SOCIETY OF ACTUARIES

A Reconciliation of the Top-Down and Bottom-Up Approaches to Risk Capital Allocations: Proportional Allocations Revisited

December 2019

A Reconciliation of the Top-Down and Bottom-Up Approaches to Risk Capital Allocations: Proportional Allocations Revisited

AUTHOR Edward Furman, Ph.D.
Yisub Kye, Ph.D.
York University

Jianxi Su, Ph.D.
Purdue University

SPONSOR Society of Actuaries Research
Expanding Boundaries Pool

Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information

Copyright © 2019 by the Society of Actuaries. All rights reserved.

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

Edward Furman* Yisub Kye† Jianxi Su‡

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 (BUs) 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 allocation (TDA) approach that entails that the allocation exercise is imposed by the corporate centre, and the bottom-up allocation (BUA) approach that implies that the allocation of the aggregate risk to business units is informed by these units. Briefly, the TDA starts with the aggregate risk capital that is then replenished among the BUs according to the views of the *centre*, thus limiting the inputs from the BUs. The BUA does start with the BUs, but it is, as a rule, too granular, and so may lead to missing the wood for the trees.

Irrespective of whether the TDA or the BUA is assumed, it is the proportional contribution of the riskiness of a stand-alone BU to the aggregate riskiness of the financial entity that is of central importance, and it is routinely computed nowadays as the quotient of the allocated risk capital due to the BU of interest and the aggregate risk capital due to the financial entity. For instance, in the simplest case when the mathematical expectation plays the role of the risk measure that generates the allocation rule, the desired proportional contribution is just a quotient of two means. Clearly in general, this quotient of means does not concur with the mean of the quotient random variable that captures the genuine stochastic proportional contribution of the riskiness of the BU of interest. Inspired by this observation, herein we reenvision the way in which the allocation problem is tackled in the state of the art. As a by-product, we unify the TDA and the BUA into one encompassing approach.

Keywords: Risk capital allocation, proportional allocation, weighted allocation, Dirichlet distribution, mixed-gamma distribution.

*Department of Mathematics and Statistics, York University, Toronto, ON, M3J 1P3, Canada.

†Department of Mathematics and Statistics, York University, Toronto, ON, M3J 1P3, Canada.

‡Corresponding author; Department of Statistics, Purdue University, West Lafayette, IN, 47906, United States.

1 Introduction

Let the random variable (RV) $X \geq 0$ denote an insurance risk and let the set $\mathcal{X}$ denote a collection of such risks. Also, for $n \in \mathbb{N}$, let the RV $X_i,\ i = 1, \ldots, n$, denote the risk due to the i-th business unit (BU) of a financial entity, and let $S = \sum_{i=1}^n X_i$ stand for the aggregate risk RV 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 Wang (1996); Artzner et al. (1999); Furman and Zitikis (2008a) and references therein for axiomatic treatments of risk measures, and to Guillén et al. (2013); Bernard et al. (2017); Miles et al. (2019) and references therein for some recent developments on risk aggregation.

After the aggregate risk capital, $H(S)$, has been determined, the question arises as to what is a meaningful way to allocate it to the BUs. 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. In the state-of-the-art, the allocation rule, A, assigns a (monetary) value in $[0,\ \infty) \cup \{+\infty\}$ to 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., Denault, 2001; Furman and Zitikis, 2008b; Balog et al., 2017, for theory and applications). From the above, it is clear that the allocation rule A is generated by the risk measure H. The only other assumption on the map A that we need - and indeed make in what follows - is *full-additivity*, which formally means $\sum_{i=1}^n A(X_i, S) = A(S, S)$, and so $\sum_{i=1}^n A(X_i, S) = H(S)$.

Clearly, there are numerous ways to allocate the aggregate risk due to the RV S having cumulative distribution function (CDF) $F_S(s),\ s \in [0,\ \infty)$, and inverse CDF $F_S^{-1}(p) = \inf\{s \in [0,\ \infty) :\ F_S(s) \geq p\},\ p \in [0,\ 1)$. Some of these ways are very simple, such as the hair-cut allocation, A_p,

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

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

$$A_g(X_i, S) = \mathbb{E}\left[X_i g'(F_S(S))\right],\ i = 1, \ldots, n, \tag{2}$$

where $g : [0,\ 1] \to [0,\ 1]$ is a continuously differentiable *distortion* function (Tsanakas and Barnett, 2003), and

$$A_w(X_i, S) = \frac{\mathbb{E}[X_i w(S)]}{\mathbb{E}[w(S)]},\ i = 1, \ldots, n, \tag{3}$$

where $w : [0,\ \infty) \to [0,\ \infty)$ is a non-decreasing *weight* function (Furman and Zitikis, 2008b); 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 (Grechuk, 2015). There are admittedly many ways to classify the existing risk capital allocation rules, the list

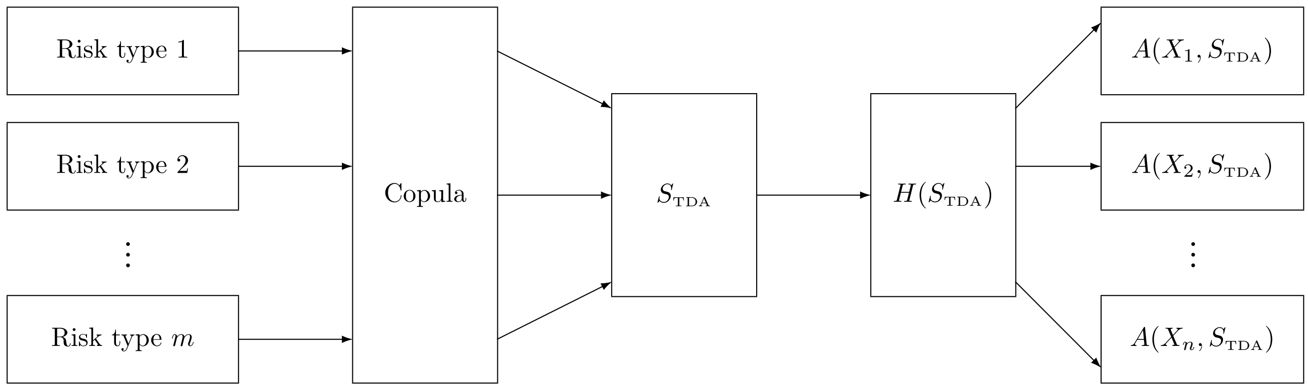

Figure 1: The top-down approach allocates the aggregate risk RV, S_{TDA}, and then the risk capital, $H(S_{\text{TDA}})$, which are to this end obtained by integrating distinct risk types via a suitable copula function and according to the views of the corporate centre.

of which is vast and grows quickly. Of particular interest to us is the simple taxonomy into the top-down allocation (TDA) rules and the bottom-up allocation (BUA) rules, which is arguably the one that stands out in applications. Specifically, the TDA allocates the aggregate risk RV, S_{TDA}, and then capital, $H(S_{\text{TDA}})$, which are to this end obtained by integrating distinct risk types (e.g., market, credit, (non-)life underwriting, health underwriting; see, EIOPA, 2010) via a suitable copula function, according to the views of the corporate centre (Figure 1). An example of the TDA is the hair-cut allocation given in Equation (1). It is fairly simple to compute and transparent to convey to the upper management. That said, the hair-cut allocation disregards the inter-dependencies among the BUs $X_1, \ldots, X_n$. As such, the hair-cut allocation fails to reflect on the fact that these BUs are constituents of a larger structure, and it treats them as stand-alone objects instead. Consequently, the TDA may miss such desired details as the inter-dependencies between BUs and, sometimes, even the BUs' stochastic characteristics. Nevertheless, with the *Own Risk and Solvency Assessment* being progressively implemented in the insurance industry, the TDA has evolved as a predominant standard (Grundke, 2010).

