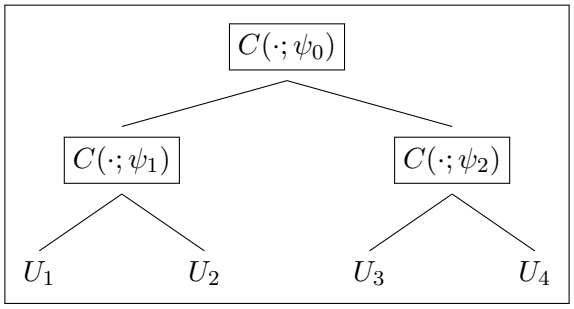


Illustration 0.12 – Représentation arborescente de la copule archimédienne hiérarchique (18)

construite par le biais des générateurs ψ_0, ψ_1 et ψ_2, et dont l'expression est

$$\begin{aligned} &C(u_1,u_2,u_3,u_4;\psi_0,\psi_1,\psi_2) \\ &= C\left(C(u_1,u_2;\psi_1),C(u_3,u_4;\psi_2);\psi_0\right) \\ &= \psi_0\left(\psi_0^{-1}\circ\psi_1\left(\psi_1^{-1}(u_1)+\psi_1^{-1}(u_2)\right)+\psi_0^{-1}\circ\psi_2\left(\psi_2^{-1}(u_3)+\psi_2^{-1}(u_4)\right)\right), \end{aligned} \tag{18}$$

où $C(\cdot;\psi)$ est une copule archimédienne avec générateur ψ. (18) est une copule imbriquée si et seulement si $(\psi_0^{-1}\circ\psi_1)'$ et $(\psi_0^{-1}\circ\psi_2)'$ sont complètement monotones. Comme l'a souligné Hofert (2010), la vérification de cette condition suffisante d'imbrication n'est pas toujours facile lorsque les générateurs appartiennent à des familles de copules archimédiennes différentes. Ceci restreint la flexibilité de ces structures. En outre, la copule hiérarchique (18) peut être représentée sous la forme d'un arbre (voir Hofert, 2012). L'Illustration 0.12 présente la structure arborescente de la copule (18). Cette représentation est inspirante pour plusieurs techniques de construction récentes des copules hiérarchiques et qui ont réussi à se passer de la nécessité de vérifier la condition d'imbrication de dépendance. Parmi ces techniques, on s'intéresse à celle basée sur une distribution de mélange exponentielle multivariée.

Soit $\boldsymbol{\Theta}=(\Theta_1,\ldots,\Theta_n)$ un vecteur de v.a. positives avec TLS conjointe $\mathcal{L}_{\boldsymbol{\Theta}}$ et TLS univariée $\mathcal{L}_{\Theta_i}$, $i=1,\ldots,n$. On définit $\mathbf{Y}=(\mathbf{Y}_1,\ldots,\mathbf{Y}_n)$, avec $\mathbf{Y}_i=(Y_{i,1},\ldots,Y_{i,n_i})$, pour $i=1,\ldots,n$. On suppose que les composantes du vecteur $\mathbf{Y}$ sont conditionnellement indépendantes et que la distribution conditionnelle de $(\mathbf{Y}_i|\boldsymbol{\Theta})$ est influencée uniquement par Θ_i, avec $(Y_{i,1}|\Theta_i)\perp\cdots\perp Y_{i,n_i}|\Theta_i)$ et $(Y_{i,j}|\Theta_i)\sim Exp(\Theta_i)$, pour $j=1,\ldots,n_i$ et $i=1,\ldots,n$. On obtient

$$\begin{aligned} \overline{F}_{\mathbf{Y}}(\mathbf{y}) &= \mathrm{E}\left[\overline{F}_{\mathbf{Y}|\Theta_1,\ldots,\Theta_1}(\mathbf{y})\right] \\ &= \mathrm{E}\left[\prod_{i=1}^{n}\prod_{j=1}^{n_i}\overline{F}_{Y_{i,j}|\Theta_i}(y_{i,j})\right] \\ &= \mathrm{E}\left[\prod_{i=1}^{n}\prod_{j=1}^{n_i}\exp\left(-y_{i,j}\Theta_i\right)\right] \\ &= \mathcal{L}_{\boldsymbol{\Theta}}\left(\sum_{j=1}^{n_1}y_{1,j},\ldots,\sum_{j=1}^{n_n}y_{n,j}\right), \end{aligned} \tag{19}$$

où $\mathbf{y} = (y_{1,1}, \ldots, y_{1,n_1}, \ldots, y_{n,1}, \ldots, y_{n,n_n})$. En utilisant le Théorème de Sklar ainsi que l'expression (14), on déduit la copule associée à la fonction (19)

$$C(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \mathcal{L}_{\Theta_1,\ldots,\Theta_d}\left(\sum_{j=1}^{n_1} \mathcal{L}_{\Theta_1}^{-1}(u_{1,j}), \ldots, \sum_{j=1}^{n_n} \mathcal{L}_{\Theta_n}^{-1}(u_{n,j})\right), \tag{20}$$

où $\mathbf{u}_i = (u_{i,1}, \ldots, u_{i,n_i})$, pour $i = 1, \ldots, n$. Hering et al. (2010) et Cossette et al. (2017) se sont basés sur cette approche de construction pour définir de nouvelles familles de copules archimédiennes hiérarchiques. La différence entre ces articles est la distribution du vecteur commun $\boldsymbol{\Theta}$. Hering et al. (2010) définit $\boldsymbol{\Theta}$ via des subordonnées de Lévy, tandis que Cossette et al. (2017) utilise des lois composées.. Une comparaison détaillée de ces deux approches de construction est discutée dans la Section 4.2 de Cossette et al. (2017).

En s'inspirant de ces travaux, dans le Chapitre 3 on introduit une nouvelle famille de copules hiérarchiques dont le vecteur commun est construit selon une arborescence de concaténation de v.a. indépendantes. Cette structure permet d'obtenir des relations archimédiennes intra-groupes et des copules non échangeables de type (17) inter-groupes.

0.6 Aperçu de la thèse

Ce projet de thèse comporte trois articles scientifiques présentés dans les Chapitres 1 à 3 et qui constituent des contributions reliées aux opérations de provisionnement, solvabilité et allocation du capital. Dans cette section, on fournit un aperçu global du contenu de ces articles.

0.6.1 Chapitre 1

Comme expliqué dans la Section 0.1.2, l'exercice de provisionnement est très important pour assurer la solvabilité d'une compagnie d'assurance. Les modèles de provisionnement au niveau micro représentent une alternative plus précise des modèles traditionnels macro. Avec l'avènement de nouvelles techniques de collecte, de stockage et d'analyse des données, les chercheurs et praticiens s'intéressent de en plus en plus au développement d'approches de provisionnement granulaire et l'incorporation des données individuelles de sinistres.

Dans le Chapitre 1, on présente un nouveau modèle de provisionnement individuel à temps discret. En s'inspirant des récentes recherches discutées dans la Section 0.2.3, on s'intéresse à développer un modèle plus flexible qui incorpore plusieurs informations individuelles à la fois statiques et dynamiques. Pour ceci, on propose d'utiliser un RNR (voir Section 0.2.1), plus précisément un LSTM. Pour chaque période de développement d'un sinistre, notre réseau effectue deux tâches : une classification pour prédire la probabilité d'avoir un paiement, et une régression pour prédire le montant à payer s'il y a lieu.

On développe l'architecture de notre réseau en considérant une structure générale pour une base de données détaillées de sinistres génériques. À la date d'évaluation, les sinistres ont des

développements observés de durées différentes. Grâce à notre processus de préparation des données, le réseau est en mesure de faire des prédictions pour des sinistre à développements différents. Il permet d'obtenir des prédictions pour les paiements futurs sur des intervalles de temps fixes et en considérant une durée de développement maximale.

Pour tester la flexibilité et la performance de notre modèle, on implémente deux études de cas. La première est basée sur des données simulées obtenues à partir du générateur de Gabrielli and Wüthrich (2018). La deuxième étude utilise des données réelles d'une grande compagnie d'assurance canadienne, avec l'observation de paiements larges. Ces derniers sont censurés dans un premier temps pour réduire la variance des observations et booster l'apprentissage du modèle. Pour chaque étude, on choisit les hyperparamètres adéquats pour l'entrainement du réseau. En se basant sur les résultats obtenus pour les bases de données de test, notre modèle est plus précis en comparaison avec le modèle agrégé chain-ladder. Nos comparaisons sont basées sur les ratios de la réserve et de l'ultime. On analyse également la précision des résultats au niveau individuel.

Pour les paiements larges, on propose un ajustement des prédictions obtenues par notre réseau entrainé uniquement sur les paiements censurés. En se basant sur les paiements non censurés, on identifie le seuil ainsi que les paramètres de la loi Pareto généralisée modélisant l'excès de paiement. On modélise également la probabilité d'avoir un excès en utilisant un modèle linéaire généralisé. On illustre l'implémentation de cette approche sur les données réelles de la deuxième étude de cas.

0.6.2 Chapitre 2

En absence d'information sur la distribution de la perte totale S, l'alternative est de borner les mesures de risque associées en considérant les structures de dépendance extrême. Pour un portefeuille à deux risques, on présente dans la Section 0.4 les structures de dépendance positive (comonotonocité) et négative (antimonotonocité) extrême et leurs importantes propriétés. Pour la somme de v.a. comonotones, la représentation stochastique (5) permet de déduire facilement des expressions explicites pour diverses fonctions dont les mesures de risque VaR et $TVaR$. En contrepartie, pour la somme S^- d'une paire de v.a. antimonotones, la représentation (10) dépend des marginales des risques individuels. Dans notre projet, on s'intéresse à étudier comment la monotonocité de la fonction φ^- dépend de la structure de dépendance et du comportement de la queue des marginales.

L'objectif principal du Chapitre 2 est de développer des expressions explicites pour les mesures de risque de S^-. On considère trois familles de distributions : les distributions symétriques et absolument continues, les distributions unimodales identiques avec support sur $\mathbb{R}^+$ et les distributions unimodales avec support sur $\mathbb{R}^+$. Pour chaque distribution, on analyse la forme de la fonction φ^- (convexe ou concave) afin de développer des expressions explicites pour

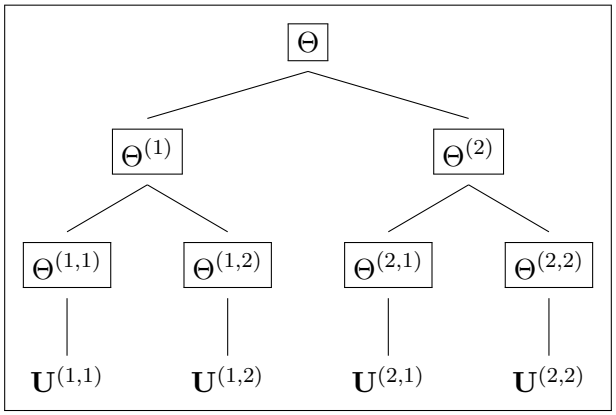


Illustration 0.13 – Exemple de structure d'arborescence à deux niveaux de la famille de copules introduite dans le Chapitre 3

$VaR_\kappa(S^-)$ et $TVaR_\kappa(S^-)$, pour $\kappa \in (0,1)$. À partir de ces expressions, on analyse la différence entre la valeur minimale (cas d'antimonotonocité) et maximale (cas de comonotonocité) de la mesure $TVaR$ de la perte agrégée sur un espace de Fréchet $\Gamma(F_1,F_2)$. En s'intéresse particulièrement aux risques marginaux distribués selon la loi Pareto. Cette dernière est une distribution subexponentielle populaire, qui joue un rôle crucial en actuariat et en gestion quantitative des risques.

0.6.3 Chapitre 3

Bien qu'il existe plusieurs copules hiérarchiques capables de modéliser des structures de dépendance complexes, elles ne peuvent généralement pas être utilisées pour des données non homogènes avec des dépendances non échangeables. À titre d'exemple, par construction, les copules bivariées au sein d'une copule archimédienne hiérarchique sont nécessairement symétriques. Ainsi, les copules introduites par Hering et al. (2010) et Cossette et al. (2017) permettent uniquement la modélisation de dépendance symétrique.

Avec l'objectif de contribuer à la modélisation adéquate des risques dépendants, dans le Chapitre 3, on définit une nouvelle famille de copules hiérarchiques potentiellement utile pour des applications pratiques. En s'inspirant de la construction (20), on définit les composantes de notre vecteur de choc commun $\boldsymbol{\Theta}$ par une concaténation descendante de v.a. positives indépendantes et appartenant à la même famille de subordinateurs de Lévy, tel que

$$\mathcal{L}_N(t) = \exp\left(-\alpha\varphi(t)\right), \quad t \geq 0. \tag{21}$$

avec $\varphi(t)$ l'exposant de Laplace. Cette condition permet de déduire des expressions explicites pour la copule conjointe ainsi que les copules inter et intra-groupes. L'Illustration 0.13 schématise un exemple de construction du vecteur commun $\boldsymbol{\Theta} = \left(\Theta^{(1,1)},\Theta^{(1,2)},\Theta^{(2,1)},\Theta^{(2,2)}\right)$ via des concaténations descendantes de v.a. indépendantes selon les différents chemins de l'arbre à

deux niveaux. Cette représentation hiérarchique ressemble à l'Illustration 0.12. Grâce à cette construction, les copules intra-groupe sont des copules archimédienne symétriques, tandis que les copules inter-groupes sont non échangeables et de la forme (17).

Nous étudions diverses propriétés de cette nouvelle famille de copules dont l'ordre de concordance. Ce dernier nous permet de vérifier la stabilité de l'ordre et la capacité d'identifier la structure en se basant sur des mesures de corrélation de rangs telle que le tau de Kendall. Pour une structure à deux niveaux, nous proposons une procédure d'identification de l'arbre ainsi qu'un algorithme d'estimation des paramètres basé sur la vraisemblance composite. Ces contributions permettent de valoriser davantage cette nouvelle famille de copules.

Afin de vérifier l'efficacité et la précision des processus d'identification de la structure et d'estimation des paramètres, on élabore plusieurs exemples avec données simulées et réelles. L'étude avec données simulées permet de vérifier la précision des algorithmes en se comparant à la vraie structure et aux valeurs réelles des paramètres. Pour les deux études de cas avec données réelles, on compare la structure d'arborescence identifiée avec l'analyse préliminaire des dépendances entre les variables (symétrie ou non, forme et force de dépendance, etc.). On compare également l'efficacité de notre copule à capter la structure de dépendance des données versus une copule vine ou une copule archimédienne hiérarchique.

0.7 Bibliographie

Abdallah, A., Boucher, J.-P., and Cossette, H. (2015). Modeling dependence between loss triangles with hierarchical archimedean copulas. *ASTIN Bulletin*, 45(3) :577–599.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3) :203–228.

Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1) :68–85.

Capéraà, P., Fougères, A.-L., and Genest, C. (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis*, 72(1) :30–49.

Charpentier, A. and Pigeon, M. (2016). Macro vs. micro methods in non-life claims reserving (an econometric perspective). *Risks*, 4(2) :12.

Cossette, H., Gadoury, S.-P., Marceau, É., and Mtalai, I. (2017). Hierarchical Archimedean copulas through multivariate compound distributions. *Insurance : Mathematics and Economics*, 76 :1–13.

Côté, M.-P., Genest, C., and Abdallah, A. (2016). Rank-based methods for modeling dependence between loss triangles. *European Actuarial Journal*, 6(2) :377–408.

Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics, Springer*.

Czado, C., Kastenmeier, R., Brechmann, E. C., and Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, 2012(4) :278–305.

Deelstra, G., Dhaene, J., and Vanmaele, M. (2011). An overview of comonotonicity and its applications in finance and insurance. *Advanced Mathematical Methods for Finance*, pages 155–179.

Delong, Ł. and Wüthrich, M. V. (2020). Neural networks for the joint development of individual payments and claim incurred. *Risks*, 8(2) :33.

Denuit, M. and Dhaene, J. (2003). Simple characterizations of comonotonicity and countermonotonicity by extremal correlations. *Belgian Actuarial Bulletin*, 3 :22–27.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial Theory for Dependent Risks : Measures, Orders and Models*. John Wiley & Sons.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1) :40–57.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59 :52–69.

Embrechts, P. and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application*, 1 :493–514.

Feller, W. (1971). Introduction to the theory of probability and its applications, vol. 2. *II (2. Ed.) New York : Wiley*.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon,* 3^e *serie, Sciences, Sect. A*, 14 :53–77.

Gabrielli, A. (2020a). An individual claims reserving model for reported claims. *Available at SSRN 3612930*.

Gabrielli, A. (2020b). A neural network boosted double overdispersed poisson claims reserving model. *ASTIN Bulletin : The Journal of the IAA*, 50(1) :25–60.

Gabrielli, A., Richman, R., and Wüthrich, M. V. (2020). Neural network embedding of the over-dispersed poisson reserving model. *Scandinavian Actuarial Journal*, 2020(1) :1–29.

Gabrielli, A. and Wüthrich, M. (2018). An individual claims history simulation machine. *Risks*, 6(2) :29.

Genest, C. and Nešlehová, J. G. (2013). Assessing and modeling asymmetry in bivariate continuous data. In *Copulae in mathematical and quantitative finance*, pages 91–114. Springer.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv :1308.0850*.

Graves, A. and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6) :602–610.

Hering, C., Hofert, M., Mai, J.-F., and Scherer, M. (2010). Constructing hierarchical Archimedean copulas with Lévy subordinators. *Journal of Multivariate Analysis*, 101(6) :1428–1433.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8) :1735–1780.

Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics & Data Analysis*, 52(12) :5163–5174.

Hofert, M. (2010). *Sampling nested Archimedean copulas with applications to CDO pricing*. PhD thesis, Universität Ulm.

Hofert, M. (2012). A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 82(9) :1239–1255.

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., Maechler, M. M., and Suggests, M. (2014). Package "copula". *R Package Manual*.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8) :2554–2558.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

Joe, H. and Kurowicka, D. (2011). *Dependence modeling : vine copula handbook*. World Scientific.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivarées*. PhD thesis, Université Laval.

Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes mathematicae*, 10(2) :152–164.

Kuo, K. (2019). Deeptriangle : A deep learning approach to loss reserving. *Risks*, 7(3) :97.

Kuo, K. (2020). Individual claims forecasting with bayesian mixture density networks. *arXiv preprint arXiv :2003.02453*.

Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate analysis*, 99(10) :2234–2250.

Lindholm, M., Verrall, R., Wahl, F., and Zakrisson, H. (2020). Machine learning, regression models, and prediction of claims reserves. In *Arlington : Casualty Actuarial Society E-Forum.[Google Scholar]*.

Marceau, E. (2013). *Modélisation et évaluation quantitative des risques en actuariat : Modèles sur une période*. Springer.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403) :834–841.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4) :115–133.

McNeil, A. J. and Nešlehová, J. (2010). From Archimedean to Liouville copulas. *Journal of Multivariate Analysis*, 101(8) :1772–1790.

Mesiar, R. and Jagr, V. (2013). d-Dimensional dependence functions and Archimax copulas. *Fuzzy Sets and Systems*, 228 :78–87.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Scarsini, M. (1984). On measures of concordance. *Stochastica : Revista de Matemática Pura y Aplicada*, 8(3) :201–218.

Shi, W., Li, K. X., Yang, Z., and Wang, G. (2017). Time-varying copula models in the shipping derivatives market. *Empirical Economics*, 53(3) :1039–1058.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8 :229–231.

Sundermeyer, M., Schlüter, R., and Ney, H. (2012). LSTM neural networks for language modeling. In *Thirteenth Annual Conference of the International Speech Communication Association*.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *The Annals of Probability*, 8(4) :814–827.

Weninger, F., Erdogan, H., Watanabe, S., Vincent, E., Le Roux, J., Hershey, J. R., and Schuller, B. (2015). Speech enhancement with LSTM recurrent neural networks and its application to noise-robust asr. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 91–99. Springer.

Wüthrich, M. V. (2018). Neural networks applied to chain–ladder reserving. *European Actuarial Journal*, 8(2) :407–436.

Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons.

Chapitre 1

Micro-level Reserving for General Insurance Claims using a Long Short-Term Memory Network

Résumé

Les informations détaillées sur les sinistres individuels sont non utilisées dans toute évaluation de réserve basée sur des données agrégées et structurées sous forme de triangles de développement. Dans l'espoir d'extraire un pouvoir prédictif des caractéristiques individuelles des sinistres, des chercheurs ont récemment proposé des méthodes de provisionnement au niveau micro. Nous introduisons une approche de réserve individuelle à temps discret incorporant les informations granulaires dans une technique d'apprentissage profond appelée réseau de neurones récurrent à mémoire à court et long terme (LSTM). Ce réseau a deux tâches à chaque période de développement : une classification visant à prédire s'il y aura un paiement et une régression pour prédire le montant associé s'il y a lieu. Nous illustrons la procédure d'estimation sur deux jeux de données simulées et réelles de sinistres en assurance générale. Nous comparons notre approche avec la méthode agréger Chain-ladder en utilisant les estimations des paiements futurs ainsi que leurs valeurs réelles. Sur la base d'un modèle de Pareto généralisé pour les paiements excédentaires dépassant un certain seuil, nous ajustons la prévision de réserve obtenu avec notre LSTM pour tenir compte des sinistres importants.

Abstract

Detailed information about individual claims are completely ignored when insurance claims data are aggregated and structured in development triangles for loss reserving. In the hope of extracting predictive power from the individual claims characteristics, researchers have recently proposed to move away from these macro-level methods in favor of micro-level loss reserving approaches. We introduce a discrete-time individual reserving framework incorporating granular information in a deep learning approach named Long Short-Term Memory (LSTM) neural network. The network has two tasks at each time period: classifying whether there will be a payment and predicting the amount, if any. We illustrate the estimation procedure on a simulated and a real general insurance dataset. We compare our approach with the chain-ladder aggregate method using the predictive outstanding loss estimates and their actual values. Based on a generalized Pareto model for excess payments over a threshold, we adjust the LSTM reserve prediction to account for large claims.

Keywords: Individual claim features, Individual claim reserving, Deep learning, Large claims, Recurrent neural networks.

1.1 Introduction

Predicting outstanding liabilities is essential to ensure an insurance company's solvency. The reserving exercise consists of accurately estimating future loss payments to be made on incurred claims. Traditional macro-level loss reserving methods rely on simple assumptions and run-off triangles with claim informations summarized by occurrence and development periods. This results in a loss of individual characteristics which may limit these models' robustness, particularly in situations when portfolio mix or claim characteristics are evolving over time. For an overview of aggregate reserving models, see, e.g., England and Verrall (2002) or Wüthrich and Merz (2008). Recently, researchers have proposed micro-level loss reserving approaches that use detailed individual claims information in order to improve prediction accuracy. Their main advantage consists in yielding the predicted outstanding loss of each claim based on its individual development and characteristics. Several individual models, in continuous or discrete-time, have been proposed in the literature. Pioneer works by Arjas (1989), Norberg (1993), Haastrup and Arjas (1996) and Norberg (1999) set the stage for micro-level reserving. See, e.g., Charpentier and Pigeon (2016) for an econometric comparison between various aggregate and individual reserving models.

In recent years, individual models have gained interest, and many contributions achieved promising results. Taylor et al. (2008) used a discrete-time framework to model individual claim developments, based on claim payment and incurred losses. Zhao et al. (2009) investigated the occurrence time and the reporting delay for individual claim loss, using regression methods, then Zhao and Zhou (2010) extended these studies by incorporating a dependence structure using copulas. Pigeon et al. (2013) and Pigeon et al. (2014) proposed a parametric individual reserving model using the multivariate skew-normal distribution for claim payments, in discrete time. In a continuous-time framework, Antonio and Plat (2014) present a semi-parametric individual reserving model using detailed information on the payment rate. All these granular models are parametric or semi-parametric, based on fixed structural forms and can potentially be over-parametrized.

Machine learning techniques are highly flexible for handling structured and unstructured data. It is thus no surprise that they are gaining in popularity as building blocks of individual loss reserving models. Many recent contributions showed how machine learning could improve an individual reserving model's prediction accuracy. Wüthrich (2018) used regression trees to predict the number of payments. Tree-based techniques like ExtraTrees and XGBoost have been applied to predict outstanding individual losses, see, e.g., Baudry and Robert (2019) and Duval and Pigeon (2019).

With the growth of individual claims data collection, storage and improved computing power, it becomes interesting to consider a more advanced form of machine learning, namely deep neural networks (NNs). The latter require few restrictions and assumptions, incorporate com-

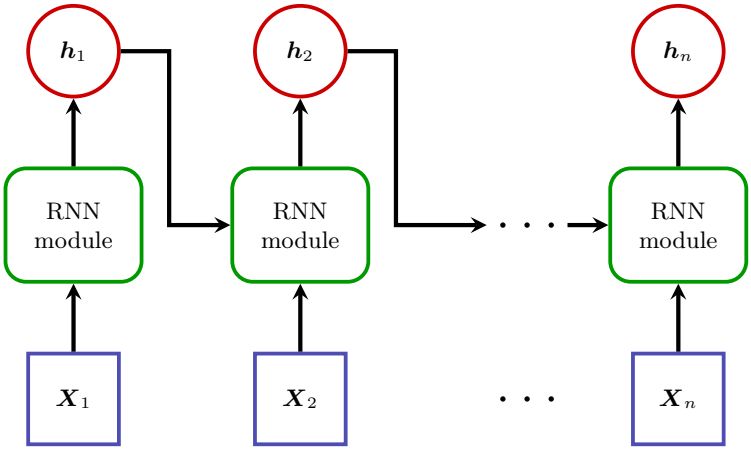


Figure 1.1 – Structure of an RNN with dynamic input.

plex non-linear trends, and have high predictive performance. NNs with various architectures have recently been applied to individual claims loss prediction. Wüthrich (2018) and Taylor (2019) surveyed recent developments in individual reserving models involving NNs. Gabrielli et al. (2020) proposed to initialize the NN with a regression model, such as an over-dispersed Poisson, to better align with traditional actuarial practice. Gabrielli (2021) structured a single NN carrying out simultaneous regression and classification tasks to predict expected payments. Summaries of the claim history are the inputs of his network, whereas Delong and Wüthrich (2020) used entire claim histories in their NN to predict the joint development of claim occurence and individual payments.

Another way to use past claim developments is through recurrent neural networks (RNN), a popular class of NNs introduced by Hopfield (1982). They are constructed as multiple copies of the same network, allowing information to be passed from one step to the successor network. Figure 1.1 shows a typical representation of a RNN, with dynamic input $\boldsymbol{X}_t$ and output $\boldsymbol{h}_t$ at period t, for $t = 1, \ldots, n$. By construction, RNNs consider relevant features to extract temporal dependencies. However, in basic RNNs, it is challenging to capture long-term dependencies, that are useful to understand claim development, because of the multiplicative gradient that can grow exponentially. To avoid exploding gradient, Hochreiter and Schmidhuber (1997) introduced Long Short-Term Memory (LSTM) networks, a class of RNNs. However, the LSTM module has a complex architecture with multiple layers interacting in a special way, and characterized by repeatedly updating memory cells. For several sequential information processing tasks, LSTM networks showed competitive results over simpler RNNs. For more details on LSTMs, see, e.g., Sundermeyer et al. (2012), Graves (2013) and Weninger et al. (2015). In the individual reserving litterature, Kuo (2020) proposed a multi-period bayesian mixture density network based on LSTMs. Unfortunately, this model does not improve the accuracy of predictions compared to the classic chain-ladder.

In this paper, we develop a new individual loss reserving model applicable for any individual

claims dataset with long-tailed development. We focus on the Reported, But Not yet Settled (RBNS) claims by modeling the development of their future payments. Our model involves an LSTM network which performs two tasks in order to predict expected future payments: classification to determine the probability of having a non-zero payment and regression to evaluate the incremental payment, given that it is non-zero. A loss-balancing technique allows the simultaneous learning of these two network tasks. We illustrate the model implementation and explore its performance on both a simulated and a real complex dataset. We compare the prediction accuracy with that of a chain-ladder model. Also, to predict large incremental payments, we design a reserve prediction approach combining the LSTM output and a generalized Pareto distribution. We study its performance on the real dataset.

Compared to Gabrielli (2021), our network is recurrent and uses exact individual claim histories instead of summarized past information. Our model has the advantage of making predictions for the entire future development rather than only yielding an ultimate value. In the case studies presented in Sections 1.3 and 1.4, our network outperforms the results obtained with the aggregate chain-ladder, which is an improvement over the model developed in Kuo (2020).

The remainder of the paper is organized as follows. In Section 1.2, we introduce some notation and the architecture and training process of our individual loss reserving model. Sections 1.3 and 1.4 present the experiments conducted on a simulated and a real detailed claim dataset, respectively. For these case studies, we explain the model construction, choose the loss function, analyze results and compare them with chain-ladder. Also, the individual reserve approach for large claims is described in Section 1.4 and illustrated on real data. We conclude in Section 1.5 and provide additional details on the simulated data, real data pre-processing and extreme value model in three Appendices.

1.2 Individual loss reserving model

In Section 1.2.1, we introduce the notation for a generic detailed insurance claims dataset. We then present the architecture and training of our LSTM network in Sections 1.2.2 and 1.2.3, respectively.

1.2.1 Notation

We consider a dataset of m reported general insurance claims. Each of these claims has a lifetime development; a typical example is depicted in Figure 1.2. A claim occurring within the policy period is usually declared to the insurer with a reporting delay. There may also be an opening lag, which is the delay between the reporting and opening dates. At opening time, specific features about the claim are known, such as the date of occurrence, reporting delay, age of the claimant (body injury insurance), or type of house (home insurance). We denote

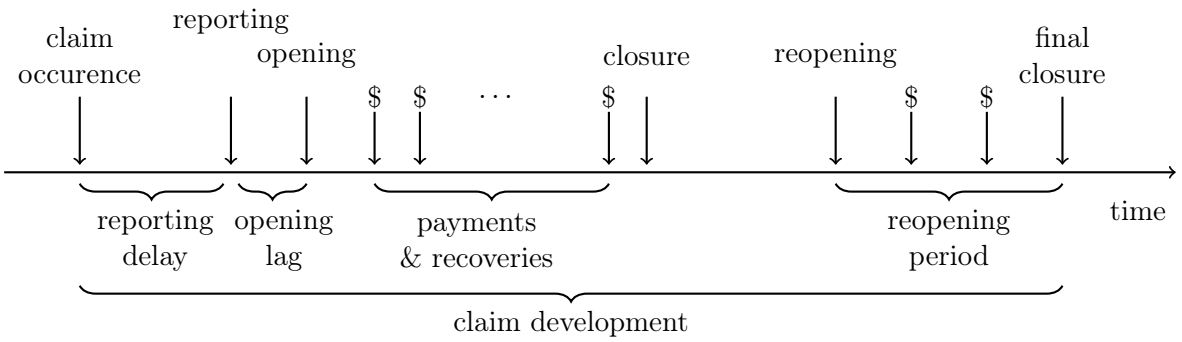


Figure 1.2 – Typical general insurance claim development.

by $\boldsymbol{S}_k$ the set including this static information for claim k, for $k = 1, \ldots, m$. Over a claim lifetime, several loss payments and recoveries may be made at different time intervals before closure. A claim can be reopened and then closed again. Throughout this settlement process, we observe additional claim characteristics.

We consider discrete-time claim developments, as the exact daily modeling is too granular and involves weekly variations that are immaterial for reserving. The development periods are of equal length and may be, for example, months, quarters, or years. We also assume that all claims are fully settled at the end of an ultimate period n. For claim k, we denote by $\boldsymbol{D}_{k,j}$ the set of dynamic information known at period j, for $j = 1, \ldots, n$. This includes the incremental payment $Y_{k,j}$, the claim status (whether the claim is open or not), and the number of claimants related to claim k.

The reserving problem amounts to estimating the outstanding payments for incurred claims at a given evaluation date T^*. The claims have a different number of known development periods at T^* depending on their occurrence date. Assume that for claim k, we have observed $t_k < n$ development periods at T^*, as illustrated in Figure 1.3. The entire dataset can be represented as

$$\{\boldsymbol{S}_k,\ \boldsymbol{D}_{k,j} : j = 1, \ldots, t_k;\ k = 1, \ldots, m\}.$$

Among the set of dynamic variables, we propose to focus on the occurrence and amounts of incremental payments. In fact, for a claim, we may observe a large number of development periods without payment or even an ultimate payment equal to \$0. Thus, we define the payment indicator $I_{k,j} = \mathbb{1}_{\{Y_{k,j} \neq 0\}}$, where $\mathbb{1}_{\{Y_{k,j} \neq 0\}} = 1$ if $Y_{k,j} \neq 0$, and 0 otherwise. The indicator $I_{k,j}$ will be used in our model to assess the probability of having a non-zero payment. The individual reserve R_k is the sum of the incremental payments forecasted for the periods $\{t_k + 1, \ldots, n\}$, and is expressed as

$$R_k = \sum_{j=t_k+1}^{n} I_{k,j} \times Y_{k,j},$$

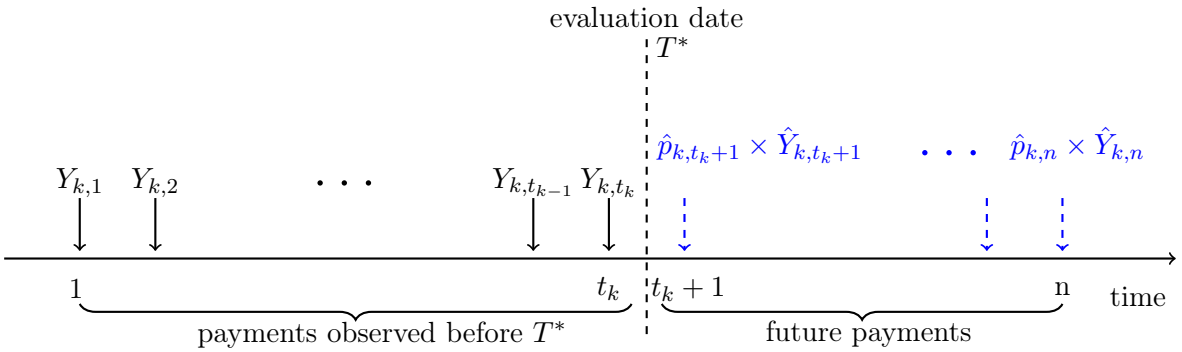


Figure 1.3 – Discrete-time claim development with observed and future payments at T^*.

which also corresponds to the ultimate loss payment minus the cumulative paid amount at period t_k.

In Section 1.2.2, we present our individual reserve approach to estimate R_k, for $k = 1, \dots, m$, but we first need to split the full dataset into training, validation, and testing datasets. We use the training dataset, denoted by $\mathcal{T}$, to find the model parameters (weights) that will be confirmed using the validation dataset $\mathcal{V}$. Our methodology is applicable for a training and validation set containing solely information that is known prior to the evaluation date T^*. Therefore, an insurer could use it as is for setting the reserves at the current evaluation date. After the training process, we use the testing dataset, including information beyond T^*, to verify the model accuracy.

1.2.2 Architecture of the LSTM individual loss reserving model

For claim $k \in \{1, \dots, m\}$, let $\hat{p}_{k,j}$ be the predicted probability of payment in period $j \in \{1, \dots, n\}$, and $\hat{Y}_{k,j}$ the corresponding predicted payment given that it is non-zero. We define the predicted expected incremental payment by $\hat{p}_{k,j} \times \hat{Y}_{k,j}$. At T^*, the insurer must predict future payments indicated in blue in Figure 1.3. We propose an individual reserving approach to estimate R_k with

$$\hat{R}_k = \sum_{j=t_k+1}^{n} \hat{p}_{k,j} \times \hat{Y}_{k,j}.$$

The reserving problem thus amounts to the prediction of the pairs $\{(\hat{p}_{k,j}, \hat{Y}_{k,j}) : j = t_k + 1, \dots, n\}$. To achieve this goal, we introduce the LSTM network depicted in Figure 1.4, where we drop the claim index k and keep only the time index j for simplicity. Since we study RBNS claims, at least one development period is known for each claim, so the LSTM network contains $n-1$ modules. At step j, we provide the network with the input vector $\boldsymbol{X}_{k,j}$ concatenating the dynamic information of period j, $\boldsymbol{D}_{k,j}$, and the static context $\boldsymbol{C}_{k,0}$, i.e., $\boldsymbol{X}_{k,j} = \{\boldsymbol{C}_{k,0}, \boldsymbol{D}_{k,j}\}$. In the following, we first explain the construction of $\boldsymbol{C}_{k,0}$ and the payment pre-processing

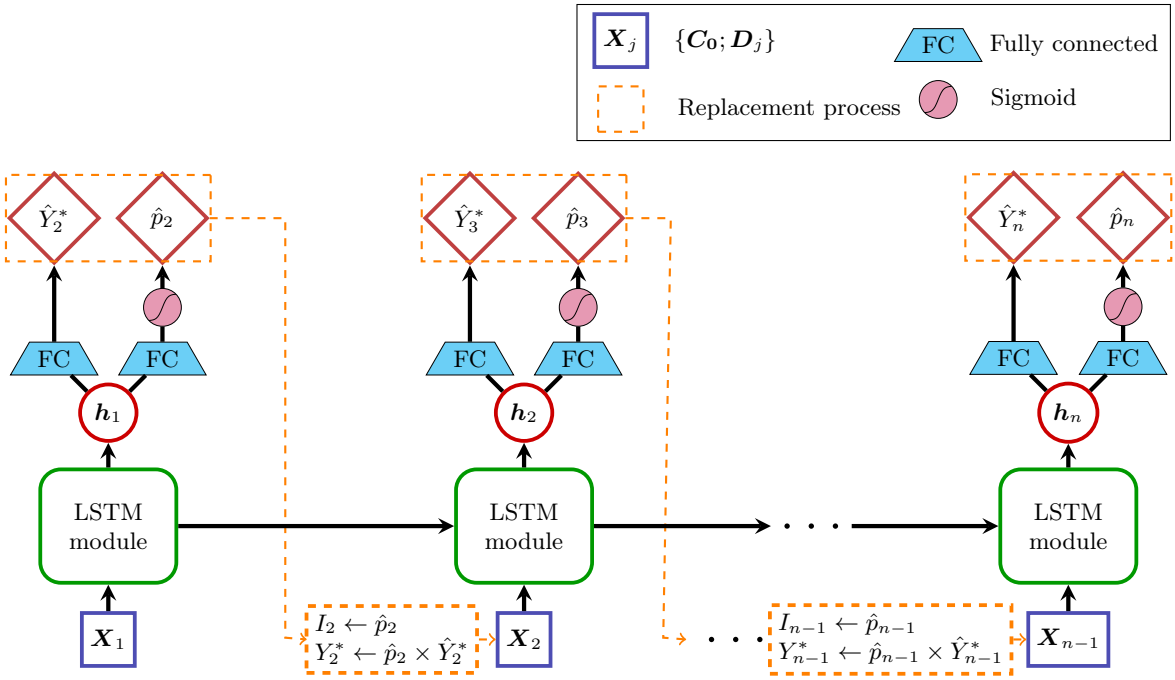


Figure 1.4 – Architecture of the LSTM individual loss reserving network.

within $\boldsymbol{D}_{k,j}$ to form $\boldsymbol{X}_{k,j}$. Then, we describe how inputs are fed to each module as illustrated in Figure 1.4 and we explain the resulting outputs.

The static features known at the opening date provide a context for the claim and may inform us on its development. The set $\boldsymbol{S}_k$ may include quantitative and/or categorical variables. Each of these types of variables must be prepared differently. Figure 1.5 illustrates the static feature engineering used in this paper. The q quantitative variables are scaled and each of the p categorical variables is indexed using a defined dictionary, then passed to an embedding layer. The latter maps each level of categorical variable $\boldsymbol{S}_{k,i}$ to an ℓ_i-dimensional vector, where $\ell_i > 1$ is a hyper-parameter that depends on the number of modalities and $i = 1, \ldots, p$ (see, e.g., Yin and Shen, 2018). After transformation, the static variables are concatenated and then passed to a linear layer for dimension reduction. We obtain the kth claim's context $C_{k,0}$, a vector of dimension c, where $c > 1$ is a hyper-parameter. Note that we choose to recall the context at each step of the LSTM in Figure 1.4 to strengthen the network, as it will keep only the relevant information through its particular internal memory cells.

For the set of dynamic features $\boldsymbol{D}_{k,j}$, two dynamic variables are appended, namely the development period j and the indicator $r_{k,j} = \mathbb{1}_{\{j \leq t_k\}}$. The incremental payments are centered and scaled for the training, validation, and testing datasets as follows

$$Y_{k,j}^* = \left(Y_{k,j} - \mu_{\mathcal{T}}\right) / \sigma_{\mathcal{T}}, \tag{1.1}$$

with mean $\mu_{\mathcal{T}}$ and scale $\sigma_{\mathcal{T}}$ computed from all payments in the training dataset. For our

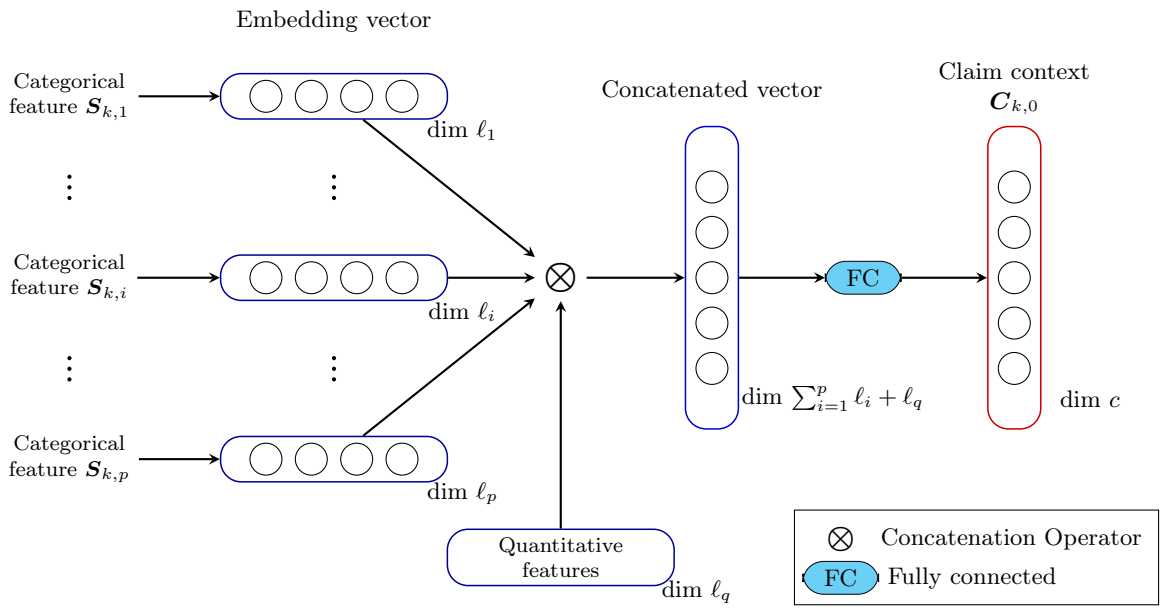


Figure 1.5 – Static variable engineering for the claim context vector.

model, we set $\boldsymbol{D}_{k,j} = \{j,\ r_{k,j},\ I_{k,j},\ Y^*_{k,j}\}$, for $j = 1, \ldots, t_k$.

Now that we have presented the construction of the inputs $\boldsymbol{X}_{k,j}$, for $j = 1, \ldots, t_k$, we recall Figure 1.4 to explain the network flow. At period j, the vector $\boldsymbol{X}_{k,j}$ passes through the LSTM module whose specific architecture (see, e.g., Yu et al., 2019) leads to the hidden state vector $\boldsymbol{h}_{k,j}$. We deduce from it the pair of predictions $(\hat{p}_{k,j+1}, \hat{Y}^*_{k,j+1})$ for the next period $j+1$ as follows. The vector $\boldsymbol{h}_{k,j}$ is passed, in parallel, to a linear layer yielding the payment $\hat{Y}^*_{k,j+1}$, and to another linear layer, then a sigmoid neuron, to obtain the predicted probability $\hat{p}_{k,j+1}$. Mathematically, we have

$$\hat{Y}^*_{k,j+1} = \boldsymbol{\beta}\boldsymbol{h}^\top_{k,j} + a, \qquad \hat{Y}_{k,j+1} = \mu_{\mathcal{T}} + \sigma_{\mathcal{T}} \hat{Y}^*_{k,j+1} \qquad \text{and} \qquad \hat{p}_{k,j+1} = [1 + \exp\{-(\boldsymbol{\beta}'\boldsymbol{h}^\top_{k,j} + a')\}]^{-1},$$

with $\boldsymbol{h}^\top_{k,j}$ the transpose of $\boldsymbol{h}_{k,j}$, $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$ two vectors of learnable weights, and a and a' are constants associated to the two linear layers. The entire network's output is the set $\{(\hat{p}_{k,j}, \hat{Y}^*_{k,j}),\ j = 2, \ldots, n\}$, from which we can retrieve the predictions for the future periods $\{(\hat{p}_{k,j}, \hat{Y}_{k,j}), j = t_k + 1, \ldots, n\}$. Moreover, for time step j, the target is the pair $(I_{k,j}, Y^*_{k,j})$ only if it is observed before the evaluation date. Otherwise, it is an empty pair. We will explain in Section 1.2.3 how the network's optimization is carried out with missing targets.

An important point is how the network deals with the missing values in the inputs to ensure information propagation. At the first step, the input $\boldsymbol{X}_{k,1}$ is observed and provided to the first LSTM module. From $\boldsymbol{h}_{k,1}$, we obtain the predicted pair $(\hat{p}_{k,2}, \hat{Y}^*_{k,2})$ for the second development period. Moving to the second step, if $t_k < 2$, then $I_{k,2}$ and $Y^*_{k,2}$ are missing values in the input $\boldsymbol{X}_{k,2}$. They are imputed by the predicted expected values, $\hat{p}_{k,2}$ and $\hat{p}_{k,2} \times \hat{Y}^*_{k,2}$, respectively. Otherwise, when $t_k \geq 2$, we keep $\boldsymbol{X}_{k,2}$ as actual observed values. We proceed in the same

manner for the other periods until time $n-1$. In Figure 1.4, orange dotted lines indicate this replacement process for the inputs. Thus, our network, which is of fixed length, may process sequences with different lengths.

1.2.3 Model training

In the following, we describe the training process, the network's two tasks, and the balanced loss function. Also, we explain the regularization techniques and metrics used.

The learning process consists of determining the optimal network weights that minimize the appropriate loss function. We update weights with a stochastic gradient descent at each epoch, relying on all mini-batches obtained by randomly splitting the training dataset. The mini-batch size is a hyper-parameter denoted b, and an epoch refers to a complete iteration through all mini-batches. At the end of each epoch, we evaluate the loss function on the validation dataset to measure the network generalization capabilities. Furthermore, given that we produce at each time period the prediction of two variables, our network performs two tasks simultaneously.

The first task is a binary classification with target $I_{k,j}$, indicating whether the claim k has a non-zero payment in period j, for $k = 1, \dots, m$ and $j = 2, \dots, n$. The indicator takes value in $\{0, 1\}$, so we choose the binary cross-entropy as loss function. For each mini-batch, the loss is defined as the weighted sum of cross-entropy losses per development period. For a given period, the loss calculation involves only observations with non-missing target $I_{k,j}$. As we move forward through the development periods, we get fewer observed data, and hence we need to weigh the sum by the number of available observations. The cross-entropy loss for a mini-batch can be expressed as

$$
\begin{aligned}
&CE\left(\hat{\boldsymbol{p}}\right) \qquad (1.2)\\
&= \sum_{j=2}^{n} \frac{\sum_{i=1}^{b} \delta_{i,j}}{\sum_{h=2}^{n}\sum_{i=1}^{b} \delta_{i,h}} \left(\sum_{k=1}^{b} \frac{\delta_{k,j}}{\sum_{i=1}^{b} \delta_{i,j}} \left[- \left\{ \mathbb{1}_{\{I_{k,j}=0\}} \log\left(1-\hat{p}_{k,j}\right) + \mathbb{1}_{\{I_{k,j}=1\}} \log\left(\hat{p}_{k,j}\right) \right\} \right] \right)\\
&= \frac{1}{\sum_{h=2}^{n}\sum_{i=1}^{b} \delta_{i,h}} \sum_{j=2}^{n} \left(\sum_{k=1}^{b} \delta_{k,j} \left[- \left\{ \mathbb{1}_{\{I_{k,j}=0\}} \log\left(1-\hat{p}_{k,j}\right) + \mathbb{1}_{\{I_{k,j}=1\}} \log\left(\hat{p}_{k,j}\right) \right\} \right] \right), \qquad (1.3)
\end{aligned}
$$

with $\hat{\boldsymbol{p}} = \{\hat{p}_{k,j} : k = 1, \dots, b, j = 2, \dots, n\}$ and to only use observations with known target, we set

$$
\delta_{k,j} = \begin{cases} 1 & , \quad I_{k,j} \in \{0,1\} \\ 0 & , \quad I_{k,j} = \text{NA} \ . \end{cases}
$$

The second task is a regression for the non-zero incremental payments $Y_{k,j}$. The regression loss is defined by a function f that can either be the squared error $f(Y^*, \hat{Y}^*) = f_{SE}(Y^*, \hat{Y}^*) = (Y^* - \hat{Y}^*)^2$ or the absolute error $f(Y^*, \hat{Y}^*) = f_{AE}(Y^*, \hat{Y}^*) = |Y^* - \hat{Y}^*|$. The choice of function f depends on the dataset and the dispersion of payment distribution. For a given mini-batch,

we evaluate the regression loss RL as a weighted sum of development period loss functions f, with

$$RL_f\left(\hat{\boldsymbol{Y}}^*\right) = \sum_{j=2}^{n} \frac{\sum_{i=1}^{b} \tilde{\delta}_{i,j}}{\sum_{h=2}^{n}\sum_{i=1}^{b} \tilde{\delta}_{i,h}} \left(\sum_{k=1}^{b} \frac{\tilde{\delta}_{k,j}}{\sum_{i=1}^{b} \tilde{\delta}_{i,j}} \; f\left(Y^*_{k,j}, \widehat{Y}^*_{k,j}\right)\right)$$
$$= \frac{1}{\sum_{h=2}^{n}\sum_{i=1}^{b} \tilde{\delta}_{i,h}} \sum_{j=2}^{n}\sum_{k=1}^{b} \tilde{\delta}_{k,j} \; f\left(Y^*_{k,j}, \widehat{Y}^*_{k,j}\right), \tag{1.4}$$

where $\hat{\boldsymbol{Y}}^* = \{\hat{Y}^*_{k,j} : k = 1, \ldots, b, j = 2, \ldots, n\}$ and to only use non-zero observed targets, we set

$$\tilde{\delta}_{k,j} = \begin{cases} 0 & , \quad Y^*_{k,j} \in \{-\mu_{\mathcal{T}}/\sigma_{\mathcal{T}}, \text{NA}\} \\ 1 & , \quad \text{otherwise} \end{cases}.$$

We recall that $Y^*_{k,j} = -\mu_{\mathcal{T}}/\sigma_{\mathcal{T}}$ correponds to the case $Y_{k,j} = 0$.

Our LSTM can learn well both tasks since they share informative features, which may induce more robust regularization and boost performance. However, for training, we must define one loss function in terms of losses (1.2) and (1.4). Multi-task network loss functions are often taken to be linear in the individual task losses. For example, the loss in Gabrielli (2021) is a weighted sum of the individual losses such that they lie on the same scale. Nevertheless, when learning regression and classification simultaneously, we should ensure that both tasks are given equal consideration to prevent the easier one from dominating. To this end, several robust automatic loss balancing techniques were proposed; for an overview of multi-task learning approaches, see, e.g., Ruder (2017).

For training our network, we use the task-dependent uncertainty weighting technique proposed in Kendall et al. (2018). This approach allows the optimal weighting of losses with different units or scales and can be applied to combine classification and regression tasks. The multi-task optimization objective is defined as a Gaussian likelihood with task-dependent uncertainty. Kendall et al. (2018) illustrated the performance of their loss balancing technique through a multi-task network learning visual scene, which outperformed a separate training of the same tasks.

In several training experiments of our LSTM network, we observed a difference in optimization challenge and measurement scale between the regression and classification loss functions. In addition to using the task-dependent uncertainty, we introduce a scale hyper-parameter α applied to the classification loss to reduce the difference with the regression loss and improve tasks learning. Hence, the Gaussian likelihood function is

$$\mathcal{L}(\hat{\boldsymbol{Y}}^*, \; \hat{\boldsymbol{p}}) = \frac{1}{\sigma_1^2} \times RL_f\left(\hat{\boldsymbol{Y}}^*\right) + \frac{\alpha}{\sigma_2^2} \times CE\left(\hat{\boldsymbol{p}}\right) + \log \sigma_1^2 + \log \sigma_2^2, \tag{1.5}$$

with σ_1^2 and σ_2^2 corresponding to the regression and classification uncertainties, respectively. The optimization problem can be expressed, in terms of the LSTM network weights vector

$\boldsymbol{W}$, as

$$\underset{\boldsymbol{W},\sigma_1,\sigma_2}{\operatorname{argmin}} \mathcal{L}(\hat{\boldsymbol{Y}}^*,\ \hat{\boldsymbol{p}}).$$

We now recall Figure 1.4 to explain several strategies used to improve the network training. Moving forward in periods, we apply the teacher forcing technique to accelerate the training process (see, e.g., Williams and Zipser, 1989; Drossos et al., 2019). According to a probability function that is exponentially increasing over epochs, at each time period we randomly choose whether to replace the observed pair $(I_{k,j},\ Y^*_{k,j})$ within the input $\boldsymbol{X}_{k,j}$ with the predictions from the previous period $(\hat{p}_{k,j},\ \hat{p}_{k,j} \times \hat{Y}^*_{k,j})$. This improves the ability of the network to predict correctly late periods even when there are imprecise inputs in early periods. Orange dotted lines in Figure 1.4 show this replacement procedure, which is similar to the missing values imputation. Also, we use stochastic gradient descent with an initial learning rate l_r, that we drop by a factor when there is no improvement in the validation loss for r_p epochs. The training process is stopped early when the validation loss does not improve for t_{es} epochs, and then we obtain the optimal weights.

Our network loss function (1.5) does not have a simple interpretation. We are thus interested in some metrics to evaluate the model's performance. For classification, we use the ROC (Receiver Operating Characteristic) curve with observations grouped by development period and the associated AUROC (Area Under the ROC). The higher the AUROC, the better the model distinguishes between having a payment or not. Other metrics are presented in Sections 1.3.3 and 1.4.4.

1.3 Simulated data

In this section, we evaluate our individual loss reserving model performance on fully developed micro-level simulated data. This allows us to compare the predictive outstanding loss estimates with their actual values. One advantage of using simulated data is research reproducibility; the complete code for this case study is available from the authors upon request. Moreover, we conduct a comparative analysis between our model and the aggregate chain-ladder approach.

1.3.1 Data description

Using NNs, Gabrielli and Wüthrich (2018) developed a stochastic generator of non-life insurance individual claim histories, based on a real property and casualty insurance dataset. The exact setup used for the generator is provided in Appendix 1.A.

