A dependence analysis within the context of risk allocations: Simplex compositions, the notion of counter-monotonicity and heavy-tailed distributions.

Nawaf M.A. Mohammed

A thesis submitted to the Faculty of Graduate Studies in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Mathematics and Statistics

York University, Ontario, Canada
June, 2023

Abstract

The remarkable development of today's financial and insurance products demands sound methodologies for the accumulation and characterization of intertwined risks. As a result, modern risk management emerges as a by product querying two key foundations. The first is concerned with the aggregation of said risks into one randomness which is consequently easily measured by a convenient risk measure and thereafter reported. The pooling is done from the different business units (BUs) composing the financial entity. The second pillar pertains to the opposite direction which concerns itself with the allocation of the total risk. It seeks to accurately and concretely attribute the riskiness of each individual BU with respect to the whole.

The aggregation process, on one hand, has been fairly well studied in the literature, implemented in the industry and even embedded into the different accords. Risk capital allocation, on the other, is generally much more involved even when a specific risk measure inducing the allocation rule is assumed, let alone the case when a class of risk measures is considered. And unlike the aggregation exercise, which is moderately determined by the collection function, attributing capital is often more heavily influenced by the dependencies among the different BUs.

In the literature, nonetheless, allocating capital can be categorized into two main camps. One is built upon the pretence that the distribution of risk should satisfy certain regulatory requirements. This leads to an axiomatic approach which is quite often mathematically tractable yet ignores the economic incentives of the market. The other school of thought is economically driven, allocating risk based on a profit-maximizing paradigm. It argues that

capital allocation should reflect the risk perception of the institution and not be imposed by any arbitrary measure, for which its selection is dubious at best. However, the economic approach suffers from complex relations that lack clear definitive forms.

At first glance the two perspectives may seem distant, as they arise naturally in their own contexts and are justified accordingly. Nonetheless, they can coincide for particular losses that enjoy certain peculiar model settings which are described thoroughly in the chapters thereafter. Surprisingly, the reconciliation comes in connection with the concept of trivial allocations. Triviality, in itself, attracts practitioners as it requires no discernible dependencies leading to a convenient yet faulty method of attributing risk. Regardless, when used in the right context it unveils surprising connections and conveys useful conclusions. The intersection of the regulatory and profit-maximizing principles, for example, mainly utilizes a milder version of triviality (proportional) which allows for distinct, albeit few, probabilistic laws that accommodate both theories. Furthermore, when a stronger triviality (absolute) condition is imposed, it yields another intriguing corollary, specifically that of restrictive extreme laws commonly known for antithetic or counter-monotonic variates.

To address the framework hitherto introduced, in the first chapter of this dissertation, we present a general class of weighted pricing functionals. This wide class covers most of the risk measures and allocations found in the literature today and adequately represents their various properties. We begin by investigating the order characteristics of the functionals under certain sufficient conditions. The results reveal interactive relationships between the weight and the aggregation make-up of the measures, which consequently, allow for effective comparison between the different risks. Then upon imposing restrictions on the allocation constituents, we establish equivalent statements for trivial allocations that uncover a novel general concept of counter-monotonicity. More significantly, similar equivalences are obtained for a weaker triviality notion that pave the path to answer the aforementioned question of allocation reconciliation.

The class of weighted functionals, though constructive, is too general to apply effectively

to the allocation theories. Thus, in the second chapter, we consider the special case of conditional tail expectation (CTE), defining its risk measure and the allocation it induces. These represent the regulatory approach to allocation as CTE is arguably one of the most prominent and frontrunner measures used and studied today. On the other side, we consider the allocation arising from the economic context that aims to maximize profit subject to other market forces as well as individual perceptions. Both allocations are taken as proportions as they are formed from compositional maps which relate to the standard simplex in either a stochastic or non-stochastic manner. Then we equate the two allocations and derive a general description for the laws that satisfy the two functionals. The Laplace transform of the multivariate size bias is used as the prime identifier delineating the general distributions and detailing subsequent corollaries and examples.

While studying the triviality nature of allocations, we focused on the central element of stochastic dependence. We showed how certain models, extremal dependence for instance, enormously influences the attribution outcome. Thus far, nonetheless, our query started from the point of allocation relations, be it proportional or absolute, then ended in law characterizations that satisfy those relations. Equally important, on the other hand, is deriving allocations expressions based on a priori assumed models. This task requires apt choices of general structures which convey the desired probabilistic nature of losses. Since constructing joint laws can be quite challenging, the compendium of probabilistic models relies heavily on leveraging the stochastic representations of known distributions. This feat allows not only for simpler computations but as well for useful interpretations. Basic mathematical operations are usually deployed to derive different joint distributions with certain desirable properties. For example, taking the minimum yields the Marshall-Olkin distribution, addition gives the additive background model and multiplication/division naturally leads to the multiplicative background model. Simultaneously, univariate manipulation through location, scale and power transforms adds to the flexibility of the margins while preserving the overall copula. In the last chapter of this dissertation, we introduce a composite of the Marshall-Olkin, additive and multiplicative models to obtain a novel multivariate Pareto-Dirichlet law possessing a profound composition capable of modelling heavy tailed events descriptor of many extremal

scenarios in insurance and finance. We study its survival function and the corresponding moments and mixed moments. Then we focus on the bivariate case, detailing the intricacies of its inherent expressions. And finally, we conclude with a thorough application to the risk and allocation functionals respectively.

To my parents

Acknowledgements

First of all, I wish to thank my doctoral supervisors, Edward Furman and Jianxi Su. I am forever grateful to both for having guided me through my wonderful Ph.D. journey. Their patience and deep insights were instrumental in shaping my understanding of the world of risk and mathematics in general. In them, I not only found distinguished mentors, rather I made life-long friends tied together by the quest of knowledge.

In addition, I would like to express my gratitude to the members of the supervisory committee, Alexey Kuznetsov and Augustine Wong, for which their embrace and genuine interest made significant contributions towards my academic maturity and general development. Lastly, many thanks to the reviewers, Ricardas Zitikis and Neal Madras, who gave useful tips and suggestions that helped shape a well-founded and coherent dissertation message.

Finally, a very special recognition to my parents, Mahmood and Jawhara, and my siblings, Nazek, Abdulrahman and Yanoof, for their unyielding support and unlimited love. Without their constant belief in me and their sacrifices, I would not have been able to reach this point in my life. Though distance separated us, but their place is forever engrained in my heart and mind.

Table of Contents

Chapter 1

Introduction

The plethora of risk measures used today aim at capturing the inherent uncertainty exhibited by the losses. The goal is to produce a scalar value capable of summarizing the underlying risk. It is evident, therefore, that this task is quite broad as there are infinitely many ways to define such a measure. Nonetheless, they all share the basic and common purpose of risk collection or aggregation. Those risks are pooled together from their constituents, generally referred to as business units (BU), into one loss that is subsequently gauged. In all accounts, the fundamental object, in which risk measures as well as allocations operate, is the randomness of the corresponding losses. If $n \in \mathbb{N}$ indicates the number of BUs with a label set $\mathcal{N} = \{1, \dots, n\}$, then the losses are represented as a non-negative random vector $\boldsymbol{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$ where $\mathcal{X}^n$ is the space of those losses. Since most measures used today are described as expectations, usually one requires each loss to have a finite mean i.e. $X_i \in L^1,\ \forall i \in \mathcal{N}$, however this is not necessary in general.

The process of aggregation, although abstract, can be explicitly defined via a collection function, call it g, which maps the the realizations of the random vector $\boldsymbol{X}$ of losses into a non-negative scalar, concretely $g : \mathbb{R}^n_+ \mapsto \mathbb{R}_+$. The choice of g and consequently the scalar value it outputs demonstrates the financial entity's decisions regarding its own structure and the internal environment it manages. It is customary to endow g with certain desirable properties, most importantly, it should reflect the monotonicity of the losses it collects. Meaning, if one loss increases, while the others are held constant, the total collection should increase.

Normally, in actuarial science, the aggregation function is chosen to be the sum of the constituent losses i.e. $g(\boldsymbol{x}) = x_1 + \cdots + x_n$. The sum, conveying the simplest form of aggregation, is often-times easier to deal with as it respects most regularity conditions and requires no additional assumptions. Finally, upon the determination of the proper compilation method g, the next stage is to appropriately map the resulting collective into a meaningful positive number. Typically, this procedure is crystallized by a functional, denoted by H, mapping the space of losses to the non-negative reals, i.e. $H : \mathcal{X}^n \mapsto \mathbb{R}_+$. Figure 1.1 shows a summary of the mechanism involving risk aggregation and measurement.

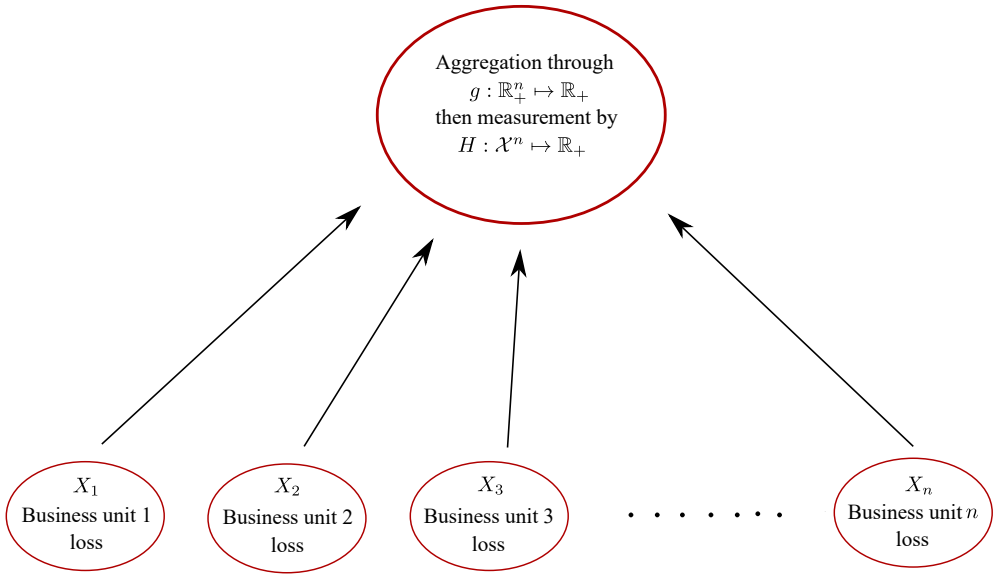


Figure 1.1: The process of risk aggregation and measurement

The choice of the risk measure H embeds all the inherent assumptions and properties that we deem prudent and sensible. The techniques used vary widely depending on the study objectives and justifications. For example, Artzner et al. (1999) defines coherent risk measures through a set of axioms. Coherence is synonyms with risk monotonicity, sub-additivity, positive homogeneity and translation invariance which mimics the established laws of modern financial theory. Another development comes from Denneberg (1990) and Wang (1996) through the introduction of a measurement method via distorting the underlying survival function of the losses. Many subsequent works expanded on this notion and produced several

characteristics under which a distortion might reflect the desirable financial incentives. Several closely related concepts emerged such as that of spectral measures (Acerbi and Tasche, 2002) and most notably, which the second chapter of this dissertation is based on and generalized upon, that of weighted premiums put forth by Furman and Zitikis (2008a).

The body of literature associated with risk measures is quite vast. To the contrary, in the industry, due to operating and regulatory reasons, adoption has been intermittent and progress is slow. Nevertheless, in recent years following the financial crises of 2007-2008, many strides were taken to mitigate the catastrophic events through appropriately measuring and allocating capital. Until recently and still today, throughout the landscape, determining risk capital is predominately done through a quantile based manner or commonly referred to as the Value-at-Risk (VaR) (Linsmeier and Pearson, 2000). Given a prudence level $q \in [0, 1)$, VaR is defined to be the smallest value of a (total) random loss $Y = g(\boldsymbol{X})$, which is usually taken to be the simple summation $Y = X_1 + \cdots + X_n$, for which the probability of not exceeding this value is at least q, formally:

$$H(\boldsymbol{X}) = \inf\{y : \mathbb{P}(Y \leq y) \geq q\}. \tag{1.1}$$

Due to its popularity, it became a useful tool for traders and risk managers alike. However it drew swaths of criticisms due to its inadequacy for non-normality and inability to properly capture the tails (Jorion, 2006). Some critics went as far as blaming the VaR for the recent financial crisis, arguing it created a false sense of security for banks as it is easily misunderstood and dangerously so.

The search for alternatives, simple computationally, yet robust enough to properly mitigate the VaR shortcomings, resulted in the recent adoption of a tail measure, frequently referred to as the conditional tail expectation (CTE) or expected shortfall (ES). It is defined to be the average of VaRs beyond the prudence level $q \in [0, 1)$, mathematically:

$$H(\boldsymbol{X}) = \frac{1}{1-q} \int_q^1 \mathrm{VaR}_t[Y] \, \mathrm{d}t, \tag{1.2}$$

and when the distribution function F_Y of the aggregate random loss Y is continuous then the CTE coincides with ES and can be compactly written as $H(\boldsymbol{X}) = \mathbb{E}\left[Y|Y > \text{VaR}_q[Y]\right]$. Remarkably, the CTE, through averaging VaRs, enjoys many advantageous attributes such as coherence. It has been implemented in the recent Basel accords and drawn praise from the academic community at large, see (e.g., Acerbi and Tasche, 2002; Tasche, 2002; Wang and Zitikis, 2021; Yamai and Yoshiba, 2005) for relevant discussions on the subject.

Not surprisingly, the volume of works dedicated to risk measurement has been growing quickly in the past decade as it has received its fair attention within the academic and non-academic circles alike. Therefore, its theory and applicability is quite well understood within the field of risk management. By comparison, when moving in the other direction, the case of allocating the pieces of the total risk to the different BUs can be quite cumbersome. Allocations, crucial to any healthy financial system, are used to gauge the riskiness, and consequently, performance of the different BUs with respect to the whole. This function serves multiple purposes. Through allocations, the financial entity can accurately distribute resources as well as expenses based on the individual BU contribution to the overall risk (Dhaene et al., 2012). Generally, allocations are represented as functionals, $\boldsymbol{A} : \mathcal{X}^n \mapsto \mathbb{R}^n_+$, delineating the link between the space of losses and aggregates to that of a non-negative vector, whose numbers, $\boldsymbol{A}(\boldsymbol{X}) = (A_1(\boldsymbol{X}), \ldots, A_n(\boldsymbol{X}))$, convey the allocations of the respective BUs. For simplicity, often times, when a class of allocations is assumed, then each functional $A_i,\ i \in \mathcal{N}$, simply maps the product space $\mathcal{X} \times \mathcal{X}$, comprised of the individual losses as the first coordinate and the aggregate as the second, to the non-negative reals, succinctly written as $A(X_i, g(\boldsymbol{X}))$. Figure 1.2 shows the reverse procedure of capital allocation starting with the joint losses and aiming at distributing the overall risk among the BUs.

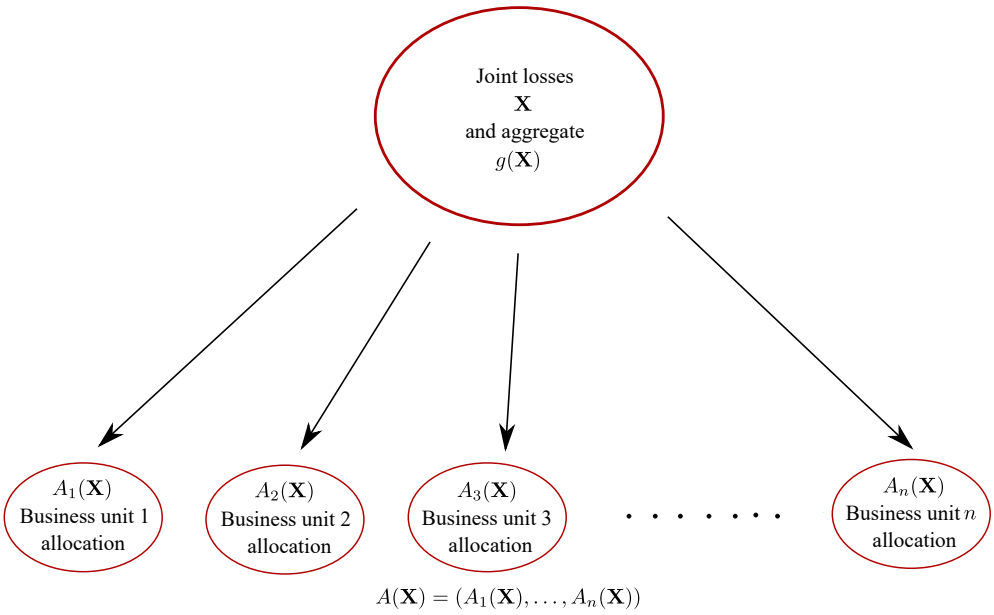


Figure 1.2: The process of risk allocation

Unlike risk measures, allocations are highly specialized mechanisms that require careful considerations. Not only the aggregate and its law are considered, but also the individual constituent relationship to them. As a result, most methodologies used today employ the machinery of risk measures as a simplistic overarching shadow which allows for a systemic derivations of the corresponding allocations. The ease of inducing allocations based on an apriori risk measure is mathematically convenient as it requires minimal assumptions. Mainly, the Euler rule, which we shall now define, is implemented in a form of a gradient method that seeks to optimally attribute the contribution of each BU to the whole. The only condition imposed on the risk measures is that of positive homogeneity. A measure H is said to be possess that property if for each scalar $\alpha > 0$, the riskiness of the scaled losses $\alpha \boldsymbol{X}$ is exactly the scaled riskiness of the original losses $\boldsymbol{X}$, expressed as $H(\alpha \boldsymbol{X}) = \alpha H(\boldsymbol{X})$. In financial terms, the scalar can represent a momentarily exchange rate and the riskiness of certain random losses should be the same regardless of the denomination. If a risk measure enjoys positive homogeneity then, for the fractional losses $\boldsymbol{X}(\boldsymbol{u}) = (u_1 X_1, \ldots, u_n X_n)$, $u_i \in [0, 1]$, $\forall i \in \mathcal{N}$, H can be expressed as a combination of the gradient components, each

weighted by the corresponding u_i, precisely:

$$H(\boldsymbol{X}(\boldsymbol{u})) = u_1 \frac{\partial H(\boldsymbol{X}(\boldsymbol{u}))}{\partial u_1} + \cdots + u_n \frac{\partial H(\boldsymbol{X}(\boldsymbol{u}))}{\partial u_n}. \tag{1.3}$$

Each term $\frac{\partial H(\boldsymbol{X}(\boldsymbol{u}))}{\partial u_i}$ is interpreted as the marginal attribution of the respective BU i to the total risk H. The final allocation, consequently, is recovered by taking the marginal attribution at full weight. In another words :

$$A_i(\boldsymbol{X}) = \frac{\partial H(\boldsymbol{X}(\boldsymbol{u}))}{\partial u_i}\bigg|_{u_j=1,\ \forall j \in \mathcal{N}}, \ \forall i \in \mathcal{N}. \tag{1.4}$$

If a risk measure is differentiable then Euler method is an elegant way to comprehensively and simultaneously measure and allocate risk. Many allocations used today follow this framework as they satisfy both positive homogeneity and differentiability. Since the VaR, defined in (1.1), is positive homogeneous, then the CTE measure in (1.2) will be positive homogeneous as well. Additionally, being differentiable, the CTE will induce the following comparable CTE allocations:

$$A_i(\boldsymbol{X}) = \mathbb{E}\left[X_i | Y > \mathrm{VaR}_q[Y]\right], \ \forall i \in \mathcal{N}, \tag{1.5}$$

where the aggregate function Y is again the sum and F_Y is assumed continuous. Regularly, allocations are reported as percentages (Belles-Sampera et al., 2016) indicating the share of each BU to the complete 100% risk. The procedure usually involves normalizing the allocation by the total measure. In the case of the CTE, one can disclose the proportional risk as:

$$\overline{A}_i(\boldsymbol{X}) = \frac{A_i(\boldsymbol{X})}{H(\boldsymbol{X})} = \frac{\mathbb{E}\left[X_i | Y > \mathrm{VaR}_q[Y]\right]}{\mathbb{E}\left[Y | Y > \mathrm{VaR}_q[Y]\right]}, \ \forall i \in \mathcal{N}. \tag{1.6}$$

The CTE allocation inherits its popularity from its inducing measure, thus it is arguably one of the most prominent rules used to attribute risk currently. It arises in multiple contexts, ranging from game theory to optimal functionals, see (e.g., Denault, 2001; Dhaene et al., 2012; Tasche, 2004) for pertinent references.

Generating allocations based on risk measures can be regarded as an indirect method bypassing the innate considerations of attributions. Certainly, starting with regulatory measures has its own advantages, however, the axioms that govern them may be not applicable within the context of allocations. Moreover, as the main object, choosing a risk measure is an unclear task that can place a disproportionate bias on the mathematical properties ignoring the evident economic forces. The selection, misleading as it can be, can have profound consequences on the institution as it ultimately determines how it perceives and controls risk.

As allocations based on regulatory measures may fail to capture economic incentives, such as profit maximization, several works have been authored in defining economic counterparts that are constructed organically. Notably, in Cummins (1988); Phillips et al. (1998) and recently in Bauer and Zanjani (2016), a thorough study of economic allocations were conducted within the various niches of market settings. Generally speaking, economic allocations are solutions to the maximization problem defined as the highest possible profit, being revenue minus costs, subject to consumer utility and regulatory solvency constraints. After a change of measure, that accounts for the marginal utility expressions, the risk measure inducing those allocations can be conveniently recovered as:

$$H(\boldsymbol{X}) = \exp\{\mathbb{E}\left[\log(Y)|Y > \mathrm{VaR}_q[Y]\right]\}, \tag{1.7}$$

where Y is the sum of the losses. The risk measure in (1.7) is not coherent as it is not translation invariant and may fail to satisfy sub-additivity. It also relates to geometric means (Hardy et al., 1988) and tail quasi-linear risk measures as in Bäuerle and Shushi (2020). Using Euler rule in (1.3), the equivalent economic allocations are expressed as:

$$\tilde{A}_i(\boldsymbol{X}) = \mathbb{E}\left[\frac{X_i}{Y}\middle| \; Y > \mathrm{VaR}_q[Y]\right], \; \forall i \in \mathcal{N}. \tag{1.8}$$

The similarities between the CTE allocation in (1.6) and the economic allocation in (1.8) is quite evident. Indeed, they both sum up to one and each express proportionality of risk attributed to the respective BU. However, in the former the ratio is non-stochastic as it is simply a division of the CTE allocation to that of the total risk, while in the latter, the

proportionality is a stochastic variate taking into account the randomness of the ratio. In terms of compositional maps, defined as $\boldsymbol{\mathcal{C}} : \mathbb{R}_+^n \mapsto \Delta^{n-1}$, $\boldsymbol{\mathcal{C}}(\boldsymbol{x}) = (\mathcal{C}_1(\boldsymbol{x}), \ldots, \mathcal{C}_n(\boldsymbol{x}))$, with $\mathcal{C}_i(x_1, \ldots, x_n) = x_i/y,\ \forall i \in \mathcal{N},\ y = \sum_{i=1}^{n} x_i$, which charts non-negative realizations to the n-dimensional standard simplex, the proportional CTE and economic allocations can be expressed as $\overline{A}_i(\boldsymbol{X}) = \mathcal{C}_i\left(\mathrm{CTE}_1(\boldsymbol{X}), \ldots, \mathrm{CTE}_n(\boldsymbol{X})\right)$, where $\mathrm{CTE}_j(\boldsymbol{X}) = \mathbb{E}\left[X_j | Y > \mathrm{VaR}_q[Y]\right]$, and $\tilde{A}_i(\boldsymbol{X}) = \mathbb{E}\left[\ \mathcal{C}_i(X_1, \ldots, X_n) \mid Y > \mathrm{VaR}_q[Y]\right]$, $\forall i \in \mathcal{N}$, respectively. The difference, made clear, lies in the order of the compositions i.e. whether placed outside or inside the expectation functional. Therefore, a natural question that arises, does there exists possible model settings which are invariant to the compositional order ? or equivalently, can the CTE and economic allocations coincide for certain choices of distributions ?

Surprisingly, the answer to the posed question is found in the language of trivial allocations. When an allocation is represented as a linear function of the related risk measure i.e. $A_i(\boldsymbol{X}) = \alpha_i + \beta_i H(\boldsymbol{X}),\ \alpha_i \geq 0, \beta_i > 0,\ \forall i \in \mathcal{N}$, and for every distortion of the underlying measure, then, under mild conditions, the laws that satisfy the linear relationship are exactly those that equate both allocations (compositionally invariant). This type of triviality can be referred to as proportional since $\frac{A_i(\boldsymbol{X})}{H(\boldsymbol{X})} = \beta_i,\ \forall i \in \mathcal{N}$, where $\alpha_i = 0, \forall i \in \mathcal{N}$, because of the aforementioned mild conditions. The constancy of proportional allocations allows for several distributions that satisfy a certain multivariate size-bias relationship. Though being stringent, it is not exceedingly so, as it allows for discernible dependence structures beyond the degenerate extremes. If, however, the allocations are represented as an absolute constant, under any measure distortion, i.e. $A_i(\boldsymbol{X}) = \alpha_i,\ \alpha_i > 0,\ \forall i \in \mathcal{N}$, then the only possible law is that of the counter-monotonic extreme. In two dimensions, counter-monotonicity is commonly known as the Fréchet lower bound which serves as the best possible dependence for diversification. In higher dimensions, however, counter-monotonicity thus far does not adhere to a unanimous definition and a unifying theory is very much needed. In this dissertation, we touch upon a possible encompassing definition, specifically that of non-increasing sets. Loosely speaking, elements of non-increasing sets have coordinates that move in "opposite directions". This means when the random losses $\boldsymbol{X}$ follow this extreme law, then absolute triviality of allocations is equivalent to their support being a non-increasing set

with $\mathbb{P}(g(\boldsymbol{X}) = c) = 1$, for some constant $c > 0$.

Both trivialities, proportional and absolute, imply the strong interplay between the allocations and the underlying dependence structures. Law characterizations, as illuminating as they can be, are not the only force operating within the capital allocation realm. Equally significant, for instance, is deriving allocation expressions based on a carefully built models. Constructing multivariate laws, therefore, is a necessary task that serves as a cornerstone tool for prudent risk management. Since scholarly attention has been drawn extensively to tail based measures and allocations, heavy-tailed laws embody a prime choice to properly model extreme events in insurance and finance (Embrechts et al., 1997). Particularly, the Pareto power law (Pareto, 1964) and its multivariate extensions (Arnold, 2015; Asimit et al., 2010; Su and Furman, 2017) stand out as ideal powerhouse befittingly capturing the intrinsic nature of the tails.

Due to the built-in complexity of multivariate distributions, scholars usually resort to exploiting stochastic representations to obtain novel distributions. Utilizing this machinery, we start, $\forall i \in \mathcal{N}$, with standalone losses X_i that possess a Pareto II distribution, which is expressed as a ratio of two independent variates, one exponential V_i and the other gamma W_i, i.e. $X_i = \frac{V_i}{W_i}$. Then using the location, scale and power transforms will yield the general class of Pareto IV:

$$X_i = \mu_i + \sigma_i \left(\frac{V_i}{W_i}\right)^{\frac{1}{\gamma_i}}, \ X_i > \mu_i, \tag{1.9}$$

where $\mu_i \in \mathbb{R}$, $\sigma_i > 0$ and $\gamma_i > 0$ are the location, scale and power parameters respectively. The representation in (1.9), has its origins in the multiplicative background model as one variate may convey a systemic while the other an idiosyncratic risk. So far, the stochastic representation describes the form of the margins with no dependence imposed. Furthermore, expressing V_i and W_i as operations of other independent variates will eventually incorporate an elegant overarching joint structure. Commencing by letting $\mathcal{M}_v, \mathcal{M}_w \subseteq \mathcal{P}(\mathcal{N})$, where $\mathcal{P}(\mathcal{N})$ is the power set of $\mathcal{N}$ without the empty set. Additionally, $\forall i \in \mathcal{N}$, we will set B_i to be a subset of either $\mathcal{M}_v$ or $\mathcal{M}_w$ representing the factors affecting the particular $i-$th BU.

Then choosing:

$$V_i = \min\left(E_B : B \in B_i\right), \tag{1.10}$$

to be the minimum of independent exponentials $E_B \sim \exp(\lambda_B)$ indexed by a power-set $\mathcal{M}_v$. The dependence of $\boldsymbol{V} = (V_1, \ldots, V_n)$, when $n = 2$, is often referred to as Marshall-Olkin (MO) (Marshall and Olkin, 1967), as it was pioneered by the two scholars studying simultaneous failure of jet engines. Nowadays, it has found its application within the actuarial practice as it is used to model concurrent deaths in joint life insurance products. This feat is characterized by the positive co-monotonic probability, $\mathbb{P}(V_1 = \cdots = V_n) > 0$, that is baked into the MO distribution.

Secondly, similar to $\boldsymbol{V}$, each gamma variate W_i is represented as the sum of a collection of independent factors i.e.:

$$W_i = \sum_{B \in B_i} Z_B, \tag{1.11}$$

such that $Z_B \sim \text{Gamma}(\alpha_B, 1)$ are gamma variables with different shapes and unit rate, and all indexed by the corresponding power-set $\mathcal{M}_w$. Thus, the random vector $\boldsymbol{W} = (W_1, \ldots, W_n)$ is tied by the additive gamma model that, due to its Laplace multiplicative structure, is extensively used in modelling losses within life and non-life insurance alike. Combining $\boldsymbol{V}$ and $\boldsymbol{W}$ results in a dependence structure of the joint losses $\boldsymbol{X}$ that is endowed with all the peculiar properties of the MO, additive and multiplicative compositions. It retains the heavy-tailed margins of Pareto IV, while serving as an encompassing rich law capable of portraying the highly non-normal world of financial risk.

Chapter 2

A primer on the generalized weighted risk functionals

2.1 Introduction

Consider non-negative random variables (RVs) $X,\ X_1, \ldots, X_n,\ n \in \mathbb{N}$, representing (insurance) losses, and let $\mathcal{X}$ denote a collection of such losses. For a Borel-measurable non-negative - and as a rule non-decreasing - 'weight' function $x \mapsto w(x),\ x \in [0,\ +\infty)$, the functionals $H_w : \mathcal{X} \to [0,\ +\infty) \cup \{+\infty\}$, such that the ratio of expectations below is well-defined and finite

$$H_w(X) = \frac{\mathbb{E}[Xw(X)]}{\mathbb{E}[w(X)]}, \tag{2.1}$$

are often called 'weighted' risk measures; also called actuarial premium calculation principles, if the bound $H_w(X) \geq \mathbb{E}[X]$ holds for those RVs $X \in \mathcal{X}$ that have finite means (e.g., Sendov et al. (2011)) i.e. non-negative loading is satisfied. Recently. the class of weighted functionals, H_w, has been connected to a theory of stress-testing, in which case weight functions play the role of 'stressing' mechanisms (e.g., Millossovich et al. (2021)). In what follows, H_w is referred to as the weighted risk functional(s) to recognize the manifold of existing applications across risk management and insurance.

In actuarial science, weighted risk functionals, H_w, were introduced by Furman and Zitikis (2007, 2008a) as a unifying class of risk functionals that comprises, e.g., the Value-

at-Risk and Conditional Tail Expectation risk measures, Esscher's, Kamps', and - under certain conditions - the distorted premiums, among other popular risk measures and actuarial premiums (we refer to, e.g., Choo and de Jong (2009); Kaluszka and Krzeszowiec (2012) and a more recent Castano-Martinez et al. (2020) and references therein, for examples of works that explore properties of the class of weighted risk functionals).

Generalizations of (2.1) have been developed in several directions, with the arguably simplest and most-popular of these directions having led to the rise of the notion of weighted risk capital allocations, put forward in Furman and Zitikis (2008b). More specifically, let $S = X_1 + \cdots + X_n$ denote the aggregate loss RV, then the functionals $A_w(X_i, S) : \mathcal{X} \times \mathcal{X} \to [0, \ +\infty) \cup \{+\infty\}$, such that the ratio of expectations below is well-defined and finite

$$A_w(X_i, S) = \frac{\mathbb{E}[X_i w(S)]}{\mathbb{E}[w(S)]}, \quad i \in \{1, \ldots, n\}, \tag{2.2}$$

are called weighted risk capital allocations (see e.g., Dhaene et al. (2012) as well as a more recent Guo et al. (2018) for details).

An alternative generalization of (2.1), which is referred to as a generalized weighted risk measure or premium in Furman and Zitikis (2009), is obtained by considering the class of functionals $H_{v,w} : \mathcal{X} \to [0, +\infty) \cup \{+\infty\}$, such that

$$H_{v,w}(X) = \frac{\mathbb{E}[v(X)w(X)]}{\mathbb{E}[w(X)]}, \tag{2.3}$$

where v, w are non-negative and Borel-measurable functions, $\mathbb{E}[v(X)w(X)] \in (0, +\infty)$, and $\mathbb{E}[w(X)] \in (0, +\infty)$ (see e.g.,Richards and Uhler (2019) for a study of the monotonicity of the class of generalized weighted risk functionals).

Yet another generalization of weighted risk functionals (2.1) was considered in Millossovich et al. (2021); Porth et al. (2014); Zhu et al. (2019) (also, Furman and Zitikis (2007) for an earlier note in this respect). This generalization hinges on the assumption that the weight function, $w(\cdot) \geq 0$ - non-decreasing in each variable and Borel-measurable - operates on vectors of loss RVs, that is $w : [0, \infty)^n \to [0, \infty)$. Clearly, if the weight function is chosen to be the simple 'sum' aggregation function, that is $w(x_1, \ldots, x_n) = x_1 + \cdots + x_n, \ x_1, \ldots x_n \geq 0$, then functional (2.2) is recovered. Zhu et al. (2019) focus on linear and log-linear combi-

nations of rate-making factors as the weight functions of interest and derive properties of what they call 'multivariate' weighted premiums (for various weight functions that arise in the context of a multivariate stress-testing theory, we refer to Millossovich et al. (2021)).

Speaking generally, aggregate financial positions are not simple sums of loss RVs (e.g., Jaworski et al. (2010), Chapter 5). Namely, let the function $g : [0, +\infty)^n \to [0, +\infty)$ be non-decreasing in each variable, Borel-measurable, and, for $\mathbf{x} = (x_1, \ldots, x_n) \in [0, +\infty)^n$, satisfy the boundary conditions

$$0 \leq \inf_{\mathbf{x}\in[0,+\infty)^n} g(\mathbf{x}) < \infty, \quad \text{and} \quad 0 < \sup_{\mathbf{x}\in[0,+\infty)^n} \mathrm{g}(\mathbf{x}) \leq +\infty,$$

that is the function $\mathbf{x} \mapsto g(\mathbf{x})$ is a general aggregation function (e.g., Grabisch et al. (2009)), and let $S_g = g(X_1, \ldots, X_n)$ denote the g-aggregate loss RV. Examples of aggregate functions are the already-mentioned (e.g., Zhu et al. (2019)) simple sum aggregation function $g(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$ and the exponential aggregation function $g(x_1, \ldots, x_n) = \sum_{i=1}^n e^{x_i}$. Other examples of aggregation functions are, e.g.,

- the maximum aggregation function - also, the largest order statistic - $g(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n)$;
- the minimum aggregation function -also, the smallest order statistic - $g(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n)$;
- the product aggregation function $g(x_1, \ldots, x_n) = x_1 \times \cdots \times x_n$;
- the log-sum-exp aggregation function $g(x_1, \ldots, x_n) = \log(e^{x_1} + \cdots + e^{x_n})$;
- the p-norm $g(x_1, \ldots, x_n) = (x_1^p + \cdots + x_n^p)^{1/p}$, where $p \in \mathbb{R}_+$.

The following arrow diagram (2.1) shows the flow of the internal model of aggregation, beginning with losses and ending with the weighted functionals (see Millossovich et al. (2021) for similar discussion).

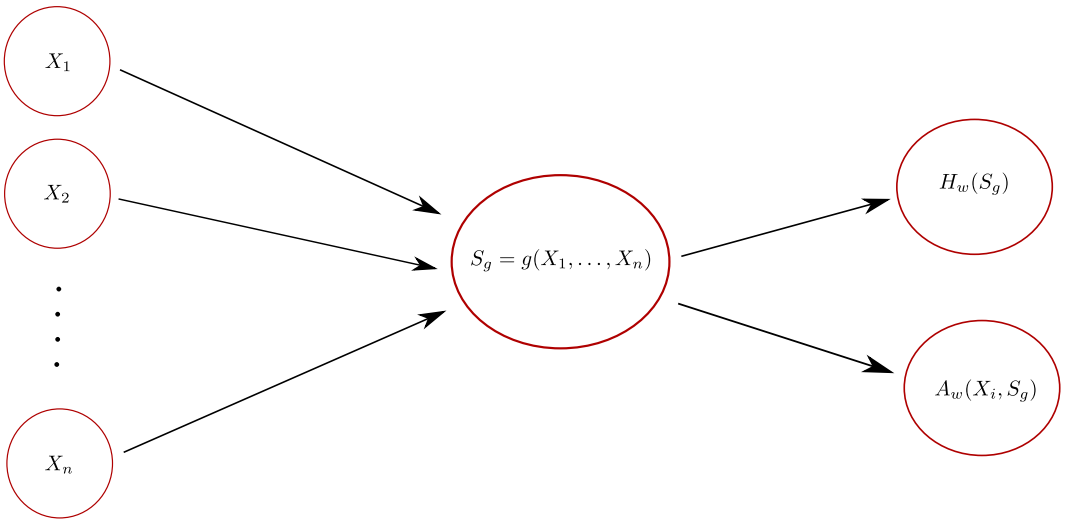


Figure 2.1: Diagram on the internal aggregation mechanism

In this paper, we work with the class of g-aggregation functions, such that the projection onto the i-th variable, P_i, $i \in \{1, \ldots, n\}$, equals that variable; namely, we require $P_i[g(x_1, \ldots, x_n)] = g(x_i) = x_i$. This additional condition, which by passing implies that the class of weight functions and the class of aggregation functions do not generally agree, is natural as an aggregation of a singleton is not really an aggregation. Keeping the above in mind, in this paper we work with the following generalized weighted risk functionals

$$H_w(S_g) = \frac{\mathbb{E}\big[S_g \times w \circ g(X_1, \ldots, X_n)\big]}{\mathbb{E}\big[w \circ g(X_1, \ldots, X_n)\big]} \tag{2.4}$$

and, for $i \in \{1, \ldots, n\}$,

$$A_w(X_i, S_g) = \frac{\mathbb{E}\big[X_i \times w \circ g(X_1, \ldots, X_n)\big]}{\mathbb{E}\big[w \circ g(X_1, \ldots, X_n)\big]}. \tag{2.5}$$

Clearly, if the g-aggregation function is the simple sum aggregation, then weighted risk functionals (2.4) and (2.5) reduce to the original ones. Summarized in Table 2.1 the weighted functionals for the popular choices of weight functions.

Name	$w(y)$	$H_w(S_g)$	$A_w(X_i, S_g)$
Net	const	$\mathbb{E}[S_g]$	$\mathbb{E}[X_i]$
Modified variance	y	$\mathbb{E}[S_g] + \frac{\mathrm{Var}(S_g)}{\mathbb{E}[S_g]}$	$\mathbb{E}[X_i] + \frac{\mathrm{Cov}(X_i,\, S_g)}{\mathbb{E}[S_g]}$
Size-biased	y^t	$\frac{\mathbb{E}\left[{S_g}^{1+t}\right]}{\mathbb{E}\left[{S_g}^{t}\right]}$	$\frac{\mathbb{E}\left[X_i \times {S_g}^{t}\right]}{\mathbb{E}\left[{S_g}^{t}\right]}$
Esscher	$e^{t\,y}$	$\frac{\mathbb{E}\left[S_g \times \exp\left(t\,S_g\right)\right]}{\mathbb{E}\left[\exp\left(t\,S_g\right)\right]}$	$\frac{\mathbb{E}\left[X_i \times \exp\left(t\,S_g\right)\right]}{\mathbb{E}\left[\exp\left(t\,S_g\right)\right]}$
Aumann-Shapley	$e^{t\,F(y)}$	$\frac{\mathbb{E}\left[S_g \times \exp\left(t\,F\left(S_g\right)\right)\right]}{\mathbb{E}\left[\exp\left(t\,F\left(S_g\right)\right)\right]}$	$\frac{\mathbb{E}\left[X_i \times \exp\left(t\,F\left(S_g\right]\right)\right]}{\mathbb{E}\left[\exp\left(t\,F\left(S_g\right)\right]\right]}$
Kamps	$1 - e^{-t\,y}$	$\frac{\mathbb{E}\left[S_g \times \left(1 - e^{-t\,S_g}\right)\right]}{\mathbb{E}\left[1 - e^{-t\,S_g}\right]}$	$\frac{\mathbb{E}\left[X_i \times \left(1 - e^{-t\,S_g}\right)\right]}{\mathbb{E}\left[1 - e^{-t\,S_g}\right]}$
Conditional tail expectation	$\mathbb{1}\{y \geq y_q\}$	$\mathbb{E}\left[S_g \middle| S_g \geq y_q\right]$	$\mathbb{E}\left[X_i \middle| S_g \geq y_q\right]$
Modified tail variance	$y\,\mathbb{1}\{y \geq y_q\}$	$\mathbb{E}\left[S_g \middle| S_g \geq y_q\right] + \frac{\mathrm{Var}\left(S_g \middle| S_g \geq y_q\right)}{\mathbb{E}\left[S_g \middle| S_g \geq y_q\right]}$	$\mathbb{E}\left[X_i \middle| S_g \geq y_q\right] + \frac{\mathrm{Cov}\left(X_i,\, S_g \middle| S_g \geq y_q\right)}{\mathbb{E}\left[S_g \middle| S_g \geq y_q\right]}$
Distorted	$h^{'}\left(\overline{F}(y)\right)$	$\mathbb{E}\left[S_g \times h^{'}\left(\overline{F}(S_g)\right)\right]$	$\mathbb{E}\left[X_i \times h^{'}\left(\overline{F}(S_g)\right)\right]$
Proportional hazard	$q\left(\overline{F}(y)\right)^{q-1}$	$q\,\mathbb{E}\left[S_g \times \left(\overline{F}(S_g)\right)^{q-1}\right]$	$q\,\mathbb{E}\left[X_i \times \left(\overline{F}(S_g)\right)^{q-1}\right]$

Table 2.1: Examples of weight function w accompanied with their associate risk measures and allocations. Within the table, we let $S_g = g(\boldsymbol{X})$, F and $\overline{F}$ denote the cumulative and decumulative distribution functions of S_g, respectively. Moreover, the distortion function $h : [0,1] \mapsto [0,1]$ is non-decreasing such that $h(0) = 0$ and $h(1) = 1$.