Unlike the TDA rules, the BUA rules start with a comprehensive multivariate CDF that describes the risks due to distinct BUs as well as the inter-dependences between these risks. Therefore, all of the: aggregate risk RV, S_{BUA}; aggregate risk capital, $H(S_{\text{BUA}})$; and allocated risk capital, $A(X_i, S_{\text{BUA}}),\ i = 1, \ldots, n$, can be computed at a stroke and with no input from the corporate centre (Figure 2). Weighted allocation (3) is an example of the BUA rule. It is considerably more granular than the already-mentioned hair-cult allocation rule in the sense that it starts with the joint multivariate CDF of the risks due to the BUs $X_1, \ldots, X_n$, and so accounts for both the inter-dependencies among these risks and the joint behavior of the pair $(X_i, S) \in \mathcal{X} \times \mathcal{X},\ i = 1, \ldots, n$, from which the allocated risk capital is obtained. The weighted allocation rule is *consistent*, satisfies *no undercut* and *consistent no undercut* properties (Furman and Zitikis, 2008b), and it is optimal in the sense of Dhaene et al. (2012). That said, unless very special distributional structures are considered (Furman et al., 2018), the weighted allocation is rather difficult

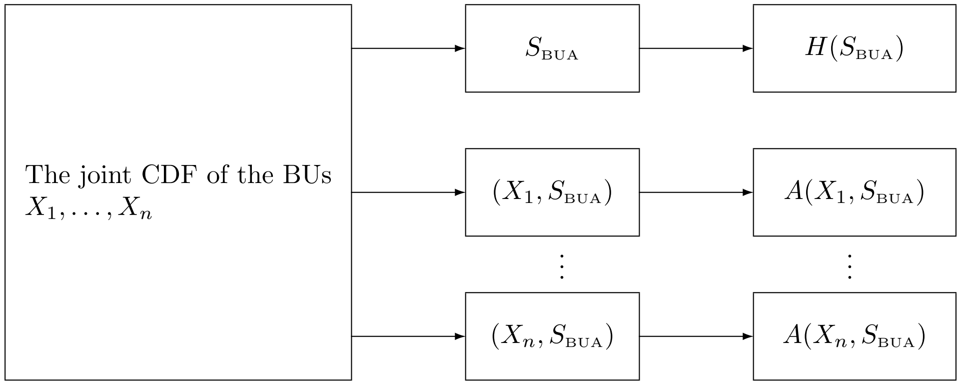

Figure 2: The bottom-up approach starts by specifying the joint CDF of the business units in a financial entity and then computes the aggregate risk capital and the allocated risk capitals at once.

to compute, even for special choices of the weight function, w, let alone in general. To illustrate the computational complexity, we refer to Dhaene et al. (2008) for elliptically distributed risks; Cai and Li (2010) for phase-type distributed risks; Furman and Landsman (2010) for Tweedie distributed risks; Vernic (2006, 2011) for skew-normal and Pareto distributed risks; Cossette et al. (2013) and Cossette et al. (2018) for the risks with the dependence structures described by the Farlie-Gumbel-Morgenstern copula and the Archimedean copula, respectively. All these works compute A_w for the special weight function $w(s) = \mathbf{1}\{s > F_S^{-1}(p)\},\ p \in [0,\ 1)$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function.

We note in passing that one may think that the RVs S_{TDA} and S_{BUA}, which are clearly not equal almost surely by construction, are at least equal in distribution. This is because the two aforementioned RVs both aim to proxy the true aggregate risk RV S. However, given the increasing complexity involved in today's insurance companies' business structure, it is rather challenging, if not practically infeasible, to match the distributions of the RVs S_{TDA} and S_{BUA} in reality, and we indeed distinguish between these two RVs in what follows.

In summary, the TDA is intuitive yet often oversimplified, and the BUA is meticulous yet may hit against too many parameters. In practice, the two approaches are often conducted separately and are sought to complement each other. The question that arises then is whether it is possible to unify the TDA and the BUA rules 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 risk factors-based aggregate risk capital, which is recommended, and sometimes even mandated by regulatory authorities. Putting forward a theoretical groundwork for such an encompassing approach to allocate risk capital is a natural call, and it is a goal that we aim to achieve in the present paper. Another goal of this paper is to revisit the very way, in which the allocation of risk capital is envisioned nowadays.

2 Motivation 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, if well-defined and finite, can be written as the proportional allocation $A(X_i, S) = H\big(Sr_i\big)$, $i = 1, \dots, n$, 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; $A_g(X_i, S) = \mathbb{E}\left[Xg'(F_S(S))\right]$ and $\sum_{i=1}^n A_g(X_i, S)$ - in the context of the distorted allocation; $A_w(X_i, S) = \mathbb{E}\left[Xw(S)\right] / \mathbb{E}\left[w(S)\right]$ and $\sum_{i=1}^n A_w(X_i, S)$ in the context of the weighted allocation. In other words, for all of (1)-(3), we have

$$r_i = \frac{A(X_i, S)}{A(S, S)} = \frac{A(X_i, S)}{H(S)},\ i = 1, \dots, n. \tag{4}$$

Ratios (4) can be reformulated with the help of the language of *compositions* (e.g., Aitchison, 1982, for details; also, 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 (see, Section 2.1 for technical details), and let $\boldsymbol{C} = (C_1, \dots, C_n) \in \mathbb{S}^n$ and $\boldsymbol{X} = (X_1, \dots, X_n) \in \mathcal{X}^n$ 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. Clearly, $C_i(\boldsymbol{X}) = X_i / \sum_{i=1}^n X_i$, $i = 1, \dots, n$, which represents the relative risk contribution of the risk due to the i-th BU in a financial entity, is a legitimate compositional map. Then we can immediately rewrite Equation (4) as

$$r_i = C_i(A(X_1, S), \dots, A(X_n, S)),\ i = 1, \dots, n,$$

and so $H\left(S \times C_i(A(X_1, S), \dots, A(X_n, S))\right)$ recovers allocation rules (1)-(3) for the appropriate choice of the map A. Further, acknowledging TDA / BUA, we arrive at

$$H\left(S_{\text{TDA}}\right) \times C_i(A(X_1, S_{\text{BUA}}), \dots, A(X_n, S_{\text{BUA}})), \tag{5}$$

which recaps, with the help of the language of compositions, the way in which the allocation exercise is usually realized nowadays.

Equation (5) is in fact an attempt to unify the TDA and the BUA rules into one universal method to allocate risk capital. However, unifications à la Equation (5) imply a somewhat naive two-stage procedure, which assumes that the TDA and the BUA are conducted independently and do not impact the outcome of each other (see, Chong et al., 2019, for another similar discussion). Also, Equation (5) reiterates that allocation rules (1)-(3) are all proportional allocations that seek to quantify the relative contribution of the risk due to the i-th BU to the aggregate risk of the financial entity. Hence, the basic stochastic object of interest when computing the allocated risk capital under allocation rules (1)-(3) must be the ratio RV $R_i = X_i/S,\ i = 1, \dots, n$.

