Multivariate exponential distributions with latent factor structure and related topics

Jan-Frederik Mai

email: janfrederik.mai@gmail.com

The present essay is a summary of ten research articles, written by the author and several co-authors. Together with the original references it constitutes the author's habilitation thesis. The central theme of the thesis is the construction of multivariate probability distributions in large dimensions. Special focus is put on the study of multivariate exponential laws. In particular, those exponential families with conditionally independent and identically distributed components are characterized. Subsequently, hierarchical dependence structures are constructed from one-factor building blocks. Besides a generic recipe for such structures, specific examples that are discussed comprise hierarchical Archimedean copulas, multivariate exponential distributions, as well as combinations thereof. Finally, some closely related topics and an application to portfolio credit risk modeling are sketched.

1 Survey of the thesis

The content of the present habilitation thesis comprises several published research articles, written jointly with German Bernhart, Marcos Escobar-Anel, Christian Hering, Marius Hofert, Pablo Olivares, Steffen Schenk, and Matthias Scherer, a detailed list of references is given below. The remaining sections aim at surveying these articles, and pointing out how they all relate to the central theme, which is the construction of tractable, high-dimensional probability distributions – with a focus on multivariate exponential laws. In order to make the present summary stringent and reader-friendly, the author purposely decided to embellish it with an elaborate introduction and some conjunctive passages. As a consequence some articles are discussed more in-depth than others, but all of them are mentioned at least once in the main body of the present summary. For technical proofs the interested reader is always referred to the original articles, except for two new lemmata, which are included to create a convenient "reading flow".

The precise content of the remaining sections is organized as follows: Section 2 provides a motivating introduction, which prepares the reader for the upcoming sections. In particular, Subsection 2.1 recalls some required mathematical background regarding De Finetti's Theorem and extendibility. Section 3 summarizes the major theoretical findings on extendible exponential distributions. Section 4 shows how to design flexible and low-parametric multi-factor models from De Finetti-type building blocks. Section

5 illustrates applications of the introduced concepts to parameter estimation, numerical density evaluation, and in the context of portfolio credit risk. Section 6 concludes. A detailed list of the published articles surveyed in the present summary is given in the sequel. In the remaining sections, these articles are labeled with "A" (standing for "article") and a number when cited, as opposed to all other citations, so that the reader can quickly realize which citation corresponds to a part of the habilitation thesis, and which does not.

- Articles surveyed in Section 3 (multivariate exponential laws):

  [A1] "The Pickands representation of survival Marshall–Olkin copulas", with M. Scherer, *Statistics and Probability Letters* **80** (2010) pp. 357–360.

  [A2] "Bivariate extreme-value copulas with discrete Pickands dependence measure", with M. Scherer, *Extremes* **14** (2011) pp. 311–324.

  [A3] "Sampling exchangeable and hierarchical Marshall–Olkin distributions", with M. Scherer, *Communications in Statistics - Theory and Methods* **42** (2013) pp. 619–632.

  [A4] "Characterization of extendible distributions with exponential minima via processes that are infinitely divisible with respect to time", with M. Scherer, *Extremes* **17** (2014) pp. 77–95.

- Articles surveyed in Section 4 (construction of hierarchical dependence structures):

  [A5] "Constructing hierarchical Archimedean copulas with Lévy subordinators", with C. Hering, M. Hofert and M. Scherer, *Journal of Multivariate Analysis* **101:6** (2010) pp. 1428–1433.

  [A6] "H-extendible copulas", with M. Scherer, *Journal of Multivariate Analysis* **110** (2012) pp. , 151–160.

  [A7] "Default models based on scale mixtures of Marshall–Olkin copulas: properties and applications", with G. Bernhart, M. Escobar-Anel and M. Scherer, *Metrika* **76** (2012) pp. 179–203.

- Articles surveyed in Section 5 (applications of the introduced concepts):

  [A8] "Moment-based estimation of extendible Marshall–Olkin distributions", with C. Hering, *Metrika* **75** (2012) pp. 601–620.

  [A9] "The density for distributions from the Bondesson class", with G. Bernhart, S. Schenk and M. Scherer, *Journal of Computational Finance, forthcoming* (2013).

  [A10] "A multivariate default model with spread and event risk", with P. Olivares, S. Schenk and M. Scherer, *Applied Mathematical Finance* **21** (2014) pp. 51–83.