The rest of this paper is devoted to the study of various properties of functionals (2.4) and (2.5). More specifically in Section 2.2, we investigate bounds for the pairs of weighted risk functionals $H_w(S_{g_1})$ and $H_w(S_{g_2})$ as well as $A_w(X_i, S_{g_1})$ and $A_w(X_i, S_{g_2})$, $i \in \{1, \dots, n\}$, where the weight function is fixed and two distinct g-aggregation functions are considered. Notably, by selecting the g-aggregation functions, g_1 and g_2, such that $g_2 = \xi \circ g_1$ with the appropriately chosen non-decreasing and Borel-measurable function $\xi : [0, +\infty) \to [0, +\infty)$, the results in this section help compare the riskiness of aggregate losses subject to coverage modifications. Then in Section 2.3, we repeat the exercise by comparing weighted risk functionals $H_{w_1}(S_g)$ and $H_{w_2}(S_g)$ as well as $A_{w_1}(X_i, S_g)$ and $A_{w_2}(X_i, S_g)$, $i \in \{1, \dots, n\}$, which this time share the same aggregation function, but have different weight functions. Not surprisingly, a departure from the simple sum aggregation function results in a significant layer of complexity both when studying properties of generalized weighted risk functionals (2.4) and (2.5) and when evaluating them. In Section 2.4, we characterise those loss RVs, for which - irrespective of the choice of the g-aggregation function and the weight function - risk functional (2.5) is either trivially obtained from risk functional (2.4) or equals a constant (e.g., Guan et al. (2021) for a similar discussion).

2.2 Orders based on different aggregation functions but the same weight function

In what follows, we fix an atomless probability space and denote by $\mathcal{X}$ and $\mathcal{X}^n$ the set of all non-negative RVs and the set of all non-negative random vectors $\boldsymbol{X} = (X_1, \dots, X_n)$; in both cases these are interpreted as losses in a portfolio of losses, $\mathcal{N} = \{1, \dots, n\}$, $n \in \mathbb{N}$. The cumulative distribution function and the decumulative distribution function of the RVs $X \in \mathcal{X}$ and $\boldsymbol{X} \in \mathcal{X}^n$ are denoted by $F_X(x) = \mathbb{P}(X \leq x), \overline{F}_X(x) = 1 - F_X(x)$ and $F_{\boldsymbol{X}}(\mathbf{x}) = \mathbb{P}(\boldsymbol{X} \leq \mathbf{x})$, $\overline{F}_{\boldsymbol{X}}(\mathbf{x}) = \mathbb{P}(\boldsymbol{X} > \mathbf{x})$, respectively, for non-negative x and $\boldsymbol{x} = (x_1, \dots, x_n)$.

It is easy to see that the generalized weighted risk capital allocation as in Equation (2.5) satisfies the no-unjustified loading property as well as the non-negative loading property (Furman and Zitikis (2008b)) given that the weight function, w, is non-decreasing and the

RVs $X_i,\ i \in \mathcal{N}$ and S_g are positively quadrant dependent (PQD) i.e. $\mathbb{P}(X_i \leq z_1, S_g \leq z_2) \geq \mathbb{P}(X_i \leq z_1)\ \mathbb{P}(S_g \leq z_2)$ (see Lehmann (1966) for details). Also, while allocations, (2.5), are fully additive, $\sum_{i=1}^{n} A_w(X_i, S_g) = H_w(S_g)$, only if the g-aggregation function is the simple sum aggregation function ($S_g = S$), they admit a special form of the no-undercut property. Recall in this respect that the no-undercut property states that stand-alone losses are riskier - require more risk capital - that those losses that are considered a part of a portfolio of losses. The no-undercut property for the class of generalized weighted risk functionals (2.5) is formulated in the following proposition, which holds due to the Jensen's inequality.

Proposition 1. *If the g-aggregation function is convex, then we have*

$$g\big(A_w(X_1, S_g), \ldots, A_w(X_n, S_g)\big) \leq H_w(S_g) \quad \textit{holds for all } \boldsymbol{X} \in \mathcal{X}^n \textit{ and } S_g = g(\boldsymbol{X}). \tag{2.6}$$

Proof.

$$\begin{aligned} g\big(A_w(X_1, S_g), \ldots, A_w(X_n, S_g)\big) &= g\big(\mathbb{E}_{w(S_g)}[X_1], \ldots, \mathbb{E}_{w(S_g)}[X_n]\big), \\ &\leq \mathbb{E}_{w(S_g)}[g(X_1, \ldots, X_n)] \\ &= H_w(S_g). \end{aligned}$$

Where $\mathbb{E}_{w(S_g)}[\,\cdot\,] = \frac{\mathbb{E}[\,\cdot\, w(g(\boldsymbol{X}))]}{\mathbb{E}[w(g(\boldsymbol{X}))]}$. □

Clearly, the p-norm, $1 \leq p \leq \infty$, and log-sum-exp g-aggregation functions mentioned in Section 2.1 satisfy the convexity condition in Proposition 1.

Next we turn to the study of how different choices of aggregation functions impact the value of generalized weighted risk functionals. Two notational conveniences are in place. First, let $w(\cdot)$ be a weight function and X and Y be two loss RVs in $\mathcal{X}$, all such that the weighted risk functional $A_w(X, Y)$ is well-defined and finite. Then, similar to the notation in the proof of Proposition 1, $H_w(X, Y) =: \mathbb{E}_{w(Y)}[X]$, where the left-hand side is a w-biased expectation. Similarly, we can write (2.4) and (2.5) as $H_w(S_g) = \mathbb{E}_{w(S_g)}[S_g]$ and $A_w(X_i, S_g) = \mathbb{E}_{w(S_g)}[X_i]$, where $S_g = g(\boldsymbol{X})$ and $i \in \mathcal{N}$. Second, let us define the following regression functions, for $\boldsymbol{X} \in \mathcal{X}^n$ and $y \geq 0$,

$$h(y) = \mathbb{E}\big[S_{g_1} \,|\, S_{g_2} = y\big], \quad \text{where } S_{g_i} = g_i(\boldsymbol{X}),\ i = 1, 2$$

and

$$\tilde{h}(y) = \mathbb{E}\big[w(S_{g_2}) \,|\, w(S_{g_1}) = y\big].$$

Theorem 1. *For a given weight function w which is assumed to be strictly increasing, let*

$$H_j = H_w(S_{g_j}), \quad \textit{with } S_{g_j} = g_j(\boldsymbol{X}),\ j = 1, 2,$$

be the weighted risk measures associated with collection function g_1 and g_2. The following relationships hold:

$$\textit{If } h(y) \left\{ \begin{matrix} \geq \\ < \end{matrix} \right\} y \textit{ and the function } y \mapsto \frac{y}{\tilde{h}(y)},\ y \in \mathbb{R}_+, \textit{ is } \left\{ \begin{matrix} \textit{increasing} \\ \textit{decreasing} \end{matrix} \right\}, \textit{ then } H_1 \left\{ \begin{matrix} \geq \\ < \end{matrix} \right\} H_2.$$

In particular, $H_1 = H_2$ holds when $h(y) = y$ and the function $y \mapsto y/\tilde{h}(y)$ is constant.

Proof. See appendix i. □

Remark 1. *The relationship between $h(y)$ and y specified in Theorem 1 compares the order of the realizations of S_{g_1} and S_{g_2}, for a given portfolio $\boldsymbol{X} \in \mathcal{X}^n$, in an average sense. Clearly, the order between S_{g_1} and S_{g_2} implies the relationship between $h(y)$ and y. Namely, if $g_1(\boldsymbol{y}) \geq g_2(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}_+^n$, then $h(y) \geq y$, $\forall y \in \mathbb{R}_+$. The same argument holds if the inequalities are reversed.*

Remark 2. *In Theorem 1, we assume the weight function w to be strictly increasing. However, this assumption is violated when it comes to the tail conditional risk measures and allocations in which the weight function is set to be $w(y) = \mathbb{1}(y > d)$ for some $d \geq 0$. In this case, Theorem 1 remains true, but the monotone condition of $y \mapsto y/\tilde{h}(y)$ needs to be replaced by that of*

$$y \mapsto \frac{w(y)}{\mathbb{E}\big[w(S_{g_2}) | S_{g_1} = y\big]}, \quad \textit{where } S_{g_j} = g_j(\boldsymbol{X}) \textit{ and } j = 1, 2.$$

Remark 3. *The function $y \mapsto y/\tilde{h}(y)$ is increasing if $\tilde{h}(y)$ grows at a slower rate than y does. Note that the monotonicity property of $\tilde{h}$ is related to the dependence structure between $\widetilde{S}_1 := w(S_{g_1})$ and $\widetilde{S}_2 := w(S_{g_2})$. Specifically, if $\widetilde{S}_2$ is stochastically increasing in $\widetilde{S}_1$, i.e.,*

$$\mathbb{P}\big(\widetilde{S}_2 > x \,|\, \widetilde{S}_1 = y\big) \quad \textit{is increasing in } y \in \mathbb{R}_+ \textit{ for all } x \in \mathbb{R}_+,$$

then the function $y \mapsto \mathbb{E}\big[\widetilde{S}_2 \,|\, \widetilde{S}_1 = y\big]$ is increasing. Whether $\tilde{h}$ increases faster or slower than y depends on the marginal distributions and dependence of $(\widetilde{S}_1, \widetilde{S}_2)$, which are jointly determined by the choices of g_1, g_2, and w.

The following assertion further clarifies the monotonicity behavior of the function $y \mapsto y/\tilde{h}(y)$ when the dependence between $\widetilde{S}_1$ and $\widetilde{S}_2$ is chosen to be co-monotonic. Two variables (X, Y) are said to be co-monotonic if they can be written as non-decreasing functions of a common variable i.e. $(X, Y) \overset{d}{=} (\xi_1(Z), \xi_2(Z))$, for non-decreasing functions ξ_1 and ξ_2.

Theorem 2. *For all $\boldsymbol{y} \in \mathbb{R}_+^n$, let $\xi \circ g_1(\boldsymbol{y}) = g_2(\boldsymbol{y})$ where $\xi : \mathbb{R}_+ \to \mathbb{R}_+$ is increasing, hence $\widetilde{S}_1$ and $\widetilde{S}_2$ are co-monotonic. Moreover, suppose that the weight function w is diffentiable and log convex.*

$$\textit{If } \xi(y) \left\{ \begin{matrix} \leq \\ > \end{matrix} \right\} y \textit{ with } \xi'(y) \left\{ \begin{matrix} \leq \\ > \end{matrix} \right\} 1 \textit{ for } y \in \mathbb{R}_+, \textit{ then we have } H_1 \left\{ \begin{matrix} \geq \\ < \end{matrix} \right\} H_2.$$

Proof. See appendix ii. □

Remark 4. *Among the examples outlined in Table 2.1, the following principles are associated with a log convex weight function: Net, Esscher, Aumann-Shapley (when F is convex), distorted (when $h' \circ \overline{F}$ is log convex), and proportional hazard (when $\overline{F}$ is log convex).*

The mere ordering of the g-aggregation functions is not sufficient in order to have the generalized weighted risk functionals ordered, as becomes evident from the following example.

Example 1. *Suppose that the risk collection $g_1(\boldsymbol{X}) \sim \mathrm{Pa(II)}(\alpha, \theta)$, the Pareto distribution of the second kind with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$, whose probability density function is given by*

$$f(x) = \frac{\alpha}{\theta}\left(1 + \frac{x}{\theta}\right)^{-(\alpha+1)}, \quad x > 0.$$

Moreover, let $\xi(y) = \max(0, c\,y - d)$ *which represents the risk reduction due to the introduction of a coinsurance factor* $c \in (0,1)$ *and a deductible* $d > 0$. *It is straightforward to check that* $\xi(y) \le y$ *and* $\xi'(y) \le 1$, *and thus the first set of assumptions about* ξ *in Theorem 2 are satisfied. Meanwhile, let* $w(y) = y^b$, $b > 0$, *hence the log-convexity condition on* w *required in Theorem 2 is violated.*

Let $S_g = g_1(\boldsymbol{X})$, *then we have*

$$H_1 = \frac{\mathbb{E}\big[{S_g}^{b+1}\big]}{\mathbb{E}\big[{S_g}^{b}\big]} = \theta\,\frac{1+b}{\alpha - b - 1},$$

in which we require $\alpha > 1 + b$ *such that the expectations above are well-defined. Under the same assumption, we have*

$$\begin{aligned} H_2 &= \frac{\mathbb{E}\big[\xi(S_g) \times w \circ \xi(S_g)\big]}{\mathbb{E}[w \circ \xi(S_g)]} \\ &= \frac{c^{b+1}\,\mathbb{E}\Big[(S_g - d/c)^{b+1} \,|\, S_g > d/c\Big]}{c^{b}\,\mathbb{E}\Big[(S_g - d/c)^{b} \,|\, S_g > d/c\Big]}. \end{aligned}$$

Note that ${S_g}^* := (S_g - d/c \,|\, S_g > d/c) \sim \mathrm{Pa(II)}\big(\alpha,\, d/c + \theta\big)$, *thus*

$$H_2 = c\,\frac{\mathbb{E}\Big[({S_g}^*)^{b+1}\Big]}{\mathbb{E}\Big[({S_g}^*)^{b}\Big]} = c\,(d/c + \theta)\,\frac{1+b}{\alpha - b - 1} = (c\,\theta + d)\,H_1.$$

Set $\theta = 1$, *then we obtain* $H_1 \ge H_2$ *if* $c + d \le 1$, *and* $H_1 < H_2$ *if* $c + d > 1$. *Collectively, this example shows that the order between* g_1 *and* g_2 *is not sufficient to determine the order between* H_1 *and* H_2.

Next we turn to study the impact of the choice of the g-aggregation function on functionals (2.5). At the outset, let us define, for $S_{g_j} = g_j(\boldsymbol{X}), j \in \{1,2\}$, $i \in \mathcal{N}$, and $y \ge 0$, the following regression function

$$\ell_{i,j}(y) = \mathbb{E}\big[w(S_{g_j}) | X_i = y\big].$$

Theorem 3. *For a given weight function w, let*

$$A_{i,j} := A_w(X_i, S_{g_j}), \quad j = 1, 2, \text{ and } i \in \mathcal{N},$$

be generalized weighted risk functionals á la (2.5) associated with the aggregation functions g_1 and g_2. The following relationships hold:

$$\text{If the function } y \mapsto \frac{\ell_{i,1}(y)}{\ell_{i,2}(y)} \text{ is } \left\{ \begin{array}{l} \text{increasing} \\ \text{decreasing} \end{array} \right\} \text{ on } y \in [0, +\infty), \text{ then } A_{i,1} \left\{ \begin{array}{l} \geq \\ < \end{array} \right\} A_{i,2}.$$

In particular, $A_{i,1} = A_{i,2}$ if $y \mapsto \ell_{i,1}(y)/\ell_{i,2}(y)$ is a constant function.

Proof. The result follows from Proposition 3.1 of Furman and Zitikis (2008b). □

Knowing the order between g_1 and g_2 may not suffice to conclude the monotonicity behavior of the ratio $y \mapsto \ell_{i,1}(y)/\ell_{i,2}(y)$. The following proposition further confirms the critical role that the weight function w plays in shaping the order between functionals $A_{i,1}$ and $A_{i,2}$.

Theorem 4. *Suppose that individual risks within a portfolio have marginal CDFs F_{X_i}, $i \in \mathcal{N}$ and are co-monotonic, namely $X_i = F_i^{-1}(U)$ almost surely, where $U \sim$ Uniform$[0, 1]$. Let the element-wise increasing aggregation functions satisfy $\xi \circ g_1(\boldsymbol{y}) = g_2(\boldsymbol{y})$ for $\boldsymbol{y} \in \mathbb{R}_+^n$, where $\xi : [0, +\infty) \to [0, +\infty)$ is increasing. Further assume that the weight function, w, is differentiable and log convex. The following relationships hold*

$$\text{If } \xi(y) \left\{ \begin{array}{l} \leq \\ > \end{array} \right\} y \text{ with } \xi'(y) \left\{ \begin{array}{l} \leq \\ > \end{array} \right\} 1 \text{ for } y \in [0, +\infty), \text{ then we have } A_{i,1} \left\{ \begin{array}{l} \geq \\ < \end{array} \right\} A_{i,2} \text{ for } i \in \mathcal{N}.$$

Proof. See appendix iii. □

In Theorem 4, the log convexity condition on the weight function w is again minimal, which is reaffirmed in the example below.

Example 2. *Consider two loss RVs distributed Pareto of the second kind, that is $X_i \sim$ Pa(II)(α, θ_i), $\alpha \in \mathbb{R}_+$, $\theta_i \in \mathbb{R}_+$, $i = 1, 2$, and assume that the RVs X_1 and X_2 are co-monotonic. Set $g_1(\boldsymbol{x}) = x_1 + x_2$ and $g_2(\boldsymbol{x}) = \xi \circ g_1(\boldsymbol{x})$, where $\xi(y) = \max(0, y - d)$ represents the risk reduction function due to the inclusion of a deductible $d > 0$. Furthermore, consider*

the same weight function as in Example 1, i.e., $w(x) = x^b$, $b \in \mathbb{R}_+$, which is log concave and thus violates the conditions on w in Theorem 4. Let $S_g = g_1(\boldsymbol{X}) = X_1 + X_2$, then since the loss RVs are co-monotonic, we have $S_g \sim \mathrm{Pa(II)}(\alpha, \theta^*)$, where $\theta^* = \theta_1 + \theta_2$. Finally, assume that the succeeding expectations are well-defined and finite, or equivalently $\alpha > b + 1$, and have

$$\begin{aligned} A_{i,1} &= \frac{\mathbb{E}\big[X_i \times w(S_g)\big]}{\mathbb{E}\big[w(S_g)\big]} \\ &= \frac{\mathbb{E}\big[X_i\,(X_1 + X_2)^b\big]}{\mathbb{E}\big[(X_1 + X_2)^b\big]} \\ &= \frac{\mathbb{E}\big[X_i^{1+b}\big]}{\mathbb{E}\big[X_i^b\big]} \\ &= \theta_i\,\frac{1+b}{\alpha - b - 1}, \qquad i = 1, 2. \end{aligned}$$

Also, we have, for $i = 1, 2,$,

$$\begin{aligned} A_{i,2} &= \frac{\mathbb{E}\big[X_i \times w \circ \xi(S_g)\big]}{\mathbb{E}\big[w \circ \xi(S_g)\big]} \\ &= \frac{\mathbb{E}\big[X_i\,(X_1 + X_2 - d)^b \,|\, X_1 + X_2 > d\big]}{\mathbb{E}\big[(X_1 + X_2 - d)^b \,|\, X_1 + X_2 > d/c\big]} \\ &= \frac{\theta_i}{\theta^*}\,\frac{\mathbb{E}\big[S_g\,(S_g - d)^b \,|\, S_g > d\big]}{\mathbb{E}\big[(S_g - d)^b \,|\, S_g > d\big]} \\ &= \frac{\theta_i}{\theta^*}\,\frac{\mathbb{E}\big[(S_g - d)^{b+1} \,|\, S_g > d\big] + d\,\mathbb{E}\big[(S_g - d)^b \,|\, S_g > d\big]}{\mathbb{E}\big[(S_g - d)^b \,|\, S_g > d\big]}. \end{aligned}$$

Note that $S_g - d \,|\, S_g > d \sim \mathrm{Pa(II)}(\alpha, d + \theta^*)$, and so we obtain

$$A_{i,2} = \frac{\theta_i}{\theta^*}\Big[\big(d + \theta^*\big)\,\frac{1+b}{\alpha - b - 1} + d\Big] = A_{i,1} + \frac{\alpha\, d\, \theta_i}{(\alpha - b - 1)\,\theta^*} \geq A_{i,1}, \qquad i = 1, 2.$$

In conclusion, we have seen in this example that if the log convexity assumption on the weight function w is violated, then the desired relationship $A_{i,1} \geq A_{i,2}$ reported in Theorem 4 can not be guaranteed. Thereby, the order between the aggregation functions g_1 and g_2 is not

sufficient to determine the order between the associated generalized weighted risk functionals. Moreover, if the weight function, w, is not log convex, then (2.5) may fail to capture the risk reduction due to the introduction of policy modifications.

2.3 Orders based on different weight functions but the same aggregation function

We are now in a position to examine the role that the weight function plays in the determination of the order of risk functionals (2.4) and (2.5), given that they share the same aggregation function. We start with the study of the former generalized weighted risk functional, and our observations are summarized in the following theorem.

Theorem 5. *For a given collection function $g : [0, +\infty)^n \to [0, +\infty)$, let*

$$\tilde{H}_j = H_{w_j}(S_g), \quad j = 1, 2$$

be two generalized weighted risk functionals associated with the weight functions w_1 and w_2. Then the following relationships holds

$$\text{If the function } y \mapsto \tfrac{w_1(y)}{w_2(y)},\ y \in \mathbb{R}_+, \text{ is } \left\{ \begin{array}{l} \text{increasing} \\ \text{decreasing} \end{array} \right\}, \text{ then } \tilde{H}_1 \left\{ \begin{array}{l} > \\ < \end{array} \right\} \tilde{H}_2.$$

Particularly, if $y \mapsto w_1(y)/w_2(y) \equiv c$, $y \in \mathbb{R}_+$, for some constant $c \in \mathbb{R}_+$, then $\tilde{H}_1 = \tilde{H}_2$.

Proof. The result follows from Theorem 4 of Patil and Rao (1978) and statement (4.3) of Furman and Zitikis (2008a). □

Interestingly, Theorem 5 shows that for a given loss position $\boldsymbol{X} \in \mathcal{X}^n$ with a fixed aggregation function g, the monotonicity behaviour order of the ratio of the two weight functions w_1 and w_2 can yield the order of the associated generalized weighted risk functionals. Furthermore, suppose that the g-aggregate RV S_g has a continuous CDF, then Table 2.2 summarizes the conditions under which the ratio $y \mapsto w_1(y)/w_2(y)$ is increasing, and thus $H_{w_1}(S_g) > H_{w_2}(S_g)$ as per Theorem 5.

w_2 \ w_1	Size-biased $w_1(y) = y^{t_1}$	Esscher $w_1(y) = e^{t_1 y}$	Aumann-Shapley $w_1(y) = e^{t_1 F(y)}$	Kamps $w_1(y) = 1 - e^{-t_1 y}$	Distorted $w_1(y) = h_1^{'}\left(\overline{F}(y)\right)$
Size-biased $w_2(y) = y^{t_2}$	$t_1 > t_2$				
Esscher $w_2(y) = e^{t_2 y}$	$y < \frac{t_1}{t_2}$, $\forall y \in \mathbb{S}_g$	$t_1 > t_2$			
Aumann-Shapley $w_2(y) = e^{t_2 F(y)}$	$f(y)\, y < \frac{t_1}{t_2}$, $\forall y \in \mathbb{S}_g$	$f(y) < \frac{t_1}{t_2}$, $\forall y \in \mathbb{S}_g$	$t_1 > t_2$		
Kamps $w_2(y) = 1 - e^{-t_2 y}$	$t_1 > 1$	$y > \frac{1}{t_2} \log\left(\frac{t_2}{t_1} + 1\right)$, $\forall y \in \mathbb{S}_g$	$y > \frac{1}{t_2} \log\left(\frac{t_2}{t_1\, f(y)} + 1\right)$, $\forall y \in \mathbb{S}_g$	$t_2 > t_1$	
Distorted $w_2(y) = h_2^{'}\left(\overline{F}(y)\right)$	$h_2^{'}(z), h_2^{''}(z) > 0$, $\forall z \in (0,1)$	$h_2^{'}(z), h_2^{''}(z) > 0$, $\forall z \in (0,1)$	$h_2^{'}(z), h_2^{''}(z) > 0$, $\forall z \in (0,1)$	$h_2^{'}(z), h_2^{''}(z) > 0$ $\forall z \in (0,1)$	$\left(h_1^{'}(z), h_2^{'}(z)\right)^{'} > 0$ $\forall z \in (0,1)$

Table 2.2: Summary of the conditions such that the function $y \mapsto w_1(y)/w_2(y)$, $y \in \mathbb{R}_+$, is increasing when the g-aggregate RV S_g has a continuous CDF. If the inequalities are reversed, then the function $y \mapsto w_1(y)/w_2(y)$ becomes decreasing. The ranges of the parameters are $t_1, t_2 \in \mathbb{R}_+$, and $f_{S_g}(y),\ y \in \mathbb{R}_+$ denotes the probability density function of S_g.

Several observations pertaining to the conditions outlined in Table 2.2 are warranted. First, note that the net premium risk functional and the modified variance risk functional are special cases of the size-biased risk functional with $t = 0$ and $t = 1$, respectively (see, Table 2.1). Thereby, Table 2.2 can be immediately used to study the order of these two risk functionals as well as other size-biased risk functionals.

Second, the diagonal cells in Table 2.2 indicate that for any two weight functions belonging to the same class, it is sufficient to use the value of the t parameter to determine the order of the associated weighted risk functionals.

Third, when comparisons are made across different classes of weight functions, then the monotonicity behavior of the function $y \mapsto w_1(y)/w_2(y)$ may depend on the support and/or the probability distribution of the RV S_g, except for the comparison between the size-biased and Kamps' risk functionals. To be specific, for the comparison between the size-biased and Esscher's risk functionals, the function $y \mapsto w_1(y)/w_2(y)$ is increasing (resp. decreasing) only when the g-aggregate RV S_g is bounded from above (resp. below) by t_1/t_2. When it comes to the comparison between the size-biased and the Aumann-Shapley risk functionals, note that commonly used continuous distributions such as the ones with unbounded supports outlined in the distribution inventory of Klugman et al. (2012), have $y \mapsto f(y)\, y$ bounded for all $y \in \mathbb{R}_+$, where f denotes the respective density. Therefore, we can find sufficiently large t_1 and/or small t_2 such that the corresponding inequality condition is satisfied, and thus the ratio $w_1(y)/w_2(y)$ is increasing. On a different note, it is also worth mentioning that $y \mapsto f(y)\, y$ is not always bounded from above. A counter example is the arcsine distribution, or more generally a Beta distribution with the second shape parameter being less than one, whose PDF is given by

$$f(y) = \frac{1}{\pi\sqrt{y(1-y)}}, \qquad y \in (0,1);$$

it is evident that $\lim_{y\uparrow 1} f(y)\, y = +\infty$. It is also possible that $y \mapsto f(y)\, y$ is bounded from below by a positive value (e.g., the right-shifted uniform distribution). In this case, we can find an appropriate pair of t_1 and t_2 such that $f(y)\, y > t_1/t_2$, thus $y \mapsto w_1(y)/w_2(y)$ is decreasing for all $y \in \mathbb{S}_g$. For common continuous distributions such as gamma, log-normal,

Pareto, and Weibull, we have $\lim_{y\downarrow 0} f(y)\, y = 0$, thus it is impossible that $f(y)\, y > t_1/t_2$ for all $y \in \mathbb{R}_+$, and $y \mapsto w_1(y)/w_2(y)$ can not be decreasing.

Turning to the Esscher functional, its comparison with Kamps functional suggests that the support of collection RV S_g must have a positive lower (resp. upper) bound ${t_2}^{-1} \log(t_2/t_1 + 1)$ such that $y \mapsto w_1(y)/w_2(y)$ is increasing (resp. decreasing). To implement the comparison between the Esscher and Aumann-Shapley functionals, we require the PDF of S_g to be bounded from above or from below by a positive value. When S_g has an unbounded support, then it is impossible that $f(y)$ has a positive lower bound, thus $y \mapsto w_1(y)/w_2(y)$ can not be increasing.

Penultimately, let us consider the comparison between the Kamps and Aumann-Shapley functionals. The inequality for ensuring the increasing property for w_1/w_2 depends on both the support of S_g and the behaviour of the PDF f. If the PDF f is unbounded at a positive point, then $\log\left(t_2(t_1\, f(y))^{-1} + 1\right) \to 0$ as y approaches to that point. So the corresponding inequality condition specified in Table 2.2 holds, and the increasing property of w_1/w_2 can be established in a neighbourhood of that positive point. However, when the PDF $f(y)$ converges to zero as y approaches to a finite point (e.g., the Gamma distribution having PDF: $f(y) = \frac{1}{2} y^2 e^{-y}$, $y \in (0, \infty)$), then the function $\log\left(t_2(t_1\, f(y))^{-1} + 1\right) \to \infty$ as y approaches zero), it is impossible that $y > \log\left(t_2(t_1\, f(y))^{-1} + 1\right)$ for all $y \in \mathbb{R}_+$, thus the function $y \mapsto w_1(y)/w_2(y)$ can not be increasing.

Finally, the comparison between the distortion functional and all others is quite simple as it can solely depend upon h. The sufficient condition of positive first and second derivatives, of h, ensures that $y \mapsto w_1(y)/w_2(y)$ is increasing. As h is usually chosen increasing, then $h^{'} > 0$ is automatically satisfied. The only remaining restriction is that imposed on the second degree. If h is convex, meaning $h^{''} > 0$, then we get the desired monotonicity behaviour of the ratio. In the literature, nonetheless, h is taken to be concave as it is equivalent to the coherence of the underlying distorted risk functional. Thus, for the coherent class, $h^{''} \leq 0$, and therefore the function $y \mapsto w_1(y)/w_2(y)$ must not be monotonic.

In what follows, we proceed to studying the conditions for determining the order of generalized weighted risk functionals (2.5) subject to different weight functions and common

aggregation function. To this end, we need the following additional notation

$$\tilde{\ell}_{i,j}(y) = \mathbb{E}[w_j(S_g)|X_i = y], \quad \text{where } S_g = g(\boldsymbol{X}),\ j \in \{1,2\},\ i \in \mathcal{N}, \text{ and } y \geq 0.$$

Theorem 6. *For a given aggregation function g, let*

$$\tilde{A}_{i,j} = A_{w_j}(X_i, S_g), \quad \textit{with } S_g = g(\boldsymbol{X}),\ j = 1,2$$

be the weighted risk functionals associated with the weight functions w_1 and w_2.

$$\textit{If the function } y \mapsto \tfrac{\tilde{\ell}_{i,1}(y)}{\tilde{\ell}_{i,2}(y)},\ y \in \mathbb{R}_+, \textit{ is } \left\{ \begin{array}{l} \textit{increasing} \\ \textit{decreasing} \end{array} \right\}, \textit{ then } \tilde{A}_{i,1} \left\{ \begin{array}{l} > \\ < \end{array} \right\} \tilde{A}_{i,2}.$$

In particular, if $\tilde{\ell}_{i,1}(y)/\tilde{\ell}_{i,2}(y) \equiv c$ for some constant $c \in \mathbb{R}_+$, then $\tilde{A}_{i,1} = \tilde{A}_{i,2}$.

Proof. The result follows from Proposition 3.1 of Furman and Zitikis (2008b). □

Interestingly, for comontonic losses Theorem 6 simplifies significantly.

Theorem 7. *Let us consider $\boldsymbol{X} = (X_1, \ldots, X_n)$ with $X_i = F_{X_i}^{-1}(U)$, where $U \sim \text{Uniform}(0,1)$ and F_{X_i} is the CDF of the RV X_i, $i \in \mathcal{N}$. Fix a component-wise increasing aggregation function g, then following relationships hold*

$$\textit{If the function } y \mapsto \frac{w_1(y)}{w_2(y)},\ y \in \mathbb{R}_+, \textit{ is } \left\{ \begin{array}{l} \textit{increasing} \\ \textit{decreasing} \end{array} \right\}, \textit{ then } \tilde{A}_{i,1} \left\{ \begin{array}{l} > \\ < \end{array} \right\} \tilde{A}_{i,2}.$$

Proof. See appendix iv. □

Theorem 7 seems to suggest that the monotonicity behavior of ratios of the weight function may be a decisive factor as to the orders of risk functionals (2.5), as it is in the context of risk functionals (2.4). The next example shows that it is not the case if the co-monotonicity assumption on the losses of interest is removed.

Example 3. *Consider the loss RV $\boldsymbol{X} = (X_1, X_2)$, whose joint probabilistic behavior is governed by a two-component mixture of gamma distribution with joint PDF (Chen et al., 2021):*

$$f_{X_1,X_2}(x_1,x_2) = p \prod_{i=1}^{2} \frac{X_i^{\alpha_{i1}-1}\theta_i^{\alpha_{i1}}}{\Gamma(\alpha_{i1})} e^{-\theta_i X_i} + (1-p) \prod_{i=1}^{2} \frac{X_i^{\alpha_{i2}-1}\theta_i^{\alpha_{i2}}}{\Gamma(\alpha_{i2})} e^{-\theta_i X_i}, \qquad x_1, x_2 > 0,\ p \in (0,1).$$

In this example, we set the aggregation function $g(\boldsymbol{x}) = x_1 + x_2$*, and consider two weight functions* $w_j(y) = y^{n_j}$*, where* $n_j \in \mathbb{N}$*,* $j = 1, 2$*, with* $n_1 \geq n_2$*. Clearly, the function* $y \mapsto w_1(y)/w_2(y)$ *is increasing. Next, let us fix* $p = 0.5$*,* $\alpha_{11} = 2$*,* $\alpha_{12} = 1$*,* $\alpha_{21} = \alpha > 0$*, and* $\alpha_{22} = 8$*. Also, let* $n_1 = 2$ *and* $n_2 = 1$*. Figure 2.2 depicts the Pearson correlation of the pair of losses,* (X_1, X_2)*, and the corresponding weighted risk functionals (2.5) as functions of* α*, which are computed based on Corollary 3 and Proposition 2 of (Chen et al., 2021). As observed, the order* $\tilde{A}_{1,1} < \tilde{A}_{1,2}$ *holds for smaller* $\alpha \in \mathbb{R}_+$*. Hence, the increasing property of* w_1/w_2 *is not sufficient to yield the desired order* $\tilde{A}_{1,1} > \tilde{A}_{1,2}$ *as per Theorem 7 after the co-monotonicity assumption is removed. Nevertheless, as the value of* α *rises, the loss RVs* X_1 *and* X_2 *become more positively correlated, as demonstrated by the increasing pattern of the Pearson correlation, and the order between* $\tilde{A}_{1,1}$ *and* $\tilde{A}_{1,2}$ *tends to coincide with the one suggested by Theorem 7 for co-monotonic losses.*

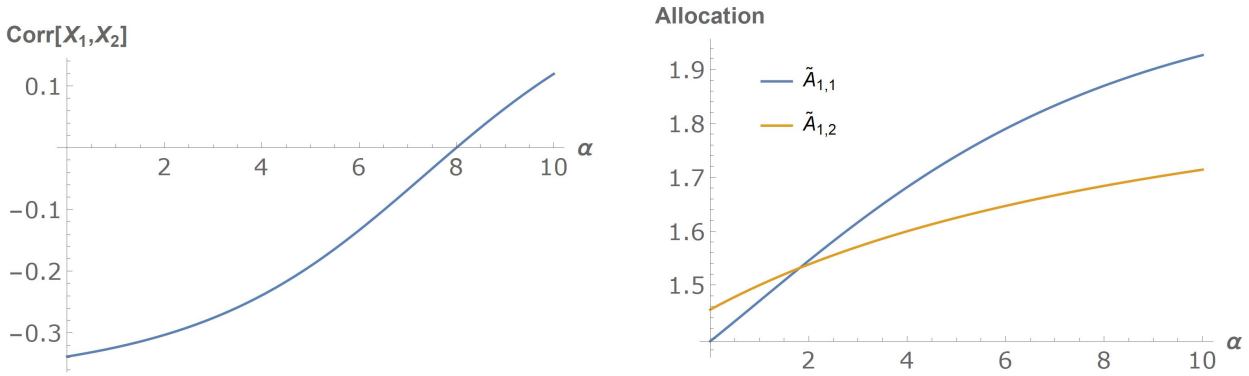


Figure 2.2: Plots of the Pearson correlation of the pair of loss RVs (X_1, X_2) and the weighted risk functional (2.5) as a function of $\alpha \in (0, 10)$.

In summary, the study in this section suggests that while the monotonicity behavior of the ratio of two weight functions, $y \mapsto w_1(y)$ and $y \mapsto w_2(y)$, can play a decisive role in the determination of the orders of risk functionals (2.4), this is not generally the case in the context of risk functionals (2.5).

2.4 Afterthoughts and related results

The results we have established thus far suggest that the choices of g and w have rather complex interactive effects on the amounts of total capital and allocations induced by the weighted functional method. Selecting the appropriate aggregation function g to work with should be driven by the business problem at hand, yet the choice of weight function w is more subjective and ad hoc. When the weighted functional method is applied, ideally the portfolio $\boldsymbol{X}$ is robust (in some senses to be specified below) to different choice of w. Characterizing such risk portfolio is what we aim to study in this section.

Consider a general random pair $(X, S_g) \in \mathcal{X}^2$, Furman and Zitikis (2008b) established a sufficient condition in terms of the linearity of the regression function $y \mapsto \mathbb{E}[X|S_g = y]$, such that

$$\frac{A_w(X, S_g) - \mathbb{E}[X]}{H_w(S_g) - \mathbb{E}[S_g]} \equiv c \tag{2.7}$$

for a constant c which depends only on the distribution of (X, S_g) but not the choice of weight function w. Notably, relationship (2.7) is reminiscent of the capital asset pricing model widely adopted in finance, in which the subjectivity of the decision maker utility function is avoided. Meanwhile, relationship (2.7) implies a linear regression of the allocation on the total capital:

$$A_w(X, S_g) = \alpha + \beta\, H_w(S_g), \tag{2.8}$$

in which α and β depend on the distribution of (X, S_g) but not the choice of w. In this paper, we consider a special case of (X, S_g) in which $X = X_i$ and $S_g = g(\boldsymbol{X})$ for a risk portfolio $\boldsymbol{X}$. The same relation in equation (2.8) has been previously studied in Furman et al. (2018b) when g is the canonical sum i.e. $g(\boldsymbol{X}) = \sum_{j=1}^{n} X_j$. This additional dependence imposed between X and S_g enables us to generalize the sufficient condition derived in Furman and Zitikis (2008b) to a necessary and sufficient result for the desirable relationships (2.7) and (2.8) to hold. Namely, fix the collection function g, and let $S_g = g(\boldsymbol{X})$, we are interested in

deriving a law characterization of the set:

$$\mathcal{T} = \big\{\boldsymbol{X} \in \mathcal{X}^n : \; A_w(X_i, S_g) = \alpha_i + \beta_i \, H_w(S_g), \text{ for any choice of } w \text{ and } i \in \mathcal{N}\big\}.$$

Where the set definition holds for some non-negative α_i and β_i, $\forall i \in \mathcal{N}$. Particularly, we are interested in characterising the set when $\alpha_i = 0, \; \forall i \in \mathcal{N}$, denoted as:

$$\mathcal{A} = \big\{\boldsymbol{X} \in \mathcal{X}^n : \; A_w(X_i, S_g) = \beta_i \, H_w(S_g), \text{ for any choice of } w \text{ and } i \in \mathcal{N}\big\},$$

and when $\beta_i = 0, \; \forall i \in \mathcal{N}$, given by:

$$\mathcal{B} = \big\{\boldsymbol{X} \in \mathcal{X}^n : \; A_w(X_i, S_g) = \alpha_i, \text{ for any choice of } w \text{ and } i \in \mathcal{N}\big\}.$$

We shall associate the terms proportional and absolute triviality with the sets $\mathcal{A}$ and $\mathcal{B}$, respectively. To begin our characterization, the following lemma is of auxiliary importance.

Lemma 1. *For a fixed collection function g, let $S_g = g(\boldsymbol{X})$. It holds that $\boldsymbol{X} \in \mathcal{T}$ if and only if*

$$\mathbb{E}[X_i \,|\, S_g] \equiv \alpha_i + \beta_i \, S_g, \quad i \in \mathcal{N}. \tag{2.9}$$

Proof. The proof follows similar arguments as in Theorem 3.1 in Furman et al. (2018b). □

Moreover, we need the following notion of multivariate size-biased transform.

Definition 1. *Let $\boldsymbol{X} \in \mathcal{X}^n$ be a loss RV with positive univariate coordinates $X_i \in L^1$, $i \in \mathcal{N}$. Then the multivariate coordinate-wise size-biased counterpart of $\boldsymbol{X}$, denoted by $\boldsymbol{X}^{(i)}$, is*

$$\mathbb{P}\big(\boldsymbol{X}^{(i)} \in \mathrm{d}\boldsymbol{x}\big) = \frac{x_i}{\mathbb{E}[X_i]} \, \mathbb{P}\big(\boldsymbol{X} \in \mathrm{d}\boldsymbol{x}\big) \quad \textit{for all } \boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_+. \tag{2.10}$$

Remark 5. *When $n = 1$, then the multivariate size-biased transform considered in Definition 1 reduces to the classical notion of univariate size-biased transform (Patil and Rao, 1978).*

Namely, the size-biased conterpart of $X \in \mathcal{X} \cap L^1$, denoted by X^*, has density function:

$$\mathbb{P}\big(X^* \in \mathrm{d}x\big) = \frac{x}{\mathbb{E}[X]}\,\mathbb{P}\big(X \in \mathrm{d}x\big), \quad x \in \mathbb{R}_+.$$

We also need some additional assumptions on the choice of collection function g. First, for a risk portfolio $\boldsymbol{X}$, suppose that

$$\mathbb{P}\big(g(\boldsymbol{X}) \in [0, \epsilon)\big) > 0, \quad \text{for any } \epsilon > 0, \tag{2.11}$$

meaning that it is possible the portfolio risk is arbitrarily small. Second, we require the collection function to satisfy the following, call it the non-neglecting condition:

$$g(\boldsymbol{x}) \downarrow 0 \quad \text{only when } \max(x_1, \ldots, x_n) \downarrow 0. \tag{2.12}$$

The non-neglecting condition ensures that no positive risks of a portfolio are neglected when the collection function returns zero.

First we will start with the characterization of proportional triviality i.e. the set $\mathcal{A}$.