Consequently, in the present paper we propose to carry out the allocation exercise in the relative, rather than absolute, terms. Namely, for $v : [0,\ \infty) \times [0,\ \infty) \to [0,\ \infty)$ such that $v(s, s) = s,\ s \in [0.\ \infty)$, we put forward the allocation rule

$$A\big(S_{\text{TDA}} \times C_i(\boldsymbol{X}), v(S_{\text{TDA}}, S_{\text{BUA}})\big), \tag{6}$$

as an alternative to Equation (5). There are a few good reasons for this substitution. First, Equation (6) allocates risk capital due to the product RV, $S_{\text{TDA}} \times X_i/S_{\text{BUA}}$, which captures the desired stochastic phenomenon of relative contribution and, also, accounts for non-trivial interactions among the random components arising from the TDA and the BUA approaches. Second, Equation (6) admits the fact that the real aggregate risk RV of the financial entity is seen differently through the lens of the TDA and the BUA approaches, and so the allocation exercise is to be realized with respect to a combination $v(S_{\text{TDA}}, S_{\text{BUA}})$ of the two aggregate risk RVs. We note in passing, that if $S_{\text{TDA}} = S_{\text{BUA}}$ almost surely and so $v(s) = s,\ s \in [0,\ \infty)$, then the values of the allocated risk capital due to Equations (5) and (6) agree, but this does not happen otherwise. Finally, Equation (6) paves the way to a genuine unification of the TDA and the BUA approaches into one consolidated method to allocate risk capital, since - unlike in Equation (5) - the syntheses now occurs at the level of the random notions of interest.

Admittedly, allocation rule (6), even when confided to the context of the weighted allocations, A_w, raises a number of questions, each of which deserves a stand-alone study. For instance, one may wonder as to how the function v is chosen; obviously, the choice of this function would involve weighting the RVs S_{TDA} and S_{BUA} with respect to, e.g., the credibility of data when constructing each one of these RVs and their CDFs. Another natural question may be as to what the dependence between the RVs S_{TDA} and $C_i(\boldsymbol{X})$ is; clearly, these two RVs must have strong positive dependence.

Mainly, and this is the question that we address in the present paper, it is critical to comprehend what the proposed shift from the allocation rules in absolute terms as per Equation (5) to the allocation rules in the relative terms as per Equation (6) ensues. In the context of the weighted risk capital allocations, which include the conditional tail expectation (CTE)-, Esscher- based allocations (Furman and Zitikis, 2008a), and under some regularity conditions even the distorted allocations, as special cases, this means examining the map $A_w^c : [0,\ 1] \times [0, \infty) \to [0,\ 1]$, such that

$$A_w^c(R_i,\ S) = \frac{\mathbb{E}[R_i w(S)]}{\mathbb{E}[w(S)]},\ i = 1, \ldots, n, \tag{7}$$

which is finite and well-defined as long as the expectation $\mathbb{E}[w(S)]$ is finite and well-defined. As the maps A_w and A_w^c operate on very different domains, the latter map requires very special distributional tools that are put forward and studied in Section 4 of this paper.

Similarly to how the indicator weight function $\mathbf{1}\{s > F_S^{-1}(p)\},\ p \in [0,\ 1)$, plays a prominent role within the framework of the weighted allocation rule, A_w, so does it in the context of the new allocation rule, A_w^c, by emphasizing the extreme scenarios in the sample space of the aggregate risk RV S. Interestingly, computing the *compositional* variant of the noble CTE-based allocation rule is not difficult (Section 5); such variant is formally given by

$$\text{CTE}_p^c(R_i,\ S) = \mathbb{E}[R_i |\ S > F_S^{-1}(p)],\ i = 1, \ldots, n, \tag{8}$$

which is finite and well-defined if $\mathbb{P}(S > F_S^{-1}(p)) \neq 0$, as opposed to the well-acclaimed $\mathrm{CTE}_p[X_i,\ S]$ that requires the finiteness of the mean of the RV $X_i \in \mathcal{X}$.

The rest of the paper capitalizes on the just-outlined ideas. Specifically in Section 3, we reveal a new multivariate distribution that is able to model an arbitrary continuous multivariate distribution with non-negative support arbitrarily well, and so is a natural choice to serve as the distribution of the risk RVs $X_1, \ldots, X_n \in \mathcal{X}$. The inventions presented in Section 3 are further employed in Section 4 to construct a flexible yet tractable framework to formulate and model the vector of random proportions $\boldsymbol{R} = (R_1, \ldots, R_n) \in \mathbb{S}^n$ as well as the corresponding joint CDF, which gives birth to a particularly versatile variant of the well-known Dirichlet distribution on the n-dimensional simplex. Finally in Section 5, we sketch an expectation maximization (EM) algorithm to facilitate applications of our constructions and provide an example borrowed from the context of the risk capital allocation problem. Section 6 concludes the paper.

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 RVs on $(\Omega, \mathcal{A}, \mathbb{P})$ with finite $r(\in [0,\ \infty))$-th moment, and let L^∞ denote the set of all essentially bounded RVs. Unless stated otherwise, we assume that RVs 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 $\boldsymbol{X} = (X_1, \ldots, X_n) \in \mathcal{X}^n$, 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 space

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

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 numbers, $\mathbb{R}_+ \cup \{0\}$; the sets $\mathbb{N}_0^n$ and $\mathbb{R}_{0,+}^n$ denote the corresponding multivariate counterparts.

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., Chang et al., 2010; Ng et al., 2011, 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)

$$f_{\Gamma_i}(x) = \frac{1}{\Gamma(\gamma)} e^{-x/\beta_i} x^{\gamma_i - 1} \beta_i^{-\gamma_i} \text{ for all } x \in \mathbb{R}_+,\ i = 1, \ldots, n. \tag{9}$$

Succinctly, we write $\Gamma_i \sim Ga(\gamma_i, \beta_i),\ i = 1, \ldots, n$. Then assume that the RVs $\Gamma_1, \ldots, \Gamma_n$ 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. Namely, the joint PDF of the RV $\boldsymbol{R}$ is

$$f_{\boldsymbol{R}}(r_1, \ldots, r_n) = \frac{1}{\mathrm{B}(\gamma_1, \ldots, \gamma_n)} \prod_{i=1}^{n} r_i^{\gamma_i - 1},\ (r_1, \ldots, r_n) \in \mathbb{S}^n, \tag{10}$$

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

$$\mathrm{B}(\gamma_1, \ldots, \gamma_n) = \frac{\Gamma(\gamma_1) \times \cdots \times \Gamma(\gamma_n)}{\Gamma(\gamma_1 + \cdots + \gamma_n)}. \tag{11}$$

The Dirichlet distribution is convenient to work with, but unfortunately, it barely suits our needs for many reasons. For example, 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 RVs $R_i = X_i/S,\ i = 1, \ldots, n$ and S are independent (e.g., Ng et al., 2011, 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 such classes of distributions that: (a) are flexible to the extent that they can model well any CDF with non-negative support; (b) allow for a dependence among $X_1, \ldots, X_n$; (c) contain the gamma distribution as a special case; (d) inhere the tractability of the gamma distributions; and (e) relax the assumption of independence of the RVs $R_i,\ i = 1, \ldots, n$, and S, as well as some other rather restrictive notions of independence on the simplex that characterize the class of Dirichlet distributions (e.g., Aitchison, 1982).

