ECONOMIC CAPITAL ANALYSIS WITHIN PORTFOLIOS OF DEPENDENT AND HEAVY-TAILED RISKS

YISUB KYE

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS
YORK UNIVERSITY,
TORONTO, ONTARIO

NOVEMBER 2019

Abstract

In the nowadays reality of prudent risk management, the problem of determining aggregate risk capital in financial entities has been intensively studied for quite long. As a result, canonical methods have been developed and even embedded in regulatory accords. While applauded by some and questioned by others, these methods provide a much desired standard benchmark for everyone. The situation is very different when the aggregate risk capital needs to be allocated to the business units of a financial entity. That is, there are overwhelmingly many ways to conduct the allocation exercise, and there is arguably no standard method to do so on the horizon.

Two overarching approaches to allocate the aggregate risk capital stand out. These are the top-down approach that entails that the allocation exercise is imposed by the corporate centre, and the bottom-up approach that implies that the allocation of the aggregate risk to business units is informed by these units. Briefly, the top-down allocations start with the aggregate risk capital that is then replenished among business units according to the views of the centre, thus limiting the inputs from the business units. The bottom-up approach does start with the business units, but it is, as a rule, too granular, and so may lead to missing the wood for the trees.

The first chapter of this dissertation is concerned with the bottom-up approach to allocating the aggregate risk capital. Namely, we put forward a general theoretical framework for the multiplicative background risk model that allows for arbitrarily distributed idiosyncratic and systemic risk factors. We reveal links between the just-mentioned general structure and the one with the exponentially distributed idiosyncratic risk factors (a key player in the modern actuarial modelling), study relevant theoretical properties of the new structure, and discuss important special cases. Also, we construct realistic numerical examples borrowed

from the context of the determination and allocation of economic capital. The examples suggest that a little departure from exponentiality can have substantial impacts on the outcome of risk analysis.

In the second chapter of this dissertation, we question the way in which the risk allocation practice is conducted in the state of the art and present an alternative that comes from the context of the distributions defined on the multidimensional simplex. More specifically, we put forward a new family of mixed-scaled Dirichlet distributions that contain the classical Dirichlet distribution as a special case, exhibit a multitude of desirable closure properties, and emerge naturally within the multivariate risk analysis context. As a by-product, our invention revisits the proportional allocation rule that is often used in applications. Interestingly, we are able to unify the top-down and the bottom-up approaches to allocating the aggregate risk capital into one encompassing method.

During the study underlying the present dissertation, we rediscovered certain problems of the standard deviation as the ubiquitous measure of variability. In particular, the standard deviation is frequently infinite for insurance risks in the Property and Casualty lines of business, and so it cannot be used to quantify variability therein. Also, the standard deviation is a questionable measure of variability when non-normal distributions are considered, and normality is rarely a reasonable assumption in insurance practice. Therefore, in the third chapter of this dissertation, we turn to an alternative measure of variability. The Gini Mean Difference, which we study in the third chapter, is finite whenever the mean is so, and it is suitable for measuring variability for non-normal risks. Nevertheless, Gini Mean Difference is by far less common in actuarial science than the standard deviation. One of the main reasons for this lies in the critics associated with the computability of the 'Gini'. We reveal convenient ways to compute the Gini Mean Difference measure of variability explicitly and often effortlessly. The thrust of our approach is a link, which we discover, between the Gini and the notion of statistical sample size-bias. Not only the just-mentioned link opens up advantageous computational routes for Gini, but also yields an alternative interpretation for it.

To my parents

Acknowledgements

First of all, I thank Professor Edward Furman for supervising my research with deep patience. He has been a great mentor and a role model in my academic endeavors. Also, I owe special thanks to Professor Jianxi Su, my co-author. Every time our research had lost direction, his advice and opinions helped me make desired progress.

Additionally, I am grateful to my Ph.D. committee members, Professor Hongmei Zhu, Professor Yang Shen, and Professor Ida Ferrara as well as to the external examiner, Professor Ri cardas Zitikis, who had offered a lot of constructive advice, which resulted in an improved work. Last, but very much not least, I am greatly indebted to my parents for their love and support, which they have shown me, as well as to my brothers, Joseph and Lucas, for supporting my parents and family while I was studying.

Table of Contents

Chapter 1

Introduction

1.1 General overview and introduction

The notion of dependence is vital to the solvency of a stand-alone (re)insurer as well as to the stability of the entire financial sector and even to the overall health of the economy. Speaking shortly, neglecting dependences within and among risks portfolios can and often does lead to solvency issues and bankruptcy of institutions. Nevertheless, traditional actuarial models rest on the assumption of independence. Unrealistic as it is, this assumption often allows for convenient simplifications and thus guarantees a desirable level of analytic tractability. In recent years, the increasing complexity of (re)insurance products as well as high competition amongst (re)insurers, have encouraged a growing interest of actuaries in modelling dependent risks, which are to this end considered random variables (RVs) $X_i,\ i = 1, \ldots, n (\in \mathbb{N})$, say. The main object of interest is then the corresponding n-variate probability distribution $F_{1,\ldots,n}$, possessing an intuitively interpretable dependence structure (copula) $C_{1,\ldots,n}$ and univariate margins $F_i,\ i = 1, \ldots, n$ that are appropriate for applications in insurance.

Whether one chooses to pursue a 'natural' way to formulate $F_{1,\ldots,n}$ by specifying the dependence of RVs $X_1, \ldots, X_n$ or the two - steps 'copula' way $C(F_1, \ldots, F_n)$, complications arise immediately. In the former case, the dependence is imposed by real-life phenomena, but the marginal distributions are not a free choice and can be rather cumbersome. In the latter case, the margins are arbitrary, but the dependence is often difficult to interpret. The choice is not trivial, but identifying a 'proper' multivariate probability model with dependence is a

fundamental pillar of the modern quantitative risk management.

Another fundamental pillar of the nowadays quantitative risk management addresses the ways of the risk assessment per se. Namely, given that a desirable multivariate probability model has been chosen, the next step is to quantify the riskiness inherent in it. To this end, let $\mathcal{X}$ denote the collection of all actuarial risks, be it the standalone $X_1, \ldots, X_n$ mentioned above, or, more generally, risk portfolios $(X_1, \ldots, X_n)'$. Then the functionals $\mathrm{H} : \mathcal{X} \to [0,\ \infty]$ and $\mathrm{A} : \mathcal{X} \times \mathcal{X} \to [0,\ \infty]$ subject to the condition that $\mathrm{A}[X,\ X] = \mathrm{H}[X]$ for every $X \in \mathcal{X}$ are called 'risk measure' and 'risk capital allocation', respectively, and the former is employed to measure the riskiness of standalone risks or their aggregates, whereas the latter is used to quantify the riskiness that distinct risks imply when they are a part of a risk portfolio [e.g., 34, 85, and references therein].

The volumes of scientific contributions on both $F_{1,\ldots,n}$ and $\mathrm{H},\ \mathrm{A}$ are vast and growing quickly. In practice, however, of utmost importance is the task of computing (preferably in an explicit way) the value of the functionals H and A in the context of the desirable cumulative distribution functions (CDFs) $F_{1,\ldots,n}$. We refer to [35] for elliptical distributions, to [23] for Phase-type distributions, to [49] for Tweedie distributions, to [115] for Skew-normal and Pareto distributions, to name just a few. Also, we refer to [56] for a general approach.

In this dissertation, we will touch on the two pillars of the successful quantitative risk management mentioned above. In other words, (1) we will propose and study dependence structures that are appropriate for actuarial modelling beyond the assumption of independence, and (2) we will evaluate some regulatory risk measures / risk capital allocations in the context of the models proposed in (1). As by products, our study will reveal a way to reconcile two cornerstone and distinct approaches to allocating risk capital, that is the so-called 'top-down' and 'bottom-up' routes [e.g., 61], as well as to handle the computational burden that has been impeding the popularization of the Gini Mean Difference measure of variability in actuarial science.

1.2 Literature review

In the following three subsections, we briefly review relevant existing literature on the multivariate reduction technique to formulating multivariate CDFs, the compendium of regulatory risk measures / capital allocations, and some relevant computational results.

1.2.1 Multivariate models

Multivariate probability distributions play a critically important role in the insurance practice - from the day to day business operations such as pricing, risk reporting, asset and liability management, to the more sophisticated level of enterprise risk management including the calculation of economic and regulatory capital requirements and risk allocations. While there are a wealth of well-established probabilistic models available for describing the stochastic nature of stand-alone risks, say $X_1, \ldots, X_n$, it is notoriously much more challenging to model the intangible dependence structure among these risks.

Let the coordinates of the non-negative $m(\in \mathbb{N})$-variate random variable $\mathbf{Y} = (Y_1 \ldots, Y_m)' \sim G_{1,\ldots,m}$ represent possibly dependent risk factors (RFs), and denote by $\mathbf{X} = (X_1, \ldots, X_n)'$ a risk portfolio (RP), whose risk components (RCs) X_i are exposed to (sub)sets $\mathcal{S}_i \subseteq \{1, \ldots, m\}$, $i = 1, \ldots, n$ of the RFs Y_j, $j = 1, \ldots, m$. In addition, let $\mathcal{Y}$ and $\mathcal{X}$ be collections of the RFs and the RPs, respectively. Then, given a functional map $T : \mathcal{Y} \to \mathcal{X}$, the object of main interest is the stochastic representations $\mathbf{X} = T(\mathbf{Y})$, as well as the involved cumulative and/or decumulative distribution functions (CDFs) and/or (DDFs), respectively. The set-up echoes real-life, where RPs face multiple RFs, such as, e.g., market, credit, systemic, demographic, operational and residual, to name a few. Popular choices of the map T have been (1) additive [e.g., 49], (2) multiplicative [104], and (3) minima [9].

Of particular interest to this dissertation are the multiplicative maps that correspond to the so-called multiplicative background risk (MBR) models [10, 44]. Namely, let R be a non-negative RV, $m = n$, and assume that the RVs $Y_1, \ldots, Y_n$ are independent mutually and on the RV R, then, under the MBR framework, we have the following stochastic representation

$$(X_1, \ldots, X_n) = (RY_1, \ldots, RY_n). \tag{1.1}$$

RVs $Y_1, \ldots, Y_n$ are as a rule viewed as some idiosyncratic RFs, and hence the dependence structure of the MBR models is stipulated by the systemic RF, R, and its interplay with the sequence of the idiosyncratic RFs. MBR models, when the idiosyncratic RVs are distributed exponentially, have been applied often times in actuarial science by, e.g., [4] in the context of ruin theory, [10] for portfolio construction, [101] for modeling auto insurance claims, and [106] for default risk. The broad range of applications of the MBR models in actuarial science (and beyound) have spurred intensive research attempting to understand intricate theoretical structure underlying the model. Namely, the dependence properties of the MBR models are studied in [108] and the aggregation properties are studied by [102]. Some computational aspects of the MBR models are studied by [27].

The aforementioned assumption of exponentiality on the idiosyncratic RFs means succinctly that $Y_i \sim Exp(\sigma_i)$, where $\sigma_i > 0$ is the scale parameter, $i = 1, \ldots, n$ [e.g., 4, 10, 102, 106, etc.]. The resulting MBR models are termed the exponential mixtures in the literature, notationally, $\boldsymbol{X} \sim \mathrm{EM}(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)'$ is a vector of scale parameters. This connects the MBR models with the Laplace transform of the systemic RV R and grants a considerable amount of mathematical tractability to the application of the model [e.g., 67]. However, mathematical convenience often comes at the expense of generality, and the MBR models are not an exception in this regard. Specifically, the exponentiality assumption mentioned above readily implies that the probability distributions of the RCs $X_1, \ldots, X_n$ agree up to the scale transformation, and, as a result, the copula function that governs the RCs of the RP $(X_1, \ldots, X_n)'$ is exchangeable. This peculiarity of the MBR models does not necessarily comply with the real life insurance data.

1.2.2 Risk measures and allocation rules

Hereafter, we interpret $X, X_1, \ldots, X_n$ as insurance risks, and let $\mathcal{X}$ denote a collection of such risks. Risk measure H is a functional that maps risk RVs in $\mathcal{X}$, or their cumulative distribution functions, to the extended non-negative real line; succinctly $H : \mathcal{X} \to [0, \infty]$. Two arguably most popular risk measures in insurance nowadays are the Value-at-Risk (VaR) and the Conditional Tail Expectation (CTE).

For a prudence level $p \in [0, 1)$ and a non-negative risk rv $X \in \mathcal{X}$ with the cdf F,

succinctly $X \sim F$, the VaR risk measure is defined as

$$\mathrm{VaR}_p[X] = \inf\{x \geq 0 : \ F(x) \geq p\}. \tag{1.2}$$

As in practice the values of interest of p are close to one, the VaR risk measure is considered a tail-based risk measure. For the sake of the notational simplicity in what follows we sometimes write x_p instead of $VaR_p[X]$. VaR is location and scale invariant, monotone, and additive for comonotonic risks but not coherent, as it may in general violate the sub-additivity axiom [7]. As a result, VaR may discourage diversification, thus providing counterintuitive risk assessments.

An alternative to the VaR risk measure, the CTE risk measure, is defined for the prudence level $p \in [0,\ 1),\ \mathbb{P}[X > x_p] > 0$ and the risk rv X with finite mean [7], as

$$\mathrm{CTE}_p[X] = \mathbb{E}[X|\ X > x_p]. \tag{1.3}$$

CTE is clearly an example of a tail-based risk measure; it is coherent for risk rv's with continuous CDFs, and additive for comonotonic risks [1, 79].

Speaking briefly, VaR is merely the generalized p-th quantile of the CDF F, and as such it cannot shed light on the severity of the tail-risk. This, along with the already-mentioned possible violation of sub-additivity, have caused pressure to replace the VaR with the CTE risk measure in regulatory accords. In turn, CTE that can be seen as the average tail-risk, cannot capture the variability of the tail-risk beyond the quantile x_p, and yet variability has been pivotal in risk management at least since 1952 [e.g., 81].

More generally, risk measures that hinge on the higher-order-conditional-moments, given these moments are finite and well-defined, are of the form

$$\mathrm{CTE}_p^k[X] = \mathbb{E}[X^k|\ X > x_p], \tag{1.4}$$

where $k \geq 0$ and $p \in [0,\ 1)$ were considered in Kim [71]. In particular, to include variability in risk measurement [48] proposed the tail-standard-deviation (TSD) risk measure, which, for the risk rv X with finite variance, prudence level $p \in [0,\ 1)$ and non-negative loading

parameter $\lambda \geq 0$, is given by

$$\mathrm{TSD}_p^{\lambda}[X] = \mathrm{CTE}_p[X] + \lambda\ \mathrm{SD}_p[X], \tag{1.5}$$

where

$$\mathrm{SD}_p[X] = \sqrt{\mathbb{E}\big[(X - \mathrm{CTE}_p[X])^2|\ X > x_p\big]}. \tag{1.6}$$

Risk measures (1.2)-(1.4) may look ad-hoc at the first glance, but they can be treated in a holistic manner by the machinery of the weighted risk measures [55]. Specifically, for $w : (-\infty,\ \infty) \to (0,\ \infty)$, the weighted risk measure indexed by w is defined as

$$H_w[X] = \frac{\mathbb{E}[Xw(X)]}{\mathbb{E}[w(X)]}, \tag{1.7}$$

subject to the fact that the expectations above are well-defined and finite. Then the choices of the Direc delta, as well as $w(x) = \mathbf{1}\{x > x_p\}$ and $w(x) = x\mathbf{1}\{x > x_p\}$ weight functions result in the VaR, CTE and TSD risk measures, respectively.

It often happens in mathematical sciences that generalizing an object highlights its characteristics and helps to understand it better. The formulation of VaR, CTE and TSD risk measures as (1.7) does exactly that, and also immediately shows that all these risk measures can be extended into allocation rules as following

$$A_w[X_i] = \frac{\mathbb{E}[X_i w(S)]}{\mathbb{E}[w(S)]}, \tag{1.8}$$

where $S = X_1 + \cdots + X_n$, and the expectations are assumed to be well-defined and finite.

Recall the VaR, CTE and TSD weighted risk measures presented in (1.2), (1.3) and (1.5), respectively. These are extended into capital allocation rules, given that all the involved expectations are well-defined and finite, using, for $s_p = VaR_p[S],\ p \in [0,\ 1)$,

$$VaR_p\left[X_i, S\right] = \mathbb{E}\left[X_i|\ S = s_p\right],\ p \in [0,\ 1), \tag{1.9}$$

$$\mathrm{CTE}_p[X_i,\ S] = \mathbb{E}[X_i|\ S > s_p] \tag{1.10}$$

and

$$\mathrm{TCov}_p[X_i,\ S] = \mathrm{Cov}[X_i,\ S|\ S > s_p], \tag{1.11}$$

where $i = 1, \ldots, n$. Clearly

$$\sum_{i=1}^{n} \mathrm{VaR}_p[X_i,\ S] = \mathrm{VaR}_p[S] \tag{1.12}$$

and

$$\sum_{i=1}^{n} \mathrm{CTE}_p[X_i,\ S] = \mathrm{CTE}_p[S] \tag{1.13}$$

as well as

$$\sum_{i=1}^{n} \mathrm{TCov}_p[X_i,\ S] = \mathrm{TSD}_p^2[S], \tag{1.14}$$

and so the desired additivity of the allocation functional is achieved in the all cases.

Another important weighted risk measure is Gini Shortfall (GS), which was introduced by [53] to fix the non-monotonicity of TSD, as well as other disadvantages of this tail-based risk measure of variability that hinges on the standard deviation. Specifically, for risk RVs with finite mean, $p \in [0,\ 1)$, loading parameter $\lambda \geq 0$, the GS risk measure is defined as

$$\mathrm{GS}_p^{\lambda}[X] = CTE_p[X] + \lambda TGini_p[X], \tag{1.15}$$

where, for X^* and X^{**} denoting two independent copies of X,

$$\mathrm{TGini}_p[X] = \mathbb{E}[X^* - X^{**}|\ X^* > x_p, X^{**} > x_p]. \tag{1.16}$$

Remarkably, the GS risk measure has turned out additive for comonotonic risks, and even coherent for $\lambda \in [0,\ 1/2)$ [53].

1.2.3 Computational results

Given MBR models (1.1) and a weighted risk measure / risk capital allocation, the desired end-result is often the value of the latter in the context of the former. Two streams of research are worthy to notice. On the one hand, [28], and recently [64], employed the machinery of divided difference to evaluate the CTE and TSD risk measures, as well as the

economic capital allocations based on them in the context of exponential mixtures. In the just-mentioned papers the final results are written in terms of divided differences of higher orders with repeated arguments. Hence, one needs to evoke recursive algorithms [e.g., 31] as well as accurate algorithms for computing derivatives of higher order to get precise outcomes. On the other hand, [114], and more recently [10], developed recursive formulas for various weighted risk measures under the MBR models framework. These recursive methods are as a rule computationally demanding and may encounter difficulties when tackling risk portfolios of high dimensions.

Chapter 2

Multiplicative background risk models with non-exchangeable dependencies

2.1 Introduction

In this paper, insurance risks are represented by non-negative random variables (rv's), $X_1, \ldots, X_n$, and $\mathcal{X}$ denotes the collection of such risk rv's. We routinely assume that all rv's are defined on the same probability space, and that this space is atomless in the sense that it is possible to define sequences of independent rv's with arbitrary distributions on it.

In practice, the risk rv's $X_1, \ldots, X_n$, which can also be viewed as risk components of a risk portfolio $\mathbf{X} = (X_1, \ldots, X_n)$, are generally dependent. The dependence can be, e.g., due to a common (systemic) risk factor or due to a set of such risk factors. We remind in passing that according to the Financial Stability Board and the International Monetary Fund, the systemic risk can be caused by impairment of all or parts of the financial system [46]. In the following, we assume that all risk components are exposed to one random systemic risk factor, R. In addition, we assume that each risk component X_i is characterized by its own idiosyncratic (also, specific) risk factor, denoted by the rv Y_i, $i = 1, \ldots, n$. It is common to assume [e.g., 10] that the idiosyncratic risk factors and the systemic risk factor are independent, and we indeed do so here.

The exponential multiplicative background risk models are a specialization of the set-up described above. Namely, let $E_1, \ldots, E_n$ be a sequence of exponentially distributed and

mutually independent rv's with arbitrary scale parameters, respectively, $\sigma_1, \ldots, \sigma_n \in \mathbb{R}_+$, and let R be an arbitrarily distributed non-negative rv independent on the just-mentioned exponential rv's. Then the distribution of the n-dimensional rv $(\mathcal{E}_1, \ldots, \mathcal{E}_n)$ that admits the following stochastic representation

$$(\mathcal{E}_1, \ldots, \mathcal{E}_n) = (RE_1, \ldots, RE_n) \tag{2.1}$$

is said to establish the class of exponential multiplicative background risk (MBR) models, or, simply exponential mixtures in the jargon of distribution theory.

Admittedly, the MBR models with the idiosyncratic risk factors distributed exponentially are very tractable technically. Namely, the decumulative distribution function (ddf) of the rv $\boldsymbol{\mathcal{E}} = (\mathcal{E}_1, \ldots, \mathcal{E}_n)$ is readily obtained as

$$\overline{F}_{\boldsymbol{\mathcal{E}}}(x_1, \ldots, x_n) = \int_0^\infty \exp\left\{r^{-1}(x_1\sigma_1^{-1} + \cdots + x_n\sigma_n^{-1})\right\} \mathrm{d}F_R(r) \quad \text{for} \quad (x_1, \ldots, x_n) \in [0,\ \infty)^n, \tag{2.2}$$

which, for $\hat{F}_{R^{-1}}$ denoting the Laplace transform of the rv R^{-1}, implies

$$\overline{F}_{\boldsymbol{\mathcal{E}}}(x_1, \ldots, x_n) = \hat{F}_{R^{-1}}\left(x_1\sigma_1^{-1} + \cdots + x_n\sigma_n^{-1}\right) \quad \text{for} \quad (x_1, \ldots, x_n) \in [0,\ \infty)^n.$$

Moreover, the joint probability density function (pdf) of the rv $(\mathcal{E}_1, \ldots, \mathcal{E}_n)$ can be obtained from the alternating sign derivative of the Laplace transform, that is we have

$$f_{\boldsymbol{\mathcal{E}}}(x_1, \ldots, x_n) = \prod_{j=1}^n \frac{1}{\sigma_j}(-1)^n \frac{\partial^n}{\partial y_1, \ldots, \partial y_n} \hat{F}_{\boldsymbol{\mathcal{E}}}(y_1, \ldots, y_n)\big|_{y_j = \frac{x_j}{\sigma_j}} \quad \text{for} \quad (x_1, \ldots, x_n) \in \mathbb{R}_+^n.$$

Hence the really colossal volume of results acquired for the Laplace transform, can be employed with a little effort to study the MBR models with exponentially distributed idiosyncratic risk factors [e.g., 4, 10, 101, 102, 106, 107].

Alongside the mathematical tractability, the exponentiality assumption is rather restrictive and may undermine the real world applicability of the exponential MBR models significantly. For instance, as the distributions of the idiosyncratic risk factors therein can differ only up to a scale transformation, the Pearson coefficients of correlation are identical for any

pair of risk components in the risk portfolios that admit the exponential MBR stochastic representation. This symmetry in the dependence has very little in common with reality; we refer to, e.g., the correlation matrices suggested in the EIOPA Quantitative Impact Study report [38]. Therefore in this paper we accept the challenge to find satisfactory generalizations of the exponential MBR models, which on one hand allow for more realistic dependencies, and on the other hand inherit, at least to some extent, the much desired tractability.

The rest of the paper is organized as follows. We set up the discussion with the most general formulation of the multiplicative background risk models in Section 2.2, where we mention very briefly the class of phase-type (PH) distributions as a natural extension of the exponential distributions. Having a recap of assorted basic properties of the PH distributions in Section 2.3, then in Section 2.4, we introduce and study the class of the multiplicative background risk models with the idiosyncratic risk factors distributed PH, succinctly PH-MBR models. Applications to actuarial risk analysis are considered throughout the paper, and illustrated numerically in Section 2.5. Specifically, in order to emphasize the tractability of the PH-MBR models and to study the implications of adopting these structures versus the classical exponential MBR models, we derive explicit expressions for some practically popular tail-based risk measures and the risk capital allocation rules based on them. Section **??** concludes the paper.

2.2 Multiplicative background risk models with general idiosyncratic and systemic risk factors

2.2.1 Definition and distributional properties

An immediate consequence of (2.1) is that in the context of the exponential MBR structures, the idiosyncratic risk factors are, up to rescaling, equal in distribution. This feature contributes greatly to the mathematical transparency of this particular class of factor models, yet it has little in common with reality [e.g., 9, 47, for examples of additive and minimum-based MBR structures, neither of which shares the aforementioned limitation].

In view of the above, our aspirations in this section are to introduce a generalization

of the exponential MBR structures that inherits the remarkable tractability of stochastic representation (2.1), and yet allows for more heterogeneity of the involved idiosyncratic risk factors. Definition 1 achieves this goal.

Definition 1. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be a vector of mutually independent rv's representing n idiosyncratic risk factors, and let R be a rv denoting a systemic risk factor independent on all idiosyncratic risk factors. The rv $\mathbf{X} = (X_1, \ldots, X_n)$ that admits the stochastic representation

$$(X_1, \ldots, X_n) = (RY_1, \ldots, RY_n), \text{ or succingtly } \mathbf{X} = R\mathbf{Y}, \tag{2.3}$$

is referred to as the general multiplicative background risk structure.

The versatility of the general MBR structures has a multitude of much welcomed implications. We document one of such implications in the next proposition. The proof of the proposition is by construction and thus omitted. At the outset, recall that, for a rv $X \in \mathcal{X}$ with finite variance, the quantity $\mathrm{SD}(X)/\mathbb{E}[X]$, where $\mathrm{SD}(X)$ represents the standard deviation, is called the coefficient of variation.

Proposition 1. *Let $(X_l, X_m) = (RY_l, RY_m)$, $1 \leq l \neq m \leq n$ be a pair of rv's coming from MBR structure (2.3), then the coefficient of Pearson correlation between the two is*

$$Corr[X_l, X_m] \;=\; \frac{1}{\left(1 + c_l^2 + c_l^2/c^2\right)^{1/2} \left(1 + c_m^2 + c_m^2/c^2\right)^{1/2}} \tag{2.4}$$

given that the correlation coefficient is well-defined and finite. Here c_l, c_m and c are the coefficients of variation of the rv's Y_l, Y_m and R, respectively.

Proposition 1 implies that the Pearson correlation is an increasing function of the coefficient of variation of the systemic risk factor R, and a decreasing function of the coefficients of variation of the idiosyncratic risk factors Y_l and Y_m. One implication of this observation is that the dependency of the general MBR structure, when measured with the Pearson coefficient of correlation, is stipulated by the interplays between the variability of the systemic and idiosyncratic risk factors.

Another implication revealed by Proposition 1 is that in the context of the general MBR structures, the Pearson coefficient of correlation can attain any value in the interval $[0,\ 1]$.

This is not so for the case of the exponential MBR structures as the next corollary clarifies, thereby providing a formal basis for the empirical critique on (2.1) [e.g., 29, and references therein]. Proposition 1 and Corollary 1 justify abandoning the class of exponential MBR structures in favour of the general ones.

Corollary 1. *Let the rv's Y_l and $Y_m, 1 \leq l \neq m \leq n$ be independent and distributed exponentially with arbitrary scale parameters, then we have*

$$Corr[X_l, X_m] \leq \frac{1}{2 + c^{-2}} < 0.5,$$

where c denotes the coefficient of variation of the rv R [e.g., the distinct exponential MBR models in, respectively, 104, 107].

We have thus far motivated the study of general MBR models (2.3) through the advantage of obtaining a more flexible dependence structure. Another way to conceptualize the general MBR structures is by considering them the *safety-loaded* exponential MBR structures. Namely, start with the $(n + 1)$ factor model formally defined by stochastic representation (2.1). It often happens that, in order to address model risk or due to other risk management purposes, actuaries, and/or risk managers, are interested to load the probability distributions of the idiosyncratic/systemic risk factors, or both. One way to accomplish the desired loading is via distorting the ddf's of the risk factors of interest [116, and references therein]. Another route for pursuing the same task is by evoking the notion of *weighted distributions* [50, and references therein]. We adopt the latter loading method in the following discussion.

First, let us elucidate the weighting approach in the simplest single risk case. Given a rv $X \in \mathcal{X}$ and a non-decreasing function $w : \mathbb{R}_+ \to \mathbb{R}_+$, denote by $X^{[w]}$ the weighted counterpart of the original rv X. It is known [e.g., 94] that

$$\mathbb{P}[X^{[w]} > x] = \frac{\mathbb{E}\left[w(X)\mathbf{1}\{X > x\}\right]}{\mathbb{E}\left[w(X)\right]} \geq \mathbb{E}\left[\mathbf{1}\{X > x\}\right] = \mathbb{P}[X > x] \text{ for all } x \in [0,\ \infty),$$

assuming that the expectation $\mathbb{E}[w(X)]$ is well-defined, finite and non-zero. This means that the weighted counterpart of the rv X dominates the original rv stochastically (1st order), succinctly $X^{[w]} \geq_{st} X$. Hence the desired loading is achieved. In the multivariate case, we

need a few more definitions.

Definition 2. Let $\mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}_+^n$ be a positive rv, then, for a Borel-measurable *weight* function $w : \mathbb{R}_+^n \to \mathbb{R}_+$ such that $0 < \mathbb{E}[w(\mathbf{X})] < \infty$, the distribution of the weighted counterpart of the rv $\mathbf{X}$, say $\mathbf{X}^{[w]}$, is defined as

$$\mathbb{P}[\mathbf{X}^{[w]} \in \mathrm{d}\mathbf{x}] = \frac{w(\mathbf{x})}{\mathbb{E}[w(\mathbf{X})]}\mathbb{P}[\mathbf{X} \in \mathrm{d}\mathbf{x}] \tag{2.5}$$

for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$. The rv's $\mathbf{X}^{[w]}$ and $\mathbf{X}$ are independent.

Definition 3. A rv $\mathbf{X} = (X_1, \ldots, X_n)$ is said to be associated if the bound

$$\mathrm{Cov}(r(\mathbf{X}),\ s(\mathbf{X})) \geq 0, \tag{2.6}$$

and therefore

$$\mathbb{E}\left[r(\mathbf{X})s(\mathbf{X}))\right] \geq \mathbb{E}\left[r(\mathbf{X})\right]\mathbb{E}\left[s(\mathbf{X})\right]$$

hold for all non-decreasing functions $r, s : \mathbb{R}_+^n \to \mathbb{R}_+$ such that the involved expectations are well-defined and finite.

We refer to [41] for a detailed discussion of the notion of association, as well as to [51, 55] for applications to insurance pricing. Definition 3 hints at the importance of the class of non-decreasing weight functions. Of all such functions, of particular interest to our present paper is the size-biased (SB) weight functions that give birth to the SB weighted distributions [93].

Definition 4. Consider the set-up of Definition 2, and let $\mathbf{h} = (h_1, \ldots, h_n)$ be a vector of non-negative constants such that $0 < \mathbb{E}\left[\prod_{j=1}^n X_j^{h_j}\right] < \infty$. Then the distribution of the size-biased counterpart of the rv $\mathbf{X}$, say $\mathbf{X}^{[\mathbf{h}]} = \left(X_1^{[h_1]}, \ldots, X_n^{[h_n]}\right)$, is defined as

$$\mathbb{P}[\mathbf{X}^{[\mathbf{h}]} \in \mathrm{d}\mathbf{x}] = \mathbb{P}[X_1^{[h_1]} \in \mathrm{d}x_1, \ldots, X_n^{[h_n]} \in \mathrm{d}x_n] = \frac{x_1^{h_1} \cdots x_n^{h_n}}{\mathbb{E}[X_1^{h_1} \cdots X_n^{h_n}]}\mathbb{P}[\mathbf{X} \in \mathrm{d}\mathbf{x}] \tag{2.7}$$

for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$. The rv's $\mathbf{X}^{[\mathbf{h}]}$ and $\mathbf{X}$ are independent.

In order to facilitate our following discussion, we generalize the notion of stochastic dominance to higher dimensions.