Proposition 2. *Consider a risk portfolio $\boldsymbol{X}$ with positive univariate coordinates $X_i \in L^1$, $i \in \mathcal{N}$. Moreover, fix the collection function g. Suppose that the conditions in (2.11) and (2.12) hold. Then $\boldsymbol{X} \in \mathcal{A}$ if and only if*

$$S_g{}^{(1)} \overset{d}{=} S_g{}^{(2)} \overset{d}{=} \cdots \overset{d}{=} S_g{}^{(n)}, \tag{2.13}$$

where $S_g{}^{(i)} := g\big(\boldsymbol{X}^{(i)}\big)$, $i \in \mathcal{N}$. Further, we have $S_g{}^{(1)} \overset{d}{=} S_g{}^*$, where $S_g{}^*$ is the size-biased counterpart of S_g. What is more, it must hold that $\beta_i = \mathbb{E}[X_i] \,/\, \mathbb{E}[S_g]$ for all $i \in \mathcal{N}$.

Proof. See appendix v □

Remark 6. *When the collection function g is additive, i.e., $g(\boldsymbol{x}) = x_1 + \ldots + x_n$, then the results in Proposition 2 are analogous to the ones in Theorem 1 of Mohammed et al. (2021).*

Remark 7. *It is evident that if $\boldsymbol{X}$ is exchangeable and the collection function is symmetric,*

then condition (2.13) *is satisfied. Moreover, under this case,* $A_w(X_i, S_g) = 1/n$ *for all* $i \in \mathcal{N}$*. Thereby,* (2.13) *implies some symmetric structure inherent in both the distribution of* $\boldsymbol{X}$ *and the choice of the collection function. When the notion of multivariate size-biased transform as per Definition 1 is interpreted in terms of loading for model/sample risk, then* (2.13) *signifies that the choice of loading direction would not impact the resulting distributing of the collection risk.*

Proposition 2 also tells that if a linear relationship between the allocated capital and total capital required holds for any choice of w, then the intercept α_i must be zero, and thus $A_w(X_i, S_g)$ is proportional to $H_w(S_g)$. One may wonder under what conditions, the other (extremal) relation in which $\alpha_i > 0$ and $\beta_i = 0$, may hold. Namely, we are now interested in studying the set $\mathcal{B}$.

Proposition 3. *Consider a risk portfolio* $\boldsymbol{X} \in \mathcal{X}^n$*. Suppose that none of the coordinates of* $\boldsymbol{X}$ *are degenerate (i.e.,* $\mathbb{P}(X_i = c) < 1$*,* $i \in \mathcal{N}$*). Then it holds that* $\boldsymbol{X} \in \mathcal{B}$ *if and only if*

$$\mathbb{P}(S_g = c) = 1 \quad \textit{with } S_g = g(\boldsymbol{X}) \textit{ and } c > 0 \textit{ is some constant.} \tag{2.14}$$

Proof. See appendix vi. □

While proposition 2 shows that proportional triviality allows for some flexibility, as the characterization includes distributions of g beyond degeneracy, the case of absolute triviality, in proposition 3, is confined to the constant aggregate i.e. degeneracy. For both cases, nonetheless, the underlying central condition is the regression function $\mathbb{E}[X_i|S_g]$ being either linear or constant in $S_g = g(\boldsymbol{X})$. The regression condition is intuitive in nature and it draws similarities from other works such as Guan et al. (2021) where an axiomatic formulation, in particular the axiom of shrinking independence, was used to reach absolute triviality.

Going back to proposition 3, there are two situations in which (2.14) can hold. The first situation is that the collection function is degenerate, i.e., $g(\boldsymbol{x}) = c$ for any $\boldsymbol{x} \in \mathbb{R}^n_+$. Such collection functions are not informative in the context of capital calculation and allocations. What is more interesting to consider pertains to the second situation in which (2.14) implies

some distributional properties of $\boldsymbol{X}$. The following definition which plays an important role in studying extremal negative dependence, is needed to facilitate the succeeding discussion.

Definition 2. *A set $\mathcal{W} = (y_1, \ldots, y_n) \in \mathbb{R}^n_+$ is said to be non-increasing if for any two points $\boldsymbol{y}_1 = (y_{11}, \ldots, y_{1n}) \in \mathcal{W}$ and $\boldsymbol{y}_2 = (y_{21}, \ldots, y_{2n}) \in \mathcal{W}$, if $y_{1i} < y_{2i}$ for some $i \in \mathcal{N}$ then there exists $j \in \mathcal{N}$, $j \neq i$, such that $y_{1j} \geq y_{2j}$.*

Proposition 4. *If the collection function g is strictly increasing coordinate-wise, then the set*

$$\mathcal{W} = \left\{ \boldsymbol{x} \in \mathbb{R}^n_+ : g(\boldsymbol{x}) = c, \quad \text{for some constant } c > 0 \right\}$$

is non-increasing.

Proof. The proof is proceeded by contradiction. Assume that $\mathcal{W}$ is not non-increasing. Then there exists two distinct points $\boldsymbol{x}_1 = (x_{11}, \ldots, x_{1n}) \in \mathcal{W}$ and $\boldsymbol{x}_2 = (x_{21}, \ldots, x_{2n}) \in \mathcal{W}$ such that $x_{1i} \leq x_{2i}$ for all $i \in \mathcal{N}$. Since g is strictly increasing coordinate-wise, we must have at least one of $g(\boldsymbol{x}_1)$ and $g(\boldsymbol{x}_2)$ not equal to c, which contracts the fact that $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{W}$. This completes the proof. □

Definition 3. *(Bignozzi and Puccetti, 2015) We say that $\boldsymbol{X} \in \mathcal{X}^n$ admits a g-mixability structure if $\mathbb{P}(\boldsymbol{X} \in \mathcal{W}) = 1$.*

Note that the dependence structure considered in definition 3 generalizes the negative dependence concepts widely studied in the probability literature (e.g., Lee and Ahn, 2014; Puccetti et al., 2012, 2013; Wang and Wang, 2011). In Bignozzi and Puccetti (2015), g was thoroughly studied under the min, max and product operators. Additionally, when $g(\boldsymbol{x}) = \sum_{i=1}^n h_i(X_i)$, for increasing functions h_i, then we get d-CTM defined in Lee and Ahn (2014). If all h_i are chosen to be the identity function i.e. we have the canonical collection function $g(\boldsymbol{x}) = x_1 + \cdots + x_n$, then the g-mixability in definition 3 reduces to the joint mixability structure. Consequently, the result of Proposition 3 yields that a constant weighted allocation irrespective of the choice of weight function w is equivalent to a risk portfolio $\boldsymbol{X}$ that admits an extremal negative dependence structure. This is precisely the law governing the random vector $\boldsymbol{X}$ whose support is a non-increasing set.

A Proofs

i Proof of Theorem 1

Proof. We only prove the first case in which $h(y) \geq y$ and the function $y \mapsto y/\tilde{h}(y)$ is increasing. The other case holds based on the same argument. Let us write

$$H_1 = \frac{\mathbb{E}\big[S_{g_1} \times w(S_{g_1})\big]}{\mathbb{E}\big[w(S_{g_1})\big]} = \frac{\mathbb{E}_{\tilde{h}\circ w(S_{g_1})}\big[S_{g_1} \times w(S_{g_1}) \times \big(\tilde{h} \circ w(S_{g_1})\big)^{-1}\big]}{\mathbb{E}_{\tilde{h}\circ w(S_{g_1})}\big[w(S_{g_1}) \times \big(\tilde{h} \circ w(S_{g_1})\big)^{-1}\big]}.$$

Since $y \mapsto y/\tilde{h}(y)$ is increasing, then using Chebyshev's sum inequality we have

$$\mathbb{E}_{\tilde{h}\circ w(S_{g_1})}\big[S_{g_1} \times w(S_{g_1}) \times \big(\tilde{h} \circ w(S_{g_1})\big)^{-1}\big] \geq \mathbb{E}_{\tilde{h}\circ w(S_{g_1})}[S_{g_1}] \times \mathbb{E}_{\tilde{h}\circ w(S_{g_1})}\big[w(S_{g_1}) \times \big(\tilde{h} \circ w(S_{g_1})\big)^{-1}\big].$$

Thereby, it holds that

$$\begin{aligned} H_1 \geq \mathbb{E}_{\tilde{h}\circ w(S_{g_1})}[S_{g_1}] = \frac{\mathbb{E}\big[S_{g_1} \times w(S_{g_2})\big]}{\mathbb{E}\big[w(S_{g_2})\big]} &= \frac{\mathbb{E}\big[h(S_{g_2}) \times w(S_{g_2})\big]}{\mathbb{E}\big[w(S_{g_2})\big]} \\ &\geq \frac{\mathbb{E}\big[S_{g_2} \times w(S_{g_2})\big]}{\mathbb{E}\big[w(S_{g_2})\big]} = H_2, \end{aligned}$$

where the second inequality holds because of the condition $h(y) \geq y$, $y \in \mathbb{R}_+$. The proof is now completed. □

ii Proof of Theorem 2

Proof. We only prove the first case in which $\xi(y) \leq y$ with $\xi'(y) \leq 1$ for all $y \in \mathbb{R}_+$. A repeated application of the same argument would yield the desired conclusion for the second case.

Recall that $S_{g_j} = g_j(\boldsymbol{X})$ for $j = 1, 2$. Since $\xi \circ g_1(\boldsymbol{y}) = g_2(\boldsymbol{y})$, we can write

$$\frac{w(y)}{\mathbb{E}\big[w(S_{g_2}) \,|\, S_{g_1} = y\big]} = \frac{w(y)}{w \circ \xi(y)},$$

which has the same monotonicity behavior as $\log\big(w(y)\big) - \log\big(w \circ \xi(y)\big)$. Consider

$$\frac{\mathrm{d}}{\mathrm{d}y}\Big[\log\big(w(y)\big) - \log\big(w \circ \xi(y)\big)\Big] = \frac{\mathrm{d}}{\mathrm{d}t}\log\big(w(t)\big)\big|_{t=y} - \frac{\mathrm{d}}{\mathrm{d}t}\log\big(w(t)\big)\big|_{t=\xi(y)}\,\xi'(y).$$

By assumption, we have $\xi'(y) \in (0, 1]$ for $y > 0$. Moreover, since $\xi(y) \le y$ and $t \mapsto w(t)$ is log convex, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\log\big(w(t)\big)\big|_{t=y} \ge \frac{\mathrm{d}}{\mathrm{d}t}\log\big(w(t)\big)\big|_{t=\xi(y)}.$$

Together with the assumption that $y \mapsto w(y)$ is increasing, we conclude

$$\frac{\mathrm{d}}{\mathrm{d}y}\Big[\log\big(w(y)\big) - \log\big(w \circ \xi(y)\big)\Big] \ge 0,$$

thus the function $y \mapsto w(y)/\mathbb{E}[w(S_{g_2}) \,|\, S_{g_1} = y]$ is increasing.

Further, note that $\xi(y) \le y$ implies $h(y) \ge y$ for all $y \in \mathbb{R}_+$ based on Remark 1. According to Theorem 1 and Remark 2, we obtain $H_1 \ge H_2$. The proof is now completed. □

iii Proof of Theorem 4

Proof. Let $S_g = g_1(\boldsymbol{X})$, then the following string of equations holds for $x \ge 0$

$$\begin{aligned}
\frac{\ell_{i,1}(x)}{\ell_{i,2}(x)} &= \frac{\mathbb{E}\big[w(S_g)|X_i = x\big]}{\mathbb{E}\big[w \circ \xi(S_g)|X_i = x\big]} \\
&= \frac{\mathbb{E}\big[w(S_g)|U = F_{X_i}(x)\big]}{\mathbb{E}\big[w \circ \xi(S_g)|U = F_{X_i}(x)\big]} \\
&= \frac{w \circ \tilde{g}(x)}{w \circ \xi \circ \tilde{g}(x)},
\end{aligned} \tag{15}$$

where $\tilde{g}(x) = g_1(\boldsymbol{x})$ with $\boldsymbol{x} = \big(F^{-1}_{X_j}(F_{X_i}(x))\big)_{j \in \mathcal{N}}$.

Since the aggregation function is element-wise increasing, the monotonicity behaviour of the ratio in (15) is same as that of $y \mapsto w(y)/w \circ \xi(y)$. Evoking the argument used in the proof of Theorem 2, we conclude that $x \mapsto \ell_{i,1}(x)/\ell_{i,2}(x)$ is increasing if $\xi(y) \leq y$ and $\xi'(y) \leq 1$, and the function is decreasing if $\xi(y) > y$ and $\xi'(y) > 1$. Applying Theorem 3 yields the desired result and thus completes the proof. □

iv Proof of Theorem 7

Proof. Fix an aggregation function g. Then let us write

$$\begin{aligned}\frac{\tilde{\ell}_{i,1}(x)}{\tilde{\ell}_{i,2}(x)} &= \frac{\mathbb{E}\big[w_1(S_g)|X_i = x\big]}{\mathbb{E}\big[w_2(S_g)|X_i = x\big]} \\ &= \frac{\mathbb{E}\big[w_1 \circ g(\boldsymbol{X})|U = F_{X_i}(x)\big]}{\mathbb{E}\big[w_2 \circ g(\boldsymbol{X})|U = F_{X_i}(x)\big]} \\ &= \frac{w_1 \circ \tilde{g}(x)}{w_2 \circ \tilde{g}(x)}, \end{aligned} \tag{16}$$

where $\tilde{g}(x) = g(\boldsymbol{x})$ with $\boldsymbol{x} = \big(F^{-1}_{X_k}(F_{X_i}(x))\big)_{k\in\mathcal{N}}$.

Since the aggregation function g is component-wise increasing, the function $x \mapsto \tilde{g}(x)$ is increasing. We can conclude that the function $x \mapsto \tilde{\ell}_{i,1}(x)/\tilde{\ell}_{i,2}(x)$ has the same monotonicity behavior as the function $y \mapsto w_1(y)/w_2(y)$. An application of Theorem 6 yields the desired result, which completes the proof. □

v Proof of Proposition 2

Proof. Let us begin with the sufficiency of the statement. First note that conditions (2.11) and (2.12) together imply $\mathbb{E}[X_i|S_g = y] \downarrow 0$ when $y \downarrow 0$, $i \in \mathcal{N}$. Then evoking Lemma 1 yields that if $\boldsymbol{X} \in \mathcal{A}$, then α_i in (2.9) must be zero, and thus

$$\mathbb{E}[X_i|S_g] = \beta_i\, S_g, \quad i \in \mathcal{N}.$$

Taking expectation on both sides of the above equation also implies $\beta_i = \mathbb{E}[S_g] \,/\, \mathbb{E}[X_i]$. Collectively, the Laplace transform of ${S_g}^{(i)}$ can be computed via, for $t > 0$,

$$\begin{aligned}\frac{\mathbb{E}\left[X_i\, e^{-tS_g}\right]}{\mathbb{E}[X_i]} &= \frac{\mathbb{E}\left[\mathbb{E}[X_i|S_g]\, e^{-tS_g}\right]}{\mathbb{E}[X_i]} \\ &= \frac{\beta_i}{\mathbb{E}[X_i]}\,\mathbb{E}\left[S_g\, e^{-tS_g}\right] \\ &= \frac{\mathbb{E}\left[S_g e^{-tS_g}\right]}{\mathbb{E}[S_g]}.\end{aligned}$$

This readily implies the desirable relationship ${S_g}^{(1)} \overset{d}{=} {S_g}^{(2)} \overset{d}{=} \cdots \overset{d}{=} {S_g}^{(n)} \overset{d}{=} {S_g}^{*}$.

To prove the necessity direction, note that for any $t > 0$,

$$\frac{\mathbb{E}\left[\mathbb{E}[X_i|S_g]\, e^{-tS_g}\right]}{\mathbb{E}[X_i]} - \frac{\mathbb{E}\left[S_g e^{-tS_g}\right]}{\mathbb{E}[S_g]} = 0$$

is equivalent to

$$\mathbb{E}\left[\left(\mathbb{E}[X_i|S_g] - \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_g]} S_g\right) e^{-tS_g}\right] = 0, \tag{17}$$

which in turn gives

$$\mathbb{E}[X_i|S_g] \equiv \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_g]}\, S_g.$$

So $\boldsymbol{X} \in \mathcal{A}$, which finishes the proof. □

vi Proof of Proposition 3

Proof. Base on Lemma 1 with $\beta_i = 0$, $i \in \mathcal{N}$, we can obtain that $\boldsymbol{X} \in \mathcal{B}$ is equivalent to $\mathbb{E}[X_i|S_g] \equiv \alpha_i$. Moreover, $\alpha_i = \mathbb{E}[X_i]$.

Now, we are in the position to prove the sufficiency direction. Consider the Laplace transform of ${S_g}^{(i)}$:

$$\frac{\mathbb{E}\left[X_i\, e^{-tS_g}\right]}{\mathbb{E}[X_i]} = \frac{\mathbb{E}\left[\mathbb{E}[X_i|S_g]\, e^{-tS_g}\right]}{\mathbb{E}[X_i]}$$

$$= \frac{\alpha_i}{\mathbb{E}[X_i]} \mathbb{E}\left[e^{-tS_g}\right]$$

$$= \mathbb{E}\left[e^{-tS_g}\right], \quad t > 0. \tag{18}$$

Since X_i's are non-degenerate, (18) implies $\mathbb{P}(S_g = c) = 1$ for some constant $c > 0$. The implication holds since otherwise ${S_g}^{(i)} \stackrel{d}{=} S_g$ which leads to a contradiction.

To prove the necessity direction, note that for any $t > 0$,

$$\frac{\mathbb{E}\left[\mathbb{E}[X_i|S_g]\, e^{-tS_g}\right]}{\mathbb{E}[X_i]} - \mathbb{E}\left[e^{-tS_g}\right] = 0,$$

which is equivalent to

$$\mathbb{E}\Big[\big(\mathbb{E}[X_i|S_g] - \mathbb{E}[X_i]\big)\, e^{-tS_g}\Big] = 0.$$

This implies

$$\mathbb{E}[X_i|S_g] \equiv \mathbb{E}(X_i),$$

or equivalently, $\boldsymbol{X} \in \mathcal{B}$, which finishes the proof. □

Chapter 3

Can a regulatory risk measure induce profit-maximizing risk capital allocations? The case of Conditional Tail Expectation

3.1 Introduction

Consider positive random variables (RVs) $X_1, \ldots, X_n,\ n \in \mathbb{N}$, which represent losses due to distinct business units (BUs) of an insurer, and denote by the sets $\mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{X}$ the collections of these BUs and losses, respectively. Then, for the aggregate loss RV $S_X := X_1 + \cdots + X_n$, the map $A : \mathcal{X} \times \mathcal{X} \to [0,\ \infty)$, which assigns non-negative values to random pairs $(X, S) \in \mathcal{X} \times \mathcal{X}$, is called a risk capital (RC) allocation rule (e.g., Denault, 2001; Dhaene et al., 2012; Furman and Zitikis, 2008b). Additionally, if $A(X, X) = H(X)$, where the map $H : \mathcal{X} \to [0,\ \infty)$ is called a risk measure and assigns non-negative values to the random loss $X \in \mathcal{X}$, then the allocation rule A is said to be induced by the risk measure H.

RC allocation rules have gained major importance in risk management and insurance applications in the context of price determination, profitability assessment, budgeting decision

making, to name just a few (Guo et al., 2018; Venter, 2004). Similarly, the academic significance of - also, interest in - the subject of RC allocations have been strong, as evidenced by the large and growing body of scholarly literature (e.g., Boonen et al., 2019; Furman et al., 2018a, 2020a; Kim and Kim, 2019; Shushi and Yao, 2020, for recent references in the *Insurance: Mathematics and Economics* journal, alone).

Not surprisingly, therefore, numerous RC allocation rules have been proposed and studied, with the RC allocation rule induced by the conditional tail expectation (CTE) risk measure being arguably the most popular (Kalkbrener, 2005). More specifically, for $q \in [0,\ 1)$, $s_q := \mathrm{VaR}_q(S_X) = \inf\left\{s \in [0,\ \infty) :\ \mathbb{P}(S_X \leq s) \geq q\right\}$, loss portfolio $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ and BU $i \in \mathcal{N}$, the CTE-based RC allocation rule, when well-defined and finite, is given by

$$\mathrm{CTE}_q(X_i, S_X) = \mathbb{E}\left[X_i | \, S_X > s_q\right]. \tag{3.1}$$

As it is common in real applications to use RC allocation (3.1) - also, other RC allocation rules - to attribute the exogenous aggregate risk capital, say $\kappa \in \mathbb{R}_+$, to distinct BUs in the set $\mathcal{N}$, and, in order to guarantee the total additivity of the RC allocation rule, it is beneficial to explore (e.g., Dhaene et al., 2012) the quantity, $\kappa_i = \kappa \times r_{q,i}$, where

$$r_{q,i} = \frac{\mathrm{CTE}_q(X_i, S_X)}{\mathrm{CTE}_q(S_X)}, \quad i \in \mathcal{N} \tag{3.2}$$

is the associated *proportional* RC allocation rule induced by the CTE risk measure $\mathrm{CTE}_q(X) = \mathrm{CTE}_q(X, X)$ for any $X \in \mathcal{X}$ and $q \in [0, 1)$.

The RC allocation rule based on the CTE risk measure has been thoroughly studied on a variety of fronts. Namely, allocation rule (3.1) was obtained as the gradient and the Aumann-Shapley allocation induced by the CTE risk measure by Tasche (2004) and Denault (2001), respectively. Also, for loss RVs with continuous cumulative distribution functions (CDFs), allocation rule (3.1) coincides with the RC allocation rule induced by the Expected Shortfall risk measure (Kalkbrener, 2005; Wang and Zitikis, 2021). Last but not least, allocation rule (3.1) belongs to the class of distorted (Tsanakas and Barnett, 2003) and weighted (Furman and Zitikis, 2008b) RC allocation rules and is optimal in the sense of Laeven and Goovaerts (2004) and Dhaene et al. (2012).

The number of works that evaluate allocation rule (3.1) for random losses having distinct joint CDFs is really overwhelming. For just a few examples, we refer to: Panjer (2002) and Landsman and Valdez (2003) for, respectively, normal and elliptical distributions; Cai and Li (2010) for phase-type distributions; Furman and Landsman (2005) for gamma distributions; Vernic (2006) and Hendriks and Landsman (2017); Vernic (2011) for, respectively, skew normal and Pareto distributions; Furman and Landsman (2010) for Tweedie distributions; Cossette et al. (2012) for compound distributions with positive severities; Cossette et al. (2013); Ratovomirija et al. (2017) for mixed Erlang distributions; Furman et al. (2018a) for Generalized Gamma Convolutions; this list is by no means exhaustive.

Despite the abundant relevant academic literature, RC allocation rule (3.1) - as well as the CTE risk measure that induces it - have been employed in practice mainly due to the inclusion in existing regulatory accords. When this aspect is put aside, the quantity $r_{q,i}$ raises a number of concerns. First, it hinges on a contentious two-step procedure, as the numerator and the denominator in Equation (3.2) must be computed separately for each BU $i \in \mathcal{N}$; similar concerns have been brought in Chong et al. (2020) in the context of the quantity κ_i. Second, it neglects the risk perception of the insurer and the economic environment in which they operate (Bauer and Zanjani, 2016).

Admittedly, it is not surprising in any way that regulations, which are driven by the notion of prudence, and insurers' targets, which are profit-oriented, diverge. Nevertheless, it is instrumental to determine whether or not there exist model settings under which the RC allocation rule induced by the CTE risk measure yields outcomes that address the just-mentioned two concerns. This is what we do in the present paper.

We have organized the rest of this papers as follows. In Section 3.2, we motivate in detail and formulate the problem of interest. We then solve this problem in Sections 3.3 and 3.4, which provide ample elucidating examples. Some of our analysis and conclusions carry over to a family of risk measures that contains the CTE risk measure as a special case, which is demonstrated in Section 3.5. Chapter 5 includes conclusions of the paper. In the sequel, we routinely work with an atomless and rich probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we let L^α and L^∞ denote, respectively, the set of all RVs that have finite α-th moment, $\alpha \in [0, \infty)$, and the set of all essentially bounded RVs on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Unless specified

otherwise, we work with the collection of integrable RVs, L^1, so that the CTE risk measure and its associated RC allocation are well-defined and finite. Finally, we denote by F_X and ϕ_X the CDF and the Laplace transform of the RV X, and we use $\mathbb{1}$ to denote the indicator function.

3.2 Compositional allocation rules induced by the Conditional Tail Expectation risk measure and a question that arises

Note that the CTE-based allocation exercise (3.2) can be framed within the context of the standard n-dimensional simplex space (Aitchison, 1986):

$$\mathfrak{S}^n = \big\{(r_1, \ldots, r_n) : \, r_i \in [0, \, 1], \, \, i = 1, \ldots, n \text{ and } r_1 + \cdots + r_n = 1\big\}.$$

Specifically, for $x_1, \ldots, x_n \in [0, \, \infty)$ and $s := x_1 + \cdots + x_n$, as well as for the special map $\mathcal{C} : [0, \, \infty)^n \to \mathfrak{S}^n$ with $\mathcal{C}_i(x_1, \ldots, x_n) = x_i/s, \, \, i = 1, \ldots, n$, proportional allocation rule (3.2) is obtained via setting $x_i = \mathrm{CTE}_q(X_i, S_X)$, and so (Belles-Sampera et al., 2016; Boonen et al., 2019)

$$r_{q,i} = \mathcal{C}_i\big(\mathrm{CTE}_q(X_1, S_X), \ldots, \mathrm{CTE}_q(X_n, S_X)\big) = \mathrm{CTE}_q(X_i, S_X) \, / \, \mathrm{CTE}_q(S_X), \, \, q \in [0, \, 1).$$

We note in passing that a similar reformulation of the RC allocation exercise in the context of the n-dimensional simplex space can be achieved effortlessly for the whole class of weighted RC allocation rules (Furman and Zitikis, 2008b), which are induced by the class of weighted risk measures (Furman and Zitikis, 2008a) and of which allocation rule (3.1) is a special case (Furman et al., 2020b). Additionally, the weighted RC allocations are those recovered when the aggregation function of the generalized functionals is set to be the canonical sum, $g(\boldsymbol{X}) = \sum_{i=1}^n X_i$, in Chapter 2.

An alternative way to determine the proportional contribution of the i-th BU of an insurer

to the aggregate risk capital - under the assumption that it is the CTE risk measure that induces the desired allocation rule - is by considering the ratio RV $R_i = X_i \,/\, S_X,\ i = 1, \ldots, n$, directly. That is, while, for a fixed $q \in [0,\ 1)$, the proportional allocation $r_{q,i}$, confined with the help of the normalizing constant $\mathrm{CTE}_q(S_X) \in \mathbb{R}_+$ to the unit interval, $\mathcal{I} = [0,\ 1]$, operates on random pairs $(X_i, S_X) \in \mathcal{X} \times \mathcal{X}$, an alternative to $r_{q,i}$ proportional allocation, call it $\tilde{r}_{q,i}$, is chosen to operate on random pairs $(R_i, S_X) \in \mathcal{X} \times \mathcal{X},\ i = 1, \ldots, n$, and so

$$\tilde{r}_{q,i} = \mathrm{CTE}_q\left(\mathcal{C}_i(X_1, \ldots, X_n), S_X\right) = \mathrm{CTE}_q\left(R_i, S_X\right),\ q \in [0,\ 1).$$

While various properties of the proportional allocation $r_{q,i}$ have been well-studied, this is not so for its counterpart, $\tilde{r}_{q,i}$. Further, we report a number of important properties of the latter quantity. In this respect, our first proposition shows that the quantity $\tilde{r}_{q,i}$ agrees with the economic capital allocation rule proposed recently by Bauer and Zanjani (2016). Namely, while the motivation for the RC allocation rule $r_{q,i}$ is the central role that the CTE risk measure plays in today's (insurance) regulation, the proportional allocation $\tilde{r}_{q,i}$ turns out to be a well-justified choice for a profit maximizing insurer with risk-averse counterparties in an incomplete market setting with frictional capital costs.

Consider the aggregate loss RV $S_X \in \mathcal{X}$ and the Geometric Tail Expectation (GTE) risk measure:

$$\mathrm{GTE}_q(S_X) := \exp\left\{\mathbb{E}\left[\log(S_X) | S_X > s_q\right]\right\},\ q \in [0,\ 1). \tag{3.3}$$

The connection of risk measure (3.3) to the notion of geometric means (e.g., Hardy et al., 1952) motivates its name; also, risk measure (3.3) is a *tail quasi-linear mean* risk measure in the sense of Bäuerle and Shushi (2020). It is not difficult to see that, for any $q \in [0,\ 1)$, risk measure (3.3) is at least as prudent as the Value-at-Risk risk measure and may be finite even if the CTE risk measure is infinite. Namely, we have the following simple result.

Proposition 5. *For any $X \in \mathcal{X}$ and $q \in [0,\ 1)$, we have the bounds*

$$\mathrm{VaR}_q(S_X) \leq \mathrm{GTE}_q(S_X) \leq \mathrm{CTE}_q(S_X).$$

Proof. By Jensen's inequality, we have, for $q \in [0,\, 1)$,

$$\exp\left\{\mathbb{E}\left[\log(S_X)|S_X > s_q\right]\right\} \leq \mathbb{E}\left[\exp\{\log(S_X)\}|S_X > s_q\right] = \mathbb{E}\left[S_X|S_X > s_q\right],$$

which proves the upper bound. In addition, for $q \in [0,\, 1)$,

$$s_q = \exp\left\{\mathbb{E}\left[\log(s_q)|S_X > s_q\right]\right\} \leq \exp\left\{\mathbb{E}\left[\log(S_X)|S_X > s_q\right]\right\},$$

establishing the lower bound and, hence, proving the proposition. □

Another immediate but worth-mentioning observation is that risk measure (3.3) is neither coherent in the sense of Artzner et al. (1999) nor convex in the sense of Föllmer and Schied (2016), as it violates translation-invariance. Consequently, risk measure (3.3) is not a monetary risk measure. Nevertheless, risk measure (3.3) belongs to the class of return risk measures (Bellini et al., 2018). Moreover, when viewed through the prism of a profit maximizing insurer, risk measure (3.3) induces the optimal RC allocation outcome, $\tilde{r}_{q,i}$, in the sense of Bauer and Zanjani (2016); intuitively, this might be due to the decreasing marginal effect of the increase in aggregate loss (Bauer and Zanjani, 2016). Our next statement about the RC allocation induced by risk measure (3.3) is formulated as a proposition.

Before stating our next result, we note that, similarly to how the CTE-based risk capital allocation is induced by the CTE risk measure, the GTE-based risk capital allocation,

$$\mathrm{GTE}_q(X_i, S_X) := \mathrm{GTE}_q(S_X) \times \tilde{r}_{q,i}, \quad q \in [0,\ 1),\ i \in \mathcal{N},$$

is induced by risk measure (3.3). The GTE-based RC is fully-additive, in the sense that the sum of the allocations is equal to the inducing risk measure as can be seen in the following equations:

$$\sum_{i=1}^{n} \mathrm{GTE}_q(X_i, S) = \sum_{i=1}^{n} \mathrm{GTE}_q\big(S_X\big) \times \tilde{r}_{q,i} = \mathrm{GTE}_q\big(S_X\big) \times \sum_{i=1}^{n} \tilde{r}_{q,i} = \mathrm{GTE}_q\big(S_X\big),$$

for $q \in [0,\, 1)$.

Proposition 6. *The GTE-based RC allocation is the gradient allocation in the direction of the loss RV $X_i \in \mathcal{X},\ i \in \mathcal{N}$ induced by risk measure (3.3).*

Proof. Since the GTE risk measure is positively homogeneous, by Euler's theorem (1.3) we have, for $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathcal{I}^n$ and $S_X(\boldsymbol{u}) := u_1 X_1 + \cdots + u_n X_n$,

$$\mathrm{GTE}_q\big(S_X(\boldsymbol{u})\big) = \sum_{i=1}^{n} u_i \frac{\partial}{\partial u_i} \mathrm{GTE}_q\big(S_X(\boldsymbol{u})\big), \quad \text{with } q \in [0,\ 1),\ i \in \mathcal{N}.$$

Therefore, for $\mathbf{1}$ denoting the n-variate vector of ones, we obtain

$$\begin{aligned}
\frac{\partial}{\partial u_i} \mathrm{GTE}_q\big(S_X(\boldsymbol{u})\big)\Big|_{\boldsymbol{u}=\mathbf{1}} &= \frac{\partial}{\partial u_i} \exp\Big\{ \mathbb{E}\left[\log(S_X(\boldsymbol{u})) | S_X(\boldsymbol{u}) > \mathrm{VaR}_q(S_X(\boldsymbol{u}))\right] \Big\}\Big|_{\boldsymbol{u}=\mathbf{1}}, \\
&= \left(\mathrm{GTE}_q\big(S_X(\boldsymbol{u})\big) \times \mathbb{E}\left[\frac{X_i}{S_X(\boldsymbol{u})} \middle|\ S_X(\boldsymbol{u}) > \mathrm{VaR}_q(S_X(\boldsymbol{u})) \right] \right) \Big|_{\boldsymbol{u}=\mathbf{1}} \\
&= \mathrm{GTE}_q\big(S_X\big) \times \tilde{r}_{q,i}, \quad i \in \mathcal{N}.
\end{aligned}$$

This completes the proof of the proposition. □

The next assertion demonstrates that the proportional allocation induced by the CTE risk measure is an approximation of the GTE allocation induced by risk measure (3.3).

Proposition 7. *The proportional allocation $r_{q,i}$ is a linear approximation of the proportional allocation $\tilde{r}_{q,i}$ for $i \in \mathcal{N}$.*

Proof. Consider the function $g(x_i, s) = x_i/s$ for $x_i, s \in \mathbb{R}_+$, $i = 1, \ldots, n$, and denote its partial derivatives by

$$g_i(x_i, s) = \frac{\partial}{\partial x_i} g(x_i, s) \quad \text{and} \quad g_s(x_i, s) = \frac{\partial}{\partial s} g(x_i, s).$$

Then the first-order Taylor expansion of g around $(x_0, s_0) = \big(\mathrm{CTE}_q(X_i\,,\, S_X),\, \mathrm{CTE}_q(S_X)\big)$, yields

$$x_i \,/\, s = g(x_0, s_0) + g_i(x_0, s_0)\,(x_i - x_0) + g_s(x_0, s_0)\,(s - s_0) + R_1(x_i, s),$$

where $R_1(x_i, s)$ is the reminder term for all $x_i, s \in \mathbb{R}_+,\ q \in [0,\ 1),\ i = 1, \ldots, n$. Consequently, we have

$$\tilde{r}_{q,i} \approx r_{q,i} + g_i(x_0, s_0)\,\mathbb{E}\big[(X_i - x_0)|\ S_X > s_q\big] + g_s(x_0, s_0)\,\mathbb{E}\big[(S_X - s_0)|\ S_X > s_q\big] = r_{q,i},$$

which establishes the desired approximation and thus completes the proof of the proposition. □

Finally, our last proposition - which can be considered a follow-up on Proposition 7 - delineates the difference between the allocations $r_{q,i}$ and $\tilde{r}_{q,i}$. The proof is an immediate consequence of the identity

$$\mathrm{Cov}(R_i, S_X|\ S_X > s_q) = \mathbb{E}[X_i|\ S_X > s_q] - \mathbb{E}[R_i|\ S_X > s_q] \times \mathbb{E}[S_X|\ S_X > s_q],$$

which holds for all $q \in [0,\ 1)$ and $(X_i, S_X) \in \mathcal{X} \times \mathcal{X},\ i = 1, \ldots, n$.

Proposition 8. *Given that all the quantities below are well-defined and finite, we have*

$$r_{q,i} = \tilde{r}_{q,i} + \frac{\mathrm{Cov}(R_i, S_X|\ S_X > s_q)}{\mathrm{CTE}_q(S_X)},\ q \in [0,\ 1),\ i \in \mathcal{N}.$$

Proposition 8 implies that it is the sign of the covariance between the RVs R_i and S_X, that determines the order of the allocations $r_{q,i}$ and $\tilde{r}_{q,i}$; note that, as $R_1 + \cdots + R_n = 1$ almost surely, we have $\sum_{i=1}^n \mathrm{Cov}(R_i, S_X|\ S_X > s_q) = 0$ (Furman and Zitikis, 2010, for examples, albeit in a different context, of the importance of covariances in insurance and finance). Also, and more importantly, Proposition 8 suggests that the proportional allocation rules induced by the CTE risk measure and those induced by risk measure (3.3) coincide if the aforementioned covariance is nil. This motivates the following question that engages us in the rest of this paper.

Question 1. *For $X_i \in \mathcal{X},\ i \in \mathcal{N}$, can we characterize those portfolios of losses $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$, for which the proportional allocations $r_{q,i}$ and $\tilde{r}_{q,i}$ agree for every $q \in [0,\ 1)$?*

At the outset, let us note that if the answer to the question above were in the affirmative, then this would imply that the regulatory (e.g., Swiss Solvency Test) CTE risk measure induces an optimal RC allocation rule for a profit maximizing insurer; the richer the class of joint CDFs of the RV $\boldsymbol{X} \in \mathcal{X}^n$ sought in Question 1, the more common the just-mentioned and apparently desirable agreement between the regulatory requirements and the risk perceptions of insurers.

Speaking formally, our goal is to characterize the following collection of loss RVs:

$$\mathfrak{W} = \big\{ \boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n \ : \ r_{q,i} = \widetilde{r}_{q,i} \ \text{ for all } q \in [0,1) \text{ and } i \in \mathcal{N} \big\}. \tag{3.4}$$

It is to be noted that the set $\mathfrak{W}$ can not be empty. To see a trivial case in which $\tilde{r}_{q,i} = r_{q,i}$ for every $q \in [0,\ 1)$ and $i \in \mathcal{N}$, let the loss RV $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ have identically distributed coordinates, $X_i \in \mathcal{X}_i$, and an exchangeable copula function, then we have $\mathbb{E}[R_i |\ S_X = s] = 1/n$ resulting in $\mathrm{Cov}(R_i, S_X |\, S_X = s) \equiv 0$, and therefore $\mathrm{Cov}(R_i, S_X |\, S_X > s_q) = 0$ for all $q \in [0,\ 1)$ and $i = 1, \ldots, n$. Consequently, the set of all loss RVs $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$, such that the proportional allocation rules $r_{q,i}$ and $\tilde{r}_{q,i}$ coincide has at least one portfolio of losses in it.

We devote the following sections of this paper to studying what other random loss RVs - besides the trivial example above - are members of the set $\mathfrak{W}$.

3.3 General considerations

In this section, we devise the necessary and sufficient conditions for the equality, $\tilde{r}_{q,i} = r_{q,i}$, for all $q \in [0,\ 1)$ and $i \in \mathcal{N}$. For this, we need a few auxiliary notions first. That is, Definitions 4 and 5 below introduce the univariate size-biased transform and its multivariate extension (Arratia et al., 2019; Furman et al., 2020a; Patil and Ord, 1976), both playing major roles in our analysis.

Definition 4. *Let $X \in L^\alpha$ be a positive loss RV, then the size-biased counterpart of order*

$\alpha \in \mathbb{R}_+$ of the loss RV X, call it $X^{[\alpha]}$, is defined via:

$$\mathbb{P}\left(X^{[\alpha]} \in dx\right) = \frac{x^\alpha}{\mathbb{E}[X^\alpha]}\mathbb{P}\left(X \in dx\right) \qquad \text{for all } x \in \mathbb{R}_+. \tag{3.5}$$

When $\alpha = 1$, we simply write X^* for the size-biased of order one variant of the RV $X \in L^1$. The RVs X and $X^{[\alpha]}$ are independent by construction for all $\alpha \in \mathbb{R}_+$.

Definition 5. *Let $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ be a loss RV with positive univariate coordinates $X_i \in L^{\alpha_i}$, $i = 1, \ldots, n$, then the multivariate size-biased counterpart of order $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n_+$ of the loss RV $\boldsymbol{X}$, call it $\boldsymbol{X}^{[\boldsymbol{\alpha}]}$, is defined via*

$$\mathbb{P}\left(\boldsymbol{X}^{[\boldsymbol{\alpha}]} \in d\boldsymbol{x}\right) = \frac{x_1^{\alpha_1} \times \cdots \times x_n^{\alpha_n}}{\mathbb{E}[X_1^{\alpha_1} \times \cdots \times X_n^{\alpha_n}]}\mathbb{P}\left(\boldsymbol{X} \in d\boldsymbol{x}\right) \qquad \text{for all } \boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_+. \tag{3.6}$$

The RVs $\boldsymbol{X}$ and $\boldsymbol{X}^{[\boldsymbol{\alpha}]}$ are independent by construction (e.g. Arratia et al., 2019; Furman et al., 2020a; Patil and Ord, 1976, for a similar discussion).

We next define the *partial size-biased transform*, which is a useful special case of the one presented in Definition 5 (e.g., Arratia et al., 2019; Furman et al., 2020a, for a few recent results in which the partial size-biased transform plays a central role but is not explicitly defined).

Definition 6. *Consider the size-biased RV $\boldsymbol{X}^{[\boldsymbol{\alpha}]}$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ as per Definition 5. Then, in the special case where the i-th coordinate of the vector $\boldsymbol{\alpha}$ is equal to $\alpha_i = \alpha \in \mathbb{R}_+$, $i = 1, \ldots, n$, whereas all other coordinates of the vector $\boldsymbol{\alpha}$ are equal to zero, we call the implied transform, the i-th partial size-biased transform of order α, and denote the corresponding RV by $\boldsymbol{X}^{[(\alpha)_i]}$. Namely, we have*

$$\mathbb{P}\left(\boldsymbol{X}^{[(\alpha)_i]} \in d\boldsymbol{x}\right) = \frac{x_i^\alpha}{\mathbb{E}[X_i^\alpha]}\mathbb{P}\left(\boldsymbol{X} \in d\boldsymbol{x}\right) \qquad \text{for all } \boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_+. \tag{3.7}$$

The RVs $\boldsymbol{X}$ and $\boldsymbol{X}^{[(\alpha)_i]}$ are independent by construction (e.g. Arratia et al., 2019; Furman et al., 2020a; Patil and Ord, 1976, for a similar discussion). For the case $\alpha = 1$ and $X_i \in L^1$, we simply write, $\boldsymbol{X}^{(*)_i}$, for the partial size-biased counterpart of the RV $\boldsymbol{X}$. The notation $\boldsymbol{X}^{(*)_i}$ is equivalent to $\boldsymbol{X}^{(i)}$ from Definition 1 of Section 2.4.

The operation of size-biasing has an important interpretation in the context of actuarial science and, more generally, in quantitative risk management, where it is considered *loading* for model/sample risk. Indeed, it is not difficult to see that the size-biased loss RVs $X^{[\alpha]}$ and $\boldsymbol{X}^{[\boldsymbol{\alpha}]}$ (also, $\boldsymbol{X}^{(\alpha)_i}$) dominate stochastically the loss RVs X and $\boldsymbol{X}$, respectively.