3.1 A multivariate mixed-gamma distribution

The class of univariate mixed-Erlang distributions (e.g., Tijms, 1994; Willmot and Lin, 2011) 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 CDFs with non-negative support, fairly tractable, and ensure straightforward multivariate extensions (e.g., Lee and Lin, 2010). That said, when chosen as a basis for a compositional map, mixed-Erlang distributions cannot incorporate PDF (10), hence adjustments have to be made. This is achieved in the following definition.

Definition 1. *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*

$$f_{\Gamma^{(\kappa)}}(x) = \sum_{k=0}^{\infty} p_\kappa(k) \frac{1}{\Gamma(\gamma_k)} e^{-x/\beta} x^{\gamma_k - 1} \beta^{-\gamma_k} \text{ for all } x \in \mathbb{R}_+. \tag{12}$$

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 (e.g., Patil and Ord, 1976, for a thorough discussion of the notion of size-biasing)*

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

In view of this, the class of mixed-gamma distributions can be considered a size-biased mixture, so the notation $\Gamma^{(\kappa)}$ where κ is the random order of the size-bias operation, is natural.

Definition 1 leads to a variety of attractive properties for the MG class of distributions. We start with the Laplace transform of PDF (12). To this end, let $P_\kappa(z) = \mathbb{E}\left[z^\kappa\right], \; |z| \leq 1$, denote the probability generating function (PGF) of the RV κ and recall that, for $\Gamma \sim Ga(\gamma, \beta)$, the Laplace transform is

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

Then we have the following assertion; the proof of which is relegated to the Appendix.

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

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

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 RVs distributed exponentially with the scale parameter $\beta \in \mathbb{R}_+$; here "$\stackrel{d}{=}$" means equality in distribution.

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

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

We next use the Laplace transform of the RV $\Gamma^{(\kappa)}$ to show that the class of MG distributions is a good modeling tool. The proof is again relegated to the Appendix.

Theorem 2. *The MG distributions are dense in the class of all continuous distributions with non-negative support.*

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 1. The extension that we put forward next is inspired by the multivariate mixed-Erlang distributions studied in Lee and Lin (2012); Willmot and Woo

(2014); Verbelen et al. (2016). Namely, the multivariate mixed-gamma distributions presented in Definition 2 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 RVs, $\kappa_i \in N_0,\ i = 1, \ldots, n$, and denote the joint PMF of $\boldsymbol{k} = (k_1, \ldots, k_n) \in \mathbb{N}_0^n$ by $p_{\boldsymbol{\kappa}}(\boldsymbol{k}) = \mathbb{P}\left(\kappa_1 = k_1, \ldots, \kappa_n = k_n\right)$.

Definition 2. *The RV $\mathbf{\Gamma}^{(\boldsymbol{\kappa})} = (\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)})$ is said to be distributed n-variate mixed-gamma (MG_n) if the corresponding joint PDF is given by*

$$f_{\mathbf{\Gamma}^{(\boldsymbol{\kappa})}}(x_1, \ldots, x_n) = \sum_{\boldsymbol{k} \in \mathbb{N}_0^n} p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \prod_{i=1}^{n} \frac{1}{\Gamma(\gamma_{k_i})} e^{-x_i/\beta_i} x_i^{\gamma_{k_i}-1} \beta_i^{-\gamma_{k_i}},\ (x_1, \ldots, x_n) \in \mathbb{R}_+^n, \tag{13}$$

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, γ_{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 of 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. The proof of the Theorem 3 is in the Appendix.

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 $(t_1, \ldots, t_n) \in \mathbb{R}_{0,+}^n$ 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 RVs 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 RVs $\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)}$ are independent.

(iv) Choose $1 \le i \ne j \le n$ 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

$$\mathrm{Corr}(\Gamma_i^{(\kappa_i)}, \Gamma_j^{(\kappa_j)}) = \mathrm{Corr}(\kappa_i, \kappa_j) \frac{\sqrt{\mathrm{Var}(\kappa_i)\mathrm{Var}(\kappa_j)}}{\sqrt{\left(\mathrm{Var}(\kappa_i) + \mathbb{E}[\kappa_i] + \gamma_i\right)\left(\mathrm{Var}(\kappa_j) + \mathbb{E}[\kappa_j] + \gamma_j\right)}}. \tag{14}$$

Note 2. *Correlation formula (14) 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* $\mathrm{Corr}(\kappa_i, \kappa_j)$, $1 \le i \ne j \le n$. *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 are dense, and so can model any multivariate distribution with positive support arbitrarily well. The denseness property guarantees a desirable level of flexibility for the proposed MG_n distributions, which helps to mitigate the miss-specification risk in the model selection process. The proof of the next 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 RVs distributed gamma with arbitrary shape and scale parameters are mixed-gamma. The next theorem is reported for completeness of exposition (e.g., Moschopoulos, 1985, 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., Hürlimann, 2005; Furman and Landsman, 2005; Su and Furman, 2017, and references therein).

Theorem 5. *For* $i = 1, \ldots, n$, *let* $\Gamma_i \sim Ga(\gamma_i,\ \beta_i)$ *denote independent RVs 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*

$$c = \prod_{i=1}^n \left(\frac{\beta^*}{\beta_i}\right)^{\gamma_i} \text{ and } \delta_k = k^{-1} \sum_{l=1}^k \sum_{i=1}^n \gamma_i \left(1 - \frac{\beta^*}{\beta_i}\right)^l \delta_{k-l}, \text{ for } k \in \mathbb{N}, \text{ and } \delta_0 = 1. \tag{15}$$

We now generalize Theorem 5 - the proof of the validity of the generalization is in the Appendix - 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*

$$p_{\kappa^*}(m) = \sum_{j=0}^{m} \sum_{k_1+\cdots+k_n=j} \left(p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \sum_{\substack{y_1+\cdots+y_n=m-j \\ (y_1,\ldots,y_n)\in\mathbb{N}_0^n}} \prod_{i=1}^{n} \frac{\Gamma(\gamma_{k_i}+y_i)}{\Gamma(\gamma_{k_i})\, y_i!} \left(\frac{\beta^*}{\beta_i}\right)^{\gamma_{k_i}} \left(1-\frac{\beta^*}{\beta_i}\right)^{y_i} \right) \tag{16}$$

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

A by-product of Theorem 6 is that it demystifies the recursive formula presented in Moschopoulos (1985) 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 7. *Within the setup in Theorem 5, we have* $\kappa^{**} \stackrel{d}{=} \sum_{i=1}^{n} N_i$, *where* $N_i \sim NB(\gamma_i, \beta^*/\beta_i)$ *are mutually independent RVs having negative binomial distributions. The PMF of the RV* κ^{**} *admits the following (non-recursive) form*

$$p_{\kappa^{**}}(k) = \sum_{\substack{y_1+\cdots+y_n=k \\ (y_1,\ldots,y_n)\in\mathbb{N}_0^n}} \prod_{i=1}^{n} \frac{\Gamma(\gamma_i+y_i)}{\Gamma(\gamma_i)\, y_i!} (\beta^*/\beta_i)^{\gamma_i} (1-\beta^*/\beta_i)^{y_i},\ k \in \mathbb{N}_0. \tag{17}$$