2 Introduction and Motivation

The present thesis deals with the stochastic modeling of multivariate distributions. More precisely, a random vector $(X_1, \ldots, X_d)$ with real components is considered. The investigation is motivated by applications where the components are interpreted as time points in the future, i.e. X_k is a future time point at which a certain event E_k happens, $k = 1, \ldots, d$. An example in a financial context is the modeling of a portfolio of d credit-risky assets, and one is interested in the events $E_k = \{\text{asset } k \text{ defaults}\}$. This interpretation makes clear the particular interest for the case that the components of the random vector are all non-negative, i.e. distributions on $[0, \infty)^d$ are studied. Moreover, the focus lies on situations when the components $X_1, \ldots, X_d$ are stochastically dependent, and when the dimension d might be very large, e.g. $d = 125$ or even larger. However, concrete examples are avoided in the sequel (except for Subsection 5.3, which illustrates an application in the context of Mathematical Finance), because the modeling recipe might as well apply to various other situations, and the findings are also interesting for pure theorists.

One goal of the remaining sections is to provide the reader with a generic recipe of how to build multivariate models for dependent time points, which satisfy the following two practical demands:

(a) *The model is intuitive.* A factor-model way of thinking is condoned because this is what people can grasp intuitively very well. As a consequence, it will be possible to add or remove components of the model without destroying its structure. I.e. the model is independent of the dimension d to some degree.

(b) *The number of model parameters can be controlled.* In particular – and this distinguishes the present approach from many other multivariate distributions – the number of parameters in the presented models does not explode with increasing dimension. Again, this goes along with a certain invariance of the models with respect to their dimension.

As mentioned in (a) above, a factor-model approach is pursued. To this end, it is first specified in a mathematically precise way what kind of factor-models are meant, which is done in Subsection 2.1 below. The idea is to build multi-factor models from simpler one-factor models, so that – to a large extent – it suffices to understand the one-factor case. Moreover, a latent factor will always be a non-decreasing, cadlàg stochastic process $H = \{H_t\}$. Thus, multi-factor models are constructed from multiple, independent stochastic processes. The approach is by definition constructive in the sense that the probability space on which the random vector is defined can be written down explicitly, i.e. the components $X_1, \ldots, X_d$ are specified as certain functionals of the factor processes and an independent sequence of random variables which are independent and identically distributed (iid). In particular, a simulation of the model along this construction is straightforward in many cases. The models constructed in this way are then further investigated, e.g. several stochastic properties of the random vector are derived, and it is explored how they stem from stochastic properties of the factor processes.

2.1 De Finetti's Theorem and its implications

We are interested in the stochastic modeling of a random vector $(X_1, \ldots, X_d)$ with real components on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in such a way that the constructed model is convenient to work with and easy to understand. In particular, we would like to model the components stochastically dependent. An intuitive approach to tackle this modeling task is to start from the bottom up, i.e. begin with the simplest kind of dependence and then add more and more structure to it. This is the approach that is pursued in the sequel. The simplest setup is definitely when all components have the same univariate distribution and are independent, we say the random variables $X_1, \ldots, X_d$ are *iid*. This setup is so simple because the whole modeling task boils down to the modeling of one univariate distribution, say of X_1, which determines the overall distribution of the vector. One obvious extension from the iid setup to dependent components is to consider a setup which is *conditionally iid*. What does that mean? In words, conditionally iid means that the probability space supports one stochastic factor, say H, which affects all otherwise identical components alike. Mathematically speaking, conditioned on the (σ-algebra generated by the) latent factor H the components $X_1, \ldots, X_d$ are iid. An example would be if the components $X_1, \ldots, X_d$ were $\{0, 1\}$-valued, Bernoulli variables with success probability $H \in [0, 1]$, but H itself is a random variable which is drawn before the (same) Bernoulli experiment is run d times. Speaking in terms of Bayesian statistics, there is a prior distribution on the success probability H. Clearly, if H is not constant, the components of $(X_1, \ldots, X_d)$ are dependent because, e.g.,

$$\begin{aligned}\mathbb{P}(X_1 = 1, X_2 = 1) &= \mathbb{E}\big[\mathbb{P}(X_1 = 1, X_2 = 1 \,|\, H)\big] = \mathbb{E}\big[\mathbb{P}(X_1 = 1 \,|\, H)\, \mathbb{P}(X_2 = 1 \,|\, H)\big] \\ &= \mathbb{E}[H^2] > \mathbb{E}[H]^2 = \mathbb{P}(X_1 = 1)\, \mathbb{P}(X_2 = 1).\end{aligned}$$