The simulated dataset contains $998,807$ claims with occurrence dates between years 1994 and 2005. For claim k, several static characteristics are contained in vector $\boldsymbol{S}_k$, for instance, claim number, accident year, claimant age, line of business, and injured body part. Through the yearly claim development, we know the incremental payments and indicators denoting whether

Table 1.1 – List of simulated dataset features and their pre-processing.

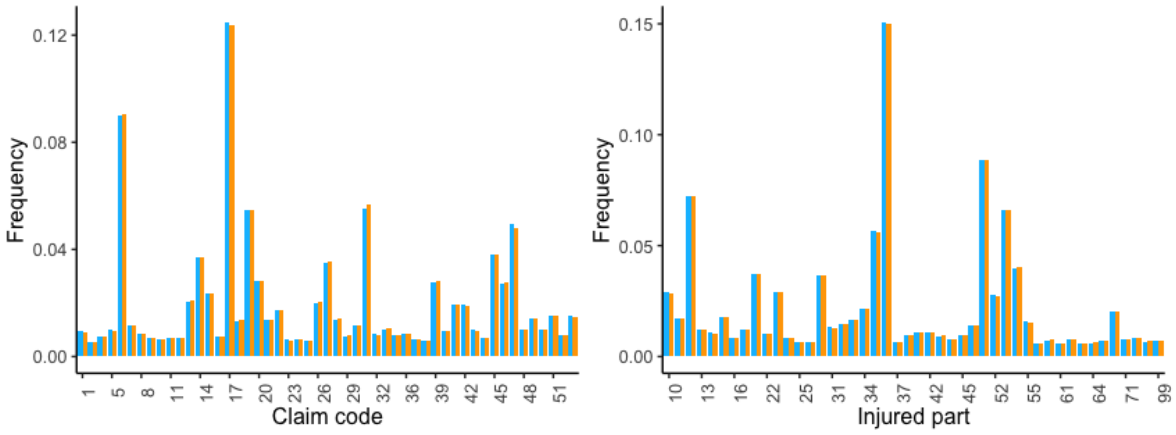
Type	Features	Pre-processing
Static	Accident year	Expressed in number of years from 1994
	Reporting delay	—
	Claimant age	Scaled to $[0, 1]$
	Line of business	4 categories indexed
	Claim code	51 categories indexed
	Injured part	46 categories indexed
Dynamic	Incremental yearly payment	Centered and scaled as in Eq. (1.1)

the claim is open or closed. These two dynamic informations are reported at the end of each development year $j = 1, \dots, 12$, and claims are assumed to be fully developed after $n = 12$ years.

For our reserving model, we drop claim status and keep the payment information within the dynamic vectors, with $\boldsymbol{D}_{k,j} = \{j,\ r_{k,j},\ I_{k,j},\ Y^*_{k,j}\}$, for $j = 1, \dots, t_k$, and $k = 1, \dots, m$. Table 1.1 lists the static and dynamic variables used in our reserving model and their pre-processing. Note that the simulated claims have some realistic features such as negative payments representing recoveries (0.3%), late reporting, reopenings (0.3%), and settlement at zero (28%). However, the data contain only few covariates to properly describe the characteristics of claims, which may harm the performance of micro-level reserving techniques.

We set $T^* = 2005$ as the evaluation year. To train our reserving model, we keep only claims reported before T^* and hide payments made beyond T^*. Hence, we retain $m = 991,904$ claims leading to a deletion of only 0.7% of them. This small percentage is due to the fact that 92% of claims are reported during their first development year. Also, claims close on average after 1.5 year. Therefore, we expect to forecast few future non-zero outstanding payments. We stratify the claims into training (60%), validation (20%), and testing (20%) datasets. The hidden payments associated to each of the three datasets made after T^* are used later to assess the network's performance.

We carried out the same preliminary analysis on both training and validation datasets to assess the split homogeneity. Among the available static information in $\{\boldsymbol{S}_k : k = 1, \dots, m\}$, we analyze the categorical covariates to index their categories and set the hyper-parameters of the embedding layers ℓ_1, ℓ_2 and ℓ_3. Figure 1.6(a) depicts the frequency of claim code. We identify 51 categories with labels in $\{1, \dots, 53\}$. For the injured body part, we observe 46 categories in the barplot illustrated in Figure 1.6(b), with labels within $\{10, \dots, 99\}$. Both figures show almost the same probability distribution for each variable for training and validation datasets, which confirm the similarity of the two datasets structures. The third categorical variable is the line of business with four categories and empirical frequency almost identical to the probability distribution given in Table 1.A.1.



(a) (b)

Figure 1.6 – Frequency of claim code (a) and injured part (b) for training (blue) and validation (orange) sets.

Table 1.2 – Hyper-parameters used to train the LSTM network on the simulated dataset.

Hyper-parameter	Value	Hyper-parameter	Value
Context size c	32	Hidden size	128
Batch size d	2048	Learning rate l_r	0.05
Reduce on plateau r_p	10 epochs	Early stopping t_{es}	15 epochs

1.3.2 Model training

Due to the ultimate development period being 12 years, we set $n = 12$ in our LSTM network depicted in Figure 1.4 and outputting the set of pairs $\{(\hat{p}_{k,j}, \hat{Y}^*_{k,j}), j = 2, \dots, 12\}$. Since we do not observe large claims or high variance for incremental payments, we choose the squared error function $f = f_{SE}$ in regression loss (1.4) and LSTM network loss function (1.5). After the pre-processing of variables described in Table 1.1, we train our net using the training dataset. We perform several experiments to set the hyper-parameters on the validation dataset. The optimal model performance is obtained with the hyper-parameter combination given in Table 1.2.

To determine the scale hyper-parameter α in loss function (1.5), we train our network with $\alpha \in \{0.1, 0.2, \dots, 2.5\}$. To set the best value, we need to compare the predictions with the actual payments while using only the information known at evaluation time $T^* = 2005$. We explain in the following how to do so. We set the end of the year 2003 as our reference date, denoted T^{**} and, for claim k, we define $t^{**}_k = t_k - (T^* - T^{**})$ as the number of development periods observed at T^{**}, as illustrated in Figure 1.7. In the validation dataset $\mathcal{V}$, we select the claims for which at least one development year is observed at T^{**}, i.e., $t^{**}_k > 0$, leading to a loss of 18% of the observations.

Our choice is based on two performance metrics. First, the ratio of the aggregate predicted

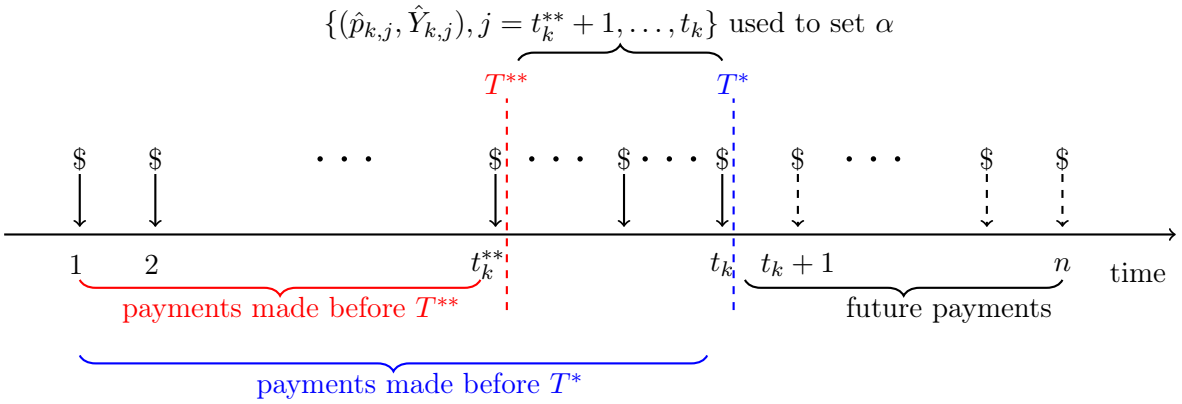


Figure 1.7 – Development of claim k over n equal periods.

Table 1.3 – Ratios on the simulated validation dataset for the selection of α.

α	0.2	0.6	0.8
$RR(2003, 2005)$	0.9556	0.9145	0.9413
$RU(2003, 2005)$	0.9950	0.9904	0.9934

over the aggregate observed between T^{**} and T^{*} is

$$RR(T^{**}, T^{*}) = \frac{\sum_{k\in\mathcal{V}} \sum_{j=t_k^{**}+1}^{t_k} \mathbb{1}_{\{t_k^{**}>0\}} \times \hat{p}_{k,j} \times \hat{Y}_{k,j}}{\sum_{k\in\mathcal{V}} \sum_{j=t_k^{**}+1}^{t_k} \mathbb{1}_{\{t_k^{**}>0\}} \times I_{k,j} \times Y_{k,j}}.$$

The second is the ratio of the aggregate ultimate assuming the ultimate time is T^{*}, given by

$$RU(T^{**}, T^{*}) = \frac{\sum_{k\in\mathcal{V}} \mathbb{1}_{\{t_k^{**}>0\}} \left(\sum_{j=1}^{t_k^{**}} I_{k,j} \times Y_{k,j} + \sum_{j=t_k^{**}+1}^{t_k} \hat{p}_{k,j} \times \hat{Y}_{k,j} \right)}{\sum_{k\in\mathcal{V}} \sum_{j=1}^{t_k} \mathbb{1}_{\{t_k^{**}>0\}} \times I_{k,j} \times Y_{k,j}}.$$

In the training experiments, we observe that both ratios have a non-monotonic relationship with α. LSTMs trained with $\alpha \in \{0.2, 0.6, 0.8\}$ outperform the other networks on the validation subset. The ratios given in Table 1.3 are closest to one in the case $\alpha = 0.2$, which is what we select.

1.3.3 LSTM results and comparison with chain-ladder

To assess our LSTM network efficiency, we use the validation and testing datasets. The latter has not contributed to the training process in any way, and the network will see it for the first time. In the following, we analyze the prediction accuracy at both the aggregate and individual levels. Also, we compare the results of the LSTM with those of the aggregate chain-ladder model.

Using the payments that were previously hidden and $\mathcal{D}$ being either the validation or testing

Table 1.4 – Aggregate ratios using $\alpha = 0.2$ on the simulated datasets.

Datasets	Ratios of aggregate reserve		Ratios of aggregate ultimate	
	LSTM	Chain-ladder	LSTM	Chain-ladder
Validation	1.0271	1.0245	1.0033	1.0030
Testing	1.0385	0.9806	1.0048	0.9975

dataset, we evaluate the ratios of the aggregate reserve and ultimate given, respectively, by

$$\frac{\sum_{k\in\mathcal{D}}\sum_{j=t_k+1}^{n}\hat{p}_{k,j}\times\hat{Y}_{k,j}}{\sum_{k\in\mathcal{D}}\sum_{j=t_k+1}^{n}I_{k,j}\times Y_{k,j}} \quad \text{and} \quad \frac{\sum_{k\in\mathcal{D}}\left(\sum_{j=1}^{t_k}I_{k,j}\times Y_{k,j}+\sum_{j=t_k+1}^{n}\hat{p}_{k,j}\times\hat{Y}_{k,j}\right)}{\sum_{k\in\mathcal{D}}\sum_{j=1}^{n}I_{k,j}\times Y_{k,j}}.$$

Note that actual reserve refers to the actual outstanding payments between T^* and n, while actual ultimate corresponds to the observed total paid over n periods. Table 1.4 presents the ratios evaluated with the predictions of the LSTM and chain-ladder. For the validation dataset, we obtain very similar results with both models. For the testing dataset, the chain-ladder underestimates the reserve by 2% while our model overstates it by 3.9%. Both models perform well on the ultimate with ratios very close to one. However, the LSTM has the advantage of producing individual predictions. Moreover, the characteristics of the simulated data are very close to the chain-ladder hypotheses, which explains these results. Section 1.4 presents a case study with real data where we observe a more important difference between the performance of these models.

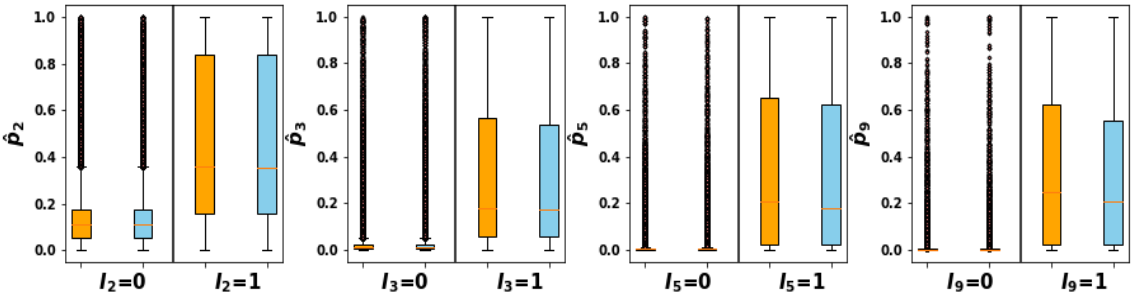


Figure 1.8 – Boxplots of the predicted probability $\hat{p}_{k,j}$ in terms of the observed $I_{k,j}$, for $j \in \{2, 3, 5, 9\}$ and for validation (orange) and testing (blue) simulated datasets.

At the individual level, we proceed to investigate the performance on the classification and regression tasks. Figure 1.8 depicts the boxplot of the predicted probability of payment in terms of the observed indicator, for periods $j \in \{2, 3, 5, 9\}$ from left to right, for validation and testing datasets. As the claim development progresses, we better predict $\Pr(I_{k,j} = 0)$ compared to $\Pr(I_{k,j} = 1)$. This is due to the short average lifetime of claims, 1.5 year, which implies a large proportion of zero values for the indicator $\{I_{k,j}, j = 3, \ldots, n\}$. Figure 1.9 illustrates the ROC curves for periods $j \in \{2, 3, 5, 9\}$. The closer the curve is to point $(0, 1)$, the better the model distinguishes between categories. We once again observe an improvement of LSTM probability predictions over periods.

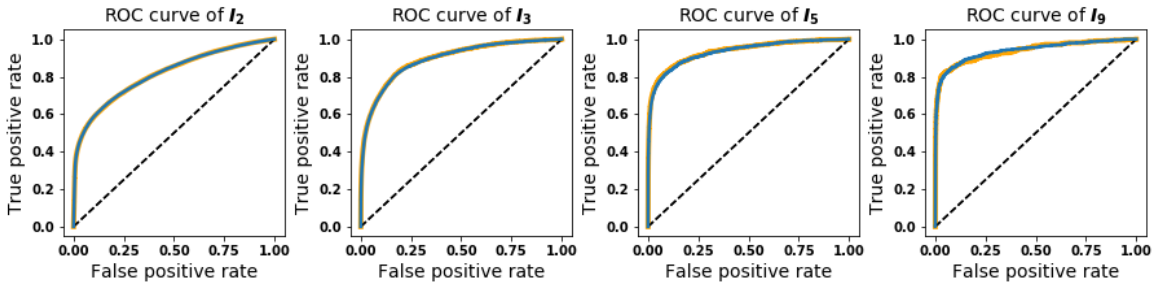


Figure 1.9 – ROC curves on the payment indicator classification, for periods $j \in \{2, 3, 5, 9\}$ for validation (orange) and testing (blue) simulated datasets.

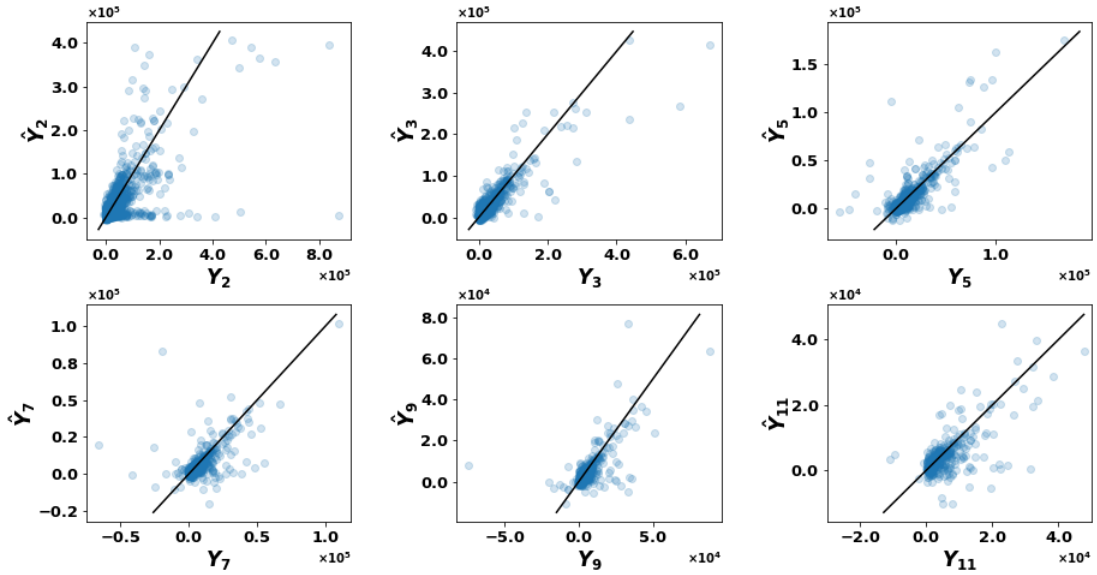


Figure 1.10 – Predicted payments in function of observed non-zero payments, for periods $j \in \{2, 3, 5, 7, 9, 11\}$ for the simulated testing dataset.

We illustrate in Figure 1.10 the scatterplots of incremental predicted payments, only for non-zero observed payments as that is how we train the regression task. We observe that the majority of points are close to the diagonal, which reflects the precision of predictions. In period $j = 2$, some predictions are close to zero whatever the value of the observed; this is expected as, in that case, the LSTM knows only one development period. The predictions improve as development progresses. From the LSTM output $\{(\hat{p}_{k,j}, \hat{Y}^*_{k,j}), j = 2, \dots, n\}$, we compute the expected individual reserve $\hat{R}_k = \sum_{j=t_k+1}^{n} \hat{p}_{k,j} \times \hat{Y}_{k,j}$. Figure 1.11 presents the scatterplots of $\hat{R}_k$ in terms of the observed reserves R_k per line of business. In general, the predictions of our network are satisfying. We note a few cases of over or under-estimation and a difference in claim sizes by line of business.

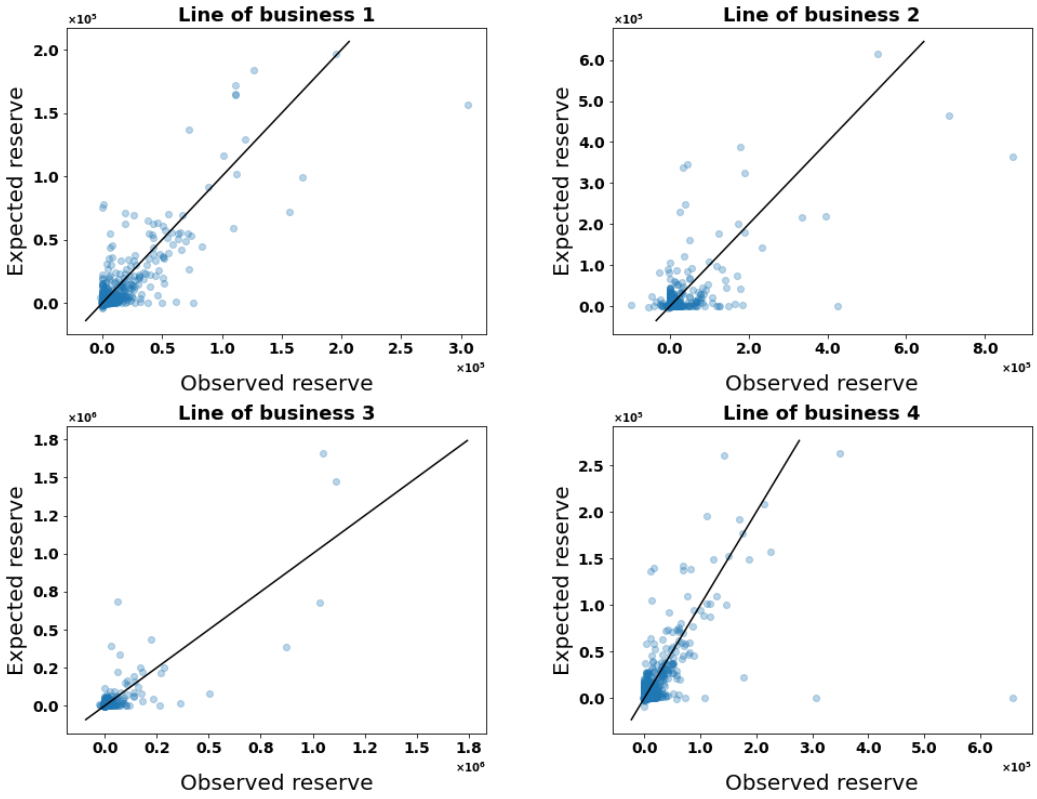


Figure 1.11 – Expected versus observed individual reserves for simulated testing dataset per line of business.

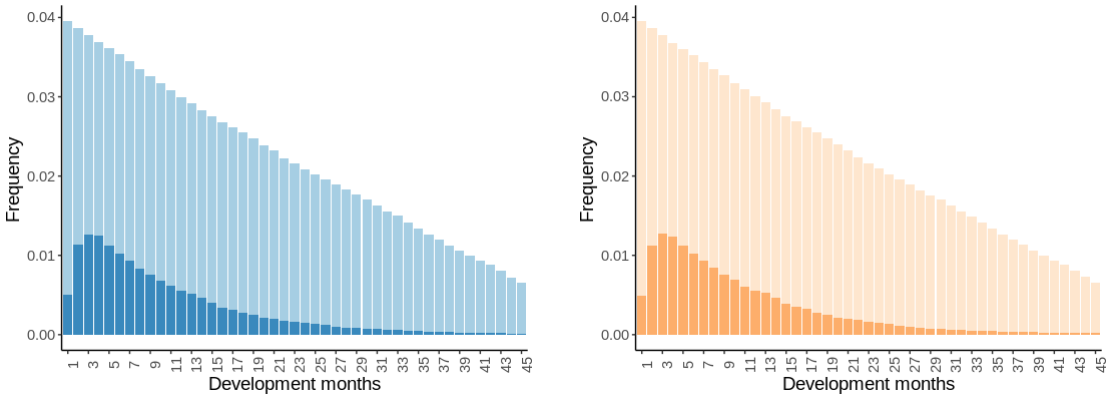
1.4 Real dataset

In this section, we illustrate our procedure on a real individual claims dataset provided by a large Canadian insurance company. First, we provide a brief description of the data. Then, we explain how we adapt our reserving model to the data and we describe the training process. As some extreme payments are hard to capture with NNs, we propose a large claim forecasting procedure. Our reserve estimates are compared with those obtained with the aggregate chain-ladder method.

1.4.1 Data description

We consider a real general insurance individual claims dataset from a Canadian insurance company. The detailed dataset relates to auto insurance policies and contains bodily injury claims that occurred from January 1^{st}, 2006 to August 31^{st}, 2010. There are $m = 53,677$ claimants associated to $41,916$ distinct claims. Note that we implicitly consider common aspects among claimants of the same claim through common covariates such as vehicle characteristics, and number of claimants.

From the reporting date, each claim's development is described through a monthly evolution



(a) (b)

Figure 1.12 – Frequency of observations without payment (light color) or non-zero payment (dark color) for training (a) and validation (b) datasets by development month.

of its individual information, divided into two categories. First, the static characteristics of claimant file k are contained in a vector $\boldsymbol{S}_k$, for $k = 1, \ldots, m$. This may include information on the claim (date, location, number of claimants, driver experience), the claimant (age, gender), and the vehicle (age, number of cylinders). The second category groups dynamic information in $\boldsymbol{D}_{k,j}$ at each period j, for instance, the number of claimants, the nature and number of injuries and the payment. Tables 1.B.1 and 1.B.2 list the static and dynamic features used in our model.

First, we pre-process the real dataset. We correct wrong entries, e.g., an accident date greater than the reporting date. To deal with missing values, we apply a multiple imputation by chained equations (MICE), using the R package *mice* (Van Buuren and Groothuis-Oudshoorn, 2011). For categorical variables with many modalities (e.g., postal code), this procedure is computationally intensive. We hence apply the MICE approach to only ten of the 17 available variables with missing values, and we drop the other ones. See Table 1.B.1 for a description of the variable pre-processing.

We set the evaluation date T^* as August 31$^{\text{st}}$, 2010. By analyzing the development of claimant files closed as of T^*, we observe that their lifetime is on average three months and they have a 0.3% probability of reopening. The probability of a claimant file lasting longer than 45 months is very low. Therefore, we are interested in studying only the first 45 months of claimant file development, i.e., we set $n = 45$ for our LSTM. This development censorship allows to have enough observations for the network to train on the last period predictions. Also, we hide all information observed beyond T^*, then stratify the dataset into training (60%), validation (20%), and testing (20%) datasets. To ensure an homogeneous dataset split, the same analyses are carried out on the training and validation datasets. According to Figure 1.B.1, we observe similar curves for both.

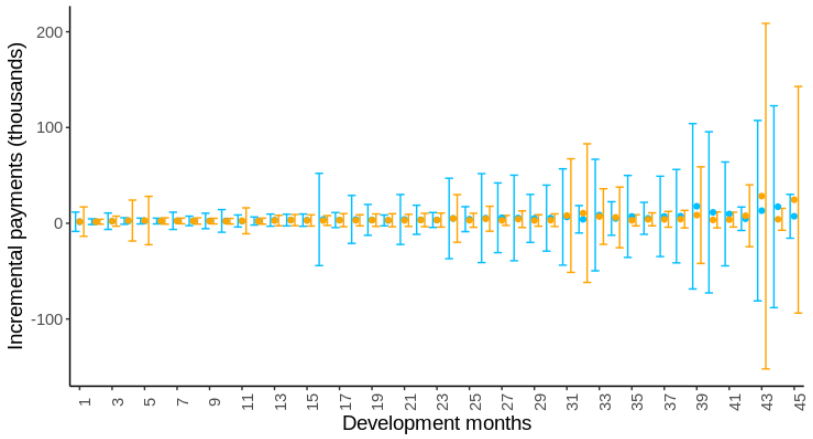


Figure 1.13 – Mean and standard deviation of non-zero incremental payments at the evaluation date T^* for training (blue) and validation (orange) datasets by development month.

The real dataset has several features in common with the simulated dataset discussed in Section 1.3. In both, we observe reopenings, and around 30% of claimants files settled without any payment even though they may remain open for several periods. However, compared to the simulated dataset, the real data has the advantage of having a monthly (rather than yearly) detailed development with a greater number of dynamic features (other than incremental payments). If we wanted to include the future development of these features in the network, it would imply an additionnal task. To avoid adding complexity, we consider the evolution of dynamic claimant file characteristics other than the payments until T^* and then keep the last observed value for the remaining development months. Thus, we take advantage of this information to improve the network predictions for the first development months.

The number of observed targets used to learn the classification and regression tasks differs and may impact the complexity of the optimization of functions (1.2) and (1.4). For the training and validation datasets, Figure 1.12 illustrates two frequency distributions. The first one, in light color, is related to observed zero incremental payments per development month. The second distribution, in dark, relates only to non-zero incremental payments among all observed payments employed to compute the regression loss (1.4). The concatenation of the two colors corresponds to the distribution of observed indicators $I_{k,j}$, for $j = 1, \ldots, t_k$, used to compute the classification loss (1.2). There are fewer observations within both datasets to train the regression task than the classification. This is expected given the large percentage of claimant files settled without any payment.

To investigate in details the incremental payments, we depict in Figure 1.13 error bars representing the standard deviation around the mean for the non-zero payments, per development month. We observe huge variances for some periods indicating the presence of extremely large payments. Recall that we did not observe extreme claims in the simulated dataset from Section 1.3. We propose in Section 1.4.2 an approach to handle large payments.

1.4.2 Model for extreme payments

Large or catastrophic claims are extreme events leading to very high payments and can be modeled using Extreme Value Theory (EVT). Given a high threshold $u > 0$, EVT allows to model the conditional distribution of the excess loss over u with a generalized Pareto distribution (GPD) as limiting distribution. For a general introduction to EVT, see, e.g., Coles et al. (2001).

Using only the training dataset, we consider $\boldsymbol{Y} = \{Y_{k,j} : Y_{k,j} \neq 0, j = 1, \ldots, t_k,\ k = 1, \ldots, m\}$ the vector of non-zero incremental payments known at the evaluation date. We aim to model the tail behavior of $\boldsymbol{Y}$ using EVT. We set a sufficiently high threshold u, and estimate the parameters of the GPD associated to the random excess $(Y_{k,j} - u | Y_{k,j} > u)$. Then, we verify that the empirical behavior of $\boldsymbol{Y}$ can be described by the selected GPD.

To choose a threshold u, we use the mean excess plot, and u is selected as the lowest threshold for which the curve becomes approximately linear. Based on the detailed analysis provided in Appendix 1.C, we set $u = 32,000$. With the selected threshold, we estimate the GPD parameters by maximum likelihood through a numerical optimization with the R package *ismev* (Heffernan, 2018). We obtain shape $\hat{\sigma} = 0.957851$ and scale $\hat{\lambda} = 14,044$. Therefore, $\hat{E}[Y_{k,j} - u | Y_{k,j} > u] = \hat{\lambda}/(1 - \hat{\sigma}) = 333,199$, i.e., given that a payment exceeds the threshold $u = 32,000$, it will cost on average \$365,199. The goodness-of-fit of the estimated GPD is verified in Appendix 1.C.

Now that we have the expected excess amount, we move to accurately estimating the probability that a payment exceeds u. Note that we have only 0.5% of available non-zero payments $\boldsymbol{Y}$ above the threshold $u = 32,000$. In addition to the development period as a predictive variable, an analysis of the payment history, inspired by Côté et al. (2021), revealed that it is less likely to observe large payments after multiple consecutive periods without payment. We use a generalized linear model (GLM) for the probability of large payment knowing the development period and the number of previous months without payment. Given the few observations (sometimes less than five) exceeding u for some development periods, we group months in quarter periods and also merge development months 39 to 45. We keep the first month apart given its particular behavior. Figure 1.14 depicts the predicted relative probabilities of large payment obtained with our GLM, per development month. The number of previous periods without payment is represented using the color scale; as it increases (colder color), the probability of extreme payment decreases. Fitted probabilities of extreme payments also increase with development month.

The large claims analysis developed in this section confirms our interpretation of Figure 1.13. To train our LSTM network on the real dataset, we propose to censor payments in order to reduce variance and improve learning. In that case, the LSTM prediction can be interpreted as the conditional expected payment given that it does not exceed the threshold u, and expressed,

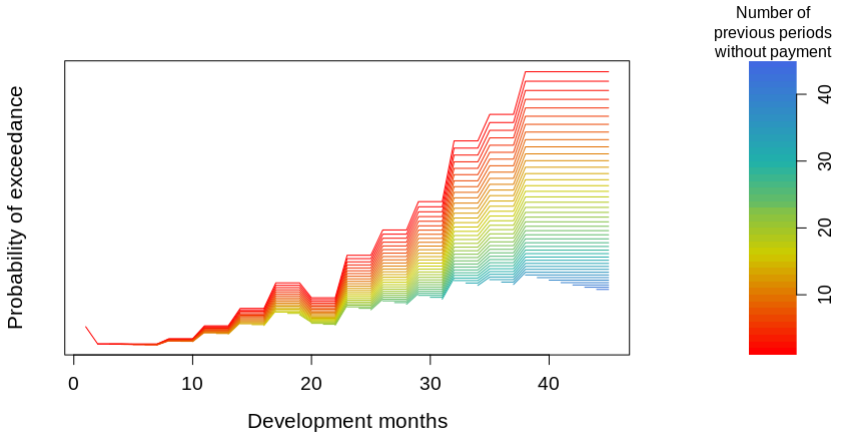


Figure 1.14 – Relative predicted probabilities of large payment in terms of development month and number of previous periods without payment.

for claimant file k and period $j = t_k + 1, \ldots, n$, by

$$\begin{aligned}\hat{E}[Y_{k,j}|Y_{k,j} \leq u] &= \hat{\Pr}\left(I_{k,j} = 1\right) \times \hat{E}[Y_{k,j}|Y_{k,j} \leq u, I_{k,j} = 1] \\ &= \hat{p}_{k,j} \times \hat{Y}_{k,j}. \end{aligned} \tag{1.6}$$

We now aim to investigate an adjustment to the network's prediction to take into account the possibility that the payment is large. For claimant file k and period j, we have

$$\begin{aligned}&\hat{E}[Y_{k,j}] \\ &= \hat{\Pr}\left(I_{k,j} = 1\right) \times \hat{E}[Y_{k,j}|I_{k,j} = 1] \\ &= \hat{p}_{k,j} \times \left\{\hat{\Pr}\left(Y_{k,j} > u\right) \hat{E}[Y_{k,j}|Y_{k,j} > u, I_{k,j} = 1] + \hat{\Pr}\left(Y_{k,j} \leq u\right) \hat{E}[Y_{k,j}|Y_{k,j} \leq u, I_{k,j} = 1]\right\} \\ &= \hat{p}_{k,j} \times \left[\hat{\Pr}\left(Y_{k,j} > u\right)\left(u + \frac{\hat{\lambda}}{1-\hat{\sigma}}\right) + \{1 - \hat{\Pr}\left(Y_{k,j} > u\right)\}\hat{Y}_{k,j}\right], \end{aligned} \tag{1.7}$$

with $\hat{\Pr}\left(Y_{k,j} > u\right)$ the probability of exceedance u obtained with our GLM. If we consider (1.7) instead of (1.6) for all forecasted payments, we may overstate the reserve as we illustrate in Section 1.4.4. Through an in-depth analysis of the results obtained for the training and validation datasets, we propose applying the adjustment given by (1.7) for selected observations with greater probability of being large if the network was fed with uncensored payments. To do so, we analyze the distribution of predictions $\hat{Y}_{k,j}$ of large censored payments $Y_{k,j}$, to choose a threshold ζ and we formulate the incremental payment prediction as follows

$$\hat{E}[Y_{k,j}] = \begin{cases} \hat{p}_{k,j} \times \hat{Y}_{k,j}, & \text{if } \hat{Y}_{i,j} \leq \zeta \\ \hat{p}_{k,j} \times \left[\hat{\Pr}\left(Y_{k,j} > u\right)\left(u + \frac{\hat{\lambda}}{1-\hat{\sigma}}\right) + \{1 - \hat{\Pr}\left(Y_{k,j} > u\right)\}\hat{Y}_{k,j}\right], & \text{otherwise.} \end{cases} \tag{1.8}$$

Table 1.5 – Hyper-parameters used to train the LSTM network on the real dataset.

Hyper-parameter	Value	Hyper-parameter	Value
Context size c	16	Hidden size	256
Batch size d	1024	Learning rate l_r	0.05
Reduce on plateau r_p	10 epochs	Early stopping t_{es}	15 epochs

Delong and Wüthrich (2020) also use a GPD for large claims with periodic parameter estimation. However, since their neural net is not recurrent, at each period, they randomly select a proportion of observations with a high propensity to generate large claims according to some key features and estimate the prediction outside the network. The choice of the percentage and key features is very sensitive and requires a good knowledge about the claims nature. In our recurrent network, the prediction process is linked from period to period and considers the claimant file past development. According to our observations, it is more efficient to set a threshold ζ instead of a percentage to adjust large claim predictions. Unlike Delong and Wüthrich (2020), we use a GLM to evaluate the probability of large claims based on some dynamic covariates.

1.4.3 Model training

Using censored payments $\min(Y_{k,j},\ u)$, we train our LSTM network to learn optimal weights and predict the individual reserve accurately. Due to the difference in the number of observations used to compute functions (1.2) and (1.4), we note a greater difficulty to learn the regression task than the classification one during our training experiments. Indeed, throughout the first epochs, we may observe a large difference between a prediction $\hat{Y}_{k,j}$ and its target $Y_{k,j}$. To reduce the regression loss scale values and make them closer to cross-entropy values, we choose the absolute error function f_{AE} instead of the squared error f_{SE} in losses (1.4) and (1.5).

After several LSTM training experiments, we set the hyper-parameter combination, given in Table 1.5, that optimizes the model performance. Compared to the LSTM network fitted to the simulated dataset in Section 1.3, the real data has inputs $\boldsymbol{X}_{k,j}$, $j = 1, \dots, n-1$, of higher dimension due to a greater number of dynamic features. Thus, we increased the size of the hidden state $\boldsymbol{h}_{k,j}$, for $j = 1, \dots, n-1$, to avoid losing information and enhance network performance. Note that a higher dimension requires more memory during the training process. Moreover, according to the dataset characteristics presented in Section 1.4.1, when the mini-batch size is too small, the network will see a non-zero payment with low probability, particularly for the last development months. Hence, we choose a batch size of 1024 to have enough targets to learn the regression task and enough mini-batches per epoch to train our model.

To set the hyper-parameter α without using the information after the evaluation date T^*,

Table 1.6 – Aggregate ratios based on censored payments, with $\alpha = 0.5$.

Datasets	Ratios of aggregate reserve		Ratios of aggregate ultimate	
	LSTM	Chain-ladder	LSTM	Chain-ladder
Validation	**1.0044**	1.1124	**1.0017**	1.0379
Testing	**0.9970**	1.1238	**0.9988**	1.0436

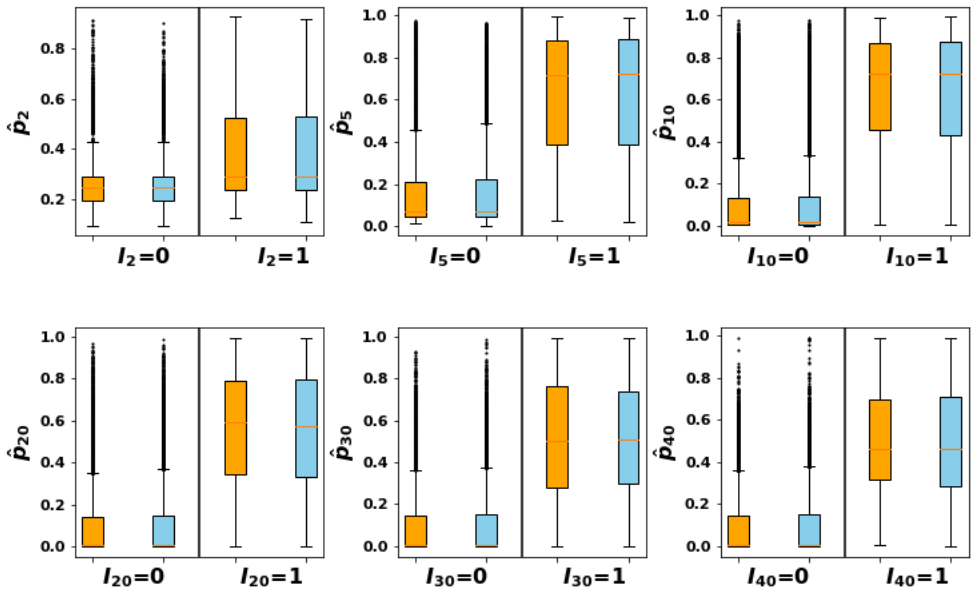


Figure 1.15 – Boxplots of the predicted probability $\hat{p}_{k,j}$ in terms of the observed indicator $I_{k,j}$, for $j \in \{2, 5, 10, 20, 30, 40\}$ and for validation (orange) and testing (blue) real datasets.

i.e., August 31$^{\text{st}}$, 2010, we follow the same approach applied in Section 1.3.2 on the simulated dataset. We set May 31$^{\text{st}}$, 2010, as a reference date, denoted T^{**}. We select the claimant files from the validation dataset with at least three months of development observed at T^{**} such that $t_k^{**} > 0$, leading to a loss of 20% of the observations. We evaluate the metric ratios $RR(T^{**}, T^{*})$ and $RU(T^{**}, T^{*})$ on this validation subset. Again, our results show no monotonic relationship between these ratios and α. LSTMs trained with either α equal to 0.5 or 1.5 lead to ratios very close to one and outperform the other trained networks. Both choices produce a very similar performance at the aggregate level. At the individual level, on the validation subset, we note an overestimation of the small actual outstanding payments for the model with $\alpha = 1.5$. Thus, we choose $\alpha = 0.5$.

1.4.4 LSTM results with censored payments

In this section, we assess the efficiency of our LSTM network trained with censored payments based on the validation and testing datasets. First, we evaluate the accuracy of predictions at the aggregate level and then we compare it to the aggregate chain-ladder model. Second, at the individual level, we investigate the performance of the regression and classification

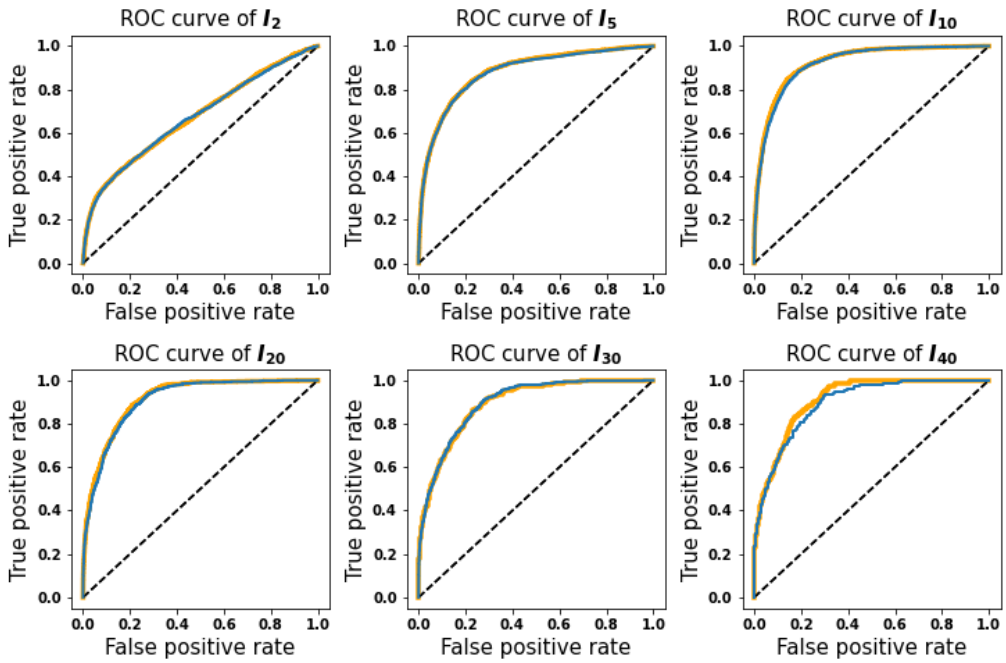


Figure 1.16 – ROC curves on the payment indicator classification, for periods $j \in \{2, 5, 10, 20, 30, 40\}$ for validation (orange) and testing (blue) real datasets.

tasks. Note that all computations in this section are made only with payments censored at $u = 32,000$.

Tables 1.6 presents the ratios of the aggregate reserve and ultimate obtained with both LSTM and chain-ladder. The expressions based on censored payments are given, respectively, by

$$\frac{\sum_{k\in\mathcal{D}}\sum_{j=t_k+1}^{n}\hat{p}_{k,j}\times\hat{Y}_{k,j}}{\sum_{k\in\mathcal{D}}\sum_{j=t_k+1}^{n}I_{k,j}\times\min\left(Y_{k,j},u\right)}$$

and

$$\frac{\sum_{k\in\mathcal{D}}\left(\sum_{j=1}^{t_k}I_{k,j}\times\min\left(Y_{k,j},u\right)+\sum_{j=t_k+1}^{n}\hat{p}_{k,j}\times\hat{Y}_{k,j}\right)}{\sum_{k\in\mathcal{D}}\sum_{j=1}^{n}I_{k,j}\times\min\left(Y_{k,j},u\right)}.$$

Note that the chain-ladder development factors are estimated using the censored training dataset. For validation and testing datasets, our network outperforms the chain-ladder with ratios very close to one. The chain-ladder overstates the reserve by more than 10%. Thus, our model can better capture claim trends, which leads to a more accurate forecast using individual information.

Figure 1.15 presents boxplots of the predicted probability of payment, equivalently the probability that the indicator $I_{k,j}$ takes value 1 or 0 for periods $j \in \{2, 5, 10, 20, 30, 40\}$. For validation and testing datasets, we observe that we better predict $\Pr(I_{k,2} = 0)$ compared to $\Pr(I_{k,2} = 1)$. As the claimant file process progresses and the network learns the development,

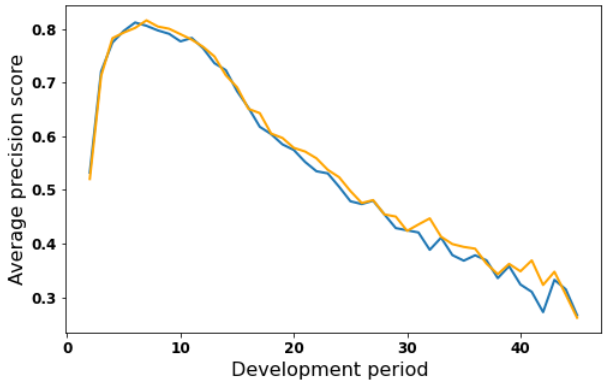


Figure 1.17 – AUROC per development month for validation (orange) and testing (blue) real datasets.

we note an improvement in the prediction of probabilities. However, there are very few payments to be made in last development months leading to an unbalanced dataset with a large presence of zero values for indicators. This can bias the interpretation of the boxplots for last periods.

The ROC curves for the classification are shown in Figure 1.16 for periods $j \in \{2, 5, 10, 20, 30, 40\}$, for both validation and testing datasets. As for Figure 1.15, we observe that the more the network knows the claimant file's history, the better is the prediction performance. Besides, to better evaluate the classification task without the impact of unbalanced observations, we use the AUROC per development period illustrated as a curve in Figure 1.17. We note an improvement during the first ten development periods, followed by a decrease in prediction accuracy. It is most probably due to the number of observations available for training. We can improve the network's performance by providing more observations, as in the simulated dataset studied in Section 1.3.

Now, we move on to analyze the predicted payments. We illustrate in Figure 1.18 the scatterplots of incremental predicted payments for non-zero observed payments in the testing dataset. Moving forward in the development, we note an improvement in the prediction accuracy with points closer to the diagonal. On the other hand, beyond month 20, there is an underestimation of large observed payments, but, as discussed in Section 1.4.1, few observations are available to train the regression task for the latest developments.

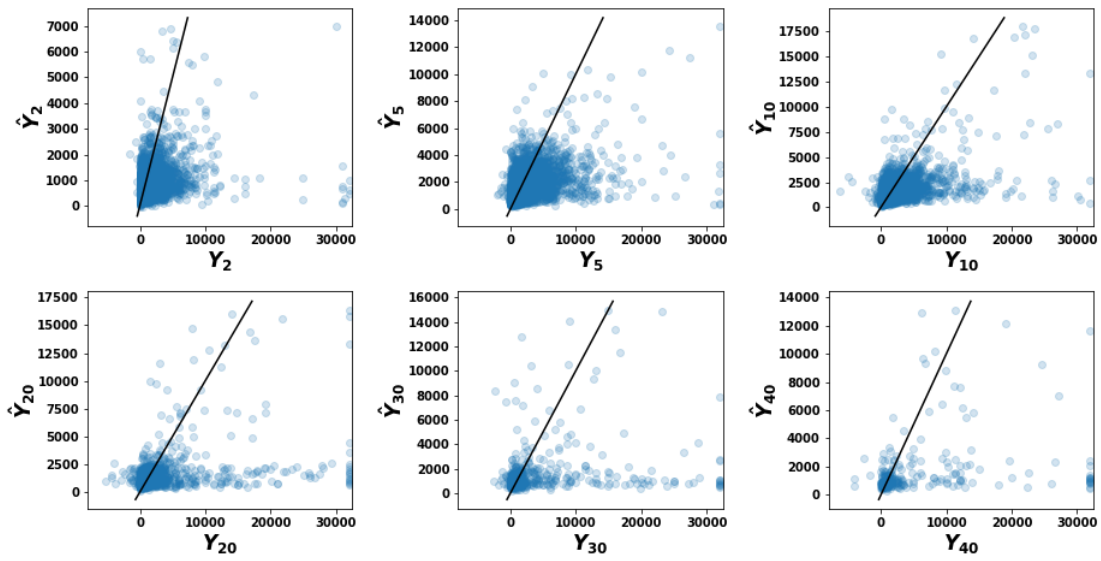


Figure 1.18 – Predicted payments in function of observed non-zero payments for development month $j \in \{2, 5, 10, 20, 30, 40\}$ for real testing dataset.

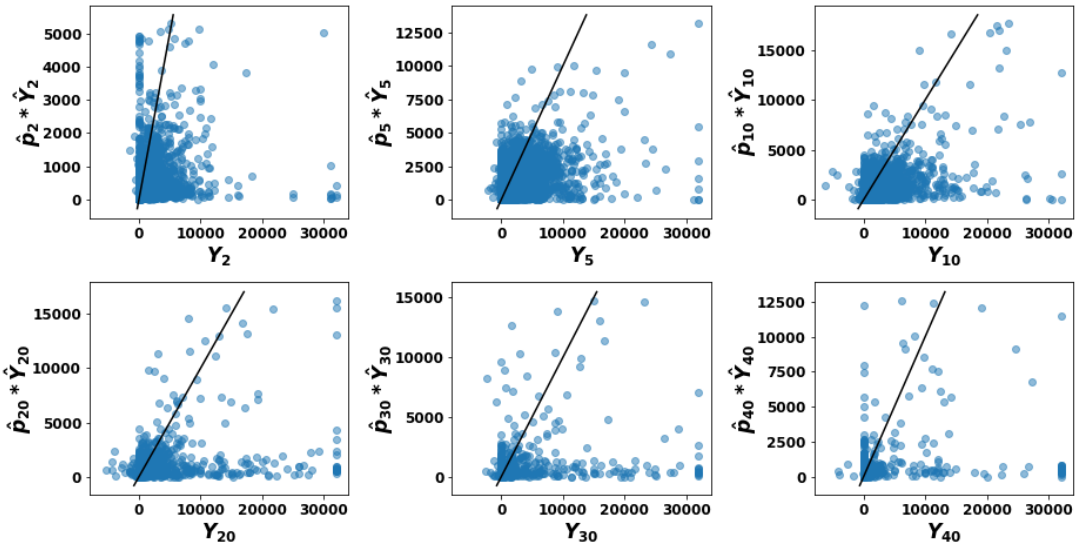


Figure 1.19 – Expected versus observed payments for month $j \in \{2, 5, 10, 20, 30, 40\}$ in the real testing dataset.

By multiplying the predictions of the regression and classification tasks, we obtain the expected payments $\hat{p}_{k,j} \times \hat{Y}_{k,j}$. They are presented for some periods in terms of the observed in Figure 1.19. We note an overestimation of zero payments for period $j = 2$. After month 20, we have an underestimation of the large observed payments and an overestimation of zero payments. By summing the expected incremental payments $\hat{p}_{k,j} \times \hat{Y}_{k,j}$ over future periods, we deduce the censored reserve $\hat{R}_k^c = \sum_{j=t_k+1}^{n} \hat{p}_{k,j} \times \hat{Y}_{k,j}$ for claimant file k. Figure 1.20 illustrates the scatterplots of $\hat{R}_k^c$ versus the observed censored reserves given by $R_k^c = \sum_{j=t_k+1}^{n} I_{k,j} \times$

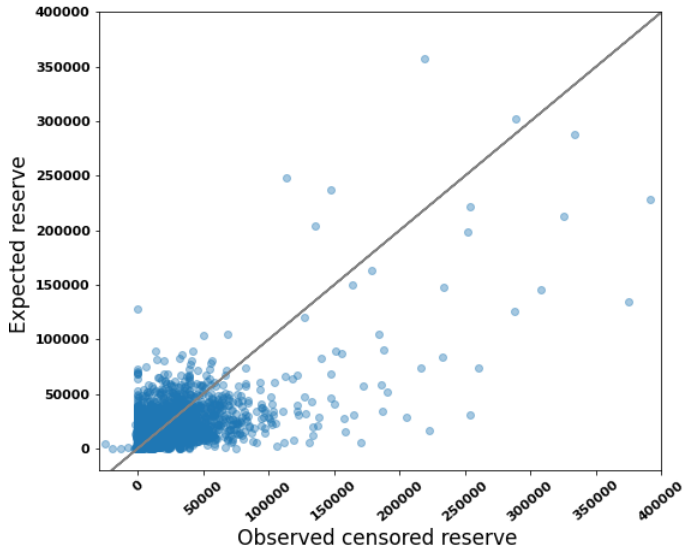


Figure 1.20 – Expected versus observed censored reserves for real testing dataset.

Table 1.7 – Summary statistics of LSTM predictions for censored observed large payments.

	mean	std	$q_{0.25}$	$q_{0.50}$	$q_{0.75}$
Predicted payment $\hat{Y}$	3220	4126	1002	1510	2761
Expected payment $\hat{p} \times \hat{Y}$	2682	4216	334	842	2479

$\min(Y_{k,j},\ u)$. Most of the points are close to the diagonal with a few over or under-estimation for large claims.

1.4.5 Results for individual reserving model with large claims

In the following, we adjust the LSTM outputs, presented in Section 1.4.4, using the procedure described in Section 1.4.2 which considers large claims. For the training dataset, Table 1.7 presents the mean, the standard deviation and some quantiles of the LSTM predictions associated to large observed payments censored at \$32,000. The values in Table 1.7 are much smaller than the target $u = 32,000$. Indeed, the model prediction is like an expectation of observed incremental payments. Since there are few censored observations, the network tends to understate them. If we had not censored the payments, the discrepancies would have been even greater.

From Table 1.7, we can see the order of magnitude of the threshold ζ. We propose to apply Eq. (1.8) according to different thresholds ζ, and to compare the adjusted predictions with the uncensored observed payments. For the training, validation and testing datasets, Tables 1.8 and 1.9 present the ratio of the aggregate reserve and ultimate, respectively, for several ζ values. Both are computed from the adjusted predictions and the uncensored observed payments as

follows:

$$\frac{\sum_{k\in\mathcal{D}}\sum_{j=t_k+1}^{n}\hat{E}[Y_{k,j}]}{\sum_{k\in\mathcal{D}}\sum_{j=t_k+1}^{n}I_{k,j}\times Y_{k,j}} \quad \text{and} \quad \frac{\sum_{k\in\mathcal{D}}\left(\sum_{j=1}^{t_k}I_{k,j}\times Y_{k,j}+\sum_{j=t_k+1}^{n}\hat{E}[Y_{k,j}]\right)}{\sum_{k\in\mathcal{D}}\sum_{j=1}^{n}I_{k,j}\times Y_{k,j}}.$$

When $\zeta = 0$, Eq. (1.8) amounts to adjusting the network outputs to consider large payments occurence for all positive predictions $\hat{Y}_{k,j}$. Unfortunately, this approach grossly overestimates payments, especially for small ones. For $\zeta = \infty$, no adjustment is applied and the network predictions are compared to the actual uncensored payments. Obviously, the ratios are very small since our predictions underestimate large payments that the network has never seen during its training. In addition, we set three values of ζ that are reasonable based on Table 1.7. For $\zeta = 2000$, the ratios are closer to one compared to $\zeta = 0$, but we still overestimate. Applying Eq. (1.8) with $\zeta = 3000$, we underestimate the aggregate reserve by 10% for the three datasets. We obtain the best results for threshold $\zeta = 2500$, where the ratios of the aggregate reserve and ultimate are very close to one.