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

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

Dirichlet PDF (10) is remarkably tractable. For example, let $\gamma^* = \gamma_1+\cdots+\gamma_n$, $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_n) \in \mathbb{R}_+^n$, and given that $\boldsymbol{R} = (R_1,\ldots,R_n) \sim Dir(\boldsymbol{\gamma})$, 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

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

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. For example, 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 (see, Aitchison, 1986, for details). Further, 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 (10) (e.g., Ng et al., 2011, among others). The task is, however, not an easy call. Namely, a slight generalization of the setup 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 RVs (note that the scale parameters are arbitrary now), and let Γ_+ denote the sum of these RVs. Then the RV $\boldsymbol{R} = (R_1, \ldots, R_n),\ R_i = \Gamma_i/\Gamma_+$ is distributed scaled Dirichlet (e.g., Ng et al., 2011), which is far less tractable than the one with PDF (10). In particular, even an analytic expression for the covariance was not known for the scaled Dirichlet distribution.

In this section, we use the class of multivariate MG distributions as the basis to formulate a generalization of Dirichlet distribution, which is suitable for studying capital allocation problems. Recall that we write $\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})} \sim MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$ when $\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})}$ 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 RVs distributed multivariate MG.

Following the language of Aitchison (1986), we call the RV $\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})} = (\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)}) \in \mathcal{X}_+^n$, basis. Then we are interested in mapping collections of RVs in $\mathcal{X}^n$ to the n-dimensional simplex $\mathbb{S}^n$ (see, Section 2.1 for a definition). In this paper, because of the nature of the capital allocation exercise, our working choice is the map $C : \mathcal{X}_+^n \to \mathbb{S}^n$, such that

$$C_i(\Gamma_1^{(\kappa_1)}, \ldots, \Gamma_n^{(\kappa_n)}) = \Gamma_i^{(\kappa_i)}/\Gamma_+^{(\kappa^*)} = R_i. \tag{18}$$

Random compositions (18) 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 (18). The proof of this assertion is in the Appendix.

Theorem 8. *Let $\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})} \sim MG_n(\boldsymbol{\gamma}, \boldsymbol{\beta}, p_{\boldsymbol{\kappa}})$, namely 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 (18). Then the joint PDF of the RV $\boldsymbol{R}$ is given by*

$$f_{\boldsymbol{R}}(r_1, \ldots, r_n) = \sum_{\boldsymbol{k} \in \mathbb{N}_0} \frac{p_{\boldsymbol{\kappa}}(\boldsymbol{k})}{\mathrm{B}\left(\boldsymbol{\gamma}_{\boldsymbol{k}}\right)} \prod_{i=1}^n \frac{1}{\beta_i} \left(\frac{r_i}{\beta_i}\right)^{\gamma_{k_i} - 1} \left(\sum_{i=1}^n \frac{r_i}{\beta_i}\right)^{-\sum_{i=1}^n \gamma_{k_i}} \tag{19}$$

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

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 (19) reduces to the PDF of the scaled Dirichlet distribution, heuristically studied in, e.g., Monti et al. (2011a) and Monti et al. (2011b). If an additional assumption that the scale parameters are chosen such that $\beta_1 = \cdots = \beta_n$ is made, then PDF (19) coincides with PDF (10). 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*

(18), 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 (19). 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 (19) 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 (19) directly, we make use of the associated stochastic representation instead (e.g., Ng et al., 2011, 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 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 $\gamma^*_{\mathcal{A}}=\sum_{i\in\mathcal{A}}\gamma_i$ and $\beta^*_{\mathcal{A}}=\bigwedge_{i\in\mathcal{A}}\beta_i$ for any $\mathcal{A}\subseteq\{1,\ldots,n\}$. We are now ready to prove - the formal proof is in the Appendix - that the class of mixed-scaled Dirichlet distributions is closed under the marginalization of any order.

Theorem 9. *The RV $\boldsymbol{R}\sim Dir(\boldsymbol{\gamma},\boldsymbol{\beta},p_{\boldsymbol{\kappa}})$ with PDF (19) 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 $\Box_{\mathcal{I}}=\{\Box_i: i\in\mathcal{I}\}$, "$\Box$" can be any one of $\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{\kappa}$, and the joint PMF*

$$p_{(\boldsymbol{\kappa}_{\mathcal{I}},\kappa^*_{\mathcal{I}^c})}(\boldsymbol{k}_{\mathcal{I}},m)=\sum_{j=0}^{m}\sum_{\sum_{v\in\mathcal{I}^c}k_v=j}\left[p_{\boldsymbol{\kappa}}(\boldsymbol{k})\sum_{\substack{\sum_{v\in\mathcal{I}^c}y_v=m-j\\ y_v\in\mathbb{N}_0}}\prod_{i\in\mathcal{I}^c}\frac{\Gamma(\gamma_{k_i}+y_i)}{\Gamma(\gamma_{k_i})\,y_i!}\left(\frac{\beta^*_{\mathcal{I}^c}}{\beta_i}\right)^{\gamma_{k_i}}\left(1-\frac{\beta^*_{\mathcal{I}^c}}{\beta_i}\right)^{y_i}\right] \tag{20}$$

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

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 PDFs, $k<n$, can be derived with the help of Theorems 8 and 9. For an illustration, we next report the

univariate and bivariate PDFs. Marginal PDFs 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 (20).

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

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

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 (20).

Note 4. *RV $X \in L^\infty$ is said to be distributed generalized three-parameter beta if the associated PDF is given by (Libby and Novick, 1982)*

$$f_X(x) \ = \ \frac{\lambda^a}{B(a,b)} \frac{x^{a-1}(1-x)^{b-1}}{(1+(\lambda-1)x)^{a+b}},\ \ x \in [0,1], \tag{21}$$

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 (Gupta and Nadarajah, 2004). 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

$$\kappa^*_{\mathcal{N}_{(i)}} \ = \ \sum_{j \in \mathcal{N}_{(i)}} (\kappa_j + N_j(\kappa_j)),$$

with the RVs $N_j(\kappa_j) \sim NB(\gamma_{\kappa_j}, \beta^*_{\mathcal{N}_{(i)}}/\beta_j),\ j \in \mathcal{N}_{(i)}$ are conditionally independent given the RV $\boldsymbol{\kappa}$.

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

$${}_{q+1}\mathrm{F}_q(a_1, \ldots, a_{q+1}; b_1, \ldots, b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k, \ldots, (a_{q+1})_k}{(b_1)_k, \ldots, (b_q)_k} \frac{z^k}{k!},\ \ |z| < 1, \tag{22}$$

where $(x)_n = \Gamma(x+n)/\Gamma(x)$ denotes the Pochhammer symbol (Gradshteyn and Ryzhik, 2014). We also need the

Appell's F_1 function, which is given by

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

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 10. *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*

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

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 (20).

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

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

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 (20).