Definition 5. For n-dimensional rv's $\mathbf{U} = (U_1, \ldots, U_n)$ and $\mathbf{V} = (V_1, \ldots, V_n)$, we say that $\mathbf{U} \geq_{uo} \mathbf{V}$, that is $\mathbf{U}$ stochastically dominates $\mathbf{V}$ in the upper orthant order, if

$$\mathbb{P}[\mathbf{U} \geq \mathbf{u}] \geq \mathbb{P}[\mathbf{V} \geq \mathbf{u}]$$

for all $\mathbf{u} \in [0,\ \infty)^n$.

Proposition 2. *Let the rv $\mathbf{X} = (X_1, \ldots, X_n)$ admit the general MBR structure stochastic representation, then $\mathbf{X}^{[\mathbf{h}]} \geq_{uo} \mathbf{X}$.*

Proof. Since $Y_1, \ldots, Y_n$ are mutually independent and hence associated, and because $\mathbf{X} = R\mathbf{Y}$, we obtain that $X_1, \ldots, X_n$ are associated. Therefore

$$\mathbb{P}\left[\mathbf{X}^{[\mathbf{h}]} > \mathbf{x}\right] = \frac{\mathbb{E}\left[X_1^{h_1} \cdots X_n^{h_n} \mathbf{1}\{\mathbf{X} > \mathbf{x}\}\right]}{\mathbb{E}\left[X_1^{h_1} \cdots X_n^{h_n}\right]} \geq \mathbb{E}\left[\mathbf{1}\{\mathbf{X} > \mathbf{x}\}\right] = \mathbb{P}\left[\mathbf{X} > \mathbf{x}\right]$$

for all $\mathbf{x} \in [0,\ \infty)^n$. This completes the proof. □

We find the following general result frequently useful in the sequel.

Proposition 3. *Let the rv $\mathbf{X} = (X_1, \ldots, X_n)$ admit stochastic representation (2.3), then, for every Borel-measurable function $g : \mathbb{R}_+^n \to \mathbb{R}_+$ such that the expectations below are well-defined and finite, we have*

$$\mathbb{E}\left[g\left(\mathbf{X}^{[\mathbf{h}]}\right)\right] = \int_0^\infty \mathbb{E}\left[g\left(r\mathbf{Y}^{[\mathbf{h}]}\right)\right] dF_{R^{[h_+]}}(r), \tag{2.8}$$

where $h_+ = h_1 + \cdots + h_n$.

Proof. Start by noticing that under the conditions of the proposition, we have

$$\mathbb{E}\left[g\left(\mathbf{X}^{[\mathbf{h}]}\right)\right] = \frac{1}{\mathbb{E}\left[X_1^{h_1} \cdots X_n^{h_n}\right]} \mathbb{E}\left[X_1^{h_1} \cdots X_n^{h_n} g(\mathbf{X})\right].$$

Then, for $h_+ = h_1 + \cdots + h_n$, we have

$$\mathbb{E}\left[X_1^{h_1} \cdots X_n^{h_n} g(\mathbf{X})\right] \quad = \quad \mathbb{E}\left[R^{h_+}\right] \int_0^\infty \mathbb{E}\left[Y_1^{h_1} \cdots Y_n^{h_n} g(r\mathbf{Y})\right] dF_{R^{[h_+]}}(r)$$

$$= \mathbb{E}\left[R^{h_+}\right]\mathbb{E}\left[Y_1^{h_1}\cdots Y_n^{h_n}\right]\int_0^\infty \mathbb{E}\left[g\left(r\mathbf{Y}^{[\mathbf{h}]}\right)\right]dF_{R^{[h_+]}}(r),$$

as well as

$$\mathbb{E}\left[X_1^{h_1}\cdots X_n^{h_n}\right] = \mathbb{E}\left[R^{h_+}\right]\mathbb{E}\left[Y_1^{h_1}\cdots Y_n^{h_n}\right].$$

This completes the proof of the proposition. □

The next corollary along with Proposition 2 shows that, as we have mentioned hitherto, general MBR structures (2.3) can be considered safety-loaded exponential MBR structures (2.1).

Corollary 2. *Let the rv* $\mathbf{X} = (X_1, \ldots, X_n)$ *admit the general MBR structure stochastic representation* (2.3)*, then* $\mathbf{X}^{[\mathbf{h}]} = \left(X_1^{[h_1]}, \ldots, X_n^{[h_1]}\right)$ *also admits the general MBR structure stochastic representation with the vector of idiosyncratic risk factors* $\mathbf{Y}^{[\mathbf{h}]} = \left(Y_1^{[h_1]}, \ldots, Y_n^{[h_1]}\right)$ *and the systemic risk factor* $R^{[h_+]}$, $h_+ = h_1 + \cdots + h_n$.

Proof. The asserted result follows immediately from Proposition 3 by choosing the identity function for g. □

At this point, it is a natural question as to whether the rv's $\mathbf{X}_1$ and $\mathbf{X}_2$ admitting the following stochastic representations

$$\mathbf{X}_1 = \left(RY_1^{[h_1]}, \ldots, RY_n^{[h_n]}\right) \text{ and } \mathbf{X}_2 = \left(R^{[h_+]}Y_1, \ldots, R^{[h_+]}Y_n\right)$$

dominate the original *rv* $\mathbf{X} = (RY_1, \ldots, RY_n)$ in the upper orthant order. The answer is in affirmative. Indeed we have $\mathbf{X}_1 \geq_{uo} \mathbf{X}$ since

$$\overline{F}_{\mathbf{X}_1}(\mathbf{x}) = \int_0^\infty \overline{F}_{\mathbf{Y}^{[\mathbf{h}]}}\left(r^{-1}\mathbf{x}\right)\mathrm{d}F_R(r) \geq \int_0^\infty \overline{F}_{\mathbf{Y}}\left(r^{-1}\mathbf{x}\right)\mathrm{d}F_R(r) = \overline{F}_{\mathbf{X}}(\mathbf{x}) \text{ for all } \mathbf{x} \in [0,\ \infty),$$

and also $\mathbf{X}_2 \geq_{uo} \mathbf{X}$ because $R^{[h_+]} \geq_{st} R$.

The following corollary is another immediate consequence of Proposition 3 by setting $h_1 = \cdots = h_n = 0$ and with the convention that the zero-SB rv is equal in distribution to

the original rv.

Corollary 3. *For the rv* $\mathbf{X} = (X_1, \ldots, X_n)$ *that admits* (2.3)*, we have*

- *the ddf of the rv* $X_j,\ j = 1, \ldots, n$ *is*

$$\overline{F}_{X_j}(x) = \int_0^\infty \overline{F}_{Y_j}(r^{-1}x) dF_R(r),\ x \in [0,\ \infty); \tag{2.9}$$

- *the joint ddf of the rv* $(X_1, \ldots, X_n)$ *is*

$$\overline{F}_{\mathbf{X}}(x_1, \ldots, x_n) = \int_0^\infty \overline{F}_{\mathbf{Y}}(r^{-1}x_1, \ldots, r^{-1}x_n) dF_R(r),\ (x_1, \ldots, x_n) \in [0,\ \infty)^n; \tag{2.10}$$

- *the joint product moment of higher order is, for positive integers* $k_1, \ldots, k_n$ *and* $k_+ = k_1 + \cdots + k_n$*,*

$$\mathbb{E}\left[\prod_{j=1}^n X_j^{k_j}\right] = \int_0^\infty r^{k_+} \mathbb{E}\left[\prod_{j=1}^n Y_j^{k_j}\right] dF_R(r) = \mathbb{E}[R^{k_+}] \prod_{j=1}^n \mathbb{E}[Y_j^{k_j}]. \tag{2.11}$$

A somewhat more interesting consequence of Proposition 3 is stated next.

Corollary 4. *Consider the rv* $\mathbf{X} = (X_1, \ldots, X_n)$ *that admits stochastic representation* (2.3)*, and let* $S_{\mathbf{X}} = X_1 + \cdots + X_n$ *and* $S_{\mathbf{X}}^{[\mathbf{h}]} = X_1^{[h_1]} + \cdots + X_n^{[h_n]}$*. Also, in a similar fashion, let* $S_{\mathbf{Y}}^{[\mathbf{h}]} = Y_1^{[h_1]} + \cdots + Y_n^{[h_n]}$*, then*

$$\begin{aligned} \mathbb{P}[S_{\mathbf{X}}^{[\mathbf{h}]} > s] &= \int_0^\infty \mathbb{P}\left[S_{\mathbf{Y}}^{[\mathbf{h}]} > r^{-1}s\right] dF_{R^{[h_+]}}(r) && (2.12) \\ &= \int_0^\infty \frac{1}{\mathbb{E}\left[\prod_{j=1}^n Y_j^{h_j}\right]} \mathbb{E}\left[\prod_{j=1}^n Y_j^{h_j} \mathbf{1}\left\{S_{\mathbf{Y}} > r^{-1}s\right\}\right] dF_{R^{[h_+]}}(r) && (2.13) \end{aligned}$$

for all $s \in [0,\ \infty)$*. Then we have* $S_{\mathbf{X}}^{[\mathbf{h}]} \geq_{st} S_{\mathbf{X}}$ *for any vector of non-negative constants* $\mathbf{h} = (h_1, \ldots, h_n)$*. Furthermore, for another vector of non-negative constants* $\mathbf{h}_1 = (h_1, \ldots, h_{n-1}, h_n')$*, that is* $\mathbf{h}$ *and* $\mathbf{h}_1$ *agree up to the* $(n-1)$*-th coordinate, we have* $S_{\mathbf{X}}^{[\mathbf{h}]} \geq_{st} S_{\mathbf{X}}^{[\mathbf{h}_1]}$ *if* $h_n \geq h_n'$*.*

Proof. Equation (2.12) is a direct consequence of Proposition 3. Also as $Y_j^{[h_j]} \geq_{st} Y_j$ for all $j = 1, \ldots, n$ and since $Y_1, \ldots, Y_n$ are mutually independent, we have

$$\mathbb{P}\left[S_{\mathbf{Y}}^{[\mathbf{h}]} > s\right] \geq \mathbb{P}[S_{\mathbf{Y}} > s] \text{ for all } s \in [0,\ \infty),$$

and this implies that $S_{\mathbf{X}}^{[\mathbf{h}]} \geq_{st} S_{\mathbf{X}}$. Further note that, for $S_{\mathbf{X},-n}^{[\mathbf{h}]} = X_1^{[h_1]} + \cdots + X_{n-1}^{[h_{n-1}]}$ and $h_n \geq h_n'$, we have

$$\begin{aligned}\mathbb{P}\left[S_{\mathbf{X}}^{[\mathbf{h}]} > s\right] = \mathbb{P}\left[S_{\mathbf{X},-n}^{[\mathbf{h}]} + X_n^{[h_n]} > s\right] &= \int_0^s \mathbb{P}\left[X_n^{[h_n]} > s - x\right] dF_{S_{\mathbf{X},-n}^{[\mathbf{h}]}}(x) \\ &\geq \int_0^s \mathbb{P}\left[X_n^{[h_n']} > s - x\right] dF_{S_{\mathbf{X},-n}^{[\mathbf{h}]}}(x) = \mathbb{P}\left[S_{\mathbf{X},-n}^{[\mathbf{h}]} + X_n^{[h_n']} > s\right] = \mathbb{P}\left[\right.\end{aligned}$$

for all $s \in [0,\ \infty)$. Finally, to prove Equation (2.13), we explore the Laplace transform of the rv $S_{\mathbf{Y}}^{[\mathbf{h}]} = Y_1^{[h_1]} + \cdots + Y_n^{[h_n]}$ and obtain

$$\begin{aligned}\mathbb{E}\left[\exp\left\{-t\left(S_{\mathbf{Y}}^{[\mathbf{h}]}\right)\right\}\right] &= \prod_{j=1}^n \mathbb{E}\left[\exp\left\{-t\left(Y_j^{[h_j]}\right)\right\}\right] = \prod_{j=1}^n \frac{1}{\mathbb{E}\left[Y_j^{h_j}\right]} \mathbb{E}\left[Y_j^{h_j} \exp\left\{-tY_j\right\}\right] \\ &= \frac{1}{\mathbb{E}\left[\prod_{j=1}^n Y_j^{h_j}\right]} \mathbb{E}\left[\prod_{j=1}^n Y_j^{h_j} \exp\left\{-tS_{\mathbf{Y}}\right\}\right] \quad \text{for} \quad Re(t) > 0.\end{aligned}$$

This completes the proof of the corollary. □

2.2.2 Applications to risk management

The discussion hitherto has a strong flavour of distribution theory, and yet it is remarkably connected to real world applications in risk management. Two examples that are motivated by recent developments in the actuarial literature are presented below.

Example 1. [Proposition 1 in 47] Consider a rv with mutually independent coordinates $\mathbf{Y} = (Y_1, \ldots, Y_n) \in \mathbb{R}_+^n$. In an attempt to derive expressions for the CTE risk measure and the allocation rule based on it, the following two formulas were reported, for ${Y_k}^*,\ k = 1, \ldots, n$ and $S_{\mathbf{Y}}^* = (Y_1 + \cdots + Y_n)^*$ denoting the SB of order one variants of the rv's Y_k and $S_{\mathbf{Y}}$, respectively, and assuming that the involved conditional expectations are well-defined and finite,

$$\mathbb{P}[S_{\mathbf{Y}} - Y_k + Y_k^* > s] = \frac{\mathbb{E}[Y_k |\ S_{\mathbf{Y}} > s]}{\mathbb{E}[Y_k]} \mathbb{P}[S_{\mathbf{Y}} > s] \quad \text{for} \quad s \in [0,\ \infty)$$

and

$$\mathbb{P}[S^*_{\mathbf{Y}} > s] = \frac{\mathbb{E}[S_{\mathbf{Y}}|\ S_{\mathbf{Y}} > s]}{\mathbb{E}[S_{\mathbf{Y}}]}\mathbb{P}[S_{\mathbf{Y}} > s] \quad \text{for} \quad s \in [0,\ \infty).$$

As a result it was noted that the following equality must hold, yet the intuition as to why it does eluded the authors back in 2005

$$\mathbb{P}[S^*_{\mathbf{Y}} > s] = \sum_{k=1}^{n} \frac{\mathbb{E}[Y_k]}{\mathbb{E}[S_{\mathbf{Y}}]}\mathbb{P}[S_{\mathbf{Y}} - Y_k + Y^*_k > s] \quad \text{for} \quad s \in [0,\ \infty). \tag{2.14}$$

Given the change of measure noted in Corollary 4, the *distributional* proof of Equation (2.14) is transparent

$$\mathbb{E}\left[e^{-tS^*_{\mathbf{Y}}}\right] = \frac{\mathbb{E}\left[S_{\mathbf{Y}}e^{-tS_{\mathbf{Y}}}\right]}{\mathbb{E}[S_{\mathbf{Y}}]} = \sum_{k=1}^{n} \frac{\mathbb{E}[Y_k]}{\mathbb{E}[S_{\mathbf{Y}}]}\frac{\mathbb{E}\left[Y_k e^{-tS_{\mathbf{Y}}}\right]}{\mathbb{E}[Y_k]} = \sum_{k=1}^{n} \frac{\mathbb{E}[Y_k]}{\mathbb{E}[S_{\mathbf{Y}}]}\mathbb{E}\left[e^{-t\left(S_{\mathbf{Y}}-Y_k+Y^*_k\right)}\right],\ Re(t) > 0.$$

This fills in the missing part in [47], and in particular implies that Lemma 1 and Theorem 3 therein are redundant.

Example 2. [Theorem 7 in 64] In the context of the exponential MBR structure

$$(\mathcal{E}_1, \ldots, \mathcal{E}_n) = R(E_1, \ldots, E_n), \tag{2.15}$$

where $E_j \sim Exp(\sigma_j)$ it was proved that it is the order of the scale parameters that determines the order of the risk capital allocation rules based on the CTE risk measure. Specifically, with the help of the theory of divided differences and multiply monotonic functions, it was shown that, for $S_{\boldsymbol{\mathcal{E}}} = \mathcal{E}_1 + \cdots + \mathcal{E}_n$,

$$\sigma_i \geq \sigma_j \Rightarrow \mathbb{E}[\mathcal{E}_i|\ S_{\boldsymbol{\mathcal{E}}} > s] \geq \mathbb{E}[\mathcal{E}_j|S_{\boldsymbol{\mathcal{E}}} > s],\ 1 \leq i, j \leq n,\ s \in \mathbb{R}_+. \tag{2.16}$$

Remarkably, (2.16) easily follows from Proposition 3 and the well-known fact that $\sigma_i \geq \sigma_j \Leftrightarrow E_i \geq_{st} E_j$. Specifically, for $S_{\boldsymbol{E}} = E_1 + \cdots + E_n$ and since $E^*_i \overset{d}{=} E_i + E'_i$ where $E'_i \sim Exp(\sigma_i)$ and E'_i and E_i are independent, we have $\sigma_i \geq \sigma_j$ implying

$$\mathbb{E}[\mathcal{E}_i|\ S_{\boldsymbol{\mathcal{E}}} > s] = \mathbb{E}[\mathcal{E}_i]\int_0^\infty \mathbb{P}\left[S_{\boldsymbol{E}} - E_i + E^*_i > r^{-1}s\right] dF_{R^*}(r) = \mathbb{E}[\mathcal{E}_i]\int_0^\infty \mathbb{P}\left[S_{\boldsymbol{E}} + E_i > r^{-1}s\right] dF_{R^*}(r)$$

$$\begin{aligned} &\geq \mathbb{E}[\mathcal{E}_j] \int_0^\infty \mathbb{P}\left[S_{\boldsymbol{E}} + E_j > r^{-1}s\right] dF_{R^*}(r) = \mathbb{E}[\mathcal{E}_j] \int_0^\infty \mathbb{P}\left[S_{\boldsymbol{E}} - E_j + E_j^* > r^{-1}s\right] dF_{R^*}(r) \\ &= \mathbb{E}[\mathcal{E}_j |\ S_{\boldsymbol{\mathcal{E}}} > s] \end{aligned}$$

for all $s \in \mathbb{R}_+$. Furthermore, the reverse implication is also true. That is if $\mathbb{E}[\mathcal{E}_i |\ S_{\boldsymbol{\mathcal{E}}} > s] \geq \mathbb{E}[\mathcal{E}_j |\ S_{\boldsymbol{\mathcal{E}}} > s]$ for all $s \in \mathbb{R}_+$, then $\sigma_i \geq \sigma_j,\ 1 \leq i, j \leq n$. To see that the latter assertion is true, assume that the conditional expectations are ordered, that is

$$\mathbb{E}[\mathcal{E}_i |\ S_{\boldsymbol{\mathcal{E}}} > s] \geq \mathbb{E}[\mathcal{E}_j |\ \mathcal{S} > s] \Rightarrow \int_0^\infty \left(\mathbb{E}[\mathcal{E}_i]\mathbb{P}\left[S_{\mathbb{E}} + E_i > r^{-1}s\right] - \mathbb{E}[\mathcal{E}_j]\mathbb{P}\left[S_{\mathbb{E}} + E_j > r^{-1}s\right]\right) dF_{R^*}(r) \geq 0,$$

and then note that this bound holds only if the integrand is non-negative, which is so only if $\sigma_i \geq \sigma_j$. To summarize, we have shown that in the context of exponential mixtures, Theorem 7 of [64] can be strengthened to state that $\sigma_i \geq \sigma_j \Leftrightarrow \mathbb{E}[\mathcal{E}_i |\ S_{\boldsymbol{\mathcal{E}}} > s] \geq \mathbb{E}[\mathcal{E}_j |\ S_{\boldsymbol{\mathcal{E}}} > s]$ for all $s \in \mathbb{R}_+$.

We conclude this section by deriving an expression for certain conditional higher order mixed moments of the rv $(X_k, X_1 + \cdots + X_n)$ where $\mathbf{X} = (X_1, \ldots, X_n)$ admits stochastic representation (2.3) and $k = 1, \ldots, n$. The result is summarized in the next proposition, and it is of pivotal importance latter on in this paper.

Proposition 4. *Let the rv* $\mathbf{X}$ *admit the generalized MBR structure stochastic representation* (2.3)*, and let* $S_{\mathbf{X}} = X_1 + \cdots + X_n$*,* $S_{\mathbf{X},-k} = X_1 + \cdots + X_{k-1} + X_{k+1} + \cdots + X_n$*, and similarly* $S_{\mathbf{Y},-k} = Y_1 + \cdots + Y_{k-1} + Y_{k+1} + \cdots + Y_n$ *then, for* $a,\ b \in \mathbb{N}$*, and assuming that the expectations below are well-defined and finite, we have*

$$\mathbb{E}\left[X_k^a S_{\mathbf{X}}^b \mathbf{1}\{S_{\mathbf{X}} > s\}\right] = \sum_{j=0}^{b} \binom{b}{j} \mathbb{E}\left[S_{\mathbf{X},-k}^j\right] \mathbb{E}\left[X_k^{a+b-j}\right] \int_0^\infty \mathbb{P}\left[(S_{\mathbf{Y},-k})^{[j]} + Y_k^{[a+b-j]} > r^{-1}s\right] dF_{R^{[a+b]}}(r) \tag{2.17}$$

for $k = 1, \ldots, n$ *and* $s \in \mathbb{R}_+$.

Proof. By construction, we have the following string of equations

$$\mathbb{E}\left[X_k^a S_{\mathbf{X}}^b \mathbf{1}\{S_{\mathbf{X}} > s\}\right] = \mathbb{E}\left[\left(\sum_{l \neq k} X_l + X_k\right)^b X_k^a \mathbf{1}\{S_{\mathbf{X}} > s\}\right]$$

$$
\begin{aligned}
&= \sum_{j=0}^{b} \binom{b}{j} \mathbb{E}\left[\left(\sum_{l\neq k} X_l\right)^j X_k^{a+b-j} \mathbb{1}\{S_{\mathbf{X}} > s\}\right] \\
&= \sum_{j=0}^{b} \binom{b}{j} \int_0^\infty r^{a+b} \mathbb{E}\left[\left(\sum_{l\neq k} Y_l\right)^j Y_k^{a+b-j} \mathbb{1}\{S_{\mathbf{Y}} > s/r\}\right] \mathrm{d}F_R(r) \\
&= \mathbb{E}\left[R^{a+b}\right] \sum_{j=0}^{b} \binom{b}{j} \int_0^\infty \mathbb{E}\left[\left(\sum_{l\neq k} Y_l\right)^j Y_k^{a+b-j} \mathbb{1}\{S_{\mathbf{Y}} > s/r\}\right] \mathrm{d}F_{R^{[a+b]}}(r).
\end{aligned}
$$

The assertion is proved by the change of measure type of reasoning used earlier in Corollary 4. □

In order to obtain further practical insights about the gain in flexibility versus the loss in tractability in the context of the transition from the exponential MBR models to the general MBR models, additional distributional assumptions are required. To this end, in the following we assume that the idiosyncratic risk factors are distributed phase-type. (We recall in passing that exponential MBR structures (2.1) hinge on the assumption that the idiosyncratic risk factors have exponential distributions with arbitrary scale parameters.) The choice of the PH distributions is natural at least because (i) they provide a convenient generalization of the class of exponential distributions, and (ii) they are dense in the class of probability distributions on the non-negative half of the real line, and hence can theoretically approximate any distribution therein arbitrarily well [e.g., 14].

2.3 Phase-type distributions

In this section, we offer a brief overview of some basic properties of the phase-type distributions. Herein and in the following, matrices and vectors are denoted by bold-face upper-case and lower-case letters, respectively. Let $\mathbf{T}$ be a square matrix with real entries, then the *matrix exponential* is defined as the following power series

$$
e^{\boldsymbol{T}} = \sum_{i=0}^{\infty} \frac{\boldsymbol{T}^i}{i!},
$$

where $\mathbf{T}^0 = \mathbf{I}$ is the identity matrix of the same dimension as the matrix $\mathbf{T}$.

Definition 6 (Asmussen, 2003)**.** The rv $Y \in \mathbb{R}_+$ is said to be distributed phase-type, if its ddf is given by

$$\overline{F}_Y(y) = \boldsymbol{\alpha} e^{y\boldsymbol{T}} \mathbf{1} \quad \text{for} \quad y \in [0,\ \infty), \tag{2.18}$$

where the parameters $\boldsymbol{\alpha}, \boldsymbol{T}, \mathbf{1}$ are respectively, a row vector, a non-singular negatively-defined square matrix, and a column vector of ones of appropriate dimension. The pair of parameters $(\boldsymbol{\alpha}, \boldsymbol{T})$ is called *the representation* of the phase-type distribution, and the row dimensions of $\boldsymbol{\alpha}$ (also, $\boldsymbol{T}$) is referred to as the order of the phase-type distribution. Succinctly, we write $Y \sim PH(\boldsymbol{\alpha}, \boldsymbol{T})$.

A direct result of the definition above is that the pdf of the rv Y is given by

$$f_Y(y) \quad = \quad \boldsymbol{\alpha} e^{y\boldsymbol{T}} \boldsymbol{t} \quad \text{for} \quad y \in \mathbb{R}_+, \tag{2.19}$$

where $\boldsymbol{t} = -\boldsymbol{T}\mathbf{1}$. We refer to [5, 98, 99] for some recent applications of the phase-type distributions in actuarial science.

The exponential distribution is clearly a member of the class of PH distributions.

Example 3 (PH representation of the exponential distribution)**.** For $\sigma > 0$, set $\boldsymbol{\alpha} = (1)$ and $\boldsymbol{T} = (-\sigma^{-1})$, then $Y \sim PH(\boldsymbol{\alpha}, \boldsymbol{T})$ has pdf

$$f_Y(y) = \sigma^{-1} e^{-\sigma^{-1} y} \quad \text{for} \quad y \in \mathbb{R}_+,$$

which agrees with the exponential distribution with scale parameter $\sigma \in \mathbb{R}_+$.

Other well-known PH distributions are hypo-exponential and mixed Erlang [e.g. 14, and references therein for more examples].

Example 4 (Convolutions of n independent exponentially distributed rv's)**.** Consider $E_i \sim Exp(\sigma_i)$, for $\sigma_i > 0$, $i = 1, \ldots, n$ and assume that the rv's $E_1, \ldots, E_n$ are mutually independent; also,

set $S_{\boldsymbol{E}} = E_1 + \cdots + E_n$. Then $S_{\boldsymbol{E}} \sim PH(\boldsymbol{\alpha}, \boldsymbol{T})$ with the parameters

$$\boldsymbol{\alpha} = (1, \underbrace{0, \ldots, 0}_{n-1}) \qquad \text{and} \qquad \boldsymbol{T} = \underbrace{\begin{pmatrix} -\sigma_1^{-1} & \sigma_1^{-1} & & & \\ & \ddots & \ddots & & \\ & & & -\sigma_{n-1}^{-1} & \sigma_{n-1}^{-1} \\ & & & & -\sigma_n^{-1} \end{pmatrix}}_{n}.$$

The distribution of the rv $Y \stackrel{d}{=} S_{\boldsymbol{E}}$ is often referred to as the hypo-exponential distribution.

Example 5 (Mixed Erlang distributions)**.** It follows immediately from Example 4 that if all scale parameters are equal, then the parametrization

$$\boldsymbol{\alpha}_j = (1, \underbrace{0, \ldots, 0}_{n_j-1}) \qquad \text{and} \qquad \boldsymbol{T}_j = \underbrace{\begin{pmatrix} -\sigma_j^{-1} & \sigma_j^{-1} & & & \\ & \ddots & \ddots & & \\ & & & -\sigma_j^{-1} & \sigma_j^{-1} \\ & & & & -\sigma_j^{-1} \end{pmatrix}}_{n_j},$$

for $\sigma_j > 0$, $j = 1, \ldots, m$, n_j, $m \in \mathbb{N}$ introduces an Erlang-distributed rv, notationally Erl, with the pdf

$$f_{Erl}(y; n_j, \sigma_j) = \frac{y^{n_j - 1} \sigma_j^{-n_j}}{(n_j - 1)!} e^{-\sigma_j^{-1} y} \quad \text{for} \quad y \in \mathbb{R}_+.$$

In addition set $p_j \in [0,1]$ such that $\sum_{j=1}^{m} p_j = 1$, and let

$$\boldsymbol{\alpha} = (p_1\boldsymbol{\alpha}_1, \ldots, p_m\boldsymbol{\alpha}_m) \qquad \text{and} \qquad \boldsymbol{T} = \oplus_{j=1}^{m}\boldsymbol{T}_j$$

(the direct sum of matrices is defined in the Appendix), then $Y \sim PH(\boldsymbol{\alpha}, \boldsymbol{T})$ is distributed an m-components mixed Erlang distribution if it has pdf

$$f_Y(y) = \sum_{j=1}^{m} p_j f_{Erl}(y; n_j, \sigma_j) = \sum_{j=1}^{m} p_j \frac{y^{n_j-1}\sigma_j^{-n_j}}{(n_j-1)!} e^{-\sigma_j^{-1}y} \quad \text{for} \quad y \in \mathbb{R}_+.$$

One remarkable property of the class of the PH distributions is that they are dense in the class of all distributions with non-negative support, hence theoretically, they can approximate any such distribution arbitrarily well. Other important characteristics of the PH distributions are enumerated below [e.g., 14, for more details].

- Let $Y \sim PH(\boldsymbol{\alpha}, \boldsymbol{T})$, then the $h(\in \mathbb{N})$-th moment of the rv Y is

$$\mathbb{E}\left[Y^h\right] = h!\boldsymbol{\alpha}\left(-\mathbf{T}^{-1}\right)^h \mathbf{1}. \tag{2.20}$$

  Also, the coefficient of variation of the rv Y, denoted by c_Y, is given by

$$c_Y = \sqrt{\frac{2\boldsymbol{\alpha}(\boldsymbol{T})^{-2}\mathbf{1}}{(\boldsymbol{\alpha}\widehat{\otimes}\boldsymbol{\alpha})(\boldsymbol{T}\widehat{\otimes}\boldsymbol{T})^{-1}\mathbf{1}} - 1}, \tag{2.21}$$

  where the Kronecker product of matrices is defined in the Appendix.

- Let $Y \sim PH(\boldsymbol{\alpha}, \boldsymbol{T})$ with order $d \in \mathbb{N}$ (i.e., the row dimension of the matrix $\boldsymbol{T}$ is d), then the $h(\in \mathbb{N})$-th order size-biased variant of the rv Y, denoted by $Y^{[h]}$, is distributed

PH with the following representation

$$\boldsymbol{\alpha}^{[h]} = \big(\boldsymbol{\alpha}(\boldsymbol{T}^{-1})^h / \boldsymbol{\alpha}(\boldsymbol{T}^{-1})^h \mathbf{1}, \underbrace{0, \ldots, 0}_{hd}\big) \quad \text{and} \quad \boldsymbol{T}^{[h]} = \underbrace{\begin{pmatrix} \boldsymbol{T} & -\boldsymbol{T} & & & \\ & \boldsymbol{T} & -\boldsymbol{T} & & \\ & & \ddots & \ddots & \\ & & & & \boldsymbol{T} \end{pmatrix}}_{(h+1)d} \tag{2.22}$$

This means that the class of PH distributions is closed under size-biasing.

- Let $Y_i \sim PH(\boldsymbol{\alpha}_i, \boldsymbol{T}_i)$, $i = 1, \ldots, n$ be mutually independent rv's, and let $S_{\mathbf{Y}} = Y_1 + \cdots + Y_n$, then it can be shown that the rv $S_{\mathbf{Y}}$ is distributed phase-type with parametrization $(\boldsymbol{\alpha}_S, \boldsymbol{T}_S)$, such that, for $\tau_i = \prod_{j=i+1}^{n} \alpha_j^0$ with $\alpha_j^0 = 1 - \boldsymbol{\alpha}_j \mathbf{1}$,

$$\boldsymbol{\alpha}_S = (\boldsymbol{\alpha}_n, \tau_{n-1}\boldsymbol{\alpha}_{n-1}, \tau_{n-2}\boldsymbol{\alpha}_{n-2}, \ldots, \tau_1 \boldsymbol{\alpha}_1) \tag{2.23}$$

and

$$\boldsymbol{T}_S = \begin{pmatrix} \boldsymbol{T}_n & \boldsymbol{t}_n\boldsymbol{\alpha}_{n-1} & \frac{\tau_{n-2}}{\tau_{n-1}}\boldsymbol{t}_n\boldsymbol{\alpha}_{n-2} & \cdots & \cdots & \frac{\tau_1}{\tau_{n-1}}\boldsymbol{t}_n\boldsymbol{\alpha}_1 \\ & \boldsymbol{T}_{n-1} & \boldsymbol{t}_{n-1}\boldsymbol{\alpha}_{n-2} & \frac{\tau_{n-3}}{\tau_{n-2}}\boldsymbol{t}_{n-1}\boldsymbol{\alpha}_{n-3} & \cdots & \frac{\tau_1}{\tau_{n-2}}\boldsymbol{t}_{n-1}\boldsymbol{\alpha}_1 \\ & & \ddots & \ddots & & \vdots \\ & & & \ddots & \ddots & \vdots \\ & & & & \boldsymbol{T}_2 & \boldsymbol{t}_2\boldsymbol{\alpha}_1 \\ & & & & & \boldsymbol{T}_1 \end{pmatrix}. \tag{2.24}$$

Hence the class of PH distributions is closed under convolutions.