In this first example H is a random variable. More generally, we have a conditionally iid model if the probability space supports an arbitrary iid sequence $U_1, \ldots, U_d$ of random variables and an arbitrary independent stochastic "object" H, and the random vector $(X_1, \ldots, X_d)$ is defined via $X_k := f(U_k, H)$, $k = 1, \ldots, d$, for some measurable "functional" f. Clearly, this general model is inconvenient because neither the law of U_1, nor the nature of the stochastic object H or the functional f are given explicitly. However, there is a canonical choice for all three entities, which we are going to consider in the sequel. By definition, conditionally iid means that conditioned on the object H the random variables $X_1, \ldots, X_d$ are iid, distributed according to a univariate distribution function F, which may depend on H. A univariate distribution function F is nothing but a non-decreasing, cadlàg function $F : \mathbb{R} \to [0, 1]$ with $\lim_{t \to -\infty} F(t) = 0$ and $\lim_{t \to \infty} F(t) = 1$, see [Billingsley (1995), Theorem 12.4, p. 176]. Therefore, without loss of generality we may assume that $H = \{H_t\}_{t \in \mathbb{R}}$ already is this distribution function itself, i.e. is a random variable in the space of distribution functions – or, in other words, a non-decreasing cadlàg stochastic process with $\lim_{t \to -\infty} H_t = 0$ and $\lim_{t \to \infty} H_t = 1$. In this case, a canonical choice for the law of U_1 is the uniform distribution on $[0,1]$ and the functional f may be chosen as

$$X_k = f(U_k, H) := \inf\{t \in \mathbb{R} \,:\, H_t > U_k\}, \quad k = 1, \ldots, d. \tag{1}$$

Indeed, one verifies that $X_1, \ldots, X_d$ are iid conditioned on $\mathcal{H} := \sigma\big(\{H_t\}_{t \in \mathbb{R}}\big)$, with common univariate distribution function H, since

$$\mathbb{P}(X_1 \leq t_1, \ldots, X_d \leq t_d \,|\, \mathcal{H}) = \mathbb{P}(U_1 \leq H_{t_1}, \ldots, U_d \leq H_{t_d} \,|\, \mathcal{H}) = H_{t_1}\, H_{t_2}\, \cdots\, H_{t_d},$$

for all $t_1, \ldots, t_d \in \mathbb{R}$. Every random vector which is conditionally iid can be constructed like this, i.e. there is a one-to-one relation between such models and random variables in the space of (one-dimensional) distribution functions. Stochastic models of the form (1) are interesting for several reasons:

(a) *Few parameters.* The only model input is the stochastic nature of the process H. In particular, if H is a stochastic process whose law is parameterized by $\boldsymbol{\theta}$, then the resulting multivariate distribution of $(X_1, \ldots, X_d)$ is parameterized by $\boldsymbol{\theta}$, irrespectively of the dimension d. Hence, the number of parameters can be controlled. This can be a striking advantage of such models, for instance compared with models that are parameterized by a correlation matrix, whose number of parameters $d\,(d-1)/2$ grows rapidly with the dimension d.

(b) *Easy simulation.* The model is intuitive to understand and straightforward to simulate, provided one has a convenient simulation engine for the process H at hand. In particular, the efficiency of the simulation algorithm grows only linearly in the dimension, since one only has to draw more iid trigger variables U_k.

(c) *Basis for generalizations.* Independent conditionally iid models of the form (1) can be used as building blocks for more general multi-factor models that overcome the underlying homogeneity assumptions, which are sometimes too restrictive in applications. A generic recipe is straightforward to implement for many families of processes H. In particular, several structurally different processes H can be combined to create quite flexible dependence structures. For this generic procedure required is only (i) a repertoire of well-investigated models of the form (1), and (ii) a careful consideration of what is lost and maintained when two models are combined. Section 3 provides theoretical background related to issue (i) for a large family of processes H and associated multivariate distributions. Details regarding issue (ii) are provided in Section 4.

Now we have a precise probabilistic understanding of the notion "conditionally iid", and a canonical stochastic model. A natural question is whether multivariate distributions of such type can be characterized analytically. For a fixed dimension d, finding convenient, necessary and sufficient, analytical conditions that help to decide whether or not a given multivariate distribution can be constructed by a stochastic model like in (1) is a difficult question, unsolved in general[1]. Nevertheless, for infinite sequences of random variables (i.e. if $d \to \infty$), a seminal theorem of De Finetti provides a satisfactory solution.