Note 5. *The covariance between any pair of RVs R_i and R_j within the mixed-scaled Dirichlet class can be readily computed via the moment formulas in Theorem 10. Interestingly, unlike for the classical Dirichlet distribution with PDF (10), 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 RVs κ_1, κ_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 10 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., Monti et al., 2011a; Ng et al., 2011, for details). In this respect, Theorem 10 provides analytical and conveniently computable expressions for the desired moment formulas. Specifically, set $\kappa_i\equiv 0$ in Theorem 10, then, for $r\in\mathbb{R}_+$

and $i = 1, \dots, n$,

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

where $\gamma^* = \sum_{i=1}^n \gamma_i$, $\beta^*_{\mathcal{N}_{(i)}} = \bigwedge_{j\in\mathcal{N}_{(i)}} \beta_j$, and $\kappa^*_{\mathcal{N}_{(i)}} \stackrel{d}{=} \sum_{j\in\mathcal{N}_{(i)}} N_j$ with the RVs 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 (17) or recursively via (15).

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

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

where $\beta^*_{\mathcal{N}_{(i,j)}} = \bigwedge_{v\in\mathcal{N}_{(i,j)}} \beta_v$ and $\kappa^*_{\mathcal{N}_{(i,j)}} \stackrel{d}{=} \sum_{v\in\mathcal{N}_{(i,j)}} N_v$ with the RVs N_v being mutually independent and $N_v \sim NB(\gamma_v, \beta^*_{\mathcal{N}_{(i,j)}}/\beta_v)$.

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 3. *For $\mathcal{I} = \{i_1, \dots, i_j\} \subset \mathcal{N}$, $j < n$, the vector*

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

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

Definition 4. *Let $\{\mathcal{I}_k\}_{k=1}^m$ where $\mathcal{I}_k = \{i_{k,1}, \dots, i_{k,j_k}\} \subset \mathcal{N}$, $j, m < n$, denote a disjoint coverage of the set $\{1, \dots, n\}$, that is $\cup_k \mathcal{I}_k = \{1, \dots, n\}$ and $\mathcal{I}_k \cap \mathcal{I}_h = \emptyset$ for $k \neq h$. Each set $\mathcal{I}_k$ gives rise to the sub-composition $\mathbf{S}_{\mathcal{I}_k}$ with the corresponding basis $\left(\Gamma_{i_{k,1}}^{(\kappa_{i_{k,1}})}, \dots, \Gamma_{i_{k,j_k}}^{(\kappa_{i_{k,j_k}})}\right)$. Then the vector*

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

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. The proofs are again relegated to the Appendix.

Theorem 11. *The RV $\boldsymbol{R} = (R_1, \ldots, R_n)$ with joint PDF (19) 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

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

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

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

where the RVs 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 $\boldsymbol{k}^* = (k^*_1, \ldots, k^*_m) \in \mathbb{N}^m_0$, the joint PMF of the RV $\boldsymbol{\kappa}^*_{\boldsymbol{\mathcal{I}}}$ can be computed via

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

where $q_v(z) = \mathbb{P}(\sum_{i \in \mathcal{I}_v} N_{v,i}(k_i) = z),\ z \in \mathbb{N}_0$, with the RVs $N_{v,i}$ being mutually independent and such that $N_{v,i}(k_i) \sim NB(\gamma_i + k_i, \beta^*_{\mathcal{I}_v}/\beta_i)$ for $i \in \mathcal{I}_v,\ v = 1, \ldots, m$; the function q_v can be computed with the help of Equation (17).

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 (18), 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. Our goal in what follows is to seek to compute the compositional variant of the CTE-based weighted allocation rule

$$\mathrm{CTE}^c_p(R_i,\ S) = \mathbb{E}[R_i |\ S > F_S^{-1}(p)],\ i = 1, \ldots, n,\ p \in [0, 1), \tag{23}$$

which can be computed using the joint PDF of the RVs R_i and S:

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

where $r\in[0,1]$, $s\in\mathbb{R}_+$, $\beta^*_{\mathcal{N}(i)}=\bigwedge_{i\in\mathcal{N}(i)}\beta_i$, $\gamma_{k^*}=\gamma^*_{\mathcal{N}(i)}+k^*$, and the joint PMF $p_{(\kappa_i,\kappa^*_{\mathcal{N}(i)})}$ follows from Equation (20). 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.

5.1 Estimation of parameters

Consider observations $\boldsymbol{x}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_d)'$, with $\boldsymbol{x}_j=(x_{1j},\ldots,x_{nj})$, $j=1,\ldots,d$, which represent sample losses 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 PDF (13) can be written as

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

The expectation-maximization (EM) algorithm is a common choice for estimating the parameters of finite mixtures. It was proposed in Dempster et al. (1977) (also, Wu, 1983) for statistical estimation in the contexts with incomplete data. We refer to, e.g., Karlis (2003) and Asimit et al. (2016) 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 Lee and Lin (2012) (also, Verbelen et al., 2016) 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 $p_{\boldsymbol{\kappa}}(\boldsymbol{k})$, $\boldsymbol{k}\in\mathcal{M}\subset\mathbb{N}_0^n$. To initialize the parameters, including the choice of the set $\mathcal{M}\subset\mathbb{N}_0^n$, we adopt the procedure in Lee and Lin (2012). Then we conduct the "expectation" (E) stage. That is, for $s\in\mathbb{N}_0$, let $\boldsymbol{\Psi}^{(s)}=(p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k}),\boldsymbol{\beta}^{(s)},\boldsymbol{\gamma}^{(s)})$ 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

$$Q(\boldsymbol{\Psi}|\ \boldsymbol{\Psi}^{(s)}) = \sum_{j=1}^{d}\sum_{\boldsymbol{k}\in\mathcal{M}}\left[\log(p_{\boldsymbol{\kappa}}^{(s)}(\boldsymbol{k})) + \sum_{i=1}^{n}\left((\gamma_i+k_i-1)\log(x_{ij}) - \frac{x_{ij}}{\beta_i} - (\gamma_i+k_i)\log(\beta_i) - \log(\Gamma(\gamma_i+k_i))\right)\right] q(\boldsymbol{k}|\boldsymbol{x}_j, \boldsymbol{\Psi}^{(s)}), \tag{24}$$

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

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

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

  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`" (Varadhan and Gilbert, 2015).

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

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 RVs representing the risks due to the BUs are distributed Pareto, log-normal, and gamma. 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., Bahnemann, 2015, for examples). Furthermore, we assume that the dependencies among the RVs X_1, X_2 and X_3 are governed by the Gaussian copula with the correlation matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1.00 & 0.50 & 0.25 \\ 0.50 & 1.00 & -0.50 \\ 0.25 & -0.50 & 1.00 \end{pmatrix}, \tag{25}$$

with the entries being motivated by the matrix used in the quantitative impact study published by the European Insurance and Occupational Pensions Authority (EIOPA, 2010).

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 5.1, we estimate the parameters of the multivariate mixed-gamma distribution, which are summarized in Table 1. In addition, Figure 3 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.

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

$\gamma_{k_{i10}}$	$\gamma_{k_{i11}}$	$\gamma_{k_{i12}}$
2.98	11.98	35.98
2.13	5.13	6.13
15.19	15.19	14.19
0.0778	0.0277	0.0078

Table 1: The parameters of the multivariate mixed-gamma distribution fitted against the simulated data.

Finally, based on the obtained parameters for the multivariate mixed-gamma distribution, we obtain the mixed-scaled Dirichlet distribution that describes the joint behavior of the random proportions, and compute the values of a few risk capital allocation rules, which are presented in Table 2.