- Let $Y_i \sim PH(\boldsymbol{\alpha}_i, \boldsymbol{T}_i),\ i = 1, \ldots, n$, and let $W_{\mathbf{Y}} = \min_{i=1,\ldots,n} Y_i$, then the rv $W_{\mathbf{Y}}$ is distributed phase-type with parametrization $(\boldsymbol{\alpha}_W, \boldsymbol{T}_W)$, where

$$\boldsymbol{\alpha}_W = \widehat{\otimes}_{i=1}^{n} \boldsymbol{\alpha}_i \quad \text{as well as} \quad \boldsymbol{T}_W = \widehat{\oplus}_{i=1}^{n} \boldsymbol{T}_i \tag{2.25}$$

(the Kronecker sum and product of matrices are defined in the Appendix). This implies that the class of PH distributions is closed under the minimum operation.

2.4 Multiplicative background risk models with the idiosyncratic risk factors distributed phase-type

In this section, we put forward a special subclass of the general MBR structures, in which the idiosyncratic risk factors are distributed phase-type. We study a wide range of distributional properties and risk management applications of what we call the phase-type MBR models,

and we demonstrate that they inherit the tractability of the exponential MBR models while also allowing for considerably more diversity in dependence modelling.

2.4.1 Definition and distributional properties

Definition 7. Recall the multiplicative background risk model stochastic representation, that is

$$\boldsymbol{X} \;=\; (X_1, \ldots, X_n) = (RY_1, \ldots RY_n)\,, \tag{2.26}$$

where the rv's $Y_1, \ldots, Y_n$ are independent mutually and also on the rv R. If in addition, the rv's Y_i are distributed phase-type with the respective representations $(\boldsymbol{\alpha}_i, \boldsymbol{T}_i)$, $i = 1, \ldots, n$, then we say that the rv $\mathbf{X}$ admits the phase-type multiplicative background risk model stochastic representation, succinctly, PH-MBR.

We have already emphasized that the Laplace transform plays an important role when exploring the properties of the MBR models with exponentially-distributed idiosyncratic risk factors. In the case of the PH-MBR models, we need to introduce a slightly more general notion of Laplace transforms with matrix-valued arguments. To this end, recall that, for a rv X and the corresponding cdf F, the Laplace transform $\hat{F}$ is defined as

$$\hat{F}(s) = \mathbb{E}\left[e^{-sX}\right] = \int_0^\infty e^{-sx} dF(x),\ Re(s) \geq 0.$$

Also, recall that the alternating sign $h(\in \mathbb{Z})$-th order (anti-)derivative of the Laplace transform can be computed via [97]

$$\hat{F}^{(h)}(s) = \mathbb{E}\left[X^h e^{-sX}\right] = \begin{cases} (-1)^h \frac{\partial^h}{\partial s^h} \hat{F}(s), & \text{for } h = 1, 2, \ldots \\ \\ \int_s^\infty \frac{(t-s)^{-h-1}}{(-h-1)!} \hat{F}(t) \mathrm{d}t, & \text{for } h = -1, -2, \ldots \end{cases}. \tag{2.27}$$

We note that the existence of the anti–derivatives of the Laplace transform $\hat{F}(s)$ depends on

the left tail behaviour of the rv X, as is seen from the following

$$\begin{aligned}\hat{F}^{(h)}(s) &= \int_s^{\infty} \frac{1}{(-h-1)!} \sum_{j=0}^{-h-1} \binom{-h-1}{j} t^j (-s)^{-h-1-j} \hat{F}(t)\mathrm{d}t \\ &= \frac{1}{(-h-1)!} \sum_{j=0}^{-h-1} \binom{-h-1}{j} (-s)^{-h-1-j} \int_s^{\infty} t^j \hat{F}(t)\mathrm{d}t,\end{aligned}$$

where $Re(s) \geq 0$ and $h = -1, -2, \ldots$

Further let $\mathbf{A}$ be a negatively-defined square matrix having $m \in \mathbb{N}$ eigenvalues $\lambda_1, \ldots, \lambda_m$ and $Re(\lambda_i) < 0$ for $i = 1, \ldots, m$, then the aforementioned Laplace transform with matrix-valued arguments is introduced as following

$$\hat{F}(-s\mathbf{A}) = \mathbb{E}\left[e^{s\mathbf{A}X}\right], \ Re(s) \geq 0$$

and, more generally in a fashion similar to (2.27) and for $h \in \mathbb{Z}$,

$$\hat{F}^{(h)}(-s\mathbf{A}) = \mathbb{E}\left[X^h e^{-tX}\right]\big|_{t=-s\mathbf{A}}, \ Re(s) \geq 0.$$

Notably, most of the results in the remaining sections of the paper are formulated in terms of the transform $\hat{F}^{(h)}$, and so the way to compute it is of great importance. To this end, note that according to the Appendix, the matrix $\boldsymbol{A}$ can be written via its Jordan form as $\boldsymbol{A} = \boldsymbol{V_A}\ \boldsymbol{J}\ \boldsymbol{V_A}^{-1}$. Moreover, for $j = 1, \ldots, m$, let $k_j = \dim(\mathfrak{E}_{\lambda_j})$ be the dimension of the eigenspace associated with the eigenvalue λ_j, and set

$$\hat{\boldsymbol{F}}^{(h)}(s; \lambda_j) = \hat{\boldsymbol{F}}^{(h)}(s; \lambda_j; \dim(\mathfrak{E}_{\lambda_j}))$$

$$= \underbrace{\begin{pmatrix} \hat{F}^{(h)}(-s\lambda_j) & s\hat{F}^{(h+1)}(-s\lambda_j) & \cdots & \cdots & s^{(k_j-1)}\frac{\hat{F}^{(h+k_j-1)}(-s\lambda_j)}{(k_j-1)!} \\ & \hat{F}^{(h)}(-s\lambda_j) & s\hat{F}^{(h+1)}(-s\lambda_j) & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \hat{F}^{(h)}(-s\lambda_j) & s\hat{F}^{(h+1)}(-s\lambda_j) \\ & & & & \hat{F}^{(h)}(-s\lambda_j) \end{pmatrix}}_{k_j=\dim(\mathfrak{E}_{\lambda_j})} \tag{2.28}$$

Then according to Equation (2.48) in the Appendix (the same section also has the definition of the direct sum of matrices), we have, for $h \in \mathbb{Z}$,

$$\hat{F}^{(h)}(-s\mathbf{A}) \;=\; \boldsymbol{V_A}\ \left(\oplus_{j=1}^{m}\hat{\boldsymbol{F}}^{(h)}(s;\lambda_j)\right)\ \boldsymbol{V_A}^{-1},\ Re(s) \geq 0. \tag{2.29}$$

Remark 1. As the matrix $\boldsymbol{A}$ is assumed to be negatively-defined, all the associated eigenvalues have negative real parts and $\mathrm{Re}(-x\lambda_j) > 0$. Along with the analyticity property of the Laplace transform, this implies that the entries in (2.28) are well defined for $h \geq 0$. Special attention must be paid for $h < 0$.

Proposition 5. *Assume that the rv* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *admits the PH-MBR stochastic representation as per Definition 7, and let the idiosyncratic risk factors have representations* $(\boldsymbol{\alpha}_i, \boldsymbol{T}_i)$, $i = 1, \ldots, n$. *Then we have the following expressions for, respectively, the marginal ddf's and pdf's of the coordinates of the rv* $\mathbf{X}$

$$\overline{F}_{X_i}(x) = \boldsymbol{\alpha}_i\ \hat{F}_{R^{-1}}(-x\boldsymbol{T_i})\ \mathbf{1} \tag{2.30}$$

and

$$f_{X_i}(x) = \boldsymbol{\alpha}_i\ \hat{F}^{(1)}_{R^{-1}}(-x\boldsymbol{T}_i)\ \boldsymbol{t}_i \tag{2.31}$$

for $x \in \mathbb{R}_+$, $i = 1, \ldots, n$ *and* $\boldsymbol{t}_i = -\boldsymbol{T}_i \mathbf{1}$.

Furthermore, for $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$ *set* $\boldsymbol{U}_{\boldsymbol{x}} = \widehat{\oplus}_{i=1}^n x_i \mathbf{T}_i$, *then the joint ddf and pdf of the rv* $\boldsymbol{X}$ *are given by, respectively,*

$$\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; \hat{\boldsymbol{\alpha}}\; \hat{F}_{R^{-1}}(-\boldsymbol{U}_{\boldsymbol{x}})\; \mathbf{1}$$

and

$$f_{\boldsymbol{X}}(\boldsymbol{x}) \;=\; \hat{\boldsymbol{\alpha}}\; \hat{F}^{(n)}_{R^{-1}}(-\boldsymbol{U}_{\boldsymbol{x}})\; \hat{\boldsymbol{t}},$$

where $\hat{\boldsymbol{\alpha}} = \widehat{\otimes}_{i=1}^n \boldsymbol{\alpha}_i$ *and* $\hat{\boldsymbol{t}} = \widehat{\otimes}_{i=1}^n \boldsymbol{t}_i$ *(we refer to the Appendix for the definitions of the Kronecker sum and product of matrices).*

Proof. For $x \in [0,\ \infty)$ and evoking Proposition 3 with Equation (2.18), we have

$$\overline{F}_{X_i}(x) = \int_0^\infty \mathbb{P}[Y_i > rx] dF_{R^{-1}}(r) = \boldsymbol{\alpha}_i \left(\int_0^\infty e^{rx\boldsymbol{T}_i} dF_{R^{-1}}(r) \right) \mathbf{1},$$

as required. In a very similar fashion, but this time for $\mathbf{x} = (x_1, \ldots, x_n)' \in [0,\ \infty)^n$,

$$\begin{aligned} \overline{F}_{\mathbf{X}}(x_1, \ldots, x_n) \;&=\; \int_0^\infty \prod_{i=1}^n \mathbb{P}[Y_i > rx_i] dF_{R^{-1}}(r) = \int_0^\infty \widehat{\otimes}_{i=1}^n\; \boldsymbol{\alpha}_i e^{r(x_i \boldsymbol{T}_i)} \mathbf{1}\; dF_{R^{-1}}(r) \\ &=\; \widehat{\otimes}_{i=1}^n\; \boldsymbol{\alpha}_i \left(\int_0^\infty \exp\{r \widehat{\oplus}_{i=1}^n (x_i \boldsymbol{T}_i)\}\; dF_{R^{-1}}(r) \right) \mathbf{1}. \end{aligned}$$

The corresponding pdf's are readily obtained by routine differentiation. This completes the proof of the proposition. □

The following result about the marginal/joint moments follows immediately from Proposition 3 and Equation (2.20), thus the proof is omitted.

Proposition 6. *Assume that the rv* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *admits the PH-MBR stochastic representation as per Definition 7 with the rv's* Y_i *having representations* $(\boldsymbol{\alpha}_i, \boldsymbol{T}_i)$, $i = 1, \ldots, n$. *Also assume that the* $h(\in \mathbb{N})$*-th moment of the rv* R *is finite, i.e.,* $\mathbb{E}[R^h] < \infty$. *Then the*

h-th marginal moment of $X_i,\ i = 1, \ldots, n$ *is given by*

$$\mathbb{E}\left[X_i^h\right] = \mathbb{E}\left[R^h\right]\ h!\boldsymbol{\alpha}_i(-\boldsymbol{T}_i^{-1})^h\mathbf{1}.$$

In addition, for $1 \leq i \neq j \leq n,\ h_i, h_j \in \mathbb{N}$, *assume that* $\mathbb{E}\left[R^{h_i+h_j}\right] < \infty$. *Then the* (h_i, h_j)*-th order joint moment of the random pair* $(X_i,\ X_j)$ *is given by*

$$\mathbb{E}\left[X_i^{h_i}X_j^{h_j}\right] = (-1)^{h_i+h_j}\ \mathbb{E}\left[R^{h_i+h_j}\right]\ h_i!h_j!\ \hat{\boldsymbol{\alpha}}\ \left[\boldsymbol{T}_i^{-h_i}\widehat{\otimes}\boldsymbol{T}_j^{-h_j}\right]\ \mathbf{1}, \tag{2.32}$$

and in particular

$$Cov(X_i, X_j) \quad = \quad Var(R)\ \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{T}})^{-1}\mathbf{1}, \tag{2.33}$$

where $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_i\widehat{\otimes}\boldsymbol{\alpha}_j$ *and* $\hat{\boldsymbol{T}} = \boldsymbol{T}_i\widehat{\otimes}\boldsymbol{T}_j$.

The following corollary is an immediate consequence of Proposition 6 (also, Proposition 1).

Corollary 5. *Consider the rv* $\mathbf{X} = (X_1, \ldots, X_n)$ *that admits the PH-MBR stochastic representation, then the Pearson coefficient of correlation between the rv's* X_i *and* X_j, *where* $1 \leq i \neq j \leq n$, *is given by*

$$Corr(X_i,\ X_j) = \prod_{k\in\{l,m\}}\left[\left(1 + \left(\frac{2\boldsymbol{\alpha}_k(\boldsymbol{T}_k)^{-2}\mathbf{1}}{(\boldsymbol{\alpha}_k\widehat{\otimes}\boldsymbol{\alpha}_k)(\boldsymbol{T}_k\widehat{\otimes}\boldsymbol{T}_k)^{-1}\mathbf{1}} - 1\right)(1 + c^{-2})\right)\right]^{-1/2}, \tag{2.34}$$

where c *is the variation coefficient of the rv* R.

Expression (2.34) obtained above is somewhat cumbersome. Remarkably, the upper bound for the Pearson coefficient of correlation is contrarily very simple. The bound is stated in the following proposition, with an important lemma reported beforehand. The lemma is quite well-known, and it asserts that the Erlang distribution is the least variable member of the class of PH distributions, when the variability is measured by means of the variance.

Lemma 1 (David and Larry, 1987)**.** *Let $Y \sim PH(\boldsymbol{\alpha}, \boldsymbol{T})$ with order $d \in \mathbb{N}$ (i.e., the row dimension of the matrix $\boldsymbol{T}$ is d), then the following bound holds for the coefficient of variation of Y, succinctly c_Y,*

$$c_Y \geq d^{-1/2}.$$

The bound becomes an equality when $Y \sim Erl(d,\ \sigma)$, that is the rv Y is distributed Erlang with the shape parameter $d \in \mathbb{N}$ and scale parameter $\sigma \in \mathbb{R}_+$.

For $1 \leq i \neq j \leq n$, assume that the rv's Y_i and Y_j are distributed phase-type with the orders $d_i \in \mathbb{N}$ and $d_j \in \mathbb{N}$, respectively (see, Definition 6). Set

$$d_{(i,j)} = \max(d_i, d_j). \tag{2.35}$$

then we have that the PH-MBR models are able to cover the entire range of positive dependence when it is measured by the Pearson coefficient of correlation. This is conveyed in the next proposition.

Proposition 7. *For $1 \leq i \neq j \leq n$, let the rv (X_i, X_j) admit the PH-MBR stochastic representation as per Definition 7, and assume that $\mathbb{E}[R] < \infty$. Then, for the Pearson coefficient of linear correlation, the following bound holds*

$$Corr[X_i, X_j] \leq \frac{1}{d_{(i,j)}^{-1}(1 + c^{-2}) + 1}, \tag{2.36}$$

where $d_{(i,j)}$ is defined in (2.35) and c denotes the coefficient of variation of the rv R. Moreover, the equality in (2.36) is attained when the rv's Y_i and Y_j are both identically distributed as an Erlang with the shape parameter $d \in \mathbb{N}$.

It is evident from (2.36) that the Pearson coefficient of linear correlation has the limiting value of one, when the parameter d goes to infinity. This means that within the context of the PH-MBR models, the maximal attainable value of the Pearson coefficient of correlation can be arbitrarily close to one, as opposed to the case of the Exponential MBR models, in which the upper-bound for the Pearson coefficient of correlation is one-half (see, Corollary 1).

Let $S_{\boldsymbol{X}} = X_1 + \cdots + X_n$, $W_{\boldsymbol{X}} = \min_{i=1,\ldots,n} X_i$, and $M_{\boldsymbol{X}} = \max_{i=1,\ldots,n} X_i$. We conclude this section by deriving the distributions of the just-mentioned rv's.

Proposition 8. *Let the rv $\boldsymbol{X} = (X_1, \ldots, X_n)$ admit the PH-MBR stochastic representation with the rv's Y_i having representations $(\boldsymbol{\alpha_i}, \boldsymbol{T}_i)$, $i = 1, \ldots, n$. Then the corresponding aggregate rv $S_{\boldsymbol{X}}$ and the minimum rv $W_{\boldsymbol{X}}$ have the ddf's and pdf's given, respectively, by*

$$\overline{F}_{\diamond}(x) \;=\; \boldsymbol{\alpha}_{\diamond}\, \hat{F}_{R^{-1}}(-x\boldsymbol{T}_{\diamond})\, \mathbf{1} \quad \textit{for} \quad x \in [0,\ \infty),$$

and

$$f_{\diamond}(x) \;=\; \boldsymbol{\alpha}_{\diamond}\, \hat{F}^{(1)}_{R^{-1}}(-x\boldsymbol{T}_{\diamond})\, \boldsymbol{t}_{\diamond} \quad \textit{for} \quad x \in \mathbb{R}_+,$$

where '$\diamond$' stands for either one of '$S_{\mathbf{X}}$' or '$W_{\mathbf{X}}$', $\boldsymbol{\alpha}_{\diamond}$ and $\boldsymbol{T}_{\diamond}$ are defined correspondingly in (2.23) - (2.25), and $\boldsymbol{t}_{\diamond} = -\boldsymbol{T}_{\diamond}\mathbf{1}$.

Finally, the next assertion establishes the distribution of the maximum rv $M_{\boldsymbol{X}}$. Its proof hinges on the inclusion-exclusion principle and Proposition 8. Similar derivations can be found in, e.g., [106, 114].

Proposition 9. *Let the rv $\boldsymbol{X} = (X_1, \ldots, X_n)$ admit the PH-MBR stochastic representation with the rv's Y_i having representations $(\boldsymbol{\alpha_i}, \boldsymbol{T}_i)$, $i = 1, \ldots, n$. Then the ddf of the maximum rv $M_{\boldsymbol{X}}$ can be computed via*

$$\overline{F}_M(x) = \sum_{\mathcal{I} \subseteq \{1,\ldots,n\}} (-1)^{|\mathcal{I}|-1} \overline{F}_{W_{\mathcal{I}}}(x) \quad \textit{for} \quad x \in [0,\ \infty), \tag{2.37}$$

where $W_{\mathcal{I}} = \min_{i \in \mathcal{I} \subseteq \{1,\ldots,n\}} X_i$.

It is intuitive and well-documented that the PH distributions are closed under the maximum operation [e.g., 14]. Hence one could argue that it is possible to find the PH representation of the rv $M_{\boldsymbol{X}}$, and therefore, the parameters of its ddf $\overline{F}_M$. This is indeed so, but the results are not easy to present, and because of little connection to our paper, we have chosen to go as far as Proposition 9.

2.4.2 Applications to risk management

We now derive explicit expressions for some risk measures of practical importance within the context of the PH-MBR models. At the outset, It should be noted that factor models play an important, if not central, role in a variety of applications. Herein we refer to [42] - for asset pricing, [105] - for portfolio optimization, [57] - for credit risk. Also, the noble arbitrage pricing theory [100] and capital asset pricing model [e.g., 78], as well as the recent weighted insurance pricing model [51, 55] all hinge on factor models.

Recall that, for a non-negative rv $X_i,\ i = 1, \ldots, n$ with finite mean, the Value-at-Risk (VaR) risk measure is defined as

$$\mathrm{VaR}_p(X_i) = \inf\{x \geq 0 :\ F_{X_i}(x) \geq p\},\ p \in [0,\ 1),$$

and the conditional tail expectation risk measure of order $k \in \mathbb{N}$ is defined as

$$\mathrm{CTE}_p^k(X_i) = \mathbb{E}\left[X_i^k |\ X_i > VaR_p[X_i]\right],\ p \in [0,\ 1)$$

[e.g., 71].

Proposition 10. *Consider the PH-MBR portfolio* $\boldsymbol{X} = (X_1, \ldots, X_n) = (RY_1, \ldots, RY_n)$, *that is let* $Y_i \sim PH(\boldsymbol{\alpha}_i, \boldsymbol{T}_i)$, $i = 1, \ldots, n$. *Also, assume that* $\mathbb{E}[X_i^k] < \infty,\ i = 1, \ldots, n$. *Then the* $k(\in \mathbb{N})$*-th order conditional moment of the* i*-th risk component is formulated, for* $p \in [0,\ 1)$, *as*

$$\mathrm{CTE}_p^k(X_i) \quad = \quad \frac{(-1)^k k!}{1-p}\left(\boldsymbol{\alpha}_i \widehat{\otimes} \boldsymbol{\alpha}_i^{[k]}\right)\left(\boldsymbol{T}_i^{-k} \widehat{\otimes} \hat{F}_{R^{-1}}^{(-k)}\left(-VaR_p(X_i)\boldsymbol{T}_i^{[k]}\right)\right)\mathbf{1}, \tag{2.38}$$

where $\hat{F}_{R^{-1}}^{(-k)}$ *can be computed via (2.29), the parameters* $\boldsymbol{\alpha}_i^{[k]}$ *and* $\boldsymbol{T}_i^{[k]}$ *are defined in (2.22), and the definition of the Kronecker product of matrices is provided in the Appendix.*

Proof. For $i = 1, \ldots, n,\ k \in \mathbb{N}$ and $p \in [0,\ 1)$, we have

$$\mathrm{CTE}_p^k(X_i) \quad \overset{(1)}{=} \quad \frac{\mathbb{E}\left[X_i^k\right]}{1-p}\mathbb{P}\left(X_i^{[k]} > VaR_p(X_i)\right)$$

$$
\begin{aligned}
&\overset{(2)}{=} \frac{\mathbb{E}[Y_i^k]\mathbb{E}[R^k]}{1-p}\mathbb{P}\left(R^{[k]}Y_i^{[k]} > VaR_p(X_i)\right) \\
&= \frac{\mathbb{E}[Y_i^k]\mathbb{E}[R^k]}{1-p}\int_0^\infty \mathbb{P}\left(Y_i^{[k]} > r^{-1}VaR_p(X_i)\right) dF_R^{[k]}(r) \\
&= \frac{\mathbb{E}[Y_i^k]}{1-p}\int_0^\infty \mathbb{P}\left(Y_i^{[k]} > rVaR_p(X_i)\right) r^{-k} dF_{R^{-1}}(r) \\
&\overset{(3)}{=} \frac{\mathbb{E}[Y_i^k]}{1-p}\boldsymbol{\alpha}_i^{[k]}\left(\int_0^\infty e^{r(VaR_p(X_i)\boldsymbol{T}_i^{[k]})} r^{-k} dF_{R^{-1}}(r)\right)\mathbf{1} \\
&= \frac{\mathbb{E}[Y_i^k]}{1-p}\boldsymbol{\alpha}_i^{[k]}\hat{F}_{R^{-1}}^{(-k)}\left(-VaR_p(X_i)\boldsymbol{T}_i^{[k]}\right)\mathbf{1},
\end{aligned}
$$

where '$\overset{(1)}{=}$', '$\overset{(2)}{=}$' and '$\overset{(3)}{=}$' hold because of, respectively, Proposition 4, Corollary 2 and the fact that the class of PH distributions is closed under size-biasing of arbitrary order with the corresponding SB representation given in (2.22). Note that $\hat{F}_{R^{-1}}^{(-k)}\left(-VaR_p(X_i)\boldsymbol{T}_i^{[k]}\right)$ can be readily computed via Equation (2.29). This completes the proof of the proposition. □

As a rule, insurers are concerned with the economic capital due to the aggregate risk, $S_{\boldsymbol{X}} = \sum_{i=1}^n X_i$, rather than due to its stand-alone constituents. Since, conditionally on the rv R, the rv $S_{\boldsymbol{X}}$ is distributed PH with the representation $(\boldsymbol{\alpha}_S,\ \mathbf{T}_S)$, where the parameters $\boldsymbol{\alpha}_S$ and $\mathbf{T}_S$ are given in, correspondingly, (2.23) and (2.24), we have the following corollary immediately.

Corollary 6. *Within the set-up of Proposition 10, the k-th order CTE of the aggregate risk rv is given by*

$$
\mathrm{CTE}_p^k(S_{\boldsymbol{X}}) = \frac{(-1)^k k!}{1-p}\left(\boldsymbol{\alpha}_S \widehat{\otimes} \boldsymbol{\alpha}_S^{[k]}\right)\left(\boldsymbol{T}_S^{-k}\widehat{\otimes}\hat{F}_{R^{-1}}^{(-k)}\left(-VaR_p\left(S_{\mathbf{X}}\right)\boldsymbol{T}_S^{[k]}\right)\right)\mathbf{1}, \quad (2.39)
$$

where the parameters $\boldsymbol{\alpha}_S$, $\boldsymbol{T}_S$ and $\boldsymbol{\alpha}_S^{[k]}$, $\boldsymbol{T}_S^{[k]}$ are defined in, respectively, (2.23), (2.24) and (2.22), and $\hat{F}_{R^{-1}}^{(-k)}(\cdot)$ can be readily computed with the help of Equation (2.29).

Remark 2. It is easy to see that the expression for the CTE risk measure of the minimum risk rv, $W_{\boldsymbol{X}} = \min_{i=1,\ldots,n} X_i$, has the same form as (2.39). Indeed, when of interest, we obtain the value of $\mathrm{CTE}_p^k(W_{\boldsymbol{X}})$ directly from Corollary 6 by merely replacing the '$S_{\mathbf{X}}$' subscript

with the '$W_{\mathbf{X}}$' subscript in the right-hand-side of (2.39). In this case, the desired parameters $\boldsymbol{\alpha}_W$, $\boldsymbol{T}_W$ and $\boldsymbol{\alpha}_W^{[k]}$, $\boldsymbol{T}_W^{[k]}$ are defined in, respectively, (2.25) and (2.22).

We further study the contribution of the risk component $X_i,\ i=1,\ldots,n$ to the overall riskiness of the risk portfolio represented by the aggregate risk rv $S_{\mathbf{X}}$, that is we consider the risk capital allocation exercise next. Speaking briefly, the allocation problem is much more involved than the fairly well-studied problem of determining the aggregate risk capital. As a consequence, numerous papers have been devoted to solving this problem under very specific model and/or reference risk measure assumptions. For instance, in the case of the CTE, we refer to [47] for gamma distributions, [23] for phase-type distributions, [35] for elliptical distributions, [49] for Tweedie distributions, [114, 115] for skew-normal and Pareto distributions. We should also mention that the axiomatic approach to the notion of risk capital allocation is pursued in, e.g., [54, 69]. In addition, classes of optimal risk capital allocation rules are proposed in [36].

In the rest of this section, we develop explicit expressions for some tail-based risk capital allocation rules for the PH-MBR models. Given that in our context the systemic risk factor can have an arbitrary distribution with positive support and the idiosyncratic risk factors are distributed PH, and so can approximate fairly well any distribution with positive support, our results are quite general. We need some additional notations in order to proceed. First, let $S_{\mathbf{X}}=\sum_{j=1}^n X_j,\ S_{\mathbf{Y}}=\sum_{j=1}^n Y_j$ as before, and let $S_{\mathbf{X},-i}=\sum_{j=1,\ j\neq i}^n X_j$ and $S_{\mathbf{Y},-i}=\sum_{j=1,\ j\neq i}^n Y_j$. Second, we define, for $h_1,h_2\in\mathbb{N}_0$,

$$\mathbf{C}_{i,j}=\frac{h_1!(h_1+h_2-j)!}{(h_1-j)!}\left(\boldsymbol{\alpha}_{S_{\mathbf{Y},-i}}\ \widehat{\otimes}\ \boldsymbol{\alpha}_i\right)\times\left(\boldsymbol{T}_{S_{\mathbf{Y},-i}}^{-j}\ \widehat{\otimes}\ \boldsymbol{T}_i^{-(h_1+h_2-j)}\right)\mathbf{1},\ i=1,\ldots,n,\ j=0,\ldots,h_1,$$

where $\boldsymbol{\alpha}_{S_{\mathbf{Y},-i}}$ and $\boldsymbol{T}_{S_{\mathbf{Y},-i}}$ are given by (2.23) and (2.24). Lastly, we set

$$\boldsymbol{\alpha}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}=\left(\boldsymbol{\alpha}_i^{[h_1+h_2-j]},\ (1-\boldsymbol{\alpha}_i^{[h_1+h_2-j]}\mathbf{1})\ \boldsymbol{\alpha}_{S_{\mathbf{Y},-i}}^{[j]}\right)\tag{2.40}$$

and

$$\boldsymbol{T}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]} = \begin{pmatrix} \boldsymbol{T}_i^{[h_1+h_2-j]} & -\boldsymbol{T}_i^{[h_1+h_2-j]}\mathbf{1} \times \boldsymbol{\alpha}_{S_{\mathbf{Y},-i}}^{[j]} \\ \mathbf{0} & \boldsymbol{T}_{S_{\mathbf{Y},-i}}^{[j]} \end{pmatrix}. \tag{2.41}$$

We are now ready to state the next assertion.

Proposition 11. *Consider the PH-MBR portfolio $\boldsymbol{X} = (RY_1, \ldots, RY_n)$, that is let the risk factors Y_i have PH representations $(\boldsymbol{\alpha}_i, \boldsymbol{T}_i)$, $i = 1, \ldots, n$. For $h_1, h_2 \in \mathbb{N}_0$ and $p \in [0, 1)$, the higher order tail conditional joint moment of X_i and $S_{\boldsymbol{X}}$ is given by*

$$\mathbb{E}\left[S_{\boldsymbol{X}}^{h_1} X_i^{h_2} |\ S_{\boldsymbol{X}} > VaR_p(S_{\mathbf{X}})\right] = \frac{(-1)^{h_1+h_2}}{1-p} \sum_{j=0}^{h_1} \mathbf{C}_{i,j} \left[\boldsymbol{\alpha}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}\ \hat{F}_{R^{-1}}^{-(h_1+h_2)}\left(-VaR_p(S_{\mathbf{X}})\boldsymbol{T}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}\right)\ \mathbf{1}\right]. \tag{2.42}$$

Proof. According to Proposition 4, we only need to compute

$$I(s) = \int_0^\infty \mathbb{P}\left[(S_{\mathbf{Y},-k})^{[j]} + Y_k^{[h_1+h_2-j]} > r^{-1}s\right] \mathrm{d}F_{R^{[h_1+h_2]}}(r).$$

Note that by formulas (2.22)-(2.24), we know $(S_{\mathbf{Y},-k})^{[j]} + Y_k^{[h_1+h_2-j]} \sim PH(\boldsymbol{\alpha}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}, \boldsymbol{T}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]})$, where $\boldsymbol{\alpha}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}$ and $\boldsymbol{T}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}$ are defined as per (2.40) and (2.41). Thus

$$\begin{aligned} I(s) &= \mathbb{E}\left[R^{h_1+h_2}\right]^{-1} \boldsymbol{\alpha}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]} \left(\int_0^\infty r^{-(h_1+h_2)} \exp\left\{rs\boldsymbol{T}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}\right\} dF_{R^{-1}}(r)\right)\mathbf{1} \\ &= \mathbb{E}\left[R^{h_1+h_2}\right]^{-1} \boldsymbol{\alpha}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]} \left(\hat{F}_{R^{-1}}^{-(h_1+h_2)}\left(-s\boldsymbol{T}_{S_{\mathbf{Y},-i},Y_i}^{[j,h_1+h_2-j]}\right)\right)\mathbf{1}. \end{aligned}$$

The proof is now completed. □

Involved at first sight, expression (2.42) is explicit and recovers a notable number of special allocation rules tackled separately in the literature. For instance, put $h_1 = 1$, and the formula simplifies significantly to yield the $k(\in \mathbb{N})$-th order CTE-based allocation rule, of which the ubiquitous CTE-based allocation rule is a particular case.