Theorem 2.1 (De Finetti (1937))
Consider an infinite sequence of random variables $\{X_k\}_{k\in\mathbb{N}}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The sequence is *exchangeable*, i.e. for each $d \in \mathbb{N}$ the distribution of $(X_1, \ldots, X_d)$ is the same as the distribution of $(X_{\pi(1)}, \ldots, X_{\pi(d)})$ with an arbitrary permutation π on $\{1, \ldots, d\}$, if and only if it is conditionally iid, i.e. there is a sub-σ-algebra $\mathcal{H} \subset \mathcal{F}$ such that for each $d \in \mathbb{N}$ one has

$$\mathbb{P}(X_1 \le t_1, \ldots, X_d \le t_d \,|\, \mathcal{H}) = \prod_{k=1}^{d} \mathbb{P}(X_1 \le t_k \,|\, \mathcal{H}), \quad t_1, \ldots, t_d \in \mathbb{R}.$$

This theorem has first been established by [De Finetti (1931)] for $\{0, 1\}$-valued random variables, and in full generality by [De Finetti (1937)]. Generalizations to more abstract situations and different formulations can be found many times in the literature. Two of the most popular references are [Hewitt, Savage (1955)] and [Ressel (1985)] who deal with random variables taking values in more general spaces than $\mathbb{R}$. One of the essential contributions of De Finetti's Theorem is that exchangeability is a notion which allows for an analytical access to multivariate laws, whereas the notion "conditionally iid" is a priori purely probabilistic.

Now what is the use of De Finetti's Theorem for the scope of the present habilitation thesis? Suppose we start with a given parametric family of multivariate distribution functions in fixed dimension d, and we are interested in determining and describing the subclass of those distributions that one can construct like in (1). If this is possible, we call such a distribution *extendible*, i.e. "extendible" is really just another word for "conditionally iid". De Finetti's Theorem states that extendibility is equivalent to the existence of an infinite, exchangeable sequence $\{X_k\}_{k\in\mathbb{N}}$ on some probability space, such that the first d members of this sequence, arranged as a vector, follow the given multivariate distribution. This explains the nomenclature *extendibility*, because the given d-dimensional random vector can be extended to an infinite exchangeable sequence of random variables – at least in distribution. In particular, De Finetti's Theorem allows us to decide whether the given distribution is extendible in two subsequent steps:

(i) *Exchangeability.* It is necessary that the given d-dimensional distribution is exchangeable. Sometimes it is not too difficult to derive necessary and sufficient, analytical conditions on the parameters for exchangeability. This already simplifies the problem massively, because one only has to consider the exchangeable subclass of the given family of distributions, and the exchangeable subclass can often be analyzed much easier analytically. Providing a simple example, a d-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ is exchangeable

[1] However, such conditions can be found for specific classes of distributions. An example is provided by [Mai, Scherer (2013)] in the case of Marshall-Olkin distributions.

if and only if all components of $\boldsymbol{\mu}$, all diagonal elements of Σ, and all off-diagonal elements of Σ are identical. Hence, analyzing the exchangeable subclass of the normal law means studying a three-parametric family, whereas studying the full normal family requires coping with $d + d\,(d+1)/2$ parameters.

(ii) *Extendibility.* Given a parametric family of d-dimensional exchangeable distributions, one must find convenient analytical conditions on the parameters which guarantee that "the limiting process $d \to \infty$ is possible". In other words, based on the nature of the parameters one must now decide whether it is possible to extend the dimension arbitrarily. This task is typically more difficult than the first task (i), however, in some cases can be accomplished. If one is lucky, then the analytical solution on the parameter space can be converted back into probability theory and provides insight into the stochastic nature of the latent factor process H.

Section 3 characterizes the extendible subclasses for several families of multivariate exponential distributions. Moreover, other parametric families of multivariate distributions, for which the extendible subclass and the stochastic nature of H are known, comprise elliptical distributions and Archimedean copulas, for references and details the interested reader is referred to the respective chapters in [Mai, Scherer (2012)].

3 Multivariate exponential distributions

The present section deals with multivariate exponential distributions. It is explored which stochastic properties the process H must have in order for the random vector $(X_1, \ldots, X_d)$ from construction (1) to be exponential – a precise definition of this notion is given below. Firstly, a general introduction into multivariate exponential distributions is presented in Subsection 3.1. Secondly, the major theoretical findings are summarized in Subsection 3.2.