Moreover, we compute the ratios using the chain-ladder method, with development factors evaluated from the uncensored training dataset. According to Table 1.8, the chain-ladder overestimates the aggregate reserve by 30%. Our LSTM model with adjustment for large payments using threshold $\zeta = 2500$ outperforms the chain-ladder on the aggregate results and has the advantage of yielding individual reserves.

Table 1.8 – Ratios of the aggregate predicted reserve to observed aggregate reserve.

Datasets	Chain-ladder	$\zeta = 0$	$\zeta = 2000$	$\zeta = 2500$	$\zeta = 3000$	$\zeta = \infty$
Training	1.3479	5.5942	1.4110	**1.0309**	0.9048	0.7976
Validation	1.3349	5.6437	1.3623	**0.9911**	0.8808	0.7850
Testing	1.3176	5.3250	1.4426	**1.0547**	0.9295	0.7878

Table 1.9 – Ratios of the aggregate predicted ultimate to observed aggregate ultimate.

Datasets	Chain-ladder	$\zeta = 0$	$\zeta = 2000$	$\zeta = 2500$	$\zeta = 3000$	$\zeta = \infty$
Training	1.1264	2.9361	1.1732	**1.0130**	0.9599	0.9147
Validation	1.1227	3.0057	1.1564	**0.9961**	0.9485	0.9071
Testing	1.1193	2.8732	1.1917	**1.0237**	0.9694	0.9081

1.5 Conclusion

In this paper, we propose an individual reserving model based on an LSTM network and a large claim forecasting procedure. Our model has the advantage of being easily adaptable to the structure of any claim dataset with detailed development. Our procedure involves only the information available at the current evaluation date. Our network shows accurate predictions and outperforms the chain-ladder model in the two case studies carried out with a simulated and a real dataset.

In forthcoming research work, we are interested in adding other tasks to the network while ensuring a balanced optimization of the loss functions, and in considering model uncertainty. Also, we could further investigate the hyper-parameter selection procedure to enhance network performance.

Acknowledgments

The authors sincerely thank Christopher Blier-Wong, Julien Fagnan, Luc Lamontagne, François Laviolette, Étienne Marceau and Frédérik Paradis for useful discussions. The authors gratefully acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (CRDPJ 515901-17, Cossette: RGPIN-2017-04273, Côté: RGPIN-2019-04190) and the Big Data Research Center at Université Laval. We also thank an anonymous insurance company for providing the data.

1.A Simulated data setup

The Gabrielli and Wüthrich (2018) simulator must be provided with a few parameters such as the expected number of claims, a probability distribution for the allocation of claims to the lines of business (LoB), growth rates for the number of claims over the years for each LoB, and standard deviations for the total individual claim amounts and recoveries. Table 1.A.1 presents the values of these parameters in the case study from Section 1.3.

Table 1.A.1 – Parameters used to generate the simulated dataset.

Parameter	Value
Expected number of claims	1,000,000
Probability distribution for claim allocation	$\Pr(\text{LoB} = 1) = \Pr(\text{LoB} = 4) = 0.25$, $\Pr(\text{LoB} = 2) = 0.30$, $\Pr(\text{LoB} = 3) = 0.20$
Growth rate for the number of claims	0.01 (same for all lines of business)
Standard deviation	0.85 (for payments and recoveries)

1.B Description of the real dataset

Table 1.B.1 – List of static covariates in the real dataset used in our model.

Static features	Categorical	# of categories	Missing values	Pre-processing
Day of occurrence	✗	—	✗	—
Month of occurrence	✗	—	✗	—
Year of occurrence	✗	—	✗	Shifted with 2006
Reporting delay (days)	✗	—	✗	—
Company name	✓	5	✗	Indexed
Claimant age	✗	—	✓	Scaled to $[0, 1]$
Claimant gender (indicator)	✓	2	✓	—
Driver experience (years)	✗	—	✓	—
Vehicle type (indicator)	✓	2	✓	—
Vehicle age (years)	✗	—	✓	—
Drive train	✗	—	✓	—
Number of cylinders	✗	—	✓	—
Vehicle body	✓	11	✓	Indexed
Convertible (indicator)	✓	2	✓	—
Claim on highway	✓	4	✗	Indexed
Claim province	✓	11	✓	Indexed

Table 1.B.2 – List of dynamic covariates in the real dataset used in our model.

Dynamic features	Pre-processing
Degree of claimant injury	—
Incremental monthly payment	Centered and scaled as in Eq. (1.1)
Number of claimants	—
Number of reported injuries	—

Figure 1.B.1 presents the empirical frequency of three continuous variables in the real dataset studied in Section 1.4: claimant age, driver experience, and vehicle age. We observe similar curves for training and validation datasets. According to Figure 1.B.1(a), young drivers are the riskiest due to their short driving experience. This observation is also supported by Figure 1.B.1(b), illustrating the decrease in claim occurrence with driving experience. According to Figure 1.B.1(c), a new vehicle is riskier than an older one. This can be explained by the vehicle's power, encouraging excess speed, or fewer old cars on the road.

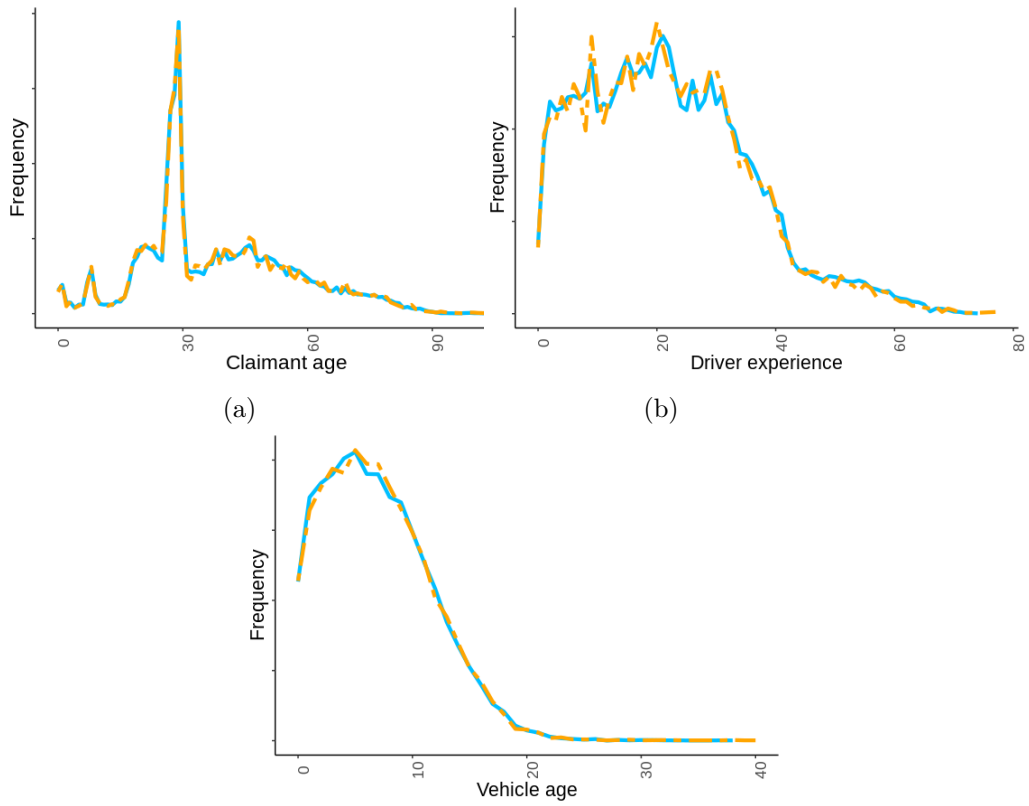


Figure 1.B.1 – Frequency of claimant age (a), driver experience (b) and vehicle age (c), for training (solid blue) and validation (dashed orange) datasets. The y axes are masked for confidentiality.

1.C Diagnostic plots for extreme value model

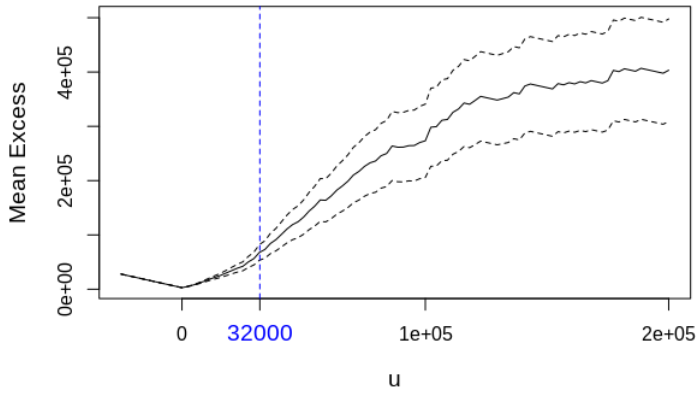


Figure 1.C.1 – Mean excess plot for incremental payments.

Figure 1.C.1 represents the mean excess plot with approximate 95% confidence intervals for the incremental payments $\boldsymbol{Y}$ as defined in Section 1.4.2. The curve is approximately linear between

32,000 and 60,000, then it decays sharply. For extreme values of u, the mean excess plot is unreliable due to the limited number of observations on which the estimate and confidence interval are based. Given that the 0.995 quantile of $\boldsymbol{Y}$ is around 31,600, we set $u = 32,000$. This choice makes it possible to keep enough extreme observations for a reliable estimation of the GPD parameters. We use a second tool to confirm this choice, which consists in fitting the GPD at a range of thresholds and looking for parameter estimates stability. Figure 1.C.2 depicts the plots of the scale σ and shape λ estimates against thresholds with 95% confidence bounds. We select the threshold u as the lowest value for which the estimates remain near-constant. Again, $u = 32,000$ appears appropriate.

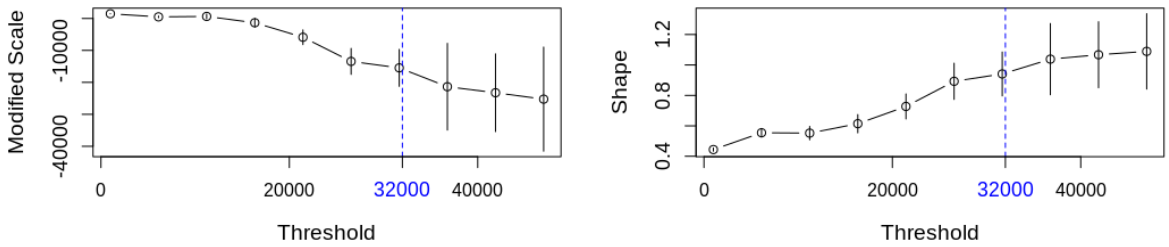


Figure 1.C.2 – Parameter estimates against threshold for incremental payments.

To assess the goodness of fit of the estimated GPD to $(Y_{k,j} - u | Y_{k,j} > u)$, we use the cumulative probability (P-P) and quantile (Q-Q) plots. If the estimated GPD model fits the exceedance observations well, then both plots should be approximately linear. Figure 1.C.3 shows both diagnostic plots, with points approximately aligned for the P-P plot. In the Q-Q plot, the points form a line for values below 200,000 (0.9995 quantile), and deviate from the diagonal for some large exceedance values outside the 95% confidence interval. However, we have too few exceedances above 200,000 to make meaningful conclusions. Thus, both plots validate the estimated GPD model.

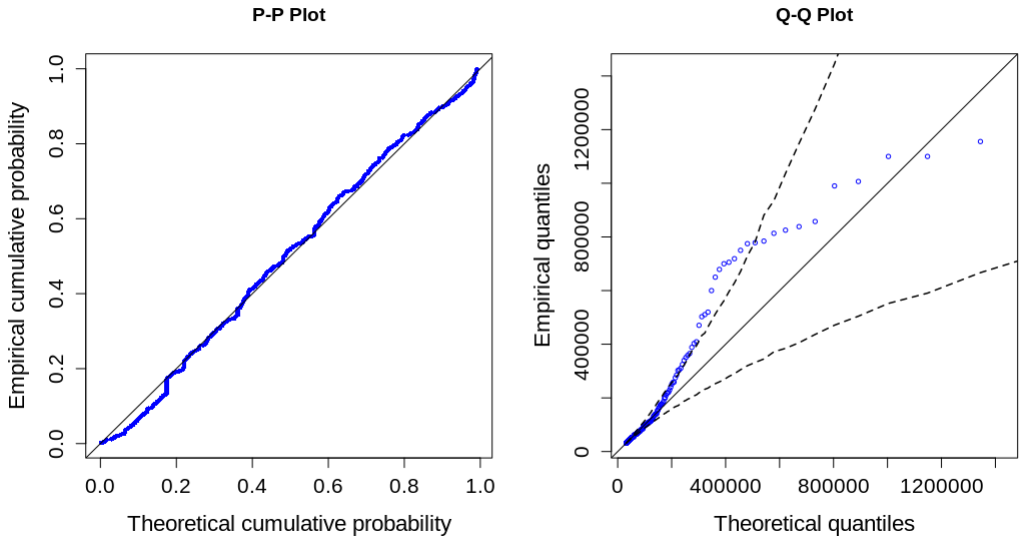


Figure 1.C.3 – Diagnostic plots for threshold excess model fitted to incremental payments.

1.6 Bibliography

Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014:649–669.

Arjas, E. (1989). The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin*, 19(2):139–152.

Baudry, M. and Robert, C. Y. (2019). A machine learning approach for individual claims reserving in insurance. *Applied Stochastic Models in Business and Industry*, 35(5):1127–1155.

Charpentier, A. and Pigeon, M. (2016). Macro vs. micro methods in non-life claims reserving (an econometric perspective). *Risks*, 4(2):12.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer, London.

Côté, M.-P., Genest, C., and Stephens, D. A. (2021). A Bayesian approach to modeling multivariate multilevel insurance claims in the presence of unsettled claims. *Bayesian Analysis*, DOI: 10.1214/20-BA1243.

Delong, Ł. and Wüthrich, M. V. (2020). Neural networks for the joint development of individual payments and claim incurred. *Risks*, 8(2):33.

Drossos, K., Gharib, S., Magron, P., and Virtanen, T. (2019). Language modelling for sound event detection with teacher forcing and scheduled sampling. *arXiv preprint arXiv:1907.08506*.

Duval, F. and Pigeon, M. (2019). Individual loss reserving using a gradient boosting-based approach. *Risks*, 7(3):79.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518.

Gabrielli, A. (2021). An individual claims reserving model for reported claims. *European Actuarial Journal*, DOI:10.1007/s13385-021-00271-4.

Gabrielli, A., Richman, R., and Wüthrich, M. V. (2020). Neural network embedding of the over-dispersed Poisson reserving model. *Scandinavian Actuarial Journal*, 2020(1):1–29.

Gabrielli, A. and Wüthrich, M. (2018). An individual claims history simulation machine. *Risks*, 6(2):29.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Haastrup, S. and Arjas, E. (1996). Claims reserving in continuous time; A nonparametric Bayesian approach. *ASTIN Bulletin*, 26(2):139–164.

Heffernan, J. E. (2018). *ISMEV: An Introduction to Statistical Modeling of Extreme Values*. R package version 1.42.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural computation*, 9(8):1735–1780.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

Kendall, A., Gal, Y., and Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491.

Kuo, K. (2020). Individual claims forecasting with Bayesian mixture density networks. *arXiv preprint arXiv:2003.02453*.

Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance I. *ASTIN Bulletin*, 23(1):95–115.

Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin*, 29(1):5–25.

Pigeon, M., Antonio, K., and Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin*, 43(3):399–428.

Pigeon, M., Antonio, K., and Denuit, M. (2014). Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics*, 58:121–131.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Sundermeyer, M., Schlüter, R., and Ney, H. (2012). LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*, pages 194–197.

Taylor, G. (2019). Loss reserving models: Granular and machine learning forms. *Risks*, 7(3):82.

Taylor, G., McGuire, G., and Sullivan, J. (2008). Individual claim loss reserving conditioned by case estimates. *Annals of Actuarial Science*, 3(1-2):215–256.

Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67.

Weninger, F., Erdogan, H., Watanabe, S., Vincent, E., Le Roux, J., Hershey, J. R., and Schuller, B. (2015). Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 91–99. Springer.

Williams, R. J. and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280.

Wüthrich, M. V. (2018). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, pages 1–16.

Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons, Chichester.

Yin, Z. and Shen, Y. (2018). On the dimensionality of word embedding. *arXiv preprint arXiv:1812.04224*.

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*, 31(7):1235–1270.

Zhao, X. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2):290–299.

Zhao, X. B., Zhou, X., and Wang, J. L. (2009). Semiparametric model for prediction of individual claim loss reserving. *Insurance: Mathematics and Economics*, 45(1):1–8.

Chapter 2

On sums of two counter-monotonic risks

Résumé

Dans la gestion des risques, les exigences de fonds propres sont le plus souvent fondées sur les mesures de risque de l'agrégation des risques individuels aléatoires. La structure de dépendance entre ces variables aléatoires a un fort impact sur la distribution de la perte globale. Il existe un nombre important de travaux de recherche portant sur la somme des risques comonotones comparer à peu de travaux qui étudients la somme des risques antimonotones. Un résultat crucial pour les risques comonotones stipule que la valeur des mesures de risque Value-at-risk et Tail Value-at-risk de la perte globale correspondent respectivement à la somme de Value-at-risk et Tail Value-at-risk des risques individuels. Notre objectif principal est de dériver des résultats aussi simples pour la somme des risques antimonotones. Pour ce faire, nous examinons séparément différents cas de distributions strictement continues bivariées pour lesquelles nous obtenons des expressions analytiques (fermées) pour la Value-at-risk et la Tail Value-at-risk de la somme de d'une paire de risques antimonotones. Les expressions de la mesure sous-additive Tail Value-at Risk nous permettent de quantifier le bénéfice de diversification maximal. De plus, nos résultats nous permettent d'analyser également la queue de la distribution de la somme d'une paire de variables antimonotones identiquement distribuées de manière sous-exponentielle.

Abstract

In risk management, capital requirements are most often based on risk measurements of the aggregation of individual risks treated as random variables. The dependence structure between such random variables has a strong impact on the behavior of the aggregate loss. One finds an extensive literature on the study of the sum of comonotonic risks but less, in comparison, has been done regarding the sum of counter-monotonic risks. A crucial result for comonotonic risks is that the Value-at-risk and the Tail Value-at-risk of their sum correspond respectively to the sum of the Value-at-risk and Tail Value-at-risk of the individual risks. In this paper, our main objective is to derive such simple results for the sum of counter-monotonic risks. To do so, we examine separately different contexts in the class of bivariate strictly continuous distributions for which we obtain closed-form expressions for the Value-at-risk and Tail Value-at-risk of the sum of two counter-monotonic risks. The expressions for the subadditive Tail Value-at risk allow us to quantify the maximal diversification benefit. Also, our findings allow us to analyze the tail of the distribution of the sum of two identically subexponentially distributed counter-monotonic random variables.

Keywords: Counter-monotonicity, Extreme negative dependence, Risk measures, Diversification Benefit, Subexponential distributions.

2.1 Introduction

In recent years, it has become crucial for insurers and financial institutions to develop efficient risk management tools to appropriately assess the total capital requirement associated with a portfolio. In general, the latter represents the sum of $n \in \mathbb{N}^+$ risks with marginal losses modeled by the random variables (rvs) $X_1, \ldots, X_n$ and respective univariate cumulative distribution functions (cdfs) $F_1, \ldots, F_n$. If the joint distribution of losses is not defined, an investigation of its extreme distributions allows to derive bounds on various risk management measures to evaluate the total capital.

Given the fixed univariate marginal cdfs $F_1, \ldots, F_n$, the set of all possible n-dimensional joint cdfs $F_{X_1,\ldots,X_n}$, is the *Fréchet space*, denoted $\Gamma(F_1, \ldots, F_n)$. It is well known (see, e.g., Denuit et al. (2006)) that $F_{X_1,\ldots,X_n} \in \Gamma(F_1, \ldots, F_n)$ is bounded by two n-dimensional functions as follows:

$$W_n(x_1, \ldots, x_n) \leq F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \leq M_n(x_1, \ldots, x_n), \tag{2.1}$$

for $(x_1, \ldots, x_n) \in \mathbb{R}^n$, where $W_n(x_1, \ldots, x_n) = \max\{\sum_{i=1}^n F_i(x_i) - n + 1; 0\}$ is referred to as the Fréchet lower bound, while $M_n(x_1, \ldots, x_n) = \min\{F_1(x_1), \ldots, F_n(x_n)\}$ corresponds to the Fréchet upper bound. The latter corresponds to the cdf of a vector of rvs whose dependence structure is the extreme positive dependence known as comonotonicity, which has been investigated in depth in the literature (see, e.g., Dhaene et al. (2000), Dhaene et al. (2002) and Dhaene et al. (2006)). Furthermore, W_2 is the cdf of a pair of rvs whose dependence structure corresponds to the extreme negative dependence known as counter-monotonicity (see, e.g., Cheung and Lo (2013a)). Much less attention has been devoted however to this bivariate negative dependence concept. In a financial context, counter-monotonicity may be useful to assess the merge of dependent assets. For example, Cheung et al. (2014) investigated under what conditions, the sum of counter-monotonic risks will decrease the overall risk. The authors proposed a necessary and sufficient condition in terms of the introduced concept of risk reducer according to convex order. For $n \geq 3$, W_n is not in general a joint cdf (see counterexamples provided by Tchen (1980) and Example 3.1 of Joe (1997)).

Several notions of extreme negative multi-dimensional dependence were proposed by researchers. In an actuarial context, Dhaene and Denuit (1999) studied the concept of mutual exclusivity, defined only under quite restrictive assumptions and leading to the safest portfolio with respect to the stop-loss order. Properties and characterization of mutually exclusive rvs are presented and generalized in Cheung and Lo (2014). Notably, they show in Theorem 4.1 that a random vector is mutually exclusive if and only if it is pairwise counter-monotonic. The link between mutual exclusivity and complete mixability, introduced in Wang and Wang (2011), is also illustrated in Cheung and Lo (2014). We refer to Puccetti et al. (2013) and Wang et al. (2013) for several properties and examples. The latter also proposed the notion of joint mixability as a generalization of complete mixability. Wang and Wang (2016) developed

in depth the theory of joint mixability, and provided necessary and sufficient conditions for the joint mixability of several distributions. In addition, Puccetti et al. (2015) introduced Σ-counter-monotonic random vectors as a notion supported in any Fréchet class and that includes counter-monotonicity, pairwise counter-monotonicity and joint mixability as particular cases. We note that Puccetti et al. (2015) presented a detailed review of the concept of comonotonocity and various concepts of extremal negative dependence through the lens of rearrangement functions and in a context of optimization problems. A multidimensional extension of counter-monotonocity named d-counter-monotonicity (d-CM), for $d > 2$, was proposed in Lee and Ahn (2014). Under a variance minimization context, Lee et al. (2017) discussed the d-CM concept as a generalization of the joint mixability. Also, the authors showed that d-CM copulas constitute a minimal class of copulas.

Referring to the literature, some researchers were interested in finding (sharp) bounds on risk measures associated with several multivariate extreme negative dependence concepts. In Wang et al. (2013), a general lower bound for the cdf of the aggregate loss S is provided and a necessary and sufficient condition for this bound to be reached is given in terms of joint mixability. When all rvs are identically distributed and have a monotone or tail-monotone density, the authors found the expression of the lower cdf of S and then deduced the worst Value-at-Risk (VaR). Cheung and Lo (2013b) found expressions of lower bounds on three convex functions (convex expectation, Tail Value-at-Risk (TVaR) and Haezendonck-Goovaerts) of a sum of rvs with arbitrary distributions. For the convex expectation and TVaR, the lower bounds are reached if and only if the rvs are mutually exclusive.

In this paper, we turn our attention to the class of bivariate strictly continuous distributions. For a portfolio with two risks (X_1, X_2) (e.g., two lines of business or two financial positions), we define the aggregate loss $S = X_1 + X_2$. It is very interesting to be able to evaluate bounds on the risk measure TVaR of S when its distribution is partially or completely unknown. The upper bound on the TVaR is reached in the case of a comonotonic dependence structure, with a closed-form expression well known in the literature. However, the expression of the lower bound is reached under a counter-monotonic dependence structure. Our goal is to provide contributions on the latter structure as the safest bivariate dependence. In specific contexts, we investigate closed-form expressions for the VaR and TVaR, which also allows us to respectively deduce expressions for the cdfs and stop-loss premiums. Note that the obtained expressions can easily be interpreted and also reflect the risk compensation between the counter-monotonic rvs. Having closed-form expressions for bounds on risk measures is very useful for actuaries and risk managers to derive closed-form expressions for the smallest and largest measures of diversification.

The paper is structured as follows. In Section 2.2, we recall several important definitions including a useful representation of the aggregate loss in terms of a function, referred to as the aggregate loss function, of a standard uniform random variable. In Section 2.3, counter-

monotonic risks with a symmetric and absolutely continuous distribution and their aggregation are discussed. The aggregate loss of two identically distributed counter-monotonic risks is studied in Section 2.4 for cases where the aggregate loss function has either a concave or convex shape. We use these results to also show that asymptotically, the tail behavior of the distribution of the sum of two counter-mononotonic risks, following a common subexponential distribution, is similar to the tail behavior of the distribution of the aggregate loss of two independent risks. In Section 2.5, we examine the more general case of the sum of non-identically distributed counter-monotonic risks in which the aggregate loss function has either a concave or convex shape. Finally, in Section 2.6, counter-monotonic risks defined in terms of two exponential functions of a symmetric rv are discussed and specific examples are treated.

2.2 Definitions

2.2.1 Extremal elements on bivariate Fréchet Class

Let $\underline{X} = (X_1, X_2)$ be a bivariate random vector, with $F_{X_1,X_2} \in \Gamma(F_1, F_2)$ and

$$W_2(x_1, x_2) \leq F_{X_1,X_2}(x_1, x_2) \leq M_2(x_1, x_2), \qquad x_1, x_2 \in \mathbb{R}, \tag{2.2}$$

where $M_2 \in \Gamma(F_1, F_2)$ is the bivariate cdf of a pair of comonotonic rvs denoted by $\left(X_1^+, X_2^+\right)$, with $F_{X_1^+,X_2^+}(x_1, x_2) = M_2(x_1, x_2)$, $x_1, x_2 \in \mathbb{R}$, while $W_2 \in \Gamma(F_1, F_2)$ is the bivariate cdf of a pair of counter-monotonic rvs denoted by $\left(X_1^-, X_2^-\right)$, with $F_{X_1^-,X_2^-}(x_1, x_2) = W_2(x_1, x_2)$, $x_1, x_2 \in \mathbb{R}$.

Pairs of rvs $\underline{X} = (X_1, X_2)$ and $\underline{X}' = (X_1', X_2')$, with $F_{X_1,X_2}, F_{X_1',X_2'} \in \Gamma(F_1, F_2)$, can be compared under the (bivariate) concordance order, written $\underline{X} \preceq_c \underline{X}'$, see, e.g., Section 2.2.1 in Joe (1997). Note that the concordance order coincides with the correlation order in Denuit et al. (2006). According to Property 6.2.8 of the latter, $F_{X_1^-,X_2^-}$ and $F_{X_1^+,X_2^+}$ correspond to the extremal elements of $\Gamma(F_1, F_2)$ under the concordance order, meaning that comonotonicity is the bivariate strongest positive dependence, whereas counter-monotonicity is the bivariate extreme negative dependence. Then from inequalities (2.2), it follows that

$$\left(X_1^-, X_2^-\right) \preceq_c (X_1, X_2) \preceq_c \left(X_1^+, X_2^+\right) \tag{2.3}$$

is satisfied for any pair of rvs (X_1, X_2) with $F_{X_1,X_2} \in \Gamma(F_1, F_2)$. Moreover, assuming $E[X_i^2] < \infty$, $i = 1, 2$, Property 5.2.8 of Denuit et al. (2006) states that

$$\mathrm{Cov}\left(X_1^-, X_2^-\right) \leq \mathrm{Cov}(X_1, X_2) \leq \mathrm{Cov}\left(X_1^+, X_2^+\right). \tag{2.4}$$

Let $S = X_1 + X_2$ be the aggregate loss rv for the random pair (X_1, X_2). Then, $S^- = X_1^- + X_2^-$ and $S^+ = X_1^+ + X_2^+$ denote, respectively, the aggregate loss rvs for the counter-monotonic pair and the comonotonic pair. Based on (2.4), the aggregate loss S^+ has the largest variability, while S^- has the smallest one. In addition, we can use convex and increasing convex orders

(see, e.g., Müller and Stoyan (2002), Denuit et al. (2006) and Shaked and Shanthikumar (2007)), respectively written as $\preceq_{cx}$ and $\preceq_{icx}$, to compare S, S^- and S^+. Extremal bounds can then be obtained for risk measures such as the TVaR and risk related quantities such as the stop-loss premium.

Theorems 2 and 3 of Dhaene and Goovaerts (1996) establish the link between the (bivariate) concordance order and the (univariate) convex order. Given (2.3), we have

$$S^- \preceq_{cx} S \preceq_{cx} S^+, \tag{2.5}$$

which implies that $F_{X_1^-,X_2^-}$ and $F_{X_1^+,X_2^+}$ are the safest and riskiest elements of $\Gamma(F_1,F_2)$, respectively. This can also be stated as follows. Let ϕ be a convex function such that $E[\phi(S)] < \infty$ for any pair (X_1,X_2) with $F_{X_1,X_2} \in \Gamma(F_1,F_2)$. Then, (2.3) and (2.5) imply that $F_{X_1^-,X_2^-}$ and $F_{X_1^+,X_2^+}$ are the solutions to the following optimization problems:

$$E[\phi(S^-)] = \inf_{F_{X_1,X_2}\in\Gamma(F_1,F_2)} \{E[\phi(S)]\} \tag{2.6}$$

and

$$E[\phi(S^+)] = \sup_{F_{X_1,X_2}\in\Gamma(F_1,F_2)} \{E[\phi(S)]\}. \tag{2.7}$$

For example, with $\phi(s) = \max(s-x;0)$, for $x \in \mathbb{R}$, $\pi_S(x) = E[\phi(S)]$ defines the stop-loss premium of the rv S. Then, from (2.6) and (2.7), we have

$$\pi_{S^-}(x) \leq \pi_S(x) \leq \pi_{S^+}(x), \qquad x \in \mathbb{R}. \tag{2.8}$$

More generally, the expectation of a supermodular function of (X_1,X_2) is maximized (minimized) under a comonotonic (counter-monotonic) dependence structure (see, e.g., Section 3.1 in Puccetti et al. (2015)). For an earlier discussion on the subject in an actuarial context, we refer the reader to section 8.3.1 of Müller and Stoyan (2002). Inequalities (2.8) may be particularly useful in a financial market setting to evaluate the upper and lower bounds for the price of basket options, whose payoff is a stop-loss function. Optimal replicating strategies can then be deduced. Hobson et al. (2005a) and Hobson et al. (2005b) derived distribution-free static arbitrage bounds in the case of basket options, using the concept of comonotonocity and counter-monotonocity. In the same context, Laurence and Wang (2005) derived sharp bounds in closed-forms under some assumptions.

It is crucial for actuaries and risk managers to identify the safest and riskiest dependence structures leading to the smallest and largest values of specific convex risk measures of the aggregate loss rv S. In the following, we recall the definition of the risk measures TVaR and Lower Tail Value-at-Risk (LTVaR).

Définition 10 *Let X be a rv defined on $\mathbb{R}$ with cdf F and quantile function*

$$F^{-1}(u) = \inf\{x, F(x) \geq u\}, u \in (0,1).$$

Assuming $E[X] < \infty$*, the TVaR and LTVaR of* X *are expressed respectively as*

$$TVaR_{\kappa}(X) = \frac{1}{1-\kappa}\int_{\kappa}^{1} VaR_{u}(X)\,\mathrm{d}u, \tag{2.9}$$

with $VaR_{\kappa}(X) = F^{-1}(\kappa)$ *and,*

$$LTVaR_{\kappa}(X) = \frac{1}{\kappa}\int_{0}^{\kappa} VaR_{u}(X)\,\mathrm{d}u,$$

for $\kappa \in (0,1)$.

The $TVaR$ represents the expected loss, given that the loss is in the worst $100(1-\kappa)\%$ cases (the right tail), while the $LTVaR$ covers values in the left tail (see, e.g., Acerbi and Tasche (2002), Inui and Kijima (2005) and McNeil et al. (2015) for properties and applications of these risk measures). Further, we note that a closed-form expression for $TVaR$ allows to obtain the stop-loss premium using

$$\pi_{X}(d) = \overline{F}_{X}(d)\,TVaR_{F_{X}(d)}(X) - d\overline{F}_{X}(d), \qquad d \in \mathbb{R}. \tag{2.10}$$

Proposition 3.4.8 of Denuit et al. (2006) states the relationship between the TVaR and the increasing convex order which proves to be very useful: for two rvs X and Y, with finite expectations, we have that

$$X \preceq_{icx} Y \Leftrightarrow TVaR_{\kappa}(X) \leq TVaR_{\kappa}(Y), \qquad \kappa \in (0,1). \tag{2.11}$$

Moreover, according to Müller and Stoyan (2002) and Shaked and Shanthikumar (2007), for two rvs with the same expectations, $\preceq_{icx}$ becomes equivalent to $\preceq_{cx}$. Combining (2.3), (2.5), and (2.11), it follows that $F_{X_1^-,X_2^-}$ and $F_{X_1^+,X_2^+}$ are the solutions to the following optimization problems: for all $\kappa \in (0,1)$, we have

$$TVaR_{\kappa}(S^-) = \inf_{F_{X_1,X_2}\in\Gamma(F_1,F_2)}\{TVaR_{\kappa}(S)\} \tag{2.12}$$

and

$$TVaR_{\kappa}(S^+) = \sup_{F_{X_1,X_2}\in\Gamma(F_1,F_2)}\{TVaR_{\kappa}(S)\}. \tag{2.13}$$

The results in (2.12) and (2.13) mean that, whatever the marginals, the inequalities

$$TVaR_{\kappa}\left(S^-\right) \leq TVaR_{\kappa}(S) \leq TVaR_{\kappa}\left(S^+\right) \tag{2.14}$$

hold for all $\kappa \in (0,1)$.

Also, given that $LTVaR_{\kappa}(S) = \frac{E[S]}{\kappa} - TVaR_{1-\kappa}(S)$, the inequalities

$$LTVaR_{\kappa}\left(S^+\right) \leq LTVaR_{\kappa}(S) \leq LTVaR_{\kappa}\left(S^-\right)$$

are valid for all $\kappa \in (0,1)$. We shall see that, the LTVaR risk measure is included in the closed-form expressions for $TVaR_{\kappa}(S^-)$ derived in Sections 2.3, 2.4 and 2.5.

2.2.2 Extremal elements on Fréchet Class and measures of diversification

The diversification gain can be defined as the capital saved when combining several risks together rather than treating them separately. We briefly recall three measures of diversification that have been introduced in the literature. For a pair of risks (X_1, X_2), with aggregate loss S, the diversification benefit at a confidence level $\kappa \in (0,1)$ can be measured by

$$DB_\kappa^\rho(S) = \rho_\kappa(X_1) + \rho_\kappa(X_2) - \rho_\kappa(S), \tag{2.15}$$

with ρ a risk measure. An important desirable property for ρ is being sub-additive, which implies a positive DB_κ^ρ when merging risks X_1 and X_2 together. If $\rho = TVaR$, then from (2.14) and (2.15) we deduce that

$$0 = DB_\kappa^{TVaR}(S^+) \leq DB_\kappa^{TVaR}(S) \leq DB_\kappa^{TVaR}(S^-), \qquad \kappa \in (0,1). \tag{2.16}$$

Assuming ρ is a sub-additive risk measure introducing a positive margin, Bürgi et al. (2008) and Dacorogna et al. (2016) propose an alternative way to quantify the diversification benefit. Given $\kappa \in (0,1)$, they define a diversification benefit index by

$$DBI_\kappa^\rho(S) = 1 - \frac{\rho_\kappa(S) - E[S]}{\sum_{i=1}^2 \rho_\kappa(X_i) - \sum_{i=1}^2 E[X_i]}. \tag{2.17}$$

For the sub-additive risk measure $TVaR$, it follows from (2.14) and (2.17) that $DBI_\kappa^{TVaR}(S) \in [0,1]$, for all $\kappa \in (0,1)$, with

$$0 = DBI_\kappa^{TVaR}(S^+) \leq DBI_\kappa^{TVaR}(S) \leq DBI_\kappa^{TVaR}(S^-) \leq 1. \tag{2.18}$$

Referring to Embrechts et al. (2009a), a diversification concentration index can be expressed as

$$CI_\kappa^\rho(S) = \frac{\rho_\kappa(S)}{\rho_\kappa(X_1) + \rho_\kappa(X_2)}, \qquad \kappa \in (0,1). \tag{2.19}$$

Then, assuming $\rho = TVaR$ in (2.19), it implies from (2.14) that the concentration index belongs to the following interval:

$$CI_\kappa^{TVaR}(S^-) \leq CI_\kappa^{TVaR}(S) \leq CI_\kappa^{TVaR}(S^+) = 1, \qquad \kappa \in (0,1). \tag{2.20}$$

Note that, in a financial context, we may also be interested to measure the degree of co-movement between stock prices as an indicator for potential systemic risk. Dhaene et al. (2012) introduced the Herd Behavior Index (HIX), which can be considered as the ratio of the variance of the real market situation $Var(S)$ to that under the comonotonic assumption $Var(S^+)$. The latter corresponds to the perfect herd behavior.

Given a pair of dependent risks, if the distribution of S is not known, it would be useful to bound the three measures of diversification using inequalities (2.16), (2.18) and (2.20).

Thus, it is important to be able to efficiently compute $TVaR_\kappa(S^-)$ and $TVaR_\kappa(S^+)$. The explicit expression of $TVaR_\kappa(S^+)$ is well known in the literature. The purpose of this article is to provide closed-form expressions of $TVaR_\kappa(S^-)$ and consequently deduce the associated measures of diversification $DB_\kappa^{TVaR}(S^-)$, $DBI_\kappa^{TVaR}(S^-)$ and $CI_\kappa^{TVaR}(S^-)$.

2.2.3 Useful representation

By definition, a stochastic representation of the pair of comonotonic and counter-monotonic rvs is given by $(X_1^+, X_2^+) \overset{d}{=} \left(F_1^{-1}(U), F_2^{-1}(U)\right)$ and $(X_1^-, X_2^-) \overset{d}{=} \left(F_1^{-1}(U), F_2^{-1}(1-U)\right)$, with $U \sim \text{Unif}(0,1)$. Based on the latter, we propose in what follows a useful representation of the aggregate loss that will be used throughout the paper in order to investigate its behavior.

Let us define the aggregate loss function $\varphi^+(u) = F_1^{-1}(u) + F_2^{-1}(u)$, for $u \in (0,1)$. Given $U \sim \text{Unif}(0,1)$, we have that $S^+ := \varphi^+(U) = F_1^{-1}(U) + F_2^{-1}(U)$. Since $\varphi^+(u)$ is a non-decreasing function of $u \in (0,1)$, we can deduce that

$$VaR_\kappa\left(S^+\right) = VaR_\kappa\left(X_1^+\right) + VaR_\kappa\left(X_2^+\right), \qquad \kappa \in (0,1), \tag{2.21}$$

which implies

$$TVaR_\kappa\left(S^+\right) = TVaR_\kappa\left(X_1^+\right) + TVaR_\kappa\left(X_2^+\right), \qquad \kappa \in (0,1). \tag{2.22}$$

These results are well known in the literature and valid for the sum of $n \geq 2$ comonotonic risks (see, e.g., Proposition 7.20 in McNeil et al. (2015)). In certain cases, we can invert $VaR_\kappa(S^+)$ to find closed-form expression of F_{S^+} or otherwise invert $VaR_\kappa(S^+)$ numerically. Such properties do not hold under counter-monotonicity. Thus, we proceed to the investigation of the behavior of the aggregate loss function under such a negative dependence structure.

Let us define the aggregate loss function $\varphi^-(u) = F_1^{-1}(u) + F_2^{-1}(1-u)$, for $u \in (0,1)$. We have that $S^- := \varphi^-(U) = F_1^{-1}(U) + F_2^{-1}(1-U)$. It is very important to note that the pattern of φ^- is influenced by the shape of the univariate cdfs F_1 and F_2 unlike φ^+, which is non-decreasing for any choice of univariate marginal distributions. It implies that $\varphi^-(u)$ is not always a monotone function of u, as it is illustrated in Figure 2.1.

In the next sections, we examine three cases: symmetric and absolutely continuous distributions, identical unimodal distributions with support on $\mathbb{R}^+$ and unimodal distributions with support on $\mathbb{R}^+$.

2.3 Symmetric absolutely continuous distributions

We recall below the definition of the dispersive order, which will be useful to compare univariate rvs by variability.

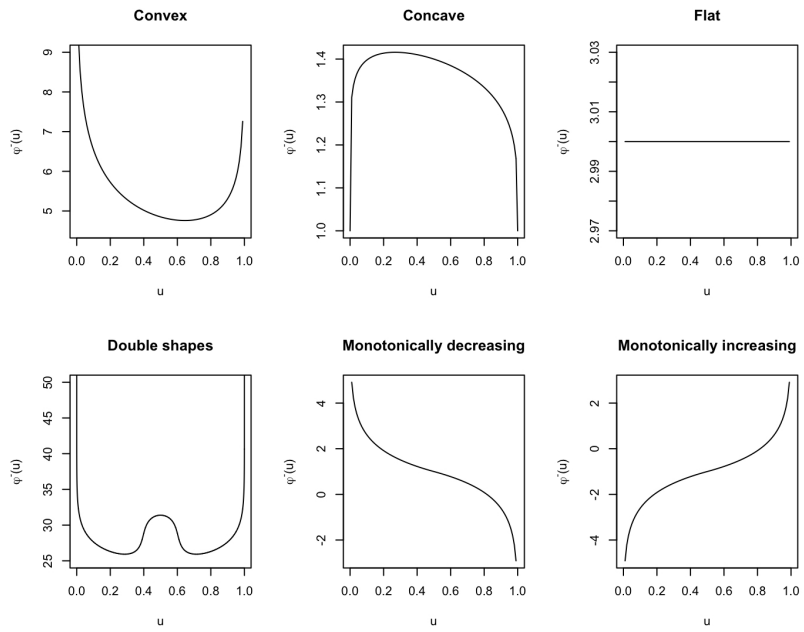


Figure 2.1 – Possible patterns of φ^- given different choices of marginals F_1 and F_2

Définition 11 *Let X and Y be two rvs, with cdfs F and G, respectively. If $F^{-1}(u)-F^{-1}(v) \leq G^{-1}(u) - G^{-1}(v)$, for $0 < v \leq u < 1$, then X is said to be smaller than Y according to the dispersive order which we denote $X \preceq_{disp} Y$.*

The dispersive order is treated in details, e.g., in Chapter 3 of Shaked and Shanthikumar (2007). Also, Jeon et al. (2006) provides applications and examples of stochastic orders. Notably, the dispersive order satisfies the following properties:

1. $X \preceq_{disp} Y \Leftrightarrow X + c \preceq_{disp} Y$, for any $c \in \mathbb{R}$.
2. $X \preceq_{disp} cX$, for $c \geq 1$.
3. $X \preceq_{disp} Y \Leftrightarrow F_1^{-1}(u) - F_2^{-1}(u)$ is increasing in $u \in (0,1)$.

Proposition 7 *Define a pair of rvs (X_1^-, X_2^-) with $X_i^- = a_i + b \times Z_i^-$, for $i = 1, 2$, with $a_1, a_2 \in \mathbb{R}$ and $b > 0$. Assume that*

(i) Z_1^- and Z_2^- are two counter-monotonic rvs with symmetric and absolutely continuous cdf F_1 and F_2;

(ii) $F_i(0) = 0.5$ and $F_i^{-1}(u) = -F_i^{-1}(1-u)$, for $u \in (0,1)$ and $i = 1, 2$;

(iii) $Z_2^- \preceq_{disp} Z_1^-$.

Then, for $\kappa \in (0,1)$, we have

$$VaR_\kappa(S^-) = VaR_\kappa(X_1^-) + VaR_{1-\kappa}(X_2^-) \tag{2.23}$$

and

$$TVaR_\kappa(S^-) = TVaR_\kappa(X_1^-) + LTVaR_{1-\kappa}(X_2^-). \tag{2.24}$$

Proof. Given the stochastic representation of X_i^-, for $i = 1, 2$, the aggregate loss function is

$$\varphi^-(u) = a_1 + b \times F_1^{-1}(u) + a_2 + b \times F_2^{-1}(1-u), \tag{2.25}$$

for $u \in (0, 1)$. With the assumption (ii) in Proposition 7, (2.25) becomes

$$\varphi^-(u) = a_1 + a_2 + b \times \left(F_1^{-1}(u) - F_2^{-1}(u)\right), \tag{2.26}$$

for $u \in (0, 1)$. Combining property 3 of the dispersive order and assumption (iii) of Proposition 7, it follows that $\varphi^-(u)$ in (2.26), and consequently, also in (2.25), is an increasing function of $u \in (0, 1)$. By Property 1.5.16(i) of the quantile function in Denuit et al. (2006), it follows that

$$\begin{aligned} VaR_\kappa\left(S^-\right) &= \varphi^-(\kappa) \\ &= a_1 + b \times F_1^{-1}(\kappa) + a_2 + b \times F_2^{-1}(1-\kappa), \end{aligned}$$

leading to (2.23). Using (2.9) with (2.23), the relation in (2.24) follows. ■

Relations given in (2.23) and (2.24) have intuitive interpretations. The expression for $VaR_\kappa\left(S^-\right)$ given in (2.23) is similar to $VaR_\kappa(S^+)$, given in (2.21), for the sum of two comonotonic rvs. In addition, (2.24) is expressed in terms of two risk measures, namely LTVaR and TVaR of the marginal distribution. We observe that (2.23) and (2.24) reflect the risk compensation between the counter-monotonic rvs X_1^- and X_2^- through the opposite variation of the components of each of them. When κ tends to 1, "small" values of the VaR of X_2^- compensate for "large" values of the VaR of X_1^-. As we shall see in Sections 2.4 and 2.5, we aim to find similar relations when the support of the marginal cdfs is $\mathbb{R}^+$.

Example 1 *We provide below a list of examples of pairs of rvs* (Z_1^-, Z_2^-)*, where* $Z_2^- \preceq_{disp} Z_1^-$*:*

1. *Example 3 of Jeon et al. (2006) : If* $Z_1^- \sim Student(\nu_1)$ *and* $Z_2^- \sim Student(\nu_2)$*, with degrees of freedom* $0 < \nu_1 < \nu_2$*, then* $Z_2 \preceq_{disp} Z_1$*.*
2. $Z_1 \sim Student(\nu)$*, where* $\nu > 0$*, and* $Z_2^- \sim Norm(0, 1)$*;*
3. $Z_1^- \sim Laplace(0, 2^{0.5})$ *and* $Z_2^- \sim Norm(0, 1)$*;*
4. $Z_1^- \sim Logistic(0, 1)$ *and* $Z_2^- \sim Norm(0, 1)$*;*
5. $Z_1^- = c_1 F^{-1}(U)$ *and* $Z_2^- = c_2 F^{-1}(1-U)$*, with* F *satisfying the assumption in Proposition 7.(ii),* $c_1 \geq c_2 > 0$ *and* $U \sim \text{Unif}(0, 1)$*.*

In the following example, we examine an interesting illustration of Proposition 7.

Example 2 *Let $Z_1^- \sim$ Logistic(0,1) and $Z_2^- \sim$ Laplace$(0,1)$ be two counter-monotonic rvs. Using Proposition 7, the VaR and TVaR of the aggregate loss rv $S^- = Z_1^- + Z_2^-$ are respectively given by*

$$VaR_\kappa(S^-) = \begin{cases} -\log(2(1-\kappa)), & 0 < \kappa < 0.5 \\ \log(2\kappa), & 0.5 \leq \kappa < 1 \end{cases} \tag{2.27}$$

and

$$TVaR_\kappa(S^-) = \begin{cases} \frac{1}{1-\kappa}\left(\log\left(2^\kappa(1-\kappa)^{\kappa-1}\right) - \kappa\right), & 0 < \kappa < 0.5 \\ \frac{1}{1-\kappa}\left(-\log(2^{\kappa-1}\kappa^\kappa) - 1 + \kappa\right), & 0.5 \leq \kappa < 1. \end{cases}$$

From (2.27), note that the rv S^- has a bounded support $(\log(0.5), \log(2))$, with cdf

$$F_{S^-}(s) = \begin{cases} 1 - 0.5\mathrm{e}^{-s}, & \log(0.5) < s < 0 \\ 0.5\mathrm{e}^{s}, & 0 \leq s < \log(2). \end{cases}$$

In the following, we study two important patterns for the shape of the aggregate loss function φ^-, namely convex and concave shapes, to develop closed-form expressions for various risk measures. We first assume that the risks are identically distributed and then we move on to the non-identically distributed case.

2.4 Identically distributed counter-monotonic rvs

In this section, we make the assumption that $X_1^- \sim X_2^- \sim X$, where X follows a continuous and strictly positive probability distribution function (pdf), denoted f, and cdf F. Hence, the aggregate loss function φ^- can be written as

$$\varphi^-(u) = F^{-1}(u) + F^{-1}(1-u), \qquad u \in (0,1). \tag{2.28}$$

To investigate (2.28) further, we develop its first and second derivatives, which are given by

$$\varphi^{-(1)}(u) = \frac{\partial \varphi^-(u)}{\partial u} = \frac{1}{F^{(1)}\left(F^{-1}(u)\right)} - \frac{1}{F^{(1)}\left(F^{-1}(1-u)\right)},$$

$$\varphi^{-(2)}(u) = \frac{\partial^2 \varphi^-(u)}{\partial u^2} = -\left(\frac{F^{(2)}\left(F^{-1}(u)\right)}{\left(F^{(1)}\left(F^{-1}(u)\right)\right)^3} + \frac{F^{(2)}\left(F^{-1}(1-u)\right)}{\left(F^{(1)}\left(F^{-1}(1-u)\right)\right)^3}\right) \tag{2.29}$$

for $u \in (0,1)$. Under the assumption that φ^- is either concave or convex, from (2.28), φ^- is symmetric and admits either a minimum or a maximum value, denoted s_0, at u_0 satisfying $\varphi^{-(1)}(u_0) = 0$, which is equivalent to $F^{(1)}\left(F^{-1}(1-u_0)\right) = F^{(1)}\left(F^{-1}(u_0)\right)$. Thus, we deduce that $u_0 = \frac{1}{2}$ and $s_0 = \varphi^-\left(\frac{1}{2}\right) = 2 \times F^{-1}\left(\frac{1}{2}\right)$. Then, if s_0 is a minimum, the values of S^- belong to the set $\left(\varphi^-\left(\frac{1}{2}\right), \sup_{u\in(0,1)} \varphi^-(u)\right)$, otherwise $S^- \in \left(\inf_{u\in(0,1)} \varphi^-(u), \varphi^-\left(\frac{1}{2}\right)\right)$. This is illustrated in Figure 2.2.

2.4.1 Convex aggregate loss function φ^-

For two identically distributed counter-monotonic risks, such that φ^- is convex and symmetric, we are able to obtain closed-form expressions for $VaR_\kappa(S^-)$ and $TVaR_\kappa(S^-)$, for $\kappa \in (0,1)$, according to the next proposition.

Proposition 8 *Assume $X_1^- \sim X_2^- \sim X$, with a convex and symmetric aggregate loss function φ^-. Then, for $\kappa \in (0,1)$, we have*

$$VaR_\kappa\left(S^-\right) = VaR_{\frac{1-\kappa}{2}}(X) + VaR_{\frac{1+\kappa}{2}}(X) \tag{2.30}$$

and

$$TVaR_\kappa\left(S^-\right) = LTVaR_{\frac{1-\kappa}{2}}(X) + TVaR_{\frac{1+\kappa}{2}}(X). \tag{2.31}$$

Proof. Since φ^- is convex and symmetric, it follows that it admits a minimum value s_0 at $u_0 = \frac{1}{2}$. For $s > s_0$, the pair of solutions (u_1, u_2) to $\varphi^-(u) = s$ are equidistant from u_0, and written as $u_1 = \frac{1}{2} - \nu$ and $u_2 = \frac{1}{2} + \nu$, with $\nu \in \left(0, \frac{1}{2}\right)$. Therefore, we have

$$F_{S^-}(s) = \Pr(u_1 < U \leq u_2) = 2\nu,$$

for $\nu \in \left(0, \frac{1}{2}\right)$. Given $\kappa \in (0,1)$, solving $F_{S^-}(s) = \kappa$ implies that $\nu = \frac{\kappa}{2}$. Thus, we have

$$VaR_\kappa\left(S^-\right) = \varphi^-\left(\frac{1}{2} - \nu\right) = \varphi^-\left(\frac{1-\kappa}{2}\right) = VaR_{\frac{1-\kappa}{2}}(X) + VaR_{\frac{1+\kappa}{2}}(X). \tag{2.32}$$

Moreover, the expression of $TVaR_\kappa\left(S^-\right)$ can be obtained by replacing (2.32) in (2.9) as follows:

$$\begin{aligned} TVaR_\kappa\left(S^-\right) &= \frac{1}{1-\kappa}\int_\kappa^1 VaR_u\left(S^-\right) du \\ &= \frac{1}{\frac{1-\kappa}{2}}\int_0^{\frac{1-\kappa}{2}} VaR_u(X)du + \frac{1}{\frac{1-\kappa}{2}}\int_{\frac{1+\kappa}{2}}^1 VaR_u(X)du \\ &= LTVaR_{\frac{1-\kappa}{2}}(X) + TVaR_{\frac{1+\kappa}{2}}(X), \end{aligned}$$

for $\kappa \in (0,1)$. ■

The expression for $VaR_\kappa\left(S^-\right)$ given in (2.30) narrows down to the sum of two VaRs of the marginal rv with $VaR_\kappa(S^-) \overset{\kappa\to 0}{\sim} VaR_{\frac{1-\kappa}{2}}(X)$ and $VaR_\kappa(S^-) \overset{\kappa\to 1}{\sim} VaR_{\frac{1+\kappa}{2}}(X)$. In addition, (2.31) is expressed in terms of two risk measures, namely LTVaR and TVaR of the marginal distribution. As κ tends to 1, (2.31) converges to $TVaR_{\frac{1+\kappa}{2}}(X)$. As in (2.23) and (2.24), (2.30) and (2.31) reflect the risk compensation between the counter-monotonic rvs X_1^- and X_2^- through the opposite variation of the components of each of them.