Corollary 7. *Consider the set-up in Proposition 11, then we have*

$$\mathrm{CTE}_p^k(X_i, S_{\boldsymbol{X}}) = \frac{(-1)^k k!}{1-p}\left(\boldsymbol{\alpha}_i \widehat{\otimes} \boldsymbol{\alpha}^{[0,k]}_{S_{\mathbf{Y},-i},Y_i}\right)\left(\boldsymbol{T}_i^{-k} \widehat{\otimes}\ \hat{F}^{(-k)}_{R^{-1}}\left(-VaR_p(S_{\mathbf{X}})\boldsymbol{T}^{[0,k]}_{S_{\mathbf{Y},-i},Y_i}\right)\right)\mathbf{1},\ i=1,\ldots,n \tag{2.43}$$

where $\boldsymbol{\alpha}^{[0,k]}_{S_{\mathbf{Y},-i},Y_i}$ *and* $\boldsymbol{T}^{[0,k]}_{S_{\mathbf{Y},-i},Y_i}$ *are defined as per (2.40) and (2.41).*

Another special allocation rule that can be easily calculated with the help of (2.42), is the tail covariance allocation proposed and studied in [48]. This is clearly so because, for $p \in [0,\ 1)$,

$$\mathrm{TCov}_p(X_i, S_{\mathbf{X}}) \quad = \quad \mathbb{E}(X_i S_{\boldsymbol{X}} | S_{\boldsymbol{X}} > VaR_p[S_{\mathbf{X}}]) - \mathrm{CTE}_p(S_{\boldsymbol{X}})\mathrm{CTE}_p(X_i, S_{\boldsymbol{X}}),$$

and each term in the left-hand-side can be computed evoking Proposition 11. The tail covariance allocation is additive as is easily seen from

$$\sum_{i=1}^{n} \mathrm{TCov}_p(X_i, S_{\mathbf{X}}) = \mathrm{Var}_p(S_{\mathbf{X}}),\ p \in [0,\ 1).$$

2.5 Numerical illustration

In this section, we elucidate the implications of the theoretical findings presented heretofore. To this end we consider three explanatory risk portfolios, all of which are particular cases of the PH-MBR structures, proposed and studied in Section 2.4. Namely, we open the discussion with an example of a risk portfolio that admits the stochastic representation of the exponential MBR models [e.g., 28, 114, for the choice of parameters]. The arguably simplest and best studied special case of the class of PH-MBR models [e.g., 4, 27, and references therein, in addition to the two other just-mentioned references], the exponential MBR structures serve as an important benchmark and confirm that our method is able to recover the existing results promptly and accurately. Then the second and third illustrative risk portfolios provide insights about the implications of the change in the distributional assumptions on, respectively, the systemic and idiosyncratic risk factors. On one hand, it is confirmed that varying the distribution of the systemic risk factor within the context of the exponential MBR structures has no effect on the order of the allocated risk capitals.

That is, the higher the value of the scale parameter of the idiosyncratic risk factor is, the more risk capital the corresponding risk component requires irrespective of the choice of the distribution of the systemic risk factor. On the other hand, it is demonstrated that a slight departure from the exponentiality of the idiosyncratic risk factors brings much more uncertainty to the allocation exercise, and, as a result, the orders of the allocated risk capitals become less predictable.

2.5.1 Portfolio 1

Consider a risk portfolio that has ten risk components and admits the stochastic representation $\boldsymbol{\mathcal{E}} = (\mathcal{E}_1, \ldots, \mathcal{E}_{10}) = (RE_1, \ldots, RE_{10})$, where the idiosyncratic risk factors E_i are distributed exponentially with the scale parameters $\sigma_i,\ i = 1, \ldots, 10$, and the systemic risk factor R is such that the rv R^{-1} is distributed gamma with the shape parameter $\alpha \in \mathbb{R}_+$ and unit scale parameter. We note in passing that this set-up is obviously a special case of the PH-MBR model (Definition 7), where the PH representations of the idiosyncratic risk factors are $(\boldsymbol{\alpha}_i,\ \boldsymbol{T}_i)$ with $\boldsymbol{\alpha}_i \equiv (1)$ and $\boldsymbol{T}_i = (-\sigma_i^{-1})$. The resulting risk components are distributed Pareto of the second kind [e.g., 8].

The CTE-based risk capital allocations in the context of the aforementioned exponential MBR structure were studied in [28] and [114] by utilizing the theory of divided differences and certain recursive techniques, respectively. Table 2.1 presents a comparative analysis of the results of [28] and [114] versus explicit expressions (2.39) and (2.43), developed in Corollaries 6 and 7, correspondingly. The subtle differences are due to the deviations in the values of the Value-at-Risk risk measure when computed in MS Excel [28, 114] versus MATLAB's '`vpasolve`' function (this paper).

To realize the potential of Corollaries 6 and 7 to further extent, we modify the shape parameter of the rv R^{-1} to $\alpha = 3$ and leave the unit scale parameter unchanged; this warrants that the TCov risk capital allocation rule is well-defined and finite. Table 2.2 displays the values of the $\mathrm{CTE}_{0.95}$ and $\mathrm{TCov}_{0.95}$ risk measures, along with the allocation rules based on them. The proportional contributions are reported in brackets.

i	1	2	3	4	5	6	7	8	9	10
σ_i	0.32	0.94	0.16	0.47	0.73	0.25	0.61	1.02	0.22	0.81
PH approach	5.7134	17.5443	2.8202	8.4883	13.4330	4.4389	11.1301	19.1368	3.8969	14.9872
Benchmark	5.7234	17.5752	2.8252	8.5031	13.4565	4.4467	11.1497	19.1705	3.9037	15.0135

Table 2.1: $\mathrm{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}],\ i = 1, \ldots, 10$ for the exponential MBR under the assumption $R^{-1} \sim Ga(1.5, 1)$.

i	1	2	3	4	5	6	7	8	9	10
σ_i	0.32	0.94	0.16	0.47	0.73	0.25	0.61	1.02	0.22	0.81
$\mathbb{E}[X_i]$	0.1600	0.4700	0.0800	0.2350	0.3650	0.1250	0.3050	0.5100	0.1100	0.4050
	(5.79%)	(17.00%)	(2.89%)	(8.50%)	(13.20%)	(4.52%)	(11.03%)	(18.44%)	(3.98%)	(14.65%)
$\mathrm{CTE}_{0.95}[X_i]$	0.9829	2.8873	0.4915	1.4437	2.2423	0.7679	1.8737	3.1331	0.6758	2.4880
	(5.79%)	(17.00%)	(2.89%)	(8.50%)	(13.20%)	(4.52%)	(11.03%)	(18.44%)	(3.98%)	(14.65%)
$\mathrm{SD}_{0.95}[X_i]$	0.7522	2.2097	0.3761	1.1049	1.7161	0.5877	1.4340	2.3978	0.5172	1.9041
	(5.79%)	(17.00%)	(2.89%)	(8.50%)	(13.20%)	(4.52%)	(11.03%)	(18.44%)	(3.98%)	(14.65%)
$\mathrm{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}]$	0.6293	2.1295	0.3025	0.9581	1.5796	0.4833	1.2846	2.3502	0.4222	1.7841
	(5.28%)	(17.86%)	(2.54%)	(8.04%)	(13.25%)	(4.05%)	(10.77%)	(19.71%)	(3.54%)	(14.96%)
$\mathrm{TCov}_{0.95}[X_i, S_{\boldsymbol{X}}]$	3.3101	11.8523	1.5717	5.1026	8.6133	2.5282	6.9270	13.1869	2.2033	9.8031
	(5.08%)	(18.21%)	(2.41%)	(7.84%)	(13.23%)	(3.88%)	(10.64%)	(20.26%)	(3.38%)	(15.06%)

Table 2.2: Some risk measures and the allocation rules based on them for an exponential MBR model with $R^{-1} \sim Ga(3, 1)$.

.

Let $pr_i = \mathbb{E}[X_i]/\mathbb{E}[S_{\mathbf{X}}]$ denote the proportional expected contributions of the risk components $X_i,\ i = 1, \ldots, n$. In our example, $n = 10$. Then we immediately observe, for $\mathrm{SD}_p(X_i) = \sqrt{\mathrm{Var}_p(X_i)}$,

$$pr_i = \frac{\mathbb{E}[X_i]}{\sum_{j=1}^n \mathbb{E}[X_j]} = \frac{\mathrm{CTE}_{0.95}(X_i)}{\sum_{j=1}^n \mathrm{CTE}_{0.95}(X_j)} = \frac{\mathrm{SD}_{0.95}(X_i)}{\sum_{j=1}^n \mathrm{SD}_{0.95}(X_j)} = \frac{\sigma_i}{\sum_{j=1}^n \sigma_j},\ i = 1, \ldots, 10,$$

which is certainly true for an arbitrary prudence level $p \in [0,\ 1)$. The equality is not surprising at all since both the CTE_p and SD_p risk measures are positively homogeneous

and the allocation rules assumed are the simple proportional allocations. It is, however, somewhat less anticipated that the magnitudes of the contributions

$$\frac{\mathrm{CTE}_p(X_i, S_{\mathbf{X}})}{\mathrm{CTE}_p(S_{\mathbf{X}})} \text{ and } \frac{\mathrm{TCov}_p(X_i, S_{\mathbf{X}})}{\mathrm{Var}_p(S_{\mathbf{X}})}$$

are also very close to $pr_i,\ i = 1, \ldots, 10$.

Our consequent elaborations concern the non-trivial allocation rules, that is the ones based on the CTE and TCov risk measures. In a hope to close up on the underlying forces that drive the shapes of the allocation values, we refine the set-up by considering a sub-portfolio of risk components, and vary the distributions of the systemic risk factor first, and the idiosyncratic risk factors thereafter.

2.5.2 Portfolio 2

In the previous illustrative portfolio (exponential MBR model), we assumed $R^{-1} \sim \mathrm{Ga}(\alpha, 1)$ for $\alpha > 0$, therefore, the risk components had Pareto distributions and consequently tails that decay polynomially. Obviously this must not be the desired characteristic, and our framework is well-suited to deal with distinct tail-heaviness.

For instance, if a portfolio of risk components with lighter tails is to be modelled, then one choice for the distribution of the rv R^{-1} is beta, succinctly $R^{-1} \sim \mathrm{Beta}(\beta)$, for $\beta \in (0, 1)$, having the pdf

$$f_{R^{-1}}(z) \quad = \quad \frac{1}{\Gamma(1-\beta)\Gamma(\beta)} z^{-1}(z-1)^{-\beta},\ z > 1. \tag{2.44}$$

In this case, the risk components are distributed gamma [104] and hence have tails that decay exponentially [e.g., 39, for a rigorous discussion].

One more convenient choice for the distribution of the rv R^{-1} is the positive stable, succinctly $R^{-1} \sim \mathrm{PS}(\tau, \gamma)$ for $\tau \in (0, 1)$ and $\gamma > 0$, having the pdf

$$f_{R^{-1}}(z) \quad = \quad -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{(-1)^k \Gamma(k\tau + 1)}{k!} (\gamma/z)^{\tau k} \sin(\tau k \pi),\ z \in \mathbb{R}_+. \tag{2.45}$$

The resulting risk components are distributed Weibull [102], and so take the spot between the gamma and Pareto distributions in terms of the tail behaviour [e.g., 39, for a rigorous discussion].

Let $(\mathcal{E}_1, \ldots, \mathcal{E}_4) = (RY_1, \ldots, RY_4)$, where $Y_i \sim \mathrm{PH}(1, -\sigma_i^{-1})$, that is the idiosyncratic risk factors Y_i, are distributed exponentially with the scale parameters σ_i for $i = 1, \ldots, 4$, with $\sigma_1 = 0.73$, $\sigma_2 = 0.81$, $\sigma_3 = 0.94$, and $\sigma_4 = 1.02$. This sub-portfolio of Portfolio 1 contains the risk components that have the highest means. Also, let

$$R_1^{-1} \sim \mathrm{Ga}(3, 1), \ R_2^{-1} \sim \mathrm{PS}(0.5, 4), \ \text{and} \ R_3^{-1} \sim \mathrm{Beta}(0.5). \tag{2.46}$$

This choice of parameters warrants that all the involved risk measures and allocation rules are well-defined and finite, as well as

$$\mathbb{E}[R_1] = \frac{1}{\alpha - 1} = \frac{1}{2}, \ \mathbb{E}[R_2] = \frac{\Gamma(1 + 1/\tau)}{\gamma} = \frac{1}{2}, \ \text{and} \ \mathbb{E}[R_3] = \beta = \frac{1}{2},$$

and hence irrespective of the choice of the distribution of the systemic risk factor, we have $\mathbb{E}[X_i] = \sigma_i/2$ for $i = 1, \ldots, 4$. Unlike similar results that hinge on divided differences or recursions [28, 114], Corollaries 6 and 7 are able to readily compute the magnitudes of the allocation rules based on either one of the CTE or TCov risk measures, among others, irrespective of the choice of the distribution of the systemic risk factor. Our finding are summarized in Table 2.3, where $pr_i, \ i = 1, \ldots, 4$ denote, as before, the proportional expected contributions.

We observe that, similarly to the case of Portfolio 1, there is no significant difference between the proportional risk contributions due to the CTE and TCov allocation rules and the trivial proportional expected contributions. In this respect, varying the distributions of the systemic risk factor has not introduced major changes. That said, the aforementioned differences are the more profound, the lighter the tail of the distribution of the systemic risk factor is. Indeed, according to Proposition 1, heavier tails of the systemic risk factor imply higher positive Pearson correlations between risk components. This, along with the fact that for comonotonic risk components, the risk contributions due to the risk capital allocation

			$R_1^{-1} \sim \text{Ga}(3,1)$		$R_2^{-1} \sim \text{PS}(1/2,4)$		$R_3^{-1} \sim \text{Beta}(1/2)$	
i	σ_i	pr_i	$\text{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{TCov}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{TCov}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{TCov}_{0.95}[X_i, S_{\boldsymbol{X}}]$
1	0.73	20.86%	1.6083	6.1071	2.2849	4.5321	1.1522	0.1808
			(19.85%)	(19.38%)	(19.95%)	(18.00%)	(18.58%)	(12.73%)
2	0.81	23.14%	1.8300	7.0281	2.5944	5.4111	1.3555	0.2548
			(22.59%)	(22.31%)	(22.65%)	(21.49%)	(21.86%)	(17.94%)
3	0.94	26.86%	2.2091	8.6482	3.1191	7.0465	1.7222	0.4230
			(27.27%)	(27.45%)	(27.23%)	(27.99%)	(27.78%)	(29.78%)
4	1.02	29.14%	2.4540	9.7243	3.4549	8.1872	1.9700	0.5617
			(30.29%)	(30.86%)	(30.17%)	(32.52%)	(31.77%)	(39.55%)

Table 2.3: Risk capital allocations based on the $\text{CTE}_{0.95}$ and $\text{TCov}_{0.95}$ risk measures for an exponential MBR portfolio with varying distributions of the systemic risk factor.

rules based on the CTE/TCov risk measures agree with the proportional expected contributions, provides an insight as to the possible reason. Note that for the choice of parameters in this example, an application of Equation (2.4) yields pair-wise Pearson correlations of $\rho_1 = 0.3333$, $\rho_2 = 0.4000$, and $\rho_3 = 0.2500$ imposed by the systemic risk factors R_1, R_2, and R_3 respectively.

In addition, and not surprisingly, the orders of magnitudes of the allocated risk capitals have carried over from Portfolio 1 to Portfolio 2. Namely, the risk measures and capital allocations of interest are all monotonic with respect to the proportional expected contributions, $pr_i,\ i = 1, \ldots, 4.$. That is, lower values of $-\boldsymbol{T}_i's$, or equivalently $\sigma_i's$, $i = 1, \ldots, n$, result in smaller *weights* of the corresponding risk components in the risk portfolio, and also lower corresponding risk capital allocation values. In the jargon of insurance business, this means that a smaller business unit must represent a smaller portion of the portfolio-wise risk, which may or may not comply with the real world practice well.

Another counterintuitive feature of the exponential MBR structures is that the Pearson correlation coefficients are identical for any pair of risk components within a portfolio of risks. As already mentioned, the more general PH-MBR structures introduced in this paper do not bear this disadvantage. We discuss an illustrative PH-MBR risk portfolio next.

2.5.3 Portfolio 3

Let us slightly depart from the exponential MBR constructions adopted in Portfolios 1 and 2. Instead, consider $Y_i \sim \mathrm{PH}(\boldsymbol{\alpha}_i, \boldsymbol{T}_i)$, $i = 1, \ldots, 4$, with the representations

$$\boldsymbol{\alpha}_i = (1, \underbrace{0, \ldots, 0}_{m_i - 1}) \qquad \text{and} \qquad \boldsymbol{T}_i = \underbrace{\begin{pmatrix} -m_i/\sigma_i & m_i/\sigma_i & & & \\ & -m_i/\sigma_i & m_i/\sigma_i & & \\ & & \ddots & \ddots & \\ & & & & -m_i/\sigma_i \end{pmatrix}}_{m_i},$$

for $\sigma_i \in \mathbb{R}_+$ and $m_i \in \mathbb{N}$. In other words, we generalize the distribution of the idiosyncratic risk factors from exponential (in Portfolios 1 and 2) to Erlang, i.e., $Y_i \sim Erl(m_i, m_i/\sigma_i)$, $i = 1, \ldots, 4$. It is noteworthy that since $\mathbb{E}[Y_i] = \sigma_i$ for $i = 1, \ldots, 4$, the proportional expected contributions within Portfolio 3 conform to these in Portfolio 2. However, unlike in Portfolio 2, the dependence structure in Portfolio 3 changes substantially, and the pair-wise Pearson coefficients of correlation are no longer identical.

For the purpose of illustration, set $m_i = i$, $i = 1, \ldots, 4$. In this case, we have

$$\mathrm{Var}(Y_1) > \mathrm{Var}(Y_2) > \mathrm{Var}(Y_3) > \mathrm{Var}(Y_4).$$

By evoking Equation (2.34), the three distinct correlation matrices, as stipulated by systemic

risk factors R_1, R_2, and R_3 in (2.46), are given by

$$\boldsymbol{P}_1 = \begin{pmatrix} 1 & 0.41 & 0.45 & 0.47 \\ & 1 & 0.55 & 0.58 \\ & & 1 & 0.63 \\ & & & 1 \end{pmatrix}, \quad \boldsymbol{P}_2 = \begin{pmatrix} 1 & 0.48 & 0.52 & 0.54 \\ & 1 & 0.62 & 0.65 \\ & & 1 & 0.70 \\ & & & 1 \end{pmatrix}, \quad \boldsymbol{P}_3 = \begin{pmatrix} 1 & 0.32 & 0.35 & 0.38 \\ & 1 & 0.45 & 0.48 \\ & & 1 & 0.53 \\ & & & 1 \end{pmatrix}.$$

			$R_1^{-1} \sim \text{Ga}(3,1)$		$R_2^{-1} \sim \text{PS}(1/2,4)$		$R_3^{-1} \sim \text{Beta}(1/2)$	
i	σ_i	pr_i	$\text{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{TCov}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{TCov}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{CTE}_{0.95}[X_i, S_{\boldsymbol{X}}]$	$\text{TCov}_{0.95}[X_i, S_{\boldsymbol{X}}]$
1	0.73	20.86%	1.7441	6.2415	2.4500	4.9237	1.4845	0.3575
			(23.30%)	(24.47%)	(23.14%)	(29.41%)	(28.68%)	(60.43%)
2	0.81	23.14%	1.7322	5.8851	2.4553	3.8230	1.1889	0.1037
			(23.15%)	(23.07%)	(23.19%)	(22.83%)	(22.97%)	(17.53%)
3	0.94	26.86%	1.9438	6.5169	2.7574	3.9770	1.2428	0.0736
			(25.97%)	(25.55%)	(26.04%)	(23.75%)	(24.01%)	(12.44%)
4	1.02	29.14%	2.0636	6.8627	2.9272	4.0195	1.2604	0.0568
			(27.57%)	(26.91%)	(27.64%)	(24.01%)	(24.35%)	(9.60%)

Table 2.4: Risk capital allocations based on the $\text{CTE}_{0.95}$ and $\text{TCov}_{0.95}$ risk measures for a PH-MBR portfolio with varying distributions of the systemic risk factor.

The values of the risk capital allocation rules in the context of Portfolio 3 are summarized in Table 2.4. Unlike earlier, the values herein are not monotonic with respect to the proportional expected contributions. For instance, although $pr_1 < pr_2$, we have

$$\text{CTE}_{0.95}(X_1, S_{\mathbf{X}}) > \text{CTE}_{0.95}(X_2, S_{\mathbf{X}})$$

for the risk portfolios with the risk components distributed Pareto and gamma. Furthermore, in the context of the risk portfolio with the risk components distributed gamma, we have

$pr_1 < pr_2 < pr_3 < pr_4$, and yet

$$\mathrm{TCov}_{0.95}(\mathrm{X}_1, \mathrm{S}_{\mathbf{X}}) > \mathrm{TCov}_{0.95}(\mathrm{X}_2, \mathrm{S}_{\mathbf{X}}) > \mathrm{TCov}_{0.95}(\mathrm{X}_3, \mathrm{S}_{\mathbf{X}}) > \mathrm{TCov}_{0.95}(\mathrm{X}_4, \mathrm{S}_{\mathbf{X}})$$

.

All in all, we have to conclude that the PH-MBR stochastic representation induces much more intricate dependence structures than the exponential MBR models do. As a result, there is seemingly no more simple rule of thumb as to how the risk capital allocation values shape up when the PH-MBR portfolios are considered.

Appendix: A recap of matrix algebra

For two matrices $\boldsymbol{A} \in \mathbb{R}^{n_1 \times m_1}$ and $\boldsymbol{B} \in \mathbb{R}^{n_2 \times m_2}$, $n_1, n_2, m_1, m_2 \in \mathbb{N}$, the direct sum of $\boldsymbol{A}$ and $\boldsymbol{B}$ is given by

$$\boldsymbol{A} \oplus \boldsymbol{B} = \begin{pmatrix} \boldsymbol{A} & \mathbf{0} \\ \\ \mathbf{0} & \boldsymbol{B} \end{pmatrix}.$$

The addition of $\boldsymbol{A}$ and $\boldsymbol{B}$ can be also defined by means of the Kronecker product and sum. In order to distinguish these two types of matrix additions, we add a 'hat' to the Kronecker operators. For any matrices $\boldsymbol{A} \in \mathbb{R}^{n_1 \times m_1}$ and $\boldsymbol{B} \in \mathbb{R}^{n_2 \times m_2}$, the Kronecker product is defined as

$$\boldsymbol{A} \widehat{\otimes} \boldsymbol{B} = \begin{pmatrix} a_{1,1}\boldsymbol{B} & \cdots & a_{1,m_1}\boldsymbol{B} \\ \\ \vdots & \ddots & \vdots \\ \\ a_{n_1,1}\boldsymbol{B} & \cdots & a_{n_1,m_1}\boldsymbol{B} \end{pmatrix},$$

where $a_{i,j}$ denotes the (i,j)-th element of $\boldsymbol{A}$, $i = 1, \ldots, n_1$, $j = 1, \ldots, m_1$. Furthermore, assume that $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices, i.e., $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_1}$ and $\boldsymbol{B} \in \mathbb{R}^{n_2 \times n_2}$, then the

Kronecker sum is given by

$$\boldsymbol{A}\widehat{\oplus}\boldsymbol{B} = (\boldsymbol{A}\widehat{\otimes}\mathbf{1}_{n_2}) + (\mathbf{1}_{n_1}\widehat{\otimes}\boldsymbol{B}).$$

Note that all the aforementioned notions of addition and multiplication can be extended to higher dimensions via a repeated application.

Let $\boldsymbol{A}$ be a square matrix, and let $\lambda \in \mathbb{C}$ be its eigenvalue with the corresponding eigenspace $\mathfrak{E}_\lambda$. The Jordan block of λ is a square, upper triangular matrix of size $k = \dim(\mathfrak{E}_\lambda)$, such that the diagonal entries are all λ, the super-diagonal entries are all 1, and zeros elsewhere, namely

$$\boldsymbol{J}(\lambda) = \boldsymbol{J}(\lambda; \dim(\mathfrak{E}_\lambda)) = \underbrace{\begin{pmatrix} \lambda & 1 & & & \\ & \ddots & \ddots & & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}}_{k=\dim(\mathfrak{E}_\lambda)}.$$

Furthermore, for $k \in \mathbb{N}$, let $\boldsymbol{A} \in \mathbb{R}^{k\times k}$ be a square matrix, with $m \in \mathbb{N}$ eigenvalues $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$, which are not necessarily distinct. Let

$$\boldsymbol{J} = \boldsymbol{J}(\lambda_1) \oplus \cdots \oplus \boldsymbol{J}(\lambda_m).$$

Then $\boldsymbol{J}$ is similar to $\boldsymbol{A}$. Namely, there exists a nonsingular matrix $\boldsymbol{V_A}$ called the generalized eigenvectors matrix, such that $\boldsymbol{A} = \boldsymbol{V_A}\ \boldsymbol{J}\ \boldsymbol{V_A}^{-1}$. The matrix $\boldsymbol{J}$ is called the Jordan form matrix, which is unique up to the permutation of the Jordan blocks on the diagonal. Now

consider a complex analytic function g. For $j = 1, \ldots, m$, let $k_j = \dim(\mathfrak{E}_{\lambda_j})$ and

$$\boldsymbol{G}(\lambda_j) = \boldsymbol{G}(\lambda_j; \dim(\mathfrak{E}_{\lambda_j})) = \underbrace{\begin{pmatrix} g(\lambda) & g^{(1)}(\lambda) & \cdots & \cdots & \frac{g^{(k_j-1)}(\lambda)}{(k_j-1)!} \\ & g(\lambda) & g^{(1)}(\lambda) & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & g(\lambda) & g^{(1)}(\lambda) \\ & & & & g(\lambda) \end{pmatrix}}_{k_j = \dim(\mathfrak{E}_{\lambda_j})}. \tag{2.47}$$

Then $g(\boldsymbol{A})$ can be computed explicitly as

$$g(\boldsymbol{A}) = \boldsymbol{V_A} \ \left(\oplus_{j=1}^{m} \boldsymbol{G}(\lambda_j)\right) \ \boldsymbol{V_A}^{-1}. \tag{2.48}$$

Chapter 3

A reconciliation of the top-down and bottom-up approaches to risk capital allocations: Proportional allocations revisited

3.1 Introduction

Let the random variable (rv) $X \geq 0$ and the set $\mathcal{X} \ni X$ denote an insurance risk and the collection of such risks, respectively. Also, for $n \in \mathbb{N}$, let the risk X_i, $i = 1, \ldots, n$ be associated with the i-th business unit (BU) of a financial entity, and let $S = \sum_{i=1}^{n} X_i$ stand for the aggregate risk in this entity. Then risk measure, H, is a map that assigns a (monetary) value in $[0, \infty) \cup \{+\infty\}$ to any risk in the set $\mathcal{X}$. We refer to [7, 50, 116] and references therein for axiomatic treatments of risk measures, and to [20, 62, 86] and references therein for some recent developments in risk aggregation.

After the aggregate risk capital, $H(S)$, has been determined, the question arises as to a meaningful way in which it can be allocated to BUs $X_1, \ldots, X_n$. This problem is significantly more involved than the one of computing $H(S)$, but an acceptable solution is of great importance as it would shed light on, e.g., profitability testing, cost sharing, pricing, among

other aspects of practical interest. Speaking formally, allocation rule, A, assigns a (monetary) value in $[0,\ \infty) \cup \{+\infty\}$ to the pairs in the Cartesian product of the set $\mathcal{X}$ with itself, such that $A(X,\ X) = H(X)$ for all $X \in \mathcal{X}$ [e.g., 16, 33, 54, for theory and applications].

The definition of the allocation rule A clearly implies that there are numerous ways to allocate the aggregate risk due to the rv S having cumulative distribution function (cdf) $F_S(s),\ s \geq 0$ and inverse cdf $F_S^{-1}(p) = \inf\{s \geq 0 :\ F_S(s) \geq p\},\ p \in [0,\ 1)$. Some are very simple, such as the hair-cut allocation, A_p,

$$A_p(X_i, S) = H(S)\frac{F_{X_i}^{-1}(p)}{\sum_{i=1}^n F_{X_i}^{-1}(p)},\ i = 1, \ldots, n, \tag{3.1}$$

where $F_{X_i}^{-1}(p),\ p \in [0,\ 1)$ are inverse cdf's of the rv's $X_i,\ i = 1, \ldots, n$. Others are more sophisticated, e.g., the allocations that hinge on, respectively, the distorted and weighted probabilities

$$A_g(X_i, S) = \mathbb{E}\left[X_i g'(F_S(S))\right],\ i = 1, \ldots, n, \tag{3.2}$$

where $g : [0,\ 1] \to [0,\ 1)$ is a continuously differentiable *distortion* function [111], and

$$A_w(X_i, S) = \frac{\mathbb{E}[X_i w(S)]}{\mathbb{E}[w(S)]},\ i = 1, \ldots, n, \tag{3.3}$$

where $w : [0,\ \infty) \to [0,\ \infty)$ is a non-decreasing *weight* function [36, 54]; we assume that all the quantities above are well-defined and finite. Yet others are even more intricate, such as the recently proposed allocation method based on finding the unique *center* of a non-empty convex weakly compact subset of a Banach space [60].

Despite the really overwhelming variety of ways to allocate the aggregate risk, and capital, there are two overarching categorise that stand out. These are the top-down and the bottom-up routes, which imply, respectively, that the allocation exercise is imposed by the corporate centre and informed by the BUs [e.g., 61, for a discussion on the top-down versus bottom-up approaches in the integrated risk management]. More specifically, hair-cur allocation (3.1) is an example of the top-down approach to allocate the aggregate risk due to the rv S. It is fairly simple to compute and transparent to convey to the upper management, and, as a result, it is very popular in applications. That said, the hair-cut allocation disregards the

inter-dependencies among the BUs $X_1, \ldots, X_n$. As such, it fails to reflect upon the fact that these BUs are constituents of a larger structure, treating them as stand-alone objects instead.

On the other hand, weighted allocation (3.3) is an example of the bottom-up approach to allocate the aggregate risk due to the rv S. It is considerably more granular in that it starts with the joint multivariate cdf of the BUs $X_1, \ldots, X_n$, and so accounts for both the inter-dependencies of these BUs and the joint behaviour of the pair $(X_i, S) \in \mathcal{X} \times \mathcal{X},\ i = 1, \ldots, n$. The allocation rule A_w is *consistent*, satisfies *no undercut* and *consistent no undercut* properties [54], and it is optimal in the sense of [36]. That said, unless very special distributional structures are considered [56], the weighted allocations are rather difficult to compute even for special choices of the weight function, w, let alone in general. To illustrate the computational complexity, we refer to [35] for elliptically distributed BUs; [23] for phase-type distributed BUs; [49] for Tweedie distributed BUs; [114, 115] for skew-normal and Pareto distributed BUs; [26] and [27] for BUs with the dependence structures described by the Farlie-Gumbel-Morgenstern copula and the Archimedean copula, respectively. All these works compute the weighted allocation A_w for the weight function $w(s) = \mathbf{1}\{s > F_S^{-1}(p)\},\ p \in [0,\ 1)$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function, and even in this very particular case the aforementioned list of references is not complete.

In summary, the top-down approach (e.g., (3.1)) is intuitive yet often oversimplified, and the bottom-up approach (e.g., (3.3)) is meticulous yet 'hits against too many parameters'. In practice, the two approaches are conducted separately and are sought to complement each other. The question that arises then is whether it is possible to unify the top-down and the bottom-up ways to allocate the aggregate risk so that the end-result is intuitive, detailed and would not add computational complexity beyond the one associated with computing the aggregate risk capital $H(S)$. Putting forward such an encompassing approach to allocate risk capital is what we pursue in the present paper.