Before we start, let us slightly reformulate the stochastic model (1) so that it more conveniently suits the setup of exponential distributions. Clearly, since exponential distributions always have non-negative components, $H_t = 0$ for all $t \leq 0$. Therefore, without loss of generality we may assume that $H = \{H_t\}_{t \geq 0}$ is indexed by $t \in [0, \infty)$. Moreover, applying the substitution $h = -\log(1 - F)$ it trivially holds true that

$$
\begin{aligned}
&\big\{\text{distribution functions } F : [0,\infty) \to [0,1] \text{ with } F(0) = 0\big\} \\
&\quad = \big\{t \mapsto 1 - \exp(-h(t)) \,\big|\, h : [0,\infty) \to [0,\infty] \text{ non-decreasing, cadlàg} \\
&\qquad\qquad \text{with } h(0) = 0 \text{ and } \lim_{t \to \infty} h(t) = \infty\big\}.
\end{aligned}
$$

One can therefore drop the boundedness assumption on the process H and rewrite the canonical construction (1) as

$$X_k := \inf\{t \geq 0 \,:\, H_t > \epsilon_k\}, \quad k = 1, \ldots, d, \tag{2}$$

where the $\epsilon_k = -\log(1-U_k)$, $k = 1, \ldots, d$, are now iid exponential random variables with unit mean, and $H = \{H_t\}_{t \geq 0}$ is a non-decreasing, cadlàg process with $H_0 = 0$ and $\lim_{t\to\infty} H_t = \infty$. This canonical probability space is visualized in Figure 1.

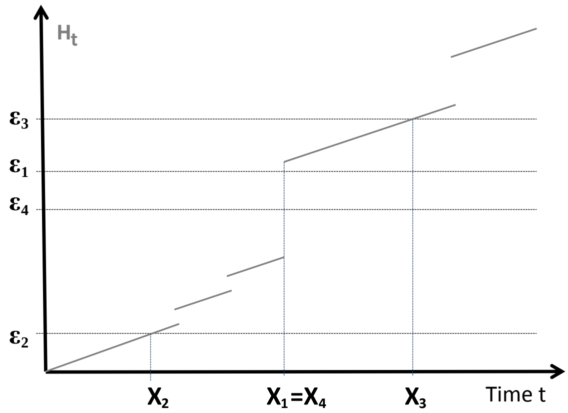


Fig. 1 Illustration of one simulation of the canonical construction (2) in dimension $d = 4$. One observes that the process $H = \{H_t\}_{t\geq 0}$ in this particular illustration has jumps, and therefore there is a positive probability that two components take the identical value. This does not happen if H is a continuous process.

3.1 A primer on multivariate exponential distributions

First of all, we must agree on a suitable definition of multivariate exponential distributions, because the one-dimensional exponential law can be lifted to the multivariate case in several ways. The seminal work of [Esary, Marshall (1974)] is briefly recalled, in which several classes of multivariate exponential distributions are considered that are obtained by lifting different univariate properties to the multivariate case. A random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ has an exponential law with rate parameter $\lambda > 0$, denoted $\mathcal{E}(\lambda)$, if its distribution function is given by

$$F(t) := \mathbb{P}(X \leq t) = \left(1 - e^{-\lambda\, t}\right) 1_{\{t>0\}}, \quad t \in \mathbb{R}.$$

The following two properties characterize the univariate exponential distribution.

(a) *Lack-of-memory:* A random variable X with support $[0,\infty)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is exponential if and only if $\mathbb{P}(X > t + s \,|\, X > s) = \mathbb{P}(X > t)$ for all $t, s \geq 0$.

(b) *Min-stability:* If two independent random variables X and Y are exponential with rates λ_X, λ_Y, then the minimum $\min\{X, Y\}$ is exponential with rate $\lambda_X + \lambda_Y$. Moreover, the parametric family of exponential distributions is the only parametric family of distributions with this closure property.

Considering the survival function $\bar{F}(t) = \exp(-\lambda\, t)$, $t \geq 0$, of $\mathcal{E}(\lambda)$, both properties (a) and (b) follow from the fact that the exponential function is essentially the only function f satisfying Cauchy's equation $f(x{+}y) = f(x)\, f(y)$, $x, y > 0$, see [Billingsley (1995), Appendix A20, p. 540]. Motivated from these two properties [Esary, Marshall (1974)] consider three different families of multivariate exponential distributions, which we adopt. We say that a random vector $(X_1, \ldots, X_d)$ with support $[0, \infty)^d$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ has a ...

(a) ... *Marshall–Olkin distribution (MO)*, if the multivariate lack-of-memory property

$$\mathbb{P}(X_{i_1} > t_1 + s, \ldots, X_{i_k} > t_k + s \,|\, X_{i_1} > s, \ldots, X_{i_k} > s) = \mathbb{P}(X_{i_1} > t_1, \ldots, X_{i_k} > t_k)$$

is satisfied for all $1 \leq i_1 < i_2 < \ldots < i_k \leq d$, $s, t_1, \ldots, t_k \geq 0$.