Furthermore, the partial size-biased RV, $\boldsymbol{X}^{[(\alpha)_i]}$, plays an important role for *size-biasing* sums of RVs. Namely, let $S_X^* = (X_1 + \cdots + X_n)^*$, then the distribution of the RV S_X^* admits a finite-mixture representation (e.g., Arratia et al., 2019) in terms of the partial size-biased RVs. Indeed, let $\phi_{S_X^*}$ denote the Laplace transform of the RV S_X^*, then, for $p_i = \mathbb{E}[X_i] \,/\, \mathbb{E}[S_X]$, we have

$$\phi_{S_X^*}(t) = \sum_{i=1}^{n} p_i \times \phi_{S_X^{(*)_i}}(t), \quad \mathrm{Re}(t) > 0, \tag{3.8}$$

where $S_X^{(*)_i}$ is the sum of the coordinates of the partially size-biased RV $\boldsymbol{X}^{(*)_i}$, $i = 1, \ldots, n$.

The following lemma is a variation of Equation (3.8) that we find useful in this paper.

Lemma 2. *Consider the RV* $\boldsymbol{X}_+ = (X_1, \ldots, X_n, Y_{n+1}, \ldots, Y_{n+m}) \in \mathcal{X}^{n+m}$, $n, m \in \mathbb{N}$, *and let* $S_X = \sum_{i=1}^n X_i$, $S_Y = \sum_{i=n+1}^{n+m} Y_i$, $S_+ = S_X + S_Y$, *and* $S_+^{(*)X} = S_X^* + S_Y$. *Then the distribution of the RV* $S_+^{(*)X}$ *admits a mixture representation in the sense that we have* $S_+^{(*)X} =_d S_+^{(*)K}$, *where the RV* $K \in \{1, \ldots, n\}$ *is such that* $\mathbb{P}(K = k) = \mathbb{E}[X_i] \,/\, \mathbb{E}[S_X]$, $k = 1, \ldots, n$

Proof. Let $\phi_{S_+^{(*)X}}$ denote the Laplace transform of the RV $S_+^{(*)X}$, then, with the help of Equation (3.7), we have, for $p_i = \mathbb{E}[X_i] \,/\, \mathbb{E}[S_X]$,

$$\phi_{S_+^{(*)X}}(t) = \frac{\mathbb{E}\left[S_X\, e^{-tS_+}\right]}{\mathbb{E}[S_X]} = \sum_{i=1}^{n} p_i \frac{\mathbb{E}\left[X_i\, e^{-tS_+}\right]}{\mathbb{E}[X_i]} = \sum_{i=1}^{n} p_i \times \phi_{S_+^{(*)_i}}(t), \quad \mathrm{Re}(t) > 0,$$

which establishes the desired result and thus completes the proof. □

The next assertion spells out the sufficient and necessary conditions for the loss portfolios $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ to belong to the set $\mathfrak{W}$, and hence it answers Question 1. The non-technical interpretation of the assertion is that for loss portfolios in the set $\mathfrak{W}$ and under the paradigm of loading for model/sample risk, the choice of the *load direction* as per Definition 6 does not impact the distribution of the loaded aggregate loss RV.

Theorem 8. *Consider the loss RV* $\boldsymbol{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$ *and assume that* $X_i \in L^1$*, then we have the equality* $r_{q,i} = \tilde{r}_{q,i}$ *for all* $q \in [0,\ 1)$*,* $i \in \mathcal{N}$*, which implies equality to* $\mathbb{E}[X_i]/\mathbb{E}[S_X]$*, if and only if* $S_X^{(*)_i} =_d S_X^{(*)_j} (=_d S_X^*)$*,* $i \neq j \in \mathcal{N}$*.*

Proof. Assume that $r_{q,i} = \tilde{r}_{q,i}$ for all $q \in [0,\ 1)$ and $i = 1, \dots, n$. By Proposition 8, this is equivalent to requesting that $\mathrm{Cov}(R_i, S_X |\, S_X > u) = 0$ for all $u \geq 0$, or, in other words with the notation $G_i(s) = \mathbb{E}[R_i |\ S = s] - \mathbb{E}[R_i],\ i = 1, \dots, n$, that

$$\mathbb{E}\big[S_X\, G_i(S_X) | S_X > u\big] = \mathbb{E}\big[G_i(S_X) | S_X > u\big]\, \mathbb{E}[S_X | S_X > u] \quad \text{for all } u \geq 0,$$

Assuming the law of S is absolutely continuous then differentiating both sides w.r.t. u we get:

$$\begin{aligned}
\big(\mathbb{E}[G_i(S) S^k \mathbf{1}_{S>u}]\big)' &= \left(\frac{\mathbb{E}[S^k \mathbf{1}_{S>u}]}{\mathbb{E}[\mathbf{1}_{S>u}]} \mathbb{E}[G_i(S) \mathbf{1}_{S>u}]\right)', \\
-u^k G_i(u) f_S(u) &= \frac{-u^k f_S(u) \mathbb{E}[\mathbf{1}_{S>u}] + f_S(u) \mathbb{E}[S^k \mathbf{1}_{S>u}]}{\mathbb{E}[\mathbf{1}_{S>u}]^2} \mathbb{E}[G_i(S) \mathbf{1}_{S>u}] - \frac{\mathbb{E}[S^k \mathbf{1}_{S>u}]}{\mathbb{E}[\mathbf{1}_{S>u}]} G_i(u) f_S(u), \\
G_i(u)(\mathbb{E}[S^k | S > u] - u^k) &= \mathbb{E}[G_i(S) | S > u](\mathbb{E}[S^k | S > u] - u^k),
\end{aligned}$$

from which we must have

$$G_i(u) = \mathbb{E}[G_i(S_X) |\ S_X > u] \quad \text{for all } u \geq 0.$$

If S has a discrete law with support $\{u_0, u_1, \dots\}$, then we follow a similar procedure and we get:

$$G_i(u_m) = \frac{\mathbb{E}[G_i(S)(S^k - u_m^k) \mathbf{1}_{S>u_m}]}{\mathbb{E}[(S^k - u_m^k) \mathbf{1}_{S>u_m}]} \quad \text{for all } m \in \mathbb{N}.$$

In both cases, $G_i(u) \equiv \text{const}$, which alongside the fact that $\mathbb{E}[G_i(S_X)] = 0$, implies $G_i(u) \equiv 0$. Furthermore, as we assumed that $r_{q,i} = \tilde{r}_{q,i}$ for all $q \in [0,\ 1)$, we have $r_{0,i} = \tilde{r}_{0,i}$, and so

$$\mathbb{E}[R_i |\, S_X] = \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_X]}$$

or, equivalently,

$$\mathbb{E}[X_i|\, S_X] = \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_X]} S_X$$

for $i = 1, \ldots, n$. Finally, we have the following implication in terms of the Laplace transform of the loss RV $S_X^{(*)i}$ and for $i = 1, \ldots, n$,

$$\phi_{S_X^{(*)i}}(t) = \frac{\mathbb{E}[X_i\, e^{-tS_X}]}{\mathbb{E}[X_i]} = \frac{\mathbb{E}\big[\mathbb{E}[X_i|S_X]\, e^{-t\,S_X}\big]}{\mathbb{E}[X_i]} = \frac{\mathbb{E}[S_X\, e^{-tS_X}]}{\mathbb{E}[S_X]} = \phi_{S_X^*}(t), \quad \mathrm{Re}(t) > 0.$$

This implies $S_X^{(*)i} =_d S_X^{(*)j}$ for all $1 \leq i \neq j \leq n$ and so completes the 'only if' direction of the theorem.

In order to prove the 'if' direction of the theorem, let us assume that the distributional equality $S_X^{(*)i} =_d S_X^{(*)j} (=_d S_X^*)$ holds for all $i = 1, \ldots, n$, which means

$$\frac{\mathbb{E}[X_i\, e^{-tS_X}]}{\mathbb{E}[X_i]} = \frac{\mathbb{E}[S_X\, e^{-tS_X}]}{\mathbb{E}[S_X]}$$

or, equivalently,

$$\mathbb{E}\left[\frac{\mathbb{E}[X_i \,|\, S_X]}{\mathbb{E}[X_i]}\, e^{-tS_X}\right] = \mathbb{E}\left[\frac{S_X}{\mathbb{E}[S_X]}\, e^{-tS_X}\right],$$

with the immediate implication

$$\mathbb{E}[R_i|S_X = u] = \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_X]} \quad \text{for all } u \geq 0.$$

Therefore, we necessarily have $\mathbb{E}[R_i] = \mathbb{E}[X_i] \,/\, \mathbb{E}[S_X]$, $i = 1, \ldots, n$. Finally, we obtain the following string of equations:

$$\begin{aligned}
\mathrm{Cov}(R_i, S_X \,|\, S_X > u) &= \mathbb{E}[R_i\, S_X \,|\, S_X > u] - \mathbb{E}[R_i \,|\, S_X > u]\, \mathbb{E}[S_X \,|\, S_X > u] \\
&= \mathbb{E}\big[\mathbb{E}[R_i|S_X]\, S_X \big| S_X > u\big] - \mathbb{E}\big[\mathbb{E}[R_i|S_X]\big|S_X > u\big]\, \mathbb{E}[S_X|S_X > u] \\
&= \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_X]}\, \mathbb{E}[S_X|S_X > u] - \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_X]}\, \mathbb{E}[S_X|S_X > u] \\
&= 0
\end{aligned}$$

for all $u \geq 0$ and $i = 1, \ldots, n$. The 'if' direction of the theorem is then proved by evoking Proposition 8. This completes the proof of the theorem. □

Some properties of the portfolios of losses $\boldsymbol{X} \in \mathfrak{W}$ are studied next. Specifically, it turns out that these portfolios are *consistent* in the sense that the answer to Question 1 must be in affirmative for all their sub-portfolios. This is formulated and proved next.

Theorem 9. *Consider the loss RV* $\boldsymbol{X}_+ = (X_1, \ldots, X_n, Y_{n+1}, \ldots, Y_{n+m}) \in \mathcal{X}^{n+m}$ *and assume that* $\boldsymbol{X}_+ \in \mathfrak{W}$*, then the sub-portfolios* $(X_1, \ldots, X_n)$ *and* $(Y_{n+1}, \ldots, Y_{n+m})$ *also belongs to the set* $\mathfrak{W}$*.*

Proof. We prove that if $\mathbf{X}_+ \in \mathfrak{W}$, then $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathfrak{W}$; the case $\boldsymbol{Y} = (Y_{n+1}, \ldots, Y_{n+m}) \in \mathfrak{W}$ follows in the same fashion. Let $S_X = \sum_{i=1}^{n} X_i$, $S_Y = \sum_{i=n+1}^{n+m} Y_i$ and $S_+ = S_X + S_Y$, as in Lemma 2. Because $\boldsymbol{X}_+ \in \mathfrak{W}$ and by Theorem 8, we have, for all $i \neq j \in \{1, \ldots, n\}$ and $\mathrm{Re}(t) > 0$,

$$\begin{aligned} \phi_{S_+^{(*)_i}}(t) := \mathbb{E}\big[\exp\{-t\, S_+^{(*)_i}\}\big] &= \frac{\mathbb{E}\big[X_i\, e^{-t(S_X+S_Y)}\big]}{\mathbb{E}[X_i]} \\ &= \frac{\mathbb{E}\big[X_j\, e^{-t(S_X+S_Y)}\big]}{\mathbb{E}[X_j]} = \mathbb{E}\big[\exp\{-t\, S_+^{(*)_j}\}\big] =: \phi_{S_+^{(*)_j}}(t). \end{aligned} \tag{3.9}$$

Therefore, we have

$$\mathbb{E}\left[e^{-t\,S_Y}\,\mathbb{E}\left[\frac{X_i}{\mathbb{E}[X_i]}e^{-tS_X}\,\middle|\,S_Y\right]\right] = \mathbb{E}\left[e^{-t\,S_Y}\,\mathbb{E}\left[\frac{X_j}{\mathbb{E}[X_j]}e^{-tS_X}\,\middle|\,S_Y\right]\right] \quad \text{for all } \mathrm{Re}(t) > 0,$$

from which we can conclude

$$\mathbb{E}\left[\frac{X_i}{\mathbb{E}[X_i]}\,e^{-tS_X}\,\middle|\,S_Y\right] = \mathbb{E}\left[\frac{X_j}{\mathbb{E}[X_j]}\,e^{-tS_X}\,\middle|\,S_Y\right] \quad \text{for all } \mathrm{Re}(t) > 0.$$

The assertion follows by the law of total expectation and evoking again Theorem 8. □

Theorem 9 remains true if a *split* results in more than two loss portfolios and implies that, when starting with a loss portfolio in the set $\mathfrak{W}$, the split operation yields loss portfolios that are also in the set $\mathfrak{W}$.

The next result emphasizes that the *merge* operation - an opposite of split - is more intricate, but that the merge of loss portfolios belonging to the set $\mathfrak{W}$ may result in a loss portfolio that also belongs to the set $\mathfrak{W}$.

Theorem 10. *Consider two independent loss portfolios,* $(X_1, \ldots, X_n) \in \mathfrak{W}$ *and* $(Y_{n+1}, \ldots, Y_{n+m}) \in \mathfrak{W}$*, and denote by* S_X *and* S_Y *the corresponding sums of coordinates. Also, let* $\boldsymbol{X}_+ = (X_1, \ldots, X_n, Y_{n+1}, \ldots, Y_{n+m}) \in \mathcal{X}^{n+m}$ *be the merged portfolio. Then,* $\boldsymbol{X}_+ \in \mathfrak{W}$ *if and only if, for* $i \in \{1, \ldots, n\}$ *and* $j \in \{n+1, \ldots, n+m\}$*,*

$$\frac{\phi_{S_X^{(*)_i}}(t)}{\phi_{S_Y^{(*)_j}}(t)} = \frac{\phi_{S_X}(t)}{\phi_{S_Y}(t)}, \quad \mathrm{Re}(t) > 0. \tag{3.10}$$

Proof. Let $S_+ = S_X + S_Y$. We need to show that $S_+^{(*)_i} =_d S_+^{(*)_j}$ for all $i \neq j \in \{1, \ldots, n+m\}$. First, consider the case in which the indices i, j belong to either one of the sets $\{1, \ldots, n\}$ or $\{n+1, \ldots, n+m\}$, say $i \in \{1, \ldots, n\}$ and $j \in \{n+1, \ldots, n+m\}$. Then by Lemma 2 with the addition of the independence assumption and since $(X_1, \ldots, X_n) \in \mathfrak{W}$ and $(Y_{n+1}, \ldots, Y_{n+m}) \in \mathfrak{W}$, we have

$$\phi_{S_+^{(*)_i}}(t) = \phi_{S_X^{(*)_i}}(t) \times \phi_{S_Y}(t) = \phi_{S_Y^{(*)_j}}(t) \times \phi_{S_X}(t) = \phi_{S_+^{(*)_j}}(t)$$

for all $\mathrm{Re}(t) > 0$ if and only if Equation (3.10) is valid.

The case when $i \neq j$ are both in $\{1, \ldots, n\}$ or both in $\{n+1, \ldots, n+m\}$ follows using Theorem 8 instead of (3.10). The "only if" part is immediate from Theorem 8. This completes the proof of the assertion. □

The assertion that concludes this section reveals that an *amalgamation* - on a BU basis - of a collection of loss portfolios, each of which belongs to the set $\mathfrak{W}$, may result in a loss portfolio that also belongs to the set $\mathfrak{W}$.

Theorem 11. *Consider two independent loss portfolios,* $(X_1, \ldots, X_n) \in \mathfrak{W}$ *and* $(Y_1, \ldots, Y_n) \in \mathfrak{W}$*. Let* $\mathcal{S} = (S_1, \ldots, S_n)$*, where* $S_i = X_i + Y_i$*,* $i = 1, \ldots, n$*. Then* $\mathcal{S} \in \mathfrak{W}$ *if and only if*

$$\frac{\mathbb{E}[X_i]}{\mathbb{E}[X_j]} = \frac{\mathbb{E}[Y_i]}{\mathbb{E}[Y_j]} \quad \textit{for all } i \neq j \in \mathcal{N}. \tag{3.11}$$

Proof. Let $S_X = X_1 + \cdots + X_n$, $S_Y = Y_1 + \cdots + Y_n$, and $S_+ = S_X + S_Y$. By Lemma 2 with the addition of the independence assumption, we have, for $i = 1, \ldots, n$,

$$\phi_{S_+^{(*)_i}}(t) = \frac{\mathbb{E}\big[S_i\, e^{-t\, S_+}\big]}{\mathbb{E}[S_i]} = \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_i]}\, \phi_{S_X^{(*)_i}}(t)\, \phi_{S_Y}(t) + \frac{\mathbb{E}[Y_i]}{\mathbb{E}[S_i]}\, \phi_{S_Y^{(*)_i}}(t)\, \phi_{S_X}(t) \quad \text{for all } \mathrm{Re}(t) > 0.$$

Then by Theorem 8:

$$\phi_{S_+^{(*)_i}}(t) = \frac{\mathbb{E}[X_i]}{\mathbb{E}[S_i]}\, \phi_{S_X^{(*)_j}}(t)\, \phi_{S_Y}(t) + \frac{\mathbb{E}[Y_i]}{\mathbb{E}[S_i]}\, \phi_{S_Y^{(*)_j}}(t)\, \phi_{S_X}(t) \quad \text{for all } \mathrm{Re}(t) > 0.$$

Condition (3.11) implies $\frac{\mathbb{E}[X_i]}{\mathbb{E}[S_i]} = \frac{\mathbb{E}[X_j]}{\mathbb{E}[S_j]}$ and $\frac{\mathbb{E}[Y_i]}{\mathbb{E}[S_i]} = \frac{\mathbb{E}[Y_j]}{\mathbb{E}[S_j]}$ thus the equality $\phi_{S_+^{(*)_i}}(t) = \phi_{S_+^{(*)_j}}(t)$ holds for all $\mathrm{Re}(t) > 0$ and $i \neq j \in \mathcal{N}$. This completes the proof of the assertion. □

We conclude this sub-section with the note that Theorems 10 and 11 stay valid even if more than two loss portfolios are considered. Specifically, consider $m(\in \mathbb{N})$ loss portfolios $\boldsymbol{X}_i = (X_{i,1}, \ldots, X_{i,n_i}) \in \mathcal{X}^{n_i}$ each have n_i BUs, where $n_i \in \mathbb{N}$, $i = 1, \ldots, m$. Assume that each portfolio $\boldsymbol{X}_i \in \mathfrak{W}$, and $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m)$ are mutually independent. Let $S_{\boldsymbol{X}_i} = X_{i,1} + \ldots + X_{i,n_i}$, then $\boldsymbol{X}_+ = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m) \in \mathfrak{W}$ if and only if

$$\frac{\phi_{S_{\boldsymbol{X}_i}^{(*)_{k_i}}}(t)}{\phi_{S_{\boldsymbol{X}_j}^{(*)_{k_j}}}(t)} = \frac{\phi_{S_{\boldsymbol{X}_i}}(t)}{\phi_{S_{\boldsymbol{X}_j}}(t)} \quad \text{for all } k_i \in \{1, \ldots, n_i\},\ k_j \in \{1, \ldots, n_j\},\ i \neq j \in \{1, \ldots, m\},\ \mathrm{Re}(t) > 0,$$

which is a multi-portfolio adjustment of Condition (3.10). Moreover, let $n_1 = \cdots = n_m = n$, $S_j = X_{1,j} + \cdots + X_{m,j}$, $j \in \{1, \ldots, n\}$, then $\mathcal{S} = (S_1, \ldots, S_n) \in \mathfrak{W}$ if and only if

$$\frac{\mathbb{E}[X_{1,i}]}{\mathbb{E}[X_{1,j}]} = \cdots = \frac{\mathbb{E}[X_{m,i}]}{\mathbb{E}[X_{m,j}]} \quad \text{for all } i \neq j \in \{1, \ldots, n\},$$

which is a multi-portfolio adjustment of Condition (3.11).

3.3.1 Examples and further elaborations

In this section, we review a few examples of those loss RVs, $\boldsymbol{X} \in \mathcal{X}^n$, for which the equality $r_{q,i} = \tilde{r}_{q,i}$ holds for all $q \in [0,\ 1)$ and $i \in \mathcal{N}$. That is, we now construct a few examples of

the loss portfolios $\boldsymbol{X} \in \mathcal{X}^n$, for which the RC allocations induced by the CTE risk measure reflect the diminishing impact of large losses on the insurers' perception of risk.

Our first example is the Liouville distributions (e.g., Gupta and Richards, 1987, for a comprehensive treatment, and Hua, 2016; McNeil and Nešlehová, 2010, for applications in the context of dependence modelling). To start with, for $\gamma \in \mathbb{R}_+$, denote by $\Gamma(\gamma)$ the complete gamma function, that is

$$\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx.$$

Also, for $\gamma_1, \ldots, \gamma_n \in \mathbb{R}_+$ and $\gamma_\bullet := \gamma_1 + \cdots + \gamma_n$, define the multivariate Beta function as

$$B(\gamma_1, \ldots, \gamma_n) = \frac{\prod_{j=1}^n \Gamma(\gamma_j)}{\Gamma(\gamma_\bullet)}.$$

Example 4. *The positive and absolutely-continuous RV, $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$, is said to be distributed Inverted-Dirichlet, succinctly $\boldsymbol{X} \sim \mathrm{ID}_n(\gamma_1, \ldots, \gamma_n, \beta)$ with the parameters $\beta, \gamma_1, \ldots, \gamma_n \in \mathbb{R}_+$, if its probability density function (PDF) is:*

$$f_{\boldsymbol{X}}(x_1, \ldots, x_n) = \frac{1}{B(\gamma_1, \ldots, \gamma_n, \beta)} \prod_{j=1}^n x_j^{\gamma_j - 1} \left(1 + \sum_{j=1}^n x_j\right)^{-(\gamma_\bullet + \beta)}, \quad x_1, \ldots, x_n \in \mathbb{R}_+,$$

(e.g., Gupta and Song, 1996; Ignatov and Kaishev, 2004, for a general discussion and applications in actuarial science, respectively).

It is not difficult to show that $\phi_{S^{(*)_i}}(t) = \phi_{S^{(*)_j}}(t)$, $\mathrm{Re}(t) > 0$ for all $1 \leq i \neq j \leq n$, and hence by Theorem 8, we have $r_{q,i} = \gamma_i/\gamma_\bullet = \tilde{r}_{q,i}$ for $q \in [0,\ 1)$ and $i \in \mathcal{N}$.

An interesting observation that paves the way for a fairly general proposition, which is stated next, is that for $\boldsymbol{X} \sim \mathrm{ID}_n(\gamma_1, \ldots, \gamma_n, \beta)$, we have the stochastic representation $X_j = Z \times Y_j$, $j = 1, \ldots, n$, where the RV $Z = \sum_{j=1}^n X_j$ has a univariate inverted beta distribution, $Z \sim IB(\gamma_\bullet, \beta)$, with the parameters $\gamma_\bullet, \beta \in \mathbb{R}_+$, and the RV $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, independent on the RV Z, is distributed multivariate Dirichlet (Ng et al., 2011).

The proof of the following assertion is readily obtained via the routine conditioning and then evoking Theorem 8 and is thus omitted.

Proposition 9. *Let the RV* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ *be independent on the RV* Z *and such that, for a constant* $b \in \mathbb{R}_+$*, the equality,* $\sum_{i=1}^n Y_i = b$*, holds almost surely. Further, let the loss portfolio* $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ *admit the stochastic representation* $X_j = Y_j \times Z,\ j \in \mathcal{N}$*. Then* $\boldsymbol{X} \in \mathfrak{W}$*.*

Proposition 9 implies that the loss RVs $X_1, \ldots, X_n$ that admit the Multiplicative Background Risk Model (MBRM) stochastic representation with the idiosyncratic risk factors (RFs), $Y_1, \ldots, Y_n$ distributed Dirichlet with parameters $\gamma_1, \ldots, \gamma_n \in \mathbb{R}_+$ and the systemic RF Z having the PDF f_Z, such that

$$f_Z(z) \propto g(z)\, z^{\gamma_\bullet - 1},\ z \in \mathbb{R}_+, \tag{3.12}$$

where $\gamma_\bullet = \sum_{i=1}^n \gamma_i$, the function $z \mapsto g(z)$ is positive, continuous and integrable in the sense of Gupta and Richards (1987), all belong to the set $\mathfrak{W}$. Some examples, in addition to the already-mentioned inverted beta distribution, of the probability distribution of the systemic RF, Z, are: the gamma distribution and the generalized mixture of exponential distributions.

The class of multivariate probability distributions that admit the stochastic representation described in Proposition 9 is called the class of *Liouville distributions*, and these distributions are one way to extend the multivariate Dirichlet distribution to the unbounded domain, $\mathbb{R}_+^n$. Another way is via the class of mixed-Gamma (MG) distributions, which has recently been presented and studied in Furman et al. (2020b). Speaking briefly and avoiding unnecessary technicalities - thus considering the simplest possible case - the loss RV $\boldsymbol{X} = (X_1, \ldots, X_n)$ is said to be distributed n-variate MG distribution if it has the PDF:

$$f_{\boldsymbol{X}}(x_1, \ldots, x_n) = \sum_{k=1}^m p_k \prod_{i=1}^n \frac{x_i^{\gamma_{i,k}-1}}{\Gamma(\gamma_{i,k})\, \theta_i^{\gamma_{i,k}}}\, e^{-x_i/\theta_i}, \quad x_1, \ldots, x_n \in \mathbb{R}_+, \tag{3.13}$$

where $\gamma_{i,k} \in \mathbb{R}_+$ and $\theta_i \in \mathbb{R}_+$ are, respectively, the shape and scale parameters, and $p_k > 0,\ k = 1, \ldots, m$ are the mixture weights satisfying $\sum_{k=1}^m p_k = 1$; succinctly, we write $\boldsymbol{X} \sim \mathrm{MG}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{p})$, where $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are the $n \times m$- and m- dimensional vectors of shape and scale parameters, respectively, and $\boldsymbol{p} = (p_1, \ldots, p_m)$.

The class of MG distributions is a generalization of the popular class of multivariate Erlang mixtures considered in Willmot and Woo (2014), albeit with (a) positive - and not positive and integer - shape parameters, and (b) possibly distinct - and not all equal - scale parameters (e.g. Lee and Lin, 2012; Verbelen et al., 2016). The class of MG distributions is connected to Question 1 in the example and proposition that follow.

Example 5. *Consider a loss portfolio* $\boldsymbol{X} \sim \mathrm{MG}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{p})$ *with the PDF as per Equation (3.13), but with* $\theta_i \equiv \theta$*. Then, for* $i \in \mathcal{N}$*, we have*

$$\phi_{S_X^{(*)_i}}(t) = \mathbb{E}[\exp\{-tS_X^{(*)_i}\}] = \sum_{k=1}^{m} p_k^{(*)_i} \, (1+\theta t)^{-\gamma_{\bullet,k}-1}, \quad \mathrm{Re}(t) > 0,$$

where

$$p_k^{(*)_i} = \frac{\gamma_{i,k}}{\sum_{l=1}^{m} \gamma_{i,l} \times p_l} \, p_k, \quad k = 1, \ldots, m,$$

which can be viewed as the i-th partial size-biased transform of the PMF underlying the stochastic shape parameters. Consequently, for the equality $S_X^{(*)_i} =_d S_X^{(*)_j}$ *to hold, we must require (due to Theorem 8)*

$$\frac{\gamma_{i,1}}{\left(\sum_{j=1}^{n} \gamma_{j,1}\right)} = \cdots = \frac{\gamma_{i,m}}{\left(\sum_{j=1}^{n} \gamma_{j,m}\right)} \tag{3.14}$$

for all $i \in \mathcal{N}$*.*

The observation presented in Example 5 is strengthened in the following proposition, which concludes this section.

Proposition 10. *Let* $\boldsymbol{X} \sim \mathrm{MG}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{p})$*. Then we have* $\boldsymbol{X} \in \mathfrak{W}$ *if and only if both of* $\theta_i \equiv \theta$ *and Equation (3.14) hold true.*

Proof. Example 5 establishes the 'if' direction. In order to prove the 'only if' direction, we pursue proof by contradiction. To this end, consider $\boldsymbol{X} \sim \mathrm{MG}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{p})$ in which the coordinates of the vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ are all distinct, and suppose $\boldsymbol{X} \in \mathfrak{W}$. (If some scale parameters were equal, then we would introduce the vector of distinct scales,

$\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_{n'})$, $n' < n$ as well as, for $d = 1, \ldots, n'$ and $\mathfrak{T}_d = \big\{i \in \{1, \ldots, n\} : \theta_i = \widehat{\theta}_d\big\}$, the corresponding shape parameters $\widehat{\gamma}_{d,l} = \sum_{i \in \mathfrak{T}_d} \gamma_{i,l},\ l = 1, \ldots, m$. We would then proceed with the proof, as it is outlined below.) Further, without loss of generality, assume that the shape parameters are ordered as $\gamma_{d,1} \leq \gamma_{d,2} \leq \cdots \leq \gamma_{d,m}$, $d = 1, \ldots, n$.

With the above in mind and for any BU, $j \in \mathcal{N}$, we have that the Laplace transform of the RV $S_X^{(*)_j}$ is:

$$\phi_{S_X^{(*)_j}}(t) = \sum_{k=1}^{m} p_k^{(*)_j} \left(1 + \theta_j\, t\right)^{-(1+\gamma_{j,k})} \prod_{d=1, d\neq j}^{n} (1 + \theta_d\, t)^{-\gamma_{d,k}}, \quad \mathrm{Re}(t) > 0.$$

Furthermore, as $\boldsymbol{X} \in \mathfrak{W}$, we have that Theorem 8 implies, for $1 \leq i \neq j \leq n$ and all $\mathrm{Re}(t) > 0$,

$$\phi_{S_X^{(*)_i}}(t) = \phi_{S_X^{(*)_j}}(t).$$

However, this is impossible, which is easily seen by comparing, e.g., the m-th terms of the Laplace transforms $\phi_{S_X^{(*)_i}}$ and $\phi_{S_X^{(*)_j}}$. Hence, we have arrived at a contradiction and the proposition is proved. □

3.4 The case of independent losses

In this section, we explore the loss portfolios $\boldsymbol{X} \in \mathcal{X}^n$ that have independent constituents. Admittedly, the assumption of independence simplifies the problem postulated in Question 1 considerably, yet nor it means that the RV $\mathbb{R} = (R_1, \ldots, R_n)$ has independent coordinates, neither that the RVs $\mathbb{R}$ and S_X are independent, thus warranting a separate discussion.

Theorem 12. *Assume that $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ is a portfolio of independent losses, then we have the equality $r_{q,i} = \tilde{r}_{q,i} (= \mathbb{E}[X_i] \,/\, \mathbb{E}[S])$ for all $q \in [0,\ 1)$ and $i \in \mathcal{N}$, if and only if the equality*

$$\phi_{X_i}(t) = \big(\phi_{X_j}(t)\big)^{\mathbb{E}[X_i]/\mathbb{E}[X_j]}$$

holds for all $i \neq j \in \mathcal{N}$ and $\mathrm{Re}(t) > 0$.

Proof. By Theorem 8 and assuming that the RVs $X_1, \ldots, X_n$ are mutually independent, we have $\boldsymbol{X} \in \mathfrak{W}$ if and only if the Laplace transforms of the RVs $X_i + X_j^*$ and $X_i^* + X_j$ agree for all $i \neq j \in \mathcal{N}$. That is, we must have

$$\frac{1}{\mathbb{E}[X_j]} \frac{\frac{d}{dt}\phi_{X_j}(t)}{\phi_{X_j}(t)} = \frac{1}{\mathbb{E}[X_i]} \frac{\frac{d}{dt}\phi_{X_i}(t)}{\phi_{X_i}(t)},$$

for all $i \neq j \in \mathcal{N}$ and $\mathrm{Re}(t) > 0$. This, in turn, is equivalent to

$$\frac{\frac{d}{dt}\log \phi_{X_j}(t)}{\frac{d}{dt}\log \phi_{X_i}(t)} = \frac{\mathbb{E}[X_j]}{\mathbb{E}[X_i]}, \tag{3.15}$$

implying, for all $\mathrm{Re}(t) > 0$,

$$\frac{\log \phi_{X_j}(t)}{\log \phi_{X_i}(t)} = \frac{\mathbb{E}[X_j]}{\mathbb{E}[X_i]}. \tag{3.16}$$

The fact that Equation (3.16) leads to Equation (3.15) is easy to check by routine differentiation in the latter equation. This completes the proof of the theorem. □

3.4.1 Examples and further elaborations

Examples of the loss portfolios $\boldsymbol{X} \in \mathcal{X}^n$ that have independent constituents and also belong to the set $\mathfrak{W}$ are really numerous. For instance, consider losses X_i, $i \in \mathcal{N}$, that have infinitely divisible distributions and such that the condition in Theorem 12 holds, then we have $\tilde{r}_{q,i} = r_{q,i} = \mathbb{E}[X_i] \,/\, \mathbb{E}[S_X]$ for any $q \in [0,\ 1)$. The next example enumerates some of the distributions of relevance, which play important roles in actuarial science and quantitative risk management.

Example 6. *Assume that the portfolio of losses $\boldsymbol{X} \in \mathcal{X}^n$ has independent constituents $X_1, \ldots, X_n$, then it belongs to the set $\mathfrak{W}$ given that these constituents have the following probability distributions:*

- $X_i \sim$ Negative-Binomial(β_i, p), $\beta_i \in \mathbb{R}_+$, $p \in (0, 1)$, *with mean* $\mathbb{E}[X_i] = \beta_i\,(1-p) \,/\, p$

and Laplace transform

$$\phi_{X_i}(t) = \Big(\frac{p}{1-(1-p)e^{-t}}\Big)^{\beta_i}, \quad \mathrm{Re}(t) > 0.$$

- $X_i \sim \mathrm{Gamma}(\gamma_i, \beta)$, $\gamma_i \in \mathbb{R}_+$, $\beta \in \mathbb{R}_+$, *with mean* $\mathbb{E}[X_i] = \gamma_i\,\beta$ *and Laplace transform*

$$\phi_{X_i}(t) = (1+\beta\,t)^{-\gamma_i}, \quad \mathrm{Re}(t) > 0.$$

- $X_i \sim \text{Inverse-Gaussian}(\mu_i, \mu_i^2)$, $\gamma_i \in \mathbb{R}_+$, *with mean* $\mathbb{E}[X_i] = \mu_i$ *and Laplace transform*

$$\phi_{X_i}(t) = \exp\Big\{\mu_i(1-\sqrt{1-2t})\Big\}, \quad \mathrm{Re}(t) > 0.$$

3.5 Further generalizations and afterthoughts

We mentioned in Section 3.2 that the CTE risk measure is a member of the class of weighted risk measures and that it induces the RC allocation $r_{q,i}$, $q \in [0,\ 1)$, $i \in \mathcal{N}$. In fact, a more encompassing class of risk measures - and hence a generalization of the CTE risk measure - can be defined as follows (Furman and Zitikis, 2008a). Let $v, w : [0,\ \infty) \to [0,\ \infty)$ be two (non-decreasing) functions, then the *generalized weighted* risk measure is the map $H_{v,w} : \mathcal{X} \to [0,\ \infty)$, which, when well-defined and finite, is given by

$$H_{v,w}(X) = \frac{\mathbb{E}\big[v(X)\,w(X)\big]}{\mathbb{E}\big[w(X)\big]}, \quad X \in \mathcal{X}. \tag{3.17}$$

This is a slight generalization of the weighted class, see chapter 2 equation (2.3) for details. For $w(x) = \mathbb{1}\{x \geq \mathrm{VaR}_q(X)\}$ and $v(x) = x$, where $x \in [0,\ \infty)$ and $q \in [0,\ 1)$, we have that the generalized weighted risk measure reduces to the CTE risk measure.

Further, for $k \in \mathbb{N}$, set $v(x) = x^k$, $x \in [0,\ \infty)$ and keep the *weight* function $x \mapsto w(x)$ equal the indicator function as before in order to emphasize the tail loss scenarios, then generalized weighted risk measure (3.17) yields the *k-th order* CTE risk measure. Furthermore,

extending the notation in Section 3.2, let, for $q \in [0,\ 1),\ k \in \mathbb{N}$ and $i \in \mathcal{N}$

$$\tilde{r}^k_{q,i} = \mathbb{E}[R^k_i |\ S > s_q] \quad \text{and} \quad r^k_{q,i} = \mathbb{E}[X^k_i |\ S > s_q] \,/\, \mathbb{E}[S^k |\ S > s_q].$$

In general, the proportional k-th order CTE-based allocation $r^k_{q,i}$ is not fully-additive i.e. the allocations do not sum up to the risk measure. Nevertheless, it is a meaningful quantity in quantitative risk management (e.g., Furman and Landsman, 2006; Kim, 2010; Landsman et al., 2016, for elaborations and applications).

It is not difficult to see that, for a fixed $k \in \mathbb{N}$, the equality $\tilde{r}^k_{q,i} = r^k_{q,i}$ holds for all $q \in [0,\ 1)$ and $i \in \mathcal{N}$, if an only if we have $\mathrm{Cov}(R^k_i, S^k \,|\, S > s_q) \equiv 0$. Therefore, it is natural to reformulate Question 1 as follows.

Question 2. *For loss RVs $X_i \in L^k$, can we characterize those loss portfolios $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$, for which the equality $\tilde{r}^k_{q,i} = r^k_{q,i}$ holds for all $q \in [0,\ 1),\ i \in \mathcal{N}$, and a fixed $k \in \mathbb{N}$?*

Question 2 seeks to characterize the RVs that belong to the set

$$\mathfrak{W}_k = \big\{\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n \ : r^k_{q,i} = \widetilde{r}^k_{q,i} \ \text{ for all } q \in [0,1) \text{ and } i \in \mathcal{N}\big\}, \tag{3.18}$$

which we do next. To start off, note that if the equality $r^k_{q,i} = \widetilde{r}^k_{q,i}$ holds for all $q \in [0,\ 1)$, then setting $q = 0$, implies that for all loss portfolios in the set $\mathfrak{W}_k$, we must have $r^k_{q,i} = \widetilde{r}^k_{q,i} = \mathbb{E}[X^k_i] \,/\, \mathbb{E}[S^k_X]$.

Theorem 13. *If the portfolio of losses $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ with $X_i \in L^k$ belongs to the set $\mathfrak{W}_k$ and so the equality $r^k_{q,i} = \tilde{r}^k_{q,i} (= \mathbb{E}[X^k_i] \,/\, \mathbb{E}[S^k_X])$ holds for all $q \in [0,\ 1),\ i \in \mathcal{N}$, and a fixed $k \in \mathbb{N}$, then $S^{(k)_i}_X =_d S^{(k)_j}_X,\ 1 \le i \neq j \le n$ (where $S^{(k)_i}_X$ is the size-bias of the sum w.r.t. x^k_i). The opposite direction does not hold.*

Proof. The proof follows the same argumentation as in Theorem 8 with the quantities $S_X,\ R_i$ and $G_i(s) = \mathbb{E}[R_i | S_X = s] - \mathbb{E}[R_i]$ replaced with the quantities $S^k_X,\ R^k_i$ and $G^k_i(s) = \mathbb{E}[R^k_i | S_X = s] - \mathbb{E}[R^k_i],\ s \in [0,\ \infty)$.

To see that the distributional equality, $S^{(k)_i}_X =_d S^{(k)_j}_X,\ 1 \le i \neq j \le n$, does not imply $\boldsymbol{X} \in \mathfrak{W}_k$, consider the RV $(X_1, X_2) \in \mathcal{X}^2$ that has independent and identically distributed

constituents; $X_i \sim \mathrm{Uni}[0,\ 1],\ i = 1, 2$. Clearly, we have $S_X^{(k)_1} =_d S_X^{(k)_2},\ k \in \mathbb{N}$. Nevertheless, with some algebra we obtain, for $i \in \{1,\ 2\}$,

$$\mathbb{E}[X_i^k | S_X = s] = \frac{s^k}{1+k} \mathbb{1}_{\{0 \le s < 1\}} + \frac{1 - (s-1)^{k+1}}{(1+k)(2-s)} \mathbb{1}_{\{1 \le s \le 2\}}, \quad s \in \mathbb{R}_+,$$

which implies $\mathbb{E}[R_i^k |\ S_X = s] \neq \text{const}$ for $k \neq 1$ and hence $\tilde{r}_{q,i}^k \neq r_{q,i}^k$. The assertion holds since the regression condition is the central link in the results of Theorem 8. This completes the proof of the theorem. □

According to Theorem 13, if the equalities $r_{q,i} = \tilde{r}_{q,i}$ and $r_{q,i}^2 = \tilde{r}_{q,i}^2$ hold for all $q \in [0,\ 1)$ and $i \in \mathcal{N}$, then we must have $\mathbb{E}[R_i |\ S_X = s] \equiv \text{const}$ and $\mathbb{E}[R_i^2 |\ S_X = s] \equiv \text{const}$, respectively. Therefore, the fact $\boldsymbol{X} \in \mathfrak{W}_1$ does not imply $\boldsymbol{X} \in \mathfrak{W}_2$ (also due to the counter example in the proof of Theorem 13). Next example demonstrates that this statement, when formulated in the opposite direction, does not hold either.

Example 7. *Consider again the MG distribution as per Example 5, i.e., let* $\boldsymbol{X} \sim \mathrm{MG}_n(\boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{p})$, *and set* $\theta_i \equiv \theta \in \mathbb{R}_+$ *and*

$$\frac{\gamma_{i,1}\,(\gamma_{i,1}+1)}{\left(\sum_{j=1}^n \gamma_{j,1}\right)\left(\sum_{j=1}^n \gamma_{j,1}+1\right)} = \cdots = \frac{\gamma_{i,m}\,(\gamma_{i,m}+1)}{\left(\sum_{j=1}^n \gamma_{j,m}\right)\left(\sum_{j=1}^n \gamma_{j,m}+1\right)}$$

for all $i \in \mathcal{N}$. *Then it is not difficult to check directly that* $r_{q,i}^2 = \tilde{r}_{q,i}^2,\ i \in \mathcal{N}$, *and therefore we have* $\mathbf{X} \in \mathfrak{W}_2$. *However, the choice of parameters above does not guarantee Equation (3.14), and consequently by Theorem 8, we do not necessarily have* $\boldsymbol{X} \in \mathfrak{W}_1$.