3.2 Outline of the paper and some preliminaries

Assume that the risk measure H is positively homogeneous, that is, for all $\lambda > 0$ and $X \in \mathcal{X}$, we have $H(\lambda X) = \lambda H(X)$. Then all of the allocation rules mentioned hitherto can be written as the proportional allocation $A_{prop}(X_i, S) = H(Sr_i),\ i = 1, \ldots, n$, where $r_i \in [0,\ 1]$ is the ratio of: $H_p(X_i) = F_{X_i}^{-1}(p)$ and $\sum_{i=1}^n H_p(X_i)$ - in the context of the hair-cut allocation; $A_g(X_i, S) = \mathbb{E}[Xg'(F_S(S))]$ and $\sum_{i=1}^n A_g(X_i, S)$ - in the context of the distorted allocation; $A_w(X_i, S) = \mathbb{E}[Xw(S)]/\mathbb{E}[w(S)]$ and $\sum_{i=1}^n A_w(X_i, S)$ in the context of the weighted allocation. Once again we assume that all of the quantities mentioned above are well-defined and finite.

The thrust of the method that we propose in the present paper is that we substitute the deterministic r_i with a ratio rv $R_i \geq 0,\ i = 1, \ldots, n$, for which $\sum_{i=1}^n R_i = 1$ almost surely. Then the joint cdf of the rv's $R_1, \ldots, R_n$ models the joint contributions of the BUs in the financial entity of interest, and so symbolizes the bottom-up approach in the risk capital allocation exercise. If the rv $X_i,\ i = 1, \ldots, n$ denotes the risk due to the i-th BU and $S = \sum_{i=1}^n X_i$ denotes the aggregate risk, then $R_i = X_i/S$. Further let the rv $Z \geq 0$ denote the overall aggregate risk that is envisioned by the upper management; the distribution of the rv Z depends on, e.g., the corporate strategic goals, business mix, risk factors matrix, among other factors, and so the rv's Z and S are not necessarily equal in distribution. Given the above, the contribution of the i-th BU is modelled by the product rv $ZR_i,\ i = 1, \ldots, n$, where $\sum_{i=1}^n ZR_i = Z$ by construction. Consequently, the allocation rule of interest is $A(ZR_i, Z)$ rather than $H(Sr_i),\ i = 1, \ldots, n$ as in (3.1)-(3.3).

The idea that we have just sketched can be reformulated with the help of the language of *compositions* (e.g., Aitchison, 1982, for details; also see Belles-Sampera et al., 2016; Boonen et al., 2019 for recent applications of compositional methods in risk management). That is, let $\mathbb{S}^n$ denote the n-dimensional simplex, and let $\mathbf{C} = (C_1, \ldots, C_n) \in \mathbb{S}^n$ and $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$ denote its elements and basis, respectively. Then we call $C : \mathcal{X}^n \to \mathbb{S}^n$ a *compositional map*, if $\sum_{i=1}^n C_i = 1$ holds almost surely. Within this context and under the special choice of the compositional map, C, such that $C_i(\mathbf{X}) = X_i / \sum_{i=1}^n X_i$, we have that $H(S \times C_i(A(X_1, S), \ldots, A(X_n, S)))$ recovers allocation rules (3.1)-(3.3) for appropriate

choices of the allocation rule A, whereas the approach proposed in the present paper suggests the allocation rule $A(Z \times C_i(\mathbf{X}), Z)$ instead. That is the deterministic composition of allocations, $C_i(A(X_1, S), \ldots, A(X_n, S))$ is substituted herein with the random composition $C_i(X_1, \ldots, X_n),\ i = 1, \ldots, n$.

Quite remarkably, when the random pair (ZR_i, Z), and not (X_i, S), is taken as the input for, e.g., the distorted or weighted allocation rule, the computational complexity of the risk capital allocation exercise reduces considerably. For instance, for the latter allocation rule, we readily have

$$A\left(ZR_i, Z\right) = \frac{\mathbb{E}\left[ZC_i(\mathbf{X})w(Z)\right]}{\mathbb{E}\left[w(Z)\right]} = \frac{\mathbb{E}\left[C_i(\mathbf{X})\mathbb{E}\left[Zw(Z)|\ \mathbf{X}\right]\right]}{\mathbb{E}\left[w(Z)\right]},\ i = 1, \ldots, n$$

for any choice of the weight function w. If the rv's $X_i,\ i = 1, \ldots, n$ are independent on the rv Z, which may become an acceptable assumption in view of what the genesis of these rv's is, and considering the already-mentioned compositional map $C_i(\mathbf{X}) = X_i/S$, then

$$A_w(ZR_i, Z) = \mathbb{E}[R_i] \times H_w(Z),\ i = 1, \ldots, n,$$

and so in this case the computational complexities involved in the aggregate and allocated risk determination are the same. Obviously, depending on what the goals of the allocation exercise are, compositions having elements, C_i, other than $R_i = X_i/S$ may be of interest in applications. One example, for s_p denoting the inverse cdf $F_S^{-1}(p)$, is the composition such that $C_i = R_i\mathbf{1}\{S > s_p\}/\mathbb{P}(S > s_p),\ p \in [0,\ 1)$ and $\mathbb{P}(S > s_p) \neq 0$, which emphasizes the extreme scenarios in the sample space of the aggregate risk rv S. In this case, and assuming again the independence of the rv's $X_1, \ldots, X_n$ on Z, we obtain

$$A_w(ZR_i, Z) = \mathbb{E}[R_i|\ S > s_p] \times H_w(Z),\ i = 1, \ldots, n,$$

which is not difficult to compute as is shown in Section 3.5.

The rest of the paper capitalizes on the just-outlined ideas. Specifically in Section 3.3, we discuss a multivariate probabilistic structure that is a natural choice to serve as the distribution of the basis rv $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$. Then in Section 3.4, we construct a flexible

yet tractable variant of the well-known Dirichlet distribution on the n-dimensional simplex, and we study some properties of the vector of random proportions $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathbb{S}^n$, which is obtained with the help of a compositional map $C : \mathcal{X}^n \to \mathbb{S}^n$. Finally in Section 3.5, we sketch an expectation maximization (EM) algorithm to estimate the parameters of the just-mentioned Dirichlet distribution and illustrate its applications in the context of the risk capital allocation exercise.

3.2.1 Preliminaries

We work with an atomless probability space $(\Omega, \mathcal{A}, \mathbb{P})$, which in our context means that there exists at least one rv with a continuous distribution in this space. Let L^r denote the set of all rv's on $(\Omega, \mathcal{A}, \mathbb{P})$ with finite $r \in [0,\ \infty)$-th moment, and let L^∞ denote the set of all essentially bounded rv's. Unless stated otherwise, we assume that rv's are in L^1. Throughout the paper, for every $X \in L^0$, we denote by F_X the cdf of the rv X. For $\mathcal{X}^n$ denoting the n-fold Cartesian product of $\mathcal{X}$ with itself, we call the rv $\mathbf{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$, basis. Besides the convex cone $\mathcal{X}$, which is a subset of L^0, in this paper we deal with the open n-dimensional simplex space

$$\mathbb{V}^n = \{(r_1, \ldots, r_n) :\ r_i \geq 0,\ i = 1, \ldots, n,\ \text{and } r_1 + \cdots + r_n < 1\}$$

and the already-mentioned boundary

$$\mathbb{S}^n = \{(r_1, \ldots, r_n) :\ r_i \geq 0,\ i = 1, \ldots, n,\ \text{and } r_1 + \cdots + r_n = 1\}\,.$$

Our main constructions are then random compositions $\boldsymbol{R} = (R_1, \ldots, R_n)$ that are special maps $C : \mathcal{X}^n \to \mathbb{S}^n$, such that $C_i(X_1, \ldots, X_n) = X_i / \sum_{i=1}^n X_i,\ i = 1, \ldots, n$. Finally, $\mathbb{N}_0$ and $\mathbb{R}_{0,+}$ denote respectively the zero-augmented sets of natural, $\mathbb{N} \cup \{0\}$, and positive real, $\mathbb{R}_+ \cup \{0\}$, numbers; the sets $\mathbb{N}_0^n$ and $\mathbb{R}_{0,+}^n$ denote the corresponding multivariate counterparts.

3.3 Constructing random compositions via a class of mixed-gamma distributions

We recall at the outset that as $R_i = X_i/S,\ i = 1, \ldots, n$, the compendium of the distributions on the simplex, and in particular the most popular member therein, the Dirichlet distributions, seem natural to evoke [e.g., 92, and references therein]. To start off, recall that the rv Γ_i is said to be distributed gamma with the shape and scale parameters $\gamma_i \in \mathbb{R}_+$ and $\beta_i \in \mathbb{R}_+$, respectively, if it has the following probability density function (pdf)

$$f_{\Gamma_i}(x) = \frac{1}{\Gamma(\gamma)} e^{-x/\beta_i} x^{\gamma_i - 1} \beta_i^{-\gamma_i} \text{ for all } x \in \mathbb{R}_+,\ i = 1, \ldots, n. \tag{3.4}$$

Succinctly, we write $\Gamma_i \sim Ga(\gamma_i, \beta_i),\ i = 1, \ldots, n$. Then assume that the rv's $\Gamma_1, \ldots, \Gamma_n$ are mutually independent, denote by $\Gamma_+ = \sum_{i=1}^n \Gamma_i$ their sum, and set $\beta_i \equiv \beta \in \mathbb{R}_+$. The joint distribution of the rv $\boldsymbol{R} = (R_1, \ldots, R_n),\ R_i = \Gamma_i/\Gamma_+,\ i = 1, \ldots, n$ is Dirichlet. That is the joint pdf of the rv $\boldsymbol{R}$ is

$$f_{\boldsymbol{R}}(r_1, \ldots, r_n) = \frac{1}{\mathrm{B}(\gamma_1, \ldots, \gamma_n)} \prod_{i=1}^n r_i^{\gamma_i - 1},\ (r_1, \ldots, r_n) \in \mathbb{S}^n, \tag{3.5}$$

where $\mathrm{B}(\gamma_1, \ldots, \gamma_n)$ is the multivariate Beta function

$$\mathrm{B}(\gamma_1, \ldots, \gamma_n) = \frac{\Gamma(\gamma_1) \times \cdots \times \Gamma(\gamma_n)}{\Gamma(\gamma_1 + \cdots + \gamma_n)}. \tag{3.6}$$

The Dirichlet distribution is convenient to work with, but unfortunately it barely suits our needs for many reasons. E.g., the assumption that the risks due to all the BUs of a financial entity are distributed gamma is very questionable, and so is doubtful the conclusion that the rv's $R_i = X_i/S,\ i = 1, \ldots, n$ and S are independent [e.g., 92, for a discussion]. Therefore, in the rest of this section we seek a suitable class of distributions to model the risks due to the BUs $X_1, \ldots, X_n$ and so to serve as a basis for the desired compositional map. In particular, we are interested in those classes of distributions that: (1) are flexible to the extent that they can model well any cdf with non-negative support; (2) allow for a dependence among $X_1, \ldots, X_n$; (3) contain the gamma distribution as a special case; (4) inhere the tractability of the gamma

distributions; (5) relax the assumption of independence of the rv's $R_i,\ i = 1, \ldots, n$ and S, as well as some other rather restrictive notions of independence on the simplex that characterize the class of Dirichlet distributions [e.g. 2].

3.3.1 A multivariate mixed-gamma distribution

The class of univariate mixed-Erlang distributions [e.g. 110, 118] is an immediate candidate to model the distribution of the risk $X_i \in \mathcal{X},\ i = 1, \ldots, n$. Indeed, mixed-Erlang distributions are dense in the space of the cdf's with non-negative support, fairly tractable, and ensue straightforward multivariate extensions [e.g., 74, 75]. That said, when chosen as a basis for a compositional map, mixed-Erlang distributions cannot incorporate pdf (3.5), hence adjustments have to be made. This is achieved in the following definition.

Definition 8. Let $\kappa \in \mathbb{N}_0$ denote a discrete rv with the probability mass function (PMF) $p_\kappa(k),\ k \in \mathbb{N}_0$. Also, let $\gamma_\kappa = \gamma + \kappa$ and $\gamma_k = \gamma + k$. Then we say that the rv $\Gamma^{(\kappa)}$ is distributed mixed-gamma (MG), succinctly $\Gamma^{(\kappa)} \sim MG(\gamma, \beta, p_\kappa)$, if its pdf is given by

$$f_{\Gamma^{(\kappa)}}(x) = \sum_{k=0}^{\infty} p_\kappa(k) \frac{1}{\Gamma(\gamma_k)} e^{-x/\beta} x^{\gamma_k - 1} \beta^{-\gamma_k} \text{ for all } x \in \mathbb{R}_+. \tag{3.7}$$

Note 1. Recall that the size-biased of order $k \in \mathbb{N}_0$ variant rv of a non-negative rv, $X \in L^k$, is defined via

$$\mathbb{P}\left(X^{(k)} \in dx\right) = \frac{x^k}{\mathbb{E}\left[X^k\right]} \mathbb{P}\left(X \in dx\right),\ x \in \mathbb{R}_+,$$

[e.g., 93, for a thorough discussion of the notion of size-biasing]. In view of this, the class of mixed-gamma distributions can be considered a size-biased mixture, and so the notation $\Gamma^{(\kappa)}$, where κ is the random order of the size-bias operation, is natural.

Definition 8 leads to a variety of attractive properties for the MG class of distributions. We start with the Laplace transform of pdf (3.7). To this end, let $P(z) = \mathbb{E}\left[z^\kappa\right],\ |z| \leq 1$ denote the probability generating function (PGF) of the rv κ and recall that, for $\Gamma \sim Ga(\gamma, \beta)$, the Laplace transform is

$$\widehat{f}_\Gamma(t) = \left(1 + \frac{t}{\beta}\right)^{-\gamma},\ t \in \mathbb{R}_{0,+}.$$

Then we have the following assertion.

Theorem 1. *The Laplace transform that corresponds to the rv* $\Gamma^{(\kappa)} \sim MG(\gamma, \beta, p_\kappa)$ *is given by*

$$\widehat{f}_{\Gamma^{(\kappa)}}(t) = \widehat{f}_{\Gamma}(t)\, P\left(\frac{1}{1+\beta t}\right),\ t \in \mathbb{R}_{0,+}.$$

Therefore, we have $\Gamma^{(\kappa)} \stackrel{d}{=} \Gamma + S_\kappa$, *where* $S_\kappa = \sum_{k=1}^{\kappa} E_k,\ S_0 = 0$, *and* $E_k,\ k \in \mathbb{N}$, *denotes a sequence of independent and identical rv's distributed exponentially with the scale parameter* $\beta \in \mathbb{R}_+$; *here "* $\stackrel{d}{=}$ *" means equality in distribution.*

Proof. By definition of the Laplace transform and interchanging the order of the summation and integration, we readily have

$$\widehat{f}_{\Gamma^{(\kappa)}}(t) = \sum_{k=0}^{\infty} p_\kappa(k)(1+\beta t)^{-(\gamma+k)} = (1+\beta t)^{-\gamma} \sum_{k=0}^{\infty} p_\kappa(k)(1+\beta t)^{-k},\ t \in \mathbb{R}_{0,+}.$$

Also, we have

$$\mathbb{E}\left[\exp(-t(\Gamma + S_\kappa))\right] \ = \ \widehat{f}_{\Gamma}(t)\mathbb{E}\left[\left(\frac{1}{1+\beta t}\right)^{\kappa}\right] = \widehat{f}_{\Gamma}(t) P\left(\frac{1}{1+\beta t}\right),\ t \in \mathbb{R}_{0,+},$$

which establishes the equality in distribution. This completes the proof of the theorem. □

The class of MG distributions is closed under rescaling. This is clearly so as is seen from

$$\widehat{f}_{\lambda\Gamma^{(\kappa)}}(t) = \mathbb{E}\left[\exp\left(-t\lambda\Gamma^{(\kappa)}\right)\right] = (1+t\lambda\beta)^{-\gamma}\left(\frac{1}{1+t\lambda\beta}\right),\ t \in \mathbb{R}_{0,+}.$$

We next use the Laplace transform of the rv $\Gamma^{(\kappa)}$ to show that the class of MG distributions is a good modelling tool. The proof of the succeeding assertion is borrowed heavily from [74].

Theorem 2. *The class of MG distributions is dense in the class of all continuous distributions with positive support.*

Proof. Fix an arbitrary positive continuous distribution with pdf f, cdf F and Laplace

transform $\widehat{f}$. Consider the sequence of Laplace transforms $\left\{\widehat{f}_n\right\}_{n\in\mathbb{N}_0}$, such that

$$\widehat{f}_n(t) = \left(1+\frac{t}{n}\right)^{-\gamma}\int_0^\infty \left(1+\frac{t}{n}\right)^{-xn} f(x)dx, \tag{3.8}$$

for $\beta = 1/n$. Then, on the one hand side, $\widehat{f}_n(t)$ is the Laplace transform of an MG pdf, that is, for $p_\kappa(k) = F((k+1)\beta) - F(k\beta),\ k\in\mathbb{N}_0$, Equation (3.8) is equivalent to

$$\sum_{k=0}^\infty \left(\int_{k\beta}^{(k+1)\beta} f(x)dx\right)(1+\beta t)^{-(\gamma+k)} = (1+\beta t)^{-\gamma}\sum_{k=0}^\infty p_\kappa(k)\ (1+\beta t)^{-k} = \widehat{f}_{\Gamma^{(\kappa)}}(t),\ t\in\mathbb{R}_{0,+}.$$

On the other hand side, by the Dominated Convergence Theorem, we obtain

$$\lim_{n\to\infty}\widehat{f}_n(t) = \int_0^\infty f(x)e^{-xt}dx = \hat{f}(t)$$

for all $t\in\mathbb{R}_{0,+}$. The assertion is thus proved by evoking Lévy's continuity theorem. □

As the risks $X_1,\ldots,X_n$ in the basis $\boldsymbol{X}$ must not be mutually independent, it is critically important for us to consider a multivariate extension of the MG distributions in Definition 8. The extension that we put forward next is inspired by the multivariate mixed-Erlang distributions studied in [75, 113, 119]. Namely, the multivariate mixed-gamma distributions presented in Definition 9 below, generalize the just-mentioned mixed-Erlang distributions by allowing for arbitrary non-negative shape parameters as well as for heterogeneous scale parameters of the margins.

Let $\boldsymbol{\kappa} = (\kappa_1,\ldots,\kappa_n)$ be a vector of discrete rv's, $\kappa_i\in N_0,\ i=1,\ldots,n$, and denote by $p_{\boldsymbol{\kappa}}(\boldsymbol{k}) = \mathbb{P}(\kappa_1 = k_1,\ldots,\kappa_n = k_n)$, where $\boldsymbol{k} = (k_1,\ldots,k_n)\in\mathbb{N}_0^n$ the corresponding joint PMF.

Definition 9. The rv $\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})} = (\Gamma_1^{(\kappa_1)},\ldots,\Gamma_n^{(\kappa_n)})$ is said to be distributed n-variate mixed-gamma (MG_n) if the corresponding joint pdf is given by

$$f_{\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})}}(x_1,\ldots,x_n) \ = \ \sum_{\boldsymbol{k}\in\mathbb{N}_0^n} p_{\boldsymbol{\kappa}}(\boldsymbol{k})\prod_{i=1}^n \frac{1}{\Gamma(\gamma_{k_i})}e^{-x_i/\beta_i}x_i^{\gamma_{k_i}-1}\beta_i^{-\gamma_{k_i}},\ (x_1,\ldots,x_n)\in\mathbb{R}_+^n, \tag{3.9}$$

where $\gamma_{k_i} = \gamma_i + k_i$ and $\beta_i, \gamma_i \in \mathbb{R}_+,\ i = 1, \ldots, n$. Succinctly, we write $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} \sim MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, where the coordinates of the vectors of parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are, respectively, $\gamma_i + k_i$ and $\beta_i,\ i = 1, \ldots, n$.

A thorough study of the class of the multivariate MG distributions is beyond the immediate interest in the present paper. Herein we only present a few basic properties that are of central importance to our subsequent study of the compositional maps that arise from the basis vectors distributed MG_n.

Theorem 3. *Consider the rv $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} \sim MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, then the following assertions hold:*

(i) The joint Laplace transform that corresponds to the rv $\mathbf{\Gamma}^{(\boldsymbol{\kappa})}$ is

$$\widehat{f}_{\mathbf{\Gamma}^{(\boldsymbol{\kappa})}}(t_1, \ldots, t_n) \;=\; \prod_{i=1}^{n} (1 + \beta_i t_i)^{-\gamma_i}\, P_{\boldsymbol{\kappa}} \left(\frac{1}{1 + \beta_1 t_1}, \ldots, \frac{1}{1 + \beta_n t_n} \right),$$

where $(t_1, \ldots, t_n) \in \mathbb{R}^n_{0,+}$ and $P_{\boldsymbol{\kappa}}$ denotes the joint PGF of the rv $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)$.

(ii) The marginal coordinate of $\mathbf{\Gamma}^{(\boldsymbol{\kappa})}$, $\Gamma_i^{(\kappa_i)} \sim MG(\gamma_i, \beta_i, p_{\kappa_i})$, $i = 1, \ldots, n$, admits the stochastic representation $\Gamma_i^{(\kappa_i)} = \Gamma_i + \sum_{j=1}^{\kappa_i} E_{i,j}$ where $\Gamma_i \sim Ga(\gamma_i, \beta_i)$ and $\{E_{i,j}\}_{j \in \mathbb{N}}$, denotes a sequence of independent and identical rv's distributed exponentially with the scale parameter $\beta_i \in \mathbb{R}_+$.

(iii) If $\kappa_1, \ldots, \kappa_n$ are independent, i.e., $p_{\boldsymbol{\kappa}}(\boldsymbol{k}) = \prod_{i=1}^{n} p_{\kappa_i}(k_i)$, then the rv's $\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)}$ are independent.

(iv) Choose $1 \leq i \neq j \leq n$ and consider the pair $(\Gamma_i^{(\kappa_i)}, \Gamma_j^{(\kappa_j)}) \sim MG_2(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, where $\boldsymbol{\gamma} = (\gamma_i, \gamma_j)$, $\boldsymbol{\beta} = (\beta_i, \beta_j)$ and $\boldsymbol{\kappa} = (\kappa_i, \kappa_j)$. Then, assuming that $\kappa_i,\ \kappa_j \in L^2$, the Pearson correlation coefficient is given by

$$\operatorname{Corr}(\Gamma_i^{(\kappa_i)}, \Gamma_j^{(\kappa_j)}) = \operatorname{Corr}(\kappa_i, \kappa_j) \frac{\sqrt{\operatorname{Var}(\kappa_i)\operatorname{Var}(\kappa_j)}}{\sqrt{(\operatorname{Var}(\kappa_i) + \mathbb{E}[\kappa_i] + \gamma_i)\,(\operatorname{Var}(\kappa_j) + \mathbb{E}[\kappa_j] + \gamma_j)}}. \tag{3.10}$$

Proof. We prove (i), as the remaining assertions either follow from it or hold by construction.

We have for $(t_1, \ldots, t_n) \in \mathbb{R}^n_{0,+}$,

$$\widehat{f}_{\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})}}(t_1, \ldots, t_n) = \sum_{\boldsymbol{k} \in \mathbb{N}_0^n}^{\infty} p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \prod_{i=1}^{n} (1 + \beta_i t_i)^{-(\gamma_i + k_i)} = \prod_{i=1}^{n} (1 + \beta_i t_i)^{-\gamma_i} \mathbb{E}\left[\prod_{i=1}^{n} (1 + \beta_i t_i)^{-\kappa_i} \right].$$

This establishes the joint Laplace transform and so proves (i). □

Note 2. Correlation formula (3.10) suggests that the multivariate MG distributions proposed herein can cover the full range of bivariate dependence, when it is measured by the Pearson coefficient of correlation. Namely, the sign of the Pearson coefficient of correlation of the pair $(\Gamma_i^{(\kappa_i)}, \Gamma_j^{(\kappa_j)})$ can be both positive and negative, stipulated by the sign of the correlation $\mathrm{Corr}(\kappa_\mathrm{i}, \kappa_\mathrm{j})$, $1 \leq \mathrm{i} \neq \mathrm{j} \leq \mathrm{n}$. Since the random pair (κ_i, κ_j) is allowed to have any dependence structure, including comonotonicity and counter-comonotonicity, $\mathrm{Corr}(\Gamma_i^{(\kappa_i)}, \Gamma_j^{(\kappa_j)})$ can attain any value in the interval $[-1, 1]$. In addition, by choosing random pairs (κ_i, κ_j) with sufficiently large variances, $\mathrm{Corr}(\Gamma_i^{(\kappa_i)}, \Gamma_j^{(\kappa_j)})$ can be made arbitrarily close to $\mathrm{Corr}(\kappa_i, \kappa_j)$.

Akin to the univariate mixed-gamma distributions, the class of MG_n distributions can model any multivariate distribution with positive support arbitrarily well. The proof of this assertion is a straightforward generalization of the proof of Theorem 2 and is thus omitted.

Theorem 4. *The multivariate MG distributions form a dense class of continuous multivariate distributions with non-negative supports.*

It is well-known that finite convolutions of the rv's distributed gamma with arbitrary shape and scale parameters are mixed-gamma. The next theorem is reported for completeness of exposition [e.g., 89, for details]. It has been frequently adopted in the actuarial literature in order to deal with general finite convolutions within the class of gamma distributions [e.g., 47, 65, 107, and referenes therein].

Theorem 5. *For $i = 1, \ldots, n$, let $\Gamma_i \sim Ga(\gamma_i, \beta_i)$ denote independent rv's distributed gamma, and let $\Gamma_+ = \Gamma_1 + \cdots + \Gamma_n$ denote their sum. Then $\Gamma_+ \sim MG(\gamma^*, \beta^*, p_{\kappa^{**}})$, where $\gamma^* = \gamma_1 + \cdots + \gamma_n$, $\beta^* = \bigwedge_{i=1}^n \beta_i$ and κ^{**} is an integer-valued non-negative rv with the PMF*

given, for $k \in \mathbb{N}_0$, *by* $p_{\kappa^{**}}(k) = c\delta_k$, *where*

$$c = \prod_{i=1}^{n} \left(\frac{\beta^*}{\beta_i}\right)^{\gamma_i} \text{ and } \delta_k = k^{-1} \sum_{l=1}^{k} \sum_{i=1}^{n} \gamma_i \left(1 - \frac{\beta^*}{\beta_i}\right)^l \delta_{k-l}, \text{ for } k \in \mathbb{N} \text{ and } \delta_0 = 1. \tag{3.11}$$

We now generalize Theorem 5 by allowing for (i) summands in the MG class of distributions, and (ii) dependence implied by the class of MG_n distributions. At the outset, we remind briefly that the rv $N \in \mathbb{N}_0$ is said to be distributed negative binomial, succinctly $N \sim NB(\gamma, p)$, where $\gamma > 0$ and $p \in (0,1)$ are parameters, if its PMF is given by

$$\mathbb{P}[N = n] = \frac{\Gamma(\gamma + n)}{\Gamma(\gamma) n!} p^\gamma (1-p)^n \text{ for all } n \in \mathbb{N}_0.$$

The corresponding PGF is

$$P_N(z) = \left(\frac{p}{1 - (1-p)z}\right)^\gamma, \ |z| < 1/(1-p).$$

Theorem 6. *Consider the rv* $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} = (\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)}) \sim MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$ *and let* $\Gamma_+^{(\kappa^*)} = \sum_{i=1}^n \Gamma_i^{(\kappa_i)}$ *denote the sum of its coordinates. Then* $\Gamma_+^{(\kappa^*)}$ *is distributed MG with the parameters* $\gamma^* = \gamma_1 + \cdots + \gamma_n$, $\beta^* = \bigwedge_{i=1}^n \beta_i$, *and* p_{κ^*} *such that*

$$p_{\kappa^*}(m) = \sum_{j=0}^{m} \sum_{k_1 + \ldots + k_n = j} \left(p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \sum_{\substack{y_1 + \ldots + y_n = m - j \\ (y_1, \ldots, y_n) \in \mathbb{N}_0^n}} \prod_{i=1}^{n} \frac{\Gamma(\gamma_{k_i} + y_i)}{\Gamma(\gamma_{k_i}) y_i!} \left(\frac{\beta^*}{\beta_i}\right)^{\gamma_{k_i}} \left(1 - \frac{\beta^*}{\beta_i}\right)^{y_i} \right) \tag{3.12}$$

for all $m \in \mathbb{N}_0$.

Proof. Let $N_i \sim NB(\gamma_i, \beta^*/\beta_i)$, $i = 1, \ldots, n$, then the associated PGF can be expressed as

$$P_{N_i}\left(\frac{1}{1+\beta^* t}\right) = \left(\frac{1+\beta_i t}{1+\beta^* t}\right)^{-\gamma_i}, \ t \in \mathbb{R}_{0,+}.$$

Furthermore, let $P_{\boldsymbol{\kappa}}(\cdot)$ denote the joint PGF of rv $\boldsymbol{\kappa}$. For $t \in \mathbb{R}_{0,+}$, we have, starting

with Point (i) of Theorem 3,

$$\begin{aligned}\mathbb{E}\left[\exp\left(-t\sum_{i=1}^{n}\Gamma_i^{(\kappa_i)}\right)\right] &= \prod_{i=1}^{n}(1+\beta_i t)^{-\gamma_i}\,P_{\boldsymbol{\kappa}}\left(\frac{1}{1+\beta_1 t_1},\ldots,\frac{1}{1+\beta_n t_n}\right)\\ &= (1+\beta^* t)^{-\gamma^*}\,\mathbb{E}\left[(1+\beta^* t)^{-\sum_{i=1}^{n}\kappa_i}\prod_{i=1}^{n}\left(\frac{1+\beta_i t}{1+\beta^* t}\right)^{-(\gamma_i+\kappa_i)}\right]\\ &= (1+\beta^* t)^{-\gamma^*}\,\mathbb{E}\left[(1+\beta^* t)^{-\sum_{i=1}^{n}\kappa_i}\prod_{i=1}^{n}P_{N_i(\kappa_i)}\left(\frac{1}{1+\beta^* t}\right)\right],\end{aligned}$$

where $N_i(\kappa_i)\sim NB(\gamma_i+\kappa_i,\beta^*/\beta_i)$, $i=1,\ldots,n$, that is the rv $N_i(\kappa_i)$ follows the negative binomial distribution with a random shape parameter. The expectation in the last line is the PGF of the rv $\kappa^*\overset{d}{=}\sum_{i=1}^{n}(\kappa_i+N_i(\kappa_i))$ evaluated at $(1+\beta^* t)^{-1}$, and so the distribution of the rv $\Gamma_+^{(\kappa^*)}$ is a mixed-gamma due to Theorem 1. Also, the PMF of the rv κ^* follows as

$$p_{\kappa^*}(m) = \sum_{j=0}^{m}\sum_{k_1+\ldots+k_n=j}\left[p_{\boldsymbol{\kappa}}(\boldsymbol{k})\,\mathbb{P}\left(\sum_{i=1}^{n}N_i(k_i)=m-j\right)\right]\ \text{for all } m\in\mathbb{N}_0.$$

This completes the proof of the theorem. □

A by-product of Theorem 6 is that it demystifies the recursive formula presented in [89] in the context of finite gamma convolutions (also, Theorem 5 above). This is stated in the following corollary, which is proved by choosing $p_{\boldsymbol{\kappa}}(0,\ldots,0)=1$ in Theorem 6.

Corollary 8. *Within the set-up in Theorem 5, we have* $\kappa^{**}\overset{d}{=}\sum_{i=1}^{n}N_i$, *where* $N_i\sim NB(\gamma_i,\beta^*/\beta_i)$ *are mutually independent rv's having negative binomial distributions. The PMF of the rv* κ^{**} *admits the following (non-recursive) form*

$$p_{\kappa^{**}}(k)=\sum_{\substack{y_1+\ldots+y_n=k\\(y_1,\ldots,y_n)\in\mathbb{N}_0^n}}\prod_{i=1}^{n}\frac{\Gamma(\gamma_i+y_i)}{\Gamma(\gamma_i)y_i!}(\beta^*/\beta_i)^{\gamma_i}(1-\beta^*/\beta_i)^{y_i},\ k\in\mathbb{N}_0. \tag{3.13}$$

In the next section, we show how the class of mixed-gamma distributions can be used as a basis for constructing random compositions $\mathbf{R}=(R_1,\ldots,R_n)\in\mathbb{S}^n$.