(b1) ... *min-stable multivariate exponential distribution (MSMVE)*, if each minimum $\min\{c_1\, X_{i_1}, \ldots, c_k\, X_{i_k}\}$ over scaled components is (univariate) exponential, for all $1 \leq i_1 < i_2 < \ldots < i_k \leq d$, $c_1, \ldots, c_k > 0$.

(b2) ... *distribution with exponential minima (EM)*, if each minimum $\min\{X_{i_1}, \ldots, X_{i_k}\}$ over components is (univariate) exponential, for all $1 \leq i_1 < i_2 < \ldots < i_k \leq d$.

It is shown in [Esary, Marshall (1974)] that MO $\subsetneq$ MSMVE $\subsetneq$ EM. Moreover, the family MO is finite-parametric with $2^d - 1$ parameters and named after the seminal reference [Marshall, Olkin (1967)], where it first appeared. The family MO has been studied intensively in the author's dissertation [Mai (2010)], and parameter estimation for the extendible subfamily is discussed in Subsection 5.1. Generally speaking, the Marshall–Olkin distribution is a paradigm example for a dependence model whose complexity grows exponentially in the dimension. For instance, there are many articles which treat the parameter estimation for bivariate Marshall–Olkin distributions, but their underlying idea is difficult to extend to larger dimensions. Already the simulation of the Marshall–Olkin distribution is a time-consuming task in large dimensions, not to say impossible on a standard PC. In order to circumvent this difficulty, the article [A3] derives a simulation algorithm for exchangeable (but not necessarily extendible) Marshall–Olkin distributions, and shows how to extend it also to hierarchical MO laws. The idea is to exploit both the exchangeability assumption (providing sufficient symmetry to join several cases by combinatorial considerations) as well as the lack-of-memory property of the Marshall–Olkin distribution (to implement the algorithm recursively). The algorithm is recursive and the runtime is random itself. However, a worst-case estimate for the runtime is provided, which is shown to outperform the classical simulation algorithm based on the canonical construction of the original reference [Marshall, Olkin (1967)] by far. For detailed information and pseudo-code of the sampling engine the interested reader is referred to [A3].

The family MSMVE is infinite-parametric. It is most convenient to study members of the family MSMVE from their (multivariate) survival functions, about which a lot is known. For instance, [Joe (1997), Theorem 6.2, p. 174] shows that a function $\bar{F}$ in d variables is the survival function of an MSMVE law if and only if it can be written as

$$\bar{F}(t_1, \ldots, t_d) = C\big(e^{-\lambda_1 t_1}, \ldots, e^{-\lambda_d t_d}\big), \quad t_1, \ldots, t_d \geq 0,$$

for some exponential rates $\lambda_1, \ldots, \lambda_d > 0$ and a so-called *extreme-value copula* C. The latter are distribution functions on $[0, 1]^d$ with uniformly distributed marginal laws on $[0, 1]$ satisfying the *extreme-value property* $C(u_1, \ldots, u_d)^t = C(u_1^t, \ldots, u_d^t)$, $u_1, \ldots, u_d \in [0, 1]$, $t \geq 0$. As the nomenclature suggests, these distributions play a fundamental role in extreme-value theory. For instance, the family of distribution functions which is obtained when plugging univariate extreme-value distribution functions into an extreme-value copula coincides with the family of multivariate extreme-value distributions. Without going into details, this implies that studying MSMVEs is equivalent to studying the dependence structure between rare events. One of the major findings in multivariate extreme-value theory, at least known since [De Haan, Resnick (1977)] but many times re-discovered and re-formulated since then, is a one-to-one relationship between d-dimensional MSMVEs and certain measures on a subspace of $\mathbb{R}^d$, which is somehow comparable with the one-to-one relationship between infinitely divisible distributions and their associated Lévy measures. A quite recent, purely analytical derivation of this result can be retrieved from [Ressel (2013)]. In terms of this measure representation the subfamily MO $\subset$ MSMVE is always given by a certain discrete measure with at most $2^d - 1$ atoms, which is explicitly stated in [A1]. Given the importance of the family MSMVE in multivariate extreme-value theory, it is of paramount interest to determine its extendible subfamily. Indeed, the upcoming Subsection 3.2 outlines how to construct conditionally iid MSMVEs from certain stochastic processes, and later on in Subsection 4.2 multi-factor MSMVEs are constructed from the conditionally iid building blocks.