2.4.2 Concave aggregate loss function φ^-

In this section, the aggregate loss function φ^- is assumed to have a smooth concave shape over $(0,1)$, which can only be achieved with bounded rvs. Such an assumption will give us the ability to develop elegant formulas for the VaR and TVaR.

Proposition 9 *Let $X_1^- \sim X_2^- \sim X$, with a concave and symmetric aggregate loss rv S^-. We have*

$$VaR_\kappa\left(S^-\right) = VaR_{\frac{\kappa}{2}}(X) + VaR_{1-\frac{\kappa}{2}}(X) \tag{2.33}$$

and

$$TVaR_\kappa\left(S^-\right) = \frac{1}{1-\kappa}\left((2-\kappa)TVaR_{\frac{\kappa}{2}}\left(X\right) - \kappa TVaR_{1-\frac{\kappa}{2}}\left(X\right)\right), \tag{2.34}$$

for $\kappa \in (0,1)$.

Proof. Since the concave function φ^- is symmetric, it reaches a maximum value s_0 for $u_0 = \frac{1}{2}$. Moreover, the two possible solutions to $\varphi^-(u) = s$, for $s < s_0$, are equidistant from u_0, with $u_2 = 1 - u_1$. Given $\nu \in \left(0, \frac{1}{2}\right)$, we can write $u_1 = \frac{1}{2} - \nu$ and $u_2 = \frac{1}{2} + \nu$. Then, we obtain

$$F_{S^-}(s) = \Pr\left(U \leq u_1\right) + \Pr\left(U \geq u_2\right) = 1 - 2\nu, \qquad 0 < s < s_0.$$

Given $\kappa \in (0,1)$, solving $F_{S^-}(s) = \kappa$, we obtain $\nu = \frac{1-\kappa}{2}$. Then, $VaR_\kappa\left(S^-\right)$ can be expressed as

$$VaR_\kappa\left(S^-\right) = \varphi^-\left(\frac{1}{2} - \nu\right) = VaR_{\frac{\kappa}{2}}(X) + VaR_{1-\frac{\kappa}{2}}(X).$$

Thus, it follows that

$$\begin{aligned}
TVaR_\kappa\left(S^-\right) &= \frac{1}{1-\kappa}\int_\kappa^1 \left(VaR_{\frac{u}{2}}(X) + VaR_{1-\frac{u}{2}}(X)\right) du \\
&= \frac{1}{1-\kappa}\left(\int_{\frac{\kappa}{2}}^{\frac{1}{2}} 2VaR_u(X)du + \int_{1-\frac{\kappa}{2}}^{\frac{1}{2}} (-2)VaR_u(X)du\right) \\
&= \frac{1}{1-\kappa}\left(2\int_{\frac{\kappa}{2}}^1 VaR_u(X)du - 2\int_{1-\frac{\kappa}{2}}^1 VaR_u(X)du\right) \\
&= \frac{1}{1-\kappa}\left((2-\kappa)TVaR_{\frac{\kappa}{2}}\left(X\right) - \kappa TVaR_{1-\frac{\kappa}{2}}\left(X\right)\right),
\end{aligned}$$

for $\kappa \in (0,1)$. ■

As in the convex case, (2.33) is also expressed as the sum of two VaRs with opposite variation. Moreover, we note that (2.34) is given only in terms of the TVaR of the marginal distribution.

2.4.3 Sufficient conditions for convex or concave aggregate loss functions φ^-

To apply either Propositions 8 or 9, we need to look for a criterion leading to either convex or concave aggregate loss functions φ^-. Such a criterion may depend on the pattern of

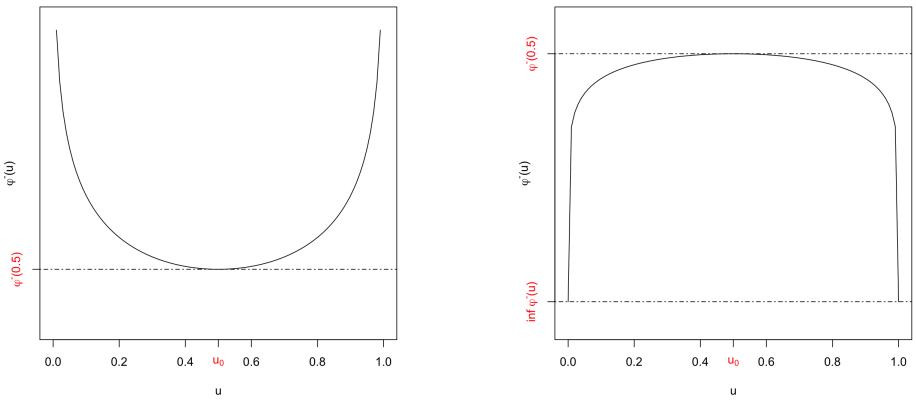


Figure 2.2 – Illustrative examples of the set of values of φ^- for both convex and concave shapes.

f. In the next proposition, we focus on distributions with a pdf that is strictly monotone (i.e., $\text{sign}\left(f^{(1)}(x)\right) = \text{sign}\left(f^{(1)}(x')\right), \forall\ (x, x') \in \mathbb{R}^{+2}$). The exponential, Pareto, gamma ($\alpha \in (0, 1]$), Weibull ($\tau \in (0, 1]$) and beta ($\alpha > 1, \beta = 1$) distributions, as well any finite or infinite mixture of them satisfy this property.

Proposition 10 *Let $X_1^- \sim X_2^- \sim X$, where X follows a continuous and strictly positive distribution.*

(i) If its pdf is strictly decreasing, then φ^- is convex.

(ii) If it has a strictly increasing pdf defined on a bounded support, then φ^- is concave.

Proof. We begin to prove the result in item (i). Since f is strictly positive and decreasing, we obtain $\text{sign}\left(F^{(1)}(x)\right) > 0$ and $f^{(1)}(x) = F^{(2)}(x) < 0$. Therefore, F is concave on its support, and $F^{(2)}\left(F^{-1}(u)\right) < 0$, for $u \in (0, 1)$. According to (2.29), we deduce that $\varphi^{-(2)}(u) > 0$, $u \in (0, 1)$. (ii) is proven similarly. ■

In the next example, we provide an application combining Propositions 8 and 10.(i).

Example 3 *For a pair of counter-monotonic risks (X_1^-, X_2^-), with $X_1^- \sim X_2^- \sim \text{Exp}(\beta)$, $\beta > 0$, using Proposition 8, the VaR and the TVaR of S^- are respectively given by*

$$VaR_\kappa\left(S^-\right) = -\frac{1}{\beta}\ln\left(\frac{1-\kappa^2}{4}\right) \tag{2.35}$$

and

$$TVaR_\kappa\left(S^-\right) = -\frac{1}{\beta}\left(-\ln\left(\frac{1+\kappa}{2}\right)\frac{1+\kappa}{1-\kappa} + \ln\left(\frac{1-\kappa}{2}\right) - 2\right), \tag{2.36}$$

for $\kappa \in (0,1)$. *By inverting (2.35), we obtain*

$$F_{S^-}(s) = \sqrt{1-4\mathrm{e}^{-\beta s}}, \qquad s \geq \frac{1}{\beta}\ln(4). \tag{2.37}$$

Combining (2.36), (2.37) and (2.10), the stop-loss premium of S^- *can be expressed as*

$$\begin{aligned}
&\pi_{S^-}(d) \\
&= \begin{cases}
\frac{2}{\beta} - d, & 0 \leq d < s_0 \\
\left(\frac{2}{\beta} - d\right) \times \left(1 - \sqrt{1-4\mathrm{e}^{-\beta d}}\right) & \\
-\frac{1}{\beta}\left(-\left(1+\sqrt{1-4\mathrm{e}^{-\beta d}}\right)\ln\left(\frac{1+\sqrt{1-4\mathrm{e}^{-\beta d}}}{2}\right)\right. & d \geq s_0. \\
\left. \quad +\left(1-\sqrt{1-4\mathrm{e}^{-\beta d}}\right)\ln\left(\frac{1-\sqrt{1-4\mathrm{e}^{-\beta d}}}{2}\right)\right), &
\end{cases}
\end{aligned} \tag{2.38}$$

with $s_0 = \frac{1}{\beta}\ln(4)$. *(2.38) can also be deduced using* $\pi_{S^-}(d) = \int_d^\infty \overline{F}_{S^-}(x)\,\mathrm{d}x$ *and Taylor expansion of* $\sqrt{1+x}$.

In Example 8.3.13 of Müller and Stoyan (2002), they discuss, in the context of Example 3, the impact on the value of the essential infimum s_0 and variance of S of an increase in the number n of aggregated dependent rvs. It is shown that s_0 increases as n increases, which translates into reducing the support of S. As for Dhaene and Goovaerts (1996), they give solely the expression of π_{S^+}.

In the next example, we give an application of Propositions 9 and 10.(ii).

Example 4 *Let* $X \sim \mathrm{Beta}(\alpha,\beta)$, *with* $\alpha > 1, \beta = 1$ *(i.e. strictly increasing pdf). For a pair of risks* (X_1^-, X_2^-), *with* $X_1^- \sim X_2^- \sim X$, *the TVaR of their aggregate loss* S^- *can be written as*

$$\begin{aligned}
TVaR_\kappa\left(S^-\right) &= \frac{2}{1-\kappa}\frac{\alpha}{\alpha+1}\left(\overline{B}\left(VaR_{\frac{\kappa}{2}}(X),\alpha+1,1\right) + \overline{B}\left(VaR_{1-\frac{\kappa}{2}}(X),\alpha+1,1\right)\right) \\
&= \frac{2}{1-\kappa}\frac{\alpha}{\alpha+1}\left(\left(1-\frac{\kappa}{2}\right)^{\frac{\alpha+1}{\alpha}} - \left(\frac{\kappa}{2}\right)^{\frac{\alpha+1}{\alpha}}\right),
\end{aligned}$$

for $\kappa \in (0,1)$, *with* $\overline{B}$ *denoting the survival function (sf) of a beta distribution and* $VaR_\kappa(X)$ *numerically-computed.*

In Section 2.6, we provide another sufficient condition on the distribution of X leading to a convex aggregate loss function φ^-.

2.4.4 Subexponential distributions and diversification benefit

Consider the sum S of two identically distributed rvs X_1 and X_2 following a common subexponential marginal distribution F and with a bivariate distribution defined by an Archimedean

copula C_θ with positive dependence parameter θ. It is shown in Proposition 2.2. of Embrechts et al. (2009b) (see also previous works by Alink et al. (2004) and Barbe et al. (2006)) that the sf of S is a function of the tail index of F, of the dependence parameter θ, and of the sf $\overline{F}$, as x tends to ∞. This asymptotic result lies between the result obtained when the rvs X_1 and X_2 are independent, and the result obtained when they are comonotonic. In this section, we seek similar results when X_1 and X_2 have a different type of negative dependence structure. Aiming to answer this question, we use Propositions 8 and 10.(i) to analyze the tail of the distribution of the sum of two identically distributed counter-monotonic rvs, more specifically when X_1^- and X_2^- follow a common subexponential distribution.

Définition 12 *A rv X follows a subexponential distribution if, for a random vector $\left(X_1^\perp, ..., X_n^\perp\right)$ of independent and identically distributed copies of X, for every $n \in \mathbb{N}$,*

$$\lim_{x\to\infty} \Pr(X_1^\perp + \cdots + X_n^\perp > x) \sim \lim_{x\to\infty} n\,\overline{F}(x) \sim \lim_{x\to\infty} \Pr(X_{max,n}^\perp > x), \tag{2.39}$$

where $X_{max,n}^\perp = \max(X_1^\perp, ..., X_n^\perp)$. See Definition 1.3.3 of Embrechts et al. (2013).

Secondly, we compare the tail behavior of the sum and the maximum of two counter-monotonic rvs.

Définition 13 *Let $X_1^- \sim X_2^- \sim X$ be two counter-monotonic rvs and define $X_{max,2}^- = \max(X_1^-, X_2^-)$. Then, the sf of $X_{max,2}^-$ is*

$$\begin{aligned}\Pr(X_{max,2}^- > x) &= \Pr(\max(F^{-1}(U), F^{-1}(1-U)) > x) \\ &= \begin{cases} 0, & x < F^{-1}(0.5) \\ 2\,\overline{F}(x), & x \geq F^{-1}(0.5) \end{cases} \\ &= \mathrm{P}(X > x \mid X > F^{-1}(0.5)).\end{aligned} \tag{2.40}$$

The sf of $X_{max,2}^-$ in (2.40) is valid no matter if the rv X follows a subexponential distribution or not. It is interesting to observe that the sf of $X_{max,2}^-$ in (2.40) is analogous to the asymptotic behavior of the sum of two independent and identically distributed rvs following a subexponential distribution F, given by (2.39).

From (2.40), we obtain

$$\mathrm{VaR}_\kappa(X_{max,2}^-) = \mathrm{VaR}_{\frac{1+\kappa}{2}}(X), \qquad \kappa \in (0,1). \tag{2.41}$$

Assuming $S^- = \varphi^-(U)$ with the aggregate loss function φ^- having a convex shape, we deduce from (2.30) of Proposition 8 and (2.41) that

$$\mathrm{VaR}_\kappa(X_{max,2}^-) = \mathrm{VaR}_{\frac{1+\kappa}{2}}(X) \leq \mathrm{VaR}_{\frac{1+\kappa}{2}}(X) + \mathrm{VaR}_{\frac{1-\kappa}{2}}(X) = \mathrm{VaR}_\kappa(S^-), \quad \forall \kappa \in (0,1). \tag{2.42}$$

From definition (2.9) of the TVaR, we also have $\text{TVaR}_\kappa(X^-_{max,2}) \leq \text{TVaR}_\kappa(S^-)$, $\forall \kappa \in (0,1)$. One sees from (2.30) that the component $\text{VaR}_{\frac{1+\kappa}{2}}(X)$ increases to ∞, while $\text{VaR}_{\frac{1-\kappa}{2}}(X)$ decreases to 0. As κ gets close to 1, the latter can be neglected, implying that only $\text{VaR}_{\frac{1+\kappa}{2}}(X)$ also equal to $\text{VaR}_\kappa(X^-_{max,2})$, contributes to $\text{VaR}_\kappa(S^-)$. In conclusion, letting $\kappa \to 1$ in (2.42), we find

$$\text{VaR}_\kappa(S^-) \overset{\kappa \to 1}{\sim} \text{VaR}_\kappa(X^-_{max,2}). \tag{2.43}$$

Combining the results in (2.43), (2.39), and (2.40) leads to the next proposition.

Proposition 11 *Let X follow a subexponential distribution with cdf F such that φ^- is a convex function, $(X_1^\perp, X_2^\perp)$ be two independent and identically distributed copies of X, and (X_1^-, X_2^-) be counter-monotonic, identically distributed copies of X. Then,*

$$\text{P}(X_1^\perp + X_2^\perp > x) \overset{x \to \infty}{\sim} \text{P}(X_1^- + X_2^- > x) \overset{x \to \infty}{\sim} \text{P}(X^-_{max,2} > x) = \text{P}(X > x \mid X > F^{-1}(0.5)). \tag{2.44}$$

From (2.43) and (2.44), we have

$$\text{VaR}_\kappa(X_1^\perp + X_2^\perp) \overset{\kappa \to 1}{\sim} \text{VaR}_\kappa(X^-_{max,2}) \overset{\kappa \to 1}{\sim} \text{VaR}_\kappa(X_1^- + X_2^-). \tag{2.45}$$

The asymptotic relations in (2.45) imply that for a pair of risks following a common subexponential distribution (e.g. for catastrophic risks), the aggregate loss, under an independence structure, behaves, asymptotically, as the aggregate loss assuming counter-monotonicity.

Now, we investigate how the aggregate loss function φ^- is affected by both the dependence structure and the tail behavior of the marginals. We focus on the popular subexponential Pareto distribution, which plays a crucial role in actuarial science and quantitative risk management. For applications, we refer to Embrechts et al. (2013), McNeil et al. (2015) and Albrecher et al. (2017).

Assuming that $X_1^- \sim X_2^- \sim X$, with $X \sim \text{Pareto}(\alpha, \lambda)$, tail index $\alpha > 1, \lambda > 0$, and based on Proposition 10, the aggregate loss function $\varphi^-(u)$ has a convex shape for $u \in (0,1)$. It means that the support of the distribution of S^- is $[s_0, \infty)$, with $s_0 = 2 \times F^{-1}\left(\frac{1}{2}\right) = 2\lambda\left(0.5^{-\frac{1}{\alpha}} - 1\right)$ being its lower bound. Then, from Proposition 8 and with a risk level of $\kappa \in (0,1)$, we obtain

$$VaR_\kappa\left(S^-\right) = \lambda\left(\left(\frac{1+\kappa}{2}\right)^{-\frac{1}{\alpha}} + \left(\frac{1-\kappa}{2}\right)^{-\frac{1}{\alpha}} - 2\right) \tag{2.46}$$

and

$$TVaR_\kappa\left(S^-\right) = 2\lambda\left(\frac{\alpha}{(\alpha-1)(1-\kappa)}\left(1 - \left(\frac{1+\kappa}{2}\right)^{-\frac{1}{\alpha}+1} + \left(\frac{1-\kappa}{2}\right)^{-\frac{1}{\alpha}+1}\right) - 1\right). \tag{2.47}$$

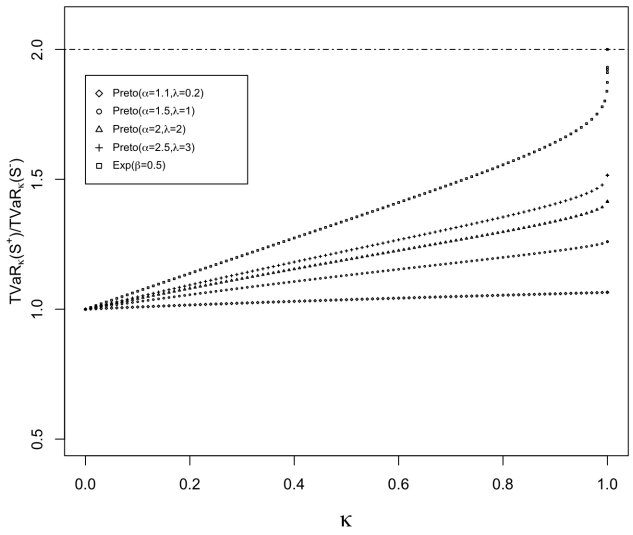


Figure 2.3 – Behavior of $\frac{TVaR_\kappa(S^+)}{TVaR_\kappa(S^-)}$, $\kappa \in (0,1)$ for Pareto and exponential marginals with identical mean.

If the pair of rvs is comonotonic, such that $X_1^+ \sim X_2^+ \sim X$, then, from (2.21) and (2.22) we have

$$VaR_\kappa\left(S^+\right) = 2\lambda\left((1-\kappa)^{-\frac{1}{\alpha}} - 1\right), \quad \kappa \in (0,1) \tag{2.48}$$

and

$$TVaR_\kappa\left(S^+\right) = 2\lambda\left(\frac{\alpha}{\alpha-1}(1-\kappa)^{-\frac{1}{\alpha}} - 1\right), \quad \kappa \in (0,1). \tag{2.49}$$

To compare the asymptotic behavior of the extreme aggregate losses S^- and S^+, we propose to analyze the asymptotic ratio of their associated risk measures. Notice that, using l'Hôpital rule with (2.46), (2.47), (2.48) and (2.49) , we have

$$\lim_{\kappa\uparrow 1}\frac{VaR_\kappa(S^+)}{VaR_\kappa(S^-)} = \lim_{\kappa\uparrow 1}\frac{TVaR_\kappa(S^+)}{TVaR_\kappa(S^-)} = 2^{1-\frac{1}{\alpha}} \in [0,2], \text{ for } \alpha > 1,$$

which is a monotone increasing function of the Pareto tail index α. In Figure 2.3, we show the curves of the ratio $\frac{TVaR_\kappa(S^+)}{TVaR_\kappa(S^-)}$ for Pareto tail index $\alpha \in \{1.1, 1.5, 2, 2.5\}$. When the value of α increases, the ratio tends to 2, which corresponds to the limit ratio when X is exponentially distributed.

Using the expression of $TVaR_\kappa(S^-)$ given in (2.47), we are able to find closed-form expressions for $DBI_\kappa^{TVaR}(S^-)$ and $CI_\kappa^{TVaR}(S^-)$, for $\kappa \in (0,1)$. Moreover, we can obtain their

corresponding limit value when κ tends to 1. Using l'Hôpital rule, we obtain

$$\lim_{\kappa\uparrow 1} DBI_{\kappa}^{TVaR}(S^{-}) = 1 - \lim_{\kappa\uparrow 1} \frac{\frac{d\left(TVaR_{\kappa}\left(S^{-}\right)-E[S]\right)}{d\kappa}}{\frac{d(2TVaR_{\kappa}(X)-2E[X])}{d\kappa}} = 1 - \lim_{\kappa\uparrow 1} \frac{VaR_{\kappa}\left(S^{-}\right)}{VaR_{\kappa}\left(S^{+}\right)} = 1 - 2^{\frac{1}{\alpha}-1},$$

$$\lim_{\kappa\uparrow 1} CI_{\kappa}^{TVaR}(S^{-}) = \lim_{\kappa\uparrow 1} \frac{TVaR_{\kappa}\left(S^{-}\right)}{TVaR_{\kappa}\left(S^{+}\right)} = 2^{\frac{1}{\alpha}-1}.$$

Therefore, for an identically Pareto distributed pair of risks, counter-monotonicity and independence lead to the same diversification benefit when the confidence level κ gets close to 1. Moreover, for the counter-monotonic case, the asymptotic expression for the diversification benefit index is a monotonic increasing function of the Pareto tail index α. If the pair of risks are independent, denoted $(X_1^{\perp}, X_2^{\perp})$ and with $X_1^{\perp} \sim X_2^{\perp} \sim X$, it can be shown, based on Corollary 1.3.2 of Embrechts et al. (2013), that $\lim_{\kappa\uparrow 1} DBI_{\kappa}^{TVaR}(S^{\perp}) = 1 - 2^{\frac{1}{\alpha}-1}$ and $\lim_{\kappa\uparrow 1} CI_{\kappa}^{TVaR}(S^{\perp}) = 2^{\frac{1}{\alpha}-1}$, which was expected according to (2.45).

In the following, we provide numerical examples to illustrate the asymptotic relation between $S^{\perp}$ and S^{-} represented by (2.45) and rewritten as

$$\overline{F}_{S^{-}}(x) \overset{x\to\infty}{\sim} \overline{F}_{S^{\perp}}(x),$$

for a pair of identically Pareto distributed risks. In Examples 5 and 6, the values of $\overline{F}_S(x) = \Pr(X_1 + X_2 > x)$ will need to be computed as precisely as possible. Cossette et al. (2019) compare two numerical methods to approximate $\overline{F}_S(x)$ for high values of x under Archimedean copula structures. It is important to note the usefulness of such approximations in the context of rare events. The first method is based on conditional Monte Carlo simulations and allows to propose four estimators with most of them having bounded relative errors. The second method is called the *Rectangles Method* and can be applied when the closed-form expression of the joint sf or cdf exists. Referring to Cossette et al. (2014), the approach is based on rectangles covering the region of interest. It produces bounds for $\overline{F}_S$ converging to the exact value as the number of rectangles increases. For the following calculations, the second method will be used to approximate $\overline{F}_S$ under the dependence structures and marginals considered. From Definition 2.1 of Cossette et al. (2014), the upper and lower bounds for the exact value of $\overline{F}_S(x)$ $(x \geq 0)$ are denoted by $\overline{A}_S^{u,m}(x)$ and $\overline{A}_S^{l,m}(x)$ respectively, with

$$\overline{A}_S^{u,m}(x) = 1 - \left[\sum_{i=1}^{2^m-1} F_{X_1,X_2}\left(\frac{i}{2^m}x, \frac{2^m-i}{2^m}x\right) - F_{X_1,X_2}\left(\frac{(i-1)}{2^m}x, \frac{2^m-i}{2^m}x\right)\right] \tag{2.50}$$

and

$$\overline{A}_S^{l,m}(x) = 1 - \left[\sum_{i=1}^{2^m} F_{X_1,X_2}\left(\frac{i}{2^m}x, \frac{2^m+1-i}{2^m}x\right) - F_{X_1,X_2}\left(\frac{(i-1)}{2^m}x, \frac{2^m+1-i}{2^m}x\right)\right], \tag{2.51}$$

where $m \in \mathbb{N}^+$ is the precision parameter. By construction, we have

$$\overline{A}_S^{l,m}(x) \leq \overline{F}_S(x) \leq \overline{A}_S^{u,m}(x),$$

for $x \geq 0$. Also, it is shown in Cossette et al. (2014) that

$$\lim_{m\to\infty} \overline{A}_S^{u,m}(x) \to \overline{F}_S(x) \text{ and } \lim_{m\to\infty} \overline{A}_S^{l,m}(x) \to \overline{F}_S(x), \text{ for } x \geq 0.$$

Example 5 *Let $X_1 \sim X_2 \sim X$, with two choices of distribution satisfying Proposition 10.(i) for X, meaning $X \sim \text{Pareto}(1.5, 1)$ and $X \sim \text{Exp}(0.5)$. We compare numerically the asymptotic behavior of $S^\perp = X_1^\perp + X_2^\perp$ and $S^- = X_1^- + X_2^-$. Exact values of $VaR_\kappa(S^-)$ are first calculated using (2.30) from Proposition 8 and then, lower and upper bounds for $\overline{F}_{S^\perp}(x)$ ($x = VaR_\kappa(S^-)$, $\kappa = 0.5, 0.9, 0.95, 1 - 10^{-j}, j = 2, ..., 9$) are computed using (2.50) and (2.51) with precision parameter $m = 12$. Numerical results are provided in Tables 2.1 and 2.2. This example allows us to analyze how extreme negative dependence plays a role in the behavior of the tail of the distribution of the aggregate loss rv, depending if X follows or not a subexponential distribution.*

In Tables 2.1 and 2.2, the values of $\overline{A}_{S^\perp}^{l,m}(x)$ and $\overline{A}_{S^\perp}^{u,m}(x)$ are very close, so that we can use them as fair approximations of $\overline{F}_{S^\perp}$. In Table 2.1, it is clear that $\overline{F}_{S^\perp}(VaR_\kappa(S^-))$ converges to $1 - \kappa$, as $\kappa \to 1$. This means that the difference between $\overline{F}_{S^\perp}(x)$ and $\overline{F}_{S^-}(x)$ becomes negligible as $x \to \infty$. In other words and as expected from (2.44) of Proposition 11, the counter-monotonicity does not have an impact on the distribution of the tail when X follows a Pareto distribution. It implies that $S^\perp$ and S^- have the same asymptotic tail distribution. When the distribution of X is not subexponential, we observe in Table 2.2 that the values of $\overline{F}_{S^\perp}(VaR_\kappa(S^-))$ are clearly different from $1 - \kappa$, for any κ. It implies that the extreme negative dependence has a significant impact on the distribution of S^-, even in its tail, when X follows a distribution that is not subexponential.

$1-\kappa$	$VaR_\kappa(S^-)$	$\overline{A}_{S^\perp}^{u,12}(VaR_\kappa(S^-))$	$\overline{A}_{S^\perp}^{l,12}(VaR_\kappa(S^-))$
5E-01	1.731256	4.85706E-01	4.85616E-01
1E-01	6.402850	1.19229E-01	1.19190E-01
5E-02	10.713093	5.81644E-02	5.81439E-02
1E-02	33.202866	1.07576E-02	1.07536E-02
1E-03	157.740439	1.01849E-03	1.01812E-03
1E-04	735.806333	1.00423E-04	1.00386E-04
1E-05	3418.95189	1.00109E-05	1.00072E-05
1E-06	15873.010519	1.00047E-06	1.00010E-06
1E-07	73679.630053	1.00037E-07	1.00001E-07
1E-08	341994.190721	1.00036E-08	1.00000E-08
1E-09	1587399.964407	1.00036E-09	1.00000E-09

Table 2.1 – Lower and upper bounds for the sf of the sum of independent and identically distributed rvs $X_1 \sim X_2 \sim \text{Pareto}(1.5, 1)$.

$1-\kappa$	$VaR_\kappa(S^-)$	$\overline{A}^{u,12}_{S^\perp}(VaR_\kappa(S^-))$	$\overline{A}^{l,12}_{S^\perp}(VaR_\kappa(S^-))$
5E-01	3.347953	5.01434E-01	5.01306E-01
1E-01	6.094051	1.92287E-01	1.92179E-01
5E-02	7.428395	11.49496E-02	11.48675E-02
1E-02	10.606660	3.13761E-02	3.13419E-02
1E-03	15.202805	4.30207E-03	4.29502E-03
1E-04	19.807075	5.45748E-04	5.44551E-04
1E-05	24.412155	6.61210E-05	6.59392E-05
1E-06	29.017316	7.76718E-06	7.74149E-06
1E-07	33.622486	8.92289E-07	8.88839E-07
1E-08	38.227656	10.07925E-08	10.03464E-08
1E-09	42.832826	11.23614E-09	11.18020E-09

Table 2.2 – Lower and upper bounds for the sf of the sum of independent and identically distributed rvs $X_1 \sim X_2 \sim \text{Exp}(0.5)$.

Example 6 *We consider a pair of identically distributed rvs, with joint behavior described under two copulas, namely the lower Fréchet bound copula and a transformed Gumbel copula with dependence parameter $\theta > 1$. We define their joint cdfs respectively by*

$$F_{X_1^-,X_2^-}(x_1,x_2) = \max\{F(x_1)+F(x_2)-1;0\},$$
$$F_{X_1^G,X_2^G}(x_1,x_2) = F(x_1) - \exp\left(-\left\{(-\ln F(x_1))^\theta + (-\ln(1-F(x_2)))^\theta\right\}^{(1/\theta)}\right),$$

for $x_1, x_2 \in \mathbb{R}^+$ and $\alpha \geq 1$. The latter is obtained assuming $X_1^G = F^{-1}(U_1)$ and $X_2^G = F^{-1}(1-U_2)$, where (U_1,U_2) is a uniformly distributed pair of rvs with Gumbel copula as joint cdf (i.e., $F_{U_1,U_2} = C_\theta^{Gumbel}$). Thus, it introduces a stronger negative dependence between values in the right tail and a weaker one between values in the left tail.

We compare numerically the asymptotic behavior of $S^G = X_1^G + X_2^G$ and $S^- = X_1^- + X_2^-$. The dependence parameter of $F_{X_1^G,X_2^G}$ is fixed at $\theta = 1.5$. We use the same numerical procedure and two choices of marginal distribution as in Example 5. Tables 2.3 and 2.4 provide respectively the numerical results when $X \sim \text{Pareto}(1.5,1)$ and $X \sim \text{Exp}(0.5)$.

$1-\kappa$	$VaR_\kappa(S^-)$	$\overline{A}^{u,12}_{S^G}(VaR_\kappa(S^-))$	$\overline{A}^{l,12}_{S^G}(VaR_\kappa(S^-))$
5E-01	1.731256	5.27194E-01	5.27079E-01
1E-01	6.402850	1.08710E-01	1.08673E-01
5E-02	10.713093	5.25253E-02	5.25069E-02
1E-02	33.202866	1.01263E-02	1.01226E-02
1E-03	157.740439	1.00210E-03	1.00173E-03
1E-04	735.806333	1.00061E-04	1.00025E-04
1E-05	3418.951897	1.00039E-05	1.00002E-05
1E-06	15873.010519	1.00036E-06	1.00000E-06
1E-07	73679.630053	1.00036E-07	1.00000E-07
1E-08	341994.190721	1.00036E-08	1.00000E-08
1E-09	1587399.964407	1.00036E-09	1.00000E-09

Table 2.3 – Lower and upper bounds for the sf of the sum of rvs $X_1 \sim X_2 \sim$ Pareto$(1.5, 1)$ linked by transformed Gumbel copula with dependence parameter $\theta = 1.5$.

$1-\kappa$	$VaR_\kappa(S^-)$	$\overline{A}^{u,12}_{S^G}(VaR_\kappa(S^-))$	$\overline{A}^{l,12}_{S^G}(VaR_\kappa(S^-))$
5E-01	3.347953	5.43763E-01	5.43591E-01
1E-01	6.094051	1.54341E-01	1.54226E-01
5E-02	7.428395	7.85721E-02	7.84993E-02
1E-02	10.606660	1.52873E-02	1.52668E-02
1E-03	15.202805	1.43601E-03	1.43328E-03
1E-04	19.807075	1.36992E-04	1.36656E-04
1E-05	24.412155	1.32580E-05	1.32181E-05
1E-06	29.017316	1.29501E-06	1.29039E-06
1E-07	33.622486	1.27230E-07	1.26705E-07
1E-08	38.227656	1.25471E-08	1.24883E-08
1E-09	42.832826	1.24057E-09	1.23407E-09

Table 2.4 – Lower and upper bounds for the sf of the sum of rvs $X_1 \sim X_2 \sim$ Exp(0.5) linked by transformed Gumbel copula with dependence parameter $\theta = 1.5$.

In Table 2.3, we observe that $\overline{F}_{S^G}(VaR_\kappa(S^-))$ *converges to* $1-\kappa$*, as* $\kappa \to 1$*. It implies that the difference between* $\overline{F}_{S^G}(x)$ *and* $\overline{F}_{S^-}(x)$ *tends to 0 as* $x \to \infty$*. In other words, the negative dependence and counter-monotonicity have a similar impact on the distribution of the tail when* X *follows a Pareto distribution (i.e., a subexponential distribution). When the distribution of* X *is not subexponential, we observe in Table 2.4 that the values of* $\overline{F}_{S^G}(VaR_\kappa(S^-))$ *are clearly different from* $1-\kappa$*, for any* κ*. It means that the extreme negative dependence and negative dependence have different impacts on the distribution of* S^- *and* S^G*, even in their tail, when* X *follows a distribution that is not subexponential.*

2.5 Non-identically distributed counter-monotonic rvs

Let us now assume that X_1^- and X_2^- are not identically distributed and follow continuous, strictly positive distributions with cdf F_i, $i = 1, 2$. The asymmetric aggregate loss function is given by

$$\varphi^-(u) = F_1^{-1}(u) + F_2^{-1}(1-u), \qquad u \in (0,1). \tag{2.52}$$

In order to study the behavior of the loss function (2.52), we develop its first and second derivatives. We obtain

$$\varphi^{-(1)}(u) = \frac{\partial \varphi^-(u)}{\partial u} = \frac{1}{F_1^{(1)}\left(F_1^{-1}(u)\right)} - \frac{1}{F_2^{(1)}\left(F_2^{-1}(1-u)\right)},$$

$$\varphi^{-(2)}(u) = \frac{\partial^2 \varphi^-(u)}{\partial u^2} = -\left(\frac{F_1^{(2)}\left(F_1^{-1}(u)\right)}{\left(F_1^{(1)}\left(F_1^{-1}(u)\right)\right)^3} + \frac{F_2^{(2)}\left(F_2^{-1}(1-u)\right)}{\left(F_2^{(1)}\left(F_2^{-1}(1-u)\right)\right)^3} \right), \tag{2.53}$$

for $u \in (0,1)$. Assume that the marginal distributions are chosen such that (2.52) is either convex or concave. Then, φ^- admits either a minimum or a maximum value, denoted s_0, at u_0 satisfying $\varphi^{-(1)}(u_0) = 0$. In addition, if we switch the marginal distributions between them such that the aggregate loss function becomes $\varphi^{-*}(u) = F_2^{-1}(u) + F_1^{-1}(1-u)$, then the shape of φ^{-*} will be the reflection of the shape of φ^-, with a minimum or a maximum value $u_0^* = 1 - u_0$. Moreover, we will obtain the same cdf, VaR and TVaR of the aggregate loss in both cases.

2.5.1 Convex aggregate loss function φ^-

Assume that the aggregate loss function φ^- has a convex shape over $(0,1)$. Then, the support of S^- will be $\left(\varphi^-(u_0), \sup_{u\in(0,1)} \varphi^-(u)\right)$. Let us define the function $\xi : (0,1) \to (0,1)$, such that $\xi(\kappa) = \underset{u\in(0,1)}{\operatorname{argmin}} \left| VaR_\kappa\left(S^-\right) - \varphi^-\left(u\right)\right|$ is the unique solution to $\varphi^-\left(u\right) = \varphi^-\left(u+\kappa\right)$. In the next proposition, we provide a closed-form expression for the TVaR of S^- using function ξ.

Proposition 12 *Let $S^- := \varphi^-(U)$, where φ^- is convex on $(0,1)$. Then, for $\kappa \in (0,1)$, we have*

$$\begin{aligned} TVaR_\kappa\left(S^-\right) &= \frac{1}{1-\kappa}\Big(\xi(\kappa)\left(LTVaR_{\xi(\kappa)}\left(X_1\right) + TVaR_{1-\xi(\kappa)}\left(X_2\right)\right) + \\ &\qquad (1-\xi(\kappa)-\kappa)\left(TVaR_{\xi(\kappa)+\kappa}\left(X_1\right) + LTVaR_{1-\xi(\kappa)-\kappa}\left(X_2\right)\right)\Big). \end{aligned} \tag{2.54}$$

Proof. Since $\varphi^-(u)$ is convex on $(0,1)$, we can define

$$\varphi^-(u) = \begin{cases} \varphi_1^-(u), & 0 < u < u_0 \\ \varphi_2^-(u), & u_0 \le u < 1, \end{cases} \tag{2.55}$$

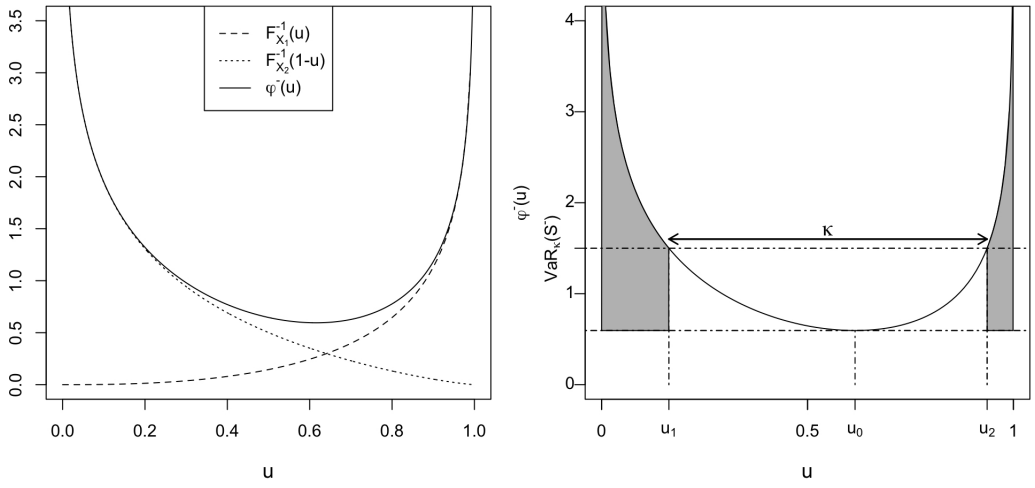


Figure 2.4 – Illustration of the convex case with non-identically distributed risks $X_1^- \sim$ Gamma$(0.4, 1)$ and $X_2^- \sim$ Gamma$(0.8, 1)$.

with $u_0 = \underset{u\in(0,1)}{\operatorname{argmin}} \varphi^-(u)$, and such that φ_1^- is strictly decreasing on $(0, u_0)$ and φ_2^- strictly increasing on $(u_0, 1)$. Then $F_{S^-}(s) = u_2 - u_1$, where $u_i = \left(\varphi_i^-\right)^{-1}(s)$, $i = 1, 2$, and $\varphi^-(u_1) = \varphi^-(u_2) = s$. It follows that, $\exists\, \xi(\kappa) \in (0, u_0)$ such that $u_1 = \xi(\kappa)$ and $u_2 = \xi(\kappa) + \kappa$ are the two solutions to $\varphi^-(u) = VaR_\kappa(S^-)$. Then, it follows that $VaR_\kappa(S^-) = \varphi^-(\xi(\kappa))$. As for the TVaR, we have

$$\begin{aligned}
TVaR_\kappa\left(S^-\right) &= \frac{1}{1-\kappa}\left(\int_0^{u_1}\varphi^-(u)\,du + \int_{u_2}^1 \varphi^-(u)\,du\right)\\
&= \frac{1}{1-(u_2-u_1)}\Big(u_1 LTVaR_{u_1}(X_1) + u_1 TVaR_{1-u_1}(X_2)\\
&\quad + (1-u_2)\,TVaR_{u_2}(X_1) + (1-u_2)\,LTVaR_{1-u_2}(X_2)\Big)\\
&= \frac{1}{1-\kappa}\Big(\xi(\kappa)\left(LTVaR_{\xi(\kappa)}(X_1) + TVaR_{1-\xi(\kappa)}(X_2)\right)\\
&\quad + (1-\xi(\kappa)-\kappa)\left(TVaR_{\xi(\kappa)+\kappa}(X_1) + LTVaR_{1-\xi(\kappa)-\kappa}(X_2)\right)\Big),
\end{aligned}$$

with $u_1 = \xi(\kappa)$ and $u_2 = \xi(\kappa) + \kappa$. This result is illustrated in Figure 2.4 in which the numerator is given by the gray area. ■

2.5.2 Concave aggregate loss function φ^-

In the following, we assume that the counter-monotonic rvs are bounded and φ^- has a concave pattern on $(0, 1)$. Thus, the support of S^- is $\left(\inf_{u\in(0,1)}\varphi^-(u), u_0\right)$, with $u_0 = \underset{u\in(0,1)}{\operatorname{argmax}}\,\varphi^-(u)$.

Proposition 13 *Let $S^- := \varphi^-(U)$, where φ^- is concave on $(0,1)$. Then we have*

$$\begin{aligned} TVaR_\kappa\left(S^-\right) &= \frac{1}{1-\kappa}\Big((1-\xi(\kappa))\left(TVaR_{\xi(\kappa)}\left(X_1\right) + LTVaR_{1-\xi(\kappa)}\left(X_2\right)\right) - \\ &\quad (\kappa-\xi(\kappa))\left(TVaR_{1+\xi(\kappa)-\kappa}\left(X_1\right) + LTVaR_{\kappa-\xi(\kappa)}\left(X_2\right)\right)\Big), \end{aligned}$$

for $\kappa \in (0,1)$.

Proof. Since $\varphi^-(u)$ is concave on $(0,1)$, it admits a maximum value $s_0 = \varphi^-(u_0)$, with $u_0 = \underset{u\in(0,1)}{\operatorname{argmax}}\,\varphi^-\left(u\right)$. Then, we can define φ^- using representation (2.55), such that φ_1^- is strictly increasing on $(0,u_0)$ and φ_2^- is strictly decreasing on $(u_0,1)$. Then, it follows that

$$\begin{aligned} F_{S^-}(s) &= \Pr(\varphi_1^-(U) < s, \varphi_2^-(U) < s) \\ &= \begin{cases} 0, & s \le 0 \\ u_1 + (1-u_2), & 0 < s < s_0 \\ 1, & s \ge s_0. \end{cases} \end{aligned} \tag{2.56}$$

We can solve $F_{S^-}(s) = \kappa \in (0,1)$ in the same manner as in the convex case. From (2.56), $\exists \xi(\kappa) \in (0,u_0)$ such that $\varphi^-(\xi(\kappa)) = \varphi^-(1+\xi(\kappa)-\kappa)$. Then, we have $VaR_\kappa\left(S^-\right) = \varphi^-(\xi(\kappa))$, implying that

$$\begin{aligned} TVaR_\kappa\left(S^-\right) &= \frac{1}{1-\kappa}\int_{u_1}^{u_2} \varphi^-\left(u\right) du \\ &= \frac{\int_{u_1}^1 F_1^{-1}\left(u\right) du - \int_{u_2}^1 F_1^{-1}\left(u\right) du + \int_{u_1}^1 F_2^{-1}\left(1-u\right) du - \int_{u_2}^1 F_2^{-1}\left(1-u\right) du}{1-\kappa} \\ &= \frac{(1-u_1)\,TVaR_{u_1}\left(X_1\right) - (1-u_2)\,TVaR_{u_2}\left(X_1\right) + (1-u_1)\,LTVaR_{1-u_1}\left(X_2\right)}{1-\kappa} \\ &\quad - \frac{(1-u_2)\,LTVaR_{1-u_2}\left(X_2\right)}{1-\kappa} \\ &= \frac{1}{1-\kappa}\Big((1-\xi(\kappa))\left(TVaR_{\xi(\kappa)}\left(X_1\right) + LTVaR_{1-\xi(\kappa)}\left(X_2\right)\right) \\ &\quad - (\kappa-\xi(\kappa))\left(TVaR_{1+\xi(\kappa)-\kappa}\left(X_1\right) + LTVaR_{\kappa-\xi(\kappa)}\left(X_2\right)\right)\Big), \end{aligned}$$

for $\kappa \in (0,1)$. Figure 2.5 illustrates the above results. ■

2.5.3 Sufficient conditions for convex or concave φ^-

A sufficient condition for having either a convex or a concave aggregate loss function φ^- may depend on the monotonicity of both pdfs f_1 and f_2. In the next proposition, we focus on distributions with a pdf that is strictly monotone.

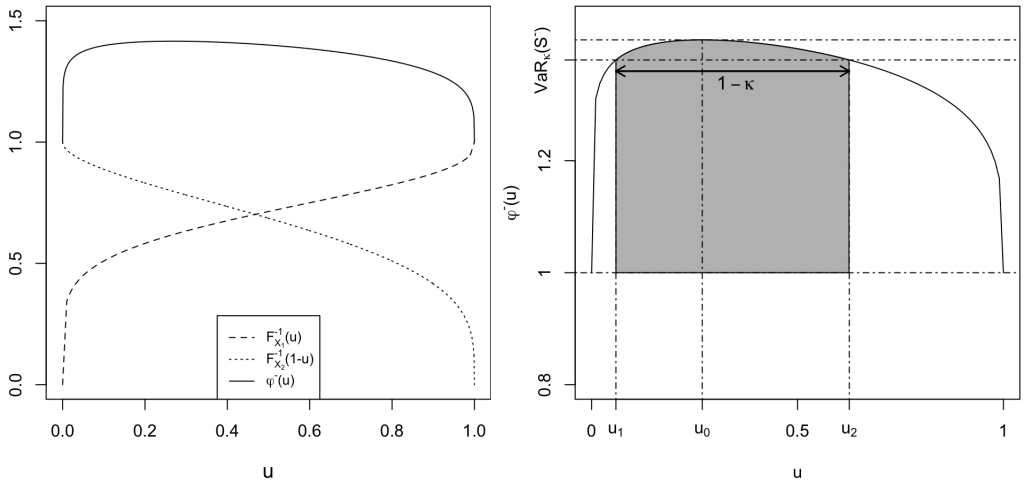


Figure 2.5 – Illustration of the concave case with non-identically distributed risks $X_1^- \sim$ Beta$(7,3)$ and $X_2^- \sim$ Beta$(4,2)$.

Proposition 14 *Let X_1^- and X_2^- be a pair of counter-monotonic rvs following continuous and strictly positive distributions.*

(i) If their pdfs f_1 and f_2 are strictly decreasing, then φ^- is convex.

(ii) If their pdfs f_1 and f_2 are strictly increasing and defined on a bounded support, then φ^- is concave.

Proof. (i) Since f_1 and f_2 are both strictly decreasing, we obtain $f_i^{(1)}(x) = F_i^{(2)}(x) < 0$, for $i = 1, 2$. Therefore, in (2.53), $\varphi^{-(2)}(u) > 0$, $u \in (0,1)$. (ii) The result is proven in a similar fashion. ■

In the next example, we illustrate a combination of Proposition 12 and Proposition 14.(i).

Example 7

We define two counter-monotonic rvs $X_1^- \sim Weibull(\tau_1, \beta_1)$ and $X_2^- \sim Weibull(\tau_2, \beta_2)$, $\beta_1, \beta_2 > 0$, $\tau_1, \tau_2 \in (0,1)$. Using Proposition 12 and Proposition 14.(i), given that $\tau_1, \tau_2 \in (0,1)$ in the Weibull distribution, we have

$$\begin{aligned} TVaR_\kappa\left(S^-\right) &= \frac{1}{1-\kappa}\left(\frac{1}{\beta_1}\Gamma\left(1+\frac{1}{\tau_1}\right)\left(H\left(-\ln(1-\xi(\kappa)); 1+\frac{1}{\tau_1}; 1\right)\right.\right. \\ &\quad \left. + \overline{H}\left(-\ln(1-\xi(\kappa)-\kappa); 1+\frac{1}{\tau_1}; 1\right)\right) \\ &\quad + \frac{1}{\beta_2}\Gamma\left(1+\frac{1}{\tau_2}\right)\left(\overline{H}\left(-\ln(\xi(\kappa)); 1+\frac{1}{\tau_2}; 1\right)\right. \\ &\quad \left.\left. + H\left(-\ln(\xi(\kappa)+\kappa); 1+\frac{1}{\tau_2}; 1\right)\right)\right), \end{aligned}$$

with $\kappa \in (0,1)$, H and $\overline{H}$ denoting, respectively, the cdf and sf of a gamma distribution. If the rvs are identically distributed, i.e. $X_1^- \sim X_2^- \sim Weibull(\tau, \beta)$, $\beta > 0$, $\tau \in (0,1)$, then

$$VaR_\kappa\left(S^-\right) = \frac{1}{\beta}\left(\left(-\ln\left(\frac{1+\kappa}{2}\right)\right)^{\frac{1}{\tau}} + \left(-\ln\left(\frac{1-\kappa}{2}\right)\right)^{\frac{1}{\tau}}\right)$$

and

$$\begin{aligned} TVaR_\kappa\left(S^-\right) &= \frac{2}{\beta(1-\kappa)}\Gamma\left(1+\frac{1}{\tau}\right)\left(H\left(-\ln\left(\frac{1+\kappa}{2}\right);1+\frac{1}{\tau};1\right)\right. \\ &\quad \left.+\overline{H}\left(-\ln\left(\frac{1-\kappa}{2}\right);1+\frac{1}{\tau};1\right)\right), \end{aligned}$$

for $\kappa \in (0,1)$, by Proposition 8.

We give an illustration of Proposition 13 with Proposition 14.(ii) in the following example.

Example 8 *Let X_1^- and X_2^- be counter-monotonic risks, with $X_1^- \sim \text{Beta}(\alpha_1, \beta_1)$, $\alpha_1 > 1, \beta_1 = 1$, and $X_2^- \sim \text{Beta}(\alpha_2, \beta_2)$, $\alpha_2 > 1, \beta_2 = 1$. Using Proposition 14.(ii), the aggregate loss function has a concave shape. According to Proposition 13, we can deduce that*

$$\begin{aligned} &TVaR_\kappa\left(S^-\right) \\ =&\frac{1}{1-\kappa}\left(\frac{\alpha_1}{\alpha_1+1}\left(\overline{B}\left(VaR_{\xi(\kappa)}(X_1),\alpha_1+1,1\right)-\overline{B}\left(VaR_{1+\xi(\kappa)-\kappa}(X_1),\alpha_1+1,1\right)\right)\right. \\ &\left.+\frac{\alpha_2}{\alpha_2+1}\left(B\left(VaR_{1-\xi(\kappa)}(X_2),\alpha_2+1,1\right)-B\left(VaR_{\kappa-\xi(\kappa)}(X_2),\alpha_2+1,1\right)\right)\right), \end{aligned}$$

for $\kappa \in (0,1)$, with $\xi(\kappa) \leq u_0$ the solution to $\varphi^-(\xi(\kappa)) = \varphi^-(1+\xi(\kappa)-\kappa)$.

In Section 2.6, we give another sufficient condition on the distributions of X_1^- and X_2^- for a convex aggregate loss function φ^-.

2.6 Additional results

In this section, we consider the sum of a pair of counter-monotonic rvs X_1^- and X_2^-, defined in terms of two exponential functions of a symmetric rv Z. In this case, even if we observe that φ^- is convex, the pdfs of the rvs X_1^- and X_2^- do not satisfy the sufficient conditions of Propositions 10 or 14. However, we go around this in the next two propositions.

Proposition 15 *Let Z be a rv following a symmetric and absolutely continuous distribution with cdf F, such that $F(0) = 0.5$ and $F^{-1}(u) = -F^{-1}(1-u)$, for $u \in (0,1)$. We define a pair*

of counter-monotonic rvs (X_1^-, X_2^-)*, with* $X_i^- = \mathrm{e}^{\mu_i+\sigma\times Z}$*, for* $i = 1, 2$*, with* $\mu_1, \mu_2 \in \mathbb{R}$ *and* $\sigma > 0$*. Then, the cdf of* S^- *is given by*

$$F_{S^-}(x) = F\left(\frac{\ln\left(\frac{x+\sqrt{x^2-4\mathrm{e}^{\mu_1+\mu_2}}}{2\mathrm{e}^{\mu_1}}\right)}{\sigma}\right) - F\left(\frac{\ln\left(\frac{x-\sqrt{x^2-4\mathrm{e}^{\mu_1+\mu_2}}}{2\mathrm{e}^{\mu_1}}\right)}{\sigma}\right), \quad x \geq 2\mathrm{e}^{\frac{\mu_1+\mu_2}{2}}.$$

Proof. The aggregate loss S^- can be expressed as

$$\begin{aligned} S^- &= \mathrm{e}^{\mu_1+\sigma F^{-1}(U)} + \mathrm{e}^{\mu_2+\sigma F^{-1}(1-U)} \\ &= \mathrm{e}^{\mu_1+\sigma Z} + \mathrm{e}^{\mu_2-\sigma Z}. \end{aligned}$$

Let $Y = \mathrm{e}^{\sigma Z}$. Then,

$$\begin{aligned} F_{S^-}(x) &= \Pr\left(\mathrm{e}^{\mu_1}Y^2 - xY + \mathrm{e}^{\mu_2} \leq 0\right) \\ &= \Pr\left(\frac{\ln\left(\frac{x-\sqrt{x^2-4\mathrm{e}^{\mu_1+\mu_2}}}{2\mathrm{e}^{\mu_1}}\right)}{\sigma} \leq Z \leq \frac{\ln\left(\frac{x+\sqrt{x^2-4\mathrm{e}^{\mu_1+\mu_2}}}{2\mathrm{e}^{\mu_1}}\right)}{\sigma}\right) \end{aligned}$$

for $x \geq 2\mathrm{e}^{\frac{\mu_1+\mu_2}{2}}$. Therefore $\underset{u\in(0,1)}{\operatorname{argmin}}\ \varphi^-(u) = F\left(\frac{\mu_2-\mu_1}{2\sigma}\right)$. ■

In Proposition 15, F can notably be the cdf of a uniform distribution, a normal distribution, a Laplace distribution, a logistic distribution, or a Student distribution.