3.4 From mixed-gamma to a general distribution on the simplex

Dirichlet pdf (3.5) is remarkably tractable. E.g., for $\gamma^* = \gamma_1 + \cdots + \gamma_n$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^n_+$ and given that $\mathbf{R} = (R_1, \ldots, R_n) \sim Dir(\boldsymbol{\gamma})$, it is easy to find

$$\mathbb{E}[R_i] = \frac{\gamma_i}{\gamma^*} \text{ and } \mathrm{Var}(R_i) = \frac{\gamma_i(\gamma^* - \gamma_i)}{\gamma^{*2}(\gamma^* + 1)}, \ i = 1, \ldots, n,$$

as well as

$$\mathrm{Cov}(R_i, R_j) = -\frac{\gamma_i \gamma_j}{\gamma^{*2}(\gamma^* + 1)}, \ 1 \leq i \neq j \leq n.$$

Hence the random pair (R_i, R_j) with the joint Dirichlet distribution must be negatively correlated, which adds an additional layer of practical inconveniences when it comes to the applications of the Dirichlet distributions in the context of risk allocations, as well as in other contexts.

In addition, with a little effort, some more intricate properties of the class of Dirichlet distributions can be derived. E.g., it is possible to show that the class of Dirichlet distributions is closed under marginalization of any order, and that the level curves, for $\gamma_i > 1$, $i = 1, \ldots, n$, are always convex sets. Further [e.g., 3, for details], rather unfortunately, the class of Dirichlet distributions can be seen as an *independence extreme* in the world of compositional data, which is the price that the Dirichlet distributions have to pay for the tractability they inherit from the class of gamma distributions.

Numerous efforts have been made to generalize the Dirichlet distribution with pdf (3.5) [e.g., 92, among others]. The task is however not an easy call. Namely, a slight generalization of the set-up leads to considerable complications. For instance, for $\Gamma_i \sim Ga(\gamma_i, \beta_i)$, $i = 1, \ldots, n$, let $\{\Gamma_i\}_{i=1}^n$ be a sequence of mutually independent rv's (note that the scale parameters are arbitrary now) and let Γ_+ denote the sum of these rv's, then the rv $\boldsymbol{R} = (R_1, \ldots, R_n)$, $R_i = \Gamma_i/\Gamma_+$ is distributed scaled Dirichlet [e.g., 92], which is by far less tractable than the one with pdf (3.5). In particular, even an analytic expression for the covariance is not known for the scaled Dirichlet distribution.

In this section, we use the class of multivariate MG distributions as the basis to formulate

a variant of a generalized Dirichlet distribution. Recall that we write $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} \sim MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$ when $\mathbf{\Gamma}^{(\boldsymbol{\kappa})}$ is distributed multivariate MG with the vectors of parameters

$$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^n_+, \ \boldsymbol{\beta} = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n_+$$

and the associated joint PMF $p_{\boldsymbol{\kappa}}$. Also, we denote by $\Gamma_+^{(\kappa^*)} = \sum_{i=1}^n \Gamma_i^{(\kappa_i)}$ the sum of the rv's distributed multivariate MG.

Following the language of [3], we call the rv $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} = (\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)}) \in \mathcal{X}^n_+$, basis. Then we are interested in mapping collections of rv's in $\mathcal{X}^n$ to the n-dimensional simplex $\mathbb{S}^n$ (see, Section 3.2.1 for a definition). In this paper, because of the nature of the capital allocation exercise and for simplicity, our working choice is the map $C : \mathcal{X}^n_+ \to \mathbb{S}^n$, such that

$$C_i(\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)}) = \Gamma_i^{(\kappa_i)} / \Gamma_+^{(\kappa^*)} = R_i, \tag{3.14}$$

yet other compositional maps are possible to have, and we indeed consider another compositional map in the applications part of the paper. Compositions (3.14) are the main object of our study in this section.

Recall that $\mathrm{B}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^n_+$, denotes the multivariate Beta function. The following assertion establishes the joint pdf of random compositions (3.14).

Theorem 7. *Let $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} \sim MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, that is the distribution of the basis vector is multivariate mixed-gamma, and let $\boldsymbol{R} = (R_1, \ldots, R_n)$ be a vector of random compositions (3.14). Then the joint pdf of the rv $\boldsymbol{R}$ is given by*

$$f_{\boldsymbol{R}}(r_1, \ldots, r_n) = \sum_{\boldsymbol{k} \in \mathbb{N}_0} \frac{p_{\boldsymbol{\kappa}}(\boldsymbol{k})}{\mathrm{B}(\boldsymbol{\gamma}_{\boldsymbol{k}})} \prod_{i=1}^n \frac{1}{\beta_i} \left(\frac{r_i}{\beta_i}\right)^{\gamma_{k_i}-1} \left(\sum_{i=1}^n \frac{r_i}{\beta_i}\right)^{-\sum_{i=1}^n \gamma_{k_i}} \tag{3.15}$$

for all $(r_1, \ldots, r_n) \in \mathbb{S}^n$, where $\gamma_{k_i} = \gamma_i + k_i$, $i = 1, \ldots, n$ and $\boldsymbol{\gamma}_{\boldsymbol{k}} = (\gamma_{k_1}, \ldots, \gamma_{k_n})$.

Proof. We begin with the joint pdf of the basis rv $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} = (\Gamma_1^{\kappa_1}, \ldots, \Gamma_n^{\kappa_n})$ (Definition 9)

$$f_{\mathbf{\Gamma}^{(\boldsymbol{\kappa})}}(x_1, \ldots, x_n) \ = \ \sum_{\boldsymbol{k} \in \mathbb{N}_0^n} p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \prod_{i=1}^n \frac{e^{-x_i/\beta_i} x_i^{\gamma_{k_i}-1}}{\beta_i^{\gamma_{k_i}} \Gamma(\gamma_{k_i})}, \ (x_1, \ldots, x_n) \in (0, \ \infty)^n$$

For $\Gamma_+^{(\kappa^*)} = \sum_{i=1}^n \Gamma_i^{(\kappa_i)}$, consider the change of variables $R_i = \Gamma_i^{(\kappa_i)}/\Gamma_+^{(\kappa^*)}$ and so $\Gamma_i^{(\kappa_i)} = R_i\,\Gamma_+^{(\kappa^*)}$, $i = 1, \ldots, n$. Since the corresponding Jacobian is $(\Gamma_+^{(\kappa^*)})^{n-1}$, we have, for $(r_1, \ldots, r_n) \in \mathbb{S}^n$ and $s \in \mathbb{R}_+$,

$$f_{\boldsymbol{R},\Gamma_+^{(\kappa^*)}}(r_1, \ldots, r_n, s) \;=\; f_{\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})}}(r_1 s, \cdots, r_n s) s^{n-1} = \sum_{\boldsymbol{k} \in \mathbb{N}_0^n} p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \prod_{i=1}^n \frac{r_i^{\gamma_{k_i}-1}}{\beta_i^{\gamma_{k_i}}\Gamma(\gamma_{k_i})}\; s^{\sum_{i=1}^n \gamma_{k_i}-1} e^{-s\sum_{i=1}^n r_i/\beta_i}. \quad (3.16)$$

The integration

$$f_{\boldsymbol{R}}(r_1, \ldots, r_n) \;=\; \int_0^\infty f_{\boldsymbol{R},\Gamma_+^{(\kappa^*)}}(r_1, \ldots, r_n, s)\mathrm{d}s$$

completes the proof of the theorem. □

Obviously, when, for all $i = 1, \ldots, n$, the rv κ_i is degenerate in the sense that there exist $\boldsymbol{k} \in \mathbb{N}_0^n$, such that $p_{\kappa}(\boldsymbol{k}) = 1$, then pdf (3.15) reduces to the pdf of the scaled Dirichlet distribution, heuristically studied in, e.g., [87] and [88]. If an additional assumption that the scale parameters are chosen such that $\beta_1 = \cdots = \beta_n$ is made, then pdf (3.15) coincides with pdf (3.5). Motivated by this observation, we call the new generalized Dirichlet put forward herein, mixed-scaled Dirichlet. Succinctly, we write $\boldsymbol{R} \sim Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ are vectors of positive parameters, and $p_{\boldsymbol{\kappa}}$ is the joint PMF of the rv $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)$.

Note 3. It is seemingly worthwhile noticing that since the rv $\boldsymbol{R} = (R_1, \ldots, R_n)$, admitting stochastic representation (3.14), must be such that $R_1 + \cdots + R_n = 1$ almost surely, we can put $r_n = 1 - \sum_{i=1}^{n-1} r_i$ in joint pdf (3.15). Then, for the last component of the just-mentioned equation, we have

$$\left(\sum_{i=1}^n \frac{r_i}{\beta_i}\right)^{-\sum_{i=1}^n \gamma_{k_i}} = \beta_n^{\sum_{i=1}^n \gamma_{k_i}} \left[1 + \sum_{i=1}^{n-1} (\beta_n/\beta_i - 1)\; r_i\right]^{-\sum_{i=1}^n \gamma_{k_i}},$$

where $(r_1, \ldots, r_n) \in \mathbb{S}^n$. It is consequently easy to notice that in the case of the equal scale

parameters, $\beta_1 = \cdots = \beta_n$, we have

$$1 + \sum_{i=1}^{n-1}(\beta_n/\beta_i - 1)r_i = 1 \text{ for all } (r_1, \ldots, r_n) \in \mathbb{S}^n,$$

and joint pdf (3.15) reduces to that of a mixed Dirichlet distribution.

We now proceed to study the marginalization properties of the class of mixed-scaled Dirichlet distributions. As it is rather challenging to integrate joint pdf (3.15) directly, we make use of the associated stochastic representation instead [e.g., 92, for a similar approach within the study of the classical Dirichlet distributions].

Clearly, as $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathbb{S}^n$, we have that its lower dimensional margins are in $\mathbb{V}^n$ (see, Section 3.2.1 for details). More formally, for $\mathcal{I} \subseteq \{1, \ldots, n\}$, let $\boldsymbol{R}_{\mathcal{I}} = \{R_i : i \in \mathcal{I}\} \in \mathbb{V}^{|\mathcal{I}|}$, where $|\mathcal{I}|$ denotes the cardinality of the set $\mathcal{I}$. When checking the marginalization property for $\mathbb{S}^n \ni \boldsymbol{R} \sim Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, we aim to explore whether the distribution of the random pair $(\boldsymbol{R}_{\mathcal{I}}, R^*_{\mathcal{I}^c}) \in \mathbb{S}^{|\mathcal{I}|+1}$, where $\mathcal{I}^c$ denotes the complement of $\mathcal{I} \subseteq \{1, \ldots, n\}$ and

$$R^*_{\mathcal{I}^c} = \frac{\sum_{i\in\mathcal{I}^c} \Gamma_i^{(\kappa_i)}}{\sum_{i\in\mathcal{I}} \Gamma_i^{(\kappa_i)} + \sum_{i\in\mathcal{I}^c} \Gamma_i^{(\kappa_i)}}.$$

is also mixed-scaled Dirichlet.

Define $\gamma^*_{\mathcal{I}^c} = \sum_{i\in\mathcal{I}^c} \gamma_i$ and $\beta^*_{\mathcal{I}^c} = \bigwedge_{i\in\mathcal{I}^c} \beta_i$, we are now ready to prove that the class of mixed-scaled Dirichlet distributions is closed under the marginalization of any order.

Theorem 8. *The rv $\boldsymbol{R} \sim Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$ with pdf (3.15) is closed under marginalizations of arbitrary order. Specifically, we have $\mathbb{S}^{|\mathcal{I}|+1} \ni (\boldsymbol{R}_{\mathcal{I}}, R^*_{\mathcal{I}^c}) \sim Dir((\boldsymbol{\gamma}_{\mathcal{I}}, \gamma^*_{\mathcal{I}^c}), (\boldsymbol{\beta}_{\mathcal{I}}, \beta^*_{\mathcal{I}^c}), p_{(\boldsymbol{\kappa}_{\mathcal{I}}, \kappa^*_{\mathcal{I}^c})})$, where $\square_{\mathcal{I}} = \{\square_i : i \in \mathcal{I}\}$, "$\square$" can be any one of $\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\kappa}$, and the joint PMF*

$$p_{(\boldsymbol{\kappa}_{\mathcal{I}}, \kappa^*_{\mathcal{I}^c})}(\boldsymbol{k}_{\mathcal{I}}, m) = \sum_{j=0}^{m} \sum_{\sum_{v\in\mathcal{I}^c} k_v = j} \left[p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \sum_{\substack{\sum_{v\in\mathcal{I}^c} y_v = m-j \\ y_v \in \mathbb{N}_0}} \prod_{i\in\mathcal{I}^c} \frac{\Gamma(\gamma_{k_i} + y_i)}{\Gamma(\gamma_{k_i}) y_i!} \left(\frac{\beta^*_{\mathcal{I}^c}}{\beta_i}\right)^{\gamma_{k_i}} \left(1 - \frac{\beta^*_{\mathcal{I}^c}}{\beta_i}\right)^{y_i} \right] \quad (3.17)$$

for $(\boldsymbol{k}_{\mathcal{I}}, m) \in \mathbb{N}_0^{|\mathcal{I}|+1}$.

Proof. We repartition the rv $\boldsymbol{R}$ as follows

$$R_i = \frac{\Gamma_i^{(\kappa_i)}}{\sum_{i\in\mathcal{I}}\Gamma_i^{(\kappa_i)} + \sum_{i\in\mathcal{I}^c}\Gamma_i^{(\kappa_i)}},\ i\in\mathcal{I} \text{ and } R^*_{\mathcal{I}^c} = \frac{\sum_{i\in\mathcal{I}^c}\Gamma_i^{(\kappa_i)}}{\sum_{i\in\mathcal{I}}\Gamma_i^{(\kappa_i)} + \sum_{i\in\mathcal{I}^c}\Gamma_i^{(\kappa_i)}}.$$

Theorem 6 implies

$$\sum_{i\in\mathcal{I}^c}\Gamma_i^{(\kappa_i)} \sim MG(\gamma^*_{\mathcal{I}^c}, \beta^*_{\mathcal{I}^c}, p_{\kappa^*_{\mathcal{I}^c}}),$$

and

$$\kappa^*_{\mathcal{I}^c} = \sum_{i\in\mathcal{I}^c}(\kappa_i + N_i(\kappa_i)), \tag{3.18}$$

where the rv's $N_i(\kappa_i) \sim NB(\gamma_{\kappa_i}, \beta^*_{\mathcal{I}^c}/\beta_i),\ i\in\mathcal{I}^c$ are conditionally independent given the rv $\boldsymbol{\kappa}$.

Therefore we conclude that

$$\mathbb{S}^{|\mathcal{I}|+1} \ni (\boldsymbol{R}_{\mathcal{I}}, R^*_{\mathcal{I}^c}) \sim Dir((\boldsymbol{\gamma}_{\mathcal{I}}, \gamma^*_{\mathcal{I}^c}), (\boldsymbol{\beta}_{\mathcal{I}}, \beta^*_{\mathcal{I}^c}), p_{(\boldsymbol{\kappa}_{\mathcal{I}}, \kappa^*_{\mathcal{I}^c})}),$$

where the joint PMF of rv $(\boldsymbol{\kappa}_{\mathcal{I}}, \kappa^*_{\mathcal{I}^c})$ can be computed via expression (3.12) in Theorem 6. This completes the proof of the theorem. □

An immediate consequence of the just-proved closure under marginalization of any order is that in the context of the mixed-scaled Dirichlet class of distributions, that is for $\boldsymbol{R} = (R_1, \ldots, R_n) \sim Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, the joint k-dimensional pdf's, $k < n$, can be derived with the help of Theorems 7 and 8. For an illustration, we next report the univariate and bivariate pdf's. Marginal pdf's of higher dimensions can be computed analogously. For notational convenience, we let $\mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{N}_{(\mathcal{I})} = \mathcal{N} \setminus \mathcal{I}$ for $\mathcal{I} \subseteq \mathcal{N}$.

Set $\mathcal{I} = \{i\}$, then the univariate pdf of the rv $R_i,\ i = 1, \ldots, n$ is

$$f_{R_i}(r) = \sum_{k_i, k^*\in\mathbb{N}_0} \frac{p_{(\kappa_i, \kappa^*_{\mathcal{N}_{(i)}})}(k_i, k^*)\left(\beta^*_{\mathcal{N}_{(i)}}/\beta_i\right)^{\gamma_{k_i}}}{\mathrm{B}(\gamma_{k_i}, \gamma_{k^*})} r^{\gamma_{k_i}-1}(1-r)^{\gamma_{k^*}-1}\left[1 + \left(\frac{\beta^*_{\mathcal{N}_{(i)}}}{\beta_i} - 1\right) r\right]^{-(\gamma_{k_i}+\gamma_{k^*})},$$

where $r \in [0,1]$, $\gamma_{k_i} = \gamma_i + k_i$, $\gamma_{k^*} = \gamma^*_{\mathcal{N}_{(i)}} + k^*$ and the joint PMF $p_{(\kappa_i, \kappa^*_{\mathcal{N}_{(i)}})}$ follows from Equation (3.17).

To find the bivariate pdf of the random pair (R_i, R_j), $i \neq j \in \{1, \ldots, n\}$, we set $\mathcal{I} = \{i,\ j\}$, and obtain

$$f_{R_i,R_j}(r_i, r_j) = \sum_{k_i,k_j,k^* \in \mathbb{N}_0} \frac{p_{(\kappa_i,\kappa_j,\kappa^*_{\mathcal{N}_{(i,j)}})}(k_i, k_j, k^*)}{\mathrm{B}(\gamma_{k_i}, \gamma_{k_j}, \gamma_{k^*})} \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_i}\right)^{\gamma_{k_i}} \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_j}\right)^{\gamma_{k_j}} r_i^{\gamma_{k_i}-1} r_j^{\gamma_{k_j}-1} (1 - r_i - r_j)^{\gamma_{k^*}-1} \left[1 + \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_i} - 1\right) r_i + \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_j} - 1\right) r_j\right]^{-(\gamma_{k_i}+\gamma_{k_j}+\gamma_{k^*})},$$

where $(r_i, r_j) \in \mathbb{V}^2$, $\gamma_{k_i} = \gamma_i + k_i$, $\gamma_{k_j} = \gamma_j + k_j$, $\gamma_{k^*} = \gamma^*_{\mathcal{N}_{(i,j)}} + k^*$ and the joint PMF $p_{(\kappa_i,\kappa_j,\kappa^*_{\mathcal{N}_{(i)}})}$ can be again formulated with the help of Equation (3.17).

Note 4. rv $X \in L^\infty$ is said to be distributed generalized three-parameter beta if the associated pdf is given by [77]

$$f_X(x) = \frac{\lambda^a}{B(a,b)} \frac{x^{a-1}(1-x)^{b-1}}{(1 + (\lambda - 1)x)^{a+b}},\ x \in [0,1], \tag{3.19}$$

where $a, b, \lambda > 0$ are parameters. Succinctly, we write $X \sim GB(a, b, \lambda)$. Some distributional properties of the class of GB distributions are discussed in [63]. It is not difficult to see that the univariate marginal distributions of the mixed-scaled Dirichlet distributions are GB with random shape parameters. Namely, we have $R_i \sim GB(\gamma_i + \kappa_i, \gamma^*_{\mathcal{N}_{(i)}} + \kappa^*_{\mathcal{N}_{(i)}}, \beta^*_{\mathcal{N}_{(i)}}/\beta_i)$ where $\kappa^*_{\mathcal{N}_{(i)}}$ is distributed as per (3.18).

Next we proceed to study the moment formulas for the mixed-scaled Dirichlet class of distributions. In this respect, the hypergeometric function plays an important role, and it is defined as [58]

$${}_{q+1}\mathrm{F}_q(a_1, \ldots, a_{q+1}; b_1, \ldots, b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k, \ldots, (a_{q+1})_k}{(b_1)_k, \ldots, (b_q)_k} \frac{z^k}{k!},\ |z| < 1, \tag{3.20}$$

where $(x)_n = \Gamma(x+n)/\Gamma(x)$ denotes the Pochhammer symbol. We also need the Appell's

F_1 function, which is given by

$$\mathrm{F}_1(a; b_1, b_2; c; x, y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{(a)_{m+n}(b_1)_m(b_2)_n}{(c)_{m+n}} \frac{x^m}{m!} \frac{y^n}{n!}, \ |x|, |y| < 1.$$

It is noteworthy that the arguments domains of the aforementioned special functions can be extended by analytic continuation. Also, there is a rich body of literature devoted to the study of both ${}_{q+1}\mathrm{F}_q$ and F_1, and the corresponding computational methods have been implemented in a variety of software packages.

Theorem 9. *Let $\boldsymbol{R} = (R_1, \ldots, R_n) \sim Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$ be a rv distributed mixed-scaled Dirichlet. Then the $r(\in \mathbb{R}_+)$-th order moment of the rv R_i, $i = 1, \ldots, n$ is given by*

$$\begin{aligned}\mathbb{E}[R_i^r] \ = \ & \sum_{k_i, k^* \in \mathbb{N}_0} p_{(\kappa_i, \kappa^*_{\mathcal{N}_{(i)}})}(k_i, k^*) \left(\frac{\beta^*_{\mathcal{N}_{(i)}}}{\beta_i} \right)^{\gamma_{k_i}} \frac{\Gamma(\gamma_{k_i} + \gamma_{k^*})\Gamma(\gamma_{k_i} + r)}{\Gamma(\gamma_{k_i} + \gamma_{k^*} + r)\Gamma(\gamma_{k_i})} \\ & {}_2\mathrm{F}_1 \left(\gamma_{k_i} + r, \gamma_{k_i} + \gamma_{k^*}; \gamma_{k_i} + \gamma_{k^*} + r; 1 - \frac{\beta^*_{\mathcal{N}_{(i)}}}{\beta_i} \right),\end{aligned}$$

where $\gamma_{k_i} = \gamma_i + k_i$, $\gamma_{k^*} = \gamma^*_{\mathcal{N}_{(i)}} + k^*$ and the joint PMF $p_{(\kappa_i, \kappa^*_{\mathcal{N}_{(i)}})}$ is defined according to Equation (3.17).

Furthermore, for $r_i, r_j \in \mathbb{R}_+$, the joint higher order moments of R_i and R_j, $i \neq j \in \{1, \ldots, n\}$, are given by

$$\begin{aligned}\mathbb{E}[R_i^{r_i} R_j^{r_j}] \ = \ & \sum_{k_i, k_j, k^* \in \mathbb{N}_0} p_{(\kappa_i, \kappa_j, \kappa^*_{\mathcal{N}_{(i,j)}})}(k_i, k_j, k^*) \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_i} \right)^{\gamma_{k_i}} \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_j} \right)^{\gamma_{k_j}} \\ & \frac{\Gamma(\gamma_{k_i} + r_i)}{\Gamma(\gamma_{k_i})} \frac{\Gamma(\gamma_{k_j} + r_j)}{\Gamma(\gamma_{k_j})} \frac{\Gamma(\gamma_{k_i} + \gamma_{k_j} + \gamma_{k^*})}{\Gamma(\gamma_{k_i} + \gamma_{k_j} + \gamma_{k^*} + r_i + r_j)} \ h(k_i, k_j, k^*),\end{aligned}$$

where

$$h(k_i, k_j, k^*) = \mathrm{F}_1 \left(\gamma_{k_i} + \gamma_{k_j} + \gamma_{k^*}; \gamma_{k_i} + r_i, \gamma_{k_j} + r_j; \gamma_{k_i} + \gamma_{k_j} + \gamma_{k^*} + r_i + r_j; 1 - \frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_i}, 1 - \frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_j} \right),$$

and $\gamma_{k_i} = \gamma_i + k_i$, $\gamma_{k_j} = \gamma_j + k_j$, $\gamma_{k^*} = \gamma^*_{\mathcal{N}_{(i,j)}} + k^*$, with $p_{(\kappa_i, \kappa_j, \kappa^*_{\mathcal{N}_{(i)}})}$ being per Equation (3.17).

Proof. The r-th order moment formula follows from Note 4 and [63], whereas the joint mo-

ment formula is obtained directly by the integral representation of the Appell's F_1 function [see, Equation (9.184) in 58]. This completes the proof of the theorem. □

Note 5. The covariance between any pair of rv's R_i and R_j within the mixed-scaled Dirichlet class can be readily computed via the moment formulas in Theorem 9. Interestingly, unlike for the classical Dirichlet distribution with pdf (3.5), the Pearson coefficient of correlation in the context of the mixed-scaled Dirichlet class of distributions is not necessarily negative. For instance, consider a simple example in which the rv's $\kappa_1,\ \kappa_2$ and κ_3 are all zero almost surely, and $\gamma_i \equiv 1,\ i = 1, \ldots, 3,\ \beta_1 = \beta_2 = 1/20,\ \beta_3 = 1$. Then an application of the moment formulas in Theorem 9 yields $\mathrm{Corr}(\mathrm{R}_1, \mathrm{R}_2) = 0.24$.

There is no known closed-form expression for computing the moments of the scaled Dirichlet distribution, that is of the mixed-scaled Dirichlet $Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$ when the rv κ_i, $i = 1, \ldots, n$ is assumed to be degenerate [e.g., 87, 92, for details]. In this respect, Theorem 9 provides analytical and conveniently computable expressions for the desired moment formulas. Specifically, set $\kappa_i \equiv 0$ in Theorem 9, then, for $r \in \mathbb{R}_+$ and $i = 1, \ldots, n$,

$$\mathbb{E}[R_i^r] \;=\; \left(\frac{\beta^*_{\mathcal{N}_{(i)}}}{\beta_i}\right)^{\gamma_i} \frac{\Gamma(\gamma_i + r)}{\Gamma(\gamma_i)} \sum_{k \in \mathbb{N}_0} p_{\kappa^*_{\mathcal{N}_{(i)}}}(k) \frac{\Gamma(\gamma^* + k)}{\Gamma(\gamma^* + k + r)}\ {}_2\mathrm{F}_1\left(\gamma_i + r, \gamma^* + k; \gamma^* + k + r; 1 - \frac{\beta^*_{\mathcal{N}_{(i)}}}{\beta_i}\right),$$

where $\gamma^* = \sum_{i=1}^n \gamma_i$, $\beta^*_{\mathcal{N}_{(i)}} = \bigwedge_{j \in \mathcal{N}_{(i)}} \beta_i$, and $\kappa^*_{\mathcal{N}_{(i)}} \overset{d}{=} \sum_{j \in \mathcal{N}_{(i)}} N_j$ with the rv's N_j being mutually independent and $N_j \sim NB(\gamma_j, \beta^*_{\mathcal{N}_{(i)}}/\beta_j)$. The PMF of $\kappa^*_{\mathcal{N}_{(i)}}$ can be computed directly via (3.13) or recursively via (3.11).

Similarly, for $r_i, r_j \in \mathbb{R}_+$, $i \neq j \in \{1, \ldots, n\}$,

$$\mathbb{E}[R_i^{r_i} R_j^{r_j}] = \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_i}\right)^{\gamma_i} \left(\frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_j}\right)^{\gamma_j} \frac{\Gamma(\gamma_i + r_i)}{\Gamma(\gamma_i)} \frac{\Gamma(\gamma_j + r_j)}{\Gamma(\gamma_j)} \sum_{k \in \mathbb{N}_0} p_{\kappa^*_{\mathcal{N}_{(i,j)}}}(k) \frac{\Gamma(\gamma^* + k)}{\Gamma(\gamma^* + k + r_i + r_j)} \tilde{h}(r_i, r_j, k),$$

where

$$\tilde{h}(r_i, r_j, k) = \mathrm{F}_1\left(\gamma^* + k; \gamma_i + r_i, \gamma_j + r_j; \gamma^* + k + r_i + r_j; 1 - \frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_i}, 1 - \frac{\beta^*_{\mathcal{N}_{(i,j)}}}{\beta_j}\right),$$

$\beta^*_{\mathcal{N}_{(i,j)}} = \bigwedge_{j\in\mathcal{N}_{(i,j)}} \beta_i$, and $\kappa^*_{\mathcal{N}_{(i,j)}} \overset{d}{=} \sum_{j\in\mathcal{N}_{(i,j)}} N_j$ with the rv's N_j being mutually independent and $N_j \sim NB(\gamma_j, \beta^*_{\mathcal{N}_{(i,j)}}/\beta_j)$.

The moment formulas above involve infinite series. For computational purposes, one may use the first $m+1$ terms of the series, where $m \in \mathbb{N}$ is such that the desired accuracy is attained. Bounds, $R_m(f) = \sum_{k=0}^{\infty} f_k - \sum_{k=0}^{m} f_k$, for the resulting truncation error can be obtained as

$$R_m(\mathbb{E}[R_i^r]) < 1 - \sum_{k=0}^{m} p_{\kappa^*_{\mathcal{N}_{(i)}}}(k) \qquad \text{and} \qquad R_m(\mathbb{E}[R_i^{r_i} R_j^{r_j}]) < 1 - \sum_{k=0}^{m} p_{\kappa^*_{\mathcal{N}_{(i,j)}}}(k).$$

We conclude the discussion in this section with a few more properties of the class of mixed-scaled Dirichlet distributions. For this, we need two additional definitions.

Definition 10. For $\mathcal{I} = \{i_1, \ldots, i_j\} \subset \mathcal{N},\ j < n$, the vector

$$\mathbf{S}_{\mathcal{I}} = \left(\frac{\Gamma_{i_1}^{(\kappa_{i_1})}}{\sum_{i\in\mathcal{I}} \Gamma_i^{(\kappa_i)}}, \ldots, \frac{\Gamma_{i_j}^{(\kappa_{i_j})}}{\sum_{i\in\mathcal{I}} \Gamma_i^{(\kappa_i)}} \right)$$

is called a *sub-composition*. The vector $(\Gamma_{i_1}, \ldots, \Gamma_{i_j})$ is called the basis of the sub-composition.

Definition 11. Let $\{\mathcal{I}_k\}_{k=1}^m$ where $\mathcal{I}_k = \{i_{k,1}, \ldots, i_{k,j_k}\} \subset \mathcal{N},\ j, m < n$, denote a disjoint coverage of the set $\{1, \ldots, n\}$, that is $\cup_k \mathcal{I}_k = \{1, \ldots, n\}$ and $\mathcal{I}_k \cap \mathcal{I}_h = \emptyset$ for $k \neq h$. Each set $\mathcal{I}_k$ gives rise to the sub-composition $\mathbf{S}_k = \mathbf{S}_{\mathcal{I}_k}$ with the corresponding basis $\left(\Gamma_{i_{k,1}}^{(\kappa_{i_{k,1}})}, \ldots, \Gamma_{i_{k,j_k}}^{(\kappa_{i_{k,j_k}})}\right)$. Then the vector

$$\boldsymbol{R}_{\boldsymbol{\mathcal{I}}} = \left(\frac{\sum_{i\in\mathcal{I}_1} \Gamma_i^{(\kappa_i)}}{\sum_{i=1}^{n} \Gamma_i^{(\kappa_i)}}, \ldots, \frac{\sum_{i\in\mathcal{I}_m} \Gamma_i^{(\kappa_i)}}{\sum_{i=1}^{n} \Gamma_i^{(\kappa_i)}} \right),\ \boldsymbol{\mathcal{I}} = \{\mathcal{I}_1, \ldots, \mathcal{I}_m\},$$

is called an *amalgamation*.

Roughly speaking, sub-compositions and amalgamations in the context of the probability distributions on $\mathbb{S}^n$ are akin to marginalizations of arbitrary order and convolutions in the context of the probability distributions on $\mathbb{R}^n_{0,+}$. We next prove that the class of mixed-scaled Dirichlet distributions is closed with respect to both notions.