We conclude this section by outlining a situation in which the fact, $\boldsymbol{X} \in \mathfrak{W}_1$, does imply the fact, $\boldsymbol{X} \in \mathfrak{W}_2$, and vice versa; curiously, this connects Questions 1 and 2 to the celebrated Lukacs's proportion-sum independence theorem (Lukacs, 1955). For this, recall that the fact that the loss portfolio $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ belongs to the set $\mathfrak{W}_k,\ k \in \mathbb{N}$, or in other words, that the RVs $R_i,\ i = 1, \ldots, n$, and S_X are uncorrelated conditionally on $S_X > s$ for all $s \in [0,\ \infty)$, does not in general imply the fact that the loss RVs R_i and S_X are independent. This statement holds true even if the constituents of the loss portfolio $\boldsymbol{X} \in \mathcal{X}^n$ are independent. The following assertion delineates the cases, in which the RVs R_i

and S_X are independent in the context of Questions 1 and 2.

Corollary 1. *Assume that the loss RVs $X_1, \ldots, X_n \in \mathcal{X}$ are independent. The loss portfolio $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ belongs to the sets $\mathfrak{W}_1$ and $\mathfrak{W}_2$ if and only if the loss RV $X_i \in \mathcal{X}$ is distributed gamma with the shape and scale parameters $\gamma_i > 0$ and $\theta > 0$, respectively, $i \in \mathcal{N}$. In this case, the RVs R_i and S_X are independent.*

Proof. In order to prove the 'if' direction, we note that by Lukacs' theorem, the assumption $X_i \sim Ga(\gamma_i, \theta)$ implies $R_i \perp\!\!\!\perp S_X$ for all $i \in \mathcal{N}$, which in turn implies $\boldsymbol{X} \in \mathfrak{W}_1 \cap \mathfrak{W}_2$.

Further, let us prove the 'only if' direction. For this, fix $i \in \mathcal{N}$ and note that by Theorems 12 and 13 - with the assumption of independence made in the latter case - we arrive at the following two equations, where $\phi_{X_i}(t)$ and $\phi_{X_j}(t)$ denote, respectively, the Laplace transforms of the loss RVs X_i and X_j, and $a := \mathbb{E}[X_j] \,/\, \mathbb{E}[X_i]$ and $b := \mathbb{E}[X_i^2] \,/\, \mathbb{E}[X_j^2]$, $1 \leq i \leq j \leq n$, for the simplicity of exposition

$$\phi_{X_j}(t) = (\phi_{X_i}(t))^a\,, \quad \mathrm{Re}(t) > 0 \tag{3.19}$$

and

$$\phi_{X_j}(t)\,\frac{d^2}{dt^2}\phi_{X_i}(t) = b\,\phi_{X_i}(t)\,\frac{d^2}{dt^2}\phi_{X_j}(t), \quad \mathrm{Re}(t) > 0. \tag{3.20}$$

Rewriting the equations above in terms of the Laplace transform ϕ_{X_i} only, we obtain, for $\mathrm{Re}(t) > 0$,

$$(\phi_{X_i}(t))^a\,\frac{d^2}{dt^2}\phi_{X_i}(t) = b\,\phi_{X_i}(t)\left[a(a-1)\,(\phi_{X_i}(t))^{a-2}\left(\frac{d}{dt}\phi_{X_i}(t)\right)^2 + a\,(\phi_{X_i}(t))^{a-1}\,\frac{d^2}{dt^2}\phi_{X_i}(t)\right],$$

which, with some algebra and the notation $c = a\,b\,(a-1)\,(1-a\,b)^{-1}$, simplifies to

$$\frac{\frac{d^2}{dt^2}\phi_{X_i}(t)}{\frac{d}{dt}\phi_{X_i}(t)} = c\,\frac{\frac{d}{dt}\phi_{X_i}(t)}{\phi_{X_i}(t)}, \quad \mathrm{Re}(t) > 0.$$

After integration and some more algebra, we arrive at the following first-order non-linear ODE

$$\frac{d}{dt}\phi_{X_i}(t) = -\mathbb{E}[X_i]\;(\phi_{X_i}(t))^c\,, \quad \mathrm{Re}(t) > 0,$$

with the solution

$$\phi_{X_i}(t) = \Big[1 + (c-1)\,\mathbb{E}[X_i]\,t\Big]^{-1/(c-1)}.$$

Finally, substituting the expressions for the first and second moments of the gamma distribution in the constant c and hence noticing that $c - 1 = 1/\gamma_i$, we arrive at

$$\phi_{X_i}(t) = (1 + \theta\,t)^{-\gamma_i}\,, \quad \mathrm{Re}(t) > 0,$$

which is the Laplace transform of the RV distributed gamma with the shape and scale parameters $\gamma_i > 0$ and $\theta > 0$, respectively. Hence, $X_i \backsim Ga(\gamma_i, \sigma),\ i \in \mathcal{N}$. Also, for $\gamma_\bullet = \gamma_1 + \cdots + \gamma_n$ as before, we have $S_X \sim Ga(\gamma_\bullet, \sigma)$, and by Lukacs' theorem, the RVs $R_i = X_i \,/\, S_X$ and S_X are independent. This completes the proof of the 'only if' direction as well as of the corollary as a whole. □

Chapter 4

One model to rule them all - A novel Pareto-Dirichlet distribution

4.1 Introduction

The power law hallmark of the Pareto distribution (Pareto, 1964) took center stage in modelling extreme events in insurance and finance. The well-known result of Balkema and de Haan (1974) and Pickands III (1975) gives a powerful approximation for the excess-of-loss probability , written as $\overline{F}_Y(x) = \mathbb{P}[X - u > x | X > u]$, in terms of a Pareto law. This probability is precisely the survival of $Y = X - u | X > u$, which is the the random variable of the conditional excess of losses, that are used throughout insurance, modelling life and non-life products alike. The approximation appears when a large enough threshold $u > 0$ is taken making the survival function expressible as:

$$\overline{F}_Y(x) \approx \left(1 + \frac{x}{\zeta(u)}\right)^{-\alpha}, \tag{4.1}$$

for $\alpha > 0,\ x > 0$ and an appropriately chosen scaling function $\zeta(u)$ (see Embrechts et al. (1997) for details).

Given the popularity of the Pareto distribution, several univariate extensions have been made, notably the Feller-Pareto - proposed by Arnold (2015) - which is a transformed Beta

II law and represented as:

$$X = \mu + \sigma \left(\frac{V}{W} \right)^{\frac{1}{\gamma}}, \ X > \mu, \ \mu \in \mathbb{R}, \ \sigma, \gamma > 0, \tag{4.2}$$

where V and W are two independent gamma variates with different shapes and common unit rate. And μ, σ, γ denote the univariate transforms of location, scale and power respectively. Many well-known distributions fall under the umbrella of Feller-Pareto which made it a useful computational tool in actuarial modelling (Arnold, 2015; Kleiber and Kotz, 2003). As univariate manipulation proved convenient when studying standalone losses, the subsequent inevitable objective is to generalize the Pareto law to higher dimensions, investigating the multitude of its possible rich structures.

In the context of insurance, the need for higher dimensional dependence can not be overstated as modelling interdependent losses require such formulation. Luckily, many advances have been made in this arena; See Asimit et al. (2010), Su and Furman (2017), Chiragiev and Landsman (2009), and Chapter 6 of Arnold (2015) for a compilation of the numerous generalizations and diverse structures of the multivariate Pareto.

Most generalizations regularly exploit the representation in equation (4.2) to construct multivariate parallels. To illustrate, let's fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{N} = \{1, 2, \ldots, n\}, \ n \in \mathbb{N}$, be a label set of constituents or frequently referred to as business units. Then set $\boldsymbol{X} = (X_1, \ldots, X_n)$ to be a random vector defined stochastically as :

$$X_i = \mu_i + \sigma_i \left(\frac{V_i}{W_i} \right)^{\frac{1}{\gamma_i}}, \ X_i > \mu_i, \ \mu_i \in \mathbb{R}, \ \sigma_i, \gamma_i > 0, \ i \in \mathcal{N}. \tag{4.3}$$

When $n = 2$, and if we take each W_i to be equal to the same gamma random variable, i.e. $W_i = W \sim \text{Gamma}(\alpha, 1), \ \alpha > 0$, and

$$\mathbb{P}(V_1 > v_1, V_2 > v_2) = \exp\{-\lambda_1 v_1 - \lambda_2 v_2 - \lambda_{12} \max(v_1, v_2)\}, \ \lambda_1, \lambda_2, \lambda_{12} > 0, \tag{4.4}$$

which is the bivariate exponential of Marshall-Olkin (MO) (see Marshall and Olkin (1967)), then the resulting survival function of the bivariate Pareto can be written as :

$$\overline{F}_{\boldsymbol{X}}(x_1, x_2) = \left[1 + \lambda_1 \left(\frac{x_1 - \mu_1}{\sigma_1}\right)^{\gamma_1} + \lambda_2 \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^{\gamma_2} + \lambda_{12} \max\left(\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^{\gamma_1}, \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^{\gamma_2}\right)\right]^{-\alpha}, \tag{4.5}$$

(see equations 6.2.2 and 6.2.3 of Arnold (2015)).

On the other hand, when V_i's are i.i.d. exponentials, i.e. $V_i \sim \exp(1)$, and W_i's are sums of independent gamma variables with common unit rate (the variables are indexed by the elements of the power set of $\mathcal{N}$) then we recover the following survival function :

$$\begin{aligned}\overline{F}_{\boldsymbol{X}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left[1 + \left(\frac{x_i - \mu_i}{\sigma_i}\right)^{\gamma_i}\right]^{-\alpha_i} \times \prod_{i_1 < i_2} \left[1 + \left(\frac{x_{i_1} - \mu_{i_1}}{\sigma_{i_1}}\right)^{\gamma_{i_1}} + \left(\frac{x_{i_2} - \mu_{i_2}}{\sigma_{i_2}}\right)^{\gamma_{i_2}}\right]^{-\alpha_{i_1 i_2}} \\ \times \cdots \times \left[1 + \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^{\gamma_i}\right]^{-\alpha_{12\cdots n}}, \; x_i > \mu_i, \; \forall i \in \mathcal{N}. \end{aligned} \tag{4.6}$$

This is precisely the multivariate Pareto defined in equation 6.2.8 of Arnold (2015). We first notice that the distribution in (4.5) allows for a singular probability, mainly co-monotonicity, while the one in (4.6) is absolutely continuous with respect to the Lebesgue measure. Furthermore, the single power and additive structure of (4.5) is restrictive, not admitting independence for instance, while (4.6) enjoys a flexible multiplicative/additive form with independence being a special case. The abundant number of parameters in (4.6), in addition, conveys a more flexible yet complex choice. Each formulation, nonetheless, when used separately, guarantees certain modelling advantages but concurrently comes with inherent drawbacks.

4.2 The Pareto-Dirichlet sub-mixture

To combine the best of (4.5) and (4.6), we extend (4.5) to have a general multivariate MO distribution while retaining the additive gamma model of (4.6). Besides the useful amalgamation, the motivation for such general construction is the interpretation of the multiplicative structure within the background economy and common shock models (see Boucher et al. (2008); Tsanakas (2008) for details). Formally, let $\mathcal{P}(\mathcal{N})$ be the power set of $\mathcal{N}$ without the

empty set s.t. $\mathcal{M}_v, \mathcal{M}_w \subseteq \mathcal{P}(\mathcal{N}),\ |\mathcal{M}_v| = m_v,\ |\mathcal{M}_w| = m_w$, and $m_v, m_w \leq 2^n - 1$. Furthermore, consider the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ defined stochastically as in equation (4.3) with $\boldsymbol{V} = (V_1, \ldots, V_n)$ defined component-wise as $V_i = \min(E_B : B \in B_i),\ \forall i \in \mathcal{N}$, where $E_B \sim \exp(\lambda_B),\ \lambda_B > 0$, are independent exponentials. Similarly, $\boldsymbol{W} = (W_1, \ldots, W_n)$ have $W_i = \sum_{B \in B_i} Z_B,\ \forall i \in \mathcal{N}$, where $Z_B \sim \text{Gamma}(\alpha_B, 1),\ \alpha_B > 0$, are independent gamma variates. The sets $B_i, \forall i \in \mathcal{N}$, are defined as the collection of the elements of $\mathcal{M}_v$ or $\mathcal{M}_w$ which include the index i i.e. $B_i = \bigcup\{B \in \mathcal{M} : i \in B\}$. In other words, B_i is the specific subset of $\mathcal{M}_v$ or $\mathcal{M}_w$ which includes the factors affecting that particular $i-$th BU. In practice, most often, the additive gamma factors are chosen so that $\boldsymbol{W}^\intercal = \boldsymbol{A}\boldsymbol{Z}^\intercal$. Meaning, $\mathcal{M}_w$ is expressible as a matrix $\boldsymbol{A}$, which is $n \times m_w$, composed of ones and zeros reflecting the choice of factors by the risk managers. And $\boldsymbol{Z} = (Z_B : B \in \mathcal{M}_w)$ is the gamma random vector. While it is more intuitive to use matrices for $\mathcal{M}_w$, as $\boldsymbol{W}$ admits the additive structure, the same can not be said for $\boldsymbol{V}$. Thus, for consistency, ease of notation, and to accommodate both $\boldsymbol{W}$ and $\boldsymbol{V}$, representing the factors choice as (subsets of) power sets, $\mathcal{M}_w$ and $\mathcal{M}_v$, is more apt.

To have a well defined random vector we assume, for each V_i and W_i, $B_i \neq \varnothing$. Lastly, we set $\boldsymbol{V}$ and $\boldsymbol{W}$ to be independent. We notice that the random vector $\boldsymbol{V} \sim \text{MO}(\lambda_B : B \in \mathcal{M}_v)$ follows a general MO structure while $\boldsymbol{W}$ follows the additive gamma model drawn from m_w independent factors. The two vectors are then joined through the multiplicative model. Together with the location, scale and power transforms of the corresponding margins, we obtain the composite vector $\boldsymbol{X}$. For what follows, for any quantity ρ, we let $\tilde{\rho}_i = \sum_{B \in B_i} \rho_B,\ \forall i \in \mathcal{N}$, and $\rho^+ = \sum_{B \in \mathcal{M}} \rho_B$ for $\mathcal{M} = \mathcal{M}_v, \mathcal{M}_w$. Finally, for ease of notation, the subscript B in ρ_B denotes an ascending enumeration of the elements, for example, if $B = \{1, 2\}$ then $\rho_B = \rho_{12}$.

Similar to many higher dimensional generalizations, the above formulation still maintains Pareto margins. The next remark is a standard result showing that in our setting the margins of (4.3) are of the Pareto IV kind.

Remark 8. *Since* $V_i \sim \exp(\tilde{\lambda}_i)$ *and* $W_i \sim \text{Gamma}(\tilde{\alpha}_i, 1)$, *then the univariate survival func-*

tion can be easily obtained as (denoting $y_i = \left(\frac{x_i-\mu_i}{\sigma_i}\right)^{\gamma_i}$*):*

$$\begin{aligned}\overline{F}_i(x_i) &= \frac{1}{\Gamma(\tilde{\alpha}_i)}\int_0^\infty \mathbb{P}\left[V_i > y_i w_i | W_i = w_i\right] w_i^{\tilde{\alpha}_i-1} e^{-w_i} \mathrm{d}w_i, \\ &= \frac{1}{\Gamma(\tilde{\alpha}_i)}\int_0^\infty w_i^{\tilde{\alpha}_i-1} e^{-(\tilde{\lambda}_i y_i+1)w_i} \mathrm{d}w_i, \\ &= \left(1+\tilde{\lambda}_i y_i\right)^{-\tilde{\alpha}_i}, \\ &= \left(1+\tilde{\lambda}_i\left(\frac{x_i-\mu_i}{\sigma_i}\right)^{\gamma_i}\right)^{-\tilde{\alpha}_i} \quad x_i > \mu_i.\end{aligned} \tag{4.7}$$

Which is the Pareto IV family (See Arnold (2015)).

Remark 9. *The k-th raw moment of* X_i *is given by:*

$$\mathbb{E}\left[X_i^k\right] = \sum_{j=0}^{k}\binom{k}{j}\mu_i^{k-j}\left(\sigma_i\tilde{\lambda}_i^{\frac{-1}{\gamma_i}}\right)^j \frac{\Gamma\left(\frac{j+\gamma_i}{\gamma_i}\right)\Gamma\left(\frac{\tilde{\alpha}_i\gamma_i-j}{\gamma_i}\right)}{\Gamma(\tilde{\alpha}_i)}, \ \tilde{\alpha}_i > \frac{k}{\gamma_i}, \tag{4.8}$$

otherwise infinite (for $\tilde{\alpha}_i \leq \frac{k}{\gamma_i}$*). In particular:*

$$\mathbb{E}[X_i] = \mu_i + \sigma_i\tilde{\lambda}_i^{\frac{-1}{\gamma_i}} \frac{\Gamma\left(\frac{1+\gamma_i}{\gamma_i}\right)\Gamma\left(\frac{\tilde{\alpha}_i\gamma_i-1}{\gamma_i}\right)}{\Gamma(\tilde{\alpha}_i)}, \tag{4.9}$$

and

$$\begin{aligned}\mathrm{Var}[X_i] = \left(\sigma_i\tilde{\lambda}_i^{\frac{-1}{\gamma_i}}\right)^2 \frac{\Gamma\left(\frac{2+\gamma_i}{\gamma_i}\right)\Gamma\left(\frac{\tilde{\alpha}_i\gamma_i-2}{\gamma_i}\right)}{\Gamma(\tilde{\alpha}_i)} - \left(\sigma_i\tilde{\lambda}_i^{\frac{-1}{\gamma_i}}\right) \frac{\Gamma\left(\frac{1+\gamma_i}{\gamma_i}\right)\Gamma\left(\frac{\tilde{\alpha}_i\gamma_i-1}{\gamma_i}\right)}{\Gamma(\tilde{\alpha}_i)} \\ \times \left[\mu_i + \left(\sigma_i\tilde{\lambda}_i^{\frac{-1}{\gamma_i}}\right) \frac{\Gamma\left(\frac{1+\gamma_i}{\gamma_i}\right)\Gamma\left(\frac{\tilde{\alpha}_i\gamma_i-1}{\gamma_i}\right)}{\Gamma(\tilde{\alpha}_i)}\right].\end{aligned} \tag{4.10}$$

(See equation 3.3.7 in Arnold (2015)).

To unveil the richness of our proposed mixture model, in the next theorem, we will derive the general dependence structure of $\boldsymbol{X}$. Since μ, σ, γ are increasing transformations, then the underlying copula is invariant. Thus, without loss of generality, we will assume $\mu_i = 0,\ \sigma_i = 1$ and $\gamma_i = 1,\ \forall i \in \mathcal{N}$. We could retrieve the distributional results in terms of

the univariate transforms simply by replacing x_i with $\left(\frac{x_i-\mu_i}{\sigma_i}\right)^{\gamma_i}$, $\forall\, i \in \mathcal{N}$, on the right hand side of the equations.

Since the composite vector $\boldsymbol{X}$ contains the case of (4.5), it means it allows for a singular co-monotonic probability. The next two definitions deal with this case by showing the necessary structural conditions on $\mathcal{M}_w$ and $\mathcal{M}_v$, under which, a singular law is possible.

Definition 7. *Let $B \subseteq \mathcal{N}$, B not a singleton, then we say the random B-vector $\boldsymbol{X}_B = (X_i : i \in B)$ has a co-monotonic (singular) component if $B \in \mathcal{M}_w \cap \mathcal{M}_v$ s.t. $\forall B^{'} \in \mathcal{M}_w$, $B^{'} \neq B$, $B^{'} \cap B = \varnothing$. This is due to the fact that the gamma B-vector $\boldsymbol{W}_B = (W_i : i \in B)$ is defined as a single common factor i.e. $W_i = Z_B$, $\forall i \in B$. We will denote the collection of those disjoint co-monotonic B's by $\mathcal{S}$.*

Remark 10. *If the random B-vector has a co-monotonic component then $\forall B^{'} \in \mathcal{M}_v$, $B^{'} \subseteq B$, $B^{'}$ not a singleton, the random $B^{'}$-subvectors have a co-monotonic part as well. Let's denote the collection of those subsets of B (including B) by $\mathcal{S}_B$ s.t. $\overline{\mathcal{S}} = \cup_{B\in\mathcal{S}}\mathcal{S}_B$.*

Now we are ready to state the main theorem of this paper which depicts the survival function of $\boldsymbol{X}$ as an elegant closed form of mixed distributions.

Theorem 14. *The survival function $\overline{F}_{\boldsymbol{X}}$ of $\boldsymbol{X}$ is given as a finite sub-mixture of multivariate Pareto distributions each weighted by a Dirichlet probability. Formally:*

$$\overline{F}_{\boldsymbol{X}}(x_1,\ldots,x_n) = \sum_{j=1}^{n!} \mathrm{MP}^{(j)}\left(x_1,\ldots,x_n\right)\ \underline{\mathrm{Dir}}^{(j)}\left(x_1,\ldots,x_n\right), \tag{4.11}$$

where :

$$\mathrm{MP}^{(j)}\left(x_1,\ldots,x_n\right) = \prod_{B\in\mathcal{M}_w}\left(1+\sum_{i\in B}\lambda_i^{(j)}x_i + \sum_{B^{'}\in\mathcal{S}_B}\lambda_{B^{'}}\max(x_i : i\in B^{'})\right)^{-\alpha_B}, \tag{4.12}$$

is a generalized multivariate Pareto distribution (which is a hybrid model of (4.5) and (4.6)) s.t. $\lambda_i^{(j)}$ is a specific sum of λ's that depend on the permutation-case (j) (see Proof i for

details). And:

$$\underline{\mathrm{Dir}}^{(j)}\left(x_1,\ldots,x_n\right)=\frac{1}{\beta(\boldsymbol{\alpha})}\int\limits_{\underline{\Delta}^{(j)}(x_1,\ldots,x_n)}\prod_{B\in\mathcal{M}_w}t_B^{\alpha_B-1}\mathrm{d}\boldsymbol{t}, \tag{4.13}$$

where $\boldsymbol{\alpha}=(\alpha_B : B\in\mathcal{M}_w)$ *and* β *is the multivariate beta function. The* $\underline{\mathrm{Dir}}^{(j)}$ *terms are integrals of the Dirichlet density over specific regions of the standard simplex denoted by* $\underline{\Delta}^{(j)}(x_1,\ldots,x_n)$, $1\leq j\leq n!$ *(see Proof i for definition). The integration regions* $\underline{\Delta}^{(j)}$*s do not span the whole simplex which consequently implies* $\sum_{j=1}^{n!}\underline{\mathrm{Dir}}^{(j)}\left(x_1,\ldots,x_n\right)\leq 1$.

Proof. See appendix i. □

The resulting distribution of theorem 14 (equation (4.11)) exhibits a peculiar form. It has two characters of dependence. One is of a Pareto type and the other one is of the quasi Dirichlet variety. It possesses a far greater flexibility through the sheer number of parameters. While the mixture constitution allows for more dependence freedom. To our knowledge, this is a novel structure that has not been studied in the literature. Henceforth, we shall term it Pareto-Dirichlet.

Remark 11. *For a fixed* $(x_1,\ldots,x_n)$ *the Dirichlet terms can be viewed as discrete distribution* $p_j=\underline{\mathrm{Dir}}^{(j)}\left(x_1,\ldots,x_n\right)$. *It is a defective distribution since* $\sum_{j=1}^{n!}p_j\leq 1$ *with* $\sum_{j=1}^{n!}p_j=1$ *if and only if there is all permutations give identical* MP *terms i.e. the sub-mixture collapses to a single* MP *term with* $\sum_{j=1}^{n!}p_j=1$.

Remark 12. *The form of the survival function of* $\boldsymbol{X}$ *is governed by both* $\boldsymbol{V}$ *and* $\boldsymbol{W}$. *Where the MO form in* $\boldsymbol{V}$ *determines the integration domain of the Dirichlet probability* $\underline{\mathrm{Dir}}$ *i.e. the cases, while the additive gamma vector* $\boldsymbol{W}$ *dictates the structure of the multivariate Pareto part* MP.

Next, we will elucidate some special cases of (4.11), explicitly, by setting $\mathcal{M}_w$ and $\mathcal{M}_v$ to have a certain form.

Corollary 2. *When* $\mathcal{M}_v=\{\{1\},\{2\},\ldots,\{n\}\}$ *(singletons) and* $\mathcal{M}_w\subseteq\mathcal{P}(\mathcal{N})$, *then the*

survival function becomes:

$$\overline{F}_{\boldsymbol{X}}(x_1,\ldots,x_n) = \prod_{B\in\mathcal{M}_w}\left(1+\sum_{i\in B}\lambda_i x_i\right)^{-\alpha_B}.$$

which is the multivariate Pareto in (4.6).

When $\mathcal{M}_w = \mathcal{M}_v = \{\{1\},\{2\},\ldots,\{n\}\}$ we recover the independent case i.e. the survival function can be written as:

$$\overline{F}_{\boldsymbol{X}}(x_1,\ldots,x_n) = \prod_{i=1}^{n}\overline{F}_i(x_i) = \prod_{i=1}^{n}(1+\lambda_i x_i)^{-\alpha_i}.$$

Corollary 3. *When $\mathcal{M}_w = \{\{1\},\{2\},\ldots,\{n\}\}$ (singletons) and $\mathcal{M}_v \subseteq \mathcal{P}(\mathcal{N}))$, $\mathcal{M}_v \neq \mathcal{M}_w$, then the multivariate Pareto part simplifies to:*

$$\mathrm{MP}^{(j)}(x_1,\ldots,x_n) = \prod_{i=1}^{n}\left(1+\lambda_i^{(j)}x_i\right)^{-\alpha_i}.$$

i.e. the multivariate Pareto part disintegrates to independence and the dependence is solely determined by the Dirichlet terms.

Corollary 4. *When $\mathcal{M}_v = \overline{\mathcal{S}}\cup\{\{1\},\{2\},\ldots,\{n\}\}$ and $\mathcal{M}_w \subseteq \mathcal{P}(\mathcal{N})$, $\mathcal{M}_w \neq \mathcal{M}_v, \mathcal{S}\subseteq\mathcal{M}_w$, then the survival function can be written as:*

$$\overline{F}_{\boldsymbol{X}}(x_1,\ldots,x_n) = \prod_{B\in\mathcal{M}_w}\left(1+\sum_{i\in B}\lambda_i x_i + \sum_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i : i\in B')\right)^{-\alpha_B}.$$

i.e. there is only one case and the sub-mixture collapses to a single multivariate Pareto distribution having co-monotonic parts. This is a generalization of the model in (4.5).

When $\mathcal{M}_v = \overline{\mathcal{S}}$ and $\mathcal{M}_w = \mathcal{S}$ then we recover the purely co-monotonic survival function:

$$\overline{F}_{\boldsymbol{X}}(x_1,\ldots,x_n) = \prod_{B\in\mathcal{S}}\left(1+\sum_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i : i\in B')\right)^{-\alpha_B}.$$

Proposition 11. *Suppose that $B = \mathcal{N}$ defines a B-vector with a co-monotonic part, and*

let $\mathcal{A}_B$ ($\mathcal{A}_{B'}$) be the event when all elements of the B-vector ($B^{'}$-sub-vectors) are equal. Additionally, assume $\mu_i = 0,\ \sigma_i = 1$ and $\gamma_i = 1,\ \forall i \in B$. Then:

$$\sum_{B^{'} \in \mathcal{S}_B} \mathbb{P}\left(\mathcal{A}_{B'}\right) = 1 - \sum_{j=1}^{k!} \left(\prod_{s=2}^{k} \frac{\lambda_{l_s}^{(j)}}{\sum_{t=1}^{s} \lambda_{l_t}^{(j)}} \right).$$

Where $\lambda_{l_p}^{(j)}$ are case and order dependent sums of λ's (see proof ii for details).

Proof. See appendix ii. □

Remark 13. *Similar to setting of proposition 11, assume $B \subseteq \mathcal{N}$ defines a B-vector with a co-monotonic part. Then:*

$$\mathbb{P}\left(\mathcal{A}_B\right) = \frac{\lambda_B}{\lambda_B + \sum\limits_{B' \in \mathcal{M}_v : B' \subset B} \lambda_{B'}}.$$

Example 8. *Let $n = 4$ i.e. $\mathcal{N} = \{1,2,3,4\}$ with $\mathcal{M}_w = \{\{1,2\},\{3,4\}\}$ and $\mathcal{M}_v = \{\{i_1, i_2\} : i_1 < i_2,\ i_1, i_2 \in \mathcal{N}\}$, then the survival function can be written as:*

$$\begin{aligned}
&\overline{F}_{\boldsymbol{X}}(x_1,x_2,x_3,x_4) = \\
&\left(1 + (\lambda_{13}+\lambda_{14})x_1 + (\lambda_{23}+\lambda_{24})x_2 + \lambda_{12}\max(x_1,x_2)\right)^{-\alpha_{12}} \left(1 + \lambda_{34}\max(x_3,x_4)\right)^{-\alpha_{34}} \\
&\times\left(1 - \beta_{\frac{\max(x_3,x_4)}{\max(x_3,x_4)+\min(x_1,x_2)}}(\alpha_{12},\alpha_{34})\right) + \left(1 + (\lambda_{12}+\lambda_{\vee 3}+\lambda_{\vee 4})\max(x_1,x_2) + \lambda_{\wedge\wedge}\min(x_1,x_2)\right)^{-\alpha_{12}} \\
&\quad\times\left(1 + (\lambda_{34}+\lambda_{\wedge\vee})\max(x_3,x_4)\right)^{-\alpha_{34}} \left(\beta_{\frac{\max(x_3,x_4)}{\max(x_3,x_4)+\min(x_1,x_2)}}(\alpha_{12},\alpha_{34})\right. \\
&\quad\left. - \beta_{\max\left(\frac{\max(x_3,x_4)}{\max(x_3,x_4)+\max(x_1,x_2)}, \frac{\min(x_3,x_4)}{\min(x_3,x_4)+\min(x_1,x_2)}\right)}(\alpha_{12},\alpha_{34})\right) \\
&+\left(1 + (\lambda_{12}+\lambda_{\vee\wedge})\max(x_1,x_2)\right)^{-\alpha_{12}} \left(1 + (\lambda_{34}+\lambda_{1\vee}+\lambda_{2\vee})\max(x_3,x_4) + \lambda_{\wedge\wedge}\min(x_3,x_4)\right)^{-\alpha_{34}} \\
&\quad\times\left(\beta_{\max\left(\frac{\max(x_3,x_4)}{\max(x_3,x_4)+\max(x_1,x_2)}, \frac{\min(x_3,x_4)}{\min(x_3,x_4)+\min(x_1,x_2)}\right)}(\alpha_{12},\alpha_{34}) - \beta_{\frac{\min(x_3,x_4)}{\min(x_3,x_4)+\max(x_1,x_2)}}(\alpha_{12},\alpha_{34})\right) \\
&\quad+\left(1 + \lambda_{12}\max(x_1,x_2)\right)^{-\alpha_{12}} \left(1 + (\lambda_{13}+\lambda_{23})x_3 + (\lambda_{14}+\lambda_{24})x_4 + \lambda_{34}\max(x_3,x_4)\right)^{-\alpha_{34}}
\end{aligned}$$

$$\times \beta_{\frac{\min(x_3,x_4)}{\min(x_3,x_4)+\max(x_1,x_2)}}(\alpha_{12},\alpha_{34})$$

$$+\left(1+(\lambda_{12}+\lambda_{\vee 3}+\lambda_{\vee 4})\max(x_1,x_2)\right)^{-\alpha_{12}}\left(1+\lambda_{\wedge 3}x_3+\lambda_{\wedge 4}x_4+\lambda_{34}\max(x_3,x_4)\right)^{-\alpha_{34}}$$

$$\times \max\left(0,\ \beta_{\frac{\min(x_3,x_4)}{\min(x_3,x_4)+\min(x_1,x_2)}}(\alpha_{12},\alpha_{34})-\beta_{\frac{\max(x_3,x_4)}{\max(x_3,x_4)+\max(x_1,x_2)}}(\alpha_{12},\alpha_{34})\right)$$

$$+\left(1+\lambda_{1\wedge}x_1+\lambda_{2\wedge}x_2+\lambda_{12}\max(x_1,x_2)\right)^{-\alpha_{12}}\left(1+(\lambda_{34}+\lambda_{1\vee}+\lambda_{2\vee})\max(x_3,x_4)\right)^{-\alpha_{34}}$$

$$\times \max\left(0,\ \beta_{\frac{\max(x_3,x_4)}{\max(x_3,x_4)+\max(x_1,x_2)}}(\alpha_{12},\alpha_{34})-\beta_{\frac{\min(x_3,x_4)}{\min(x_3,x_4)+\min(x_1,x_2)}}(\alpha_{12},\alpha_{34})\right). \quad (4.14)$$

Where $\lambda_{\vee\bullet}=\lambda_{1\bullet}$, $\lambda_{\wedge\bullet}=\lambda_{2\bullet}$ *when* $x_1>x_2$ *and vice versa when* $x_2>x_1$. *Similarly, when* $x_3>x_4$ *then* $\lambda_{\bullet\vee}=\lambda_{\bullet 3}$, $\lambda_{\bullet\wedge}=\lambda_{\bullet 4}$ *and the opposite when* $x_4>x_3$. *And* $\beta_z(p,q)$ *is the regularized incomplete beta function.*

The sub-mixture expression of example 8 reveals the modelling power of the Pareto-Dirichlet. It possesses versatile combinations with both absolutely continuous and $\{\{1,2\},\{3,4\}\}$ co-monotonic relations.

4.3 The bivariate case of Pareto-Dirichlet

Describing all cases in higher dimensions is certainly an interesting exercise. However, due to its complexity and vastness, we turn to the bivariate case to shed some light into the workings of the mixture structure of $\boldsymbol{X}$. In the following examples, we will derive the detailed expressions of the bivariate case in regards to the survival functions and the product moments.

Example 9. *Let* $n=2$ *i.e.* $\mathcal{N}=\{1,2\}$ *and set* $\mathcal{M}_w=\mathcal{M}_v=\mathcal{P}(\mathcal{N})=\{\{1\},\{2\},\{1,2\}\}$. *Then the survival function can be written as:*

$$\overline{F}_{\boldsymbol{X}}(x_1,x_2)=\sum_{j=1}^{2}\mathrm{MP}^{(j)}(x_1,x_2)\ \underline{\mathrm{Dir}}^{(j)}(x_1,x_2),$$

Where:

$$\mathrm{MP}^{(1)}(x_1,x_2) = (1+(\lambda_1+\lambda_{12})x_1)^{-\alpha_1}(1+\lambda_2 x_2)^{-\alpha_2}(1+(\lambda_1+\lambda_{12})x_1+\lambda_2 x_2)^{-\alpha_{12}}$$

When $x_1 \geq x_2$:

$$\begin{aligned}\underline{\mathrm{Dir}}^{(1)}(x_1,x_2) &= \frac{1}{\beta(\boldsymbol{\alpha})}\int_0^{a_1(x_1,x_2)}\int_0^{b_1(x_1,x_2;t_1)} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,\\ &+\frac{1}{\beta(\boldsymbol{\alpha})}\int_{a_1(x_1,x_2)}^{1}\int_0^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,\end{aligned}$$

When $x_1 < x_2$:

$$\begin{aligned}\underline{\mathrm{Dir}}^{(1)}(x_1,x_2) &= \frac{1}{\beta(\boldsymbol{\alpha})}\int_{a_1(x_1,x_2)}^{1}\int_0^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,\\ &+\frac{1}{\beta(\boldsymbol{\alpha})}\int_{c_1(x_1,x_2)}^{a_1(x_1,x_2)}\int_0^{b_1(x_1,x_2;t_1)} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,\end{aligned}$$

s.t. $a_1(x_1,x_2) = \frac{x_2((\lambda_1+\lambda_{12})x_1+1)}{x_1((\lambda_1+\lambda_2+\lambda_{12})x_2+1)+x_2}$, $c_1(x_1,x_2) = \frac{(x_2-x_1)((\lambda_1+\lambda_{12})x_1+1)}{x_2((\lambda_1+\lambda_2+\lambda_{12})x_1+1)}$
and $b_1(x_1,x_2;t_1) = \frac{(\lambda_2x_2+1)(x_1(x_2(\lambda_1(t_1-1)+\lambda_2t_1+\lambda_{12}(t_1-1))+(\lambda_1+\lambda_{12})x_1+1)+(t_1-1)x_2)}{x_1((\lambda_1+\lambda_{12})x_1+1)((\lambda_1+\lambda_2+\lambda_{12})x_2+1)}$.

Similarly, for $j = 2$*:*

$$\mathrm{MP}^{(2)}(x_1,x_2) = (1+\lambda_1x_1)^{-\alpha_1}(1+(\lambda_2+\lambda_{12})x_2)^{-\alpha_2}(1+\lambda_1x_1+(\lambda_2+\lambda_{12})x_2)^{-\alpha_{12}}$$

When $x_1 \geq x_2$:

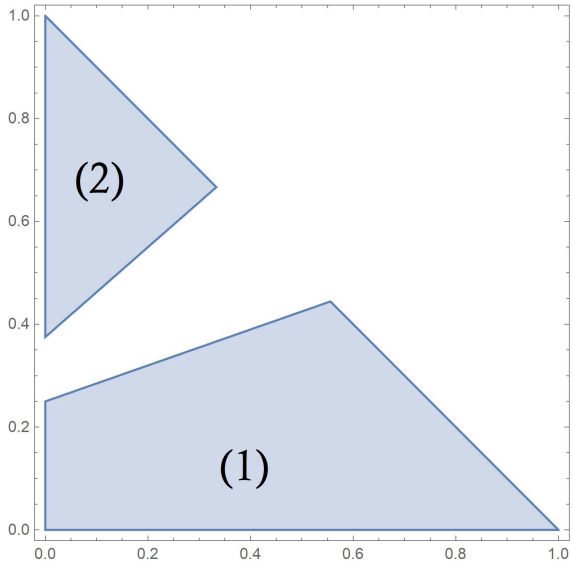
$$\underline{\mathrm{Dir}}^{(2)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_0^{a_2(x_1,x_2)}\int_{b_2(x_1,x_2;t_1)}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

When $x_1 < x_2$:

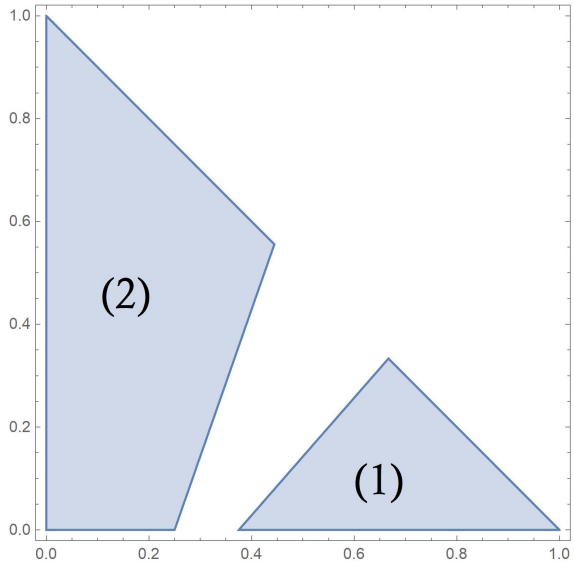
$$\begin{aligned}\underline{\mathrm{Dir}}^{(2)}(x_1,x_2) &= \frac{1}{\beta(\boldsymbol{\alpha})}\int_0^{c_2(x_1,x_2)}\int_0^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,\\ &+\frac{1}{\beta(\boldsymbol{\alpha})}\int_{c_2(x_1,x_2)}^{a_2(x_1,x_2)}\int_{b_2(x_1,x_2;t_1)}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,\end{aligned}$$

s.t. $a_2(x_1,x_2) = \frac{x_2(\lambda_1x_1+1)}{x_1((\lambda_1+\lambda_2+\lambda_{12})x_2+1)+x_2}$, $c_2(x_1,x_2) = \frac{(x_2-x_1)(\lambda_1x_1+1)}{x_2((\lambda_1+\lambda_2+\lambda_{12})x_1+1)}$
and $b_2(x_1,x_2)\frac{((\lambda_2+\lambda_{12})x_2+1)(x_1(x_2(\lambda_1(t_1-1)+(\lambda_2+\lambda_{12})t_1)+\lambda_1x_1+1)+(t_1-1)x_2)}{x_1(\lambda_1x_1+1)((\lambda_1+\lambda_2+\lambda_{12})x_2+1)}$.

The images below show the Dirichlet integration domain (on the regular 2-simplex) for the two cases $x_1 \geq x_2$ and $x_1 < x_2$. Each image labels the regions of $j = (1)$ and $j = (2)$ accordingly. For both images the parameters are $\lambda_1 = \lambda_2 = \lambda_{12} = 1$. While (x_1, x_2) are chosen to be $x_1 = 2, x_2 = 1$ for (a) and $x_1 = 1, x_2 = 2$ for (b).



(a) $x_1 \geq x_2$



(b) $x_1 < x_2$

In the following examples, we will explore the other special cases of the bivariate setting. Unlike example 9, where $\mathcal{M}_w = \mathcal{M}_v = \mathcal{P}(\mathcal{N})$, in examples 10 and 11 the aim is to show what happens when $\mathcal{M}_w \subset \mathcal{M}_v = \mathcal{P}(\mathcal{N})$ and $\mathcal{M}_v \subset \mathcal{M}_w = \mathcal{P}(\mathcal{N})$ respectively.

Example 10. *Set* $\mathcal{M}_v = \mathcal{P}(\mathcal{N}) = \{\{1\}, \{2\}, \{1, 2\}\}$. *Then:*

$$\overline{F}_{\boldsymbol{X}}(x_1, x_2) = \sum_{j=1}^{2} \mathrm{MP}^{(j)}(x_1, x_2) \ \underline{\mathrm{Dir}}^{(j)}(x_1, x_2),$$

Case 1: For $\mathcal{M}_w = \{\{1\}, \{2\}\}$*:*

$$\mathrm{MP}^{(1)}(x_1, x_2) = (1 + (\lambda_1 + \lambda_{12})x_1)^{-\alpha_1} (1 + \lambda_2 x_2)^{-\alpha_2},$$

$$\underline{\mathrm{Dir}}^{(1)}(x_1, x_2) = \frac{1}{\beta(\boldsymbol{\alpha})} \int_{a_1(x_1, x_2)}^{1} t^{\alpha_1 - 1}(1 - t)^{\alpha_2 - 1} \mathrm{d}t,$$

$$\mathrm{MP}^{(2)}(x_1, x_2) = (1 + \lambda_1 x_1)^{-\alpha_1} (1 + (\lambda_2 + \lambda_{12}) x_2)^{-\alpha_2},$$

$$\underline{\mathrm{Dir}}^{(2)}(x_1, x_2) = \frac{1}{\beta(\boldsymbol{\alpha})} \int_{a_2(x_1,x_2)}^{1} t^{\alpha_1 - 1}(1-t)^{\alpha_2 - 1} \mathrm{d}t,$$

s.t. $a_1(x_1, x_2) = \frac{x_2((\lambda_1+\lambda_{12})x_1+1)}{x_1((\lambda_1+\lambda_2+\lambda_{12})x_2+1)+x_2}$ *and* $a_2(x_1, x_2) = \frac{x_1((\lambda_2+\lambda_{12})x_2+1)}{x_1((\lambda_1+\lambda_2+\lambda_{12})x_2+1)+x_2}$.