Proposition 16 *Let* Z *be a rv following a symmetric and absolutely continuous distribution with pdf* f *and cdf* F*, such that* $F(0) = 0.5$ *and* $F^{-1}(u) = -F^{-1}(1-u)$*, for* $u \in (0, 1)$*. We define a pair of counter-monotonic rvs* (X_1^-, X_2^-)*, defined by* $X_i^- = \mathrm{e}^{\mu_i+\sigma_i\times Z}$*, for* $i = 1, 2$*, with* $\mu_1, \mu_2 \in \mathbb{R}$ *and* $\sigma_1, \sigma_2 > 0$*. Then, the aggregate loss function* φ^- *has a convex pattern on* $(0, 1)$*.*

Proof. To prove the convexity of φ^-, we investigate the sign of its first derivative expressed as

$$\varphi^{-(1)}(u) = \frac{\partial\varphi^-(u)}{\partial u} = \frac{f_2(F_2^{-1}(1-u)) - f_1(F_1^{-1}(u))}{f_1(F_1^{-1}(u)) \times f_2(F_2^{-1}(1-u))}, \qquad u \in (0, 1).$$

Given the stochastic representation of X_i, we have $f_i(x) = \frac{1}{x\sigma_i} f\left(\frac{\ln(x)-\mu_i}{\sigma_i}\right)$ and $F_i^{-1}(u) = \mathrm{e}^{\mu_i+\sigma_i F^{-1}(u)}$, for $i = 1, 2$. Now, assume that $\varphi^{-(1)} \leq 0$ on a given interval $I_1 \subset (0, 1)$. Then,

we have

$$
\begin{aligned}
& f_1(F_1^{-1}(u)) \geq f_2(F_2^{-1}(1-u)) \\
& \Leftrightarrow \frac{1}{\sigma_1 F_1^{-1}(u)} f\left(\frac{\ln\left(F_1^{-1}(u)\right)-\mu_1}{\sigma_1}\right) \geq \frac{1}{\sigma_2 F_2^{-1}(1-u)} f\left(\frac{\ln\left(F_2^{-1}(1-u)\right)-\mu_2}{\sigma_2}\right) \\
& \Leftrightarrow e^{(\mu_2-\mu_1)-(\sigma_2+\sigma_1)F^{-1}(u)} \geq \frac{\sigma_1}{\sigma_2} \\
& \Leftrightarrow u \leq F\left(\frac{\mu_2-\mu_1-\ln\left(\frac{\sigma_1}{\sigma_2}\right)}{\sigma_2+\sigma_1}\right).
\end{aligned}
$$

Thus, φ^- is decreasing on $I_1 = \left(0, F\left(\left(\mu_2-\mu_1-\ln\left(\sigma_1/\sigma_2\right)\right)/\left(\sigma_2+\sigma_1\right)\right)\right]$ and increasing on $I_2 = \left[F\left(\left(\mu_2-\mu_1-\ln\left(\sigma_1/\sigma_2\right)\right)/\left(\sigma_2+\sigma_1\right)\right), 1\right)$, which confirms its convexity on $(0,1)$. ■

Whether the pair of counter-monotonic rvs (X_1^-, X_2^-), defined in Proposition 16, are identically distributed or not, we can deduce, given the convexity of φ^-, closed-form expressions for the VaR and TVaR according to Propositions 8 and 12, respectively.

Example 9 *Let (X_1^-, X_2^-) be a pair of counter-monotonic rvs, defined by $X_i^- = \mathrm{e}^{\mu_i+\sigma_i\times Z}$, for $i=1,2$, with $Z \sim \text{Normal}(0,1)$, $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1 = \sigma_2 = \sigma > 0$. According to Propositions 15 and 16, the aggregate loss rv S^- takes value in $[\varphi^-(\Phi(\frac{\mu_1-\mu_2}{2\sigma})), \infty)$ and it is characterized by a convex aggregate loss function φ^-. Its cdf is given by*

$$
F_{S^-}(x) = \Phi\left(\frac{\ln\left(\frac{x+\sqrt{x^2-4\mathrm{e}^{\mu_1+\mu_2}}}{2\mathrm{e}^{\mu_1}}\right)}{\sigma}\right) - \Phi\left(\frac{\ln\left(\frac{x-\sqrt{x^2-4\mathrm{e}^{\mu_1+\mu_2}}}{2\mathrm{e}^{\mu_1}}\right)}{\sigma}\right), \tag{2.57}
$$

for $x \geq 2\mathrm{e}^{\frac{\mu_1+\mu_2}{2}}$. Now, assume that $\mu_1 = \mu_2 = \mu$ and hence X_1^- and X_2^- become identically distributed. Given Proposition 8, for $\kappa \in (0,1)$, the VaR and TVaR of S^- are respectively given by

$$
VaR_\kappa(S^-) = \mathrm{e}^{\mu+\sigma\Phi^{-1}\left(\frac{1-\kappa}{2}\right)} + \mathrm{e}^{\mu+\sigma\Phi^{-1}\left(\frac{1+\kappa}{2}\right)}
$$

and

$$
TVaR_\kappa(S^-) = \frac{2}{1-\kappa}\mathrm{e}^{\mu+\frac{\sigma^2}{2}}\left(1+\Phi\left(\Phi^{-1}\left(\frac{1-\kappa}{2}\right)-\sigma\right)-\Phi\left(\Phi^{-1}\left(\frac{1+\kappa}{2}\right)-\sigma\right)\right). \tag{2.58}
$$

Then we can deduce the following stop-loss premium of S^-

$$
\pi_{S^-}(d) = \begin{cases} 2\mathrm{e}^{\mu+\frac{1}{2}\sigma^2} - d, & 0 \leq d < 2\mathrm{e}^{\mu} \\ 2\mathrm{e}^{\mu+\frac{1}{2}\sigma^2}\left(1+\Phi\left(\Phi^{-1}\left(\frac{\overline{F}_{S^-}(d)}{2}\right)-\sigma\right)-\Phi\left(\Phi^{-1}\left(\frac{1+F_{S^-}(d)}{2}\right)-\sigma\right)\right) & \\ -d\overline{F}_{S^-}(d), & d \geq 2\mathrm{e}^{\mu}. \end{cases} \tag{2.59}
$$

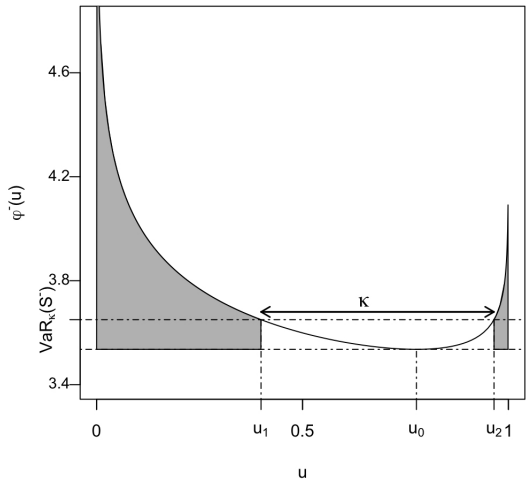


Figure 2.6 – Illustration of φ^- for $X_1^- \sim \text{LNorm}(0.6, 0.2)$ and $X_2^- \sim \text{LNorm}(0.575, 0.3)$.

The second term in (2.59) is obtained by combining $\pi_{S^-}(d) = \overline{F}_{S^-}(d)\, TVaR_{F_{S^-}(d)}(S^-) - d\overline{F}_{S^-}(d)$ *with (2.57) and (2.58).*

In the next example, we consider a pair of non-identically distributed counter-monotonic rvs following lognormal distributions. We aim to illustrate the impact of the convex order between the rvs on the asymmetric pattern of φ^-.

Example 10 *Let* X_1^- *and* X_2^- *be counter-monotonic rvs, such that* $X_1^- \sim \text{LNorm}(\mu_1 = 0.6, \sigma_1 = 0.2)$ *and* $X_2^- \sim \text{LNorm}(\mu_2 = 0.575, \sigma_2 = 0.3)$*, with* $E[X_1^-] = E[X_2^-]$*. Given the extreme negative dependence, one random component of the aggregate loss will increase, while the other one will decrease. Figure 2.6 shows the asymmetric shape of the aggregate loss function* φ^-*, which is convex on* $(0,1)$ *and takes value in* $[3.5360, \infty)$ *according to Proposition 16. We observe that* $X_1^- := F_1^-(U)$ *increases more slowly than* $X_2^- := F_2^-(1-U)$ *decreases, with* $u_0 = 0.7766 > \frac{1}{2}$*. Since* $X_1^- \preceq_{cx} X_2^-$*, the component* $F_2^-(1-U)$ *dominates the component* $F_1^-(U)$ *in their sum within* $\varphi^-(U)$*. Furthermore, according to Proposition 12,* $TVaR_\kappa(S^-)$*, for* $\kappa \in (0,1)$*, can be written as*

$$\begin{aligned} TVaR_\kappa(S^-) \;=\; & \frac{1}{1-\kappa}\Big(e^{\mu_1+\frac{\sigma_1^2}{2}} \left(1 + \Phi(\Phi^{-1}(\zeta(\kappa)) - \sigma_1) - \Phi(\Phi^{-1}(\zeta(\kappa)+\kappa) - \sigma_1)\right) + \\ & e^{\mu_2+\frac{\sigma_2^2}{2}} \left(1 + \Phi\left(\Phi^{-1}(1-\zeta(\kappa)-\kappa) - \sigma_2\right) - \Phi\left(\Phi^{-1}(1-\zeta(\kappa)) - \sigma_2\right)\right)\Big). \end{aligned}$$

2.7 Conclusion

In this paper, we present simple and nice expressions for the VaR and TVaR of the sum of two counter-monotonic rvs S^-. The results are obtained for cases where the rvs X_1^- and X_2^- have symmetric and absolutely continuous distributions or unimodal distributions with support on $\mathbb{R}^+$. Using our approach, we derived closed-form expressions with intuitive interpretations. Assuming that X_1^- and X_2^- follow a subexponential distribution F, such expressions allowed us to investigate in depth the tail of the distribution of their sum. This analysis lead to the interesting conclusion that under a negative dependence relation, even extreme, the asymptotic behavior of the aggregate loss is similar to the sum of the corresponding pair of independent rvs, i.e. $\overline{F}_{S^-}(x) \overset{x\to\infty}{\sim} \overline{F}_{S^\perp}(x)$, and it is defined solely in terms of $\overline{F}(x)$ as $x \to \infty$.

For $n > 2$, let $\underline{X} = (X_1, \ldots, X_n)$ be a vector of n rvs, with $F_{\underline{X}} \in \Gamma(F_1, \ldots, F_n)$. The Fréchet lower bound W_n, defined in (1), is in general not a cdf, and there is no universal concept of extreme negative dependence defining a joint cdf $F_{\underline{X}}$ which is the solution to the optimization problem $\inf_{F_{\underline{X}} \in \Gamma(F_1,\ldots,F_n)} \{E[\phi(S)]\}$, with ϕ a convex function such that $E[\phi(S)] < \infty$, for any marginals $F_1, \ldots, F_n$. For these reasons, different notions of extreme negative dependence have been proposed in the literature under marginals assumptions. Not surprisingly, we wonder if the proposed approach in this paper for the special case of bivariate counter-monotonic risks can be extended to some multivariate extreme negative dependence notions.

The most intuitive multivariate extension of the notion of counter-monotonicity is pairwise counter-monotonicity. Given, e.g., Proposition 3.2 of Puccetti et al. (2015), the Fréchet class $\Gamma(F_1, ..., F_n)$ supporting the joint cdfs of pairwise counter-monotonic random vectors does not contain cdfs of vectors with strictly continuous marginals. Hence, this multivariate dependence concept can not be used in the context of the present paper. The concept of complete mixability is defined for identically distributed rvs. The beta distribution example given in Section 2.4 with monotone pdf is a context to which the concept of complete mixability could be considered. Also, Gaffke and Rüscherndorf (1981) illustrate why complete mixability can not be defined when the marginals are exponential. Wang et al. (2013) also provide other examples which do not fit in the context of the present paper. The concept of Σ-counter-monotonicity introduced in Puccetti et al. (2015) seems promising. In futur research, we will explore more deeply how to adapt our approach to that notion through the investigation of an appropriate stochastic representation of the aggregate loss under such multivariate dependence.

Acknowledgments

The authors wish to thank the anonymous referees and the editor for several valuable comments and suggestions which significantly improved the manuscript. We also thank Samuel Pineault for his input and insight on earlier works which has lead to this research paper. This work was partially supported by the Natural Sciences and Engineering Research Council of Canada

(Cossette: 054993; Marceau: 053934) and by the Chaire en actuariat de l'Université Laval (Chaoubi, Cossette, Gadoury and Marceau: FO502320).

2.8 Bibliography

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.

Albrecher, H., Teugels, J. L., and Beirlant, J. (2017). *Reinsurance: Actuarial and Statistical Aspects*. John Wiley & Sons.

Alink, S., Löwe, M., and Wüthrich, M. V. (2004). Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics*, 35(1):77–95.

Barbe, P., Fougeres, A.-L., and Genest, C. (2006). On the tail behavior of sums of dependent risks. *ASTIN Bulletin: The Journal of the IAA*, 36(2):361–373.

Bürgi, R., Dacorogna, M. M., and Iles, R. (2008). Risk aggregation, dependence structure and diversification benefit. *Stress testing for financial institutions*.

Cheung, K. C., Dhaene, J., Lo, A., and Tang, Q. (2014). Reducing risk by merging counter-monotonic risks. *Insurance: Mathematics and Economics*, 54:58–65.

Cheung, K. C. and Lo, A. (2013a). Characterizations of counter-monotonicity and upper comonotonicity by (tail) convex order. *Insurance: Mathematics and Economics*, 53(2):334–342.

Cheung, K. C. and Lo, A. (2013b). General lower bounds on convex functionals of aggregate sums. *Insurance: Mathematics and Economics*, 53(3):884–896.

Cheung, K. C. and Lo, A. (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance: Mathematics and Economics*, 55:180–190.

Cossette, H., Côté, M.-P., Mailhot, M., and Marceau, E. (2014). A note on the computation of sharp numerical bounds for the distribution of the sum, product or ratio of dependent risks. *Journal of Multivariate Analysis*, 130:1–20.

Cossette, H., Marceau, E., Nguyen, Q. H., and Robert, C. Y. (2019). Tail approximations for sums of dependent regularly varying random variables under Archimedean copula models. *Methodology and Computing in Applied Probability*, 21(2):461–490.

Dacorogna, M. M., Elbahtouri, L., and Kratz, M. (2016). Explicit diversification benefit for dependent risks. *SCOR papers*, (38).

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons.

Dhaene, J. and Denuit, M. (1999). The safest dependence structure among risks. *Insurance: Mathematics and Economics*, 25(1):11–21.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, 31(1):3–33.

Dhaene, J. and Goovaerts, M. J. (1996). Dependency of risks and stop-loss order. *ASTIN Bulletin: The Journal of the IAA*, 26(2):201–212.

Dhaene, J., Linders, D., Schoutens, W., and Vyncke, D. (2012). The Herd Behavior Index: A new measure for the implied degree of co-movement in stock markets. *Insurance: Mathematics and economics*, 50(3):357–370.

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., and Vyncke, D. (2006). Risk measures and comonotonicity: A review. *Stochastic models*, 22(4):573–606.

Dhaene, J., Wang, S., Young, V. R., and Goovaerts, M. (2000). Comonotonicity and maximal stop-loss premiums. *Bulletin of the Swiss Association of Actuaries*, 2:99–113.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling Extremal Events: for Insurance and Finance*, volume 33. Springer Science & Business Media.

Embrechts, P., Lambrigger, D. D., and Wüthrich, M. V. (2009a). Multivariate extremes and the aggregation of dependent risks: Examples and counter-examples. *Extremes*, 12(2):107–127.

Embrechts, P., Nešlehová, J., and Wüthrich, M. V. (2009b). Additivity properties for value-at-risk under Archimedean dependence and heavy-tailedness. *Insurance: Mathematics and Economics*, 44(2):164–169.

Gaffke, N. and Rüscherndorf, L. (1981). On a class of extremal problems in statistics. *Mathematische Operationsforschung und Statistik. Series Optimization*, 12(1):123–135.

Hobson, D., Laurence, P., and Wang, T.-H. (2005a). Static-arbitrage optimal subreplicating strategies for sasket options. *Insurance: Mathematics and Economics*, 37(3):553–572.

Hobson, D., Laurence, P., and Wang, T.-H. (2005b). Static-arbitrage upper bounds for the prices of basket options. *Quantitative finance*, 5(4):329–342.

Inui, K. and Kijima, M. (2005). On the significance of expected shortfall as a coherent risk measure. *Journal of Banking & Finance*, 29(4):853–864.

Jeon, J., Kochar, S., and Park, C. G. (2006). Dispersive ordering – some applications and examples. *Statistical Papers*, 47(2):227–247.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

Laurence, P. and Wang, T.-H. (2005). Sharp upper and lower bounds for basket options. *Applied Mathematical Finance*, 12(3):253–282.

Lee, W. and Ahn, J. Y. (2014). On the multidimensional extension of countermonotonicity and its applications. *Insurance: Mathematics and Economics*, 56:68–79.

Lee, W., Cheung, K. C., and Ahn, J. Y. (2017). Multivariate countermonotonicity and the minimal copulas. *Journal of Computational and Applied Mathematics*, 317:589–602.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*, volume 389. Wiley New York.

Puccetti, G., Wang, B., and Wang, R. (2013). Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. *Insurance: Mathematics and Economics*, 53(3):821–828.

Puccetti, G., Wang, R., et al. (2015). Extremal dependence concepts. *Statistical Science*, 30(4):485–517.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Science & Business Media.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *The Annals of Probability*, 8(4):814–827.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, 41(3):808–826.

Wang, R., Peng, L., and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, 17(2):395–417.

Chapter 3

Hierarchical copulas with Archimedean blocks and asymmetric between-block pairs

Résumé

Une nouvelle classe de copules hiérarchiques est introduite en se basant sur des fonctions de survie conjointes de distributions de mélanges exponentiels multivariés. L'élément clé de cette construction est le vecteur de mélange aléatoire défini par des convolutions de variables associées à un subordonné de Lévy. On obient des copules hiérarchiques avec des copules intra-bloc Archimidiènnes et des paires de copules asymétriques entre blocs. Pour le cas particulier des arbres à deux niveaux, les propriétés de dépendance des paires sont étudiées, et une procédure d'estimation complète de la structure et des paramètres est proposée. L'efficacité de la procédure est illustrée à travers trois exemples de données simulées ainsi qu'une étude avec deux jeux de données réels.

Abstract

A new class of hierarchical copulas is introduced based on joint survival functions of multivariate exponential mixture distributions. The key element of this construction is the mixing random vector defined by convolutions associated to a Lévy subordinator, and leading to hierarchical copulas with Archimedean within-block copulas and asymmetric between-block pair-copulas. For the specific case of two-level trees, dependence properties of pairs are investigated, and a full estimation procedure is proposed for the tree structure and parameters of the hierarchical copulas. The efficiency of the procedure is illustrated through three simulation examples and a study with two real datasets.

Keywords: Hierarchical Copulas, Asymmetric Pair-Copulas, Archimedean Copulas, Composite Likelihood Estimation, Block-Exchangeability, Partitioning Around Medoids.

3.1 Introduction

Introduced by Sklar (1959), copulas are cumulative distribution functions (cdf) for vectors of uniformly distributed random variables (rvs) on the interval $[0,1]$. Their main advantage consists in summarizing the dependence structure between several risk factors, separately from their marginal distributions. This makes them popular tools in notably financial and insurance risk management.

Archimedean copulas are an important class of copulas, with a d-dimensional expression given by

$$C(u_1,\dots,u_d) = \psi\left(\psi^{-1}(u_1)+\cdots+\psi^{-1}(u_d)\right), \tag{3.1}$$

where $(u_1,\dots,u_d)\in[0,1]^d$ and $\psi:[0,\infty)\to[0,1]$ is a continuous and non-increasing function called the *generator* of the copula. It satisfies $\psi(0)=1$ and $\lim_{t\to\infty}\psi(t)=0$, and it is strictly decreasing on $[0,\inf\{t:\psi(t)=0\}]$. The function $\psi^{-1}:[0,1]\to[0,\infty)$ is the inverse of ψ (with $\psi^{-1}(0)=\inf\{t:\psi(t)=0\}$). Moreover, (3.1) defines an Archimedean copula in any dimension $d\geq 2$ if and only if ψ is completely monotone, see Kimberling (1974). From Bernstein's Theorem (see Feller (1971)), ψ has to be the Laplace-Stieltjes transform (LST) of a strictly positive rv. This type of copulas has the advantages of having explicit expressions, as well as capturing various dependence structures and being easily used in high dimensions (see, e.g., Schweizer (1991), Genest and Mackay (1986), Joe (2014) and McNeil et al. (2015)). However, Archimedean copulas are exchangeable copulas, i.e. for an arbitrary permutation σ of $\{1,\dots,d\}$, we have that $C(u_1,\dots,u_d)=C(u_{\sigma_1},\dots,u_{\sigma_d})$. In practice, this can be a major drawback.

To allow non-exchangeability, hierarchical Archimedean copulas (HAC) have been developed. Joe (1997) discusses the first approach to construct such copulas, leading to the well-known nested Archimedean copulas, obtained by nesting Archimedean copulas into each other. Further research has been done on this construction technique, see, e.g., Górecki et al. (2016) and McNeil (2008). As pointed out notably by Hofert (2010), the verification of the sufficient nesting condition within these copulas is not always an easy task when generators belong to different Archimedean families hence leading to less flexible structures. Note that the authors in Section 3 of Górecki et al. (2017) discuss this aspect in detail and provide sufficient nesting conditions. Recent developments on other construction techniques have successfully suppressed the need for this condition to be verified. For example, Hering et al. (2010) proposed to construct HACs using Lévy subordinators, while Cossette et al. (2017) used multivariate compound distributions. In both these papers, the construction approach relies on a multivariate exponential mixture distribution. The mixing random vector is defined using Lévy subordinators with a common random time in Hering et al. (2010) and random sums with common counting rvs in Cossette et al. (2017). A detailed comparison of these two construction approaches is discussed in Section 4.2 of Cossette et al. (2017). Although this

family of copulas offers very rich and complex structures of dependence, it can typically not be used for data as represented in Figure 3.1 because some pairs are symmetric (exchangeable) and others are not. By construction, bivariate margins of HACs are necessarily symmetric.

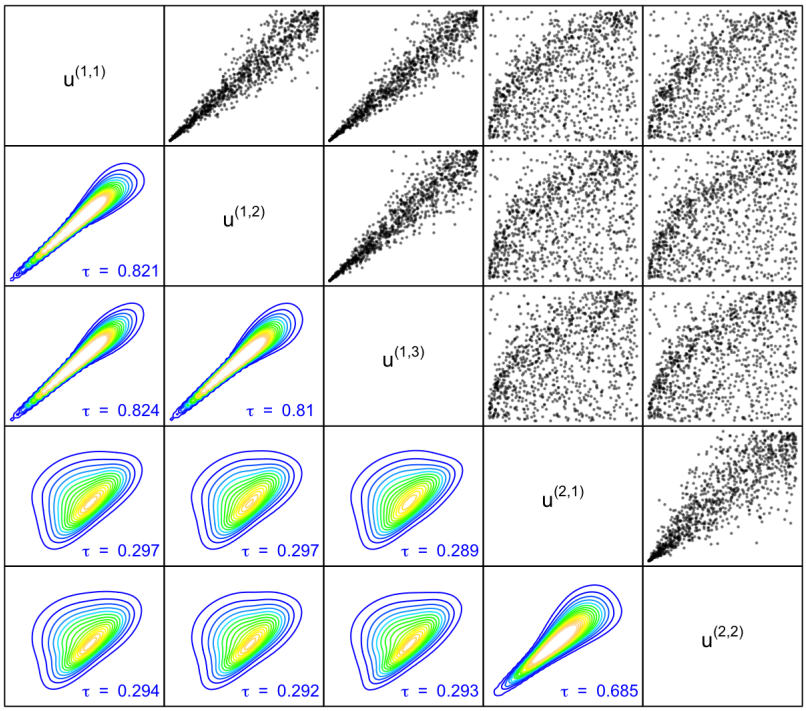


Figure 3.1 – Upper triangular matrix: scatterplots of exchangeable and non-exchangeable pairs, and lower triangular matrix: associated empirical density contour plots of bivariate normalized copula

Other extensions of Archimedean copulas have been proposed in the literature in order to avoid bivariate exchangeability. McNeil and Nešlehová (2010) introduced Liouville copulas as survival copulas of multivariate Liouville distributions, which induce an asymmetric dependence structure in general. Extremal properties of Liouville copulas were investigated in Hua (2016). In Belzile and Nešlehová (2017) the limiting extreme value copulas of Liouville copulas and of their survival counterparts were derived.

Characterized by their Pickands dependence function, extreme value copulas aim to represent the limiting dependence structure of componentwise block maxima under appropriate normalizations. An extreme value copula is asymmetric under the condition of an asymmetric Pickands dependence function. Introduced by Capéraà et al. (2000), bivariate Archimax copulas may also create asymmetric dependence. They include both the extreme value copulas and the Archimedean copulas. A multivariate extension of Archimax copulas was proposed by

Mesiar and Jagr (2013) and investigated in depth by Charpentier et al. (2014). Finally, Hofert et al. (2018) proposed a hierarchical generalization of Archimax copulas based on hierarchical stable tail dependence functions or hierarchical frailties with a discussion on possible flexible extensions to nested Archimax copulas.

In this paper, we introduce a hierarchical copula that allows to have a block exchangeability structure, in which within-block copulas are Archimedean and between-block copulas are non-exchangeable (in particular pairs that belong to different blocks are asymmetric). Inspired by Cossette et al. (2017), our hierarchical copulas are constructed from the survival function of a multivariate mixed exponential distribution through a convolution technique instead of compounding. In 1995, Khoudraji (1995) proposed a popular mechanism for generating bivariate asymmetric copulas through products of copulas (not necessarily Archimedean). We refer to Genest and Nešlehová (2013) for a review on the construction of asymmetric bivariate dependence structures (measures of asymmetry and tests of symmetry are also presented and discussed). A multivariate extension of Khoudraji (1995)'s multiplicative procedure was proposed in Liebscher (2008). We observe such types of products of copulas with our model for between-block copulas.

The remainder of the paper is organized as follows. In Section 3.2, we present the construction approach of the proposed family of hierarchical copulas. We introduce notations and provide general families of exchangeable and non-exchangeable copulas of blocks of leaves within the tree structure of our proposed family. In Section 3.3, we investigate the special case of two-level hierarchical copulas. We first present and discuss several specific families of copulas. Then we provide sufficient conditions for stochastic comparison of pair-copulas. In Section 3.4, we propose a full estimation procedure that is able to determinate the tree structure of a two-level hierarchical copula and to produce estimates of the parameters through a bottom-up composite likelihood approach. The full estimation efficiency is illustrated through three examples with simulated data and a study with two real datasets.

3.2 Construction of hierarchical copulas through convolutions

Referring to Marshall and Olkin (1988), multivariate distributions can be derived through mixture models, also known as frailty models. Using exponential mixtures, Archimedean copulas can easily be constructed and explicitly expressed in function of the LST of a strictly positive mixing rv. Inspired by this well-known approach, Cossette et al. (2017) proposed their HACs using multivariate compound distributions. In the following, we introduce a construction procedure to build hierarchical copulas, using a vector of mixing rvs, whose joint distribution is obtained through convolutions. We provide explicit expressions of the within-block Archimedean copulas and of the non-exchangeable between-block copulas. We also propose a simulation algorithm.

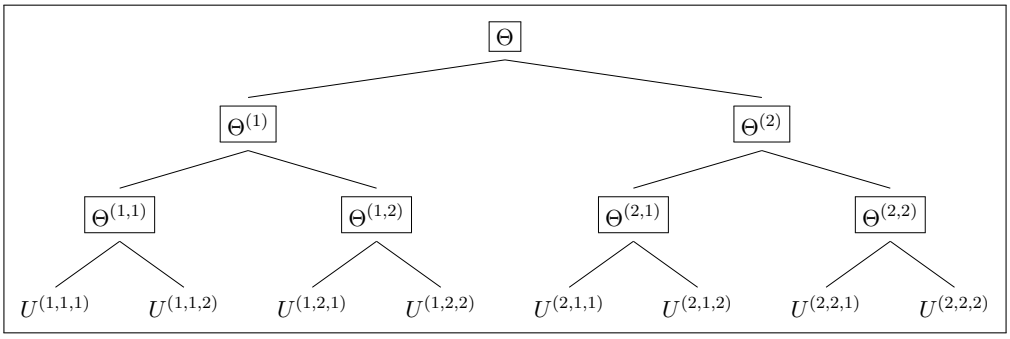


Figure 3.2 – Illustration of the tree structure discussed in Example 1 with $l = 3$ levels

3.2.1 Tree structure of mixing random variables

We first introduce notation to define the tree structure describing the joint distribution of the mixing random vector. This notation is mainly inspired by Cossette et al. (2019). For illustration purposes, we will use the tree construction example given in Figure 3.2.

The root of the tree structure will be denoted by the strictly positive rv Θ and will be assigned a level $r = 0$. The latter is the parent node of the rvs $\Theta^{(1)}, \Theta^{(2)}, \ldots$. These rvs are also the parents for nodes at level $r = 2$; for example, $\Theta^{(1,1)}, \Theta^{(1,2)}, \ldots$ are the children of $\Theta^{(1)}$. The hierarchical structure is then built by iterating until level $r = l - 1$, were $l \in \mathbb{N}_+$ is the depth of the tree. Level l is devoted to the vector of leaves corresponding to the components of the vector $\mathbf{U}$ with uniform marginals.

The path from the root node Θ (level 0) to a node at level r, for $r \in \{1, \ldots, l-1\}$, is defined by the vector $\mathbf{i}_r = (i_1, \ldots, i_r)$. For $j \in \{1, \ldots, r\}$, i_j represents the position of the node in the path at level j. For simplification purposes, we let $\Theta^{\mathbf{i}_0} = \Theta$. For $r \in \{0, \ldots, l-1\}$, the number of children directly under the parent node with path $\mathbf{i}_r$ is defined by $a_{\mathbf{i}_r}$.

Let us denote by $\mathbf{i}_{r|j}$ the vector of the first j components of a specific path $\mathbf{i}_r$. Given the hierarchical structure, the set of all possible paths at level r can be defined by

$$\mathcal{I}_r = \{\mathbf{i}_r = (i_1, \ldots, i_r) : i_j \in \{1, \ldots, a_{\mathbf{i}_{r|j-1}}\}, j = 1, \ldots, r\}.$$

For $r \in \{1, \ldots, l-1\}$, the vector of all nodes at level r is denoted by $\mathbf{\Theta}_r = \left(\Theta^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r\right)$. At the lowest level $r = l$, the vector of leaves is given by

$$\mathbf{U} = \left(\mathbf{U}^{\mathbf{i}_{l-1}} : \mathbf{i}_{l-1} \in \mathcal{I}_{l-1}\right),$$

where $\mathbf{U}^{\mathbf{i}_{l-1}} = \left(U^{(\mathbf{i}_{l-1}, j)} : j \in \{1, \ldots, a_{\mathbf{i}_{l-1}}\}\right)$, for $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$, denotes the sub-vector of leaves $\mathbf{U}$ under the parent node $\Theta^{\mathbf{i}_{l-1}}$. The total number of leaves (the dimension of $\mathbf{U}$) is given by

$$n = \sum_{\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}} a_{\mathbf{i}_{l-1}}.$$

Example 1 *Figure 3.2 corresponds to a three-level tree structure. The set of possible paths at level 1 is given by* $\mathcal{I}_1 = \{(1),(2)\}$*, with both nodes having 2 children, i.e.* $a_{(i)} = 2$*, for* $i \in \{1,2\}$*. At level 2, the set of possible paths is given by* $\mathcal{I}_2 = \{(1,1),(1,2),(2,1),(2,2)\}$*. The set* $\mathcal{I}_2$ *hence regroups all paths* $\mathbf{i}_2 = (i_1, i_2)$ *where each element* i_j *of this vector gives the position in the path at level* $j \in \{1,2\}$*. For* $\mathbf{i}_2 = (2,1)$ *for example,* $\mathbf{i}_1 = \mathbf{i}_{2|1} = (2)$ *is the first element of* $\mathbf{i}_2$ *and indicates that at level* 1*, we are under* $\Theta^{(2)}$*. The vectors of all nodes at levels 1 and 2 correspond respectively to* $\mathbf{\Theta}_1 = \left(\Theta^{(1)}, \Theta^{(2)}\right)$ *and* $\mathbf{\Theta}_2 = \left(\Theta^{(1,1)}, \Theta^{(1,2)}, \Theta^{(2,1)}, \Theta^{(2,2)}\right)$*. For all four nodes in* $\mathbf{\Theta_2}$*, we have* $a_{\mathbf{i}_2} = 2$*, for* $\mathbf{i}_2 \in \mathcal{I}_2$*. Under node* $\Theta^{(2,1)}$*, we have the vector of leaves* $\mathbf{U}^{(2,1)} = \left(U^{(2,1,1)}, U^{(2,1,2)}\right)$*.*

Let $\mathbf{\Theta} = \left(\Theta^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, \ldots, l-1\right)$. We now explain how the multi-level construction of $\mathbf{\Theta}$ is built through a tree structure and Lévy subordinators. We first associate to the root node Θ a vector of independent rvs $\mathbf{N} = \left(N^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, \ldots, l-1\right)$, that is also independent of Θ, and then define recursively

$$\Theta^{\mathbf{i}_r} = \Theta^{\mathbf{i}_{r|r-1}} + N^{\mathbf{i}_r}, \qquad \mathbf{i}_r \in \mathcal{I}_r, r = 1, \ldots, l-1. \tag{3.2}$$

Note that the definition of $\Theta^{\mathbf{i}_r}$ in (3.2) can also be written as

$$\Theta^{\mathbf{i}_r} = \Theta + \sum_{k=1}^{r} N^{\mathbf{i}_{r|k}}, \qquad \mathbf{i}_r \in \mathcal{I}_r, r = 1, \ldots, l-1. \tag{3.3}$$

Let us now characterize the distributions of the components of $\mathbf{N}$. A Lévy subordinator is a one-dimensional Lévy process, $(S_t)_{t\geq 0}$, that is non-decreasing almost surely (a.s.), and satisfies

$$E\left[\exp(-uS_t)\right] = \exp\left(-t\varphi(u)\right), \qquad t \geq 0,$$

where $\varphi(u)$ is called the Laplace exponent and has the following form:

$$\varphi(u) = bu + \int_0^\infty \left(1 - e^{-yu}\right) \nu\left(dy\right),$$

with $b > 0$, ν a measure on $(0,\infty)$ such that $\int_0^\infty (1 \wedge y)\nu\left(dy\right) < \infty$, where $a \wedge b = \min\{a,b\}$ (see, for instance, Applebaum (2009)). The pair (b,ν) is referred to as the characteristics of the subordinator $(S_t)_{t\geq 0}$ and are called the drift and the Lévy measure respectively.

Let $\boldsymbol{\alpha} = (\alpha_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, \ldots, l-1)$ be a vector of strictly positive constants. For $\mathbf{i}_r \in \mathcal{I}_r, r = 1, \ldots, l-1$, we assume that the LST of $N^{\mathbf{i}_r}$, $\mathcal{L}_{N^{\mathbf{i}_r}}(t) = E\left[\exp\left(-tN^{\mathbf{i}_r}\right)\right]$, satisfies

$$\mathcal{L}_{N^{\mathbf{i}_r}}(t) = \exp\left(-\alpha_{\mathbf{i}_r}\varphi(t)\right), \quad t \geq 0. \tag{3.4}$$

The characteristics of the Lévy subordinator are chosen such that the rvs $N^{\mathbf{i}_r}$ are strictly positive, for $\mathbf{i}_r \in \mathcal{I}_r, r = 1, \ldots, l-1$. We also assume that

$$\mathcal{L}_{\Theta}(t) = \exp\left(-\alpha_{\mathbf{i}_0}\varphi(t)\right), \quad t \geq 0.$$

Table 3.1 – Examples of distributions whose LST, $\psi(t;\delta)$, satisfy (3.4). These distributions may be characterized by additional parameters

#	Distribution	Parameters	$\psi(t;\delta)$
1	$\tilde{S}\left(\lambda, 1, (\cos(\lambda\pi/2)\,\delta)^{1/\lambda}, \delta\mathbb{1}_{\{\lambda=1\}}, h\mathbb{1}_{\{\lambda\neq 1\}}; 1\right)$	$\delta>0, \lambda\in(0,1], h\geq 0$	$\mathrm{e}^{-\delta\left((h+t)^{\lambda}-h^{\lambda}\right)}$
2	Gamma $\Gamma(\delta, 1)$	$\delta>0$	$(1+t)^{-\delta}$
3	Gamma stopped positive stable (δ,β)	$\delta>0, \beta\geq 1$	$\left(1+t^{\frac{1}{\beta}}\right)^{-\delta}$
4	Gamma stopped gamma (δ,β)	$\delta>0, \beta>0$	$(1+\beta\ln(1+t))^{-\delta}$
5	Gamma stopped Sibuya (δ,β)	$\delta>0, \beta\geq 1$	$\left(1-\ln\left(1-\left(1-\mathrm{e}^{-t}\right)^{\beta}\right)\right)^{-\delta}$
6	Shifted negative binomial (δ,δ,q)	$\delta>0, 0<q\leq 1$	$\left(\frac{q}{\mathrm{e}^{t}-(1-q)}\right)^{\delta}$
7	Shifted compound Poisson-exponential $(\delta, 1)$	$\delta>0$	$\mathrm{e}^{-\delta t\frac{2+t}{1+t}}$

We denote the LST of $N^{\mathbf{i}_r}$ by $\psi(t;\alpha_{\mathbf{i}_r})$ and thus $\mathcal{L}_{\Theta}(t)=\psi(t;\alpha_{\mathbf{i}_0})$.

Combining (3.3) and (3.4), we deduce that the LSTs of $\Theta^{\mathbf{i}_r}$ are given by

$$\begin{aligned}\mathcal{L}_{\Theta^{\mathbf{i}_r}}(t) &= \mathrm{E}\left[\exp\left(-t\left(\Theta+\sum_{k=1}^{r}N^{\mathbf{i}_{r|k}}\right)\right)\right]\\ &= \mathrm{E}\left[\exp\left(-t\Theta\right)\right]\times\prod_{k=1}^{r}\mathrm{E}\left[\exp\left(-tN^{\mathbf{i}_{r|k}}\right)\right],\end{aligned}$$

which becomes

$$\mathcal{L}_{\Theta^{\mathbf{i}_r}}(t)=\psi(t;\alpha_{\mathbf{i}_0})\times\prod_{k=1}^{r}\psi(t;\alpha_{\mathbf{i}_{r|k}})=\psi(t;\delta_{\mathbf{i}_r}),$$

where $\delta_{\mathbf{i}_r}=\alpha_{\mathbf{i}_0}+\sum_{k=1}^{r}\alpha_{\mathbf{i}_{r|k}}$, for $r=1,\ldots,l-1$. We let $\delta_{\mathbf{i}_0}=\alpha_{\mathbf{i}_0}$. Therefore, $\Theta^{\mathbf{i}_r}$, $r\in\{0,\ldots,l-1\}$, belongs to the same distribution family as the rvs of type N, with structural parameter $\delta_{\mathbf{i}_r}$.

Example 1 (Cont'd) *To illustrate the recursive definition of the nodes of* $\boldsymbol{\Theta}$*, let us consider nodes* $\Theta^{(1)}$ *and* $\Theta^{(2,1)}$ *at levels 1 and 2 respectively. We have* $\Theta^{(1)}=\Theta+N^{(1)}$ *and* $\Theta^{(2,1)}=\Theta^{(2)}+N^{(2,1)}$*, where* $\Theta^{(2)}=\Theta+N^{(2)}$*. Note, for example, that the rv* $N^{(1,2)}$ *has LST* $\psi\left(t;\alpha_{(1,2)}\right)$ *and* $\Theta^{(2,1)}$ *has LST* $\psi\left(t;\delta_{(2,1)}\right)$ *with* $\delta_{(2,1)}=\alpha+\alpha_{(2)}+\alpha_{(2,1)}$*.*

There exist several distributions whose LSTs satisfy (3.4). Some examples are given in Table 3.1. The first family (#1) is the exponential tilted stable distribution, also named tempered stable distribution, and denoted by

$$\tilde{S}\left(\lambda, 1, (\cos(\lambda\pi/2)\,\delta)^{1/\lambda}, \delta\mathbb{1}_{\{\lambda=1\}}, h\mathbb{1}_{\{\lambda\neq 1\}}; 1\right)$$

(see, e.g., Hofert (2011b) and Joe (2014)). This distribution can be efficiently sampled using, e.g., the function *retstable* in the `R` package *copula*. The second family (#2) is the well-known gamma distribution with constant scale parameter equal to 1 and denoted by $\Gamma(\delta, 1)$.

Referring to Appendix A in Joe (2014), we identify three two-parameter LST families satisfying (3.4), named gamma stopped positive stable, gamma stopped gamma and gamma stopped Sibuya (see Joe (2014) for more details and simulation algorithms for the three associated distributions). They correspond to LST families (#3), (#4) and (#5).

By shifting some distributions with a point mass at 0, we can finally obtain additional strictly positive distributions satisfying (3.4). The shifted negative binomial distribution (δ, δ, q) corresponds to a negative binomial distribution (δ, q) shifted by δ. The shifted compound Poisson-exponential $(\delta, 1)$ distribution is obtained from a compound Poisson distribution with parameter δ associated with exponential risks (with parameter β equal to 1) and then shifted by δ.

It is well-known (see, e.g., Theorem 4.5.1 in Nelsen (2006)) that, if $\psi(t)$ is a LST of a strictly positive rv then, for $\beta \geq 1$, $\psi_\beta(t) = \psi\left(t^{\frac{1}{\beta}}\right)$ is also a LST of a strictly positive rv. LST (#3) is derived from LST (#2) using this operation. Therefore, LST families (#1), (#4), (#5), (#6) and (#7) can be also widened by considering the LSTs given by

$$\psi_\beta(t;\delta) = \psi\left(t^{\frac{1}{\beta}};\delta\right) = \exp\left(-\delta\varphi_\beta(t)\right), \quad t \geq 0,$$

with $\varphi_\beta(t) = \varphi\left(t^{\frac{1}{\beta}}\right)$, instead of $\psi(t;\delta)$. The parametric LST families generated in this way are referred to as outer power LST families. They were investigated in the literature as generators of Archimedean copulas and HACs. See, e.g., Section 4.5.1 of Nelsen (2006), Hofert (2008), Hofert (2011a) and Górecki et al. (2020) for more details and sampling algorithms.

3.2.2 Hierarchy of Laplace-Stieltjes Transforms

We are interested in developing the expression of the multivariate LST of the random vector $\mathbf{\Theta}_r = \left(\Theta^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r\right)$, for $r \in \{1, \ldots, l-1\}$. To do so, we proceed recursively and use the stochastic representation of the components of $\mathbf{\Theta}_r$ in terms of the independent rvs of type N. To obtain generalized formulas, we will introduce additional notation when needed.

At level 1, the multivariate LST of $\mathbf{\Theta}_1 = \left(\Theta^{\mathbf{i}_1} : \mathbf{i}_1 \in \mathcal{I}_1\right)$ is given by

$$\mathcal{L}_{\mathbf{\Theta}_1}(\mathbf{t}_1) = \psi\left(\|\mathbf{t}_1\|; \alpha_{\mathbf{i}_0}\right) \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \psi\left(t^{\mathbf{i}_1}; \alpha_{\mathbf{i}_1}\right), \tag{3.5}$$

for the vector $\mathbf{t}_1 = \left(t^{\mathbf{i}_1} : \mathbf{i}_1 \in \mathcal{I}_1, t^{\mathbf{i}_1} \geq 0\right)$ and where $\|\mathbf{t}_1\| = \sum_{\mathbf{i}_1 \in \mathcal{I}_1} t^{\mathbf{i}_1}$. The detailed development of (3.5) is provided in Appendix A. Given the independence of the involved rvs, (3.5) can be written as a product of LSTs. Note that all components of $\mathbf{\Theta}_1$ are function of Θ, which explains the use of $\|\mathbf{t}_1\|$ in $\psi(\cdot\,; \alpha_{\mathbf{i}_0})$.

To obtain the joint LST of $\mathbf{\Theta}_r$ for lower levels of the tree structure, we have to identify common rvs of type N for components of $\mathbf{\Theta}_r$ through all the considered paths. For a given path $\mathbf{i}_k \in \mathcal{I}_k$, with $k \in \{1, \ldots, r\}$ and $r \in \{1, \ldots, l-1\}$, let $\mathcal{A}_{r,\mathbf{i}_k} = \{\mathbf{i}_r : \mathbf{i}_{r|k} = \mathbf{i}_k, \mathbf{i}_r \in \mathcal{I}_r\}$ be the subset of

$\mathcal{I}_r$ having the first k components of their paths equal to $\mathbf{i}_k$. Let $\mathbf{t}_r = \left(t^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, t^{\mathbf{i}_r} \geq 0\right)$, for $r \in \{1, \ldots, l-1\}$. We define $\mathbf{t}_{r,\mathbf{i}_k} = \left(t^{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{A}_{r,\mathbf{i}_k}\right)$ as the sub-vector of $\mathbf{t}_r$ having a common path from the root Θ to the node $\Theta^{\mathbf{i}_k}$ at level k.

Given the notation just introduced, we can express the multivariate LST of $\mathbf{\Theta}_2$ as the product of the univariate LSTs of the involved rvs of type N applied to the appropriate components of the vector $\mathbf{t}_2$. Therefore, from (3.5), we have

$$\mathcal{L}_{\mathbf{\Theta}_2}(\mathbf{t}_2) = \psi\left(\|\mathbf{t}_2\|; \alpha_{\mathbf{i}_0}\right) \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \psi\left(\|\mathbf{t}_{2,\mathbf{i}_1}\|; \alpha_{\mathbf{i}_1}\right) \times \prod_{\mathbf{i}_2 \in \mathcal{I}_2} \psi\left(t^{\mathbf{i}_2}; \alpha_{\mathbf{i}_2}\right), \tag{3.6}$$

with $\|\mathbf{t}_{2,\mathbf{i}_1}\| = \sum_{\mathbf{i}_2 \in \mathcal{A}_{2,\mathbf{i}_1}} t^{\mathbf{i}_2}$, for $\mathbf{i}_1 \in \mathcal{I}_1$. The full development of (3.6) is provided in Appendix A.

In the same manner, we can deduce a generalized formula for the multivariate LST at level r, for $r \in \{1, \ldots, l-1\}$. Thus, we have

$$\begin{aligned} \mathcal{L}_{\mathbf{\Theta}_r}(\mathbf{t}_r) &= \mathcal{L}_{\Theta}(\|\mathbf{t}_r\|) \times \prod_{k=1}^{r} \prod_{\mathbf{i}_k \in \mathcal{I}_k} \mathcal{L}_{N^{\mathbf{i}_k}}(\|\mathbf{t}_{r,\mathbf{i}_k}\|) \\ &= \psi\left(\|\mathbf{t}_r\|; \alpha_{\mathbf{i}_0}\right) \times \prod_{k=1}^{r} \prod_{\mathbf{i}_k \in \mathcal{I}_k} \psi\left(\|\mathbf{t}_{r,\mathbf{i}_k}\|; \alpha_{\mathbf{i}_k}\right), \end{aligned} \tag{3.7}$$

with $\|\mathbf{t}_{r,\mathbf{i}_k}\| = \sum_{\mathbf{i}_r \in \mathcal{A}_{r,\mathbf{i}_k}} t^{\mathbf{i}_r}$. Note that $\mathcal{A}_{r,\mathbf{i}_r} = \{\mathbf{i}_r\}$ and $\|\mathbf{t}_{r,\mathbf{i}_r}\| = t^{\mathbf{i}_r}$.

Example 1 (Cont'd) *We consider the three-level hierarchical structure depicted in Figure 3.2. At level 1, we have* $\mathbf{\Theta}_1 = \left(\Theta^{(1)}, \Theta^{(2)}\right)$, *with*

$$\begin{aligned} \mathcal{L}_{\mathbf{\Theta}_1}(\mathbf{t}_1) &= \psi\left(\|\mathbf{t}_1\|; \alpha_{\mathbf{i}_0}\right) \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \psi\left(t^{\mathbf{i}_1}; \alpha_{\mathbf{i}_1}\right) \\ &= \psi\left(t^{(1)} + t^{(2)}; \alpha_{\mathbf{i}_0}\right) \times \psi\left(t^{(1)}; \alpha_{(1)}\right) \times \psi\left(t^{(2)}; \alpha_{(2)}\right), \end{aligned}$$

for $\mathbf{t}_1 = \left(t^{(1)}, t^{(2)}\right)$. *Level 2 has 4 nodes. Following all possible paths we can identify the sets* $\mathcal{A}_{2,\mathbf{i}_k}$, *for* $\mathbf{i}_k \in \mathcal{I}_k, k = 1, 2$. *For example,* $\mathcal{A}_{2,(1)} = \{(1,1),(1,2)\}$, $\mathcal{A}_{2,(2)} = \{(2,1),(2,2)\}$ *and* $\mathcal{A}_{2,(1,2)} = \{(1,2)\}$. *From (3.7), the multivariate LST of* $\mathbf{\Theta}_2$ *can be expressed, for* $\mathbf{t}_2 = \left(t^{(1,1)}, t^{(1,2)}, t^{(2,1)}, t^{(2,2)}\right)$, *as*

$$\begin{aligned} \mathcal{L}_{\mathbf{\Theta}_2}(\mathbf{t}_2) &= \psi\left(\|\mathbf{t}_2\|; \alpha_{\mathbf{i}_0}\right) \times \prod_{k=1}^{2} \prod_{\mathbf{i}_k \in \mathcal{I}_k} \psi\left(\|\mathbf{t}_{2,\mathbf{i}_k}\|; \alpha_{\mathbf{i}_k}\right) \\ &= \psi\left(t^{(1,1)} + t^{(1,2)} + t^{(2,1)} + t^{(2,2)}; \alpha_{\mathbf{i}_0}\right) \times \psi\left(t^{(1,1)} + t^{(1,2)}; \alpha_{(1)}\right) \\ &\quad \times \psi\left(t^{(2,1)} + t^{(2,2)}; \alpha_{(2)}\right) \times \psi\left(t^{(1,1)}; \alpha_{(1,1)}\right) \times \psi\left(t^{(1,2)}; \alpha_{(1,2)}\right) \times \psi\left(t^{(2,1)}; \alpha_{(2,1)}\right) \\ &\quad \times \psi\left(t^{(2,2)}; \alpha_{(2,2)}\right). \end{aligned}$$

3.2.3 The hierarchical copula

Let $\mathbf{Y} = \left(\mathbf{Y}^{\mathbf{i}_{l-1}}, \mathbf{i}_{l-1} \in \mathcal{I}_{l-1}\right)$ be a vector of n rvs, composed of $d = \sum_{\mathbf{i}_{l-2} \in \mathcal{I}_{l-2}} a_{\mathbf{i}_{l-2}}$ blocks, where $\mathbf{Y}^{\mathbf{i}_{l-1}} = \left(Y^{(\mathbf{i}_{l-1},j)}, j = 1, \ldots, a_{\mathbf{i}_{l-1}}\right)$, $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$. Given the strictly positive mixing random vector $\mathbf{\Theta}_{l-1}$, we assume that the components of $\mathbf{Y}$ are conditionally independent and the components of $\mathbf{Y}^{\mathbf{i}_{l-1}}$ follow a conditional distribution only influenced by $\Theta^{\mathbf{i}_{l-1}}$, i.e., $\left(Y^{(\mathbf{i}_{l-1},j)} \mid \mathbf{\Theta}_{l-1} = \theta_{l-1}\right) \sim \left(Y^{(\mathbf{i}_{l-1},j)} \mid \Theta^{\mathbf{i}_{l-1}} = \theta^{\mathbf{i}_{l-1}}\right)$ for $j = 1, \ldots, a_{\mathbf{i}_{l-1}}$, $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$. Following Marshall and Olkin (1988), we assume that

$$(Y^{(\mathbf{i}_{l-1},1)} \mid \Theta^{\mathbf{i}_{l-1}} = \theta^{\mathbf{i}_{l-1}}), \ldots, (Y^{\left(\mathbf{i}_{l-1},a_{\mathbf{i}_{l-1}}\right)} \mid \Theta^{\mathbf{i}_{l-1}} = \theta^{\mathbf{i}_{l-1}})$$

are exponentially distributed with parameter $\theta^{\mathbf{i}_{l-1}}$, for $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$.

Therefore, $\mathbf{Y}$ has a multivariate mixed exponential distribution with joint survival function written as

$$\overline{F}_{\mathbf{Y}}(\mathbf{y}) = \mathcal{L}_{\mathbf{\Theta}_{l-1}}\left(\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} y^{(\mathbf{i}_{l-1},j)} : \mathbf{i}_{l-1} \in \mathcal{I}_{l-1}\right)\right), \tag{3.8}$$

where $\mathbf{y} = \left(y^{(\mathbf{i}_{l-1},j)} : j = 1, \ldots, a_{\mathbf{i}_{l-1}}, \mathbf{i}_{l-1} \in \mathcal{I}_{l-1}\right)$. See Appendix A for the detailed development of (3.8). The sub-vector $\mathbf{Y}^{\mathbf{i}_{l-1}}$ has a multivariate survival function defined only in terms of the LST of $\Theta^{\mathbf{i}_{l-1}}$, for $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$, with

$$\overline{F}_{\mathbf{Y}^{\mathbf{i}_{l-1}}}\left(\mathbf{y}^{\mathbf{i}_{l-1}}\right) = \mathcal{L}_{\Theta^{\mathbf{i}_{l-1}}}\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} y^{(\mathbf{i}_{l-1},j)}\right) = \psi\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} y^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right), \tag{3.9}$$

where $\mathbf{y}^{\mathbf{i}_{l-1}} = \left(y^{(\mathbf{i}_{l-1},j)}, j = 1, \ldots, a_{\mathbf{i}_{l-1}}\right)$. The inverse of the univariate survival function $\overline{F}_{Y^{(\mathbf{i}_{l-1},j)}}$ can be written as $\overline{F}^{-1}_{Y^{(\mathbf{i}_{l-1},j)}}\left(u^{(\mathbf{i}_{l-1},j)}\right) = \mathcal{L}^{-1}_{\Theta^{\mathbf{i}_{l-1}}}\left(u^{(\mathbf{i}_{l-1},j)}\right) = \psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right)$, for $u^{(\mathbf{i}_{l-1},j)} \in [0,1]$, $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$ and $j \in \{1, \ldots, a_{\mathbf{i}_{l-1}}\}$.