Table 2 hints at the following observations.

- The substitution of the "composition of allocations" method with the proposed in this paper "allocation of

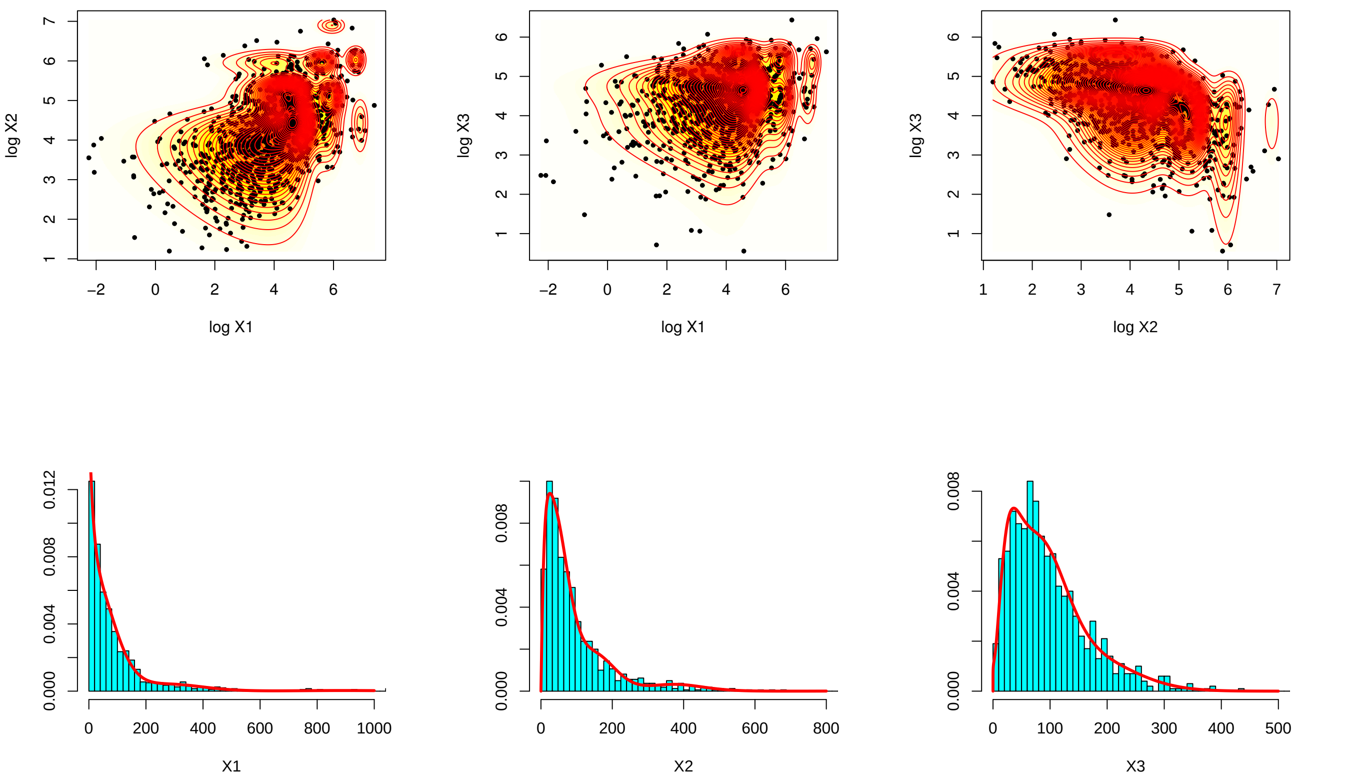

Figure 3: Bivariate log transformed density contours and marginal histograms for the fitted multivariate mixed-gamma distribution.

#	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
	$\mathrm{CTE}^{c}_{0.95}(R_i,\,S)$	0.546 (−2.3%)	0.319 (0.6%)	0.135 (8.9%)

Table 2: Comparisons of the “composition of allocations” method and the “allocation of a composition” method with the help of the fitted mixed-scaled Dirichlet distribution; $R_i = X_i/S$.

a composition” method leads to outcomes of the risk capital allocation exercise that differ in both order and magnitude (e.g., the case of 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 inappropriate risk capital allocations.

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

- A closer inspection of the CTE-based proportional allocation formula helps to elucidate a subtle differences between the two approaches studied in case #2. For this, we recall the elementary formula for computing the conditional covariance of any pair of RVs $X,\, Y \in \mathbb{R}$ and an event $\Sigma \subseteq \Omega$

$$\mathrm{Cov}(X, Y \,|\, \Sigma) = \mathbb{E}[X\,Y|\Sigma] - \mathbb{E}[X|\Sigma]\,\mathbb{E}[Y|\Sigma],$$

given that the expectations are finite and well-defined. Denote by $s_p = F_S^{-1}(p)$ the inverse CDF of the aggregate risk, then for $i = 1, \ldots, n$ and $p \in [0, 1)$, we observe

$$\frac{\mathrm{CTE}_p(X_i, S)}{\sum_{i=1}^{n} \mathrm{CTE}_p(X_i,\, S)} = \mathrm{CTE}_p^c(R_i, S) + \frac{\mathrm{Cov}(R_i, S \,|\, S > s_p)}{\mathbb{E}\big[S \,|\, S > s_p\big]} = \mathrm{TCov}_p(R_i, S), \tag{26}$$

where the functional $\mathrm{TCov}_p(\cdot)$ is known as the modified tail covariance (Furman and Landsman, 2006, and references therein). Hence, the classic CTE-based proportional allocation rule and its compositional counterpart are explicitly connected, and the sign of the covariance $\mathrm{Cov}(R_i, S \,|\, S > s_p)$, $p \in [0,\ 1)$ is the decisive factor as to the order of magnitude of the two approaches to allocate risk capital. This finding is reflected in Table 2. Namely, according to the setup of correlation matrix (25), the risk contribution due to BU 1 has the highest positive conditional correlation $\mathrm{Corr}(R_1, S|\ S > s_{0.95}) = 0.24$ among the three risks ($\mathrm{Corr}(R_2, S|\ S > s_{0.95}) = -0.03$ and $\mathrm{Corr}(R_3, S|\ S > s_{0.95}) = -0.31$), and this yields

$$\frac{\mathrm{CTE}_{0.95}(X_1,\, S)}{\sum_{i=1}^{3} \mathrm{CTE}_{0.95}(X_i,\, S)} > \mathrm{CTE}_{0.95}^c(R_1,\, S).$$

The same rationale can be used to explain the differences between the two allocation methods for BU 2 and BU 3.

- Stochastic dependence is not the only driver that dictates the orders of the outcomes of the risk capital allocation exercise within 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 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 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.

6 Conclusions

There exist a great variety of distinct ways to allocate the aggregate risk capital to constituents. While the choice of the most appropriate allocation rule should be dictated by the goals of the exercise, and so may vary from task to task, all 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.

In this work, we have discussed the idea of replacing 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 risk capital allocation. A 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.

Our numerical study suggests that the classical approach to the risk capital allocation exercise and the one studied in this paper lead to results that may differ in both order and magnitude of the obtained values of the allocated risk capital. A notable observation in this respect relates to the notion of positive dependence. Namely, in the classical approach to allocating risk capital, e.g., the CTE-based allocation rule, or the class of weighted allocation rules, stronger positive dependencies of the risk due to a business line on the aggregate risk of the financial entity assuredly imply larger shares of the aggregate risk capital. This guideline is naturally toned down when random proportions are considered.