Case 2: For $\mathcal{M}_w = \{\{1\}, \{1, 2\}\}$*:*

$$\mathrm{MP}^{(1)}(x_1, x_2) = (1 + (\lambda_1 + \lambda_{12}) x_1)^{-\alpha_1} (1 + (\lambda_1 + \lambda_{12}) x_1 + \lambda_2 x_2)^{-\alpha_{12}},$$

When $x_1 \geq x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1, x_2) = 1,$$

When $x_1 < x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1, x_2) = \frac{1}{\beta(\boldsymbol{\alpha})} \int_{a_1(x_1,x_2)}^{1} t^{\alpha_1 - 1}(1-t)^{\alpha_{12} - 1} \mathrm{d}t,$$

$$\mathrm{MP}^{(2)}(x_1, x_2) = (1 + \lambda_1 x_1)^{-\alpha_1} (1 + \lambda_1 x_1 + (\lambda_2 + \lambda_{12}) x_2)^{-\alpha_{12}},$$

When $x_1 \geq x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1, x_2) = 0,$$

When $x_1 < x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1, x_2) = \frac{1}{\beta(\boldsymbol{\alpha})} \int_{0}^{a_2(x_1,x_2)} t^{\alpha_1 - 1}(1-t)^{\alpha_{12} - 1} \mathrm{d}t,$$

s.t. $a_1(x_1, x_2) = \frac{(x_2-x_1)((\lambda_1+\lambda_{12})x_1+1)}{x_2((\lambda_1+\lambda_2+\lambda_{12})x_1+1)}$ *and* $a_2(x_1, x_2) = \frac{(x_2-x_1)(\lambda_1 x_1+1)}{x_2((\lambda_1+\lambda_2+\lambda_{12})x_1+1)}$.

Case 3: For $\mathcal{M}_w = \{\{2\}, \{1,2\}\}$*:*

$$\mathrm{MP}^{(1)}(x_1, x_2) = (1 + (\lambda_2 + \lambda_{12})x_2)^{-\alpha_2} (1 + (\lambda_2 + \lambda_{12})x_2 + \lambda_1 x_1)^{-\alpha_{12}},$$

When $x_1 > x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1, x_2) = \frac{1}{\beta(\boldsymbol{\alpha})} \int_{a_1(x_1,x_2)}^{1} t^{\alpha_2 - 1}(1-t)^{\alpha_{12}-1} \mathrm{d}t,$$

When $x_1 \leq x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1, x_2) = 1,$$

$$\mathrm{MP}^{(2)}(x_1, x_2) = (1 + \lambda_2 x_2)^{-\alpha_2} (1 + \lambda_2 x_2 + (\lambda_1 + \lambda_{12})x_1)^{-\alpha_{12}},$$

When $x_1 > x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1, x_2) = \frac{1}{\beta(\boldsymbol{\alpha})} \int_{0}^{a_2(x_1,x_2)} t^{\alpha_2 - 1}(1-t)^{\alpha_{12}-1} \mathrm{d}t,$$

When $x_1 \leq x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1, x_2) = 0,$$

s.t. $a_1(x_1, x_2) = \frac{(x_1 - x_2)((\lambda_2 + \lambda_{12})x_2 + 1)}{x_1((\lambda_1 + \lambda_2 + \lambda_{12})x_2 + 1)}$ *and* $a_2(x_1, x_2) = \frac{(x_1 - x_2)(\lambda_2 x_2 + 1)}{x_1((\lambda_1 + \lambda_2 + \lambda_{12})x_2 + 1)}$.

Example 11. *Set* $\mathcal{M}_w = \mathcal{P}(\mathcal{N}) = \{\{1\}, \{2\}, \{1,2\}\}$ *then:*

$$\overline{F}_{\boldsymbol{X}}(x_1, x_2) = \sum_{j=1}^{2} \mathrm{MP}^{(j)}(x_1, x_2) \ \underline{\mathrm{Dir}}^{(j)}(x_1, x_2),$$

Case 1: For $\mathcal{M}_v = \{\{1\}, \{1,2\}\}$*:*

$$\mathrm{MP}^{(1)}(x_1, x_2) = (1 + (\lambda_1 + \lambda_{12})x_1)^{-\alpha_1 - \alpha_{12}},$$

When $x_1 > x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_{a_1(x_1,x_2)}^{1}\int_{0}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$
$$+\frac{1}{\beta(\boldsymbol{\alpha})}\int_{0}^{a_1(x_1,x_2)}\int_{0}^{b_1(x_1,x_2;t_1)} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

When $x_1 \le x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_{a_1(x_1,x_2)}^{1}\int_{0}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$
$$+\frac{1}{\beta(\boldsymbol{\alpha})}\int_{c_1(x_1,x_2)}^{a_1(x_1,x_2)}\int_{0}^{b_1(x_1,x_2;t_1)} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

$$\mathrm{MP}^{(2)}(x_1,x_2) = (1+\lambda_1 x_1)^{-\alpha_1}(1+\lambda_{12}x_2)^{-\alpha_2}(1+\lambda_1x_1+\lambda_{12}x_2)^{-\alpha_{12}},$$

When $x_1 \ge x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_{0}^{a_2(x_1,x_2)}\int_{b_2(x_1,x_2;t_1)}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

When $x_1 < x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_{c_2(x_1,x_2)}^{a_2(x_1,x_2)}\int_{b_2(x_1,x_2;t_1)}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$
$$+\frac{1}{\beta(\boldsymbol{\alpha})}\int_{0}^{c_2(x_1,x_2)}\int_{0}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

s.t. $a_1(x_1,x_2) = \frac{x_2((\lambda_1+\lambda_{12})x_1+1)}{x_1((\lambda_1+\lambda_{12})x_2+1)+x_2}$, $a_2(x_1,x_2) = \frac{x_2(\lambda_1x_1+1)}{x_1((\lambda_1+\lambda_{12})x_2+1)+x_2}$, $b_1(x_1,x_2;t_1) = \frac{(t_1-1)x_2+x_1}{x_1((\lambda_1+\lambda_{12})x_2+1)}$, $b_2(x_1,x_2;t_1) = \frac{(\lambda_{12}x_2+1)(((t_1-1)x_2+x_1)(\lambda_1x_1+1)+\lambda_{12}t_1x_1x_2)}{x_1(\lambda_1x_1+1)((\lambda_1+\lambda_{12})x_2+1)}$, $c_1(x_1,x_2) = 1-\frac{x_1}{x_2}$ *and* $c_2(x_1,x_2) = \frac{(x_2-x_1)(\lambda_1x_1+1)}{x_2((\lambda_1+\lambda_{12})x_1+1)}$.

Case 2: For $\mathcal{M}_v = \{\{2\},\{1,2\}\}$*:*

$$\mathrm{MP}^{(1)}(x_1,x_2) = (1+\lambda_{12}x_1)^{-\alpha_1}(1+\lambda_2x_2)^{-\alpha_2}(1+\lambda_{12}x_1+\lambda_2x_2)^{-\alpha_{12}},$$

When $x_1 \ge x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_{a_1(x_1,x_2)}^{1}\int_{0}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

$$+\frac{1}{\beta(\boldsymbol{\alpha})}\int_0^{a_1(x_1,x_2)}\int_0^{b_1(x_1,x_2;t_1)} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

When $x_1 < x_2$:

$$\underline{\mathrm{Dir}}^{(1)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_{a_1(x_1,x_2)}^{1}\int_0^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$
$$+\frac{1}{\beta(\boldsymbol{\alpha})}\int_{c_1(x_1,x_2)}^{a_1(x_1,x_2)}\int_0^{b_1(x_1,x_2;t_1)} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

$$\mathrm{MP}^{(2)}(x_1,x_2) = (1+(\lambda_2+\lambda_{12})x_2)^{-\alpha_2-\alpha_{12}},$$

When $x_1 \geq x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_0^{a_2(x_1,x_2)}\int_{b_2(x_1,x_2;t_1)}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

When $x_1 < x_2$:

$$\underline{\mathrm{Dir}}^{(2)}(x_1,x_2) = \frac{1}{\beta(\boldsymbol{\alpha})}\int_{c_2(x_1,x_2)}^{a_2(x_1,x_2)}\int_{b_2(x_1,x_2;t_1)}^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$
$$+\frac{1}{\beta(\boldsymbol{\alpha})}\int_0^{c_2(x_1,x_2)}\int_0^{1-t_1} t_1^{\alpha_1-1}t_2^{\alpha_2-1}(1-t_1-t_2)^{\alpha_{12}-1}\mathrm{d}t_2\ \mathrm{d}t_1,$$

s.t. $a_1(x_1,x_2) = \frac{x_2(\lambda_{12}x_1+1)}{x_1((\lambda_2+\lambda_{12})x_2+1)+x_2}$, $a_2(x_1,x_2) = \frac{x_2}{x_1((\lambda_2+\lambda_{12})x_2+1)+x_2}$,
$b_1(x_1,x_2;t_1) = \frac{(\lambda_2x_2+1)(x_1(\lambda_2t_1x_2+\lambda_{12}((t_1-1)x_2+x_1)+1)+(t_1-1)x_2)}{x_1(\lambda_{12}x_1+1)((\lambda_2+\lambda_{12})x_2+1)}$, $b_2(x_1,x_2;t_1) = \frac{(\lambda_2+\lambda_{12})t_1x_2x_1+(t_1-1)x_2+x_1}{x_1}$,
$c_1(x_1,x_2) = \frac{(x_2-x_1)(\lambda_{12}x_1+1)}{x_2((\lambda_2+\lambda_{12})x_1+1)}$ *and* $c_2(x_1,x_2)\frac{x_2-x_1}{x_2((\lambda_2+\lambda_{12})x_1+1)}$.

Case 3: For $\mathcal{M}_v = \{\{1\},\{2\}\}$. *Then the survival functions can be written as:*

$$\overline{F}_{\boldsymbol{X}}(x_1,x_2) = (1+\lambda_1x_1)^{-\alpha_1}(1+\lambda_2x_2)^{-\alpha_2}(1+\lambda_1x_1+\lambda_2x_2)^{-\alpha_{12}},$$

In the previous examples, the survival distribution was absolutely continuous w.r.t. the Lebesgue measure. Since co-monotonicity in two dimensions can only manifest in one form, the next example shows this case.

Example 12. *Set* $\mathcal{M}_v = \{\{1\},\{2\},\{1,2\}\}$ *while* $\mathcal{M}_w = \{\{1,2\}\}$. *We notice that this is the*

only full setting with co-monotonic parts in the bivariate case. Then survival functions can be written as:

$$\overline{F}_{\boldsymbol{X}}(x_1, x_2) = (1 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_{12} \max(x_1, x_2))^{-\alpha_{12}},$$

which is the bivariate Pareto specified in (4.5).

It is possible, as in the case of the bivariate MO, to split the distribution in example 12 into an absolutely continuous and singular parts. The next proposition shows the details of the decomposition.

Proposition 12. *Let $\mathcal{M}_v$ and $\mathcal{M}_w$ be defined as in example 12 then the following decomposition holds:*

$$\overline{F}_{\boldsymbol{X}}(x_1, x_2) = \tau \overline{F}^{\,a}_{\boldsymbol{X}}(x_1, x_2) + (1 - \tau)\overline{F}^{\,s}_{\boldsymbol{X}}(x_1, x_2),$$

s.t.:

$$\begin{aligned}
\overline{F}^{\,a}_{\boldsymbol{X}}(x_1, x_2) &= \frac{\lambda_1 + \lambda_2 + \lambda_{12}}{\lambda_1 + \lambda_2}\, \overline{F}_{\boldsymbol{X}}(x_1, x_2) - \frac{\lambda_{12}}{\lambda_1 + \lambda_2} \left(1 + (\lambda_1 + \lambda_2 + \lambda_{12}) \max(x_1, x_2)\right)^{-\alpha_{12}}, \\
\overline{F}^{\,s}_{\boldsymbol{X}}(x_1, x_2) &= \left(1 + (\lambda_1 + \lambda_2 + \lambda_{12}) \max(x_1, x_2)\right)^{-\alpha_{12}}, \\
\tau &= \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_{12}}
\end{aligned}$$

with $\overline{F}_{\boldsymbol{X}}(x_1, x_2) = (1 + \lambda_1 x_1 + \lambda_2 x_2 + \lambda_{12} \max(x_1, x_2))^{-\alpha_{12}}$, $\tau = \mathbb{P}[X_1 \neq X_2] = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_{12}}$, *and*
$(1 - \tau) = \mathbb{P}[X_1 = X_2] = \frac{\lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}}$ *(see 11). And* $\overline{F}^{\,a}_{\boldsymbol{X}}$ *and* $\overline{F}^{\,s}_{\boldsymbol{X}}$ *are the absolutely continuous and singular parts of the survival function respectively.*

Proof. By differentiation and integration we get $\overline{F}^{\,a}_{\boldsymbol{X}}$, then $\overline{F}^{\,s}_{\boldsymbol{X}} = \overline{F}_{\boldsymbol{X}} - \overline{F}^{\,a}_{\boldsymbol{X}}$ (see Marshall and Olkin (1967)). □

Convolutions play a central role in risk aggregation and allocation. Admittedly, due to its complexity, it is usually not expressible as a closed form. In the special case of example 12, nonetheless, a general expression can be derived, as the next proposition delineates.

Proposition 13. *Let $\mathcal{M}_v$ and $\mathcal{M}_w$ be defined as in example 12 and set $\mu_i = 0$, $\sigma_i = \gamma_i = 1$, $\forall i \in \mathcal{N}$, and $S = X_1 + X_2$. Then the convolution survival function, for $\lambda_1 \neq \lambda_2 + \lambda_{12}$ and $\lambda_2 \neq \lambda_1 + \lambda_{12}$, is expressible as:*

$$\overline{F}_S(s) = \frac{\left((\lambda_1+\lambda_{12})\,s+1\right)^{-\alpha_{12}}\left((\lambda_2+\lambda_{12})\,s+1\right)^{-\alpha_{12}}\left((\lambda_1+\lambda_2+\lambda_{12})\,s+2\right)^{-\alpha_{12}-1}}{(\lambda_1+\lambda_2+\lambda_{12})^{\,2}\left((\lambda_1-\lambda_2)^{\,2}-\lambda_{12}^2\right)}$$

$$\times\Bigg(\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}}+\lambda_1^5 s\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}\left((\lambda_1+\lambda_2+\lambda_{12})\,s+2\right)^{\alpha_{12}}\Big(\lambda_2\left(\lambda_2+\lambda_{12}\right)^3$$
$$\left((\lambda_2+\lambda_{12})\,s+2\right)\left((\lambda_1+\lambda_2+\lambda_{12})\,s+2\right)^{\alpha_{12}}-2^{\alpha_{12}+1}\lambda_{12}\left(\lambda_2+\lambda_{12}\right)\Big(\lambda_{12}^2 s\left(2\alpha_{12}+\lambda_2+2\right)$$
$$+\lambda_2\left(-2\lambda_2\left(\alpha_{12}s+s-1\right)+\lambda_2^2 s-4\right)+2\lambda_{12}\left(\lambda_2+\lambda_2^2 s+2\right)\Big)\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}\Big)+2\lambda_1^2\Big($$
$$-\lambda_{12}\Big(\lambda_{12}\left(-2\alpha_{12}s+5\lambda_2 s-2s+4\right)+2\lambda_2\left(s\left(\alpha_{12}+\lambda_2\right)+s+1\right)+3\lambda_{12}^2 s-4\Big)\left(2\left(\lambda_1+\lambda_{12}\right)s+2\right)^{\alpha_{12}}$$
$$\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}}-\left(\lambda_2+\lambda_{12}\right)\left((\lambda_1+\lambda_2+\lambda_{12})\,s+2\right)^{\alpha_{12}}$$
$$\Big(\left(-2\lambda_{12}^2 s-\lambda_{12}\left(\lambda_2 s+3\right)+\lambda_2\left(\lambda_2 s+1\right)\right)\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}+\lambda_2\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}}\Big)\Big)$$
$$+\lambda_1^4\Big(\left((\lambda_1+\lambda_2+\lambda_{12})\,s+2\right)^{\alpha_{12}}-2^{\alpha_{12}+1}\lambda_{12}s\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}}$$
$$2\left(\lambda_2 s+2\lambda_{12}s+1\right)\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}-\lambda_2 s\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}}\Big)+2\lambda_1^3\Big($$
$$\left((\lambda_1+\lambda_2+\lambda_{12})\,s+2\right)^{\alpha_{12}}-\lambda_{12}\left(-2\alpha_{12}s+2\lambda_2 s+3\lambda_{12}s-2s+2\right)\left(2\left(\lambda_1+\lambda_{12}\right)s+2\right)^{\alpha_{12}}$$
$$\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}}\left(\lambda_2+3\lambda_{12}^2 s+3\lambda_{12}\left(\lambda_2 s+1\right)\right)\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}-\lambda_2\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}+1}\Big)$$
$$+\lambda_1\Big(-2^{\alpha_{12}+1}\lambda_{12}\Big(2\lambda_{12}^2\left(\alpha_{12}s+2\lambda_2 s+s+1\right)+\lambda_2\lambda_{12}\left(4\alpha_{12}s+5\lambda_2 s+4s+4\right)$$
$$+2\lambda_2\left(\lambda_2\left(s\left(\alpha_{12}+\lambda_2\right)+s+1\right)+4\right)+\lambda_{12}^3 s\Big)\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}}$$
$$+\left(\lambda_2+\lambda_{12}\right)^2\left((\lambda_1+\lambda_2+\lambda_{12})\,s+2\right)^{\alpha_{12}}\Big(2\lambda_2\left((\lambda_2+\lambda_{12})\,s+1\right)^{\alpha_{12}+1}$$
$$-\left(\lambda_2-\lambda_{12}\right)\left((\lambda_2+\lambda_{12})\,s+2\right)\left((\lambda_1+\lambda_{12})\,s+1\right)^{\alpha_{12}}\Big)\Big)\Bigg). \quad (4.15)$$

Proof. Integration w.r.t. Z_{12} of the MO convolution (Nadarajah and Kotz, 2005) yields the result. □

Remark 14. *The survival function of the special cases of proposition 13 can be easily derived. Specifically, 1) when one of $\lambda_i = \lambda_j + \lambda_{12}$, $i \neq j \in \mathcal{N}$, holds or 2) when both conditions hold (in which case $\lambda_{12} = 0$ and $\lambda_1 = \lambda_2 = \lambda$).*

1):

$$\overline{F}_S(s) = \frac{1}{2\lambda_{12}\left(\lambda_j+\lambda_{12}\right)^2}\Bigg(\alpha_{12}\lambda_{12}s\left(\lambda_j+\lambda_{12}\right)\left(\left(\lambda_j+\lambda_{12}\right)^2+2\lambda_{12}\right)\left(\left(\lambda_j+\lambda_{12}\right)s+1\right)^{-\alpha_{12}-1}$$
$$+\left(\left(\lambda_j s+2\lambda_{12}s+1\right)\left(\left(\lambda_j+\lambda_{12}\right)s+1\right)\right)^{-\alpha_{12}}\Big(\left(\lambda_j^3+4\lambda_{12}\lambda_j^2+\lambda_{12}^3+2\left(2\lambda_j+1\right)\lambda_{12}^2\right)\left(\lambda_j s+2\lambda_{12}s+1\right)^{\alpha_{12}}$$
$$-\lambda_j\left(\lambda_j+\lambda_{12}\right)^2\left(\left(\lambda_j+\lambda_{12}\right)s+1\right)^{\alpha_{12}}\Big)\Bigg). \quad (4.16)$$

2):

$$\overline{F}_S(s) = \left(1+\lambda s\right)^{-\alpha_{12}-1}\left(1+(1+\alpha_{12})\lambda s\right). \quad (4.17)$$

These examples are not exhaustive by any measure. Expressions of the joint distributions when $\mathcal{M}_w$ and $\mathcal{M}_v$ are strict subsets can be straightforwardly derived leading to a more parsimonious models. In the next propositions, we will turn our attention to calculating the product moments and describing the resulting correlations.

Proposition 14. *Set $\mathcal{M}_w = \mathcal{M}_v = \mathcal{P}(\mathcal{N})$ and let $k_1, k_2 \in \mathbb{N}$, then the k_1, k_2- raw product moment of (X_1, X_2) can be expressed as $\Big($ given $\alpha_1+\alpha_{12} > \frac{k_1}{\gamma_1}$, $\alpha_2+\alpha_{12} > \frac{k_2}{\gamma_2}$ and $\alpha_1+\alpha_2+\alpha_{12} > \frac{k_1}{\gamma_1}+\frac{k_2}{\gamma_2}$ $\Big)$:*

$$\mathbb{E}\left[X_1^{k_1}X_2^{k_2}\right] = \sum_{j_1=0}^{k_1}\sum_{j_2=0}^{k_2}\binom{k_1}{j_1}\binom{k_2}{j_2}\mu_1^{k_1-j_1}\mu_2^{k_2-j_2}\sigma_1^{j_1}\sigma_2^{j_2}\ \Psi\left(j_1,j_2,\gamma_1,\gamma_2,\lambda_1,\lambda_2,\lambda_{12},\alpha_1,\alpha_2,\alpha_{12}\right) \quad (4.18)$$

s.t.:

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{\left(\lambda_1+\lambda_{12}\right)^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_1+\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{\left(\lambda_2+\lambda_{12}\right)^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_2+\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{\left(\lambda_1+\lambda_2+\lambda_{12}\right)^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma(\alpha_1+\alpha_{12})\Gamma(\alpha_2+\alpha_{12})}\Bigg[\frac{j_1}{\gamma_1}\ {}_2F_1\left(1,\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2};\frac{j_2}{\gamma_2}+1;\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right) \\ \qquad +\frac{j_2}{\gamma_2}\ {}_2F_1\left(1,\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2};\frac{j_1}{\gamma_1}+1;\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\Bigg] & \text{when } j_1 \neq 0,\ j_2 \neq 0, \\ \qquad \times\ {}_3F_2\left(\alpha_{12},\frac{j_1}{\gamma_1},\frac{j_2}{\gamma_2};\alpha_1+\alpha_{12},\alpha_2+\alpha_{12};1\right), \end{cases} \quad (4.19)$$

where ${}_2F_1$ and ${}_3F_2$ are hyper-geometric functions (see Gradshteyn and Ryzhik (2014)).

Proof. See appendix iii □

Corollary 5. *The product moment of* $\boldsymbol{X} = (X_1, X_2)$ *can be written as:*

$$\mathbb{E}\left[X_1 X_2\right] = \mu_1\mu_2 + \mu_2\sigma_1 \frac{\Gamma\left(\frac{1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)}{\left(\lambda_1+\lambda_{12}\right)^{\frac{1}{\gamma_1}}\Gamma\left(\alpha_1+\alpha_{12}\right)} + \mu_1\sigma_2 \frac{\Gamma\left(\frac{1}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{\gamma_2}\right)}{\left(\lambda_2+\lambda_{12}\right)^{\frac{1}{\gamma_2}}\Gamma\left(\alpha_2+\alpha_{12}\right)}$$
$$+\sigma_1\sigma_2 \frac{\Gamma\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{\gamma_2}\right)}{\left(\lambda_1+\lambda_2+\lambda_{12}\right)^{\frac{1}{\gamma_1}+\frac{1}{\gamma_2}}\Gamma\left(\alpha_1+\alpha_{12}\right)\Gamma\left(\alpha_2+\alpha_{12}\right)} \left[\frac{1}{\gamma_1}\,{}_2F_1\left(1, \frac{1}{\gamma_1}+\frac{1}{\gamma_2}; \frac{1}{\gamma_2}+1; \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right.$$
$$\left. + \frac{1}{\gamma_2}\,{}_2F_1\left(1, \frac{1}{\gamma_1}+\frac{1}{\gamma_2}; \frac{1}{\gamma_1}+1; \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right] \times\ {}_3F_2\left(\alpha_{12}, \frac{1}{\gamma_1}, \frac{1}{\gamma_2}; \alpha_1+\alpha_{12}, \alpha_2+\alpha_{12}; 1\right). \tag{4.20}$$

s.t. $\alpha_1+\alpha_{12} > \frac{1}{\gamma_1}$, $\alpha_2+\alpha_{12} > \frac{1}{\gamma_2}$ *and* $\alpha_1+\alpha_2+\alpha_{12} > \frac{1}{\gamma_1}+\frac{1}{\gamma_2}$. *Then for* $\alpha_1+\alpha_{12} > \frac{2}{\gamma_1}$ *and* $\alpha_2+\alpha_{12} > \frac{2}{\gamma_2}$ *the Pearson correlation is given as:*

$$\mathrm{Corr}[X_1, X_2] = \frac{\mathrm{Cov}[X_1, X_2]}{\sqrt{\mathrm{Var}[X_1]}\,\sqrt{\mathrm{Var}[X_2]}}, \tag{4.21}$$

where:

$$\mathrm{Cov}[X_1, X_2] = \sigma_1\sigma_2 \frac{\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{\gamma_2}\right)}{\Gamma\left(\alpha_1+\alpha_{12}\right)\Gamma\left(\alpha_2+\alpha_{12}\right)} \left[\frac{\Gamma\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}\right)}{\left(\lambda_1+\lambda_2+\lambda_{12}\right)^{\frac{1}{\gamma_1}+\frac{1}{\gamma_2}}} \left[\frac{1}{\gamma_1}\right.\right.$$
$$\left.\times\,{}_2F_1\left(1, \frac{1}{\gamma_1}+\frac{1}{\gamma_2}; \frac{1}{\gamma_2}+1; \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right) + \frac{1}{\gamma_2}\,{}_2F_1\left(1, \frac{1}{\gamma_1}+\frac{1}{\gamma_2}; \frac{1}{\gamma_1}+1; \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right]$$
$$\left.\times\ {}_3F_2\left(\alpha_{12}, \frac{1}{\gamma_1}, \frac{1}{\gamma_2}; \alpha_1+\alpha_{12}, \alpha_2+\alpha_{12}; 1\right) - \frac{\Gamma\left(\frac{1}{\gamma_1}+1\right)\Gamma\left(\frac{1}{\gamma_2}+1\right)}{\left(\lambda_1+\lambda_{12}\right)^{\frac{1}{\gamma_1}}\left(\lambda_2+\lambda_{12}\right)^{\frac{1}{\gamma_2}}}\right].$$

$$\mathrm{Var}[X_1] = \sigma_1^2 \frac{\Gamma\left(\frac{2}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{2}{\gamma_1}\right)}{\left(\lambda_1+\lambda_{12}\right)^{\frac{2}{\gamma_1}}\Gamma\left(\alpha_1+\alpha_{12}\right)} - \sigma_1 \frac{\Gamma\left(\frac{1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)}{\left(\lambda_1+\lambda_{12}\right)^{\frac{1}{\gamma_1}}\Gamma\left(\alpha_1+\alpha_{12}\right)} \left[\mu_1 + \right.$$
$$\left. \sigma_1 \frac{\Gamma\left(\frac{1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)}{\left(\lambda_1+\lambda_{12}\right)^{\frac{1}{\gamma_1}}\Gamma\left(\alpha_1+\alpha_{12}\right)}\right].$$

$$\mathrm{Var}[X_2] = \sigma_2^2 \frac{\Gamma\left(\frac{2}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{2}{\gamma_2}\right)}{\left(\lambda_2+\lambda_{12}\right)^{\frac{2}{\gamma_2}}\Gamma\left(\alpha_2+\alpha_{12}\right)} - \sigma_2 \frac{\Gamma\left(\frac{1}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{\gamma_2}\right)}{\left(\lambda_2+\lambda_{12}\right)^{\frac{1}{\gamma_2}}\Gamma\left(\alpha_2+\alpha_{12}\right)} \left[\mu_2 + \right.$$

$$\sigma_2 \frac{\Gamma\left(\frac{1}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{\gamma_2}\right)}{\left(\lambda_2+\lambda_{12}\right)^{\frac{1}{\gamma_2}}\Gamma\left(\alpha_2+\alpha_{12}\right)}\Bigg]. \quad (4.22)$$

Example 13. *Set $\mathcal{M}_w$ and $\mathcal{M}_v$ as in example 9 s.t. $\mu_i = 0,\ \sigma_i = 1,\ i = 1,2$. Define $\text{Corr}[X_1, X_2; \frac{1}{2}]$, $\text{Corr}[X_1, X_2; 1]$ and $\text{Corr}[X_1, X_2; 2]$ to be the correlations when $\gamma_1 = \gamma_2 = \frac{1}{2}$, $\gamma_1 = \gamma_2 = 1$ and $\gamma_1 = \gamma_2 = 2$ respectively. Then:*

$$\text{Corr}\left[X_1, X_2; \tfrac{1}{2}\right] = \frac{\left(\sqrt{6\Gamma\left(\alpha_2+\alpha_{12}-4\right)\Gamma\left(\alpha_2+\alpha_{12}\right)-\Gamma\left(\alpha_2+\alpha_{12}-2\right)^2}\right)^{-1}}{\sqrt{6\Gamma\left(\alpha_1+\alpha_{12}-4\right)\Gamma\left(\alpha_1+\alpha_{12}\right)-\Gamma\left(\alpha_1+\alpha_{12}-2\right)^2}}$$
$$\times \Gamma\left(\alpha_1+\alpha_{12}-2\right)\Gamma\left(\alpha_2+\alpha_{12}-2\right)\left(\lambda_1+\lambda_2+\lambda_{12}\right)^{-3}\Big(\left(\left(\lambda_1+\lambda_2\right)^3+10\lambda_{12}^2\left(\lambda_1+\lambda_2\right)+6\lambda_{12}^3\right.$$
$$\left.+\left(5\lambda_1^2+12\lambda_2\lambda_1+5\lambda_2^2\right)\lambda_{12}\right){}_3F_2\left(2,2,\alpha_{12};\alpha_1+\alpha_{12},\alpha_2+\alpha_{12};1\right)-\left(\lambda_1+\lambda_2+\lambda_{12}\right)^3\Big). \quad (4.23)$$

$$\text{Corr}\left[X_1, X_2; 1\right] = \frac{\left(\lambda_1+\lambda_2+2\lambda_{12}\right){}_3F_2\left(1,1,\alpha_{12};\alpha_1+\alpha_{12},\alpha_2+\alpha_{12};1\right)-\left(\lambda_1+\lambda_2+\lambda_{12}\right)}{\sqrt{\frac{\alpha_1+\alpha_{12}}{\alpha_1+\alpha_{12}-2}}\sqrt{\frac{\alpha_2+\alpha_{12}}{\alpha_2+\alpha_{12}-2}}\left(\lambda_1+\lambda_2+\lambda_{12}\right)} \quad (4.24)$$

$$\text{Corr}\left[X_1, X_2; 2\right] = \frac{4\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{2}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{2}\right)}{\sqrt{\frac{4\Gamma(\alpha_1+\alpha_{12})^2}{(\alpha_1+\alpha_{12}-1)(\lambda_1+\lambda_{12})}-\frac{\pi\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{2}\right)^2}{(\lambda_1+\lambda_{12})}}\sqrt{\frac{4\Gamma(\alpha_2+\alpha_{12})^2}{(\alpha_2+\alpha_{12}-1)(\lambda_2+\lambda_{12})}-\frac{\pi\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{2}\right)^2}{(\lambda_2+\lambda_{12})}}}$$
$$\times\left(\frac{1}{2}\left(\frac{\sin^{-1}\left(\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_1+\lambda_2+\lambda_{12}}}\right)}{\sqrt{\lambda_1}\sqrt{\lambda_2+\lambda_{12}}}+\frac{\sin^{-1}\left(\frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1+\lambda_2+\lambda_{12}}}\right)}{\sqrt{\lambda_2}\sqrt{\lambda_1+\lambda_{12}}}\right){}_3F_2\left(\frac{1}{2},\frac{1}{2},\alpha_{12};\alpha_1+\alpha_{12},\alpha_2+\alpha_{12};1\right)\right.$$
$$\left.-\frac{\pi}{4\sqrt{\lambda_1+\lambda_{12}}\sqrt{\lambda_2+\lambda_{12}}}\right). \quad (4.25)$$

Proposition 14 deals with the case of full parameters i.e. when $\mathcal{M}_w = \mathcal{M}_v = \mathcal{P}(\mathcal{N})$. The other possibilities involve either $\mathcal{M}_v \subset \mathcal{P}(\mathcal{N})$ or $\mathcal{M}_w \subset \mathcal{P}(\mathcal{N})$. Simply taking $\lambda_B = 0$ or $\alpha_B = 0$ will not reflect the simplified form. The following two propositions investigate these two limiting cases.

Proposition 15. *Let $\mathcal{M}_v = \mathcal{P}(\mathcal{N})$ and $\mathcal{M}_w \subset \mathcal{P}(\mathcal{N})$ then the k_1, k_2- raw product moment of (X_1, X_2) can be expressed similarly to Equation (4.18) s.t.:*

Case 1) $\mathcal{M}_w = \{\{1\},\{2\}\}$ $\left(\text{ for } \alpha_1 > \frac{k_1}{\gamma_1},\ \alpha_2 > \frac{k_2}{\gamma_2}\ \right)$*:*

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_1-\frac{j_1}{\gamma_1}\right)}{(\lambda_1+\lambda_{12})^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_1)}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_2-\frac{j_2}{\gamma_2}\right)}{(\lambda_2+\lambda_{12})^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_2)}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_1-\frac{j_1}{\gamma_1}\right)\Gamma\left(\alpha_2-\frac{j_2}{\gamma_2}\right)}{(\lambda_1+\lambda_2+\lambda_{12})^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma(\alpha_1)\Gamma(\alpha_2)} \left[\frac{j_1}{\gamma_1}\ {}_2F_1\left(1, \frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}; \frac{j_2}{\gamma_2}+1; \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right. & \\ \left. \quad +\frac{j_2}{\gamma_2}\ {}_2F_1\left(1, \frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}; \frac{j_1}{\gamma_1}+1; \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right] & \text{when } j_1 \neq 0,\ j_2 \neq 0, \end{cases} \tag{4.26}$$

Case 2) $\mathcal{M}_w = \{\{1\},\{1,2\}\}$ $\left(\text{ for } \alpha_{12} > \frac{k_2}{\gamma_2},\ \alpha_1 + \alpha_{12} > \frac{k_1}{\gamma_1} + \frac{k_2}{\gamma_2}\ \right)$*:*

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_1+\lambda_{12})^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_1+\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_2+\lambda_{12})^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}-\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_1+\lambda_2+\lambda_{12})^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_2}{\gamma_2}\right)\Gamma(\alpha_{12})} \left[\frac{j_1}{\gamma_1}\ {}_2F_1\left(1, \frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}; \frac{j_2}{\gamma_2}+1; \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right. & \\ \left. \quad +\frac{j_2}{\gamma_2}\ {}_2F_1\left(1, \frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}; \frac{j_1}{\gamma_1}+1; \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right] & \text{when } j_1 \neq 0,\ j_2 \neq 0, \end{cases} \tag{4.27}$$

Case 3) $\mathcal{M}_w = \{\{2\},\{1,2\}\}$ $\left(\text{ for } \alpha_{12} > \frac{k_1}{\gamma_1},\ \alpha_2+\alpha_{12} > \frac{k_1}{\gamma_1}+\frac{k_2}{\gamma_2}\ \right)$:

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_1+\lambda_{12})^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_2+\lambda_{12})^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_2+\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_1}{\gamma_1}-\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_1+\lambda_2+\lambda_{12})^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_1}{\gamma_1}\right)\Gamma(\alpha_{12})}\left[\frac{j_1}{\gamma_1}\ {}_2F_1\left(1,\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2};\frac{j_2}{\gamma_2}+1;\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right. & \\ \qquad \left. +\frac{j_2}{\gamma_2}\ {}_2F_1\left(1,\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2};\frac{j_1}{\gamma_1}+1;\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right] & \text{when } j_1 \neq 0,\ j_2 \neq 0, \end{cases} \tag{4.28}$$

Case 4) $\mathcal{M}_w = \{\{1,2\}\}$ $\left(\text{ for } \alpha_{12} > \frac{k_1}{\gamma_1}+\frac{k_2}{\gamma_2}\ \right)$:

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_1+\lambda_{12})^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_2+\lambda_{12})^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_{12}-\frac{j_1}{\gamma_1}-\frac{j_2}{\gamma_2}\right)}{(\lambda_1+\lambda_2+\lambda_{12})^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma(\alpha_{12})}\left[\frac{j_1}{\gamma_1}\ {}_2F_1\left(1,\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2};\frac{j_2}{\gamma_2}+1;\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right. & \\ \qquad \left. +\frac{j_2}{\gamma_2}\ {}_2F_1\left(1,\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2};\frac{j_1}{\gamma_1}+1;\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right] & \text{when } j_1 \neq 0,\ j_2 \neq 0, \end{cases} \tag{4.29}$$

Proof. Similar to proof iii of proposition 14. □

Proposition 16. *Let* $\mathcal{M}_w = \mathcal{P}(\mathcal{N})$ *and* $\mathcal{M}_v \subset \mathcal{P}(\mathcal{N})$ *then the* k_1, k_2*- raw product moment of* (X_1, X_2) *can be expressed similarly to Equation* (4.18) *s.t.* $\left(\text{ given } \alpha_1+\alpha_{12} > \frac{k_1}{\gamma_1},\ \alpha_2+\alpha_{12} > \frac{k_2}{\gamma_2}\right.$ *and* $\left.\alpha_1+\alpha_2+\alpha_{12} > \frac{k_1}{\gamma_1}+\frac{k_2}{\gamma_2}\ \right)$:

Case 1) $\mathcal{M}_v = \{\{1\},\{2\}\}$*:*

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_1)^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_1+\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_2)^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_2+\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_1)^{\frac{j_1}{\gamma_1}}(\lambda_2)^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_1+\alpha_{12})\Gamma(\alpha_2+\alpha_{12})} & \text{when } j_1 \neq 0,\ j_2 \neq 0 \\ \quad \times {}_3F_2\left(\alpha_{12}, \frac{j_1}{\gamma_1}, \frac{j_2}{\gamma_2}; \alpha_1+\alpha_{12}, \alpha_2+\alpha_{12}; 1\right), & \end{cases} \tag{4.30}$$

Case 2) $\mathcal{M}_v = \{\{1\},\{1,2\}\}$*:*

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_1+\lambda_{12})^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_1+\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_{12})^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_2+\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_1+\lambda_{12})^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma(\alpha_1+\alpha_{12})\Gamma(\alpha_2+\alpha_{12})} \left[\frac{j_1}{\gamma_1}\right. & \\ \quad \left. + \frac{j_2}{\gamma_2}\ {}_2F_1\left(1, \frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}; \frac{j_1}{\gamma_1}+1; \frac{\lambda_1}{\lambda_1+\lambda_{12}}\right)\right] & \text{when } j_1 \neq 0,\ j_2 \neq 0, \\ \quad \times\ {}_3F_2\left(\alpha_{12}, \frac{j_1}{\gamma_1}, \frac{j_2}{\gamma_2}; \alpha_1+\alpha_{12}, \alpha_2+\alpha_{12}; 1\right), & \end{cases} \tag{4.31}$$

Case 3) $\mathcal{M}_v = \{\{2\},\{1,2\}\}$*:*

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_{12})^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_1+\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_2+\lambda_{12})^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_2+\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_2+\lambda_{12})^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma(\alpha_1+\alpha_{12})\Gamma(\alpha_2+\alpha_{12})} \left[\frac{j_2}{\gamma_2}\right. & \\ \quad \left. + \frac{j_1}{\gamma_1}\ {}_2F_1\left(1, \frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}; \frac{j_2}{\gamma_2}+1; \frac{\lambda_2}{\lambda_2+\lambda_{12}}\right)\right] & \text{when } j_1 \neq 0,\ j_2 \neq 0, \\ \quad \times\ {}_3F_2\left(\alpha_{12}, \frac{j_1}{\gamma_1}, \frac{j_2}{\gamma_2}; \alpha_1+\alpha_{12}, \alpha_2+\alpha_{12}; 1\right), & \end{cases} \tag{4.32}$$

Case 4) $\mathcal{M}_v = \{\{1,2\}\}$*:*

$$\Psi = \begin{cases} 1, & \text{when } j_1 = j_2 = 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+1\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)}{(\lambda_{12})^{\frac{j_1}{\gamma_1}}\Gamma(\alpha_1+\alpha_{12})}, & \text{when } j_1 \neq 0,\ j_2 = 0, \\ \frac{\Gamma\left(\frac{j_2}{\gamma_2}+1\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_{12})^{\frac{j_2}{\gamma_2}}\Gamma(\alpha_2+\alpha_{12})}, & \text{when } j_1 = 0,\ j_2 \neq 0, \\ \frac{\Gamma\left(\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}\right)\Gamma\left(\alpha_1+\alpha_{12}-\frac{j_1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{j_2}{\gamma_2}\right)}{(\lambda_{12})^{\frac{j_1}{\gamma_1}+\frac{j_2}{\gamma_2}}\Gamma(\alpha_1+\alpha_{12})\Gamma(\alpha_2+\alpha_{12})} \left[\frac{j_1}{\gamma_1}\right. & \text{when } j_1 \neq 0,\ j_2 \neq 0, \\ \left. \quad +\frac{j_2}{\gamma_2}\right] {}_3F_2\left(\alpha_{12}, \frac{j_1}{\gamma_1}, \frac{j_2}{\gamma_2}; \alpha_1+\alpha_{12}, \alpha_2+\alpha_{12}; 1\right). & \end{cases} \tag{4.33}$$

Proof. Similar to that of proposition 15. □

Remark 15. *Similar to Corollary 5 we can derive the product moments* ($k_1 = k_2 = 1$) *and consequently the correlations for the different cases of propositions 15 & 16.*

Remark 16. *When* $\mathcal{M}_w \subset \mathcal{P}(\mathcal{N})$ *and* $\mathcal{M}_v \subset \mathcal{P}(\mathcal{N})$ *i.e. both are strict subsets, then their* k_1, k_2*-raw product moments are a combination of those in propositions 15 and 16. This works since* $\boldsymbol{V}$ *and* $\boldsymbol{W}$ *are independent.*

Example 14. *Set* $\mu_i = 0,\ \sigma_i = 1, \gamma_i = 1,\ i = 1, 2$*. Then the correlations for examples 10, 11 and 12 are (given* $\alpha_1 > 2,\ \alpha_2 > 2,\ \alpha_1 + \alpha_{12} > 2$ *and/or* $\alpha_2 + \alpha_{12} > 2$*):*

Example 10:

$$\text{Case 1}: \text{Corr}[X_1X_2] = \frac{\lambda_{12}}{(\lambda_1+\lambda_2+\lambda_{12})}\sqrt{\frac{(\alpha_1-2)(\alpha_2-2)}{\alpha_1\alpha_2}}.$$

$$\text{Case 2}: \text{Corr}[X_1X_2] = \frac{\lambda_1+\lambda_2+(\alpha_1+\alpha_{12})\lambda_{12}}{(\alpha_1+\alpha_{12})(\lambda_1+\lambda_2+\lambda_{12})}\sqrt{\frac{(\alpha_1+\alpha_{12})(\alpha_{12}-2)}{\alpha_{12}(\alpha_1+\alpha_{12}-2)}}.$$

$$\text{Case 3}: \text{Corr}[X_1X_2] = \frac{\lambda_1+\lambda_2+(\alpha_2+\alpha_{12})\lambda_{12}}{(\alpha_2+\alpha_{12})(\lambda_1+\lambda_2+\lambda_{12})}\sqrt{\frac{(\alpha_2+\alpha_{12})(\alpha_{12}-2)}{\alpha_{12}(\alpha_2+\alpha_{12}-2)}}.$$

Example 11:

$$\text{Case 1}: \text{Corr}[X_1X_2] = \left({}_3F_2\left(1,1,\alpha_{12};\alpha_1+\alpha_{12},\alpha_2+\alpha_{12};1\right)-1\right)\sqrt{\frac{(\alpha_1+\alpha_{12}-2)(\alpha_2+\alpha_{12}-2)}{(\alpha_1+\alpha_{12})(\alpha_2+\alpha_{12})}}$$

$$\text{Case 2}: \text{Corr}[X_1X_2] = \left((\lambda_1 + 2\lambda_{12})\ {}_3F_2\left(1,1,\alpha_{12};\alpha_1+\alpha_{12},\alpha_2+\alpha_{12};1\right) - (\lambda_1+\lambda_{12})\right) \times \sqrt{\frac{(\alpha_1+\alpha_{12}-2)(\alpha_2+\alpha_{12}-2)}{(\alpha_1+\alpha_{12})(\alpha_2+\alpha_{12})}}$$

$$\text{Case 3}: \text{Corr}[X_1X_2] = \left((\lambda_2 + 2\lambda_{12})\ {}_3F_2\left(1,1,\alpha_{12};\alpha_1+\alpha_{12},\alpha_2+\alpha_{12};1\right) - (\lambda_2+\lambda_{12})\right) \times \sqrt{\frac{(\alpha_1+\alpha_{12}-2)(\alpha_2+\alpha_{12}-2)}{(\alpha_1+\alpha_{12})(\alpha_2+\alpha_{12})}}$$

Example 12:

$$\text{Corr}[X_1X_2] = \frac{\lambda_1+\lambda_2+\alpha_{12}\lambda_{12}}{\alpha_{12}(\lambda_1+\lambda_2+\lambda_{12})}.$$

Granted that the Pareto-Dirichlet extension suffers from the usual non-negative correlations. However, in its generality, it reflected assorted forms capable of grasping different modelling choices. We conclude our study with an application to the pricing functionals, calculating the different risk measures and allocations when the joint law is of the Pareto-Dirichlet type.