Given the findings from the author's dissertation [Mai (2010)] on the subfamily MO $\subset$ MSMVE, one might hope that some results easily extend from MO to MSMVE in an obvious manner. Unfortunately, there are a couple of difficulties preventing this strategy from being straightforward – predominantly the fact that MO is finite-parametric and MSMVE is not. One way of thinking, which turned out not to work, is illustrated briefly, because it is educational and, luckily, provides a result of independent interest in dimension $d = 2$. In terms of the aforementioned measure representations for MSMVE, one could have hope that each representing measure of an MSMVE law can be constructed as a convex mixture of representing measures from the class MO. Since the latter have finite support, finite convex combinations still remain finite measures, so representation measures with infinite support cannot be obtained without considering closure properties of such a construction. One hope is that an arbitrary discrete representing measure can be attained by such convex mixtures, and a second hope is that discrete measures are dense in the set of all measures in some sense. Although there is a positive answer to the second hope at least in dimension $d = 2$, the article [A2] shows that already in the bivariate case $d = 2$ the first hope is too ambitious. It is shown that an arbitrary discrete representing measure of a bivariate MSMVE can be attained as the convex combination of certain measures with at most two atoms. Unfortunately, the discrete measures associated with Marshall–Olkin distributions turn out not to be the relevant building blocks in this construction. Rather, another two-parametric family of distributions, termed BC_2 in [A2], is found to be essential, which is a result of independent interest. Unfortunately, this finding makes explicit use of a construction idea that only works in dimension $d = 2$.

3.2 Extendible exponential distributions

In the author's dissertation [Mai (2010)] it has already been shown that if the stochastic process $H = \{H_t\}_{t \geq 0}$ in construction (2) is a (possibly killed) Lévy subordinator[2], then the random vector $(X_1, \ldots, X_d)$ has a Marshall–Olkin distribution. Even more, every extendible Marshall–Olkin distribution can be obtained by this construction. This result is generalized to the superclasses MSMVE and EM in Theorem 3.3 below. In order to formulate it, required is the notion of stochastic processes which are *infinitely divisible with respect to time (IDT)*. We distinguish between processes that are *strong IDT* and *weak IDT*, the former have first been studied in [Mansuy (2005), Es-Sebaiy, Ouknine (2008)] and the latter in [Hakassou, Ouknine (2012)]. In the following definition, $\stackrel{d}{=}$ means equality in distribution. We recall that $(X_1, \ldots, X_d) \stackrel{d}{=} (Y_1, \ldots, Y_d)$ means $\mathbb{E}[f(X_1, \ldots, X_d)] = \mathbb{E}[f(Y_1, \ldots, Y_d)]$ for all bounded, continuous functions $f : \mathbb{R}^d \to \mathbb{R}$, where the expectation values $\mathbb{E}$ are taken on the respective probability spaces of $(X_1, \ldots, X_d)$ and $(Y_1, \ldots, Y_d)$, which might be different. Equality in law for two stochastic processes $X = \{X_t\}_{t \geq 0}$ and $Y = \{Y_t\}_{t \geq 0}$ means that $(X_{t_1}, \ldots, X_{t_d}) \stackrel{d}{=} (Y_{t_1}, \ldots, Y_{t_d})$ for arbitrary $d \in \mathbb{N}$ and $t_1, t_2, \ldots, t_d \geq 0$.

Definition 3.1 (Strong and weak IDT process)
A stochastic process $H = \{H_t\}_{t \geq 0}$ is called *weak IDT* if for each $n \in \mathbb{N}$, each $t \geq 0$, and independent copies $H^{(1)}, \ldots, H^{(n)}$ of H it holds that

$$H_t \stackrel{d}{=} H^{(1)}_{t/n} + H^{(2)}_{t/n} + \ldots + H^{(n)}_{t/n}. \tag{3}$$

Furthermore, if the equality (3) is even satisfied by whole paths, i.e. if for each $n \in \mathbb{N}$ and independent copies $H^{(1)}, \ldots, H^{(n)}$ of H we have

$$\{H_t\}_{t \geq 0} \stackrel{d}{=} \big\{H^{(1)}_{t/n} + H^{(2)}_{t/n} + \ldots + H^{(n)}_{t/n}\big\}_{t \geq 0},$$

then H is called *strong IDT*.