Proposition 1 *Let* $\mathbf{U}$ *be an* n*-dimensional vector with uniform marginals and with joint cdf corresponding to the survival copula* $C_{\mathbf{U}}$ *associated with the survival function* $\overline{F}_{\mathbf{Y}}$ *of* $\mathbf{Y}$*. Then,* $C_{\mathbf{U}}$ *is an* n*-dimensional hierarchical copula with* d *blocks given by*

$$\begin{aligned} C_{\mathbf{U}}(\mathbf{u}) &= \psi\left(\sum_{\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}} \sum_{j=1}^{a_{\mathbf{i}_{l-1}}} \psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right); \alpha_{\mathbf{i}_0}\right) \\ &\quad \times \prod_{k=1}^{l-1} \prod_{\mathbf{i}_k \in \mathcal{I}_k} \psi\left(\sum_{\mathbf{i}_{l-1} \in \mathcal{A}_{l-1|\mathbf{i}_k}} \sum_{j=1}^{a_{\mathbf{i}_{l-1}}} \psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right); \alpha_{\mathbf{i}_k}\right), \end{aligned} \tag{3.10}$$

where $\mathbf{u} = \left(u^{(\mathbf{i}_{l-1},j)} : j = 1, \ldots, a_{\mathbf{i}_{l-1}}, \mathbf{i}_{l-1} \in \mathcal{I}_{l-1}\right) \in [0,1]^n$.

Proof. The expression of $C_{\mathbf{U}}$ in (3.10) is easily derived by using Sklar's theorem, (3.7) and the joint survival distribution function (3.8) since

$$\begin{aligned} C_{\mathbf{U}}(\mathbf{u}) &= \overline{F}_{\mathbf{Y}}\left(\left(\overline{F}^{-1}_{Y^{(\mathbf{i}_{l-1},j)}}\left(u^{(\mathbf{i}_{l-1},j)}\right) : j=1,\dots,a_{\mathbf{i}_{l-1}}, \mathbf{i}_{l-1}\in\mathcal{I}_{l-1}\right)\right) \\ &= \overline{F}_{\mathbf{Y}}\left(\left(\psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)};\delta_{\mathbf{i}_{l-1}}\right) : j=1,\dots,a_{\mathbf{i}_{l-1}}, \mathbf{i}_{l-1}\in\mathcal{I}_{l-1}\right)\right) \\ &= \psi\left(\sum_{\mathbf{i}_{l-1}\in\mathcal{I}_{l-1}}\sum_{j=1}^{a_{\mathbf{i}_{l-1}}}\psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)};\delta_{\mathbf{i}_{l-1}}\right);\alpha_{\mathbf{i}_0}\right) \\ &\quad\times\prod_{k=1}^{l-1}\prod_{\mathbf{i}_k\in\mathcal{I}_k}\psi\left(\sum_{\mathbf{i}_{l-1}\in\mathcal{A}_{l-1|\mathbf{i}_k}}\sum_{j=1}^{a_{\mathbf{i}_{l-1}}}\psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)};\delta_{\mathbf{i}_{l-1}}\right);\alpha_{\mathbf{i}_k}\right). \end{aligned}$$

■

It is worth mentioning that, although we also use Lévy subordinators in our construction principle, our hierarchical copulas differ from the Archimedean copulas in Hering et al. (2010). For example, in the latter reference, all pairs are joined by an Archimedean copula contrarily to our construction in which not all copulas are Archimedean. Our approach would coincide with Hering et al. (2010) if we had considered a two-level structure with a degenerate distribution for Θ.

To efficiently sample a realization of $\mathbf{U}$ with hierarchical copula $C_{\mathbf{U}}$ as multivariate cdf, we propose to use Algorithm 3 based on the stochastic representation (3.2) of nodes.

Algorithme 3 : Simulation procedure

Result : A realization of $\mathbf{U} = \left(\mathbf{U}^{\mathbf{i}_{l-1}} : \mathbf{i}_{l-1}\in\mathcal{I}_{l-1}\right)$

Sample the root Θ

for $r \leftarrow 1$ **to** $l-1$ **by** 1 **do**
  foreach *path* $\mathbf{i}_r \in \mathcal{I}_r$ **do**
    Sample $N^{\mathbf{i}_r}$;
    Compute $\Theta^{\mathbf{i}_r} \leftarrow \Theta^{\mathbf{i}_{r|r-1}} + N^{\mathbf{i}_r}$;
  end
end
foreach *path* $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$ **do**
  for $j \leftarrow 1$ **to** $a_{\mathbf{i}_{l-1}}$ **by** 1 **do**
    Sample $R^{(\mathbf{i}_{l-1},j)} \sim Exp(1)$;
    Compute $U^{(\mathbf{i}_{l-1},j)} = \psi\left(\frac{R^{(\mathbf{i}_{l-1},j)}}{\Theta^{\mathbf{i}_{l-1}}};\delta_{\mathbf{i}_{l-1}}\right)$;
  end
end

3.2.4 Within-block and between-block copulas

Given the proposed hierarchical construction, we examine thoroughly, the copulas associated to leaves within-blocks or between-blocks. Since there are a lot of between-block copulas, we focus on those that will be of interest for the estimation of the vector of parameters $\boldsymbol{\alpha} = (\alpha_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, \ldots, l-1)$ by composite likelihood (see Section 3.4.1).

For this purpose, we introduce the following notation:

$$\begin{aligned}
&\mathcal{D}_{\mathbf{i}_r} \\
&= \left\{(\mathbf{i}_r, i_{r+1}, \ldots, i_{l-1}, j) : i_{r+1} \in \{1, \ldots, a_{\mathbf{i}_r}\}, \ldots, i_{l-1} \in \{1, \ldots, a_{(\mathbf{i}_r, i_{r+1}, \ldots, i_{l-2})}\}, j \in \{1, \ldots, a_{\mathbf{i}_{l-1}}\}\right\} \\
&= \left\{(\mathbf{i}_{l-1}, j) : \mathbf{i}_{l-1|r} = \mathbf{i}_r, \mathbf{i}_{l-1} \in \mathcal{I}_{l-1}, j = 1, \ldots, a_{\mathbf{i}_{l-1}}\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\mathcal{S}_{\mathbf{i}_r} \\
&= \left\{ \left\{ \begin{pmatrix} \mathbf{i}_{l-1}^{(1)}, j^{(1)} \end{pmatrix}, \ldots, \begin{pmatrix} \mathbf{i}_{l-1}^{(a_{\mathbf{i}_r})}, j^{(a_{\mathbf{i}_r})} \end{pmatrix} \right\} : \begin{pmatrix} \mathbf{i}_{l-1}^{(m)}, j^{(m)} \end{pmatrix} \in \mathcal{D}_{\mathbf{i}_r}, m \in \{1, \ldots, a_{\mathbf{i}_r}\}, \mathbf{i}_{l-1|r+1}^{(1)} < \ldots < \mathbf{i}_{l-1|r+1}^{(a_{\mathbf{i}_r})} \right\}.
\end{aligned}$$

$D_{\mathbf{i}_r}$ represents the set of paths of leaves under $\Theta^{\mathbf{i}_r}$, and $\mathcal{S}_{\mathbf{i}_r}$ is the subset of $a_{\mathbf{i}_r}$ paths belonging to $D_{\mathbf{i}_r}$, such that each component within a subset has a different node $\Theta^{\mathbf{i}_{r+1}}$. In the following, $\mathcal{S}_{\mathbf{i}_r}$ will be used to assess the between-block copula of leaves $\mathbf{U}^{\{\sigma\}}$, $\sigma \in \mathcal{S}_{\mathbf{i}_r}$.

Example 1 (Cont'd) *Let us consider the tree in Figure 3.2 and its sets of paths $\mathcal{D}_{(1)}$ and $\mathcal{S}_{(1)}$ specifically. We have $\mathcal{D}_{(1)} = \{(1,1,1), (1,1,2), (1,2,1), (1,2,2)\}$ which consists of all possible paths from node $\Theta^{(1)}$ and*

$$\mathcal{S}_{(1)} = \{\{(1,1,1), (1,2,1)\}, \{(1,1,1), (1,2,2)\}, \{(1,1,2), (1,2,1)\}, \{(1,1,2), (1,2,2)\}\}$$

a subset of 2 paths of $\mathcal{D}_{(1)}$ with each component within a subset having node either $\Theta^{(1,1)}$ or $\Theta^{(1,2)}$.

For $\sigma = \{\sigma^{(m)} : m \in \{1, \ldots, a_{\mathbf{i}_r}\}\} \in \mathcal{S}_{\mathbf{i}_r}$, where $\sigma^{(m)} = \begin{pmatrix} \mathbf{i}_{l-1}^{(m)}, j^{(m)} \end{pmatrix}$, let $\mathbf{U}^{\{\sigma\}} = (U^{\sigma^{(m)}} : m \in \{1, \ldots, a_{\mathbf{i}_r}\})$ be the associated sub-vector of $\mathbf{U}$.

Proposition 2 ***Families of within-block and between-block copulas.***

(i) For block $\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}$, the joint cdf of $\mathbf{U}^{\mathbf{i}_{l-1}}$ is the Archimedean copula with generator $\psi\left(\cdot\,; \delta_{\mathbf{i}_{l-1}}\right)$ given by

$$C_{\mathbf{U}^{\mathbf{i}_{l-1}}}\left(\mathbf{u}^{\mathbf{i}_{l-1}}\right) = \psi\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} \psi^{-1}\left(u^{(\mathbf{i}_{l-1}, j)}; \delta_{\mathbf{i}_{l-1}}\right); \delta_{\mathbf{i}_{l-1}}\right), \tag{3.11}$$

with $\mathbf{u}^{\mathbf{i}_{l-1}} = \left(u^{(\mathbf{i}_{l-1}, j)}, j = 1, \ldots, a_{\mathbf{i}_{l-1}}\right)$.

(ii) For $r \in \{0, \ldots, l-2\}$, $\mathbf{i}_r \in \mathcal{I}_r$ and $\sigma \in \mathcal{S}_{\mathbf{i}_r}$, the joint cdf of $\mathbf{U}^{\{\sigma\}}$ is given by

$$\begin{aligned}
&C_{\mathbf{U}^{\{\sigma\}}}\left(\mathbf{u}^{\{\sigma\}}\right) \\
&= \psi\left(\sum_{m=1}^{a_{\mathbf{i}_r}} \psi^{-1}\left(u^{\sigma^{(m)}}; \delta_{\mathbf{i}_{l-1}^{(m)}}\right); \delta_{\mathbf{i}_r}\right) \times \prod_{m=1}^{a_{\mathbf{i}_r}} \psi\left(\psi^{-1}\left(u^{\sigma^{(m)}}; \delta_{\mathbf{i}_{l-1}^{(m)}}\right); \delta_{\mathbf{i}_{l-1}^{(m)}} - \delta_{\mathbf{i}_r}\right).
\end{aligned} \tag{3.12}$$

Proof. See Appendix B. ■

For $\sigma \in \mathcal{S}_{\mathbf{i}_r}$, $C_{\mathbf{U}^{\{\sigma\}}}$ is an $a_{\mathbf{i}_r}$-dimensional hierarchical copula with parameters $(\delta_{\mathbf{i}_r}, \delta_{\mathbf{i}_{l-1}^{(1)}}, \ldots, \delta_{\mathbf{i}_{l-1}^{(a_{\mathbf{i}_r})}})$ depending on the descendent set of $\Theta^{\mathbf{i}_r}$ involved given σ, i.e. $\left\{\Theta^{\mathbf{i}_{l-1}^{(1)}}, \ldots, \Theta^{\mathbf{i}_{l-1}^{(a_{\mathbf{i}_r})}}\right\}$.

Example 1 (Cont'd) *Consider the three-level tree structure shown in Figure 3.2. At level $r = 0$, there are 4 possible asymmetric pair-copulas defined with Θ, depending on its considered descendent set. Hence, for $u^{(1,k^{(1)},l^{(1)})}, u^{(2,k^{(2)},l^{(2)})} \in [0,1]$, $k^{(1)}, k^{(2)}, l^{(1)}, l^{(2)} \in \{1,2\}$, we have*

$$\begin{aligned} & C_{\left(U^{(1,k^{(1)},l^{(1)})}, U^{(2,k^{(2)},l^{(2)})}\right)}\left(u^{(1,k^{(1)},l^{(1)})}, u^{(2,k^{(2)},l^{(2)})}\right) \\ = \; & \psi\left(\sum_{m=1}^{2} \psi^{-1}\left(u^{(m,k^{(m)},l^{(m)})}; \delta_{\mathbf{i}_2^{(m)}}\right); \delta_{\mathbf{i}_0}\right) \times \prod_{m=1}^{2} \psi\left(\psi^{-1}\left(u^{(m,k^{(m)},l^{(m)})}; \delta_{\mathbf{i}_2^{(m)}}\right); \delta_{\mathbf{i}_2^{(m)}} - \delta_{\mathbf{i}_0}\right). \end{aligned}$$

For example, for $k^{(1)} = k^{(2)} = 1$, $l^{(1)} = 1$ and $l^{(2)} = 2$, we have

$$\begin{aligned} C_{\left(U^{(1,1,1)}, U^{(2,1,2)}\right)}(u^{(1,1,1)}, u^{(2,1,2)}) = \; & \psi\left(\psi^{-1}\left(u^{(1,1,1)}; \delta_{(1,1)}\right) + \psi^{-1}\left(u^{(2,1,2)}; \delta_{(2,1)}\right); \delta_{(0)}\right) \\ & \times \psi\left(\psi^{-1}\left(u^{(1,1,1)}; \delta_{(1,1)}\right); \delta_{(1,1)} - \delta_{(0)}\right) \\ & \times \psi\left(\psi^{-1}\left(u^{(2,1,2)}; \delta_{(2,1)}\right); \delta_{(2,1)} - \delta_{(0)}\right), \end{aligned}$$

and for $k^{(1)} = 2$, $k^{(2)} = 1$, $l^{(1)} = 1$ and $l^{(2)} = 2$, we have

$$\begin{aligned} C_{\left(U^{(1,2,1)}, U^{(2,1,2)}\right)}(u^{(1,2,1)}, u^{(2,1,2)}) = \; & \psi\left(\psi^{-1}\left(u^{(1,2,1)}; \delta_{(1,2)}\right) + \psi^{-1}\left(u^{(2,1,2)}; \delta_{(2,1)}\right); \delta_{(0)}\right) \\ & \times \psi\left(\psi^{-1}\left(u^{(1,2,1)}; \delta_{(1,2)}\right); \delta_{(1,2)} - \delta_{(0)}\right) \\ & \times \psi\left(\psi^{-1}\left(u^{(2,1,2)}; \delta_{(2,1)}\right); \delta_{(2,1)} - \delta_{(0)}\right). \end{aligned}$$

At level $r = 1$, the asymmetric pair-copula associated to node $\Theta^{(1)}$ is given by

$$\begin{aligned} & C_{\left(U^{(1,1,l^{(1)})}, U^{(1,2,l^{(2)})}\right)}\left(u^{(1,1,l^{(1)})}, u^{(1,2,l^{(2)})}\right) \\ = \; & \psi\left(\psi^{-1}\left(u^{(1,1,l^{(1)})}; \delta_{(1,1)}\right) + \psi^{-1}\left(u^{(1,2,l^{(2)})}; \delta_{(1,2)}\right); \delta_{(1)}\right) \\ & \times \psi\left(\psi^{-1}\left(u^{(1,1,l^{(1)})}; \delta_{(1,1)}\right); \delta_{(1,1)} - \delta_{(1)}\right) \times \psi\left(\psi^{-1}\left(u^{(1,2,l^{(2)})}; \delta_{(1,2)}\right); \delta_{(1,2)} - \delta_{(1)}\right), \end{aligned}$$

with $u^{(1,1,l^{(1)})}, u^{(1,2,l^{(2)})} \in [0,1]$, and $l^{(1)}, l^{(2)} \in \{1,2\}$.

3.3 Two-level hierarchical copulas: families of copulas and stochastic comparisons

In this section, we propose to illustrate some examples of within-block and between-block pair-copulas within the two-level hierarchical copulas constructed through the approach proposed in Section 3.2. Then, we investigate the dependence properties of these pair-copulas. Note that such a two-level structure is used in the studies of Section 3.4 and illustrated notably in Figures 3.6 and 3.8.

First note that, based on (3.10), (3.11) and (3.12), the two-level hierarchical copula, the within-block copulas and the between-block copulas are, respectively, given by

$$\begin{aligned} C_{\mathbf{U}}\left(\mathbf{u}\right) &= \psi\left(\sum_{\mathbf{i}_1\in\mathcal{I}_1}\sum_{j=1}^{a_{\mathbf{i}_1}}\psi^{-1}\left(u^{(\mathbf{i}_1,j)};\delta_{\mathbf{i}_1}\right);\delta_{\mathbf{i}_0}\right) \\ &\quad\times\prod_{\mathbf{i}_1\in\mathcal{I}_1}\psi\left(\sum_{j=1}^{a_{\mathbf{i}_1}}\psi^{-1}\left(u^{(\mathbf{i}_1,j)};\delta_{\mathbf{i}_1}\right);\alpha_{\mathbf{i}_1}\right), \qquad (3.13) \\ C_{\mathbf{U}^{\mathbf{i}_1}}\left(\mathbf{u}^{\mathbf{i}_1}\right) &= \psi\left(\sum_{j=1}^{a_{\mathbf{i}_1}}\psi^{-1}\left(u^{(\mathbf{i}_1,j)};\delta_{\mathbf{i}_1}\right);\delta_{\mathbf{i}_1}\right) \qquad (3.14) \end{aligned}$$

and

$$C_{\mathbf{U}^{\{\sigma\}}}\left(\mathbf{u}^{\{\sigma\}}\right) = \psi\left(\sum_{m=1}^{a_{\mathbf{i}_0}}\psi^{-1}\left(u^{\sigma^{(m)}};\delta_{\mathbf{i}_1^{(m)}}\right);\delta_{\mathbf{i}_0}\right)\times\prod_{m=1}^{a_{\mathbf{i}_0}}\left(u^{\sigma^{(m)}}\right)^{1-\frac{\delta_{\mathbf{i}_0}}{\delta_{\mathbf{i}_1^{(m)}}}}, \qquad (3.15)$$

for $\mathbf{i}_1\in\mathcal{I}_1$, and $\sigma\in\mathcal{S}_{\mathbf{i}_0}$. We deduce that the within-block pair-copulas are Archimedean copulas with generator $\psi\left(\cdot\,;\delta_{\mathbf{i}_1}\right)$, i.e.

$$C_{\left(U^{(\mathbf{i}_1,j)},U^{(\mathbf{i}_1,j')}\right)} = \psi\left(\psi^{-1}\left(u^{(\mathbf{i}_1,j)};\delta_{\mathbf{i}_1}\right)+\psi^{-1}\left(u^{(\mathbf{i}_1,j')};\delta_{\mathbf{i}_1}\right);\delta_{\mathbf{i}_1}\right), \qquad (3.16)$$

for $\mathbf{i}_1\in\mathcal{I}_1, j, j'\in\{1,\ldots,a_{\mathbf{i}_1}\}, j\neq j'$ and that the between-block pair-copulas are asymmetric copulas given by

$$C_{\left(U^{(\mathbf{i}_1,j)},U^{(\mathbf{i}'_1,j')}\right)} = \psi\left(\psi^{-1}\left(u^{(\mathbf{i}_1,j)};\delta_{\mathbf{i}_1}\right)+\psi^{-1}\left(u^{(\mathbf{i}'_1,j')};\delta_{\mathbf{i}'_1}\right);\delta_{\mathbf{i}_0}\right)\times\left(u^{(\mathbf{i}_1,j)}\right)^{1-\frac{\delta_{\mathbf{i}_0}}{\delta_{\mathbf{i}_1}}}\times\left(u^{(\mathbf{i}'_1,j')}\right)^{1-\frac{\delta_{\mathbf{i}_0}}{\delta_{\mathbf{i}'_1}}}, \qquad (3.17)$$

for $\mathbf{i}_1,\mathbf{i}'_1\in\mathcal{I}_1, \mathbf{i}_1\neq\mathbf{i}'_1, j\in\{1,\ldots,a_{\mathbf{i}_1}\}, j'\in\{1,\ldots,a_{\mathbf{i}'_1}\}$.

3.3.1 Families of within-block and between-block pair-copulas

We want to clarify the closed-form expressions of the within-block and between-block pair-copulas. We first focus on the between-block pair-copulas, and then on the within-block pair-copulas since they appear as a special case when all parameters of the between-block pair-copulas are equal.

According to (3.17), the asymmetric pair-copulas belong to the following family of bivariate copulas

$$C_{\delta_0,\delta_1,\delta_2}(u_1,u_2) = \psi\left(\psi^{-1}\left(u_1;\delta_1\right)+\psi^{-1}\left(u_2;\delta_2\right);\delta_0\right)u_1^{1-\frac{\delta_0}{\delta_1}}u_2^{1-\frac{\delta_0}{\delta_2}}, \quad (u_1,u_2)\in[0,1]^2, \qquad (3.18)$$

with $0<\delta_0\leq\min\left(\delta_1,\delta_2\right)$. The pair-copula (3.18) has the asymmetric form proposed in Chapter 4 of Khoudraji (1995). Indeed it can be rewritten in the following way

$$C_{\delta_0,\delta_1,\delta_2}(u_1,u_2) = C_{\delta_0,\delta_0,\delta_0}\left(u_1^{\frac{\delta_0}{\delta_1}},u_2^{\frac{\delta_0}{\delta_2}}\right)u_1^{1-\frac{\delta_0}{\delta_1}}u_2^{1-\frac{\delta_0}{\delta_2}}, \quad (u_1,u_2)\in[0,1]^2.$$

In the same way, $C_{\mathbf{U}^{\{\sigma\}}}$ has the asymmetric form proposed in Liebscher (2008). Based on (3.18) and the LST families of Table 3.1, Table 3.2 shows several examples of asymmetric pair-copulas.

Table 3.2 – Families of between-block asymmetric pair-copulas based on LST families of Table 3.1

#	$C_{\delta_0,\delta_1,\delta_2}(u_1,u_2)$
1	$\mathrm{e}^{-\left(\left(\left(-\frac{\delta_0}{\delta_1}\ln(u_1)+\delta_0 h^\lambda\right)^{\frac{1}{\lambda}}+\left(-\frac{\delta_0}{\delta_2}\ln(u_2)+\delta_0 h^\lambda\right)^{\frac{1}{\lambda}}-\delta_0^{\frac{1}{\lambda}}h\right)^{\lambda}-\delta_0 h^\lambda\right)} u_1^{1-\frac{\delta_0}{\delta_1}} u_2^{1-\frac{\delta_0}{\delta_2}}$
2	$\left(u_1^{-\frac{1}{\delta_1}}+u_2^{-\frac{1}{\delta_2}}-1\right)^{-\delta_0} u_1^{1-\frac{\delta_0}{\delta_1}} u_2^{1-\frac{\delta_0}{\delta_2}}$
3	$\left(1+\left(\left(u_1^{-\frac{1}{\delta_1}}-1\right)^{\beta}+\left(u_2^{-\frac{1}{\delta_2}}-1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\delta_0} u_1^{1-\frac{\delta_0}{\delta_1}} u_2^{1-\frac{\delta_0}{\delta_2}}$
4	$\left(1+\beta\ln\left(\mathrm{e}^{\frac{1}{\beta}\left(u_1^{-\frac{1}{\delta_1}}-1\right)}+\mathrm{e}^{\frac{1}{\beta}\left(u_2^{-\frac{1}{\delta_2}}-1\right)}-1\right)\right)^{-\delta_0} u_1^{1-\frac{\delta_0}{\delta_1}} u_2^{1-\frac{\delta_0}{\delta_2}}$
5*	$\left(1-\ln\left(1-(1-g_1(u_1)\times g_2(u_2))^{\beta}\right)\right)^{-\delta_0} u_1^{1-\frac{\delta_0}{\delta_1}} u_2^{1-\frac{\delta_0}{\delta_2}}$
6	$\left(\frac{q}{\left(qu_1^{-\frac{1}{\delta_1}}+1-q\right)\times\left(qu_2^{-\frac{1}{\delta_2}}+1-q\right)-(1-q)}\right)^{\delta_0} u_1^{1-\frac{\delta_0}{\delta_1}} u_2^{1-\frac{\delta_0}{\delta_2}}$
7**	$\mathrm{e}^{-\delta_0 t(u_1,u_2)\frac{2+t(u_1,u_2)}{1+t(u_1,u_2)}} u_1^{1-\frac{\delta_0}{\delta_1}} u_2^{1-\frac{\delta_0}{\delta_2}}$

* $g_i(u_i)=1-\left(1-\mathrm{e}^{1-u_i^{-\frac{1}{\delta_i}}}\right)^{\frac{1}{\beta}}$, for $i\in\{1,2\}$

** $t(u_1,u_2)=\frac{-\delta_1^{-1}\ln(u_1)+\sqrt{\left(-\delta_1^{-1}\ln(u_1)\right)^2+4}}{2}+\frac{-\delta_2^{-1}\ln(u_2)+\sqrt{\left(-\delta_2^{-1}\ln(u_2)\right)^2+4}}{2}-2$

To illustrate the shape of these asymmetric pair-copulas, let us consider Figure 3.3 that shows both normalized density contour plots and scatterplots of 2,000 samples for three bivariate asymmetric copulas. These copulas are generated with LSTs of the gamma family (#2), the gamma stopped positive stable family (#3) and the gamma stopped gamma family (#4) (see Table 3.2). The three scatterplots have a concave shape because the parameters of the three asymmetric copulas have been chosen such that $\delta_1 > \delta_2$. If we let $\delta_1 < \delta_2$, the scatterplots would have a convex shape. Note that parameter values were chosen to illustrate the asymmetrical shapes of the copulas rather than having the same Kendall's tau. We can also see, by comparing plots obtained with LST families (#2) and (#3), how adding a parameter $\beta > 1$ affects the Kendall's tau value.

The asymmetric pair-copulas (#2) of Table 3.2 have been explicitely introduced in Khoudraji (1995) and can be considered as an asymmetric adaptation of the Archimedean Clayton copula. Referring to Proposition 18 of Bernardoff (2018), we can in particular deduce the following expressions of its Spearman's rho and Kendall's tau

$$\rho\left(C_{\delta_0,\delta_1,\delta_2}\right)=3\times\left({}_3F_2\left([1,1,\delta_0],[1+2\delta_1,1+2\delta_2],1\right)-1\right),$$

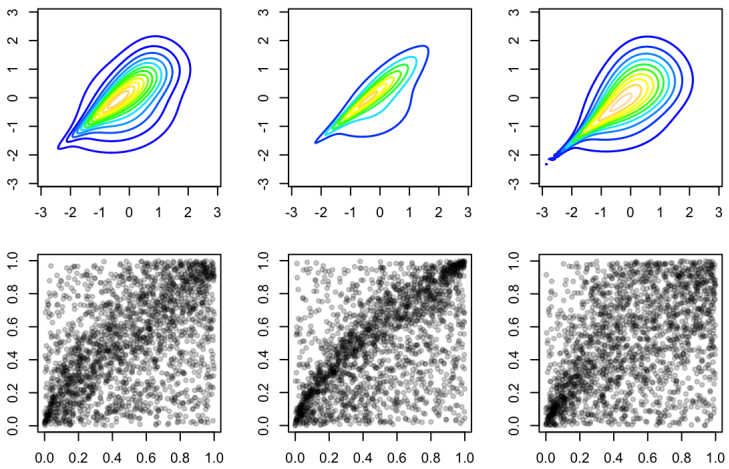


Figure 3.3 – First column: bivariate copula (#2) with $\tau = 0.337$ ($\delta_0 = 0.15, \delta_1 = 0.3$ and $\delta_2 = 0.2$), second column: bivariate copula (#3) with $\tau = 0.403$ ($\delta_0 = 0.15, \delta_1 = 0.3, \delta_2 = 0.2$ and $\beta = 1.7$), and third column: bivariate copula (#4) with $\tau = 0.317$ ($\delta_0 = 1.2, \delta_1 = 2, \delta_2 = 1.5$ and $\beta = 0.5$). Top row: normalized density contour plots, and bottom row: scatterplots of 2000 samples

and

$$\begin{aligned}\tau\left(C_{\delta_0,\delta_1,\delta_2}\right) &= 1 - {}_3F_2([1,1,2\delta_0],[1+2\delta_1,1+2\delta_2],1) \\ &\quad + \frac{4\delta_0}{(1+2\delta_1)(1+2\delta_2)}\,{}_3F_2([1,2,2\delta_0+1],[2+2\delta_1,2+2\delta_2],1) \\ &\quad - \frac{\delta_0^2}{(1+2\delta_1)(1+\delta_1)(1+2\delta_2)(1+\delta_2)}\,{}_3F_2([2,2,2\delta_0+2],[3+2\delta_1,3+2\delta_2],1),\end{aligned}$$

where ${}_3F_2$ is the generalized hyper-geometric function with 3 parameters of type 1 and 2 parameters of type 2 (see Abramowitz and Stegun (1948) for details).

Figure 3.4 illustrates the change in the normalized density contour plots of this copula (#2) according to the three copula parameters δ_0, δ_1 and δ_2. As δ_0 increases (first row in Figure 3.4), the normalized density contour spreads out more and becomes more asymmetrical. In contrast, when δ_1 (resp. δ_2) increases, the second row (resp. third row) in Figure 3.4 shows a contraction of the normalized density contour. Note that, when $\delta_1 < \delta_2$, the left tail of the normalized density contour is below the diagonal, otherwise it is above.

As already explained, LST (#3) generalizes LST (#2) in Table 3.1 and therefore copula (#3) generalizes copula (#2) in Table 3.2. According to Figure 3.5, as β increases, the normalized density contour of copula (#3) tends to shrink and the convex curve in the scatterplot becomes more intense. Note that the parameters were chosen such that $\delta_0 < \delta_1 < \delta_2$. If one had chosen $\delta_1 > \delta_2$, then the scatterplots would have a concave curve.

To end this subsection, let us consider the within-block Archimedean pair-copulas. According to (3.16), they have the following form

$$C_{\delta,\delta,\delta}(u_1,u_2) = \psi\left(\psi^{-1}\left(u_1;\delta\right) + \psi^{-1}\left(u_2;\delta\right);\delta\right), \quad (u_1,u_2) \in [0,1]^2,$$

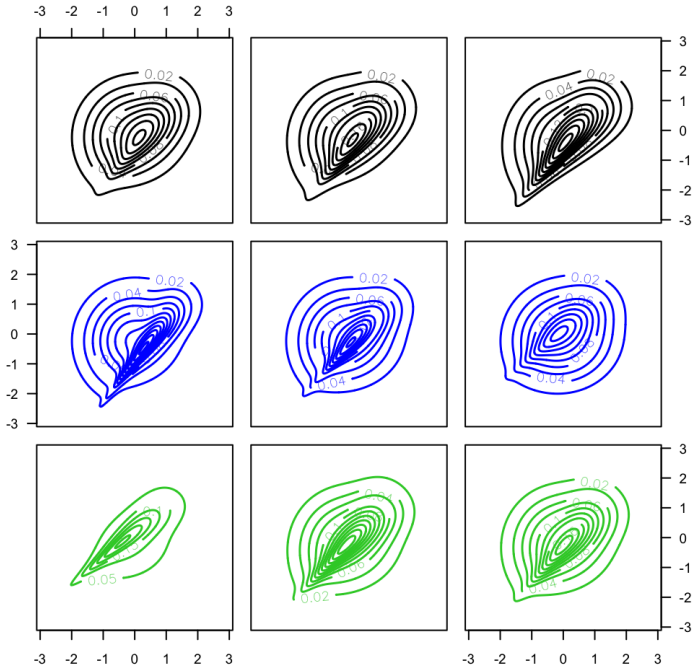


Figure 3.4 – Illustration of the normalized density contour plots of copula (#2) for increasing positive value of δ_0 (first row, with $\delta_1 = 0.2$ and $\delta_2 = 0.4$), δ_1 (second row, with $\delta_0 = 0.05$ and $\delta_2 = 0.2$), and δ_2 (third row, with $\delta_0 = 0.1$ and $\delta_1 = 0.2$)

with $\delta > 0$. It is noteworthy that they are derived from the asymmetric pair-copula (3.18) by letting the three parameters to be equal. For a large number of LST families considered in Table 3.2, we can deduce their associated well-known Archimedean copula given in Table 3.3.

3.3.2 Stochastic comparison of bivariate copulas

In the following, we investigate the dependence properties of the two-level hierarchical structure through the stochastic comparison of between-block and within-block pair-copulas.

Let

$$\Psi_\delta(x,y) = -\delta\varphi\left(\varphi^{-1}\left(\frac{x}{\delta}\right) + \varphi^{-1}\left(\frac{y}{\delta}\right)\right) + x + y, \quad x,y \geq 0,$$

where φ is defined in (3.4) and φ^{-1} is its inverse function. Then, (3.18) can be expressed as

$$C_{\delta_0,\delta_1,\delta_2}(u_1,u_2) = \exp\left(\Psi_{\delta_0}\left(-\ln\left(u_1^{\frac{\delta_0}{\delta_1}}\right), -\ln\left(u_2^{\frac{\delta_0}{\delta_2}}\right)\right)\right) \times u_1 u_2, \quad (u_1,u_2) \in [0,1]^2.$$

From (3.17) and (3.16), we see that the asymmetric between-block pair-copula (3.18) has to be compared with the two following within-block symmetric pair-copulas

$$C_{\delta_1,\delta_1,\delta_1}(u_1,u_2) = \exp\left(\Psi_{\delta_1}\left(-\ln(u_1), -\ln(u_2)\right)\right) \times u_1 u_2$$

and

$$C_{\delta_2,\delta_2,\delta_2}(u_1,u_2) = \exp\left(\Psi_{\delta_2}\left(-\ln(u_1), -\ln(u_2)\right)\right) \times u_1 u_2.$$

We consider the concordance order as defined in Section 2.2.1 of Joe (1997) for the comparison.

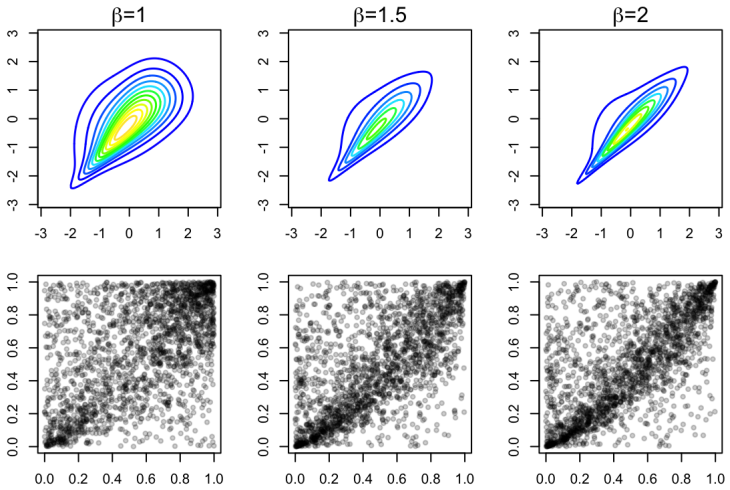


Figure 3.5 – Bivariate asymmetric copula #3. First column: $\beta = 1$ and $\tau = 0.347$, second column: $\beta = 1.5$ and $\tau = 0.456$, and third column: $\beta = 2$ and $\tau = 0.49$. Top row: normalized density contour plots, and bottom row: scatterplots of 2000 samples

Definition 1 *Let C_1 and C_2 be two pair-copulas. C_2 is more **concordant** that C_1, written $C_1 \prec_c C_2$, if $C_1(u_1, u_2) \leq C_2(u_1, u_2)$, for $(u_1, u_2) \in [0,1]^2$.*

Remark 1 *As mentioned in Section 2.2.1 of Joe (1997), a very useful consequence of the concordance order is its relation with dependence measures. If C_1 and C_2 are pair-copulas with respective Kendall taus τ_1, τ_2, Spearman rhos ρ_1, ρ_2, upper tail dependence parameters λ_1, λ_2, and such that $C_1 \prec_c C_2$, then $\tau_1 \leq \tau_2$, $\rho_1 \leq \rho_2$, and $\lambda_1 \leq \lambda_2$.*

Let, for $0 < \delta_0 \leq \delta$, $x \geq 0$,

$$f_{\delta_0,\delta}(x) = \varphi^{-1}\left(\frac{\delta_0}{\delta}\varphi(x)\right).$$

The following proposition leads us to compare the within-blocks pair-copulas and between-blocks pair-copulas with the concordance order.

Proposition 3 *For $0 < \delta_0 \leq \min(\delta_1, \delta_2)$, we have*

$$C_{\delta_0,\delta_1,\delta_2} \prec_c C_{\delta_1,\delta_1,\delta_1} \qquad \text{and} \qquad C_{\delta_0,\delta_1,\delta_2} \prec_c C_{\delta_2,\delta_2,\delta_2},$$

if and only if

$$\Psi_{\delta_1}(x,y) \geq \Psi_{\delta_0}\left(\frac{\delta_0}{\delta_1}x, \frac{\delta_0}{\delta_2}y\right) \qquad \text{and} \qquad \Psi_{\delta_2}(x,y) \geq \Psi_{\delta_0}\left(\frac{\delta_0}{\delta_1}x, \frac{\delta_0}{\delta_2}y\right), \tag{3.19}$$

for all $x, y \geq 0$. Moreover, if f_{δ_0,δ_1} and f_{δ_0,δ_2} are subadditive functions, and Ψ_{δ_0} is an increasing function in each of its arguments, then (3.19) is satisfied.

Proof. See Appendix C. ■

Table 3.3 – Families of within-block Archimedean pair-copulas based on Table 3.2

#	Special cases	$C_{\delta,\delta,\delta}(u_1,u_2)$	Name
1	–	$\mathrm{e}^{-\left(\left(\left(-\ln(u_1)+\delta h^\lambda\right)^{\frac{1}{\lambda}}+\left(-\ln(u_2)+\delta h^\lambda\right)^{\frac{1}{\lambda}}-\delta^{\frac{1}{\lambda}}h\right)^{\lambda}-\delta h^\lambda\right)}$	
1	$h=1,\lambda=0.5$	$\mathrm{e}^{\delta\left(1-\sqrt{\left(1-\ln\left(u_1^{\frac{1}{\delta}}\right)\right)^2+\left(1-\ln\left(u_2^{\frac{1}{\delta}}\right)\right)^2-1}\right)}$	
1	$h=0$	$\mathrm{e}^{-\left((-\ln(u_1))^{\frac{1}{\lambda}}+(-\ln(u_2))^{\frac{1}{\lambda}}\right)^{\lambda}}$	Gumbel
1	$\delta h^\lambda=1$	$\mathrm{e}^{-\left(\left((-\ln(u_1)+1)^{\frac{1}{\lambda}}+(-\ln(u_2)+1)^{\frac{1}{\lambda}}-1\right)^{\lambda}-1\right)}$	#13 (Nelsen, 2006)
2	–	$\left(u_1^{-\frac{1}{\delta}}+u_2^{-\frac{1}{\delta}}-1\right)^{-\delta}$	Clayton
3	–	$\left(1+\left(\left(u_1^{-\frac{1}{\delta}}-1\right)^{\beta}+\left(u_2^{-\frac{1}{\delta}}-1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\delta}$	BB1 (Joe, 2014)
4	–	$\left(1+\beta\ln\left(\mathrm{e}^{\frac{1}{\beta}\left(u_1^{-\frac{1}{\delta}}-1\right)}+\mathrm{e}^{\frac{1}{\beta}\left(u_2^{-\frac{1}{\delta}}-1\right)}-1\right)\right)^{-\delta}$	BB2 (Joe, 2014)
5*	–	$\left(1-\ln\left(1-(1-g(u_1)\times g(u_2))^{\beta}\right)\right)^{-\delta}$	
6	–	$\left(\frac{q}{\left(qu_1^{-\frac{1}{\delta}}+1-q\right)\times\left(qu_2^{-\frac{1}{\delta}}+1-q\right)-(1-q)}\right)^{\delta}$	
7**	–	$\mathrm{e}^{-\delta t(u_1,u_2)\frac{2+t(u_1,u_2)}{1+t(u_1,u_2)}}$	

* $g(u)=1-\left(1-\mathrm{e}^{1-u^{-\frac{1}{\delta}}}\right)^{\frac{1}{\beta}}$

** $t(u_1,u_2)=\frac{-\delta^{-1}\ln(u_1)+\sqrt{\left(-\delta^{-1}\ln(u_1)\right)^2+4}}{2}+\frac{-\delta^{-1}\ln(u_2)+\sqrt{\left(-\delta^{-1}\ln(u_2)\right)^2+4}}{2}-2$

We deduce from Proposition 3 that Kendall's tau (resp. Spearman's rho) values of within-blocks pair-copulas are higher than Kendall's tau (resp. Spearman's rho) values of between-blocks pair-copulas. This is crucial in the estimation procedure described in Section 3.4 where we will identify blocks through clustering algorithms that use a matrix of dependence measures.

Corollary 1 *For LST families (#1), (#2), (#3), the within-blocks pair-copulas are more concordant than the between-blocks pair-copulas.*

Proof. See Appendix D. ■

3.4 Two-level hierarchical copulas: estimation procedure and case studies

For the two-level hierarchical copulas (3.13), we propose an appropriate estimation procedure. We investigate its performance through three simulation examples and two real dataset studies.

3.4.1 Estimation procedure

The structure determination and parameter estimation of HACs have attracted significant attention in recent literature. Savu and Trede (2010) is among the first to provide an estimation approach for HACs assuming prior knowledge on the structure, while Okhrin et al. (2013) appears as the first paper to simultaneously estimate both the structure and the parameters using a multi-stage procedure. The latter presents several approaches to identify the structure and proposes an estimation procedure mainly based on maximum-likelihood. Kendall's tau inversion technique is briefly mentioned as an alternative to estimate the parameters. Several research papers proposed other approaches to perform both tasks, e.g., Górecki and Holeňa (2013) and Górecki et al. (2016) suggest to use Kendall's tau and its pairwise inversion. Cossette et al. (2019) propose an agglomerative hierarchical clustering method based on Spearman's rho matrix to determinate the tree structure combined with a top-down composite likelihood estimation procedure for the parameter estimation of the family of HACs introduced in Cossette et al. (2017).

After dealing in Section 3.3 with the concordance properties of the two-level hierarchical copulas, we now propose an estimation procedure based on these properties. The components of the hierarchical copula $C_{\mathbf{U}}$ are gathered into $d = a_{\mathbf{i}_0}$ blocks. The first step is to estimate the number of blocks as well as leaves belonging to each block. A first approach consists in using the identification algorithm proposed by Perreault et al. (2019), which is based on an investigation of block exchangeability in the Kendall rank correlation matrix. A second approach consists in using a clustering algorithm. Both methods use a matrix of dependence measures, and hence they are useful to estimate the structure of our hierarchical copula when Proposition 3 is satisfied. Let us now discuss briefly both approaches.

The iterative algorithm introduced in Perreault et al. (2019) is appropriate for data with a structure having exchangeable blocks. It aims to identify elements of the correlation matrix that are equal. In the specific case of our model, considering several simulated examples we have tried, we noted that the algorithm succeeds in identifying blocks with a great accuracy, but as the number of observations or variables increases, the procedure requires more and more computational time. Moreover, when the differences between Kendall's tau values within and between blocks were small, the algorithm encountered difficulties to group the variables into blocks.

Clustering algorithms also appear to be appropriate because they can use the property that Kendall's tau (resp. Spearman's rho) values of within-blocks pair-copulas are higher than Kendall's tau (resp. Spearman's rho) values of between-blocks pair-copulas. A two-step approach is necessary to determine the structure. First, the number of groups is estimated using the "silhouette coefficient", which we will use in the application of the Partitioning Around Medoids (PAM) algorithm, introduced by Rousseeuw and Kaufman (1990), on the Kendall's tau matrix. Contrary to the K-means algorithm introduced by MacQueen et al. (1967) and which represents each cluster center by its mean, the PAM algorithm consists first in finding a set of d medoids and then constructing d clusters by assigning each variable

to the nearest medoid while minimizing the sum of the dissimilarities of the variables to their closest representative medoid. See, e.g., Bernard et al. (2013) and Barbakh et al. (2009) for more details on the PAM algorithm.

Once the structure is properly identified and corresponds to the structure of a hierarchical copula $C_{\mathbf{U}}$ given by (3.13), the second step of the estimation procedure is to estimate the parameters of the convolution distributions and the Lévy subordinator. There are two vectors of parameters to be estimated. The first one is the vector of parameters $\boldsymbol{\alpha} = (\alpha_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, 1)$ or equivalently the vector of parameters $\boldsymbol{\delta} = (\delta_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, 1)$ that appear in (3.4) and can be seen as the times the Lévy process is sampled. There is one parameter (one component of the vector) for each node of the structure. The second vector of parameters is the vector of parameters that characterizes the Lévy subordinator and its function φ. This vector is shared by all nodes and we denote it by $\boldsymbol{\kappa}$. For example, in the case (#1) of Table 3.1, $\boldsymbol{\kappa} = (\lambda, h)$.

Given m observations of $\mathbf{U} \sim C_{\mathbf{U}}$, denoted by $\mathbf{u}_k, k \in \{1, \ldots, m\}$, the maximum-likelihood estimation method is defined as the maximization of

$$L_{C_{\mathbf{U}}}(\boldsymbol{\delta}, \boldsymbol{\kappa}) = \prod_{k=1}^{m} c_{\mathbf{U}}(\mathbf{u}_k; \boldsymbol{\delta}, \boldsymbol{\kappa}),$$

where $c_{\mathbf{U}}$ is the density of the copula $C_{\mathbf{U}}$. As the dimension of $C_{\mathbf{U}}$ becomes larger, finding an algebraic expression for $c_{\mathbf{U}}$ becomes more tedious and, when it is possible to derive it, the optimization requires more computational time. Thus, we turn our attention to a composite likelihood approach, which reduces the overall complexity of the problem.

Named by Lindsay (1988), a composite likelihood function is formed by multiplying likelihoods of a set of conditional or marginal events. The latter involves smaller subsets of data, which reduces the computational challenge, while the results remain asymptotically unbiased. For more details, we refer to Zhao and Joe (2005), Varin and Vidoni (2005) and Varin et al. (2011). Moreover, recent works use composite likelihood methods in several areas of application leading to efficient results for, e.g., multivariate survival data (Parner (2001)), frailty models (Henderson and Shimakura (2003)) or multi-level insurance claims (Shi et al. (2016)). Cossette et al. (2019) proposed a top-down composite likelihood procedure to estimate HACs proposed in Cossette et al. (2017). Inspired by the latter, we propose in the following a bottom-up composite likelihood estimation procedure assuming the hierarchical structure of $C_{\mathbf{U}}$ is known.

To illustrate our approach, let us begin with the simpler case where there is no vector of parameters $\boldsymbol{\kappa}$ as in case (#2) of Table 3.1, i.e., $\psi(\cdot\,; \delta_{\mathbf{i}_1})$ is a one-parameter LST. Since the copula associated to the children node $\Theta^{\mathbf{i}_1}$, $\mathbf{i}_1 \in \mathcal{I}_1$, is a $a_{\mathbf{i}_1}$-dimensional Archimedean copula with generator $\psi(.; \delta_{\mathbf{i}_1})$, its density can be written as

$$c_{\mathbf{U}^{\mathbf{i}_1}}\left(\mathbf{u}^{\mathbf{i}_1}; \delta_{\mathbf{i}_1}\right) = \psi^{(a_{\mathbf{i}_1})}\left(\sum_{j=1}^{a_{\mathbf{i}_1}} \psi^{-1}\left(u^{(\mathbf{i}_1,j)}; \delta_{\mathbf{i}_1}\right); \delta_{\mathbf{i}_1}\right) \prod_{j=1}^{a_{\mathbf{i}_1}} \left(\psi^{-1}\right)'\left(u^{(\mathbf{i}_1,j)}; \delta_{\mathbf{i}_1}\right), \tag{3.20}$$

where $\psi^{(a)}(\cdot\,; \delta_{\mathbf{i}_1})$ is the a^{th} derivative of the univariate function $\psi(\cdot\,; \delta_{\mathbf{i}_1})$, for $a \in \mathbb{N}_+$. See, e.g., Hofert et al. (2012) for more details. We can therefore estimate the dependence parameter $\delta_{\mathbf{i}_1}$ using the maximum likelihood approach applied to the block of observations $\mathbf{u}_k^{\mathbf{i}_1} = \left(u_k^{(\mathbf{i}_1,j)} : j = 1, \ldots, a_{\mathbf{i}_1}\right)$, $k = 1, \ldots, m$. We then go into the structure up to $r = 0$, and consider the asymmetric copula

associated to the root node for which only one unknown parameter, $\delta_{\mathbf{i}_0}$, has to be estimated while the other parameters, $\left(\delta_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)$, are replaced by their estimates of the previous step, $\left(\hat{\delta}_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)$,

$$\begin{aligned} & C_{\mathbf{U}^{\{\sigma\}}}\left(\mathbf{u}^{\{\sigma\}}; \delta_{\mathbf{i}_0}, \left(\hat{\delta}_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)\right) \\ & = \psi\left(\sum_{j=1}^{a_{\mathbf{i}_0}} \psi^{-1}\left(u^{\sigma^{(j)}}; \hat{\delta}_{\mathbf{i}_1^{(j)}}\right); \delta_{\mathbf{i}_0}\right) \times \prod_{j=1}^{a_{\mathbf{i}_0}} \psi\left(\psi^{-1}\left(u^{\sigma^{(j)}}; \hat{\delta}_{\mathbf{i}_1^{(j)}}\right); \hat{\delta}_{\mathbf{i}_1^{(j)}} - \delta_{\mathbf{i}_0}\right), \end{aligned} \tag{3.21}$$

for $\sigma \in S_{\mathbf{i}_0}$. To estimate $\delta_{\mathbf{i}_0}$, we use all possible combinations of observations with indices belonging to the set $S_{\mathbf{i}_0}$, and the likelihood function given by

$$L_C\left(\delta_{\mathbf{i}_0}\right) = \prod_{k=1}^{m} \prod_{\sigma \in S_{\mathbf{i}_0}} c\left(\mathbf{u}_k^{\{\sigma\}}; \delta_{\mathbf{i}_0}, \left(\hat{\delta}_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)\right).$$

Since the copula given in (3.21) is not Archimedean, the closed-form expression of its density $c_{\mathbf{U}^{\{\sigma\}}}\left(\mathbf{u}^{\{\sigma\}}\right)$ is not known and it may be difficult to find it theoretically in some cases. The parameter estimation procedure for hierarchical copulas with a one-parameter LST is summarized in Algorithm 4.

Algorithme 4 : Bottom-up composite likelihood procedure for hierarchical copulas with a one-parameter LST

Input : $\mathbf{u}_k, k \in \{1, \ldots, m\}$
Result : $\left(\hat{\delta}_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r \in \{0, 1\}\right)$

foreach *path* $\mathbf{i}_1 \in \mathcal{I}_1$ **do**
  Compute $L_C\left(\delta_{\mathbf{i}_1}\right) = \prod_{k=1}^{m} c_{\mathbf{U}^{\mathbf{i}_1}}\left(u_k^{(\mathbf{i}_1,1)}, \ldots, u_k^{(\mathbf{i}_1,a_{\mathbf{i}_1})}\right)$;
  Solve $\hat{\delta}_{\mathbf{i}_1} = \underset{\delta_{\mathbf{i}_1}}{\operatorname{argmax}} \ \ln\left(L_C\left(\delta_{\mathbf{i}_1}\right)\right)$;
end
Compute $L_C\left(\delta_{\mathbf{i}_0}\right) = \prod_{k=1}^{m} \prod_{\sigma \in S_{\mathbf{i}_0}} c_{\mathbf{U}^{\{\sigma\}}}\left(u_k^{\sigma^{(1)}}, \ldots, u_k^{\sigma^{(a_{\mathbf{i}_0})}} | \hat{\delta}_{\mathbf{i}_1} : \mathbf{i}_1 \in \mathcal{I}_1\right)$;
Solve $\hat{\delta}_{\mathbf{i}_0} = \underset{\delta_{\mathbf{i}_0}}{\operatorname{argmax}} \ \ln\left(L_C\left(\delta_{\mathbf{i}_0}\right)\right)$.

Now, we consider the general case where $C_{\mathbf{U}}$ is a two-level hierarchical copula generated with a two-parameter LST family. For each node $\Theta^{\mathbf{i}_r}$ of the tree structure, we have to estimate the associated parameter $\delta_{\mathbf{i}_r}$, and for all nodes we have to estimate the common parameter $\boldsymbol{\kappa}$. To do so, we start from $r = 1$. The copula associated to the children node $\Theta^{\mathbf{i}_1}$, $\mathbf{i}_1 \in \mathcal{I}_1$, is an $a_{\mathbf{i}_1}$-dimensional Archimedean copula with generator $\psi\left(.; \delta_{\mathbf{i}_1}, \boldsymbol{\kappa}\right)$ and its density can be written as

$$c_{\mathbf{U}^{\mathbf{i}_1}}\left(\mathbf{u}^{\mathbf{i}_1}; \delta_{\mathbf{i}_1}, \boldsymbol{\kappa}\right) = \psi^{(a_{\mathbf{i}_1})}\left(\sum_{j=1}^{a_{\mathbf{i}_1}} \psi^{-1}\left(u^{(\mathbf{i}_1,j)}; \delta_{\mathbf{i}_1}, \boldsymbol{\kappa}\right); \delta_{\mathbf{i}_1}, \boldsymbol{\kappa}\right) \prod_{j=1}^{a_{\mathbf{i}_1}} \left(\psi^{-1}\right)'\left(u^{(\mathbf{i}_1,j)}; \delta_{\mathbf{i}_1}, \boldsymbol{\kappa}\right),$$

where $\psi^{(a)}\left(\cdot\,; \delta_{\mathbf{i}_1}, \boldsymbol{\kappa}\right)$ is the a^{th} derivative of the univariate function $\psi\left(\cdot\,; \delta_{\mathbf{i}_1}, \boldsymbol{\kappa}\right)$, for $a \in \mathbb{N}_+$. For each $\mathbf{i}_1 \in \mathcal{I}_1$, we estimate $(\delta_{\mathbf{i}_1}, \boldsymbol{\kappa})$ using the maximum likelihood approach applied to the block of observations $\mathbf{u}_k^{\mathbf{i}_1} = \left(u_k^{(\mathbf{i}_1,j)} : j = 1, \ldots, a_{\mathbf{i}_1}\right)$, $k = 1, \ldots, m$, and we denote by $(\hat{\delta}_{\mathbf{i}_1}, \hat{\boldsymbol{\kappa}}_{\mathbf{i}_1})$ the result of the optimisation program. We then go into the structure up to $r = 0$, and consider the asymmetric

copula associated to the root node for which the parameters, $\delta_{\mathbf{i}_0}$ and $\boldsymbol{\kappa}$, also have to be estimated, assuming that the parameters, $\left(\delta_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)$, are replaced by their estimates of the previous step, $\left(\hat{\delta}_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)$,

$$\begin{aligned} & C_{\mathbf{U}^{\{\sigma\}}}\left(\mathbf{u}^{\{\sigma\}}; \delta_{\mathbf{i}_0}, \boldsymbol{\kappa}, \left(\hat{\delta}_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)\right) \\ = \; & \psi\left(\sum_{j=1}^{a_{\mathbf{i}_0}} \psi^{-1}\left(u^{\sigma^{(j)}}; \hat{\delta}_{\mathbf{i}_1^{(j)}}, \boldsymbol{\kappa}\right); \delta_{\mathbf{i}_0}, \boldsymbol{\kappa}\right) \times \prod_{j=1}^{a_{\mathbf{i}_0}} \psi\left(\psi^{-1}\left(u^{\sigma^{(j)}}; \hat{\delta}_{\mathbf{i}_1^{(j)}}\right); \hat{\delta}_{\mathbf{i}_1^{(j)}} - \delta_{\mathbf{i}_0}, \boldsymbol{\kappa}\right), \end{aligned}$$

for $\sigma \in S_{\mathbf{i}_0}$. We estimate $(\delta_{\mathbf{i}_0}, \boldsymbol{\kappa})$ using the asymmetric copula with all possible combinations of observations and likelihood function given by

$$L_C\left(\delta_{\mathbf{i}_0}, \boldsymbol{\kappa}\right) = \prod_{k=1}^{m} \prod_{\sigma \in S_{\mathbf{i}_0}} c\left(\mathbf{u}_k^{\{\sigma\}}; \delta_{\mathbf{i}_0}, \boldsymbol{\kappa}, \left(\hat{\delta}_{\mathbf{i}_1^{(j)}} : j = 1, \ldots, a_{\mathbf{i}_0}\right)\right).$$

Once the vector of parameters $(\hat{\delta}_{\mathbf{i}_r}, \hat{\boldsymbol{\kappa}}_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, 1)$ is estimated, we set the estimator of the common parameter equal to $\hat{\boldsymbol{\kappa}} = \text{mean}\left(\hat{\boldsymbol{\kappa}}_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, 1\right)$. Then, we estimate again $(\delta_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, 1)$ using Algorithm 4 and fixing $\boldsymbol{\kappa}$ to $\hat{\boldsymbol{\kappa}}$. The whole parameter estimation procedure is summarized in Algorithm 5.