4.4 Application in premium pricing functionals

Risk measures are of utmost importance for risk managers, especially in insurance. They are used to gauge the riskiness of the financial entity determining its economic capital requirements. Two of the most prominent measures are the Value-at-Risk (VaR) (Linsmeier and Pearson, 2000) and the Conditional Tail Expectation (CTE) (Acerbi and Tasche, 2002; Tasche, 2002; Wang and Zitikis, 2021; Yamai and Yoshiba, 2005) with the latter superseding the former.

Definition 8. *For a prudence level $q \in [0,1)$ the VaR and CTE are defined as:*

$$\text{VaR}_q[Y] = \inf\{y : F_Y(y) \geq q\}, \tag{4.34}$$

$$\mathrm{CTE}_q[Y] = \frac{1}{1-q}\int_q^1 \mathrm{VaR}_t[Y]\ \mathrm{d}t, \tag{4.35}$$

When the CDF F_Y of Y is continuous, then $\mathrm{CTE}_q[Y] = \mathbb{E}\left[Y|Y > \mathrm{VaR}_q[Y]\right]$.

Proposition 17. *Since $X_i \sim$ Pareto IV then the two measures for standalone losses can be written as:*

$$\mathrm{VaR}_q[X_i] = \mu_i + \sigma_i\left(\frac{(1-q)^{-\frac{1}{\tilde{\alpha}_i}} - 1}{\tilde{\lambda}_i}\right)^{\frac{1}{\gamma_i}}. \tag{4.36}$$

$$\mathrm{CTE}_q[X_i] = \mu_i + \frac{\sigma_i\tilde{\alpha}_i}{\tilde{\lambda}_i^{\frac{1}{\gamma_i}}(1-q)}\,\beta_{(1-q)^{\frac{1}{\tilde{\alpha}_i}}}\left(\tilde{\alpha}_i - \frac{1}{\gamma_i}, \frac{1}{\gamma_i}+1\right),\ \tilde{\alpha}_i > \frac{1}{\gamma_i}. \tag{4.37}$$

Where $\beta_z(p,q)$ is the regularized incomplete beta function.

Proof. Straightforward evaluations. □

Corollary 6. *When $\mu_i = 0$ and $\sigma_i = \gamma_i = 1$ then the two measures simplify to:*

$$\mathrm{VaR}_q[X_i] = \frac{(1-q)^{-\frac{1}{\tilde{\alpha}_i}} - 1}{\tilde{\lambda}_i}, \tag{4.38}$$

$$\mathrm{CTE}_q[X_i] = \frac{1}{\tilde{\alpha}_i - 1}\left[\tilde{\alpha}_i\mathrm{VaR}_q[X_i] + \frac{1}{\tilde{\lambda}_i}\right],\ \tilde{\alpha}_i > 1. \tag{4.39}$$

Albeit the forms of VaR and CTE are compact for each X_i, risk measures usually deal with aggregates of losses. Regrettably, under the Pareto-Dirichlet structure, the convolution $S = \sum_{j=1}^n X_j$ is quite cumbersome to calculate, let alone computing the respective risk measures. However, in some particular cases, such as the one in example 12, the CTE measure can be discerned as a function of VaR.

Proposition 18. *Let $\boldsymbol{X} = (X_1, X_2)$ be Pareto-Dirichlet, s.t. $\mathcal{M}_w$, $\mathcal{M}_v$ and the univariate transformations as in proposition 13. Additionally, set $\lambda_1 = \lambda_2 = \lambda$, $\lambda_{12} \neq 0$, and $\alpha_{12} > 1$. Then the CTE risk measure of the aggregate $S = X_1 + X_2$, in terms of VaR denoted by $s_q = \mathrm{VaR}_q[S]$ for a prudence level $q \in [0,1)$, is given by:*

$$\mathrm{CTE}_q[S] = s_q + \frac{2}{(1-q)(\alpha_{12}-1)\lambda_{12}(2\lambda+\lambda_{12})^3}\Big(-\lambda(2\lambda+\lambda_{12})^3$$

$$\times(\lambda+\lambda_{12})^{-\alpha_{12}-1}(\lambda s_q+\lambda_{12}s_q+1)\left(\frac{1}{\lambda+\lambda_{12}}+s_q\right)^{-\alpha_{12}}+2\left(\lambda+\frac{\lambda_{12}}{2}\right)^{-\alpha_{12}}\left(\frac{2}{2\lambda+\lambda_{12}}+s_q\right)^{-\alpha_{12}}$$
$$\lambda_{12}\left(\lambda_{12}\left(2\alpha_{12}\lambda s_q+\lambda_{12}\left(\alpha_{12}+\lambda+1\right)s_q+\lambda\left((5\lambda+2)s_q+2\right)+4\right)+2\lambda^2\left(4\lambda s_q+3\right)\right)+4\lambda^3\left(\lambda s_q+1\right)\Bigg). \tag{4.40}$$

Proof. Integration of the survival function in proposition 13 yields the result. □

Corollary 7. *The CTE of the convolution for the special cases in remark 14 can be straightforwardly obtained. Given $\alpha>1$ and $q\in[0,1)$:*

1) When one of $\lambda_i=\lambda_j+\lambda_{12},\ i\neq j\in\mathcal{N}$ holds, then:

$$\begin{aligned}\mathrm{CTE}_q[S]=s_q+\frac{1}{2(1-q)\left(\alpha_{12}-1\right)\lambda_{12}\left(\lambda_j+\lambda_{12}\right)^3}\Bigg(&\lambda_{12}\left(\left(\lambda_j+\lambda_{12}\right)^2+2\lambda_{12}\right)\left(\left(\lambda_j+\lambda_{12}\right)s_q+1\right)^{-\alpha_{12}}\\ &\times\left(\alpha_{12}\left(\lambda_j+\lambda_{12}\right)s_q+1\right)+\left(\lambda_{12}^2\left(4\lambda_j+2\right)+4\lambda_{12}\lambda_j^2+\lambda_j^3+\lambda_{12}^3\right)\left(\left(\lambda_j+\lambda_{12}\right)s_q+1\right)^{1-\alpha_{12}}\\ &-\frac{\lambda_j\left(\lambda_j+\lambda_{12}\right)^3\left(\left(\lambda_j+2\lambda_{12}\right)s_q+1\right)^{1-\alpha_{12}}}{\lambda_j+2\lambda_{12}}\Bigg).\end{aligned} \tag{4.41}$$

2) When both conditions hold, then:

$$\mathrm{CTE}_q[S]=s_q+\frac{\left(\lambda s_q+1\right)^{-\alpha_{12}}\left(\left(\alpha_{12}+1\right)\lambda s_q+2\right)}{\left(1-q\right)\left(\alpha_{12}-1\right)\lambda}. \tag{4.42}$$

For similar reasons, inverting the survival function of S in proposition 13 is not attainable. Thus the VaR can only be deduced numerically. Nonetheless, a compact form is obtainable for some choices of parameters. For example, let $\alpha_{12}=1$ then the VaR of (4.17), as a function of q, is straightforwardly derived as $\mathrm{VaR}_q[S]=\frac{1}{\lambda}\frac{\sqrt{q}}{1-\sqrt{q}}$.

On the opposite direction of risk aggregation, when the goal is to gauge the riskiness of distinct business units or conduct marginal economic capital analysis, then we revert to risk capital allocations (CAs) (Denault, 2001; Dhaene et al., 2012; Furman and Zitikis, 2008b). CAs are an important tool in quantitative risk management, as they capture the performance of the individual with respect to the others. As a consequence, interdependence plays a crucial role in determining their outcome. Henceforth, we consider two main allocations, the

first gauges the riskiness of one BU to another, also called a regression CA, and the second, induced by the CTE measure, which deals with the relationship of the respective BU to the whole (aggregate).

Definition 9. *A regression CA (RCA) is defined as:*

$$A_{i_1}(X_{i_1}, X_{i_2}) = \mathbb{E}\left[X_{i_1} | X_{i_2} > y\right], \ y \geq 0. \tag{4.43}$$

Proposition 19. *Let* $\boldsymbol{X} = (X_1, X_2)$ *be Pareto-Dirichlet s.t.* $\mathcal{M}_w = \mathcal{M}_v = \mathcal{P}(\mathcal{N})$. *Furthermore set* $\mu_i = 0$ *and* $\sigma_i = \gamma_i = 1, \ \forall i \in \mathcal{N}$. *Then for* $y \geq 0$*:*

$$\begin{aligned}\mathbb{E}\left[X_{i_1} | X_{i_2} > y\right] = \frac{(1 + (\lambda_{i_2} + \lambda_{12})y\ )^{\alpha_{i_2}+\alpha_{12}}}{\lambda_{i_1}(\lambda_{i_1} + \lambda_{12})(\alpha_{i_1} + \alpha_{12} - 1)}\Big[(\lambda_{i_1} + \lambda_{12})\left(1 + (\lambda_{i_2} + \lambda_{12})y\ \right)^{-\alpha_{i_2}} \\ \times {}_2F_1\left(\alpha_{12}, \alpha_{i_1} + \alpha_{12} - 1; \alpha_{i_1} + \alpha_{12}; -(\lambda_{i_2} + \lambda_{12})y\right) - \lambda_{12}\left(1 + (\lambda_{i_1} + \lambda_{i_2} + \lambda_{12})y\ \right)^{-\alpha_{i_2}} \\ \times {}_2F_1\left(\alpha_{12}, \alpha_{i_1} + \alpha_{12} - 1; \alpha_{i_1} + \alpha_{12}; -(\lambda_{i_1} + \lambda_{i_2} + \lambda_{12})y\right)\Big],\end{aligned} \tag{4.44}$$

given $\alpha_{i_1} + \alpha_{12} > 1$.

Proof. See appendix iv. □

Following the setting of proposition 19, the next assertions discuss the limiting cases of $\mathcal{M}_w \subset \mathcal{P}(\mathcal{N})$ and $\mathcal{M}_v \subset \mathcal{P}(\mathcal{N})$. Their proofs are similar to that of iv.

Proposition 20. *If* $\mathcal{M}_w \subset \mathcal{P}(\mathcal{N})$ *and* $\mathcal{M}_v = \mathcal{P}(\mathcal{N})$*, then:*

Case 1: $\mathcal{M}_w = \{\{i_1\}, \{i_2\}\}$ *(given* $\alpha_{i_1} > 1$ *) :*

$$\begin{aligned}\mathbb{E}\left[X_{i_1} | X_{i_2} > y\right] = \frac{(1 + (\lambda_{i_2} + \lambda_{12})y\ )^{\alpha_{i_2}}}{\lambda_{i_1}(\lambda_{i_1} + \lambda_{12})(\alpha_{i_1} - 1)}\Big[(\lambda_{i_1} + \lambda_{12})\left(1 + (\lambda_{i_2} + \lambda_{12})y\ \right)^{-\alpha_{i_2}} \\ - \lambda_{12}\left(1 + (\lambda_{i_1} + \lambda_{i_2} + \lambda_{12})y\ \right)^{-\alpha_{i_2}}\Big],\end{aligned} \tag{4.45}$$

Case 2: $\mathcal{M}_w = \{\{i_1\}, \{1, 2\}\}$ *(given* $\alpha_{i_1} + \alpha_{12} > 1$ *) :*

$$\mathbb{E}\left[X_{i_1} | X_{i_2} > y\right] =$$

$$\frac{(1+(\lambda_{i_2}+\lambda_{12})y\ )^{\alpha_{12}}}{\lambda_{i_1}(\lambda_{i_1}+\lambda_{12})(\alpha_{i_1}+\alpha_{12}-1)}\Big[(\lambda_{i_1}+\lambda_{12})\ {}_2F_1\left(\alpha_{12},\alpha_{i_1}+\alpha_{12}-1;\alpha_{i_1}+\alpha_{12};-(\lambda_{i_2}+\lambda_{12})y\right)$$
$$-\lambda_{12}\ {}_2F_1\left(\alpha_{12},\alpha_{i_1}+\alpha_{12}-1;\alpha_{i_1}+\alpha_{12};-(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12})y\right)\Big]. \quad (4.46)$$

Case 3: $\mathcal{M}_w=\{\{i_2\},\{1,2\}\}$ *(given* $\alpha_{12}>1$*):*

$$\mathbb{E}\left[X_{i_1}|X_{i_2}>y\right]=\frac{(1+(\lambda_{i_2}+\lambda_{12})y\ )^{\alpha_{i_2}+\alpha_{12}}}{\lambda_{i_1}(\lambda_{i_1}+\lambda_{12})(\alpha_{12}-1)}\Big[(\lambda_{i_1}+\lambda_{12})\left(1+(\lambda_{i_2}+\lambda_{12})y\ \right)^{1-\alpha_{i_2}-\alpha_{12}}$$
$$-\lambda_{12}\left(1+(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12})y\ \right)^{1-\alpha_{i_2}-\alpha_{12}}\Big], \quad (4.47)$$

Case 4: $\mathcal{M}_w=\{\{1,2\}\}$ *(given* $\alpha_{12}>1$*):*

$$\mathbb{E}\left[X_{i_1}|X_{i_2}>y\right]=\frac{(1+(\lambda_{i_2}+\lambda_{12})y\ )^{\alpha_{12}}}{\lambda_{i_1}(\lambda_{i_1}+\lambda_{12})(\alpha_{12}-1)}\Big[(\lambda_{i_1}+\lambda_{12})\left(1+(\lambda_{i_2}+\lambda_{12})y\ \right)^{1-\alpha_{12}}$$
$$-\lambda_{12}\left(1+(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12})y\ \right)^{1-\alpha_{12}}\Big], \quad (4.48)$$

Proposition 21. *If* $\mathcal{M}_v\subset\mathcal{P}(\mathcal{N})$ *and* $\mathcal{M}_w=\mathcal{P}(\mathcal{N})$*, then (given* $\alpha_{i_1}+\alpha_{12}>1$ *):*

Case 1: $\mathcal{M}_v=\{\{i_1\},\{i_2\}\}$*:*

$$\mathbb{E}\left[X_{i_1}|X_{i_2}>y\right]=\frac{(1+\lambda_{i_2}y\ )^{\alpha_{12}}}{\lambda_{i_1}(\alpha_{i_1}+\alpha_{12}-1)}\ {}_2F_1\left(\alpha_{12},\alpha_{i_1}+\alpha_{12}-1;\alpha_{i_1}+\alpha_{12};-\lambda_{i_2}y\right). \quad (4.49)$$

Case 2: $\mathcal{M}_v=\{\{i_1\},\{1,2\}\}$*:*

$$\mathbb{E}\left[X_{i_1}|X_{i_2}>y\right]=\frac{(1+\lambda_{12}y\ )^{\alpha_{i_2}+\alpha_{12}}}{\lambda_{i_1}(\lambda_{i_1}+\lambda_{12})(\alpha_{i_1}+\alpha_{12}-1)}\Big[(\lambda_{i_1}+\lambda_{12})\left(1+\lambda_{12}y\ \right)^{-\alpha_{i_2}}$$
$$\times\ {}_2F_1\left(\alpha_{12},\alpha_{i_1}+\alpha_{12}-1;\alpha_{i_1}+\alpha_{12};-\lambda_{12}y\right)-\lambda_{12}\left(1+(\lambda_{i_1}+\lambda_{12})y\ \right)^{-\alpha_{i_2}}$$
$$\times\ {}_2F_1\left(\alpha_{12},\alpha_{i_1}+\alpha_{12}-1;\alpha_{i_1}+\alpha_{12};-(\lambda_{i_1}+\lambda_{12})y\right)\Big]. \quad (4.50)$$

Case 3: $\mathcal{M}_v=\{\{i_2\},\{1,2\}\}$*:*

$$\mathbb{E}\left[X_{i_1}|X_{i_2}>y\right]=\frac{\left(1+(\lambda_{i_2}+\lambda_{12})y\ \right)^{\alpha_{12}-1}}{\lambda_{12}}\Big[\frac{\left(1+(\lambda_{i_2}+\lambda_{12}(1+\alpha_2))y\ \right)}{(\alpha_{i_1}+\alpha_{12}-1)}$$
$$\times\ {}_2F_1\left(\alpha_{12},\alpha_{i_1}+\alpha_{12}-1;\alpha_{i_1}+\alpha_{12};-(\lambda_{i_2}+\lambda_{12})y\right)+\frac{\alpha_{12}\lambda_{12}\left(1+(\lambda_{i_2}+\lambda_{12})y\ \right)\ y}{\alpha_{i_1}+\alpha_{12}}$$
$$\times\ {}_2F_1\left(\alpha_{12}+1,\alpha_{i_1}+\alpha_{12};\alpha_{i_1}+\alpha_{12}+1;-(\lambda_{i_2}+\lambda_{12})y\right)\Big],\quad(4.51)$$

Case 4: $\mathcal{M}_v=\{\{1,2\}\}$*:*

$$\mathbb{E}\left[X_{i_1}|X_{i_2}>y\right]=\frac{\left(1+\lambda_{12}y\ \right)^{\alpha_{12}-1}}{\lambda_{12}}\Big[\frac{\left(1+\lambda_{12}(1+\alpha_2)y\ \right)}{(\alpha_{i_1}+\alpha_{12}-1)}$$
$$\times\ {}_2F_1\left(\alpha_{12},\alpha_{i_1}+\alpha_{12}-1;\alpha_{i_1}+\alpha_{12};-\lambda_{12}y\right)+\frac{\alpha_{12}\lambda_{12}\left(1+\lambda_{12}y\ \right)\ y}{\alpha_{i_1}+\alpha_{12}}$$
$$\times\ {}_2F_1\left(\alpha_{12}+1,\alpha_{i_1}+\alpha_{12};\alpha_{i_1}+\alpha_{12}+1;-\lambda_{12}y\right)\Big],\quad(4.52)$$

Parallel to remark 16 when $\mathcal{M}_w\subset\mathcal{P}(\mathcal{N})$ and $\mathcal{M}_v\subset\mathcal{P}(\mathcal{N})$, then their RCA regressions are a combination of those in propositions 20 and 21. The next corollary show how to retrieve the regression functions from their RCAs counterparts.

Remark 17. *We can easily deduce the regression function from the results of propositions 19, 20 and 21, simply by differentiation. Namely:*

$$\mathbb{E}[X_{i_1}|X_{i_2}=y]=\frac{\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}\left[X_{i_1}\mathbf{1}_{\{X_{i_2}>y\}}\right]}{\frac{\mathrm{d}}{\mathrm{d}y}\ \mathbb{P}[X_{i_2}>y]}=-\frac{1}{f_{X_{i_2}}(y)}\ \frac{\mathrm{d}}{\mathrm{d}y}\left(\mathbb{P}[X_{i_2}>y]\times\mathbb{E}[X_{i_1}|X_{i_2}>y]\right).\quad(4.53)$$

Example 15. *Set $\mathcal{M}_w$ and $\mathcal{M}_v$ as in example 12 (case 4 of proposition 20), then the regression function is (given $\alpha_{12}>1$):*

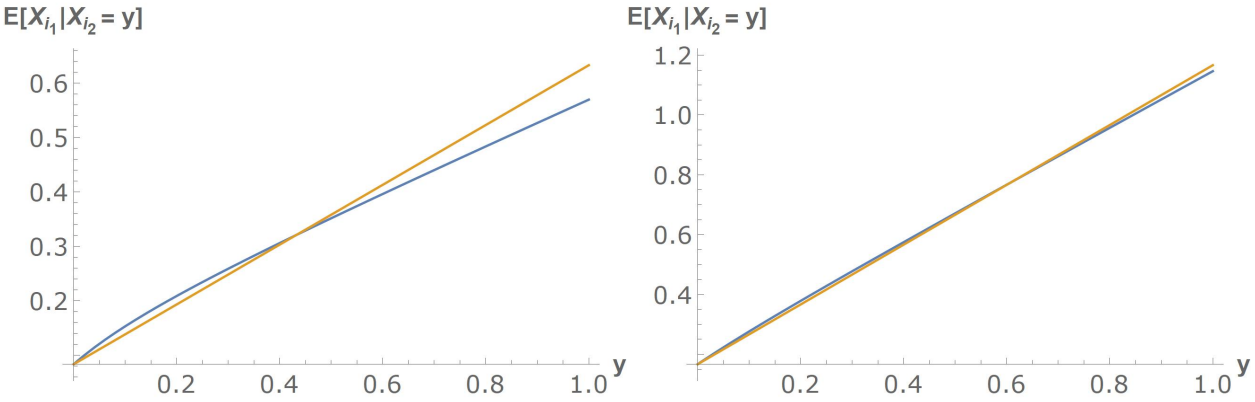
$$\mathbb{E}[X_{i_1}|X_{i_2}=y]=\frac{1}{\alpha_{12}\lambda_{i_1}\left(\lambda_{i_1}+\lambda_{12}\right)\left(\lambda_{i_2}+\lambda_{12}\right)}\Big(\left(\lambda_{i_1}+\lambda_{12}\right)\left(\lambda_{i_2}+\lambda_{12}\right)\left(1+\left(\lambda_{i_2}+\lambda_{12}\right)y\right)$$
$$-\lambda_{12}\left(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12}\right)\left(1+\left(\lambda_{i_2}+\lambda_{12}\right)y\right)^{\alpha_{12}+1}\left(1+\left(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12}\right)y\right)^{-\alpha_{12}}\Big).\quad(4.54)$$

We notice that $\mathbb{E}[X_{i_1}|X_{i_2}=y]$ is increasing for $y\geq 0$, and when $\lambda_{i_2}+\lambda_{12}>>\lambda_{i_1}$ then

$\mathbb{E}[X_1|X_2 = y]$ *behaves almost linearly, specifically:*

$$\mathbb{E}[X_{i_1}|X_{i_2} = y] \approx \frac{\lambda_{i_2}\left(1 + \left(\lambda_{i_2} + \lambda_{12}\right) y\right)}{\alpha_{12}\left(\lambda_{i_1} + \lambda_{12}\right)\left(\lambda_{i_2} + \lambda_{12}\right)}. \tag{4.55}$$

Similar conclusions can be made for the other cases of proposition 20. The following figure show two plots of the regression function when a) normal b) almost linear choice of parameters. The blue curves represent the regression functions, each compared with a straight line in orange.



(a) $\alpha_{12} = \lambda_{i_1} = 2$ and $\lambda_{i_2} = \lambda_{12} = 1$ (b) $\alpha_{12} = \lambda_{i_2} = 2$ and $\lambda_{i_1} = \lambda_{12} = 1$

We notice that in 4.2a the regression function is concave while in 4.2b is almost linear compared to the respective straight lines.

The second, and the most prominent CA used today, is the one induced by the CTE risk measure (see Mohammed et al. (2021)). The CTE allocation quantifies the sensitivity of the part with respect to the aggregate S.

Definition 10. *A CTE capital allocation (CCA) is defined as:*

$$A_i(X_i, S; q) = \mathbb{E}\left[X_i | S > s_q\right], \ s_q \geq 0. \tag{4.56}$$

Where $S = \sum_{j=1}^{n} X_j$ *and* $s_q = \mathrm{VaR}_q[S]$, $q \in [0, 1)$.

Due to the complexity and intractability of the aggregate distribution, computing (4.56)

and s_q (as the VaR of equation (4.15) illustrates) is generally confined to numerical evaluations. There are, however, some special cases where closed forms for (4.56) in terms of s_q are possible. The next remark and subsequent example provide an illustration.

Remark 18. *If $\boldsymbol{X} = (X_1, X_2)$ is Pareto-Dirichlet, s.t. $\mu_i = 0,\ \sigma_i = \gamma_i = 1,\ \forall i \in \mathcal{N}$, with $\mathcal{M}_w = \{\{1,2\}\}$ and $\mathcal{M}_v = \{\{1\},\{2\},\{1,2\}\}$ then the CCA allocation is already quite complex. For any $n > 2$, the CCA allocation is very cumbersome and involves integrations of exponentials along different domains. Since those domains depend on $Z_{\mathcal{N}}$ as well (which is integrated over in order to obtain some tractable form) representing the CCA as a general expression is almost impossible. However when $n = 2$, and under simplified settings, such as if $\lambda_1 = \lambda_2 = \lambda_{12} = \lambda$, then a closed form is obtainable. Specifically:*

$$\mathbb{E}[X_i|S > s_q] = \frac{\left(\left(\frac{3\lambda s_q}{2}+1\right)(2\lambda s_q+1)\right)^{-\alpha}\left((2\lambda s_q+1)^{\alpha}(3\alpha\lambda s_q+2) - \left(\frac{3\lambda s_q}{2}+1\right)^{\alpha}(2\alpha\lambda s_q+1)\right)}{2(1-q)(\alpha-1)\lambda}. \tag{4.57}$$

The following example, under different simplified assumptions, provides additional illustrations.

Example 16. *Let $\boldsymbol{X} = (X_1, X_2)$ be Pareto-Dirichlet, s.t. $\mu_i = 0,\ \sigma_i = \gamma_i = 1,\ \forall i \in \mathcal{N}$. Set $\mathcal{M}_w = \{\{1,2\}\}$ with $\alpha_{12} > 1$ then for the different choices of $\mathcal{M}_v$, the CCA allocation can be written, in terms of the VaR s_q, as:*

Case 1: $\mathcal{M}_v = \{\{i_1\},\{i_2\}\}$:

$$\begin{aligned}&\mathbb{E}\left[X_{i_1}|S > s_q\right] = \\ &\frac{\left((1+\lambda_{i_1}s_q)(1+\lambda_{i_2}s_q)\right)^{-\alpha_{12}}}{(1-q)(\alpha_{12}-1)\lambda_{i_1}(\lambda_{i_1}-\lambda_{i_2})^2}\Big(\lambda_{i_2}^2(1+\lambda_{i_2}s_q)^{\alpha_{12}} + \lambda_{i_1}\lambda_{i_2}(\alpha_{12}\lambda_{i_2}s_q-2)(1+\lambda_{i_2}s_q)^{\alpha_{12}} \\ &+ \lambda_{i_1}^2\left((1+\lambda_{i_1}s_q)^{\alpha_{12}} + \lambda_{i_2}s_q\left((1+\lambda_{i_1}s_q)^{\alpha_{12}} - \alpha_{12}(1+\lambda_{i_2}s_q)^{\alpha_{12}} - (1+\lambda_{i_2}s_q)^{\alpha_{12}}\right)\right)\Big).\end{aligned} \tag{4.58}$$

Case 2: $\mathcal{M}_v = \{\{i_1\},\{1,2\}\}$:

$$\mathbb{E}\left[X_{i_1}|S > s_q\right] = \frac{\left((1+\lambda_{12}s_q)\left(1+\frac{1}{2}(\lambda_{i_1}+\lambda_{12})s_q\right)\right)^{-\alpha_{12}}}{2(1-q)(\alpha_{12}-1)(\lambda_{i_1}-\lambda_{12})^2(\lambda_{i_1}+\lambda_{12})}$$

$$\times\left(\lambda_{12}^{2}\left(2+\alpha_{12}\lambda_{12}s_q\right)\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}+2\lambda_{i_1}\lambda_{12}\left(-3\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}+\left(1+\frac{1}{2}\left(\lambda_{i_1}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}\right.\right.$$
$$+\lambda_{12}s_q\left(\left(1+\frac{1}{2}\left(\lambda_{i_1}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}-\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}\right)\Bigg)+\lambda_{i_1}^{2}\left(2\left(1+\frac{1}{2}\left(\lambda_{i_1}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}\right.$$
$$\left.\left.-\lambda_{12}s_q\left(\alpha_{12}\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}+2\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}-2\left(1+\frac{1}{2}\left(\lambda_{i_1}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}\right)\right)\right). \quad (4.59)$$

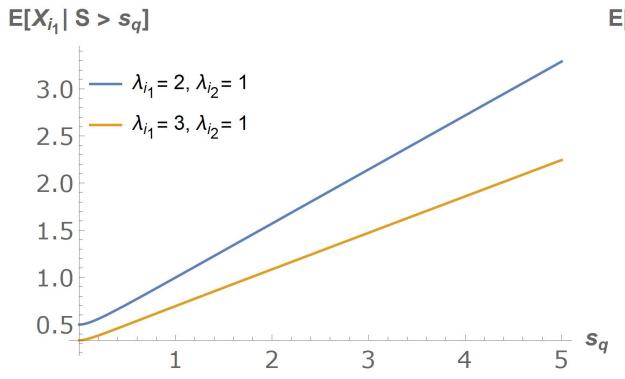
Case 3: $\mathcal{M}_v=\{\{i_2\},\{1,2\}\}$ *:*

$$\mathbb{E}\left[X_{i_1}|S>s_q\right]=\frac{2^{\alpha_{12}-1}\left(\left(1+\lambda_{12}s_q\right)\left(2+\left(\lambda_{i_2}+\lambda_{12}\right)s_q\right)\right)^{-\alpha_{12}}}{\left(1-q\right)\left(\alpha_{12}-1\right)\left(\lambda_{i_2}-\lambda_{12}\right)^{2}\lambda_{12}}$$
$$\times\left(\lambda_{12}^{2}\left(2+\alpha_{12}\lambda_{12}s_q\right)\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}+2\lambda_{i_2}^{2}\left(1+\alpha_{12}\lambda_{12}s_q\right)\left(1+\frac{1}{2}\left(\lambda_{i_2}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}\right.$$
$$-\lambda_{i_2}\lambda_{12}\left(4\left(1+\frac{1}{2}\left(\lambda_{i_2}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}+\lambda_{12}s_q\left(2\left(\left(1+\frac{1}{2}\left(\lambda_{i_2}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}-\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}\right)\right.\right.$$
$$\left.\left.\left.+\alpha_{12}\left(\left(1+\lambda_{12}s_q\right)^{\alpha_{12}}+2\left(1+\frac{1}{2}\left(\lambda_{i_2}+\lambda_{12}\right)s_q\right)^{\alpha_{12}}\right)\right)\right)\right). \quad (4.60)$$

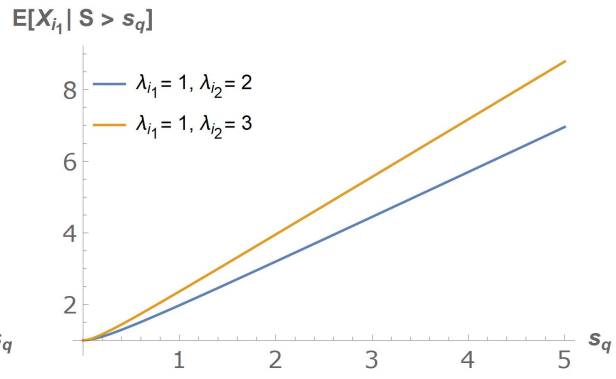
Case 4: $\mathcal{M}_v=\{\{1,2\}\}$ *:*

$$\mathbb{E}\left[X_{i_1}|S>s_q\right]=\frac{\left(1+\frac{\lambda_{12}s_q}{2}\right)^{-\alpha_{12}}\left(2+\alpha_{12}\lambda_{12}s_q\right)}{2(1-q)\left(\alpha_{12}-1\right)\lambda_{12}}. \quad (4.61)$$

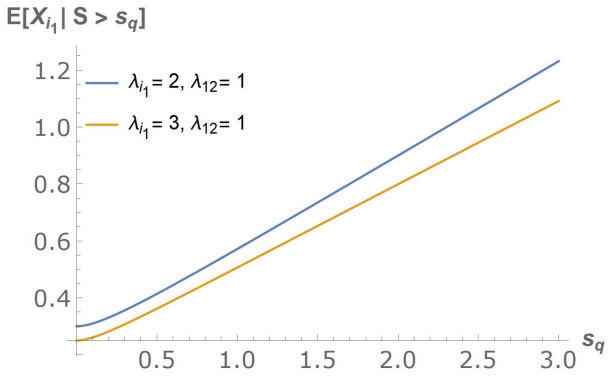
The following figure includes plots of $s_q\mapsto\mathbb{E}\left[X_{i_1}|S>s_q\right]$ *for the 4 cases of example 16. The exponential parameters are chosen to reflect the respective cases. For all plots, the gamma shape* α_{12} *is set at 2.*



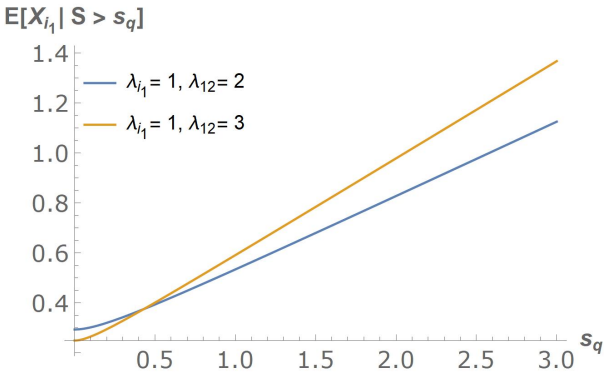
(a) Case 1a



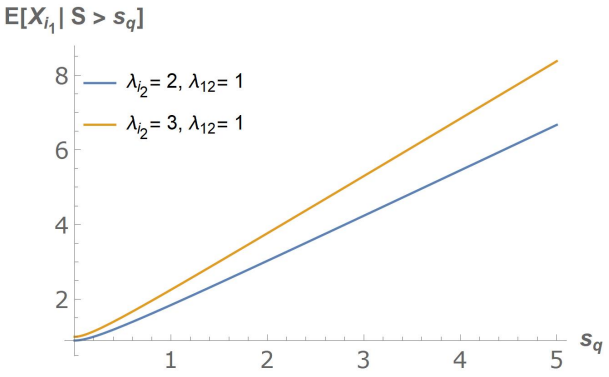
(b) Case 1b



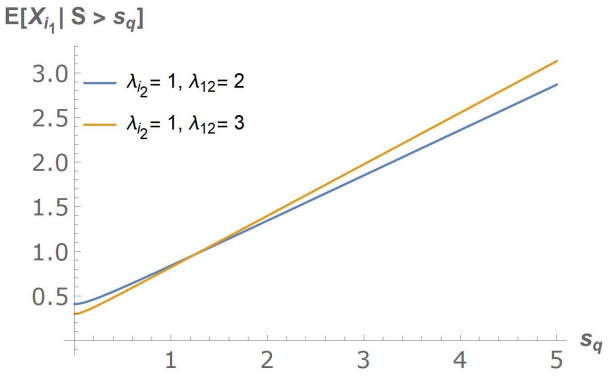
(c) Case 2a



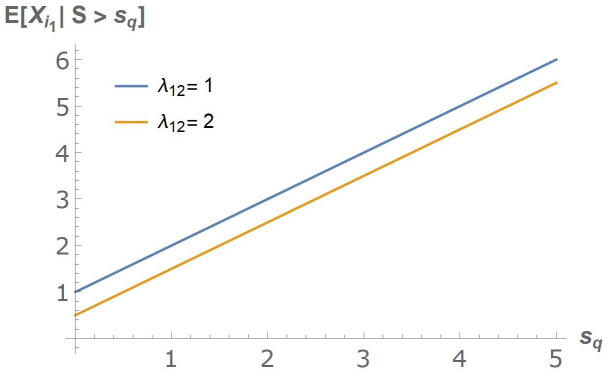
(d) Case 2b



(e) Case 3a



(f) Case 3b



(g) Case 4

As expected, for all cases, $s_q \mapsto \mathbb{E}\left[X_{i_1} | S > s_q\right]$ *is an increasing function. Interestingly, the behaviour of the allocation varies for the different choices of* λ*. In addition, since* W *is a common independent factor, then the allocation structure , for the individual loss given the sum* $S = \frac{1}{W}(V_{i_1} + V_{i_2})$*, is mostly influenced by the MO joint law of the random vector* $\boldsymbol{V} = (V_{i_1}, V_{i_2}) = (\min(E_{i_1}, E_{12}), \min(E_{i_2}, E_{12}))$.

For case 1 (independent $V_{i_1} = E_{i_1}$ and $V_{i_2} = E_{i_2}$): a) as λ_{i_1} goes up, the allocation curve shifts downward for all s_q due to the lower overall exponential mean of the corresponding random variable E_{i_1}. b) On the other hand, the opposite happens when λ_{i_2} is augmented. This is a consequence of E_{i_2} taking smaller values and most contribution for the sum comes from E_{i_1}, i.e. V_{i_1} takes larger values to compensate, rendering a larger expectation.

In case 2 ($V_{i_1} = \min(E_{i_1}, E_{12})$ and $V_{i_2} = E_{12}$): a) the role of λ_{i_1} is similar to that of case 1a. b) When it comes to varying λ_{12}, the effect on $\mathbb{E}[X_{i_1}|S > s_q]$ is quite mixed. If s_q is on the lower-end then $\mathbb{E}[X_{i_1}|S > s_q] \approx \mathbb{E}\left[\frac{\min(E_{i_1}, E_{12})}{W}\right]$. This implies a larger allocation for a smaller λ_{12} due to the larger overall mean. As s_q increases, and crosses a certain threshold, the reverse order holds. This is justified by the co-monotonic effect between (X_{i_1}, S). It is higher when λ_{12} is larger.

Case 3 ($V_{i_1} = E_{12}$ and $V_{i_2} = \min(E_{i_2}, E_{12})$): a) the role of λ_{i_2} is similar to that of case 1b. b) similar to case 2b.

For case 4 ($V_{i_1} = V_{i_2} = E_{12}$): as λ_{12} increases the mean decreases, which in turn yields smaller allocation.