Proposition 12. *The rv* $\boldsymbol{R} = (R_1, \ldots, R_n)$ *with joint pdf* (3.15) *is closed under sub-compositions and amalgamations. Specifically, we have, for* $\boldsymbol{R} \sim Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$,

(i) $\mathbb{S}^{|\mathcal{I}|} \ni \boldsymbol{S}_{\mathcal{I}} \sim Dir(\boldsymbol{\gamma}_{\mathcal{I}}, \boldsymbol{\beta}_{\mathcal{I}}, p_{\boldsymbol{\kappa}_{\mathcal{I}}})$, *where* $\square_{\mathcal{I}} = \{\square_i : i \in \mathcal{I}\}$ *and "*$\square$*" can be any one of* γ, κ *and* β*;*

(ii) $\mathbb{S}^m \ni \boldsymbol{R}_{\boldsymbol{\mathcal{I}}} \sim Dir(\boldsymbol{\gamma}^*_{\boldsymbol{\mathcal{I}}}, \boldsymbol{\beta}^*_{\boldsymbol{\mathcal{I}}}, p_{\boldsymbol{\kappa}^*_{\boldsymbol{\mathcal{I}}}})$, *where* $\square_{\boldsymbol{\mathcal{I}}} = \{\square_{\mathcal{I}_j} : j = 1, \ldots, m\}$, *"*$\square$*" can be any one of* γ^*, β^* *and* κ^**, such that*

$$\gamma^*_{\mathcal{I}_j} = \sum_{i \in \mathcal{I}_j} \gamma_i \quad \textit{and} \quad \beta^*_{\mathcal{I}_j} = \bigwedge_{i \in \mathcal{I}_j} \beta_i \quad \textit{for} \quad j = 1, \ldots, m.$$

Also, the rv $\boldsymbol{\kappa}^*_{\mathcal{I}} = (\kappa^*_{\mathcal{I}_1}, \ldots, \kappa^*_{\mathcal{I}_m})$ *has the coordinates*

$$\kappa^*_{\mathcal{I}_j} \overset{d}{=} \sum_{i \in \mathcal{I}_j} (\kappa_i + N_i(\kappa_i)),$$

where the rv's N_i *are conditionally independent given the rv* $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_m)$ *and such that* $N_i(\kappa_i) \sim NB(\gamma_i, {\beta^*}_{\mathcal{I}_i}/\beta_i)$, $i \in \mathcal{I}_j$. *For* $\boldsymbol{k}^* = (k^*_1, \ldots, k^*_m) \in \mathbb{N}_0^m$, *the joint PMF of the rv* $\boldsymbol{\kappa}^*_{\boldsymbol{\mathcal{I}}}$ *can be computed via*

$$p_{\boldsymbol{\kappa}^*_{\boldsymbol{\mathcal{I}}}}(\boldsymbol{k}^*) = \sum_{\substack{j_v \in \{0, \ldots, k^*_v\} \\ 1 \le v \le m}} \sum_{\substack{\sum_{i \in \mathcal{I}_v} k_i = j_v \\ 1 \le v \le m}} p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \prod_{v=1}^{m} q_v(k^*_v - j_v),$$

where $q_v(z) = \mathbb{P}(\sum_{i \in \mathcal{I}_v} N_{v,i}(k_i) = z)$ *with the rv's* $N_{v,i}$ *being mutually independent and such that* $N_{v,i}(k_i) \sim NB(\gamma_i + k_i, \beta^*_{\mathcal{I}_v}/\beta_i)$, $z \in \mathbb{N}_0$, $v = 1, \ldots, m$, $i \in \mathcal{I}_v$*; the function* q_v *can be computed with the help of Equation (3.13).*

Proof. Assertion (i) follows immediately from, e.g., stochastic representation (3.14). To confirm Assertion (ii), recall that we have already shown that sums of mixed-gamma distributions are also mixed-gamma. That is, due to Theorem 6, we have

$$\sum_{i \in \mathcal{I}_j} \Gamma_i^{(\kappa_i)} \sim MG(\gamma^*_{\mathcal{I}_j}, \beta^*_{\mathcal{I}_j}, p_{\kappa^*_{\mathcal{I}_j}}), \ j = 1, \ldots, m,$$

where $\kappa^*_{\mathcal{I}_j} \overset{d}{=} \sum_{i \in \mathcal{I}_j} (\kappa_i + N_i(\kappa_i))$ with the rv's N_i being mutually independent and such that

$N_i(\kappa_i) \sim NB(\gamma_i + \kappa_i, \beta^*_{\mathcal{I}_v}/\beta_i)$ that are conditionally independent given the rv $\boldsymbol{\kappa}$. The joint PMF of $\boldsymbol{\kappa}^*_{\boldsymbol{\mathcal{I}}} = (\kappa^*_{\mathcal{I}_1}, \ldots, \kappa^*_{\mathcal{I}_m})$ can be computed by conditioning as follows

$$p_{\boldsymbol{\kappa}^*_{\boldsymbol{\mathcal{I}}}}(\boldsymbol{k}^*) = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{\boldsymbol{\kappa}^*_{\boldsymbol{\mathcal{I}}} = \boldsymbol{k}^*\}\middle|\boldsymbol{\kappa}\right]\right] = \mathbb{E}\left[\prod_{j=1}^{m} \mathbb{P}\left(\sum_{i\in\mathcal{I}_j} N_{j,i}(\kappa_i) = k^*_j - \sum_{i\in\mathcal{I}_j}\kappa_i\right)\right],$$

where the rv's $N_{j,i}$ are mutually independent and such that $N_{j,i}(k_i) \sim NB(\gamma_i + k_i, \beta^*_{\mathcal{I}_j}/\beta_i)$ for $i \in (1, \ldots, n)$, $j \in (1, \ldots, m)$, $k_i \in \mathbb{N}_0$. This yields the closure under amalgamations property. The proof is completed. □

3.5 Applications

To summarize the discussion hitherto, we have assumed that $n(\in \mathbb{N})$ BUs of a financial entity are formally described by a rv $\boldsymbol{X} = (X_1, \ldots, X_n)$ that has a mixed-gamma distribution, $MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$. With the help of compositional map (3.14) (other maps of practical interest can also be used), we have obtained the random proportions $\boldsymbol{R} = (R_1, \ldots, R_n)$ distributed mixed-scaled Dirichlet, $Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ are vectors of positive parameters, and $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)$ is a rv. The just-mentioned parameters of the distribution of the rv $\boldsymbol{R} = (R_1, \ldots, R_n)$ reflect the BUs' inputs, and therefore have represented the bottom-up approach to allocating risk capital. The top-down approach has been modelled by the rv Z.

We have mentioned that, if the rv Z is assumed independent on the rv's $R_1, \ldots, R_n$, as well as if compositional map (3.14) is considered, the weighted allocation admits the following remarkably simple form

$$A_w(X_i, Z) = \mathbb{E}[R_i] \times H_w(Z), \ i = 1, \ldots, n, \tag{3.21}$$

for an arbitrary legitimate weight function w. The simple form of Equation (3.21) carries over to other compositional maps when the assumption of independence of the rv's Z and $R_1, \ldots, R_n$ is kept in place. For instance, for $p \in [0,\ 1)$, $s_p = \mathrm{VaR_p(S)}$ and $S = \sum_{i=1}^{n} X_i$,

assume that $C : \mathcal{X}^n \to \mathbb{S}^n$ is such that

$$C_{i,p}(X_1, \ldots, X_n) = \frac{1}{1-p} \frac{X_i}{S} \mathbf{1}\{S > s_p\}. \tag{3.22}$$

Obviously, the map C_p is of importance when those random proportions $R_1, \ldots, R_n$ that occur under the extreme scenarios that the aggregate risk rv exceeds the Value-at-Risk, $\mathrm{VaR}_p(S)$, are of interest. In this more general case, indeed C_p coincides with compositional map (3.14) for $p \downarrow 0$, we have, for $\mathbb{P}(S > s_p) = 1 - p \neq 0$,

$$A_{w,p}(X_i, Z) = \mathbb{E}[R_i |\ S > s_p] \times H_w(Z),\ i = 1, \ldots, n, \tag{3.23}$$

which can be computed using (3.16), that is given that the joint pdf of the rv's R_i and S is

$$f_{R_i,S}(r,s) = \sum_{k_i,k^* \in \mathbb{N}_0} p_{(\kappa_i, \kappa^*_{\mathcal{N}_{(i)}})}(k_i, k^*) \frac{r^{\gamma_{k_i}-1}(1-r)^{\gamma_{k^*}-1}}{\beta_i^{\gamma_{k_i}} (\beta^*_{\mathcal{N}_{(i)}})^{\gamma_{k_i}} \Gamma(\gamma_{k_i}) \Gamma(\gamma_{k^*})}\ s^{\gamma_{k_i}+\gamma_{k^*}-1} e^{-s[r/\beta_i + (1-r)/\beta^*_{\mathcal{N}_{(i)}}]},$$

where $\beta^*_{\mathcal{N}_{(i)}} = \wedge_{i \in \mathcal{N}_{(i)}} \beta_i$, $\gamma_{k^*} = \gamma^*_{\mathcal{N}_{(i)}} + k^*$, $\kappa^*_{\mathcal{N}_{(i)}} \overset{d}{=} \sum_{j \in \mathcal{N}_{(i)}} N_j$, where $N_j \sim NB(\gamma_j, \beta^*_{\mathcal{N}_{(i)}}/\beta_j)$, PMF $p_{(\kappa_i, \kappa^*_{\mathcal{N}_{(i)}})}$ follows from (3.17), and $r \in [0,1]$, $s \in \mathbb{R}_+$. This line of reasoning yields a counterpart of the CTE-based risk capital allocation defined for distributions on the simplex. In a similar fashion, other members of the class of weighted risk capital allocations can be computed for the random proportions $\boldsymbol{R} = (R_1, \ldots, R_n)$.

The rest of this section is divided into two subsections. Namely, first, we outline a method to estimate the parameters of the mixed-scaled Dirichlet distributions put forward in this paper, and second, we present a few applications to the risk capital allocation problem.

3.5.1 Estimation of parameters

Consider observations $\boldsymbol{x} = (x_{11}, \ldots, x_{nd})$, which represent risks arising from $n (\in \mathbb{N})$ BUs of a financial entity. Our goal is to estimate the parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ as well as the PMF of the rv $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)$ that characterize the mixed-gamma distributions $MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, and so the mixed-scaled Dirichlet distributions $Dir(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$. To this end, assume that the rv $\boldsymbol{\kappa}$ has a bounded support, $\mathcal{M} \subset \mathbb{N}_0^n$, say. Then the multivariate

mixed-gamma distributions establish a class of finite mixtures and, e.g., pdf (3.9) can be written as

$$f_{\mathbf{\Gamma}^{(\boldsymbol{\kappa})}}(x_1,\ldots,x_n) \;=\; \sum_{\boldsymbol{k}\in\mathcal{M}} p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \prod_{i=1}^{n} \frac{1}{\Gamma(\gamma_{k_i})} e^{-x_i/\beta_i} x_i^{\gamma_{k_i}-1} \beta_i^{-\gamma_{k_i}},\ (x_1,\ldots,x_n)\in\mathbb{R}_+^n.$$

The Expectation-Maximization (EM) algorithm is a common choice for estimating the parameters of finite mixtures. It was proposed in [32] [also, 120] for statistical estimation in the contexts with incomplete data. We refer to, e.g., [70] and [11] for the applications of the EM algorithm to certain multivariate Exponential and Pareto distributions. For obvious reasons, we ground the estimation procedure herein in the one developed in [75] [also, 113] for the class of mixed-Erlang distributions. However, there are some differences. Namely, besides the natural restriction on the space of shape parameters, the estimation procedures presented in ibid assume common scale parameters $\beta_1=\cdots=\beta_n$ and so have to be adjusted to fit the context of the mixed-gamma distributions proposed in this paper. We sketch the algorithm next.

Recall that we need to estimate the parameters $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_n)$, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_n)$ and $p_{\boldsymbol{\kappa}}(\boldsymbol{k})$, $\boldsymbol{k}\in\mathcal{M}\subset\mathbb{N}_0^n$. In order to initialize the parameters, including the choice of the set $\mathcal{M}\subset\mathbb{N}_0^n$, we adopt the procedure in [75]. Then we conduct the "expectation" (E) stage. That is, for $s\in\mathbb{N}_0$, let $\mathbf{\Psi}^{(s)}=(p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k}),\boldsymbol{\beta}^{(s)},\boldsymbol{\gamma}^{(s)})$ denote the vector of parameters that results from the s-th iteration of the algorithm. The conditional expectation of the complete-data likelihood can be computed via

$$\begin{aligned} Q(\mathbf{\Psi}|\ \mathbf{\Psi}^{(s)}) &= \sum_{j=1}^{d}\sum_{\boldsymbol{k}\in\mathcal{M}} \log(p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k}))q(\boldsymbol{k}|\boldsymbol{x}_j,\mathbf{\Psi}^{(s)}) \\ &\quad+\sum_{j=1}^{d}\sum_{\boldsymbol{k}\in\mathcal{M}}\left[\sum_{i=1}^{n}\left((\gamma_i+k_i-1)\log(x_{ij})-\frac{x_{ij}}{\beta_i}-(\gamma_i+k_i)\log(\beta_i)-\log(\Gamma(\gamma_i+k_i))\right)q(\boldsymbol{k}|\boldsymbol{x}_j,\mathbf{\Psi}^{(s)})\right], \end{aligned} \tag{3.24}$$

where, for $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})$, $j = 1, \ldots, d$,

$$q(\boldsymbol{k}|\boldsymbol{x}_j, \boldsymbol{\Psi}^{(s)}) = p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k}) \prod_{i=1}^{n} \frac{e^{-x_{ij}/\beta_i^{(s)}} x_{ij}^{\gamma_i^{(s)}+k_i-1} \beta_i^{-\gamma_i^{(s)}-k_i}}{\Gamma(\gamma_i^{(s)}+k_i)} \Big/ \sum_{\boldsymbol{k}\in\mathcal{M}} p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k}) \prod_{i=1}^{n} \frac{e^{-x_{ij}/\beta_i^{(s)}} x_{ij}^{\gamma_i^{(s)}+k_i-1} \beta_i^{-\gamma_i^{(s)}-k_i}}{\Gamma(\gamma_i^{(s)}+k_i)}$$

is the posterior probability function. The aforementioned conditional expectation, $Q(\boldsymbol{\Psi}|\, \boldsymbol{\Psi}^{(s)})$, serves as the input for the "maximization" (M) stage of the estimation procedure. Namely, in order to find the vector of updated parameters that maximizes (3.24) subject to the constraint $\sum_{\boldsymbol{k}\in\mathcal{M}} p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k}) = 1$, we compute the partial derivatives of $Q(\boldsymbol{\Psi}|\, \boldsymbol{\Psi}^{(s)})$ with respect to $p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k})$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. Equating these partial derivatives to zero leads to the following equations, and thereafter to the parameter vector $\boldsymbol{\Psi}^{(s+1)} = (p_{\boldsymbol{\kappa}}^{(s+1)}(\boldsymbol{k}), \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\gamma}^{(s+1)})$, associated with the $(s+1) \in \mathbb{N}$ iteration of the EM algorithm.

- For $p_{\boldsymbol{\kappa}}^{(s+1)}(\boldsymbol{k})$, we have

$$p_{\boldsymbol{\kappa}}^{(s+1)}(\boldsymbol{k}) = \frac{1}{d} \sum_{j=1}^{d} q(\boldsymbol{k}|\boldsymbol{x}_j, \boldsymbol{\Psi}^{(s)}), \;\; \boldsymbol{k} \in \mathcal{M}.$$

- For $\boldsymbol{\beta}^{(s+1)}$, we solve

$$\beta_i^{(s+1)} = \frac{\sum_{j=1}^{d} x_{ij}}{d \sum_{\boldsymbol{k}\in\mathcal{M}} p_{\boldsymbol{\kappa}}^{(s+1)}(\boldsymbol{k})(\gamma_i^{(s+1)}+k_i)}, \;\; i = 1, \ldots, n.$$

- For $\boldsymbol{\gamma}^{(s+1)}$, we arrive at

$$\sum_{j=1}^{d} \log(x_{ij}) - d \left[\log\left(\sum_{j=1}^{d} \frac{x_{ij}}{d} \right) - \log\left(\sum_{\boldsymbol{k}\in\mathcal{M}} \alpha_{\boldsymbol{k}}^{(s+1)} (\gamma_i^{(s+1)}+k_i) \right) + \sum_{\boldsymbol{k}\in\mathcal{M}} p_{\boldsymbol{\kappa}}^{(s+1)}(\boldsymbol{k}) \; \psi(\gamma_i^{(s+1)}+k_i) \right] = 0,$$

  where $i = 1, \ldots, n$ and $\psi(\cdot)$ denotes the digamma function. The latter system of non-linear equations can be solved numerically with the help of, e.g., the `R` package "`BB`" [112].

The E and M stages iterate unless the improvement in the partial log-likelihood between two consecutive stages falls below a pre-specified threshold.

3.5.2 A numerical example

In this subsection, we offer a numerical example to illustrate the method to allocate risk capital proposed in this paper. We briefly recall that the gist of our method is the suggestion to substitute the commonly employed 'composition of allocations' $C_i(A(X_1, S), \ldots, A(X_n, S))$ with an allocation of the composition $C_i(X_1, \ldots, X_n)$. where $i = 1, \ldots, n$.

In order to construct the desired illustration, we consider an insurance portfolio which comprises three BUs. The rv's representing the risks due to the BUs are distributed Pareto, Log-normal, and gamma, and, more specifically, we set $X_1 \sim Pa(3, 200)$, $X_2 \sim Log\text{-}N(4.1, 1)$, and $X_3 \sim Ga(2, 50)$. The distributions are chosen such that the means are all equal, that is $\mathbb{E}[X_i] = 100$, $i = 1, 2, 3$. Also, these distributions are common choices in actuarial practice [e.g., 15, for examples]. Furthermore, we assume that the dependencies among the rv's X_1, X_2, and X_3 are governed by the Gaussian copula with the correlation matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1.00 & 0.50 & 0.25 \\ 0.50 & 1.00 & -0.50 \\ 0.25 & -0.50 & 1.00 \end{pmatrix},$$

with the entries being motivated by the matrix used in the Quantitative Impact Study published by the Basel Committee.

Then we simulate 1000 samples from the aforementioned set-up, and we fit the proposed multivariate mixed-gamma distribution to the simulated samples, pretending that the true distributions are unknown. Using the estimation method described in Section 3.5.1, we estimate the parameters of the multivariate mixed-gamma distribution, which are summarized in Table 3.1. In addition, Figure 3.1 depicts the pair-wise log transformed density contours and the marginal histograms for the fitted multivariate mixed-gamma distribution, which visually confirm that this class of distributions fits the simulated data well.

Finally, based on the obtained parameters for the multivariate mixed-gamma distribution,

i	β_i	$\gamma_{k_{i1}}$	$\gamma_{k_{i2}}$	$\gamma_{k_{i3}}$	$\gamma_{k_{i4}}$	$\gamma_{k_{i5}}$	$\gamma_{k_{i6}}$	$\gamma_{k_{i7}}$	$\gamma_{k_{i8}}$	$\gamma_{k_{i9}}$	$\gamma_{k_{i10}}$	$\gamma_{k_{i11}}$	$\gamma_{k_{i12}}$
1	27.53	0.98	2.98	13.98	1.98	9.98	0.98	3.98	12.98	30.98	2.98	11.98	35.98
2	13.76	4.13	13.13	72.13	28.13	28.13	2.13	6.13	12.13	30.13	2.13	5.13	6.13
3	14.96	3.19	3.19	3.19	1.19	3.19	8.19	7.19	6.19	6.19	15.19	15.19	14.19
$p_{\boldsymbol{\kappa}}$		0.2156	0.1396	0.0040	0.0260	0.0238	0.2241	0.2074	0.0377	0.0085	0.0778	0.0277	0.0078

Table 3.1: The parameters of the multivariate mixed-gamma distribution fitted against the simulated data.

we obtain the mixed-scaled Dirichlet distribution that describes the joint behaviour of the random proportions, and compute the values of a few risk capital allocation rules, which are presented in Table 3.2.

Table 3.2 hints at the following observations.

- The substitution of the 'composition of allocations' method with the proposed in this paper 'allocation of a composition' method leads to the outcomes of the risk capital allocation exercise that differ in both order and magnitude; e.g., the case of the allocation rule #1. The reason, in that particular case, is that the ratio of expected values, $\mathbb{E}[X_i]/\mathbb{E}[S]$, disregards the interdependencies among the risks due to the various BUs, and hence may yield inadequate risk capital requirements.

- In the case of the allocation rule #2, the orders, as stipulated by the two approaches,

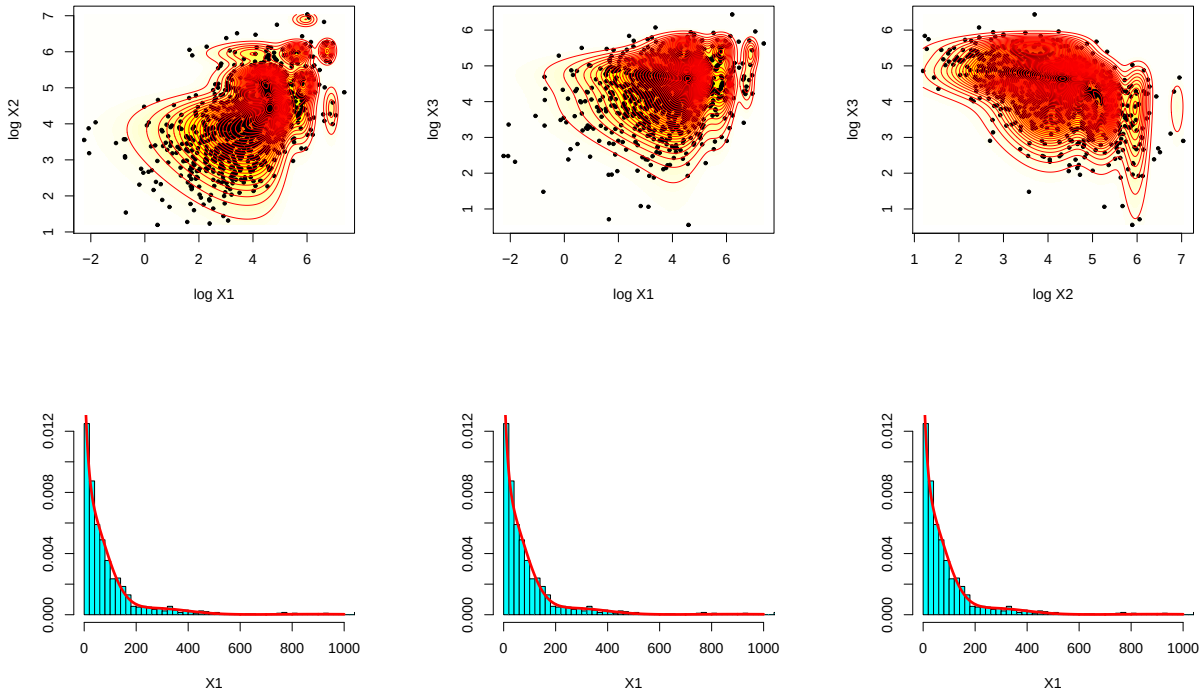

Figure 3.1: Bivariate log transformed density contours and marginal histograms for the fitted multivariate mixed-gamma distribution.

#	Risk capital allocation	Business unit 1	Business unit 2	Business unit 3
1	$\mathbb{E}[X_i]/\sum_{i=1}^{3}\mathbb{E}[X_i]$	0.335	0.335	0.330
	$\mathbb{E}[R_i]$	0.262 (-21.8%)	0.335 (0%)	0.403 (22.1%)
2	$\mathrm{CTE}_{0.95}[X_i, S]/\sum_{i=1}^{3}\mathrm{CTE}_{0.95}[X_i, S]$	0.559	0.317	0.124
	$\mathrm{CTE}_{0.95}[R_i, S]$	0.546 (-2.3%)	0.319 (0.6%)	0.135 (8.9%)

Table 3.2: Comparisons of the 'composition of allocations' method and the 'allocation of a composition' method for compositional maps (3.14) and (3.22) with the help of the fitted mixed-scaled Dirichlet distribution; $R_i = X_i/S$.

are aligned. The cause is arguably that the CTE-based risk capital allocation rule accounts for the joint dependence of the risks due to the BUs of interest, as well as for the dependence of each risk on the aggregate risk.

- While the traditional approaches to the risk capital allocation exercise, e.g., the CTE-based allocation rule or, more generally, the class of weighted allocation rules gauge the effect of the interdependencies among risks as well as of the dependence on the aggregate risk, the method proposed in this paper considers one more consequence of these dependencies, that is the behaviour of the random proportion $R_i = X_i/S$. As a result, e.g., the positive dependence on the aggregate risk may provide a cushion against allocating excessive risk capital. This observation is reflected in Table 3.2. Namely, note that the risk due to BU 1 has the highest positive conditional correlation $\text{Corr}(X_1, S|\ S > s_{0.95}) = 0.75$ among the three risks ($\text{Corr}(X_2, S|\ S > s_{0.95}) = 0.38$ and $\text{Corr}(X_3, S|\ S > s_{0.95}) = 0.01$), and this yields

$$\frac{CTE_{0.95}(X_1, S)}{CTE_{0.95}(S)} > CTE_{0.95}(R_1, S).$$

- Stochastic dependence is not the only driver that dictates the orders of the outcomes of the risk capital allocation exercise within each one of the risk capital allocation rules, #1 and #2. These orders are also determined by the shapes of the distributions of the risks due to the three BUs. Namely, in the context of the risk capital allocation rule #1, the risk distributed gamma draws the largest proportion of the aggregate risk capital, as this distribution has its mass concentrated around the mean rather than in the tails. On the other hand, in the case of the allocation rule #2, the order flips, and the risk distributed Pareto drags the largest portion of the aggregate risk capital, since Pareto is the most heavy-tailed of the three distributions employed in the example.

Chapter 4

Computing the Gini index: A note

4.1 Introduction

The Gini index is a popular measure of wealth inequality (also of, e.g., income, education and opportunity inequalities). Speaking briefly, the Gini index, G in what follows, is a relative measure of variability that equals twice the distance between the curve of the actual distribution of wealth (aka the Lorenz curve) and the curve of the total equality. The purpose of this note is to present an alternative expression and interpretation of the Gini index that hinge on the notion of the size-biased distribution [95]. Our observation reveals convenient ways to compute Gini for a great variety of statistical distributions, and, in particular, simplifies the expressions originally reported in [80] and then reproduced untouched in a number of sources including [72], [37] and [6]. As such, Sections 2 and 3, 4 of this note are kindred in spirit to [76] and [80], respectively, since in the former paper an interpretation of Gini in terms of the covariance is reported, and in the latter paper the generalized hypergeometric function is employed to derive Gini for a number of statistical distributions.

4.2 The Gini index through the lens of the size-biased sampling

Let the realizations of the positive random variable (RV) $X > 0$ represent wealth in a population and assume that the mean of the RV X is finite and denoted by m. As the cumulative distribution function (CDF), $F(x) := \mathbb{P}(X \leq x),\ x \in [0, \infty)$, is not observable, one is forced to work with a random sample RV in lieu of the RV X. Ideally, the just-mentioned random sample RV, call it Y for the sake of the utopian set-up, has the same distribution as the RV X. In such situations, and assuming that the RVs X and Y are independent and have continuous CDFs, we readily have $\mathbb{P}(Y \leq X) = 0.5$. (From now and on, the RV Y stands for the independent copy of the RV X.)

Unfortunately, the distribution of the random sample RV that is really observed in applications is not the same as the distribution of the desired RV X. This is because of, e.g., the sample size-bias associated with the sampling procedure, which assigns higher probabilities to those wealth values that are possessed by larger groups of individuals in the population [95]. Let us denote by Y^*, the size-biased random sample RV that, as the name suggests, accounts for the sampling size-bias; let F^* denote the CDF of the RV Y^*. Then we have

$$F^*(x) = \frac{1}{m}\mathbb{E}[X\mathbf{1}\{X \leq x\}] \quad \text{for all} \quad x \geq 0, \tag{4.1}$$

where $\mathbf{1}$ denotes the indicator function.

Clearly, the probability $\mathbb{P}(Y^* \leq X)$ is in general not equal to 0.5. Moreover, as the RV Y^* has first-order stochastic dominance over the RV X, and therefore over the RV Y, we have $\mathbb{P}(Y^* \leq X)/\mathbb{P}(Y \leq X) =: R_X \in (0,\ 1)$. We next assert that the Gini index is closely related to R_X. To start off, we remind in passing that the classical expression for the Gini index is

$$G_X = \frac{1}{2m}\mathbb{E}[|X - Y|].$$

Proposition 13. *Let X be a positive RV with continuous CDF, F, and finite mean, m, and*

let Y denote an independent copy of X. Then

$$G_X = 1 - R_X, \tag{4.2}$$

where $R_X = \mathbb{P}(Y^* \leq X)/\mathbb{P}(Y \leq X)$.

Proof. It suffices to note that, for $\mathbf{1}$ denoting the indicator function, we have

$$\mathrm{Cov}(X,\ F(X)) = \mathbb{E}[XF(X)] - 0.5m = \mathbb{E}[X\mathbb{E}[\mathbf{1}\{Y \leq X\}\big|X]] - 0.5m,$$

and then the desired statement follows by evoking the covariance representation of the Gini index [76] along with Equation (4.1). This concludes the proof of the proposition. □

Equation (4.2) provides a probabilistic interpretation of the Gini index. Specifically, it means that the Gini index quantifies the size-bias concealed in the distribution of the random sample RV Y^*, as opposed to the distribution of the actual wealth RV X. The closer the value of the Gini index is to zero, the more accurate the sampling procedure (with respect to size-bias) is.

In the remaining sections of this note, we make use of the just-stated connection between the Gini index and the notion of size-biasing in order to compute G for a great variety of wealth distributions. One of these distributions is a generalization of the so-called generalized hypergeometric distribution, on which Professors Mathai and Saxena wrote: "... general family of statistical probability distributions from which almost all the classical probability distributions are obtained as special cases" [82].

4.3 Applications to classes of distributions with the unbounded support

In view of Equation (4.2), in order to compute the Gini index explicitly, we need to determine the distribution of the size-biased RV Y^* as well as the distribution of either one of the RVs

$(Y^* - X)$ or Y^*/X, where the RVs $X,\ Y,\ Y^*$ are mutually independent. In the rest of the note, we show that these distributions can be readily obtained in many interesting cases.

We start with a very simple example in which the distribution of wealth is log-normal.

Example 6. Let the RV X have the log-normal distribution with the parameters $\mu \in (-\infty, \infty)$ and $\sigma > 0$; succinctly, $X \sim LN(\mu, \sigma^2)$. The probability density function (PDF) of the RV X is

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x) - \mu}{\sigma}\right)^2\right) \text{ for } x > 0.$$

Then we have $Y^* \sim LN(\mu + \sigma^2, \sigma^2)$. Therefore

$$0.5\, R_X = \mathbb{P}\left(\log(Y^*) \leq \log(X)\right) = 1 - \Phi\left(\sigma/\sqrt{2}\right),$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal RV. Hence the expression for the Gini index is immediate. This concludes Example 6.

In the same elementary fashion, Proposition 13 yields an expression for the Gini index within the fairly large class of the generalized gamma distributions. Since the Gini index is scale-invariant, that is $G_{cX} = G_X$ for $c > 0$, in the following examples we work with the unit scale parameters only.

Example 7. Let the RV X have the generalized gamma distribution, that is its PDF is

$$f(x) = \frac{a}{\Gamma(p)} \exp(-x^a)\, x^{ap-1} \text{ for } x > 0,$$

where $a, p > 0$ are parameters, and $\Gamma(\cdot)$ denotes the complete gamma function; succinctly, we write $X \sim GG(a, p)$. All of the Weibull, gamma, half-normal distributions, among other distributions, are particular cases in the class of the generalized gamma distributions.

It is well-known that the RVs distributed generalized gamma are closed under power transforms and size-biasing. More specifically, we readily obtain $X^a \sim Ga(p, 1)$, that is the RV X^a is distributed gamma with the shape parameter $p > 0$ and unit scale, and also $Y^* \sim GG(a, p + 1/a)$. Therefore, we immediately have $\mathbb{P}(Y^*/X \leq 1) = \mathbb{P}(Z < 1)$, where the

RV Z is distributed beta prime and hence

$$0.5\, R_X = I\left(\frac{1}{2}; p + \frac{1}{a}, p\right),$$

where $I(\cdot)$ denotes the regularized incomplete beta function. The expression for the Gini index follows at a stroke, and it is more concise than the one presented in [80] [also, 72]. This concludes Example 7.

In the next example, we discuss another class of distributions that has positive support and contains such very popular in the economic theory distributions as the Singh-Maddala and Dagum distributions as well as the Lomax and log-logistic distributions as special cases. Proposition 13 readily yields an expression for the Gini index.