Algorithme 5 : Bottom-up composite likelihood procedure for hierarchical copulas with a two-parameter LST

Input : $\mathbf{u}_k, k \in \{1, \ldots, m\}$
Result : $\left(\hat{\delta}_{\mathbf{i}_r}, \hat{\boldsymbol{\kappa}} : \mathbf{i}_r \in \mathcal{I}_r, r \in \{0, 1\}\right)$

foreach *path* $\mathbf{i}_1 \in \mathcal{I}_1$ **do**

Compute $L_C\left(\delta_{\mathbf{i}_1}, \boldsymbol{\kappa}_{\mathbf{i}_1}\right) = \prod_{k=1}^{m} c_{\mathbf{U}^{\mathbf{i}_1}}\left(u_k^{(\mathbf{i}_1,1)}, \ldots, u_k^{(\mathbf{i}_1,a_{\mathbf{i}_1})}\right)$;

Solve $\left(\hat{\delta}_{\mathbf{i}_1}, \hat{\boldsymbol{\kappa}}_{\mathbf{i}_1}\right) = \underset{\delta_{\mathbf{i}_1}, \boldsymbol{\kappa}_{\mathbf{i}_1}}{\text{argmax}} \; \ln\left(L_C\left(\delta_{\mathbf{i}_1}, \boldsymbol{\kappa}_{\mathbf{i}_1}\right)\right)$;

end

Compute $L_C\left(\delta_{\mathbf{i}_0}, \boldsymbol{\kappa}_{\mathbf{i}_0}\right) = \prod_{k=1}^{m} \prod_{\sigma \in S_{\mathbf{i}_0}} c_{\mathbf{U}^{\{\sigma\}}}\left(u_k^{\sigma^{(1)}}, \ldots, u_k^{\sigma^{(a_{\mathbf{i}_0})}} | \hat{\delta}_{\mathbf{i}_1}, \hat{\boldsymbol{\kappa}}_{\mathbf{i}_1} : \mathbf{i}_1 \in \mathcal{I}_1\right)$;

Solve $\left(\hat{\delta}_{\mathbf{i}_0}, \hat{\boldsymbol{\kappa}}_{\mathbf{i}_0}\right) = \underset{\delta_{\mathbf{i}_0}, \boldsymbol{\kappa}_{\mathbf{i}_0}}{\text{argmax}} \; \ln\left(L_C\left(\delta_{\mathbf{i}_0}, \boldsymbol{\kappa}_{\mathbf{i}_0}\right)\right)$;

Set $\hat{\boldsymbol{\kappa}} = \text{mean}\left(\hat{\boldsymbol{\kappa}}_{\mathbf{i}_r} : \mathbf{i}_r \in \mathcal{I}_r, r = 0, 1\right)$;

Estimate again $\left(\hat{\delta}_{\mathbf{i}_r}, \mathbf{i}_r \in \mathcal{I}_r, r \in \{0, 1\} \middle| \hat{\boldsymbol{\kappa}}\right)$ using Algorithm 4.

3.4.2 Simulated data

Based on simulated datasets, we provide illustrations of the full estimation procedure described in Section 3.4.1.

Example 2 *We consider the tree structure depicted in Figure 3.6, with* $\Theta \sim \Gamma(\delta_{(0)} = 0.1, 1)$, $\Theta^{(1)} \sim \Gamma(\delta_{(1)} = 0.11, 1)$, *and* $\Theta^{(2)} \sim \Gamma(\delta_{(2)} = 0.25, 1)$. *We carried out 1,000 samples of size 5 from the structure. Figure 3.1 shows the asymmetric and symmetric pairs within the sampled observations.*

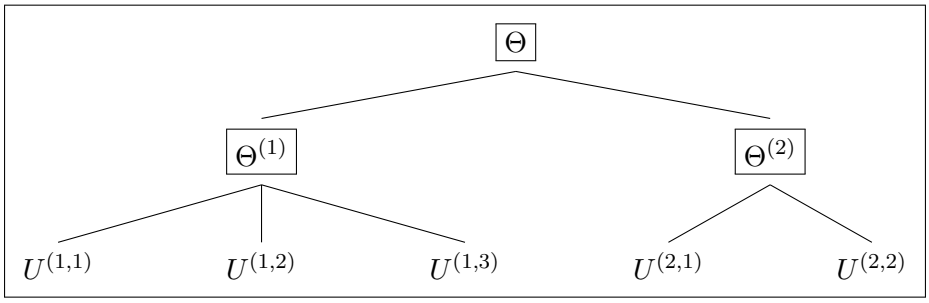


Figure 3.6 – Tree structure of Examples 2 and 3, with $l = 2$, $a_{(0)} = 2$, $a_{(1)} = 3$ and $a_{(2)} = 2$

Using Perreault et al. (2019)'s procedure, we identified adequately the two clusters $(U^{(1,1)}, U^{(1,2)}, U^{(1,3)})$ *and* $(U^{(2,1)}, U^{(2,2)})$*. For the parameter estimation, we performed 200 replicates of 1,000 simulations of* $\mathbf{U}$ *according to the tree in Figure 3.6 and at each replicate, we estimated the parameters according to Algorithm 4. Note that, over all the simulation replicates, we obtained the same approximated tree through the structure determination process. Table 3.4 provides the resulting empirical confidence intervals and standard deviations of the estimators, with* $q_{0.25}$ *and* $q_{0.75}$ *denoting, respectively, the 25th and 75th empirical quantiles. Note that all true parameter values belong to the corresponding confidence interval.*

Table 3.4 – Estimated parameters with simulated data according to tree structure in Figure 3.6, with assumptions of Example 2

Parameters	Value	$q_{0.25}$	Mean	$q_{0.75}$	Std
$\delta_{(0)}$	0.1	0.0981	0.1000	0.1018	0.0029
$\delta_{(1)}$	0.11	0.1084	0.1100	0.1115	0.0022
$\delta_{(2)}$	0.25	0.2460	0.2516	0.2569	0.0085

Example 3 *As in Example 2, we consider the tree structure depicted in Figure 3.6. We now assume that the distributions of the nodes belong to the gamma stopped positive stable (GSPS) distribution family, such that* $\Theta \sim GSPS(\delta_{(0)} = 0.2, \beta = 1.7)$*,* $\Theta^{(1)} \sim GSPS(\delta_{(1)} = 0.45, \beta = 1.7)$*, and* $\Theta^{(2)} \sim GSPS(\delta_{(2)} = 0.25, \beta = 1.7)$*. We carried out 1,000 samples of size 5 from the structure. Figure 3.7 shows the asymmetric and symmetric pairs within the sampled observations, through pairwise scatterplots and density contour plots. Note that, scatterplots have a convex curve since* $\delta_{(1)} > \delta_{(2)}$*.*

Based on Perreault et al. (2019)'s procedure, we identified adequately the two clusters with their appropriate components. For the parameter estimation, we performed 200 replicates of 1,000 simulations of $\mathbf{U}$ *according to the tree in Figure 3.6, and then we estimated the parameters according to Algorithm 5 for each simulation replicate. For all simulated observations, we obtained the same approximated tree through the structure determination process. Table 3.5 provides the resulting empirical confidence intervals and standard deviations of the estimators. Note that, the latter are greater than the ones obtained in Table 3.4. This illustrates the impact of adding a common parameter in the structure on the estimation precision of the other parameters.*

Compared to Example 2, where ψ is a one-parameter LST, the estimation process requires almost 10

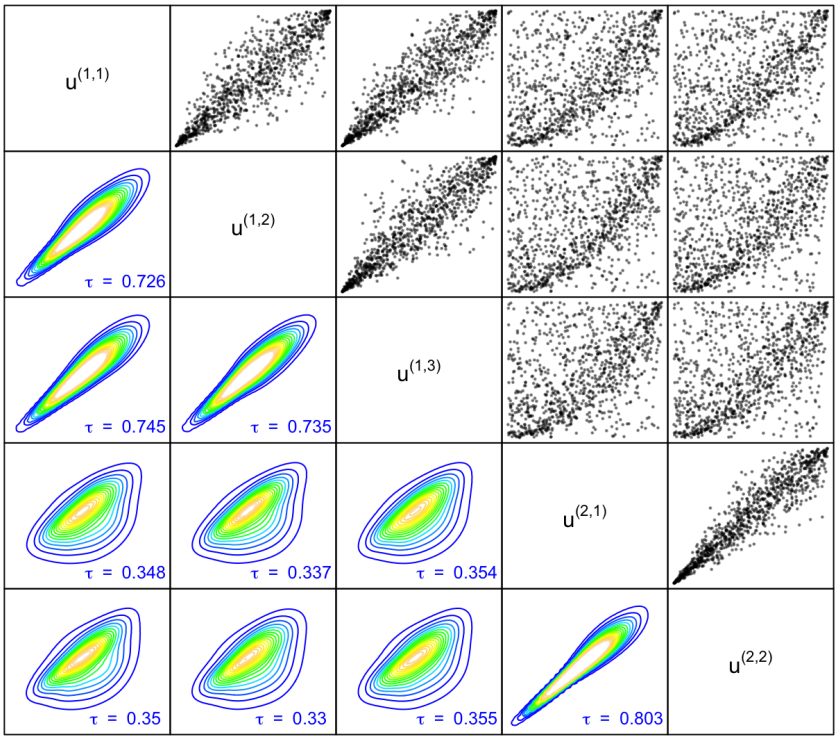


Figure 3.7 – Upper triangular matrix: pairwise scatterplots of 1,000 observations of Example 3, and lower triangular matrix: empirical density contour plots of bivariate normalized copula for all pairs

Table 3.5 – Estimated parameters with simulated data according to tree structure in Figure 3.6, with assumptions of Example 3

Parameters	Value	$q_{0.25}$	Mean	$q_{0.75}$	Std
β	1.70	1.6524	1.6896	1.7328	0.0624
$\delta_{(0)}$	0.20	0.1900	0.1980	0.2068	0.0119
$\delta_{(1)}$	0.45	0.4272	0.4468	0.4629	0.0252
$\delta_{(2)}$	0.25	0.2391	0.2481	0.2578	0.0143

times more computing time for Example 3, where ψ belongs to a two-parameter LST family.

Example 4 *We consider the dependence structure defined in Figure 3.8, with* $\Theta \sim \Gamma(\delta_{(0)} = 0.05, 1)$, $\Theta^{(1)} \sim \Gamma(\delta_{(1)} = 0.06, 1)$, $\Theta^{(2)} \sim \Gamma(\delta_{(2)} = 0.13, 1)$, *and* $\Theta^{(3)} \sim \Gamma(\delta_{(3)} = 0.21, 1)$. *We sampled 1,000 realizations of size 6 from the structure. Figure 3.9 shows both the pairwise scatterplots and density contour plots of the sampled observations and illustrates the asymmetric and symmetric pairs. To approximate the tree structure, we use once again Perreault et al. (2019)'s algorithm which identified accurately the number of clusters and with the right components within them. To estimate the param-*

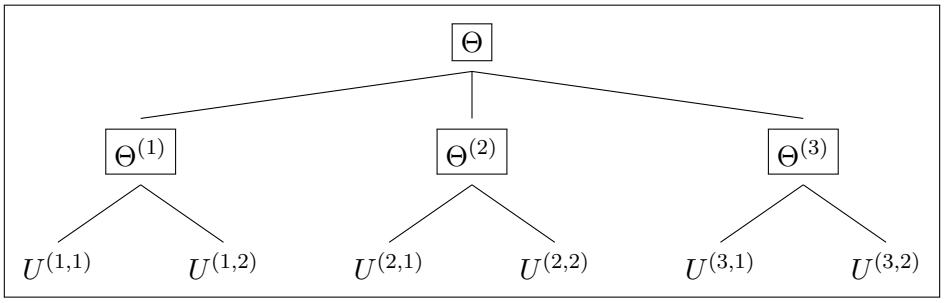


Figure 3.8 – Tree structure of Example 4, with $l = 2$, $a_{(0)} = 3$ and $a_{(1)} = a_{(2)} = a_{(3)} = 2$

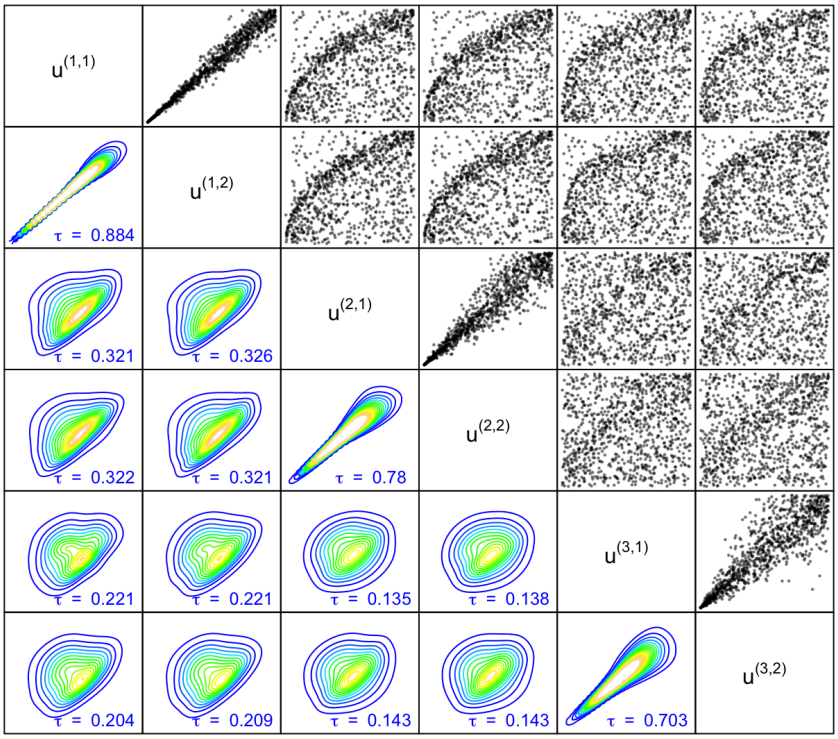


Figure 3.9 – Upper triangular matrix: pairwise scatterplots of 1,000 observations of Example 4, and lower triangular matrix: empirical density contour plots of bivariate normalized copula for all pairs

eters, we performed 200 replicates of 1,000 simulations of $\mathbf{U}$ *according to the tree in Figure 3.8, and at each replicate we applied the full estimation procedure. For all 200 replicates, we obtained the same approximated structure, which confirms the efficiency of the structure determination procedure. Table 3.6 provides the empirical resulting confidence intervals and standard deviations of the estimators.*

Compared to Example 2, where the tree structure has two nodes at level 1, the estimation process

Table 3.6 – Estimated parameters with simulated data according to tree structure in Figure 3.8, with assumptions of Example 4

Parameters	Value	$q_{0.25}$	Mean	$q_{0.75}$	Std
$\delta_{(0)}$	0.05	0.0483	0.0495	0.0506	0.0016
$\delta_{(1)}$	0.06	0.0589	0.0599	0.0611	0.0016
$\delta_{(2)}$	0.13	0.1270	0.1299	0.1331	0.0043
$\delta_{(3)}$	0.21	0.2053	0.2105	0.2147	0.0074

requires almost twice the computing time for Example 4, where we have three nodes at level 1. It is obvious that the computational time will be more important as the number of nodes or levels increases. The dimension of non-exchangeable between-block copulas is directly linked to the number of nodes and thus the maximization problem becomes more complex. However, as the within-block copulas are Archimedean copulas, we can increase the number of leaves within these blocks without changing the complexity of the computation.

3.4.3 Nutrient data

We consider a Nutrient dataset from a study of nutrient intake conducted by the United States Department of Agriculture in 1985. For $m = 737$ women, five variables were measured, namely the daily Calcium (Ca), Iron (Ir), Protein (Pr), Vitamin C and Vitamin A (VitA) intake. This dataset has already been used in the literature to study possible asymmetries within dependence structures. More precisely, McNeil and Nešlehová (2010) have shown improvement in fitting the asymmetric Clayton Liouville class to Nutrient data instead of the Archimedean Clayton copula. Genest et al. (2012) and Quessy and Bahraoui (2013) illustrated their tests for detecting asymmetry of the underlying copula for all pairs of daily intake. In the following, we investigate the fit of our proposed two-level hierarchical copula to the following four daily intakes: Ca, Ir, Pr and VitA.

Figure 3.10 shows both the pairwise scatterplots and density contour plots of pseudo-observations. The latter correspond to the ranks of the selected variables within each sample divided by $m+1$. We observe that for each pair, the bivariate empirical copula has more dependence in its lower tail than its upper tail, which is similar to the shape of the symmetric BB1 copula (which is also the outer power transformation of the Clayton copula). Also, we notice lack of exchangeability and hence we propose to fit the data to a two-level copula of type (3.13) with the two-parameter LST of gamma stopped positive stable distribution.

Applying Perreault et al. (2019)'s clustering procedure, we did not succeed in identifying any grouping of variables. This may be due to low differences in the values of Kendall's tau. Hence, we rather used the second structure determination approach discussed in Section 3.4.1. Using the function *silhouette* implemented in the `R` package *cluster*, we deduced a silhouette coefficient equal to 2. Given the empirical Kendall's tau values, the PAM algorithm with $d = 2$ identified clusters (U_{Ca}, U_{VitA}) and (U_{Ir}, U_{Pr}). Figure 3.11 shows the estimated two-level tree structure. We identify four asymmetric pairs, namely (U_{Ca}, U_{Ir}), (U_{Ca}, U_{Pr}), (U_{VitA}, U_{Ir}) and (U_{VitA}, U_{Pr}), and two symmetric pairs, namely (U_{Ir}, U_{Pr}) and (U_{Ca}, U_{VitA}), which are all in agreement with the test of symmetry done by Genest

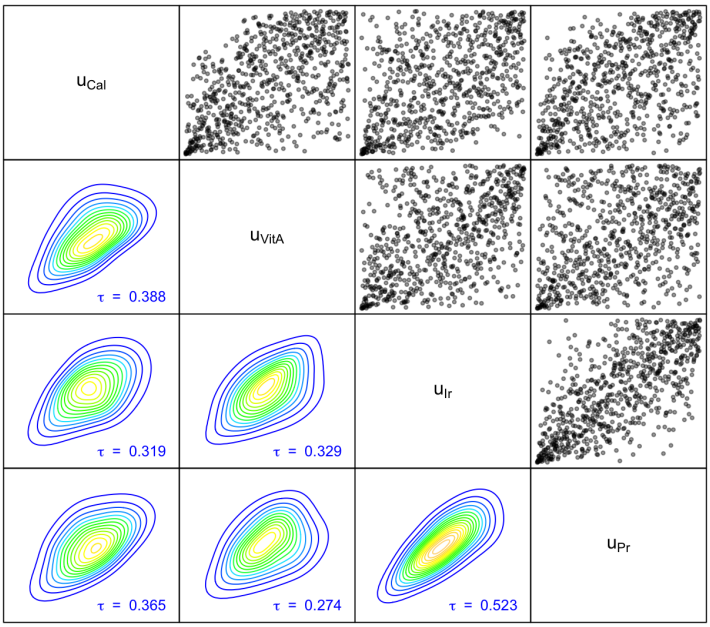


Figure 3.10 – Upper triangular matrix: pairwise scatterplots of Nutrient data, and lower triangular matrix: empirical density contour plots of bivariate normalized copula for all pairs

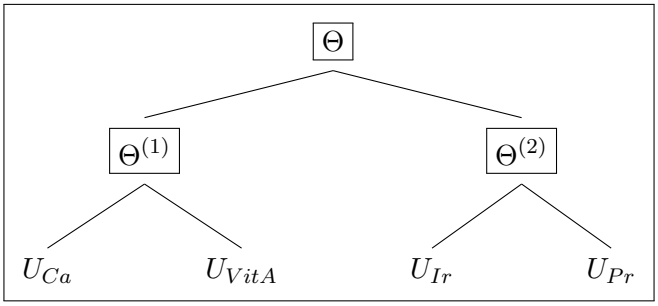


Figure 3.11 – Tree structure derived from the PAM algorithm with the empirical Kendall's tau values of Nutrient dataset

et al. (2012), except for the last pair. Based on this approximated tree structure, we assume that the

4-dimensional hierarchical copula of $\mathbf{U} = (U_{Ca}, U_{VitA}, U_{Ir}, U_{Pr})$ is given by

$$
\begin{aligned}
&C_{\mathbf{U}}\left(u_{Ca}, u_{VitA}, u_{Ir}, u_{Pr}\right) \\
&= \left(1 + \left(\left(u_{Ca}^{-\frac{1}{\delta_{(1)}}} - 1\right)^{\beta} + \left(u_{VitA}^{-\frac{1}{\delta_{(1)}}} - 1\right)^{\beta} + \left(u_{Ir}^{-\frac{1}{\delta_{(2)}}} - 1\right)^{\beta} + \left(u_{Pr}^{-\frac{1}{\delta_{(2)}}} - 1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\delta_{(0)}} \\
&\times \left(1 + \left(\left(u_{Ca}^{-\frac{1}{\delta_{(1)}}} - 1\right)^{\beta} + \left(u_{VitA}^{-\frac{1}{\delta_{(1)}}} - 1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\left(\delta_{(1)} - \delta_{(0)}\right)} \\
&\times \left(1 + \left(\left(u_{Ir}^{-\frac{1}{\delta_{(2)}}} - 1\right)^{\beta} + \left(u_{Pr}^{-\frac{1}{\delta_{(2)}}} - 1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\left(\delta_{(2)} - \delta_{(0)}\right)}. \qquad (3.22)
\end{aligned}
$$

Table 3.7 – Estimated parameters with Nutrient data using steps 1 to 6 of Algorithm 5

Parameters	$\hat{\delta}_{(0)}$	$\hat{\delta}_{(1)}$	$\hat{\delta}_{(2)}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{(1)}$	$\hat{\beta}_{(2)}$
Value	0.8891	1.0200	0.9768	1.0853	1.0667	1.3324

Table 3.8 – Estimated parameters with Nutrient data using results from Table 3.7 and step 8 of Algorithm 5

Parameters	$\hat{\delta}_{(0)}$	$\hat{\delta}_{(1)}$	$\hat{\delta}_{(2)}$
Value	0.7545	1.2177	0.7623

Applying steps 1 to 6 of Algorithm 5, we obtain the estimated parameters shown in Table 3.7. Therefore, we choose $\hat{\beta} = 1.1615$, and then from step 8 of Algorithm 5 the estimates of the parameters of (3.22) are finally given in Table 3.8. Based on these last estimation results and tree structure in Figure 3.11, the Kendall's tau matrix derived from the two-level hierarchical copula model is given in Figure 3.13a. For the pair (U_{Ca}, U_{VitA}), the estimated copula allows to obtain a Kendall's tau close to the one of the real data. However, for the pair (U_{Ir}, U_{Pr}), the model underestimates a little bit its empirical Kendall's tau.

To further adjust the estimated hierarchical copula to the real pseudo-observations, we propose to add an additional step to our Algorithm 5 that takes into account a better loss function to calibrate β than individual composite likelihoods. For this purpose, we introduce a L_2-measure between empirical pair-copulas and those derived from our model defined by

$$
L_2^{pair}(\beta) = \sum_{i=1}^{3} \sum_{j=i+1}^{4} \sum_{k=1}^{m} \left(C_{U^i,U^j}\left(u_k^i, u_k^j; \hat{\boldsymbol{\delta}}, \beta\right) - \hat{C}_{U^i,U^j}\left(u_k^i, u_k^j\right)\right)^2, \qquad (3.23)
$$

where C_{U^i,U^j} is the theoretical pair-copula between U^i and U^j given the vector of estimated parameters $\hat{\boldsymbol{\delta}}$, and $\hat{C}_{U^i,U^j}\left(u^i, u^j\right) = \frac{1}{m}\sum_{h=1}^{m} \mathbb{1}_{\left\{u_h^i \leq u^i, u_h^j \leq u^j\right\}}$ is the empirical pair-copula between U^i and U^j. Then, we choose $\hat{\beta}$ as the value that minimizes $L_2^{pair}(\beta)$ and obtain $\hat{\beta} = 1.2210$. The new Kendall's tau matrix with the new value for β is also given in Figure 3.13b. One sees that this choice provides better approximation for the rank correlations within the dataset. Figure 3.12 illustrates the scatterplots of

737 realizations of $(U_{Ca}, U_{VitA}, U_{Ir}, U_{Pr})$ with the estimated two-level hierarchical copula (3.22). We note a strong similarity between the shapes of bivariate copulas in both Figures 3.10 and 3.12.

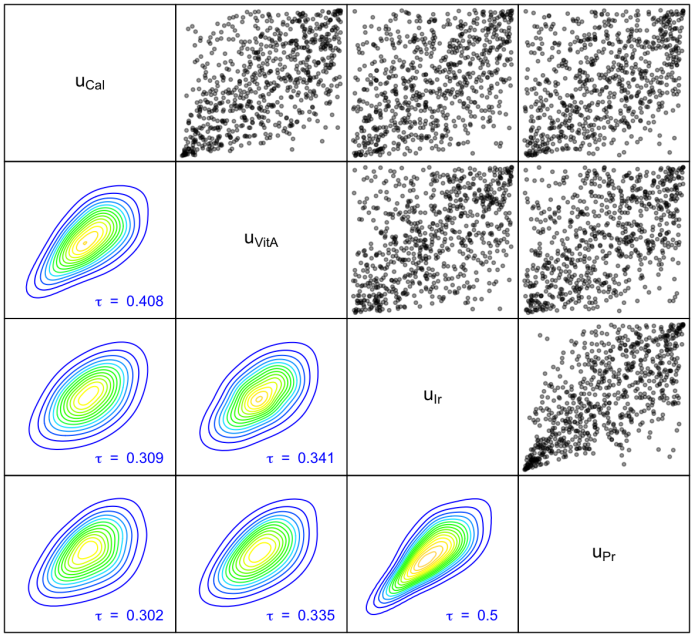


Figure 3.12 – Upper triangular matrix: pairwise scatterplots of 737 observations from the tree in Figure 3.11 and the two-level copula (3.22), and lower triangular matrix: empirical density contour plots of bivariate normalized copula for all pairs

Figure 3.13 – Theoretical Kendall's tau matrices according to the hierarchical copula (3.22), with (a) $\hat{\beta} = 1.1615$ and (b) $\hat{\beta} = 1.2210$

(a)

$$\begin{array}{c} \begin{array}{cccc} U_{Ca} & U_{VitA} & U_{Ir} & U_{Pr} \end{array} \\ \begin{bmatrix} . & 0.3896 & 0.3062 & 0.3062 \\ . & . & 0.3062 & 0.3062 \\ . & . & . & 0.4800 \\ . & . & . & . \end{bmatrix} \end{array}$$

(b)

$$\begin{array}{c} \begin{array}{cccc} U_{Ca} & U_{VitA} & U_{Ir} & U_{Pr} \end{array} \\ \begin{bmatrix} . & 0.4194 & 0.3243 & 0.3243 \\ . & . & 0.3243 & 0.3243 \\ . & . & . & 0.5054 \\ . & . & . & . \end{bmatrix} \end{array}.$$

To check the performance of our hierarchical copula approach to fit the Nutrient dataset, we propose to compare it with vine copula models. Initially proposed by Joe (1996), vine copula models are built up based on a cascade decomposition of a multivariate density into a product of conditional pair-copula densities. They can be represented as a sequence of trees, where at each node, an arbitrary bivariate copula can be used to model the conditional dependence. Thus, they are flexible and can take into account asymmetries and tail dependence. However, they may not perform as well as hierarchical models when data reveal hierarchical structure dependencies. For more details about vine copulas,

see, e.g., Bedford and Cooke (2002), Joe and Kurowicka (2011) and Czado (2019). In the following, we focus on the two classes C-vine and D-vine copulas.

Figures 3.14 and 3.15 show the trees of the estimated C-vine and D-vine copulas, respectively, obtained using the `R` package *VineCopula*. Figure 3.16 illustrates simulations from these estimated vine copulas. According to tree $\mathcal{T}_1$ of the C-vine copula, we note that symmetric pair-copulas were selected for (U_{Ca}, U_{Ir}) and (U_{VitA}, U_{Ir}), while Figure 3.10 shows asymmetries for these pairs. In addition, tree $\mathcal{T}_1$ of the D-vine copula shows a symmetric copula for the pair (U_{Ca}, U_{Ir}), while it was proven to be asymmetric in Genest et al. (2012). Thus, even if vines provide rank correlations very close to those of the real data, the selected pair-copulas do not necessarily model perfectly the symmetric or asymmetric structure of the bivariate distributions.

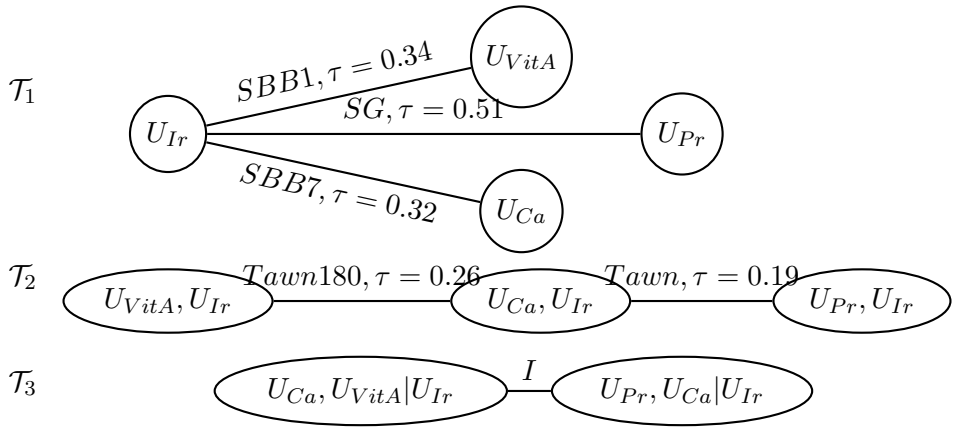


Figure 3.14 – Trees of the estimated 4-dimensional C-vine copula for the Nutrient data, with pair-copula families and Kendall's tau values (See Appendix E for a list of copula family names)

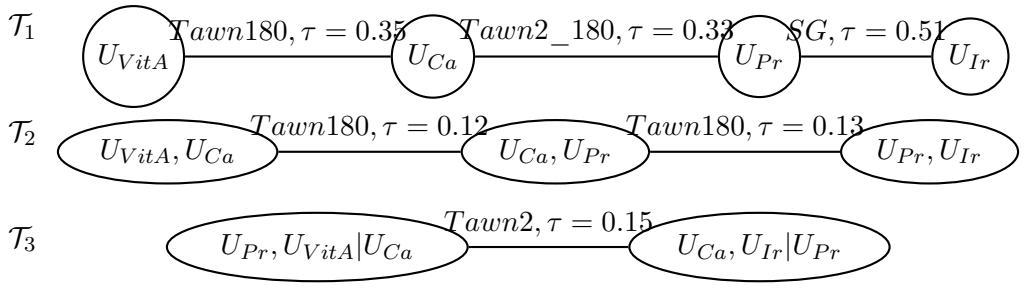


Figure 3.15 – Trees of the estimated D-vine copula for the Nutrient data, with pair-copula families and Kendall's tau values (See Appendix E for a list of copula family names)

The comparison between the vine copulas and our two-level hierarchical copula can not be made with the Akaike information criterion since our model is not estimated through the maximum likelihood method. To compare the copula models, we propose to use the L_2-measure to evaluate the distance

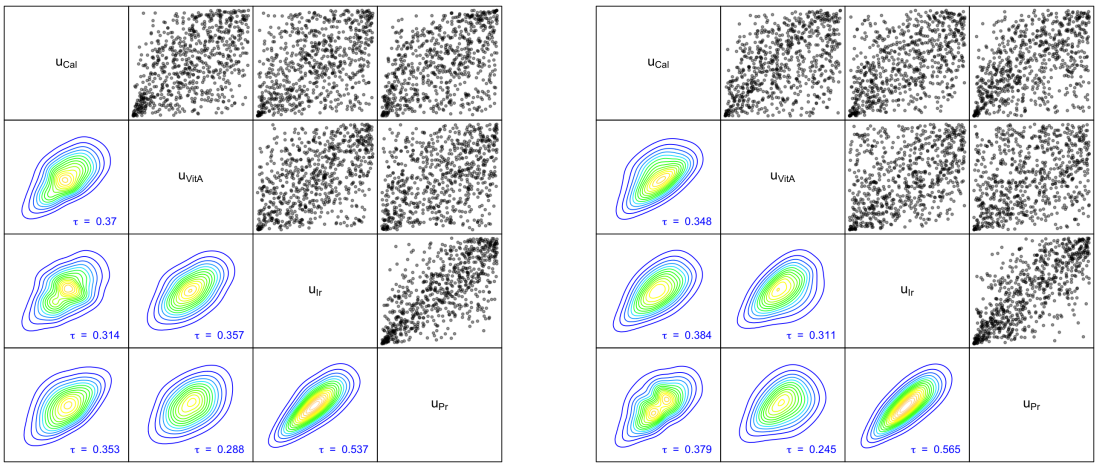


Figure 3.16 – Left panel: 4-dimensional C-vine copula with trees in Figure 3.14, and right panel: 4-dimensional D-vine copula with trees in Figure 3.15. Upper triangular matrices: pairwise scatterplots of 737 observations, and lower triangular matrices: empirical density contour plots of bivariate normalized copula for all pairs

between the estimated 4-dimensional copulas and the empirical copula of the Nutrient data, using

$$L_2(\mathbf{U}_k, k = 1, \ldots, m) = \sum_{k=1}^{m} \left(C_{\mathbf{U}}(\mathbf{u}_k) - \hat{C}_{\mathbf{U}}(\mathbf{u}_k) \right)^2, \tag{3.24}$$

with $C_{\mathbf{U}}$ and $\hat{C}_{\mathbf{U}}$, respectively, the estimated parametric and empirical copulas. Note that for the C-vine and D-vine copulas, $C_{\mathbf{U}}$ is approximated by the empirical copula estimated over 1,000,000 simulated observations. From Table 3.9, we can see that the three copulas are close to the real data structure and that the C-vine copula has the smallest measure. However, when also taking into account the number of parameters (as well as the sets of families of pair-copulas which are used by the `R` package *VineCopula*), there is a significant advantage to consider the hierarchical copula to avoid overfitting.

Table 3.9 – L_2-measure for models fitted to Nutrient data

Copula model	Hierarchical copula	C-vine copula	D-vine copula
Number of parameters	4	9	11
L_2-measure	0.0649	0.0438	0.0629

In order to test the stability and the efficiency of the considered procedure to estimate the two-level hierarchical copula, we performed 200 replicates of 737 simulations of $\mathbf{U}$. At each replicate, we estimated both the tree structure and the parameters. Over all samples, we succeeded in obtaining the same estimated tree. The resulting confidence intervals and standard deviation of the parameters are given in Table 3.10.

Table 3.10 – Estimated parameters with simulated data from the approximated structure of Nutrient data

Parameters	Value	$q_{0.25}$	Mean	$q_{0.75}$	Std
β	1.2210	1.2044	1.2273	1.2500	0.0381
$\delta_{(0)}$	0.7545	0.7123	0.7518	0.7879	0.0626
$\delta_{(1)}$	1.2177	1.1381	1.2363	1.3187	0.1417
$\delta_{(2)}$	0.7623	0.7316	0.7736	0.8128	0.0633

3.4.4 Air quality data

Extracted from Paris region air quality datasets during year 2019 (see https://www.airparif.asso.fr/), NO_2 concentrations provide a good example where our hierarchical structure fits very well data. For several locations in Paris and in its suburbs, hourly NO_2 values in the air were continuously measured during the whole year, for a total of 8760 observations. We dropped some air quality monitoring stations and some hours of observation with missing values and we kept 4 stations and 7847 observations. These stations are: Gonesse (GN), Paris 12e (PR_12), Boulevard Haussmann (BH) and Route Nationale 2 Pantin (RN). Since we have a large dataset, we decided to split it into a training dataset and a testing dataset stratified randomly with proportion 75% and 25% in order to compare in all honesty our approach with the vine copula approach on the testing dataset. We first investigate the estimation of the different copula models on the training dataset with $m = 5885$ pseudo-observations.

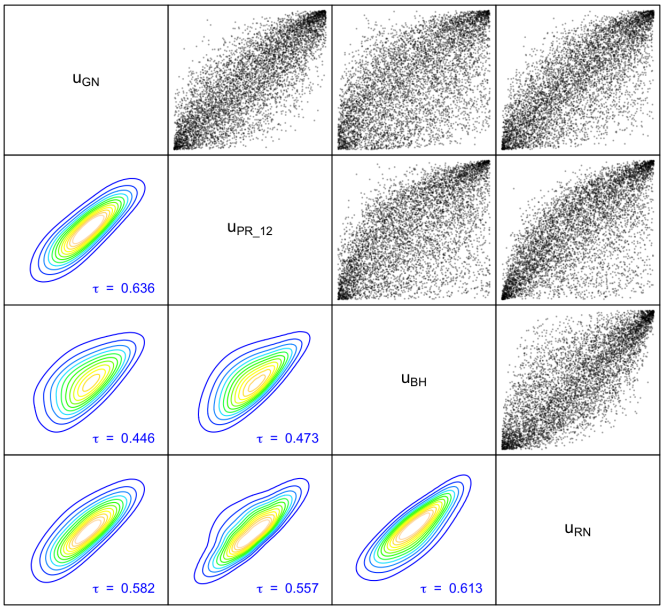


Figure 3.17 – Upper triangular matrix: pairwise scatterplots of NO_2 training dataset, and lower triangular matrix: empirical density contour plots of bivariate normalized copula for all pairs

We can identify from Figure 3.17 the symmetric pairs (U_{GN}, U_{PR_12}) and (U_{BH}, U_{RN}), and the asymmetric ones. Moreover, we observe that for symmetric pairs, the bivariate empirical copula has strong upper and lower tail dependence and a shape similar to the one of the symmetric BB1 copula. The asymmetric pair-copulas have a shape similar to the one shown in Figure 3.5.

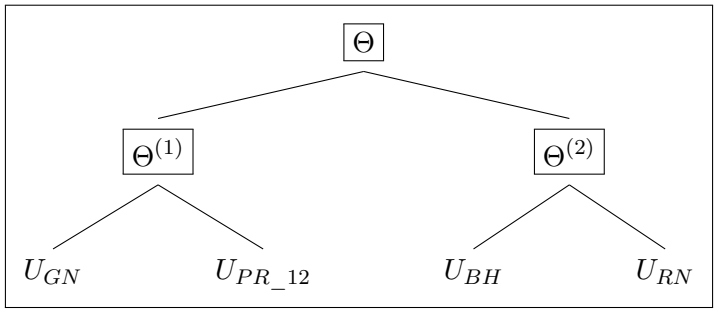


Figure 3.18 – Tree structure derived from the PAM algorithm with the empirical Kendall's tau values of NO_2 training dataset

We therefore propose to fit the training dataset with a two-level hierarchical copula of type (3.13) whose LST belongs to the gamma stopped positive stable family. Given the empirical Kendall's tau values and an estimated silhouette coefficient equal to 2, the PAM algorithm identified clusters (U_{GN}, U_{PR_12}) and (U_{BH}, U_{RN}) (which is consistent with the fact that these pairs are symmetric in our model). Based on this result, we build the estimated two-level tree structure shown in Figure 3.18, and we assume that the 4-dimensional hierarchical copula of $\mathbf{U} = \left(U_{GN}, U_{PR\_12}, U_{BH}, U_{RN}\right)$ is given by

$$
\begin{aligned}
& C_{\mathbf{U}}\left(u_{GN}, u_{PR\_12}, u_{BH}, u_{RN}\right) \\
= & \left(1+\left(\left(u_{GN}^{-\frac{1}{\delta_{(1)}}}-1\right)^{\beta}+\left(u_{PR\_12}^{-\frac{1}{\delta_{(1)}}}-1\right)^{\beta}+\left(u_{BH}^{-\frac{1}{\delta_{(2)}}}-1\right)^{\beta}+\left(u_{RN}^{-\frac{1}{\delta_{(2)}}}-1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\delta_{(0)}} \\
& \times\left(1+\left(\left(u_{GN}^{-\frac{1}{\delta_{(1)}}}-1\right)^{\beta}+\left(u_{PR\_12}^{-\frac{1}{\delta_{(1)}}}-1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\left(\delta_{(1)}-\delta_{(0)}\right)} \\
& \times\left(1+\left(\left(u_{BH}^{-\frac{1}{\delta_{(2)}}}-1\right)^{\beta}+\left(u_{RN}^{-\frac{1}{\delta_{(2)}}}-1\right)^{\beta}\right)^{\frac{1}{\beta}}\right)^{-\left(\delta_{(2)}-\delta_{(0)}\right)} .
\end{aligned}
\tag{3.25}
$$

Table 3.11 – Estimated parameters with NO_2 training dataset using steps 1 to 6 of Algorithm 5

Parameters	$\hat{\delta}_{(0)}$	$\hat{\delta}_{(1)}$	$\hat{\delta}_{(2)}$	$\hat{\beta}_{(0)}$	$\hat{\beta}_{(1)}$	$\hat{\beta}_{(2)}$
Value	3.0223	3.0223	4.1930	2.2151	2.1439	2.1854

By applying the first phase of parameter estimation, involving steps 1 to 6 of Algorithm 5, we obtain parameter estimates given in Table 3.11. Note that the three β estimations have very close values.

Table 3.12 – Estimated parameters with NO_2 training dataset using results from Table 3.11 and step 8 of Algorithm 5

Parameters	$\hat{\delta}_{(0)}$	$\hat{\delta}_{(1)}$	$\hat{\delta}_{(2)}$
Value	3.2397	3.2398	4.1544

Given $\hat{\beta} = \text{mean}\left(\hat{\beta}_{(0)}, \hat{\beta}_{(1)}, \hat{\beta}_{(2)}\right) = 2.1815$, from step 8 of Algorithm 5, the adjusted parameters of (3.25) are then given in Table 3.8. With these parameters, the hierarchical copula (3.25) leads to a dependence structure with Kendall's tau matrix given in Figure 3.19a. To further improve the calibration of the Kendall's tau matrix, we solve the minimization problem

$$\underset{\beta}{\text{argmin}}\ L_2^{pair}(\beta),$$

with L_2^{pair} defined by (3.23). Then, we obtain $\hat{\beta} = 2.3122$, and deduce the theoretical Kendall's tau matrix given in Figure 3.19b. Figure 3.20 illustrates the scatterplots of 5885 realizations of $\left(U_{GN}, U_{PR\_12}, U_{BH}, U_{RN}\right)$ with the estimated two-level hierarchical copula (3.25).

Figure 3.19 – Kendall's tau matrices according to Copula (3.25), with (a) $\hat{\beta} = 2.1815$ and (b) $\hat{\beta} = 2.3122$

(a)

	U_{GN}	U_{PR_2}	U_{BH}	U_{RN}
U_{GN}	.	0.6028	0.4965	0.4965
U_{PR_2}	.	.	0.4965	0.4965
U_{BH}	.	.	.	0.5908
U_{RN}	.	.	.	.

(b)

	U_{GN}	U_{PR_2}	U_{BH}	U_{RN}
U_{GN}	.	0.6253	0.5140	0.5140
U_{PR_2}	.	.	0.5140	0.5140
U_{BH}	.	.	.	0.6139
U_{RN}	.	.	.	.

As mentioned previously, we also estimate vine copulas on the NO_2 training dataset. Figures 3.21 and 3.22 represent, respectively, the sequence of trees for the estimated C-vine and D-vine copulas. Rank correlations appear to be close to those of the real data. However, from Figure 3.23, the simulated pair-copulas do not have similar shapes as those of the empirical pair-copulas in Figure 3.17. In particular, we observe less asymmetry patterns in Figure 3.23 compared to Figure 3.17.

As in the previous case study and for the same reasons, we use the L_2-measure given in (3.24). Table 3.13 shows the good performance of the two-level hierarchical copula (3.25) compared to the alternative vine copulas, while its number of parameters (4) is much smaller than the number of parameters of its competitors (11). With 1962 new observations, the testing set is now used to observe how each copula model is capable of generalizing on new data. Table 3.13 also shows that the two-level hierarchical copula (3.25) allows a better generalization with a smaller L_2-measure on the testing set.

Eventually, we tried to model the dependency structure within the training dataset with a HAC ignoring the observed asymmetries. Hence, using the `R` package *HAC*, the resulting estimated HAC has the same tree structure as ours and and has for family of generators the family of stable distributions.

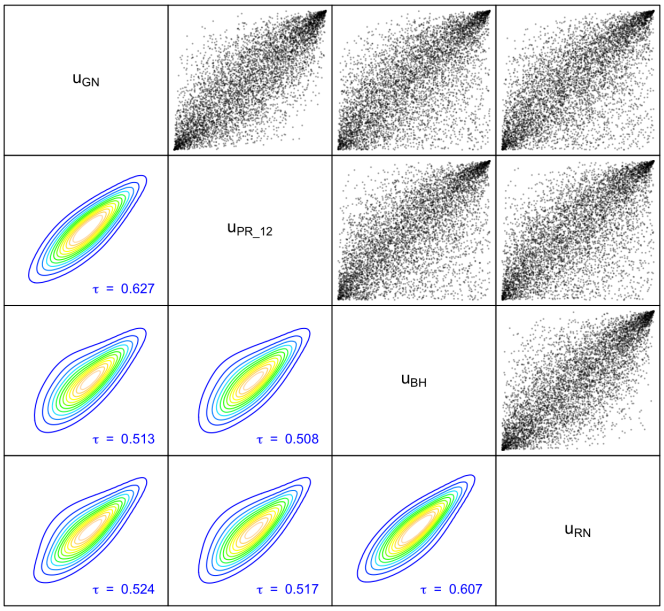


Figure 3.20 – Upper triangular matrix: pairwise scatterplots of 5885 observations from the estimated two-level copula (3.25), and lower triangular matrix: empirical density contour plots of bivariate normalized copula for all pairs

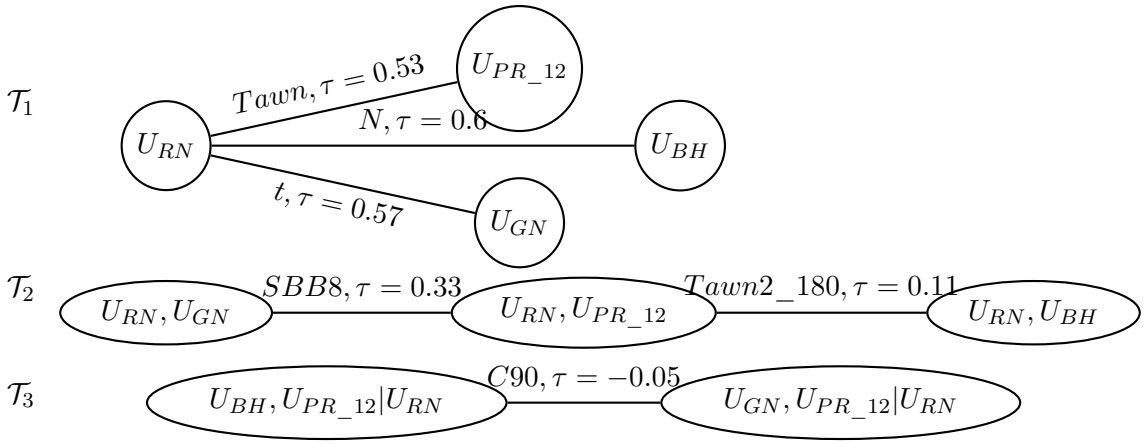


Figure 3.21 – Trees of the estimated C-vine copula for NO_2 training dataset, with pair-copula families and Kendall's tau values (See Appendix E for a list of copula family names)

With three estimated parameters, we obtained $L_2 = 1.5109$ for the training dataset, which is a considerable larger value compared to the values in Table 3.13.

We carried out 200 replicates with 5885 simulations of $(U_{GN}, U_{PR\_12}, U_{BH}, U_{RN})$, in order to confirm both the stability of the considered hierarchical copula and the performance of Algorithm 5. Over all the replicates, we obtained the same estimated tree structure. The resulting confidence intervals and

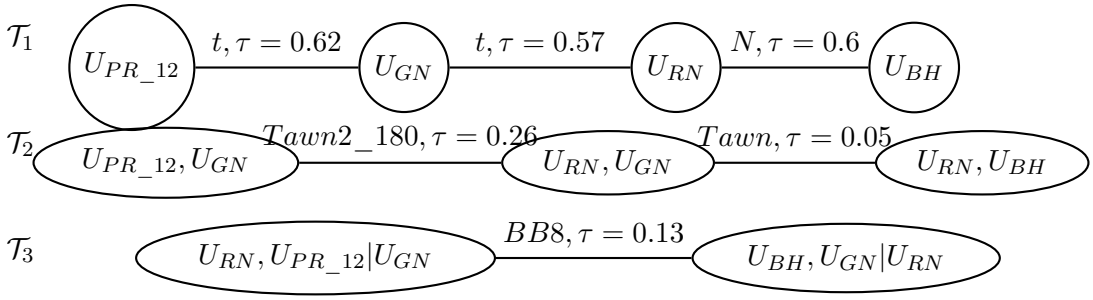


Figure 3.22 – Trees of the estimated D-vine copula for NO_2 training dataset, with pair-copula families and Kendall's tau values (See Appendix E for a list of copula family names)

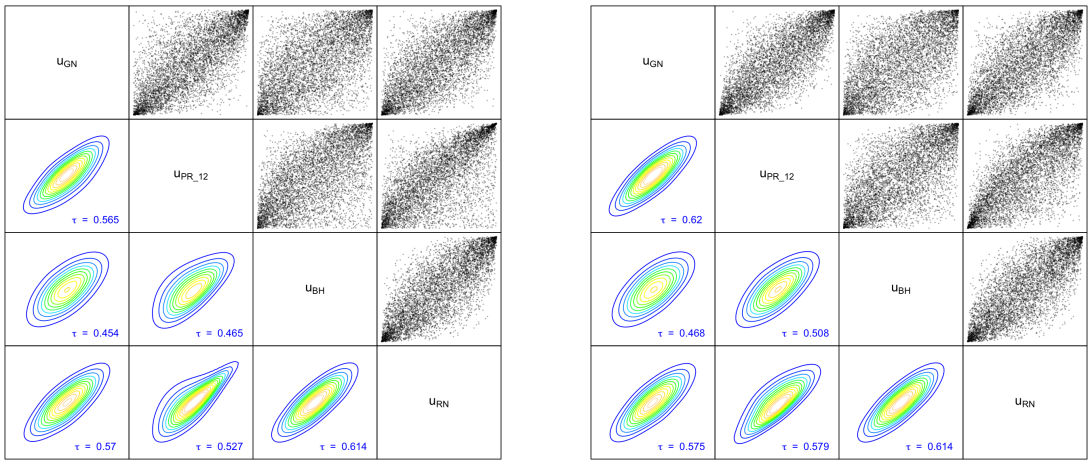


Figure 3.23 – Left panel: 4-dimensional C-vine copula with trees in Figure 3.21, and right panel: 4-dimensional D-vine copula with trees in Figure 3.22. Upper triangular matrices: pairwise scatterplots of 5885 observations, and lower triangular matrices: empirical density contour plots of bivariate normalized copula for all pairs

Table 3.13 – L_2-measure for copula models fitted to NO_2 datasets

Model	hierarchical copula	C-vine copula	D-vine copula
Number of parameters	4	11	11
L_2-measure for training dataset	0.3064	0.3454	0.3129
L_2-measure for testing dataset	0.0735	0.0825	0.2000

standard deviation of the estimated parameters are given in Table 3.14.

3.5 Conclusion

In this paper, we propose a new hierarchical block-exchangeable structure leading to non-exchangeable pair-copulas between-blocks and Archimedean copulas within-blocks. We focus on the particular case of a two-level structure for which we investigate the pair dependence properties and propose a full

Table 3.14 – Estimated parameters with simulated data from the approximated structure of NO_2 training dataset

Parameters	Value	$q_{0.25}$	Mean	$q_{0.75}$	Std
β	2.3122	2.2847	2.3175	2.3500	0.0420
$\delta_{(0)}$	3.2397	3.0509	3.2476	3.4417	0.2794
$\delta_{(1)}$	3.2398	3.1180	3.2957	3.4839	0.2716
$\delta_{(2)}$	4.1544	3.7784	4.2839	4.7525	0.6600

estimation procedure with several illustrations.

For a multi-level structure, the question of the identification of the tree structure is still an open problem, although the proposed estimation procedure by a bottom-up composite likelihood approach is easily generalized.

Acknowledgments

We would like to thank two reviewers and an Associate Editor for all their comments which considerably helped to improve the paper. The authors also thank Christian Genest and Johanna Nešlehová for providing us the Nutrient dataset. We also thank Samuel Perreault for his helpful explanations regarding the use of the algorithm presented in Perreault et al. (2019). The authors gratefully acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada (Cossette: 04273; Marceau: 05605), from the Chaire en actuariat de l'Université Laval (Chaoubi, Cossette, and Marceau: FO502323), and from the Chaire d'Excellence DAMI (Robert).