A Proofs

i Proof of Theorem 14

Proof.

$$\overline{F}_{\boldsymbol{X}}(x_1,\dots,x_n) = \prod_{B\in\mathcal{M}_w}\frac{1}{\Gamma(\alpha_B)}\int_{\mathbb{R}_+^{m_w}}\mathbb{P}\left[\bigcap_{i=1}^{n}\{V_i > x_i\tilde{z}_i | W_i=\tilde{z}_i\}\right]\prod_{B\in\mathcal{M}_w} z_B^{\alpha_B-1}e^{-z_B}\mathrm{d}\boldsymbol{z},$$

$$= \prod_{B\in\mathcal{M}_w}\frac{1}{\Gamma(\alpha_B)}\int_{\mathbb{R}_+^{m_w}}\exp\left\{-\sum_{\{i\}\in\mathcal{M}_v}\lambda_i x_i\tilde{z}_i - \sum_{\{i_1,i_2\}\in\mathcal{M}_v}\lambda_{i_1i_2}\max(x_{i_1}\tilde{z}_{i_1}, x_{i_2}\tilde{z}_{i_2})\right.$$

$$\left. - \cdots - \mathbb{1}_{\{\mathcal{N}\in\mathcal{M}_v\}}\lambda_{\mathcal{N}}\max\left(x_i\tilde{z}_i : i\in\mathcal{N}\right)\right\}\prod_{B\in\mathcal{M}_w} z_B^{\alpha_B-1}e^{-z_B}\mathrm{d}\boldsymbol{z},$$

By the change of variables $(Z_B : B\in\mathcal{M}_w) \to ((R_B : B\in\mathcal{M}_w), S)$, s.t. $S = Z^+$ and $R_B = \frac{Z_B}{S}$, we get:

$$\overline{F}_{\boldsymbol{X}}(x_1,\dots,x_n) = \prod_{B\in\mathcal{M}_w}\frac{1}{\Gamma(\alpha_B)}\int_{\Delta^{m_w-1}}\int_{\mathbb{R}_+} s^{\alpha^+-1}\exp\left\{-s\left(\sum_{\{i\}\in\mathcal{M}_v}\lambda_i x_i\tilde{r}_i\right.\right.$$

$$\left.\left. + \sum_{\{i_1,i_2\}\in\mathcal{M}_v}\lambda_{i_1i_2}\max(x_{i_1}\tilde{r}_{i_1}, x_{i_2}\tilde{r}_{i_2}) + \cdots + \mathbb{1}_{\{\mathcal{N}\in\mathcal{M}_v\}}\lambda_{\mathcal{N}}\max(x_i\tilde{r}_i : i\in\mathcal{N})\right)\right\}$$

$$\times\prod_{B\in\mathcal{M}_w} r_B^{\alpha_B-1}e^{-sr_B}\mathrm{d}s\mathrm{d}\boldsymbol{r},$$

$$= \prod_{B\in\mathcal{M}_w}\frac{1}{\Gamma(\alpha_B)}\int_{\Delta^{m_w-1}}\prod_{B\in\mathcal{M}_w} r_B^{\alpha_B-1}\int_{\mathbb{R}_+} s^{\alpha^+-1}\exp\left\{-s\left(1+\sum_{\{i\}\in\mathcal{M}_v}\lambda_i x_i\tilde{r}_i\right.\right.$$

$$\left.\left. + \sum_{\{i_1,i_2\}\in\mathcal{M}_v}\lambda_{i_1i_2}\max(x_{i_1}\tilde{r}_{i_1}, x_{i_2}\tilde{r}_{i_2}) + \cdots + \mathbb{1}_{\{\mathcal{N}\in\mathcal{M}_v\}}\lambda_{\mathcal{N}}\max(x_i\tilde{r}_i : i\in\mathcal{N})\right)\right\}\mathrm{d}s\mathrm{d}\boldsymbol{r},$$

$$= \frac{1}{\beta(\boldsymbol{\alpha})}\int_{\Delta^{m_w-1}}\frac{\prod_{B\in\mathcal{M}_w} r_B^{\alpha_B-1}}{\left(1+\sum_{\{i\}\in\mathcal{M}_v}\lambda_i x_i\tilde{r}_i + \cdots + \mathbb{1}_{\{\mathcal{N}\in\mathcal{M}_v\}}\lambda_{\mathcal{N}}\max(x_i\tilde{r}_i : i\in\mathcal{N})\right)^{\alpha^+}}\mathrm{d}\boldsymbol{r}$$

Where Δ^{m_w-1} is the standard (m_w-1)-simplex.

The number of terms of the denominator $1+\sum_{\{i\}\in\mathcal{M}_v}\lambda_i x_i\tilde{r}_i+\cdots+\mathbb{1}_{\{\mathcal{N}\in\mathcal{M}_v\}}\lambda_{\mathcal{N}}\max(x_i\tilde{r}_i : i\in\mathcal{N})$ is m_v+1 and they depend on the elements of the set $\mathcal{M}_v$.

Let's assume $x_1\tilde{r}_1 \geq x_2\tilde{r}_2 \geq \cdots \geq x_n\tilde{r}_n$, we will call this case $j=1$, and denote $\lambda_1^{(1)} = \left\{\sum_{B\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}\lambda_B : 1\in B\right\}$, $\lambda_2^{(1)} = \left\{\sum_{B\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}\lambda_B : 2\in B \text{ and } 1\notin B\right\}$, $\lambda_3^{(1)} = \left\{\sum_{B\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}\lambda_B : 3\in B \text{ and } 1,2\notin B\right\}$ and so on until $\lambda_n^{(1)} = \lambda_n\mathbb{1}_{\{n\}\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}$. For $j = 2,3,\ldots,n!$ we follow the standard permutation ordering and follow the same procedure as above. For instance, for the last case $j=n!$ we have $\lambda_n^{(n!)} = \left\{\sum_{B\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}\lambda_B : n\in B\right\}$, $\lambda_{n-1}^{(n!)} = \left\{\sum_{B\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}\lambda_B : n-1\in B \text{ and } n\notin B\right\}$, and so on until $\lambda_1^{(n!)} = \lambda_1\mathbb{1}_{\{1\}\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}$. We will follow the convention $\lambda_i^{(j)}=0$ whenever $\lambda_i^{(j)} = \left\{\sum_{B\in\mathcal{M}_v\setminus\overline{\mathcal{S}}}\lambda_B : \ \ldots\right\} = \varnothing$.

When $n=3$, then the permutation cases (and order) are $j=1,\ x_1\tilde{r}_1\geq x_2\tilde{r}_2\geq x_3\tilde{r}_3$, $j=2,\ x_1\tilde{r}_1\geq x_3\tilde{r}_3\geq x_2\tilde{r}_2$, $j=3,\ x_2\tilde{r}_2\geq x_1\tilde{r}_1\geq x_3\tilde{r}_3$, $j=4,\ x_2\tilde{r}_2\geq x_3\tilde{r}_3\geq x_1\tilde{r}_1$, $j=5,\ x_3\tilde{r}_3\geq x_1\tilde{r}_1\geq x_2\tilde{r}_2$ and $j=6,\ x_3\tilde{r}_3\geq x_2\tilde{r}_2\geq x_1\tilde{r}_1$. The number of cases will depend on $\mathcal{M}_v$ and $\mathcal{S}$. It is less than $n!$, i.e. some cases will fuse together, whenever $\mathcal{M}_v\subset\mathcal{P}(\mathcal{N})$ or $\mathcal{S}\neq\varnothing$.

Define $\Delta^{(1)}(x_1,\ldots,x_n) = \left\{(r_B)_{B\in\mathcal{M}_w}\in\Delta^{m_w-1} : x_1\tilde{r}_1\geq x_2\tilde{r}_2\geq\cdots\geq x_n\tilde{r}_n\right\}$,
$\Delta^{(2)}(x_1,\ldots,x_n) = \left\{(r_B)_{B\in\mathcal{M}_w}\in\Delta^{m_w-1} : x_1\tilde{r}_1\geq x_2\tilde{r}_2\geq\cdots\geq x_n\tilde{r}_n\geq x_{n-1}\tilde{r}_{n-1}\right\}$ and so on until case $j=n!$ we have $\Delta^{(n!)}(x_1,\ldots,x_n) = \left\{(r_B)_{B\in\mathcal{M}_w}\in\Delta^{m_w-1} : x_n\tilde{r}_n\geq x_{n-1}\tilde{r}_{n-1}\geq\cdots\geq x_1\tilde{r}_1\right\}$.
Each
$\Delta^{(j)}(x_1,\ldots,x_n)$ is a region (subset) of the standard simplex determined by the tuple $(x_1,\ldots,x_n)$ s.t. $\cup_{j=1}^{n!}\Delta^{(j)}(x_1,\ldots,x_n) = \Delta^{m_w-1}$.

Using all of the above we get:

$$\overline{F}_{\boldsymbol{X}}(x_1,\ldots,x_n)=\frac{1}{\beta(\boldsymbol{\alpha})}\left[\sum_{j=1}^{n!}\int_{\Delta^{(j)}(x_1,\ldots,x_n)}\frac{\prod\limits_{B\in\mathcal{M}_w}r_B^{\alpha_B-1}}{\left(1+\sum\limits_{i=1}^{n}\lambda_i^{(j)}x_i\tilde{r}_i+\sum\limits_{B\in\mathcal{S}}r_B\sum\limits_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i:i\in B')\right)^{\alpha^+}}\mathrm{d}\boldsymbol{r}\right],$$

With the change of variables $(R_B : B\in\mathcal{M}_w)\to(T_B : B\in\mathcal{M}_w)$ s.t.

$$T_B=\frac{\left(1+\sum\limits_{i\in B}\lambda_i^{(j)}x_i+\sum\limits_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i:i\in B')\right)R_B}{1+\sum\limits_{i=1}^{n}\lambda_i^{(j)}x_i\tilde{R}_i+\sum\limits_{A\in\mathcal{S}}R_A\sum\limits_{A'\in\mathcal{S}_A}\lambda_{A'}\max(x_i:i\in A')}.$$

We notice that $\sum\limits_{B\in\mathcal{M}_w}T_B=1$ *a.s.* Let $\prod\limits^{\star}_{B\in\mathcal{M}_w}\left(1+\sum\limits_{i\in B}\lambda_i^{(j)}x_i+\sum\limits_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i:i\in B')\right)$ be the unique product (without repetitions) of $\prod\limits_{B\in\mathcal{M}_w}\left(1+\sum\limits_{i\in B}\lambda_i^{(j)}x_i+\sum\limits_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i:i\in B')\right)$. Then define:

$$\theta_B=\frac{\prod\limits^{\star}_{A\in\mathcal{M}_w}\left(1+\sum\limits_{i\in A}\lambda_i^{(j)}x_i+\sum\limits_{A'\in\mathcal{S}_A}\lambda_{A'}\max(x_i:i\in A')\right)}{\left(1+\sum\limits_{i\in B}\lambda_i^{(j)}x_i+\sum\limits_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i:i\in B')\right)},$$

which changes depending on the corresponding case j. Then R_B in terms of T_B is:

$$R_B=\frac{\theta_BT_B}{\sum\limits_{A\in\mathcal{M}_w}\theta_AT_A}.$$

Furthermore, define the transformed region of the simplex as:

$$\underline{\Delta}^{(j)}(x_1,\ldots,x_n) = \left\{ (t_B)_{B\in\mathcal{M}_w} \in \Delta^{m_w-1} : \left(\frac{\theta_B t_B}{\sum\limits_{A\in\mathcal{M}_w} \theta_A t_A} \right)_{B\in\mathcal{M}_w} \in \Delta^{(j)}(x_1,\ldots,x_n) \right\}.$$

We notice that $\{\underline{\Delta}^{(j)}(x_1,\ldots,x_n)\}_{1\leq j\leq n!}$ are disjoint except on a subset of measure zero. Putting everything together we get::

$$\begin{aligned}
\overline{F}_{\boldsymbol{X}}(x_1,\ldots,x_n) &= \frac{1}{\beta(\boldsymbol{\alpha})}\Bigg[\sum_{j=1}^{n!}\prod_{B\in\mathcal{M}_w}\left(1+\sum_{i\in B}\lambda_i^{(j)}x_i+\sum_{B'\in\mathcal{S}_B}\lambda_{B'}\max(x_i : i\in B')\right)^{-\alpha_B} \\
&\qquad\qquad \times \int\limits_{\underline{\Delta}^{(j)}(x_1,\ldots,x_n)} \prod_{B\in\mathcal{M}_w} t_B^{\alpha_B-1}\mathrm{d}\boldsymbol{t}\Bigg], \\
&= \sum_{j=1}^{n!} \mathrm{MP}^{(j)}(x_1,\ldots,x_n)\ \frac{1}{\beta(\boldsymbol{\alpha})} \int\limits_{\underline{\Delta}^{(j)}(x_1,\ldots,x_n)} \prod_{B\in\mathcal{M}_w} t_B^{\alpha_B-1}\mathrm{d}\boldsymbol{t}, \\
&= \sum_{j=1}^{n!} \mathrm{MP}^{(j)}(x_1,\ldots,x_n)\ \underline{\mathrm{Dir}}^{(j)}(x_1,\ldots,x_n).
\end{aligned}$$

To prove the last claim of $\sum\limits_{j=1}^{n!}\underline{\mathrm{Dir}}^{(j)}(x_1,\ldots,x_n) \leq 1$. Since $\tilde{r}_i \propto \sum\limits_{i\in B\in\mathcal{M}_w}\theta_B t_B$, we notice that going through the cases, $j=1$ to $j=n!$ i.e. $x_1\tilde{r}_1 \geq \cdots \geq x_n\tilde{r}_n$ to $x_n\tilde{r}_n \geq \cdots \geq x_1\tilde{r}_1$, the θ_B's decrease in the increasing $\tilde{r}_i$ terms while they increase in the decreasing terms. This implies that the union of the (disjoint) cases do not span the whole simplex i.e. $\cup_{j=1}^{n!}\underline{\Delta}^{(j)}(x_1,\ldots,x_n) \subseteq \Delta^{m_w-1}$. The inclusion is strict with the equality attained if and only if there is only one case. All of which implies $\sum\limits_{j=1}^{n!}\underline{\mathrm{Dir}}^{(j)}(x_1,\ldots,x_n) \leq 1$. □

ii Proof of Proposition 11

Proof. Similar to proof i of theorem 14, let's assume $x_1 > x_2 > \cdots > x_n$. We will call this case $j = 1$, and denote $\lambda_{l_1}^{(1)} = \left\{ \sum_{B' \in \overline{\mathcal{S}}_B} \lambda_{B'} : 1 \in B' \right\}$ where $\overline{\mathcal{S}}_B = \mathcal{S}_B \cup \mathcal{G}$, $\mathcal{G} \subseteq \{\{1\}, \{2\}, \ldots, \{n\}\}$, $\lambda_{l_2}^{(1)} = \left\{ \sum_{B' \in \overline{\mathcal{S}}_B} \lambda_{B'} : 2 \in B' \text{ and } 1 \notin B' \right\}$, $\lambda_{l_3}^{(1)} = \left\{ \sum_{B' \in \overline{\mathcal{S}}_B} \lambda_{B'} : 3 \in B' \text{ and } 1, 2 \notin B' \right\}$, and so on until $\lambda_{l_n}^{(1)} = \lambda_n \mathbb{1}_{\{n\} \in \overline{\mathcal{S}}_B}$. For $j = 2, 3, \ldots, n!$ we follow the standard permutation ordering and follow the same procedure as above. The subscripts l_s of each $\lambda^{(j)}$ correspond to the position of the case order and not of the index shared (unlike theorem 14). Finally, integrating the absolutely continuous parts of the Marshall-Olkin B-vector $(V_i : i \in B)$, we get the desired result. □

iii Proof of Proposition 14

Proof. We first notice that $\mathbb{E}\left[X_1^{k_1} X_2^{k_2}\right]$ involves combinations of $\mathbb{E}\left[\left(\frac{V_1}{W_1}\right)^{\frac{j_1}{\gamma_1}}\right]$, $\mathbb{E}\left[\left(\frac{V_2}{W_2}\right)^{\frac{j_2}{\gamma_2}}\right]$ and $\mathbb{E}\left[\left(\frac{V_1}{W_1}\right)^{\frac{j_1}{\gamma_1}}\left(\frac{V_2}{W_2}\right)^{\frac{j_2}{\gamma_2}}\right]$. Since the first two are a direct result of remark 9, and absorbing the j_1, j_2 into the γ_1, γ_2, it suffices to prove the expression of $\mathbb{E}\left[\left(\frac{V_1}{W_1}\right)^{\frac{1}{\gamma_1}}\left(\frac{V_2}{W_2}\right)^{\frac{1}{\gamma_2}}\right]$. By independence:

$$\mathbb{E}\left[\left(\frac{V_1}{W_1}\right)^{\frac{1}{\gamma_1}}\left(\frac{V_2}{W_2}\right)^{\frac{1}{\gamma_2}}\right] = \mathbb{E}\left[(V_1)^{\frac{1}{\gamma_1}}(V_2)^{\frac{1}{\gamma_2}}\right] \mathbb{E}\left[\left(\frac{1}{W_1}\right)^{\frac{1}{\gamma_1}}\left(\frac{1}{W_2}\right)^{\frac{1}{\gamma_2}}\right].$$

We will handle each expectation separately.

$$\begin{aligned}
\mathbb{E}\left[(V_1)^{\frac{1}{\gamma_1}}(V_2)^{\frac{1}{\gamma_2}}\right] &= \frac{1}{\gamma_1\gamma_2}\int_{\mathbb{R}_+^2} u_1^{\frac{1}{\gamma_1}-1} u_2^{\frac{1}{\gamma_2}-1} \exp\{-\lambda_1 u_1 - \lambda_2 u_2 - \lambda_{12}\max(u_1, u_2)\}\mathrm{d}u_1\mathrm{d}u_2, \\
&= \frac{1}{\gamma_1\gamma_2}\left[\int_0^\infty u_2^{\frac{1}{\gamma_2}-1}\exp\{-\lambda_2 u_2\}\int_{u_2}^\infty u_1^{\frac{1}{\gamma_1}-1}\exp\{-(\lambda_1+\lambda_{12})u_1\}\mathrm{d}u_1\mathrm{d}u_2\right. \\
&\quad \left. + \int_0^\infty u_2^{\frac{1}{\gamma_2}-1}\exp\{-(\lambda_2+\lambda_{12})u_2\}\int_0^{u_2} u_1^{\frac{1}{\gamma_1}-1}\exp\{-\lambda_1 u_1\}\mathrm{d}u_1\mathrm{d}u_2\right],
\end{aligned}$$

$$= \frac{1}{\gamma_1\gamma_2}\left[\frac{1}{(\lambda_1+\lambda_{12})^{\frac{1}{\gamma_1}}}\int_0^\infty u_2^{\frac{1}{\gamma_2}-1}\exp\{-\lambda_2 u_2\}\ \mathbf{\Gamma}\left(\frac{1}{\gamma_1},(\lambda_1+\lambda_{12})u_2\right)\mathrm{d}u_2\right.$$
$$\left.+\frac{1}{\lambda_1^{\frac{1}{\gamma_1}}}\int_0^\infty u_2^{\frac{1}{\gamma_2}-1}\exp\{-(\lambda_2+\lambda_{12})u_2\}\ \boldsymbol{\gamma}\left(\frac{1}{\gamma_1},\lambda_1 u_2\right)\mathrm{d}u_2,\right.$$

where $\mathbf{\Gamma}(\cdot,\cdot)$, $\boldsymbol{\gamma}(\cdot,\cdot)$ are the upper and lower incomplete gamma functions respectively. Then using equations 6.455 - (1)(2) in Gradshteyn and Ryzhik (2014) we get the desired result i.e.:

$$\mathbb{E}\left[(V_1)^{\frac{1}{\gamma_1}}(V_2)^{\frac{1}{\gamma_2}}\right] = \frac{\Gamma\left(\frac{1}{\gamma_1}+\frac{1}{\gamma_2}\right)}{(\lambda_1+\lambda_2+\lambda_{12})^{\frac{1}{\gamma_1}+\frac{1}{\gamma_2}}}\left[\frac{1}{\gamma_1}\ {}_2F_1\left(1,\frac{1}{\gamma_1}+\frac{1}{\gamma_2};\frac{1}{\gamma_2}+1;\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right.$$
$$\left.+\frac{1}{\gamma_2}\ {}_2F_1\left(1,\frac{1}{\gamma_1}+\frac{1}{\gamma_2};\frac{1}{\gamma_1}+1;\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}\right)\right].$$

We now turn to the second expectation:

$$\mathbb{E}\left[\left(\frac{1}{W_1}\right)^{\frac{1}{\gamma_1}}\left(\frac{1}{W_2}\right)^{\frac{1}{\gamma_2}}\right] = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_{12})}\int_{\mathbb{R}_+^3}\frac{u_1^{\alpha_1-1}u_2^{\alpha_2-1}u_{12}^{\alpha_{12}-1}}{(u_1+u_{12})^{\frac{1}{\gamma_1}}(u_2+u_{12})^{\frac{1}{\gamma_2}}}$$
$$\times\exp\{-u_1-u_2-u_{12}\}\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}u_{12},$$
$$= \frac{1}{\Gamma\left(\frac{1}{\gamma_1}\right)\Gamma\left(\frac{1}{\gamma_2}\right)}\int_0^\infty\int_0^\infty\frac{s_1^{\frac{1}{\gamma_1}-1}s_2^{\frac{1}{\gamma_2}-1}(1+s_1)^{-\alpha_1}(1+s_2)^{-\alpha_2}}{(s_1+s_2+1)^{\alpha_{12}}}\mathrm{d}s_1\mathrm{d}s_2,$$
$$= \frac{1}{\Gamma\left(\frac{1}{\gamma_1}\right)\Gamma\left(\frac{1}{\gamma_2}\right)}\int_0^\infty\frac{s_2^{\frac{1}{\gamma_2}-1}}{(1+s_2)^{\alpha_2}}\ {}_2F_1\left(\alpha_{12},\alpha_1+\alpha_{12}-\frac{1}{\gamma_1};\alpha_1+\alpha_{12};-s_2\right)\mathrm{d}s_2,$$
$$= \frac{\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)}{\Gamma\left(\alpha_1+\alpha_{12}\right)\Gamma\left(\frac{1}{\gamma_2}\right)}\int_0^1 t_2^{\frac{1}{\gamma_2}-1}(1-t_2)^{\alpha_2-\frac{1}{\gamma_2}-1}$$
$$\times\ {}_2F_1\left(\alpha_{12},\alpha_1+\alpha_{12}-\frac{1}{\gamma_1};\alpha_1+\alpha_{12};\frac{1}{t_2}\right)\mathrm{d}t_2,$$
$$= \frac{\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)}{\Gamma\left(\alpha_1+\alpha_{12}\right)\Gamma\left(\frac{1}{\gamma_2}\right)}\int_0^1 t_2^{\frac{1}{\gamma_2}-1}(1-t_2)^{\alpha_{12}+\alpha_2-\frac{1}{\gamma_2}-1}$$

$$\times {}_2F_1\left(\alpha_{12}, \frac{1}{\gamma_1}; \alpha_1+\alpha_{12}; t_2\right) \mathrm{d}t_2,$$

$$= \frac{\Gamma\left(\alpha_1+\alpha_{12}-\frac{1}{\gamma_1}\right)\Gamma\left(\alpha_2+\alpha_{12}-\frac{1}{\gamma_2}\right)}{\Gamma\left(\alpha_1+\alpha_{12}\right)\Gamma\left(\alpha_2+\alpha_{12}\right)}$$

$$\times {}_3F_2\left(\alpha_{12}, \frac{1}{\gamma_1}, \frac{1}{\gamma_2}; \alpha_1+\alpha_{12}, \alpha_2+\alpha_{12}; 1\right),$$

given $\alpha_1+\alpha_{12} > \frac{1}{\gamma_1}$, $\alpha_2+\alpha_{12} > \frac{1}{\gamma_2}$ and $\alpha_1+\alpha_2+\alpha_{12} > \frac{1}{\gamma_1}+\frac{1}{\gamma_2}$. Where in the 3rd, 5th and last equalities we used equations 9.111, 9.131 and 7.512 - (2), respectively, of Gradshteyn and Ryzhik (2014). □

iv Proof of Proposition 19

Proof.

$$\mathbb{E}[X_{i_1}|X_{i_2} > y] = \frac{1}{\mathbb{P}[X_{i_2} > y]}\, \mathbb{E}\left[X_{i_1}\mathbf{1}_{\{X_{i_2}>y\}}\right],$$

The first part is $\frac{1}{\mathbb{P}[X_{i_2}>y]} = (1+(\lambda_{i_2}+\lambda_{12})y\ )^{\alpha_{i_2}+\alpha_{12}}$ and the second one:

$$\begin{aligned}
\mathbb{E}\left[X_{i_1}\mathbf{1}_{\{X_{i_2}>y\}}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{V_{i_1}}{w_{i_1}}\mathbf{1}_{\{V_{i_2}>yw_{i_2}\}}\middle| W_{i_1}=w_{i_1},\ W_{i_2}=w_{i_2}\right]\right], \\
&= \frac{1}{\lambda_{i_1}(\lambda_{i_1}+\lambda_{12})}\mathbb{E}\left[\frac{1}{W_{i_1}}\left[(\lambda_{i_1}+\lambda_{12})e^{-y(\lambda_{i_2}+\lambda_{12})W_{i_2}} - \lambda_{12}e^{-y(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12})W_{i_2}}\right]\right], \\
&= \frac{1}{\lambda_{i_1}(\lambda_{i_1}+\lambda_{12})(\alpha_{i_1}+\alpha_{12}-1)}\Big[(\lambda_{i_1}+\lambda_{12})\left(1+(\lambda_{i_2}+\lambda_{12})y\ \right)^{-\alpha_{i_2}} \\
&\times {}_2F_1\left(\alpha_{12}, \alpha_{i_1}+\alpha_{12}-1; \alpha_{i_1}+\alpha_{12}; -(\lambda_{i_2}+\lambda_{12})y\right) - \lambda_{12}\left(1+(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12})y\ \right)^{-\alpha_{i_2}} \\
&\times {}_2F_1\left(\alpha_{12}, \alpha_{i_1}+\alpha_{12}-1; \alpha_{i_1}+\alpha_{12}; -(\lambda_{i_1}+\lambda_{i_2}+\lambda_{12})y\right)\Big],
\end{aligned}$$

given $\alpha_{i_1}+\alpha_{12} > 1$. □

Chapter 5

Conclusions

The deep connection between the allocation exercise and the inherent dependence structures of the losses is unmistakable and most fascinating to say the least. In this dissertation we first uncovered several associations pertaining to the interplay between these concepts. Mainly, through the lens of trivial allocations, interesting questions were answered and many others arose. Often dismissed as a futile and meaningless practice, triviality, used in the right context, elucidated both illuminating mathematical and profound economical consequences. Second, a quest was undertaken to construct a new joint distribution, possessing versatile properties, which can be used to capture the fickle character of financial risk. The outcome was a novel multivariate Pareto law that conveyed a peculiar mixture. It was fairly general to accommodate the bulk of the Pareto distributions studied thus far, while tractable enough to allow for useful expressions for the moments, measures and allocations.

To achieve the aforementioned outline, our investigation began in chapter 2. A class of generalized weighted functionals were introduced and studied. It covered most of the functionals used and examined today and the extension is done via an arbitrarily chosen aggregation function g. The functionals were primarily used as an overarching object, drawing a broad picture and allowing for meaningful subsequent results. Within the second chapter, additionally, a thorough analysis of the relationship between the aggregation and weight functions, g and w, were considered. Specifically, several theorems were proved detailing conditions for which the order between the functionals could or couldn't be possible. The

comparison was delineated in two facets. First, when g was varied and w was taken to be common, and second, when the reverse set-up is carried out. Lastly, the chapter concludes with a study of the different characterizations for both proportional and absolute triviality of allocations. The multivariate size-bias and the Laplace transform were deployed as the principal tools in obtaining the corresponding representations. Proportional triviality showed a leeway that allowed for non-degenerate distributions, while absolute triviality implied a general extreme law of negative association suggesting a possible lower bound for all joint dependencies.

In chapter 3, a more careful consideration was conducted on proportional triviality. Particularly, when the weight function is set to be the indicator function of the sum above the VaR s_q threshold, written as $\mathbf{1}_{S>s_q}$, while g is given as the canonical sum $\sum_{i=1}^{n} x_i$, then the functional outcome becomes that of the conditional tail expectation. The CTE allocation, normalized by the CTE risk measure, represented the regulatory perspective of allocations. It is axiomatic driven, consequently mathematical, with no discernible economic interpretation. On the other side of the coin sits an organic view of allocation. It is a result of an insurers profit maximization problem subject to market and utility constraints. Therefore, it embodies the economic justification of capital allocation. We, then, undertook the obvious task of reconciling the two paradigms. Their alignment was shown to exactly coincides with the concept of proportional triviality. Under a certain size-bias symmetry of the joint law, if a random vector of losses possesses such dependence, then the two allocations are the same. Several examples for both exchangeable and non-exchangeable risk were illustrated as well as a thorough characterization of the independence case. Thereafter, we generalized the size-bias identification for a polynomial function of the risks and delineated their commonalities and differences with the original linear case. Finally, we showed, through Luckacs theorem, an interesting intersection between the linear and the second order case, where the only possible laws were those of the independent common rate gammas.

Seeing how allocations and dependence interact, finally, in chapter 4, the second part of the objectives was contemplated. We employed the powerful machinery of stochastic

representations to unveil a rich structure descriptor of many extreme financial phenomena. Distinctly, we put forth a novel multivariate law, termed Pareto-Dirichlet, that is a composite of three elementary mathematical procedures. The simple quotient, constituting the multiplicative background model, between the Marshall-Olkin and the additive gamma laws yielded a mixture of multivariate Pareto, where each term is weighted by a corresponding Dirichlet probability. Through the MO joint law, we utilized the minimum operator while addition served as the binary operation in the additive gamma model. Consequently, both the stochastic depiction and the compact form of the Pareto-Dirichlet allowed for the computation of ample distributional quantities as well as explicit expressions of the joint laws. We, then, touched on the two dimensional case and derived almost all possible distributional forms with the respective product moments and correlations. Finally, we applied the new model to the standard risk measures and allocations illustrating several of the special cases for CTE risk measure as well as for both the regression and CTE allocations.

To conclude, the quest for prudent risk practices is an on-going endeavour. Much of it, is walked on uncharted territory. Expectedly, the different paths should be investigated with caution but concurrently explored with vigour to the lowest depths. The answers to the different financial and economical challenges do not only inform a robust ecosystem but reach out to include intriguing mathematical discoveries.

Bibliography

Acerbi, C. and Tasche, D. (2002). Expected Shortfall: A Natural Coherent Alternative to Value at Risk. *Economic Notes*, 31(2):379–388.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, New York.

Arnold, B. C. (2015). *Pareto Distributions (2nd ed.)*. CRC Press, Boca Raton.

Arratia, R., Goldstein, L., and Kochman, F. (2019). Size bias for one and all. *Probability Surveys*, 16(none):1 – 61.

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

Asimit, A. V., Furman, E., and Vernic, R. (2010). On a multivariate Pareto distribution. *Insurance: Mathematics and Economics*, 46(2):308–316.

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

Bauer, D. and Zanjani, G. (2016). The marginal cost of risk, risk measures, and capital allocation. *Management Science*, 62(5):1431–1457.

Bäuerle, N. and Shushi, T. (2020). Risk management with Tail Quasi-Linear Means. *Annals of Actuarial Science*, 14:170–187.

Belles-Sampera, J., Guillen, M., and Santolino, M. (2016). Compositional methods applied to capital allocation problems. *The Journal of Risk*, 19(2):15–30.

Bellini, F., Laeven, R. J. A., and Rosazza Gianin, E. (2018). Robust return risk measures. *Mathematics and Financial Economics*, 12(1):5–32.

Bignozzi, V. and Puccetti, G. (2015). Studying mixability with supermodular aggregating functions. *Statistics & Probability Letters*, 100:48–55.

Boonen, T. J., Guillen, M., and Santolino, M. (2019). Forecasting compositional risk allocations. *Insurance: Mathematics and Economics*, 84:79–86.

Boucher, J. P., Denuit, M., and Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of poisson and negative binomial distributions. *Variance*, 2(1):135–162.

Cai, J. and Li, H. (2010). Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, 42(3):810–825.

Castano-Martinez, A., Lopez-Blazquez, F., Pigueiras, G., and Sordo, M. A. (2020). A method for constructing and interpreting some weighted premium principles. *ASTIN Bulletin: The Journal of the IAA*, 50(3):1037–1064.

Chen, Y., Song, Q., and Su, J. (2021). On a class of multivariate mixtures of gamma distributions: Actuarial applications and estimation via stochastic gradient methods. *Variance*.

Chiragiev, A. and Landsman, Z. (2009). Multivariate flexible Pareto model: Dependency structure, properties and characterizations. *Statistics and Probability Letters*, 79(16):1733–1743.

Chong, W. F., Feng, R., and Jin, L. (2020). Holistic Principle for Risk Aggregation and Capital Allocation. *SSRN Electronic Journal*, pages 1–41.

Choo, W. and de Jong, P. (2009). Loss reserving using loss aversion functions. *Insurance: Mathematics and Economics*, 45(2):271–277.

Cossette, H., Côté, M.-P., Marceau, E., and Moutanabbir, K. (2013). Multivariate distribution defined with Farlie-Gumbel-Morgenstern copula and mixed Erlang marginals:

Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 52(3):560–572.

Cossette, H., Mailhot, M., and Marceau, É. (2012). TVaR-based capital allocation for multivariate compound distributions with positive continuous claim amounts. *Insurance: Mathematics and Economics*, 50(2):247–256.

Cummins, J. D. (1988). Risk-based premiums for insurance guaranty funds. *The Journal of Finance*, 43(4):823–839.

Denault, M. (2001). Coherent allocation of risk capital. *The Journal of Risk*, 4(1):1–34.

Denneberg, D. (1990). Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin: The Journal of the IAA*, 20(2):181–190.

Dhaene, J., Tsanakas, A., Valdez, E. A., and Vanduffel, S. (2012). Optimal capital allocation principles. *Journal of Risk and Insurance*, 79(1):1–28.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Heidelberg.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance: An Introduction in Discrete Time*. De Gruyter, Berlin.

Furman, E., Hackmann, D., and Kuznetsov, A. (2018a). On log-normal convolutions: An analytical-numerical method with applications to economic capital determination. Technical report, York University & Johannes Kepler University.

Furman, E., Kuznetsov, A., and Zitikis, R. (2018b). Weighted risk capital allocations in the presence of systematic risk. *Insurance: Mathematics and Economics*, 79:75–81.

Furman, E., Kye, Y., and Su, J. (2020a). Multiplicative background risk models: Setting a course for the idiosyncratic risk factors distributed phase-type. *Insurance: Mathematics and Economics*, 96:153–167.

Furman, E., Kye, Y., and Su, J. (2020b). A reconciliation of the top-down and bottom-up approaches to risk capital allocations: Proportional allocations revisited. *North American Actuarial Journal*, 0(0):1–22.

Furman, E. and Landsman, Z. (2005). Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics*, 37(3):635–649.

Furman, E. and Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin*, 36(2):433–462.

Furman, E. and Landsman, Z. (2010). Multivariate Tweedie distributions and some related capital-at-risk analyses. *Insurance: Mathematics and Economics*, 46(2):351–361.

Furman, E. and Zitikis, R. (2007). An actuarial premium pricing model for nonnormal insurance and financial risks in incomplete markets, zinoviy landsman and michael sherris, january 2007. *North American Actuarial Journal*, 11(3):174–176.

Furman, E. and Zitikis, R. (2008a). Weighted premium calculation principles. *Insurance: Mathematics and Economics*, 42(1):459–465.

Furman, E. and Zitikis, R. (2008b). Weighted risk capital allocations. *Insurance: Mathematics and Economics*, 43(2):263–269.

Furman, E. and Zitikis, R. (2009). Weighted pricing functionals with applications to insurance: An overview. *North American Actuarial Journal*, 13(4):483–496.

Furman, E. and Zitikis, R. (2010). General Stein-type covariance decompositions with applications to insurance and finance. *ASTIN Bulletin*, 40(1):369–375.

Grabisch, M., Marichal, J.-L., Mesiar, R., and Pap, E. (2009). *Aggregation Functions*. Cambridge University Press.

Gradshteyn, I. S. and Ryzhik, I. M. (2014). *Table of Integrals, Series, and Products*. Academic Press, New York.

Guan, Y., Tsanakas, A., and Wang, R. (2021). An impossibility theorem on capital allocation. *SSRN*.

Guo, Q., Bauer, D., and Zanjani, G. (2018). Capital allocation techniques: Review and comparison. Technical report, Variance.

Gupta, A. K. and Song, D. (1996). Generalized liouville distribution. *Computers & Mathematics with Applications*, 32(2):103–109.

Gupta, R. D. and Richards, D. S. (1987). Multivariate liouville distributions. *Journal of Multivariate Analysis*, 23(2):233–256.

Hardy, G., Littlewood, J., and Polya, G. (1952). *Inequalities*. Cambridge University Press, Cambridge.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1988). *Inequalities*. Cambridge University Press.

Hendriks, H. and Landsman, Z. (2017). A generalization of multivariate Pareto distributions: tail risk measures, divided differences and asymptotics. *Scandinavian Actuarial Journal*, 2017(9):785–803.

Hua, L. (2016). A Note on Upper Tail Behavior of Liouville Copulas. *Risks*, 4(40).

Ignatov, Z. G. and Kaishev, V. K. (2004). A Finite-Time Ruin Probability Formula for Continuous Claim Severities. *Journal of Applied Probability*, 41(2):570–578.

Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors (2010). *Copula Theory and Its Applications*, volume 198 of *Lecture Notes in Statistics*. Springer, Heidelberg.

Jorion, P. (2006). *Value at Risk, 3rd Ed.* McGraw-Hill.

Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, 15(3):425–437.

Kaluszka, M. and Krzeszowiec, M. (2012). Pricing insurance contracts under cumulative prospect theory. *Insurance: Mathematics and Economics*, 50(1):159–166.

Kim, J. H. and Kim, S.-Y. (2019). Tail risk measures and risk allocation for the class of multivariate normal mean–variance mixture distributions. *Insurance: Mathematics and Economics*, 86:145–157.

Kim, J. H. T. (2010). Conditional tail moments of the exponential family and its related distributions. *North American Actuarial Journal*, 14(2):198–216.

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley-Interscience.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions (4th ed.)*. Wiley, Hoboken.

Laeven, R. J. and Goovaerts, M. J. (2004). An optimization approach to the dynamic allocation of economic capital. *Insurance: Mathematics and Economics*, 35(2):299–319.

Landsman, Z., Makov, U., and Shushi, T. (2016). Tail conditional moments for elliptical and log-elliptical distributions. *Insurance: Mathematics and Economics*, 71:179–188.

Landsman, Z. and Valdez, E. A. (2003). Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71.

Lee, S. C. K. and Lin, X. S. (2012). Modeling dependent risks with multivariate Erlang mixtures. *ASTIN Bulletin*, 42(1):153–180.

Lee, W. and Ahn, J. Y. (2014). On the multidimensional extension of countermonotonicity and its applications. *Insurance: Mathematics and Economics*, 56:68–79.

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5):1137–1153.

Linsmeier, T. J. and Pearson, N. D. (2000). Value at risk. *Financial Analysts Journal*, 56(2):47–67.

Lukacs, E. (1955). A characterization of the gamma distribution. *Annals of Mathematical Statistics*, 26(2):319–324.

Marshall, A. W. and Olkin, I. (1967). A generalized bivariate exponential distribution. *Journal of Applied Probability*, 4(2):291–302.

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

Millossovich, P., Tsanakas, A., and Wang, R. (2021). A theory of multivariate stress testing. *SSRN Electronic Journal*.

Mohammed, N., Furman, E., and Su, J. (2021). Can a regulatory risk measure induce profit-maximizing risk capital allocations? the case of conditional tail expectation. *Insurance: Mathematics and Economics*, 101:425–436.

Nadarajah, S. and Kotz, S. (2005). Sums and ratios for marshall and olkin's bivariate exponential distribution. *Mathematical Scientist*, 30(2):112–119.

Ng, K. W., Tian, G. L., and Tang, M. L. (2011). *Dirichlet and Related Distributions: Theory, Methods and Applications*. Wiley, Chichester.

Panjer, H. (2002). Measurement of risk, solvency requirements and allocation of capital within financial conglomerates. Technical report, SOA.

Pareto, V. (1964). *Cours d'économie politique*. Librairie Droz.

Patil, G. and Ord, J. (1976). On size-biased sampling and related form-invariant weighted distributions. *Sankhyā: The Indian Journal of Statistics Series B*, 38(1):48–61.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, pages 179–189.

Phillips, R. D., Cummins, J. D., and Allen, F. (1998). Financial pricing of insurance in the multiple-line insurance company. *The Journal of Risk and Insurance*, 65(4):597–636.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131.

Porth, L., Zhu, W., and Tan, K. S. (2014). A credibility-based erlang mixture model for pricing crop reinsurance. *Agricultural Finance Review*, 74(2):162–187.

Puccetti, G., Wang, B., and Wang, R. (2012). Advances in complete mixability. *Journal of Applied Probability*, 49(2):430–440.

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.

Ratovomirija, G., Tamraz, M., and Vernic, R. (2017). On some multivariate Sarmanov mixed Erlang reinsurance risks: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 74:197–209.

Richards, D. and Uhler, C. (2019). Loading monotonicity of weighted premiums, and total positivity properties of weight functions. *Journal of Mathematical Analysis and Applications*, 475(1):532–553.

Sendov, H. S., Wang, Y., and Zitikis, R. (2011). Log-supermodularity of weight functions, ordering weighted losses, and the loading monotonicity of weighted premiums. *Insurance: Mathematics and Economics*, 48(2):257–264.

Shushi, T. and Yao, J. (2020). Multivariate risk measures based on conditional expectation and systemic risk for Exponential Dispersion Models. *Insurance: Mathematics and Economics*, 93:178–186.

Su, J. and Furman, E. (2017). Multiple risk factor dependence structures: Distributional properties. *Insurance: Mathematics and Economics*, 76:56–68.

Tasche, D. (2002). Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7):1519–1533.

Tasche, D. (2004). Allocating portfolio economic capital to sub-portfolios. *Economic Capital: A Practitioner's Guide*, page 275 – 302. Risk.net.

Tsanakas, A. (2008). Risk measurement in the presence of background risk. *Insurance: Mathematics and Economics*, 42(2):520–528.

Tsanakas, A. and Barnett, C. (2003). Risk capital allocation and cooperative pricing of insurance liabilities. *Insurance: Mathematics and Economics*, 33:239–254.

Venter, G. (2004). Capital allocation survey with commentary. *North American Actuarial Journal*, 8(2):96–107.

Verbelen, R., Antonio, K., and Claeskens, G. (2016). Multivariate mixtures of Erlangs for density estimation under censoring. *Lifetime Data Analysis*, 22(3):429–455.

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics*, 38(2):413–426.

Vernic, R. (2011). Tail conditional expectation for the multivariate Pareto distribution of the second kind: Another approach. *Methodology and Computing in Applied Probability*, 13(1):121–137.

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, R. and Zitikis, R. (2021). An axiomatic foundation for the expected shortfall. *Management Science*, 67(3):1413–1429.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(1):71–92.

Willmot, G. E. and Woo, J.-K. (2014). On some properties of a class of multivariate Erlang mixtures with insurance applications. *ASTIN Bulletin*, 45(01):151–173.

Yamai, Y. and Yoshiba, T. (2005). Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking & Finance*, 29(4):997–1015.

Zhu, W., Tan, K. S., and Porth, L. (2019). Agricultural insurance ratemaking: Development of a new premium principle. *North American Actuarial Journal*, 23(4):512–534.