Recall at the outset that the $(q+1) \times q$ hypergeometric function is defined as

$${}_{q+1}F_q\left[\begin{matrix} a_1, \ldots, a_{q+1} \\ b_1, \ldots, b_q \end{matrix}; z\right] = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_{q+1})_k}{(b_1)_k \cdots (b_q)_k} \times \frac{z^k}{k!}, \tag{4.3}$$

where $(p)_k = \Gamma(p+k)/\Gamma(p)$ for $k \in \{1,\ 2,\ \ldots\}$, and $(p)_0 = 1$ denotes the Pochhammer symbol. Regarding the domain of convergence for the hypergeometric function, we refer the reader to [58] for details. Also, recall that for two independent RVs $Z_1(p_1, q_1)$ and $Z_2(p_2, q_2)$ distributed beta, and so all of $p_1,\ p_2,\ q_1,\ q_2$ are positive shape parameters, we have [96]

$$\mathbb{P}\left(\frac{Z_1(p_1,q_1)}{Z_2(p_2,q_2)} \leq 1\right) = \frac{1}{p_1 \Gamma(q_1)} \left[\prod_{i=1,2} \frac{\Gamma(p_i + q_i)}{\Gamma(p_i)}\right] \frac{\Gamma(p_+)}{\Gamma(p_+ + q_2)}\ {}_3F_2\left[\begin{matrix} p_1, 1-q_1, p_+ \\ p_1+1, p_+ + q_2 \end{matrix}; 1\right], \tag{4.4}$$

where $p_+ = p_1 + p_2$.

Example 8. Let the RV X have the generalized beta of the 2nd kind distribution with the

PDF

$$f(x) = \frac{a}{B(p,q)} x^{ap-1} \left(1 + x^a\right)^{-p-q} \text{ for } x > 0,$$

where $a, p, q > 0$ are shape parameters, and $B(\cdot)$ denotes the complete beta function; succinctly we write $X \sim GB2(a, p, q)$. As in Example 7, we set the scale parameter to one.

It is well-known that, for $X \sim GB2(a, p, q)$, we have $X^a \sim B2(p, q)$, that is powers of the RVs distributed generalized beta of the 2nd kind are distributed the simple beta of the 2nd kind (or equivalently, beta prime). Also, it is trivial to confirm that $Y^* \sim GB2(a, p + 1/a, q - 1/a)$. As $Z_i(p, q),\ i = 1, 2$ denote independent RVs distributed beta and having the shape parameters $p, q > 0$, we immediately obtain

$$0.5\, R_X = \mathbb{P}(Y^*/X \leq 1) = \mathbb{P}\left(\frac{Z_1(p + 1/a, q - 1/a)}{Z_2(p, q)} \leq 1\right).$$

By evoking Equation (4.4), we end up with

$$0.5\, R_X = 1 - \frac{1}{p} \frac{1}{B(p,q)} \frac{\Gamma(p+q)\Gamma(2p+1/a)}{\Gamma(p+1/a)\Gamma(2p+q)} {}_3F_2 \left[\begin{matrix} p, 1-q, 2p+1/a \\ p+1, 2p+q \end{matrix} ; 1 \right],$$

which is less computationally demanding than the expressions that have been reported in the literature thus far [e.g., 80]. This concludes Example 8.

Examples 6, 7, and 8 unify and simplify the majority of the existing expressions for the Gini index. In the following example, we demonstrate how Proposition 13 can be employed within an even more encompassing class of distributions. To this end, note that if the size-biased RV Y^* having CDF F^* (e.g., Equation (4.1)) is considered a size-biased RV of order one, then a more general size-biased RV of order $a > 0$, call it $Y^{(a)}$ say, can be defined via its corresponding CDF, $F^{(a)}$ say, as following

$$F^{(a)}(x) = \frac{1}{\kappa(a)} \mathbb{E}[X^a \mathbf{1}\{X \leq x\}] \text{ for } x \geq 0, \tag{4.5}$$

where $\kappa(a) = \mathbb{E}[X^a] < +\infty$. We mention in passing that CDF (4.5) introduces the so-called

log-exponential family of distributions, of which all of the log-normal, generalized gamma and generalized beta of the 2nd kind distributions discussed thus far are special cases [e.g. 55, for details].

Further, let $H(\cdot)$ denote the Fox-H function [see, 83]. More specifically, for a suitable contour of integration C, $i = \sqrt{-1}$, $(\boldsymbol{a}_p, \boldsymbol{b}_p) = \{(a_j, b_j)\}_{j=1,\ldots,p}$ and $(\boldsymbol{c}_q, \boldsymbol{d}_q) = \{(c_j, d_j)\}_{j=1,\ldots,q}$, the Fox-H function admits the following Mellin-Barnes type integral representation,

$$H_{p,q}^{m,n}\left[\begin{array}{c} (\boldsymbol{a}_p, \boldsymbol{b}_p) \\ \\ (\boldsymbol{c}_q, \boldsymbol{d}_q) \end{array}; x\right] = \frac{1}{2\pi i}\int_C \frac{\prod_{i=1}^{m}\Gamma(c_i + d_i s)\prod_{i=1}^{n}\Gamma(1 - a_i - b_i s)}{\prod_{i=m+1}^{q}\Gamma(1 - c_i - d_i s)\prod_{i=n+1}^{p}\Gamma(a_i + b_i s)}\, x^{-s}\mathrm{d}s, \tag{4.6}$$

where $0 \leq m \leq q$, $0 \leq n \leq p$, $b_i > 0,\ i = 1, \ldots, p$, $d_i > 0,\ i = 1, \ldots, q$ and $a_i,\ i = 1, \ldots, p$ and $c_i,\ i = 1, \ldots, q$ are real or complex numbers such that the complete gamma functions $\Gamma(c_i + d_i s),\ i = 1, \ldots, m$ and $\Gamma(1 - a_i - b_i s),\ i = 1, \ldots, n$ do not have common poles.

It is not difficult to verify that, for any real or complex σ, the following identity holds

$$x^{\sigma} H_{p,q}^{m,n}\left[\begin{array}{c} (\boldsymbol{a}_p, \boldsymbol{b}_p) \\ \\ (\boldsymbol{c}_q, \boldsymbol{d}_q) \end{array}; x\right] = H_{p,q}^{m,n}\left[\begin{array}{c} (\boldsymbol{a}_p + \sigma\boldsymbol{b}_p, \boldsymbol{b}_p) \\ \\ (\boldsymbol{c}_q + \sigma\boldsymbol{d}_q, \boldsymbol{d}_q) \end{array}; x\right],$$

where $(\boldsymbol{a}_p + \sigma\boldsymbol{b}_p, \boldsymbol{b}_p) = \{(a_j + \sigma b_j, b_j)\}_{j=1,\ldots,p}$ and $(\boldsymbol{c}_q + \sigma\boldsymbol{d}_q, \boldsymbol{d}_q) = \{(c_j + \sigma d_j, d_j)\}_{j=1,\ldots,q}$. Therefore, the Fox-H function can be used very naturally to establish distributions that belong to the log-exponential class of distributions defined via (4.5) by setting the PDF of the RV X therein to be proportional to the Fox-H function.

Recall that the Mellin transform is defined as $\{\mathcal{M} f\}(s) = \int_0^\infty x^{s-1} f(x)dx$ for functions f and the values s, complex or real, such that the integral converges [e.g., 83, for more details]. Then the integral representation of the Fox-H function (4.6) can be interpreted as the inverse Mellin transform. Hence, we readily have that the normalizing constant in Equation (4.5) is

given by

$$\kappa\left(\boldsymbol{a}_p, \boldsymbol{b}_p, \boldsymbol{c}_q, \boldsymbol{d}_q\right) = \int_0^\infty H_{p,q}^{m,n}\left[\begin{array}{c} (\boldsymbol{a}_p, \boldsymbol{b}_p) \\ \\ (\boldsymbol{c}_q, \boldsymbol{d}_q) \end{array}; x\right] \mathrm{d}x = \frac{\prod_{i=1}^{m}\Gamma(c_i+d_i)\prod_{i=1}^{n}\Gamma(1-a_i-b_i)}{\prod_{i=m+1}^{q}\Gamma(1-c_i-d_i)\prod_{i=n+1}^{p}\Gamma(a_i+b_i)}. \tag{4.7}$$

Example 9. Let the RV X have the Fox-H distribution, that is the PDF of the RV X is given by

$$f(x) = \frac{1}{\kappa\left(\boldsymbol{a}_p, \boldsymbol{b}_p, \boldsymbol{c}_q, \boldsymbol{d}_q\right)} \times H_{p,q}^{m,n}\left[\begin{array}{c} (\boldsymbol{a}_p, \boldsymbol{b}_p) \\ \\ (\boldsymbol{c}_q, \boldsymbol{d}_q) \end{array}; x\right] \text{ for } x > 0, \tag{4.8}$$

where $0 \le m \le q$, $0 \le n \le p$, $b_i > 0,\ i = 1, \ldots, p$, $d_i > 0,\ i = 1, \ldots, q$ and $a_i,\ i = 1, \ldots, p$ and $c_i,\ i = 1, \ldots, q$ are real or complex numbers such that the Fox-H function is well-defined, and integral (4.7) converges. The PDF above includes as special cases the majority of the distributions with the support $(0,\ \infty)$ known nowadays. In fact, an example of a common PDF that is not a particular case of (4.8) has eluded us thus far.

In spite of the utmost generality of the Fox-H distributions, the route to compute the Gini index for PDF (4.8) is not different from the one we pursued in Examples 6, 7, and 8. Indeed, in addition to the already-mentioned closure under size-biasing, note that the quotients of the independent RVs distributed Fox-H are distributed Fox-H [83]. Hence, we have

$$0.5\, R_X = \frac{1}{\kappa\left(\boldsymbol{a}_p + \boldsymbol{b}_p, \boldsymbol{b}_p, \boldsymbol{c}_q + \boldsymbol{d}_q, \boldsymbol{d}_q\right)} \times \frac{1}{\kappa\left(\boldsymbol{a}_p, \boldsymbol{b}_p, \boldsymbol{c}_q, \boldsymbol{d}_q\right)} \times H_{p+q+1,p+q+1}^{m+n,m+n+1}\left[\begin{array}{c} (0,1),\ (\boldsymbol{a}_{p+q}^*, \boldsymbol{b}_{p+q}^*) \\ \\ (\boldsymbol{c}_{p+q}^*, \boldsymbol{d}_{p+q}^*),\ (-1,1) \end{array}; 1\right],$$

where the parameters $(\boldsymbol{a}^*_{p+q}, \boldsymbol{b}^*_{p+q}) = \{(a^*_j, b^*_j)\}_{j=1,\ldots,p+q}$ are such that

$$(a^*_j, b^*_j) = \begin{cases} (a_j + b_j, b_j), & j \in \{1, \ldots, n\} \\ (1 - c_{j-n} - 2d_{j-n}, d_{j-n}), & j \in \{n+1, \ldots, n+m\} \\ (a_{j-m} + b_{j-m}, b_{j-m}), & j \in \{n+m+1, \ldots, m+p\} \\ (1 - c_{j-p} - 2d_{j-p}, d_{j-p}), & j \in \{m+p+1, \ldots, p+q\} \end{cases},$$

and the parameters $(\boldsymbol{c}^*_{p+q}, \boldsymbol{d}^*_{p+q}) = \{(c^*_j, d^*_j)\}_{j=1,\ldots,p+q}$ are such that

$$(c^*_j, d^*_j) = \begin{cases} (c_j + d_j, d_j), & j \in \{1, \ldots, m\} \\ (1 - a_{j-m} - 2b_{j-m}, b_{j-m}), & j \in \{m+1, \ldots, n+m\} \\ (c_{j-n} + d_{j-n}, d_{j-n}), & j \in \{n+m+1, \ldots, n+q\} \\ (1 - a_{j-q} - 2b_{j-q}, b_{j-q}), & j \in \{n+q+1, \ldots, p+q\} \end{cases}.$$

4.4 Applications to classes of distributions with bounded supports

In the previous section, we elucidated the usefulness of Proposition 13 when computing the Gini index for a multitude of wealth distributions with the support $(0, \infty)$. Below we consider a few examples in which the wealth distributions have bounded supports, and Proposition 13 allows for an elementary derivation of G.

The unit-gamma distribution which we are going to discuss next, can be viewed as the exponential transform of the classical gamma distribution as well as a limiting distribution

of the generalized beta distribution of the first kind [80].

Example 10. The RV X is said to be distributed unit-gamma, if its PDF is given by

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\beta-1}(-\ln x)^{\alpha-1} \text{ for } x \in (0,1),$$

where $\alpha, \beta > 0$ are parameters; succinctly, we write $X \sim UG(\alpha, \beta)$.

The unit-gamma distributions are closed under size-biasing, and in particular we have $Y^* \sim UG(\alpha, \beta+1)$, where the RV Y is an independent copy of the RV X as before. Further, denote by Z_1 and Z_2 two independent copies of the RV $Z \sim Ga(\alpha, 1)$. Then the quotient Z_1/Z_2 is distributed beta prime, and we readily obtain

$$0.5\, R_X = \mathbb{P}\left(-\ln(Y^*) \geq -\ln(X)\right) = \mathbb{P}\left(\frac{Z_1}{Z_2} \leq \frac{\beta}{1+\beta}\right) = I\left(\frac{\beta}{1+2\beta}; \alpha, \alpha\right),$$

where $I(\cdot)$ denotes the regularized incomplete beta function. This concludes Example 10.

A more general class of distributions with bounded supports is the class of the generalized beta distributions of the first kind [80].

Example 11. Let the RV X have the generalized beta of the first kind distribution, that is the PDF of the RV X is given by

$$f(x) = \frac{a}{B(p,q)} x^{ap-1}\left(1-x^a\right)^{q-1} \text{ for } x \in (0,1),$$

where $a, p, q >$ are shape parameters; succinctly, we write $X \sim GB1(a, p, q)$.

To see that Proposition 13 is again immediately applicable, observe that if $X \sim GB1(a, p, q)$, then $X^a \sim B1(p, q)$ that is, the latter RV is distributed beta of the first kind. Also, we have $Y^* \sim GB1(a, p+1/a, q)$. Finally, let $Z_i(p, q),\ i = 1, 2$ denote independent RVs distributed beta with the shape parameters $p, q > 0$, then we have

$$0.5\, R_X = \mathbb{P}\left(\frac{Z_1(p+1/a, q)}{Z_2(p,q)} \leq 1\right) = 1 - \frac{1}{p}\frac{1}{B(p,q)}\frac{\Gamma(p+q+1/a)\Gamma(2p+1/a)}{\Gamma(p+1/a)\Gamma(2p+q+1/a)}\, {}_3F_2\left[\begin{matrix} p, 1-q, 2p+1/a \\ p+1, 2p+q+1/a \end{matrix}; 1\right]$$

in which the left-most expression follows by evoking Equation (4.4). This concludes Example 11.

Chapter 5

Conclusions

In Chapter 2, we studied the class of general multiplicative background risk models and then introduced and investigated its special subclass, the multiplicative background risk models with idiosyncratic risk factors distributed phase-type and the systemic risk factor distributed arbitrarily on the non-negative half of the real line. The new constructions are a generalization of the noble exponential multiplicative background risk models (also known as exponential mixtures), and as such allow for much more intricate dependencies in the sense that, e.g., the Pearson coefficients of correlation of the involved risk components must not be identical for any pair of risk components. Also, when constructed to this purpose, the new constructions can be considered exponential mixtures loaded for model risk. In spite of quite remarkable generality, i.e., phase-type distributions can approximate any non-negative distribution fairly well, the new constructions are surprisingly tractable technically. To reflect on this fact, we derived explicit expressions for some practical tail-based risk measures and the risk capital allocation rules based on them. Our message to practitioners is briefly this: when modelling portfolios of dependent risks, a slight departure from exponential mixtures may trigger significant changes in the conclusions as to the magnitudes and orders of the required risk capitals. Hence, within exponential mixtures, mistakes are easy to make, and the consequences may be very hard to predict.

In Chapter 3, we argued that all risk capital allocation rules nowadays aim at determining the percentages of the aggregate risk capital that have to be set aside for the business units of a financial entity. These percentages are risk capital allocations due to the business units,

normalized in order to ensure the full-additivity of the end-result. We then revealed a way to replace the aforementioned deterministic percentages with the random proportions that sum up to one almost surely, thus getting hands directly on the stochastic phenomenon that underpins the allocation procedure. In order to study the random proportions, we have introduced, in the reverse order, a new class of multivariate mixed-scaled Dirichlet distributions that govern the stochastic characteristics of the random proportions, also known as compositions, as well as a class of multivariate mixed-gamma distributions that serve as a basis for these compositions. We have studied some relevant (closure) properties of the two just-mentioned classes of probability models and demonstrated that they provide versatile yet surprisingly tractable tools for risk analysis, and in particular, for the purpose of the risk capital allocation exercise. An important by-product of our approach to allocating the aggregate risk capital is that it allows to unify the bottom-up and the top-down threads in the allocations' state-of-the-art into one encompassing method.

In Chapter 4, we revealed a connection between the Gini Mean Difference measure of variability and the concept of statistical size-biasing to compute this measure of variability explicitly and often effortlessly for a great variety of probability distributions, starting from those that are as simple as the log-normal distribution and ending with such encompassing ones as the Fox-H distribution that seemingly covers all known probability distributions on the non-negative half of the real line. We hope this will popularize Gini Mean Difference in actuarial science, where it has been understudied.

Bibliography

[1] Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:1487-1503.

[2] Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44:139–177.

[3] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, New York.

[4] Albrecher, H., Constantinescu, C., and Loisel, S. (2011). Explicit ruin formulas for models with dependence among risks. *Insurance: Mathematics and Economics*, 48(2):265–270.

[5] Amin, H. Z. and Stanford, D. A. (2016). Bayesian and Bühlmann credibility for phase-type distributions with a univariate risk parameter. *Scandinavian Actuarial Journal*, 2016(4):338–355.

[6] Arnold, B.C. and Sarabia, M.J. (2018). *Majorization and the Lorenz Order with Applications in Applied Mathematics and Economics*. Springer, New York.

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

[8] Arnold, B. C. (1983). *Pareto Distributions*. International Co-operative Publishing House, Fairland.

[9] Asimit, V., Furman, E. and Vernic, R. (2010). On a multivariate Pareto distribution. *Insurance: Mathematics and Economics*, 46(2), 308–316.

[10] Asimit, A. V., Vernic, R., and Zitikis, R. (2016). Background risk models and stepwise portfolio construction. *Methodology and Computing in Applied Probability*, 18(3):805–827.

[11] Asimit, A. V., Furman, E., and Vernic, R. (2016). Statistical inference for a new class of multivariate Pareto distributions. *Communications in Statistics - Simulation and Computation*, 45:458 – 471.

[12] Asmussen, S. and Rolski, T. (1992). Computational methods in risk theory: A matrix-algorithmic approach. *Insurance Mathematics and Economics*, 10(4):259–274.

[13] Asmussen, S., Nerman, O., and Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm.

[14] Asmussen, S. (2003). *Applied Probability and Queues*. Springer, New York.

[15] Bahnemann, D. (2015). *Distributions for actuaries*. Casualty Actuarial Society.

[16] Balog, D., Bátyi, T. L., Csóka, P., and Pintér, M. (2017). Properties and comparison of risk capital allocation methods. *European Journal of Operational Research*, 259(2):614–625.

[17] Belles-Sampera, J., Guillen, M., and Santolino, M. (2016). Compositional methods applied to capital allocation problems. *The Journal of Risk*, 19(2):15–30.

[20] Bernard, C., Rüschendorf, L., Vanduffel, S., and Wang, R. (2017). Risk bounds for factor models. *Finance and Stochastics*, 21(3):631–659.

[19] Boonen, T. J., Guillen, M., and Santolino, M. (2019). Forecasting compositional risk allocations. *Insurance: Mathematics and Economics*, 84:79–86.

[20] Bernard, C., Rüschendorf, L., Vanduffel, S., and Wang, R. (2017). Risk bounds for factor models. *Finance and Stochastics*, 21(3):631–659.

[21] Bladt, M. (2005). A review on phase-type distributions and their use in risk theory. *ASTIN Bulletin*, 35(1):145–161.

[22] Bondesson, L.(1992). Generalized gamma convolutions and related classes of distributions and densities. Lecture Notes in Statistics, 76.. Springer-Verlag, New York, 1992.

[23] Cai, J. and Li, H. (2005). Conditional tail expectations for multivariate Phase-type distributions. *Journal of Applied Probability*, 42(3):809–825.

[24] Carter, B.D. and Springer, M.D. (1978). The distribution of products, quotients and powers of independent H-function variates. SIAM Journal of Applied Mathematics 33(4) 542 – 558.

[25] Corrigan, J., de Decker, J., Hoshino, T., van Delft, L., and Verheugen, H. (2009). Aggregation of risks and allocation of capital. Technical report, Milliman, Seattle.

[26] 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.

[27] Cossette, H., Marceau, E., Mtalai, I., and Veilleux, D. (2018). Dependent risk models with Archimedean copulas: A computational strategy based on common mixtures and applications. *Insurance: Mathematics and Economics*, 78:53–71.

[28] Chiragiev, A., and Landsman, Z. (2007). Multivariate Pareto portfolios: TCE-based capital allocation and divided differences. *Insurance: Mathematics and Economics*, 4:261–280.

[29] Das, S. R., Duffie, D., Kapadia, N., and Saita, L. (2007). Common failings: How corporate defaults are correlated. *Journal of Finance*, 62(1):93–117.

[30] David, A. and Larry, S. (1987). The least variable phase type distribution is Erlang. *Communications in Statistics. Stochastic Models*, 3(3):467–473.

[31] De Boor, C. (1978). *A practical guide to splines*. New York: Springer-Verlag.

[32] Dempster, A. P., Laird, M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1 – 38.

[33] Denault, M. (2001). Coherent allocation of risk capital. *The Journal of Risk*, 4(1):1–34.

[34] Denuit, M., Dhaene, J., Goovaerts, M.J. and Kaas, R. (1997). *Actuarial Theory for Dependent Risk: Measures, Orders and Models.*. Wiley, New York

[35] Dhaene, J., Henrard, L., Landsman, Z., Vandendorpe, A., and Vanduffel, S. (2008). Some results on the CTE-based capital allocation rule. *Insurance: Mathematics and Economics*, 42(2):855–863.

[36] Dhaene, J., Tsanakas, A., Valdez, E. A., and Vanduffel, S. (2012). Optimal capital allocation principles. *Journal of Risk and Insurance*, 79(1):1–28.

[37] Duangkamon, C. (2008). Modeling income distributions and Lorenz curves. Springer, New York.

[38] EIOPA (2010). QIS5 technical specifications. Technical report, European Insurance and Occupational Pensions Authority.

[39] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Heidelberg.

[40] Erlang, A.K. (1909). The Theory o Probabilities and Telephone Conversations. *Nyt Tidsskrift for Matematik*, 20(B):33–39.

[41] Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics*, 38:1466–1474.

[42] Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.

[43] Fang, K.-T., Kotz, S., and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. CRC Press, Boca Raton.

[44] Franke, G., Schlesinger, H., and Stapleton, R.C. (2006). Multiplicative background risk. *Management Science*, 52:146–153.

[45] Franke, G., Schlesinger, H., and Stapleton, R.C. (2011). Risk taking with additive and multiplicative background risks.. *Econ. Theory*, 146:1547–1568.

[46] FSB (2009). Guidance to assess the systemic importance of financial institutions, markets and instruments: Initial considerations. Technical report, Report to the G-20 Finance Ministers and Central Bank Governors.

[47] Furman, E. and Landsman, Z. (2005). Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics*, 37(3):635–649.

[48] Furman, E. and Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin*, 36(2):433–462.

[49] Furman, E. and Landsman, Z. (2010). Multivariate Tweedie distributions and some related capital-at-risk analysis. *Insurance: Mathematics and Economics*, 46(2):351–361.

[50] Furman, E. and Zitikis, R. (2008a). Weighted premium calculation principles. *Insurance: Mathematics and Economics*, 42(1):459–465.

[51] Furman, E. and Zitikis, R. (2010). General Stein-type covariance decompositions with applications to insurance and finance. *ASTIN Bulletin*, 40(1):369–375.

[52] Furman, O. and Furman, E. (2010). On some layer-based risk measures with applications to exponential dispersion models. *Journal of Probability and Statistics*, 2010:1–19.

[53] Furman, E., Wang, R. and Zitikis, R(2017). Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking and Finance*, 83:70–84.

[54] Furman, E. and Zitikis, R. (2008b). Weighted risk capital allocations. *Insurance: Mathematics and Economics*, 43(2):263–269.

[55] Furman, E. and Zitikis, R. (2009). Weighted pricing functionals with applications to insurance. *North American Actuarial Journal*, 13(4):483–496.

[56] Furman, E., Kuznetsov, A. and Zitikis, R. (2018). Weighted risk capital allocations in the presence of systematic risk. *Insurance: Mathematics and Economics*, 79:75–81.

[57] Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, 12:199–232.

[58] Gradshteyn, I. and Ryzhik, I. (2014). *Table of Integrals, Series, and Products*. Academic Press, New York.

[59] Grassia, A. (1977). On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions. Australian Journal of Statistics 19(2) 108 – 114.

[60] Grechuk, B. (2015). The center of a convex set and capital allocation. *European Journal of Operational Research*, 243:628–636.

[61] Grundke, P. (2010). Top-down approaches for integrated risk management: How accurate are they? *European Journal of Operational Research*, 203(3):662–672.

[62] Guillén, M., Sarabia J.-M., and Prieto, F. (2013). Simple risk measure calculations for sums of positive random variables. *Insurance: Mathematics and Economics*, 51(1):273–280.

[63] Gupta, A. K. and Nadarajah, S. (2004). *Handbook of Beta Distribution and Its Applications*. Marcel Dekker, New York.

[64] 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.

[65] Hürlimann, W. (2005). Analytical evaluation of economic risk capital for portfolios of gamma risks. *ASTIN Bulletin*, 31(01):107–122.

[66] Jensen, Arne. (1953). Application of stochastic processes to an investment plan. *Metroeconomica*, 5(3):97–151.

[67] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman and Hall, London.

[68] Johnson, M. A. (1988). The denseness of phase distributions. Technical report, Purdue University, West Lafyette.

[69] Kalkbrener, M. (2005). An axiomatic approach to capital allocation. *Mathematical Finance*, 15(3):425–437.

[70] Karlis, D. (2003). ML estimation for multivariate shock models via an EM algorithm. *Annals of the Institute of Statistical Mathematics*, 55:817–830.

[71] Kim, J. H. (1988). Conditional tail moments of the exponential family and its related distributions. *North American Actuarial Journal*, 14(2):198–216.

[72] Kleiber, C and Kotz, S. (2003). Statistical Size Distributions in Economics and Actuarial Sciences. Wiley, New York.

[73] Latouche, G. and Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Society for Industrial and Applied Mathematics, Philadelphia.

[74] Lee, S. C. K. and Lin, X. S. (2010). Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal*, 14(1):107–130.

[75] Lee, S. C. and Lin, X. S. (2012). Modeling dependent risks with multivariate Erlang mixtures. *ASTIN Bulletin*, 42(1):153–180.

[76] Lerman, R.I. and Yitzhaki, S. (1984). A note on the calculation and interpretation of the Gini index. Economics Letters 15, 363 – 368.

[77] Libby, D. L. and Novick, M. R. (1982). Multivariate generalized beta distributions with applications to utility assessment. *Journal of Educational and Behavioral Statistics*, 7(4):271–294.

[78] Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–37.

[79] Liu, H. and Wang, R. (2017). Collective risk models with dependence uncertainty. *ASTIN Bulletin*, 47(02):1–29.

[80] McDonald, J.B. (1984). Some generalized functions for the size distribution of income. *Econometrica* 52(3), 647 – 665.

[81] Markowitz, H. M. (1991). Foundations of portfolio theory. *The Journal of Finance*, 46(2):469.

[82] Mathai, A.M. and Saxena, R.K. (1967). On a generalized hypergeometric distribution. *Metrika* 11(1), 127–132.

[83] Mathai, A.M., Haubold, H.J. and Saxena, R.K. (2010). *The H-Function: Theory and Applications*. Springer, New York.

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

[85] McNeil, A.J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. (Revised Edition.) Princeton University Press, Princeton, NJ.

[86] Miles, J., Furman, E., and Kuznetsov, A. (2019). Risk aggregation: A general approach via the class of generalized gamma convolutions. *Variance, in press*.

[87] Monti, G. S., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (2011a). Notes on the scaled Dirichlet distribution. In *Compositional Data Analysis: Theory and Applications*, pages 128–138. Wiley, Chichester.

[88] Monti, G. S., Mateu-Figueras, G., Pawlowsky-Glahn, V., and Egozcue, J. J. (2011b). The shifted-scaled Dirichlet distribution in the simplex. In *the 4th International Workshop on Compositional Data Analysis*.

[89] Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, 37(1):541–544.

[90] Neuts, M.F. (1995). *Algorithmic Probability: A Collection of Problems*. CRC Press, Boca Raton, Florida.

[91] Neuts, M.F. (1981). *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Johns Hopkins University Press, Baltmore, MD.

[92] Ng, K. W., Tian, G. L., and Tang, M. L. (2011). *Dirichlet and Related Distributions: Theory, Methods and Applications*. Wiley, Chichester.

[93] 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.

[94] Patil, G. P., Rao, C. R., and Zelen, M. (1986). A computerized bibliography of weighted distributions and related weighted methods for statistical analysis and interpretations of encountered data, observational studies, representativeness issues, and reading inference. Technical report, Pennsylvania State University, University Park, PA.

[95] Patil, G.P. and Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics 34(2),179 – 189.

[96] Pham-Gia, T. and Turkkan, N. (2002). The product and quotient of general beta distributions. *Statistical Papers* 43(2002), 537 – 550.

[97] Prudnikov, A., Brychkov, Y. A., and Marichev, O. I. (1992). *Integrals and Series (vol.5: Inverse Laplace Transforms)*. Gordon and Breach, New York.

[98] Ren, J. (2010). Recursive formulas for compound phase distributions - univariate and bivariate cases. *ASTIN Bulletin*, 40(2):615–629.

[99] Ren, J. and Zitikis, R. (2017). CMPH: A multivariate phase-type aggregate loss distribution. *Dependence Modeling*, 5(1):304–315.

[100] Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13:341–360.

[101] Sarabia, J. M., Gómez-Déniz, E., Prieto, F., and Jordá, V. (2016). Risk aggregation in multivariate dependent Pareto distributions. *Insurance: Mathematics and Economics*, 71:154–163.

[102] Sarabia, J. M., Gómez-Déniz, E., Prieto, F., and Jordá, V. (2018). Aggregation of dependent risks in mixtures of exponential distributions and extensions. *ASTIN Bulletin*, 48(3):1079–1107.

[103] Sarabia, J. M., Gómez-Déniz, E., Prieto, F., and Jordá, V. (2018). Modelling dependent risks with heavy-tail marginals. *ASTIN Bulletin*, in press.

[104] Semenikhine, V., Furman, E., and Su, J. (2018). On a multiplicative multivariate gamma distribution with applications in insurance. *Risks*, 6(3):1–20.

[105] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425–442.

[106] Su, J. and Furman, E. (2017a). A form of multivariate Pareto distribution with applications to financial risk measurement. *ASTIN Bulletin*, 47(1):331–357.

[107] Su, J. and Furman, E. (2017b). Multiple risk factor dependence structures: Distributional properties. *Insurance: Mathematics and Economics*, 76:56–68.

[108] Su, J. and Hua, L. (2017). A general approach to full-range tail dependence copulas. *Insurance: Mathematics and Economics*, 77:49–64.

[109] Sweeting, P. (2011). *Financial Enterprise Risk Management*. Cambridge University Press, Cambridge.

[110] Tijms, H. C. (1994). *Stochastic Models: An Algorithmic Approach*. Wiley, Chichester.

[111] Tsanakas, A. and Barnett, C. (2003). Risk capital allocation and cooperative pricing of insurance liabilities. *Insurance: Mathematics and Economics*, 33:239–254.

[112] Varadhan, R. and Gilbert, P. (2015). BB : An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software*, 32(4):1–26.

[113] Verbelen, R., Antonio, K., and Claeskens, G. (2016). Multivariate mixtures of Erlangs for density estimation under censoring. *Lifetime Data Analysis*, 22(3):429–455.

[114] 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.

[115] Vernic, R. (2006). Multivariate skew-normal distributions with applications to insurance. *Insurance: Mathematics and Economics*, 38(2):413–426.

[116] Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(01):71–92.

[117] Wang, S. (1998). Aggregation of correlated risk portfolios: Models and algorithms. *Proceedings of the Casualty Actuarial Society*, pages 848–939.

[118] Willmot, G. E. and Lin, X. S. (2011). Risk modelling with the mixed Erlang distribution. *Applied Stochastic Models in Business and Industry*, 27(1):8–22.

[119] 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.

[120] Wu, C. (1983). On the convergence properties of the em algorithm. *Annals of Statistics*, 11:95 – 103.

[121] Yitzhaki, S and Schechtman, E. (2013). *The Gini Methodology: A Primer on a Statistical Methodology*. Springer, New York.