Appendix

A Detailed developments

- Details on the derivation of (3.5) are the following ones: For $\mathbf{t}_1 = \left(t^{\mathbf{i}_1} : \mathbf{i}_1 \in \mathcal{I}_1\right)$, we have

$$\begin{aligned}
\mathcal{L}_{\boldsymbol{\Theta}_1}(\mathbf{t}_1) &= \mathrm{E}\left[\prod_{\mathbf{i}_1 \in \mathcal{I}_1} \exp\left(-t^{\mathbf{i}_1}\Theta^{\mathbf{i}_1}\right)\right] \\
&= \mathrm{E}\left[\exp\left(-\sum_{\mathbf{i}_1 \in \mathcal{I}_1} t^{\mathbf{i}_1}\Theta\right)\right] \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \mathrm{E}\left[\exp\left(-t^{\mathbf{i}_1}N^{\mathbf{i}_1}\right)\right] \\
&= \mathcal{L}_{\Theta}\left(\|\mathbf{t}_1\|\right) \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \mathcal{L}_{N^{\mathbf{i}_1}}\left(t^{\mathbf{i}_1}\right) \\
&= \psi\left(\|\mathbf{t}_1\|; \alpha_{\mathbf{i}_0}\right) \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \psi\left(t^{\mathbf{i}_1}; \alpha_{\mathbf{i}_1}\right).
\end{aligned}$$

- Details on the derivation of (3.6) are the following ones: For $\mathbf{t}_2 = \left(t^{\mathbf{i}_2} : \mathbf{i}_2 \in \mathcal{I}_2\right)$, we have

$$\begin{aligned}
\mathcal{L}_{\mathbf{\Theta}_2}\left(\mathbf{t}_2\right) &= \mathrm{E}\left[\prod_{\mathbf{i}_2 \in \mathcal{I}_2} \exp\left(-t^{\mathbf{i}_2}\Theta^{\mathbf{i}_2}\right)\right] \\
&= \mathrm{E}\left[\exp\left(-\sum_{\mathbf{i}_2 \in \mathcal{I}_2} t^{\mathbf{i}_2}\Theta\right)\right] \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \mathrm{E}\left[\exp\left(-\sum_{\mathbf{i}_2 \in \mathcal{A}_{2,\mathbf{i}_1}} t^{\mathbf{i}_2} N^{\mathbf{i}_1}\right)\right] \\
&\quad \times \prod_{\mathbf{i}_2 \in \mathcal{I}_2} \mathrm{E}\left[\exp\left(-t^{\mathbf{i}_2} N^{\mathbf{i}_2}\right)\right] \\
&= \mathcal{L}_\Theta\left(\|\mathbf{t}_2\|\right) \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \mathcal{L}_{N^{\mathbf{i}_1}}\left(\|\mathbf{t}_{2,\mathbf{i}_1}\|\right) \times \prod_{\mathbf{i}_2 \in \mathcal{I}_2} \mathcal{L}_{N^{\mathbf{i}_2}}\left(t^{\mathbf{i}_2}\right) \\
&= \psi\left(\|\mathbf{t}_2\|\,; \alpha_{\mathbf{i}_0}\right) \times \prod_{\mathbf{i}_1 \in \mathcal{I}_1} \psi\left(\|\mathbf{t}_{2,\mathbf{i}_1}\|\,; \alpha_{\mathbf{i}_1}\right) \times \prod_{\mathbf{i}_2 \in \mathcal{I}_2} \psi\left(t^{\mathbf{i}_2}; \alpha_{\mathbf{i}_2}\right).
\end{aligned}$$

- Details on the derivation of (3.8) are the following ones: For $\mathbf{y} = (y^{(\mathbf{i}_{l-1},j)} : j = 1, \ldots, a_{\mathbf{i}_{l-1}}, \underline{i}_{l-1} \in \mathcal{I}_{l-1})$, we have

$$\begin{aligned}
\overline{F}_{\mathbf{Y}}\left(\mathbf{y}\right) &= \mathrm{E}\left[\overline{F}_{\mathbf{Y}|\mathbf{\Theta}_{l-1}}\left(\mathbf{y}\right)\right] \\
&= \mathrm{E}\left[\prod_{\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}} \prod_{j=1}^{a_{\mathbf{i}_{l-1}}} \overline{F}_{Y^{(\mathbf{i}_{l-1},j)}|\Theta^{\mathbf{i}_{l-1}}}\left(y^{(\mathbf{i}_{l-1},j)}\right)\right] \\
&= \mathrm{E}\left[\prod_{\mathbf{i}_{l-1} \in \mathcal{I}_{l-1}} \prod_{j=1}^{a_{\mathbf{i}_{l-1}}} \exp\left(y^{(\mathbf{i}_{l-1},j)}\Theta^{\mathbf{i}_{l-1}}\right)\right] \\
&= \mathcal{L}_{\mathbf{\Theta}_{l-1}}\left(\left(\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} y^{(\mathbf{i}_{l-1},j)} : \mathbf{i}_{l-1} \in \mathcal{I}_{l-1}\right)\right)\right).
\end{aligned}$$

B Proof of Proposition 2

(i) For $\mathbf{u}^{\mathbf{i}_{l-1}} = \left(u^{(\mathbf{i}_{l-1},j)}, j = 1, \ldots, a_{\mathbf{i}_{l-1}}\right)$, we have

$$\begin{aligned}
C_{\mathbf{U}^{\mathbf{i}_{l-1}}}\left(\mathbf{u}^{\mathbf{i}_{l-1}}\right) &= C_{\mathbf{U}}\left(\tilde{\mathbf{u}}^{\mathbf{i}_{l-1}}\right) \\
&= \psi\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} \psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right); \alpha_{\mathbf{i}_0}\right) \\
&\quad \times \prod_{k=1}^{l-1} \psi\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} \psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right); \alpha_{\mathbf{i}_{l-1|k}}\right) \\
&= \psi\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} \psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right); \alpha_{\mathbf{i}_0} + \sum_{k=1}^{l-1} \alpha_{\mathbf{i}_{l-1|k}}\right) \\
&= \psi\left(\sum_{j=1}^{a_{\mathbf{i}_{l-1}}} \psi^{-1}\left(u^{(\mathbf{i}_{l-1},j)}; \delta_{\mathbf{i}_{l-1}}\right); \delta_{\mathbf{i}_{l-1}}\right),
\end{aligned}$$

with $\tilde{\mathbf{u}}^{\mathbf{i}'_{l-1}} = \begin{cases} \mathbf{u}^{\mathbf{i}_{l-1}} & \text{if } \mathbf{i}_{l-1} = \mathbf{i}'_{l-1}, \\ \mathbf{1}_{a_{\mathbf{i}'_{l-1}}} & \text{if } \mathbf{i}_{l-1} \neq \mathbf{i}'_{l-1}, \end{cases}$ and $\mathbf{1}_{a_{\mathbf{i}'_{l-1}}} = (1, \ldots, 1) \in \mathbb{R}^{a_{\mathbf{i}'_{l-1}}}$.

(ii) The multivariate survival function of $\mathbf{Y}^{\{\sigma\}}$ can be expressed as

$$\begin{aligned}
\overline{F}_{\mathbf{Y}^{\{\sigma\}}}\left(\mathbf{y}^{\{\sigma\}}\right) &= \mathrm{E}\left[\overline{F}_{\mathbf{Y}^{\{\sigma\}}|\Theta^{\mathbf{i}_{l-1}^{(1)}},\ldots,\Theta^{\mathbf{i}_{l-1}^{(a_{\mathbf{i}_r})}}}\left(\mathbf{y}^{\{\sigma\}}\right)\right] \\
&= \mathrm{E}\left[\prod_{m=1}^{a_{\mathbf{i}_{l-1}}}\overline{F}_{Y^{\sigma^{(m)}}|\Theta^{\mathbf{i}_{l-1}^{(m)}}}\left(y^{\sigma^{(m)}}\right)\right] \\
&= \mathrm{E}\left[\prod_{m=1}^{a_{\mathbf{i}_{l-1}}}\exp\left(-y^{\sigma^{(m)}}\Theta^{\mathbf{i}_{l-1}^{(m)}}\right)\right] \\
&= \mathrm{E}\left[\exp\left(-\left\|\mathbf{y}^{\{\sigma\}}\right\|\Theta^{\mathbf{i}_r}\right)\right]\times\prod_{m=1}^{a_{\mathbf{i}_r}}E\left[\exp\left(-y^{\sigma^{(m)}}\sum_{k=r+1}^{l-1}N^{\mathbf{i}_{l-1|k}^{(m)}}\right)\right] \\
&= \psi\left(\left\|\mathbf{y}^{\{\sigma\}}\right\|;\delta_{\mathbf{i}_r}\right)\times\prod_{m=1}^{a_{\mathbf{i}_r}}\psi\left(y^{\sigma^{(m)}};\sum_{k=r+1}^{l-1}\alpha_{\mathbf{i}_{l-1|k}^{(m)}}\right).
\end{aligned}$$

Using Sklar's theorem, we deduce the copula of $\mathbf{U}^{\{\sigma\}}$

$$\begin{aligned}
C_{\mathbf{U}^{\{\sigma\}}}\left(\mathbf{u}^{\{\sigma\}}\right) &= \overline{F}_{\mathbf{Y}^{\{\sigma\}}}\left(\overline{F}_{Y^{\sigma^{(m)}}}^{-1}\left(u^{\sigma^{(m)}}\right),m\in\{1,\ldots,a_{\mathbf{i}_r}\}\right) \\
&= \psi\left(\sum_{m=1}^{a_{\mathbf{i}_r}}\psi^{-1}\left(u^{\sigma^{(m)}};\delta_{\mathbf{i}_{l-1}^{(m)}}\right);\delta_{\mathbf{i}_r}\right) \\
&\quad\times\prod_{m=1}^{a_{\mathbf{i}_r}}\psi\left(\psi^{-1}\left(u^{\sigma^{(m)}};\delta_{\mathbf{i}_{l-1}^{(m)}}\right);\sum_{k=r+1}^{l-1}\alpha_{\mathbf{i}_{l-1|k}^{(m)}}\right) \\
&= \psi\left(\sum_{m=1}^{a_{\mathbf{i}_r}}\psi^{-1}\left(u^{\sigma^{(m)}};\delta_{\mathbf{i}_{l-1}^{(m)}}\right);\delta_{\mathbf{i}_r}\right) \\
&\quad\times\prod_{m=1}^{a_{\mathbf{i}_r}}\psi\left(\psi^{-1}\left(u^{\sigma^{(m)}};\delta_{\mathbf{i}_{l-1}^{(m)}}\right);\delta_{\mathbf{i}_{l-1}^{(m)}}-\delta_{\mathbf{i}_r}\right),
\end{aligned}$$

with $\mathbf{u}^{\{\sigma\}}=\left(u^{\sigma^{(m)}}:m\in\{1,\ldots,a_{\mathbf{i}_r}\}\right)\in[0,1]^{a_{\mathbf{i}_r}}$ and $\sigma\in\mathcal{S}_{\mathbf{i}_r}$.

C Proof of Proposition 3

It is easily seen that the condition

$$C_{\delta_1,\delta_1,\delta_1}(u_1,u_2)\geq C_{\delta_0,\delta_1,\delta_2}(u_1,u_2)\qquad\text{and}\qquad C_{\delta_2,\delta_2,\delta_2}(u_1,u_2)\geq C_{\delta_0,\delta_1,\delta_2}(u_1,u_2),$$

for all $(u_1,u_2)\in[0,1]^2$, is equivalent to the condition that

$$\Psi_{\delta_1}(x,y)\geq\Psi_{\delta_0}\left(\frac{\delta_0}{\delta_1}x,\frac{\delta_0}{\delta_2}y\right)\qquad\text{and}\qquad\Psi_{\delta_2}(x,y)\geq\Psi_{\delta_0}\left(\frac{\delta_0}{\delta_1}x,\frac{\delta_0}{\delta_2}y\right).$$

for all $x,y\geq 0$. Moreover, if we have

$$\Psi_{\delta_1}(x,y)\quad\geq\quad\Psi_{\delta_0}(x,y)\geq\Psi_{\delta_0}\left(\frac{\delta_0}{\delta_1}x,\frac{\delta_0}{\delta_2}y\right)$$

and

$$\Psi_{\delta_2}(x,y)\quad\geq\quad\Psi_{\delta_0}(x,y)\geq\Psi_{\delta_0}\left(\frac{\delta_0}{\delta_1}x,\frac{\delta_0}{\delta_2}y\right),$$

for all $x, y \geq 0$, then (3.19) will be satisfied. But we have $\Psi_{\delta_1}(x,y) \geq \Psi_{\delta_0}(x,y)$ if and only if

$$\delta_1 \varphi\left(\varphi^{-1}\left(\frac{x}{\delta_1}\right) + \varphi^{-1}\left(\frac{y}{\delta_1}\right)\right) \leq \delta_0 \varphi\left(\varphi^{-1}\left(\frac{x}{\delta_0}\right) + \varphi^{-1}\left(\frac{y}{\delta_0}\right)\right),$$

or equivalently, if and only if

$$\varphi_1\left(\varphi_1^{-1}(x) + \varphi_1^{-1}(y)\right) \leq \varphi_0\left(\varphi_0^{-1}(x) + \varphi_0^{-1}(y)\right), \tag{A.26}$$

with $\varphi_1(x) = \delta_1 \varphi(x)$ and $\varphi_0(x) = \delta_0 \varphi(x)$. According to Theorem 4.4.2 of Nelsen (2006), (A.26) is satisfied if and only if $\varphi_1^{-1} \circ \varphi_0$ is subadditive, or equivalently if and only if the function

$$f_{\delta_0,\delta_1}(x) = \varphi^{-1}\left(\frac{\delta_0}{\delta_1}\varphi(x)\right)$$

is subadditive, i.e. $f_{\delta_0,\delta_1}(x+y) \leq f_{\delta_0,\delta_1}(x) + f_{\delta_0,\delta_1}(y)$, for all $x, y \geq 0$.

D Proof of Corollary 1

(i) Let $\psi(t;\delta) = \mathrm{e}^{-\delta\left((h+t)^\lambda - h^\lambda\right)}$ be the LST of the exponential tilted stable distribution, with $\varphi(t) = (h+t)^\lambda - h^\lambda$. Then, we have

$$\Psi_\delta(x,y) = -\delta\left(\left(\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{y}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} - h\right)^\lambda - h^\lambda\right) + x + y,$$

with partial derivative with respect to x given by

$$\begin{aligned}
\frac{\partial}{\partial x}\Psi_\delta(x,y) &= -\delta\lambda\frac{1}{\lambda}\frac{1}{\delta}\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}-1}\left(\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{y}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} - h\right)^{\lambda-1} + 1 \\
&= -\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}-1}\left(\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{y}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} - h\right)^{\lambda-1} + 1.
\end{aligned}$$

Given that $\lambda \in (0,1]$, we have

$$\begin{aligned}
\frac{y}{\delta} + h^\lambda &\geq h^\lambda \\
\left(\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{y}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} - h\right) &\geq \left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} \\
\left(\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{y}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} - h\right)^{\lambda-1} &\leq \left(\frac{x}{\delta} + h^\lambda\right)^{\frac{\lambda-1}{\lambda}} \\
\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}-1}\left(\left(\frac{x}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} + \left(\frac{y}{\delta} + h^\lambda\right)^{\frac{1}{\lambda}} - h\right)^{\lambda-1} &\leq 1.
\end{aligned}$$

Thus, $\frac{\partial}{\partial x}\Psi_\delta(x,y) \geq 0$. With the same manner, we can prove that $\frac{\partial}{\partial y}\Psi_\delta(x,y) \geq 0$. In addition, we have

$$f_{\delta_0,\delta}(x) = \left(\frac{\delta_0}{\delta}\left((h+x)^\lambda - h^\lambda\right) + h^\lambda\right)^{\frac{1}{\lambda}} - h,$$

with the first derivative given by

$$\frac{\partial}{\partial x} f_{\delta_0,\delta}(x) = \frac{\delta_0}{\delta}(h+x)^{\lambda-1}\left(\frac{\delta_0}{\delta}\left((h+x)^\alpha - h^\lambda\right) + h^\lambda\right)^{\frac{1}{\lambda}-1},$$

and the second derivative

$$\begin{aligned}\frac{\partial^2}{\partial x^2} f_{\delta_0,\delta}(x) &= \frac{\delta_0}{\delta}(\lambda-1)(h+x)^{\lambda-2}\left(\frac{\delta_0}{\delta}\left((h+x)^\lambda - h^\lambda\right)+h^\lambda\right)^{\frac{1}{\lambda}-1} \\ &\quad + \frac{\delta_0}{\delta}(h+x)^{\lambda-1}\left(\frac{1}{\lambda}-1\right)\frac{\delta_0}{\delta}\lambda(h+x)^{\lambda-1}\left(\frac{\delta_0}{\delta}\left((h+x)^\lambda - h^\lambda\right)+h^\lambda\right)^{\frac{1}{\lambda}-2} \\ &= \frac{\delta_0}{\delta}(\lambda-1)(h+x)^{\lambda-2}\left(\frac{\delta_0}{\delta}\left((h+x)^\lambda - h^\lambda\right)+h^\lambda\right)^{\frac{1}{\lambda}-2} \times \left(\frac{\delta_0}{\delta}-1\right)h^\lambda.\end{aligned}$$

Thus, $\frac{\partial^2}{\partial x^2} f_{\delta_0,\delta}(x) \leq 0$ since $\lambda \in (0,1]$. Since $f_{\delta_0,\delta}(0) = 0$, we deduce that $f_{\delta_0,\delta}$ is subadditive, for all $x \geq 0$.

(ii) Let $\psi(t;\delta) = (1+t)^{-\delta}$ be the LST of the gamma distribution, with $\varphi(t) = \ln(1+t)$. Then, we obtain, for $x, y \geq 0$,

$$\begin{aligned}\Psi_\delta(x,y) &= -\delta \ln\left(e^{\frac{x}{\delta}} + e^{\frac{y}{\delta}} - 1\right) + x + y, \\ \frac{\partial}{\partial x}\Psi_\delta(x,y) &= -\frac{e^{\frac{x}{\delta}}}{e^{\frac{x}{\delta}} + e^{\frac{y}{\delta}} - 1} + 1 \geq 0, \\ \frac{\partial}{\partial y}\Psi_\delta(x,y) &= -\frac{e^{\frac{y}{\delta}}}{e^{\frac{x}{\delta}} + e^{\frac{y}{\delta}} - 1} + 1 \geq 0, \\ f_{\delta_0,\delta}(x) &= (1+x)^{\frac{\delta_0}{\delta}} - 1\end{aligned}$$

and

$$\frac{\partial^2}{\partial x^2} f_{\delta_0,\delta}(x) = \frac{\delta_0}{\delta}\left(\frac{\delta_0}{\delta}-1\right)(1+x)^{\frac{\delta_0}{\delta}-2} \leq 0,$$

and we conclude in the same way as in the previous case.

(iii) Let $\psi(t;\delta) = \left(1+t^{\frac{1}{\beta}}\right)^{-\delta}$ be the LST of the gamma stopped positive stable distribution, with $\varphi(t) = \ln\left(1+t^{\frac{1}{\beta}}\right)$. Then, we obtain

$$\Psi_\delta(x,y) = -\delta \ln\left(1 + \left((e^{\frac{x}{\delta}}-1)^\beta + (e^{\frac{y}{\delta}}-1)^\beta\right)^{\frac{1}{\beta}}\right) + x + y,$$

and

$$\frac{\partial}{\partial x}\Psi_\delta(x,y) = -\frac{e^{\frac{x}{\delta}}(e^{\frac{x}{\delta}}-1)^{\beta-1}\left((e^{\frac{x}{\delta}}-1)^\beta + (e^{\frac{y}{\delta}}-1)^\beta\right)^{\frac{1}{\beta}-1}}{1 + \left((e^{\frac{x}{\delta}}-1)^\beta + (e^{\frac{y}{\delta}}-1)^\beta\right)^{\frac{1}{\beta}}} + 1.$$

Note that

$$1 + \left((e^{\frac{x}{\delta}}-1)^\beta + (e^{\frac{y}{\delta}}-1)^\beta\right)^{\frac{1}{\beta}} = e^{\frac{x}{\delta}} + \left((e^{\frac{x}{\delta}}-1)^\beta + (e^{\frac{y}{\delta}}-1)^\beta\right)^{\frac{1}{\beta}} - \left((e^{\frac{x}{\delta}}-1)^\beta\right)^{\frac{1}{\beta}} \geq e^{\frac{x}{\delta}},$$

and

$$(e^{\frac{x}{\delta}}-1)^{\beta-1}\left((e^{\frac{x}{\delta}}-1)^\beta + (e^{\frac{y}{\delta}}-1)^\beta\right)^{\frac{1}{\beta}-1} = \left(1 + \frac{(e^{\frac{y}{\delta}}-1)^\beta}{(e^{\frac{x}{\delta}}-1)^\beta}\right)^{\frac{1}{\beta}-1} \leq 1,$$

since $\beta \geq 1$. Then, we conclude that $\frac{\partial}{\partial x}\Psi_\delta(x,y) \geq 0$, and with the same manner $\frac{\partial}{\partial y}\Psi_\delta(x,y) \geq 0$. Moreover, for $x \geq 0$, we have

$$\begin{aligned} f_{\delta_0,\delta}(x) &= \left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{\beta}, \\ \frac{\partial}{\partial x}f_{\delta_0,\delta}(x) &= \frac{\delta_0}{\delta}\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{\beta-1}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}-1} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial x^2}f_{\delta_0,\delta}(x) &= (\beta-1)\frac{1}{\beta}\left(\frac{\delta_0}{\delta}\right)^2\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{\beta-2}\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}-1}\right)^2 \\ &\quad +\frac{1}{\beta}\frac{\delta_0}{\delta}\left(\frac{\delta_0}{\delta}-1\right)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{\beta-1}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-2}\left(x^{\frac{1}{\beta}-1}\right)^2 \\ &\quad +\left(\frac{1}{\beta}-1\right)\frac{\delta_0}{\delta}\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{\beta-1}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}-2}. \end{aligned}$$

We can rewrite the second derivative as

$$\begin{aligned} &\frac{\partial^2}{\partial x^2}f_{\delta_0,\delta}(x) \\ &= \frac{1}{\beta}\frac{\delta_0}{\delta}\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{\beta-2}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}-1}\Bigg[(\beta-1)\frac{\delta_0}{\delta}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}-1} \\ &\quad +\left(\frac{\delta_0}{\delta}-1\right)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)\left(1+x^{\frac{1}{\beta}}\right)^{-1}x^{\frac{1}{\beta}-1}+(1-\beta)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)x^{-1}\Bigg] \\ &= \frac{1}{\beta}\frac{\partial}{\partial x}f_{\delta_0,\delta}(x)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{-1}\Bigg[\left(\frac{\delta_0}{\delta}-1\right)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)\left(1+x^{\frac{1}{\beta}}\right)^{-1}x^{\frac{1}{\beta}-1} \\ &\quad +(\beta-1)\frac{\delta_0}{\delta}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}-1}+(1-\beta)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)x^{-1}\Bigg] \\ &= \frac{1}{\beta}\frac{\partial}{\partial x}f_{\delta_0,\delta}(x)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{-1}\Bigg[\left(\frac{\delta_0}{\delta}-1\right)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)\left(1+x^{\frac{1}{\beta}}\right)^{-1}x^{\frac{1}{\beta}-1} \\ &\quad +(\beta-1)x^{-1}\left(\frac{\delta_0}{\delta}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}}-\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)\right)\Bigg]. \end{aligned}$$

For $x \geq 0$ and $\beta \geq 1$, we have

$$\frac{1}{\beta}\frac{\partial}{\partial x}f_{\delta_0,\delta}(x)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)^{-1} \geq 0$$

and

$$\left(\frac{\delta_0}{\delta}-1\right)\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)\left(1+x^{\frac{1}{\beta}}\right)^{-1}x^{\frac{1}{\beta}-1} \leq 0,$$

and we want to prove that

$$\frac{\delta_0}{\delta}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}}-\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right) \leq 0.$$

We have

$$\begin{aligned}\frac{\delta_0}{\delta}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}}-\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right) &= \left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}\left(\frac{\delta_0}{\delta}x^{\frac{1}{\beta}}-\left(1+x^{\frac{1}{\beta}}\right)\right)+1\\ &= \left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}\left(x^{\frac{1}{\beta}}\left(\frac{\delta_0}{\delta}-1\right)-1\right)+1.\end{aligned}$$

Let $g(x)=\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}\left(x^{\frac{1}{\beta}}\left(\frac{\delta_0}{\delta}-1\right)-1\right)+1$, then

$$\begin{aligned}g'(x) &= \frac{1}{\beta}\left(\frac{\delta_0}{\delta}-1\right)x^{\frac{1}{\beta}-1}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-2}\left(x^{\frac{1}{\beta}}\left(\frac{\delta_0}{\delta}-1\right)-1\right)\\ &\quad+\frac{1}{\beta}\left(\frac{\delta_0}{\delta}-1\right)x^{\frac{1}{\beta}-1}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}\\ &= \frac{1}{\beta}\left(\frac{\delta_0}{\delta}-1\right)x^{\frac{1}{\beta}-1}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-2}\left(x^{\frac{1}{\beta}}\left(\frac{\delta_0}{\delta}-1\right)-1+1+x^{\frac{1}{\beta}}\right)\\ &= \frac{1}{\beta}\frac{\delta_0}{\delta}\left(\frac{\delta_0}{\delta}-1\right)x^{\frac{2}{\beta}-1}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-2}\leq 0.\end{aligned}$$

In addition $g(0)=0$. Therefore, $g(x)\leq 0$, which means that

$$\frac{\delta_0}{\delta}\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}-1}x^{\frac{1}{\beta}}-\left(\left(1+x^{\frac{1}{\beta}}\right)^{\frac{\delta_0}{\delta}}-1\right)\leq 0.$$

Thus, we conclude that $\frac{\partial^2}{\partial x^2}f_{\delta_0,\delta}(x)\leq 0$.

E List of bivariate copula family names

List of copula family names used in C-vine and D-vine trees in Figures 3.14, 3.15, 3.21, and 3.22 is given in Table 3.E.1.

Table 3.E.1 – Bivariate copula family names

Copula short name	Copula long name
I	Independence
N	Gaussian
t	T-Student
BB8	Frank-Joe
SG	Survival Gumbel
SBB1	Survival BB1
SBB7	Survival BB7
C90	Rotated Clayton with 90 degrees
Tawn	Tawn type 1
Tawn180	Rotated Tawn type 1 with 180 degrees
Tawn2	Tawn type 2
Tawn2_180	Rotated Tawn type 2 with 180 degrees

3.6 Bibliography

Abramowitz, M. and Stegun, I. A. (1948). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office.

Applebaum, D. (2009). *Lévy processes and stochastic calculus*. Cambridge university press.

Barbakh, W. A., Wu, Y., and Fyfe, C. (2009). Review of clustering algorithms. In *Non-Standard Parameter Adaptation for Exploratory Data Analysis*, pages 7–28. Springer.

Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30:1031–1068.

Belzile, L. R. and Nešlehová, J. G. (2017). Extremal attractors of Liouville copulas. *Journal of Multivariate Analysis*, 160:68–92.

Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, 26(20):7929–7937.

Bernardoff, P. (2018). Marshall–Olkin Laplace transform copulas of multivariate gamma distributions. *Communications in Statistics-Theory and Methods*, 47(3):655–670.

Capéraà, P., Fougères, A.-L., and Genest, C. (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis*, 72(1):30–49.

Charpentier, A., Fougères, A.-L., Genest, C., and Nešlehová, J. (2014). Multivariate Archimax copulas. *Journal of Multivariate Analysis*, 126:118–136.

Cossette, H., Gadoury, S.-P., Marceau, É., and Mtalai, I. (2017). Hierarchical Archimedean copulas through multivariate compound distributions. *Insurance: Mathematics and Economics*, 76:1–13.

Cossette, H., Gadoury, S.-P., Marceau, E., and Robert, C. Y. (2019). Composite likelihood estimation method for hierarchical Archimedean copulas defined with multivariate compound distributions. *Journal of Multivariate Analysis*, 172:59–83.

Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics, Springer*.

Feller, W. (1971). Introduction to the theory of probability and its applications, vol. 2. *II (2. Ed.) New York: Wiley*.

Genest, C. and Mackay, R. J. (1986). Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14(2):145–159.

Genest, C., Nešlehová, J., and Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64(4):811–834.

Genest, C. and Nešlehová, J. G. (2013). Assessing and modeling asymmetry in bivariate continuous data. In *Copulae in mathematical and quantitative finance*, pages 91–114. Springer.

Górecki, J., Hofert, M., and Holeňa, M. (2016). An approach to structure determination and estimation of hierarchical Archimedean copulas and its application to Bayesian classification. *Journal of Intelligent Information Systems*, 46(1):21–59.

Górecki, J., Hofert, M., and Holeňa, M. (2017). On structure, family and parameter estimation of hierarchical Archimedean copulas. *Journal of Statistical Computation and Simulation*, 87(17):3261–3324.

Górecki, J., Hofert, M., and Okhrin, O. (2020). Outer power transformations of hierarchical Archimedean copulas: Construction, sampling and estimation. *arXiv preprint arXiv:2003.13301*.

Górecki, J. and Holeňa, M. (2013). Structure determination and estimation of hierarchical Archimedean copulas based on kendall correlation matrix. In *International Workshop on New Frontiers in Mining Complex Patterns*, pages 132–147. Springer.

Henderson, R. and Shimakura, S. (2003). A serially correlated gamma frailty model for longitudinal count data. *Biometrika*, 90(2):355–366.

Hering, C., Hofert, M., Mai, J.-F., and Scherer, M. (2010). Constructing hierarchical Archimedean copulas with Lévy subordinators. *Journal of Multivariate Analysis*, 101(6):1428–1433.

Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics & Data Analysis*, 52(12):5163–5174.

Hofert, M. (2010). *Sampling nested Archimedean copulas with applications to CDO pricing*. PhD thesis, Universität Ulm.

Hofert, M. (2011a). Efficiently sampling nested Archimedean copulas. *Computational Statistics & Data Analysis*, 55(1):57–70.

Hofert, M. (2011b). Sampling exponentially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):1–11.

Hofert, M., Huser, R., and Prasad, A. (2018). Hierarchical Archimax copulas. *Journal of Multivariate Analysis*, 167:195–211.

Hofert, M., Mächler, M., and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 110:133–150.

Hua, L. (2016). A note on upper tail behavior of Liouville copulas. *Risks*, 4(4):40.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

Joe, H. and Kurowicka, D. (2011). *Dependence modeling: vine copula handbook*. World Scientific.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivarées*. PhD thesis, Université Laval.

Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes mathematicae*, 10(2):152–164.

Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate analysis*, 99(10):2234–2250.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, 80(1):221–239.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pages 281–297. Oakland, CA, USA.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American statistical association*, 83(403):834–841.

McNeil, A. J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 78(6):567–581.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press.

McNeil, A. J. and Nešlehová, J. (2010). From Archimedean to Liouville copulas. *Journal of Multivariate Analysis*, 101(8):1772–1790.

Mesiar, R. and Jagr, V. (2013). d-Dimensional dependence functions and Archimax copulas. *Fuzzy Sets and Systems*, 228:78–87.

Nelsen, R. (2006). An introduction to copulas, 2nd. *New York: Springer Science Business Media*.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013). On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, 173(2):189–204.

Parner, E. (2001). A composite likelihood approach to multivariate survival data. *Scandinavian Journal of Statistics*, 28(2):295–302.

Perreault, S., Duchesne, T., and Nešlehová, J. G. (2019). Detection of block-exchangeable structure in large-scale correlation matrices. *Journal of Multivariate Analysis*, 169:400–422.

Quessy, J.-F. and Bahraoui, T. (2013). Graphical and formal statistical tools for the symmetry of bivariate copulas. *Canadian Journal of Statistics*, 41(4):637–656.

Rousseeuw, P. J. and Kaufman, L. (1990). Finding groups in data. *Hoboken: Wiley Online Library*.

Savu, C. and Trede, M. (2010). Hierarchies of Archimedean copulas. *Quantitative Finance*, 10(3):295–304.

Schweizer, B. (1991). Thirty years of copulas. In *Advances in probability distributions with given marginals*, pages 13–50. Springer.

Shi, P., Feng, X., Boucher, J.-P., et al. (2016). Multilevel modeling of insurance claims using copulas. *The Annals of Applied Statistics*, 10(2):834–863.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3):335–356.

Conclusion

L'objectif de cette thèse était de proposer des modèles susceptibles de résoudre des problématiques de provisionnement et solvabilité d'une compagnie d'assurance. Plus précisément, nous avons étudié une approche granulaire pour l'évaluation de la réserve et différentes structures de dépendance bivariées et multivariées.

Pour une estimation plus précise des engagements futurs d'une compagnie d'assurance, nous avons proposé un modèle de réserve individuel dans le Chapitre 1. Nous avons utilisé un réseau de neurones récurrent permettant d'incorporer le développement d'un sinistre avec plusieurs informations individuelles statiques et dynamiques. Cette approche permet d'intégrer des covariables individuelles structurées ou non et de modéliser des relations complexes. Notre réseau a la particularité d'avoir deux tâches : une classification pour estimer la probabilité d'avoir un paiement et une régression pour prédire le montant, s'il y a lieu. Ceci nous a menés à utiliser une technique de balancement pour l'apprentissage simultané des deux tâches durant le processus d'entrainement. À travers deux études de cas avec jeux de données réelles et simulées, nous avons illustré la démarche pour la préparation des observations, le processus d'entrainement du réseau et le choix des hyperparamètres. Les estimations obtenues sont plus précises comparativement à celles obtenues avec le modèle agrégé chain-ladder. De plus, pour les paiements larges nous avons proposé une approche d'estimation basée sur les résultats du réseau et la loi Pareto généralisée.

Dans les Chapitres 2 et 3, nous avons développé des modèles en relation avec la solvabilité et l'allocation du capital. Pour le cas d'un portefeuille à deux risques, nous nous sommes intéressés à la dépendance négative extrême et à l'évaluation des mesures de risque VaR et $TVaR$ pour la perte totale dans le Chapitre 2. Nous avons présenté des expressions explicites pour ces mesures en considérant plusieurs familles de distributions continues en appuyant nos démonstrations avec des illustrations graphiques. Cette contribution permet de déduire facilement d'autres mesures importantes telles que le bénéfice de diversification et la prime stop-loss.

Dans le Chapitre 3, nous avons introduit une nouvelle famille de copules hiérarchiques. Notre construction est basée sur une loi mélange exponentielle multivariée avec vecteur de chocs communs obtenus par des sommes de variables indépendantes dont les distributions appartiennent à des familles de distributions fermées sous la convolution. Cette approche nous a permis d'obtenir des copules caractérisées par des distributions non échangeables entre les composantes de blocs ce qui les distingue des autres copules hiérarchiques proposées dans la littérature. Nous avons introduit un algorithme de détermination de la structure et une procédure d'estimation des paramètres accompagnés de plusieurs exemples et de deux études de cas. Dans ces derniers, notre famille de copules a permis de mieux reproduire la

structure complexe des données en comparaison aux copules en vignes et aux copules archimédiennes hiérarchiques.

À travers les études de cas présentées dans cette thèse, on a démontré l'intérêt des approches et modèles proposés. On a également recensé des extensions potentielles, comportant plusieurs défis, à étudier dans des travaux futurs. Pour notre modèle de réserve individuel présenté au Chapitre 1, on peut ajouter des tâches de régression et de classification pour prédire le développement d'autres variables stochastiques telles que le statut ouvert/fermé des dossiers. Il est également intéressant d'ajouter au réseau un mécanisme d'attention et changer la fonction de perte. Ces modifications peuvent améliorer la performance des prédictions individuelles. L'interprétation des résultats du réseau ainsi que l'obtention des intervalles de confiance représentent un volet important de nos travaux de recherche futurs. Les approches présentées dans le Chapitre 2 peuvent être utilisées pour développer des expressions explicites pour plusieurs mesures de risque et être adaptées pour d'autres familles de distributions continues et discrètes. On vise aussi à adapter notre approche à des vecteurs de plus de deux variables aléatoires définies selon les différentes relations de dépendance négative extrême proposées dans la littérature. Pour la famille des copules hiérarchiques présentée dans le Chapitre 3, on envisage d'investiguer davantage les propriétés de dépendance dans les arbres à plusieurs niveaux et proposer une généralisation des algorithmes d'identification de la structure et d'estimation des paramètres.

Bibliographie intégrale

Abdallah, A., Boucher, J.-P., and Cossette, H. (2015). Modeling dependence between loss triangles with hierarchical archimedean copulas. *ASTIN Bulletin*, 45(3):577–599.

Abramowitz, M. and Stegun, I. A. (1948). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office.

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503.

Albrecher, H., Teugels, J. L., and Beirlant, J. (2017). *Reinsurance: Actuarial and Statistical Aspects*. John Wiley & Sons.

Alink, S., Löwe, M., and Wüthrich, M. V. (2004). Diversification of aggregate dependent risks. *Insurance: Mathematics and Economics*, 35(1):77–95.

Antonio, K. and Plat, R. (2014). Micro-level stochastic loss reserving for general insurance. *Scandinavian Actuarial Journal*, 2014:649–669.

Applebaum, D. (2009). *Lévy processes and stochastic calculus*. Cambridge university press.

Arjas, E. (1989). The claims reserving problem in non-life insurance: Some structural ideas. *ASTIN Bulletin*, 19(2):139–152.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Barbakh, W. A., Wu, Y., and Fyfe, C. (2009). Review of clustering algorithms. In *Non-Standard Parameter Adaptation for Exploratory Data Analysis*, pages 7–28. Springer.

Barbe, P., Fougeres, A.-L., and Genest, C. (2006). On the tail behavior of sums of dependent risks. *ASTIN Bulletin: The Journal of the IAA*, 36(2):361–373.

Baudry, M. and Robert, C. Y. (2019). A machine learning approach for individual claims reserving in insurance. *Applied Stochastic Models in Business and Industry*, 35(5):1127–1155.

Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, 30:1031–1068.

Belzile, L. R. and Nešlehová, J. G. (2017). Extremal attractors of Liouville copulas. *Journal of Multivariate Analysis*, 160:68–92.

Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, 26(20):7929–7937.

Bernardoff, P. (2018). Marshall–Olkin Laplace transform copulas of multivariate gamma distributions. *Communications in Statistics-Theory and Methods*, 47(3):655–670.

Brechmann, E. C., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85.

Bürgi, R., Dacorogna, M. M., and Iles, R. (2008). Risk aggregation, dependence structure and diversification benefit. *Stress testing for financial institutions*.

Capéraà, P., Fougères, A.-L., and Genest, C. (2000). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis*, 72(1):30–49.

Charpentier, A., Fougères, A.-L., Genest, C., and Nešlehová, J. (2014). Multivariate Archimax copulas. *Journal of Multivariate Analysis*, 126:118–136.

Charpentier, A. and Pigeon, M. (2016). Macro vs. micro methods in non-life claims reserving (an econometric perspective). *Risks*, 4(2):12.

Cheung, K. C., Dhaene, J., Lo, A., and Tang, Q. (2014). Reducing risk by merging counter-monotonic risks. *Insurance: Mathematics and Economics*, 54:58–65.

Cheung, K. C. and Lo, A. (2013a). Characterizations of counter-monotonicity and upper comonotonicity by (tail) convex order. *Insurance: Mathematics and Economics*, 53(2):334–342.

Cheung, K. C. and Lo, A. (2013b). General lower bounds on convex functionals of aggregate sums. *Insurance: Mathematics and Economics*, 53(3):884–896.

Cheung, K. C. and Lo, A. (2014). Characterizing mutual exclusivity as the strongest negative multivariate dependence structure. *Insurance: Mathematics and Economics*, 55:180–190.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer, London.

Cossette, H., Côté, M.-P., Mailhot, M., and Marceau, E. (2014). A note on the computation of sharp numerical bounds for the distribution of the sum, product or ratio of dependent risks. *Journal of Multivariate Analysis*, 130:1–20.

Cossette, H., Gadoury, S.-P., Marceau, É., and Mtalai, I. (2017). Hierarchical Archimedean copulas through multivariate compound distributions. *Insurance: Mathematics and Economics*, 76:1–13.

Cossette, H., Gadoury, S.-P., Marceau, E., and Robert, C. Y. (2019a). Composite likelihood estimation method for hierarchical Archimedean copulas defined with multivariate compound distributions. *Journal of Multivariate Analysis*, 172:59–83.

Cossette, H., Marceau, E., Nguyen, Q. H., and Robert, C. Y. (2019b). Tail approximations for sums of dependent regularly varying random variables under Archimedean copula models. *Methodology and Computing in Applied Probability*, 21(2):461–490.

Côté, M.-P., Genest, C., and Abdallah, A. (2016). Rank-based methods for modeling dependence between loss triangles. *European Actuarial Journal*, 6(2):377–408.

Côté, M.-P., Genest, C., and Stephens, D. A. (2021). A Bayesian approach to modeling multivariate multilevel insurance claims in the presence of unsettled claims. *Bayesian Analysis*, DOI: 10.1214/20-BA1243.

Czado, C. (2019). Analyzing dependent data with vine copulas. *Lecture Notes in Statistics, Springer*.

Czado, C., Kastenmeier, R., Brechmann, E. C., and Min, A. (2012). A mixed copula model for insurance claims and claim sizes. *Scandinavian Actuarial Journal*, 2012(4):278–305.

Dacorogna, M. M., Elbahtouri, L., and Kratz, M. (2016). Explicit diversification benefit for dependent risks. *SCOR papers*, (38).

Deelstra, G., Dhaene, J., and Vanmaele, M. (2011). An overview of comonotonicity and its applications in finance and insurance. *Advanced Mathematical Methods for Finance*, pages 155–179.

Delong, Ł. and Wüthrich, M. V. (2020). Neural networks for the joint development of individual payments and claim incurred. *Risks*, 8(2):33.

Denuit, M. and Dhaene, J. (2003). Simple characterizations of comonotonicity and countermonotonicity by extremal correlations. *Belgian Actuarial Bulletin*, 3:22–27.

Denuit, M., Dhaene, J., Goovaerts, M., and Kaas, R. (2006). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. John Wiley & Sons.

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, 93(1):40–57.

Dhaene, J. and Denuit, M. (1999). The safest dependence structure among risks. *Insurance: Mathematics and Economics*, 25(1):11–21.

Dhaene, J., Denuit, M., Goovaerts, M. J., Kaas, R., and Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, 31(1):3–33.

Dhaene, J. and Goovaerts, M. J. (1996). Dependency of risks and stop-loss order. *ASTIN Bulletin: The Journal of the IAA*, 26(2):201–212.

Dhaene, J., Linders, D., Schoutens, W., and Vyncke, D. (2012). The Herd Behavior Index: A new measure for the implied degree of co-movement in stock markets. *Insurance: Mathematics and economics*, 50(3):357–370.

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., and Vyncke, D. (2006). Risk measures and comonotonicity: A review. *Stochastic models*, 22(4):573–606.

Dhaene, J., Wang, S., Young, V. R., and Goovaerts, M. (2000). Comonotonicity and maximal stop-loss premiums. *Bulletin of the Swiss Association of Actuaries*, 2:99–113.

Dissmann, J., Brechmann, E. C., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Drossos, K., Gharib, S., Magron, P., and Virtanen, T. (2019). Language modelling for sound event detection with teacher forcing and scheduled sampling. *arXiv preprint arXiv:1907.08506*.

Duval, F. and Pigeon, M. (2019). Individual loss reserving using a gradient boosting-based approach. *Risks*, 7(3):79.

Embrechts, P. and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application*, 1:493–514.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling Extremal Events: for Insurance and Finance*, volume 33. Springer Science & Business Media.

Embrechts, P., Lambrigger, D. D., and Wüthrich, M. V. (2009a). Multivariate extremes and the aggregation of dependent risks: Examples and counter-examples. *Extremes*, 12(2):107–127.

Embrechts, P., Nešlehová, J., and Wüthrich, M. V. (2009b). Additivity properties for value-at-risk under Archimedean dependence and heavy-tailedness. *Insurance: Mathematics and Economics*, 44(2):164–169.

England, P. D. and Verrall, R. J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, 8(3):443–518.

Feller, W. (1971). Introduction to the theory of probability and its applications, vol. 2. *II (2. Ed.) New York: Wiley*.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon, 3^e serie, Sciences, Sect. A*, 14:53–77.

Gabrielli, A. (2020). A neural network boosted double overdispersed poisson claims reserving model. *ASTIN Bulletin: The Journal of the IAA*, 50(1):25–60.

Gabrielli, A. (2021). An individual claims reserving model for reported claims. *European Actuarial Journal*, DOI:10.1007/s13385-021-00271-4.

Gabrielli, A., Richman, R., and Wüthrich, M. V. (2020). Neural network embedding of the over-dispersed Poisson reserving model. *Scandinavian Actuarial Journal*, 2020(1):1–29.

Gabrielli, A. and Wüthrich, M. (2018). An individual claims history simulation machine. *Risks*, 6(2):29.

Gaffke, N. and Rüscherndorf, L. (1981). On a class of extremal problems in statistics. *Mathematische Operationsforschung und Statistik. Series Optimization*, 12(1):123–135.

Genest, C. and Mackay, R. J. (1986). Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14(2):145–159.

Genest, C., Nešlehová, J., and Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64(4):811–834.

Genest, C. and Nešlehová, J. G. (2013). Assessing and modeling asymmetry in bivariate continuous data. In *Copulae in mathematical and quantitative finance*, pages 91–114. Springer.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Górecki, J., Hofert, M., and Holeňa, M. (2016). An approach to structure determination and estimation of hierarchical Archimedean copulas and its application to Bayesian classification. *Journal of Intelligent Information Systems*, 46(1):21–59.

Górecki, J., Hofert, M., and Holeňa, M. (2017). On structure, family and parameter estimation of hierarchical Archimedean copulas. *Journal of Statistical Computation and Simulation*, 87(17):3261–3324.

Górecki, J., Hofert, M., and Okhrin, O. (2020). Outer power transformations of hierarchical Archimedean copulas: Construction, sampling and estimation. *arXiv preprint arXiv:2003.13301*.

Górecki, J. and Holeňa, M. (2013). Structure determination and estimation of hierarchical Archimedean copulas based on kendall correlation matrix. In *International Workshop on New Frontiers in Mining Complex Patterns*, pages 132–147. Springer.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Graves, A. and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6):602–610.

Haastrup, S. and Arjas, E. (1996). Claims reserving in continuous time; A nonparametric Bayesian approach. *ASTIN Bulletin*, 26(2):139–164.

Heffernan, J. E. (2018). *ISMEV: An Introduction to Statistical Modeling of Extreme Values*. R package version 1.42.

Henderson, R. and Shimakura, S. (2003). A serially correlated gamma frailty model for longitudinal count data. *Biometrika*, 90(2):355–366.

Hering, C., Hofert, M., Mai, J.-F., and Scherer, M. (2010). Constructing hierarchical Archimedean copulas with Lévy subordinators. *Journal of Multivariate Analysis*, 101(6):1428–1433.

Hiabu, M., Margraf, C., Martínez-Miranda, M. D., and Nielsen, J. P. (2016). The link between classical reserving and granular reserving through double chain ladder and its extensions. *British Actuarial Journal*, 21(1):97–116.

Hobson, D., Laurence, P., and Wang, T.-H. (2005a). Static-arbitrage optimal subreplicating strategies for sasket options. *Insurance: Mathematics and Economics*, 37(3):553–572.

Hobson, D., Laurence, P., and Wang, T.-H. (2005b). Static-arbitrage upper bounds for the prices of basket options. *Quantitative finance*, 5(4):329–342.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural computation*, 9(8):1735–1780.

Hofert, M. (2008). Sampling Archimedean copulas. *Computational Statistics & Data Analysis*, 52(12):5163–5174.

Hofert, M. (2010). *Sampling nested Archimedean copulas with applications to CDO pricing*. PhD thesis, Universität Ulm.

Hofert, M. (2011a). Efficiently sampling nested Archimedean copulas. *Computational Statistics & Data Analysis*, 55(1):57–70.

Hofert, M. (2011b). Sampling exponentially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):1–11.

Hofert, M. (2012). A stochastic representation and sampling algorithm for nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, 82(9):1239–1255.

Hofert, M., Huser, R., and Prasad, A. (2018). Hierarchical Archimax copulas. *Journal of Multivariate Analysis*, 167:195–211.

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., Maechler, M. M., and Suggests, M. (2014). Package "copula". *R Package Manual*.

Hofert, M., Mächler, M., and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 110:133–150.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

Hua, L. (2016). A note on upper tail behavior of Liouville copulas. *Risks*, 4(4):40.

Inui, K. and Kijima, M. (2005). On the significance of expected shortfall as a coherent risk measure. *Journal of Banking & Finance*, 29(4):853–864.

Jeon, J., Kochar, S., and Park, C. G. (2006). Dispersive ordering – some applications and examples. *Statistical Papers*, 47(2):227–247.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. *Lecture Notes-Monograph Series*, pages 120–141.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

Joe, H. (2014). *Dependence modeling with copulas*. Chapman and Hall/CRC.

Joe, H. and Kurowicka, D. (2011). *Dependence modeling: vine copula handbook*. World Scientific.

Kendall, A., Gal, Y., and Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491.

Khoudraji, A. (1995). *Contributions à l'étude des copules et à la modélisation de valeurs extrêmes bivarées*. PhD thesis, Université Laval.

Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. *Aequationes mathematicae*, 10(2):152–164.

Kuo, K. (2019). Deeptriangle: A deep learning approach to loss reserving. *Risks*, 7(3):97.

Kuo, K. (2020). Individual claims forecasting with Bayesian mixture density networks. *arXiv preprint arXiv:2003.02453*.

Laurence, P. and Wang, T.-H. (2005). Sharp upper and lower bounds for basket options. *Applied Mathematical Finance*, 12(3):253–282.

Lee, W. and Ahn, J. Y. (2014). On the multidimensional extension of countermonotonicity and its applications. *Insurance: Mathematics and Economics*, 56:68–79.

Lee, W., Cheung, K. C., and Ahn, J. Y. (2017). Multivariate countermonotonicity and the minimal copulas. *Journal of Computational and Applied Mathematics*, 317:589–602.

Liebscher, E. (2008). Construction of asymmetric multivariate copulas. *Journal of Multivariate analysis*, 99(10):2234–2250.

Lindholm, M., Verrall, R., Wahl, F., and Zakrisson, H. (2020). Machine learning, regression models, and prediction of claims reserves. In *Arlington: Casualty Actuarial Society E-Forum.[Google Scholar]*.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary mathematics*, 80(1):221–239.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pages 281–297. Oakland, CA, USA.

Marceau, E. (2013). *Modélisation et évaluation quantitative des risques en actuariat: Modèles sur une période*. Springer.

Marshall, A. W. and Olkin, I. (1988). Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403):834–841.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton university press.

McNeil, A. J. and Nešlehová, J. (2010). From Archimedean to Liouville copulas. *Journal of Multivariate Analysis*, 101(8):1772–1790.

Mesiar, R. and Jagr, V. (2013). d-Dimensional dependence functions and Archimax copulas. *Fuzzy Sets and Systems*, 228:78–87.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*, volume 389. Wiley New York.

Nelsen, R. (2006). An introduction to copulas, 2nd. *New York: Springer Science Business Media*.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Norberg, R. (1993). Prediction of outstanding liabilities in non-life insurance I. *ASTIN Bulletin*, 23(1):95–115.

Norberg, R. (1999). Prediction of outstanding liabilities II. Model variations and extensions. *ASTIN Bulletin*, 29(1):5–25.

Okhrin, O., Okhrin, Y., and Schmid, W. (2013). On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, 173(2):189–204.

Parner, E. (2001). A composite likelihood approach to multivariate survival data. *Scandinavian Journal of Statistics*, 28(2):295–302.

Perreault, S., Duchesne, T., and Nešlehová, J. G. (2019). Detection of block-exchangeable structure in large-scale correlation matrices. *Journal of Multivariate Analysis*, 169:400–422.

Pigeon, M., Antonio, K., and Denuit, M. (2013). Individual loss reserving with the multivariate skew normal framework. *ASTIN Bulletin*, 43(3):399–428.

Pigeon, M., Antonio, K., and Denuit, M. (2014). Individual loss reserving using paid–incurred data. *Insurance: Mathematics and Economics*, 58:121–131.

Puccetti, G., Wang, B., and Wang, R. (2013). Complete mixability and asymptotic equivalence of worst-possible VaR and ES estimates. *Insurance: Mathematics and Economics*, 53(3):821–828.

Puccetti, G., Wang, R., et al. (2015). Extremal dependence concepts. *Statistical Science*, 30(4):485–517.

Quessy, J.-F. and Bahraoui, T. (2013). Graphical and formal statistical tools for the symmetry of bivariate copulas. *Canadian Journal of Statistics*, 41(4):637–656.

Rousseeuw, P. J. and Kaufman, L. (1990). Finding groups in data. *Hoboken: Wiley Online Library*.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Savu, C. and Trede, M. (2010). Hierarchies of Archimedean copulas. *Quantitative Finance*, 10(3):295–304.

Scarsini, M. (1984). On measures of concordance. *Stochastica: Revista de Matemática Pura y Aplicada*, 8(3):201–218.

Schweizer, B. (1991). Thirty years of copulas. In *Advances in probability distributions with given marginals*, pages 13–50. Springer.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Science & Business Media.

Shi, P., Feng, X., Boucher, J.-P., et al. (2016). Multilevel modeling of insurance claims using copulas. *The Annals of Applied Statistics*, 10(2):834–863.

Shi, W., Li, K. X., Yang, Z., and Wang, G. (2017). Time-varying copula models in the shipping derivatives market. *Empirical Economics*, 53(3):1039–1058.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

Sundermeyer, M., Schlüter, R., and Ney, H. (2012). LSTM neural networks for language modeling. In *Thirteenth annual conference of the international speech communication association*, pages 194–197.

Taylor, G. (2019). Loss reserving models: Granular and machine learning forms. *Risks*, 7(3):82.

Taylor, G., McGuire, G., and Sullivan, J. (2008). Individual claim loss reserving conditioned by case estimates. *Annals of Actuarial Science*, 3(1-2):215–256.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *The Annals of Probability*, 8(4):814–827.

Van Buuren, S. and Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.

Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360.

Wang, B. and Wang, R. (2016). Joint mixability. *Mathematics of Operations Research*, 41(3):808–826.

Wang, R., Peng, L., and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, 17(2):395–417.

Weninger, F., Erdogan, H., Watanabe, S., Vincent, E., Le Roux, J., Hershey, J. R., and Schuller, B. (2015). Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 91–99. Springer.

Williams, R. J. and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280.

Wüthrich, M. V. (2018a). Machine learning in individual claims reserving. *Scandinavian Actuarial Journal*, pages 1–16.

Wüthrich, M. V. (2018b). Neural networks applied to chain–ladder reserving. *European Actuarial Journal*, 8(2):407–436.

Wüthrich, M. V. and Merz, M. (2008). *Stochastic claims reserving methods in insurance*, volume 435. John Wiley & Sons, Chichester.

Yin, Z. and Shen, Y. (2018). On the dimensionality of word embedding. *arXiv preprint arXiv:1812.04224*.

Yu, Y., Si, X., Hu, C., and Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation*, 31(7):1235–1270.

Zhao, X. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics*, 46(2):290–299.

Zhao, X. B., Zhou, X., and Wang, J. L. (2009). Semiparametric model for prediction of individual claim loss reserving. *Insurance: Mathematics and Economics*, 45(1):1–8.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3):335–356.