Acknowledgments

We are grateful to the Society of Actuaries for the generous financial support of our research, as well as for appointing a very knowledgeable project oversight group, whose insightful comments and thought-provoking questions resulted in a much-improved version of this paper. Edward Furman acknowledges the continuous support of his research by the Natural Sciences and Engineering Council of Canada.

References

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

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

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

Asimit, A. V., Furman, E., and Vernic, R. (2016). Statistical inference for a new class of multivariate Pareto distributions. *Communications in Statistics - Simulation and Computation*, 45(2):458–471.

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

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.

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

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

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

Cai, J. and Li, H. (2010). Conditional tail expectations for multivariate phase-type distributions. *Journal of Applied Probability*, 42(3):810–825.

Chang, W.-Y., Gupta, R. D., and Richards, D. S. P. (2010). Structural properties of the generalized Dirichlet distributions. *Contemporary Mathematics*, 516:109–124.

Chong, W., Feng, R., and Jin, L. (2019). Holistic principle for risk aggregation and capital allocation. *Working paper*, pages 1–41.

Cossette, H., Côté, M.-P., Marceau, E., and Moutanabbir, K. (2013). Multivariate distribution defined with Farlie-Gumbel-Morgenstern copula and mixed Erlang marginals: Aggregation and capital allocation. *Insurance: Mathematics and Economics*, 52(3):560–572.

Cossette, H., 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.

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):1–38.

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

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.

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

EIOPA (2010). QIS 5 technical specifications. Technical report, European Insurance and Occupational Pensions Authority.

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.

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

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

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

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

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

Gradshteyn, I. S. and Ryzhik, I. M. (2014). *Table of Integrals, Series, and Products*. Academic Press, New York.

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

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

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.

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

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

Karlis, D. (2003). ML estimation for multivariate shock models via an EM algorithm. *Annals of the Institute of Statistical Mathematics*, 55(4):817–830.

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.

Lee, S. C. K. and Lin, X. S. (2012). Modeling dependent risks with multivariate Erlang mixtures. *ASTIN Bulletin*, 42(1):153–180.

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.

Miles, J., Furman, E., and Kuznetsov, A. (2019). Risk aggregation: A general approach via the class of generalized gamma convolutions. *Variance, in press*.

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.

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*.

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.

Ng, K. W., Tian, G. L., and Tang, M. L. (2011). *Dirichlet and Related Distributions: Theory, Methods and Applications*. Wiley, Chichester.

Patil, G. and Ord, J. (1976). On size-biased sampling and related form-invariant weighted distributions. *Sankhya: The Indian Journal of Statistics Series B*, 38(1):48–61.

Su, J. and Furman, E. (2017). Multiple risk factor dependence structures: Distributional properties. *Insurance: Mathematics and Economics*, 76:56–68.

Tijms, H. C. (1994). *Stochastic Models: An Algorithmic Approach*. Wiley, Chichester.

Tsanakas, A. and Barnett, C. (2003). Risk capital allocation and cooperative pricing of insurance liabilities. *Insurance: Mathematics and Economics*, 33:239–254.

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.

Verbelen, R., Antonio, K., and Claeskens, G. (2016). Multivariate mixtures of Erlangs for density estimation under censoring. *Lifetime Data Analysis*, 22(3):429–455.

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance: Mathematics and Economics*, 38(2):413–426.

Vernic, R. (2011). Tail conditional expectation for the multivariate Pareto distribution of the second kind: Another approach. *Methodology and Computing in Applied Probability*, 13(1):121–137.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26(1):71–92.

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.

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.

Wu, C. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, 11(1):95–103.

Appendix A Technical proofs

Proof of Theorem 1. 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

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

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

Proof of Theorem 2. The proof of the succeeding assertion is borrowed heavily from Lee and Lin (2010). 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

$$\widehat{f}_n(t) = \left(1+\frac{t}{n}\right)^{-\gamma} \int_0^\infty \left(1+\frac{t}{n}\right)^{-\lfloor xn \rfloor} f(x)dx, \tag{27}$$

where '$\lfloor\cdot\rfloor$' denotes the flooring function. Then, on the one hand side, $\widehat{f}_n(t)$ is the Laplace transform of an MG PDF, that is, for $\beta = 1/n$ and $p_\kappa(k) = F((k+1)\beta) - F(k\beta),\ k \in \mathbb{N}_0$, Equation (27) 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. □

Proof of Theorem 3. 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). □

Proof of Theorem 6. 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 $\kappa^* \stackrel{d}{=} \sum_{i=1}^{n}(\kappa_i + N_i(\kappa_i))$ evaluated at $(1 + \beta^* t)^{-1}$, 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

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

This completes the proof of the theorem. □

Proof of Theorem 8. We begin with the joint PDF of the basis RV $\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})} = (\Gamma_1^{\kappa_1}, \ldots, \Gamma_n^{\kappa_n})$ (Definition 2)

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

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 $s \in \mathbb{R}_+$,

$$f_{\boldsymbol{R},\Gamma_+^{(\kappa^*)}}(r_1, \ldots, r_n, s) = f_{\boldsymbol{\Gamma}^{(\boldsymbol{\kappa})}}(r_1 s, \cdots, r_n s) s^{n-1} = \sum_{\boldsymbol{k} \in \mathbb{N}_0^n} p_{\boldsymbol{\kappa}}(\boldsymbol{k}) \prod_{i=1}^{n} \frac{r_i^{\gamma_{k_i}-1}}{\beta_i^{\gamma_{k_i}} \Gamma(\gamma_{k_i})} \, s^{\sum_{i=1}^n \gamma_{k_i} - 1} e^{-s \sum_{i=1}^n r_i/\beta_i}. \quad (28)$$

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. □

Proof of Theorem 9. 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

$$\kappa_{\mathcal{I}^c}^* = \sum_{i \in \mathcal{I}^c} (\kappa_i + N_i(\kappa_i)), \quad (29)$$

where the RVs $N_i(\kappa_i) \sim NB(\gamma_{\kappa_i}, \beta_{\mathcal{I}^c}^*/\beta_i), \; i \in \mathcal{I}^c$ 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 (16) in Theorem 6. This completes the proof of the theorem. □

Proof of Theorem 10. The r-th order moment formula follows from Note 4 and Gupta and Nadarajah (2004), whereas the joint moment formula is obtained directly by the integral representation of the Appell's F_1 function (see, Equation (9.184) in Gradshteyn and Ryzhik, 2014). This completes the proof of the theorem. □

Proof of Theorem 11. Assertion (i) follows immediately from, e.g., stochastic representation (18). 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 $\kappa^*_{\mathcal{I}_j} \stackrel{d}{=} \sum_{i\in\mathcal{I}_j}(\kappa_i + N_i(\kappa_i))$ with the RVs 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 RVs $N_{j,i}$ are mutually independent and such that $N_{j,i}(k_i) \sim NB(\gamma_i + k_i, \beta^*_{\mathcal{I}_j}/\beta_i)$ for $i\in\mathcal{I}_j$, $j\in(1,\ldots,m)$, $k_i\in\mathbb{N}_0$. This yields the closure under amalgamations property. The proof is completed. □