Every strong IDT process is also weak IDT by definition, but the converse needs not hold. For an example of a weak IDT process which is not strong IDT we refer to [A4, Example 5.1]. Every Lévy process is strong IDT, but the converse needs not hold. For instance, if $S = \{S_t\}_{t \geq 0}$ is a non-trivial Lévy subordinator and $a > b > 0$, then

[2]For background on these processes the reader is referred to the textbook [Applebaum (2004)].

the stochastic process $\{S_{a\,t} + S_{b\,t}\}_{t \geq 0}$ is strong IDT by Lemma 3.2 below, but not a Lévy subordinator. For further examples we refer the interested reader to [A4]. For a cadlàg, non-decreasing stochastic process $\{H_t\}_{t \geq 0}$ with $H_0 = 0$ the Laplace transforms $x \mapsto \mathbb{E}[\exp(-x\,H_t)]$, $x \geq 0$, are well-defined for each $t \geq 0$. Using standard arguments from the theory of infinite divisibility, such as [Sato (1999), Theorem 7.10, p. 35], it can be shown that a non-decreasing process $H = \{H_t\}_{t \geq 0}$ with $H_0 = 0$ is weak IDT if and only if there exists a function $\Psi : [0, \infty) \to [0, \infty)$ such that

$$\mathbb{E}\Big[e^{-x\,H_t}\Big] = e^{-t\,\Psi(x)}, \quad x, t \geq 0.$$

In such a situation, it is well-known that Ψ must be a so-called *Bernstein function*, i.e. $\Psi \in C^{\infty}(0, \infty)$, $(-1)^{k+1}\,\Psi^{(k)}(x) \geq 0$, $x > 0$, $k \in \mathbb{N}$, with $\Psi(0) = 0$ and a possible jump at zero. There exists vast literature on the study of such functions, because they play a dominant role at numerous places in Probability Theory and Analysis. For instance, the textbooks [Berg et al. (1984), Schilling et al. (2010)] contain extensive analytical studies of Bernstein functions. Many results in the present thesis, e.g. the proof of Theorem 3.3 below, rely on the analytical treatment of infinite divisibility via Bernstein functions. For more information on the latter the interested reader is also referred to Subsection 5.2, where a convenient Laplace inversion algorithm for densities of infinitely divisible laws associated with certain Bernstein functions is developed.

In order to get a feeling for the notion of IDT processes, the following (new) lemma points out a distinctive closure property of strong IDT processes, which is neither shared by weak IDT processes nor by Lévy subordinators. It is required later on in Lemma 4.4.

Lemma 3.2 (Strong IDT processes form a cone in time and space)
Let $H = \{H_t\}_{t \geq 0}$ be a strong IDT process, $m \in \mathbb{N}$, and $a_j, b_j > 0$ constants, $j = 1, \ldots, m$. Then the process $\big\{a_1\,H_{b_1\,t} + \ldots + a_m\,H_{b_m\,t}\big\}_{t \geq 0}$ is also strong IDT.

Proof
Consider $n \in \mathbb{N}$ independent copies of the stochastic process H, denoted $H^{(i)}$, $i = 1, \ldots, n$. The following observation is needed:

($*$) If $X = \{X_t\}_{t \geq 0}$ and $Y = \{Y_t\}_{t \geq 0}$ are arbitrary stochastic processes with $\{X_t\}_{t \geq 0} \stackrel{d}{=} \{Y_t\}_{t \geq 0}$, it follows that $\big\{a_1\,X_{b_1\,t} + \ldots + a_m\,X_{b_m\,t}\big\}_{t \geq 0} \stackrel{d}{=} \big\{a_1\,Y_{b_1\,t} + \ldots + a_m\,Y_{b_m\,t}\big\}_{t \geq 0}$, as will be shown in the sequel: fixing some $d \in \mathbb{N}$ and $t_1, t_2, \ldots, t_d \geq 0$, the random vectors $(X_{b_1\,t_1}, \ldots, X_{b_1\,t_d}, X_{b_2\,t_1}, \ldots, X_{b_m\,t_d})$ and $(Y_{b_1\,t_1}, \ldots, Y_{b_1\,t_d}, Y_{b_2\,t_1}, \ldots, Y_{b_m\,t_d})$ have the same distribution by assumption. Hence,

$$\Big(\sum_{j=1}^{m} a_j\,X_{b_j\,t_1}, \ldots, \sum_{j=1}^{m} a_j\,X_{b_j\,t_d}\Big) \stackrel{d}{=} \Big(\sum_{j=1}^{m} a_j\,Y_{b_j\,t_1}, \ldots, \sum_{j=1}^{m} a_j\,Y_{b_j\,t_d}\Big),$$

since these random vectors arise by applying the same measurable functional to $(X_{b_1\,t_1}, \ldots, X_{b_1\,t_d}, X_{b_2\,t_1}, \ldots, X_{b_m\,t_d})$, respectively $(Y_{b_1\,t_1}, \ldots, Y_{b_1\,t_d}, Y_{b_2\,t_1}, \ldots, Y_{b_m\,t_d})$. This proves the claim.

Defining $X_t := H_t$ and $Y_t := \sum_{i=1}^{n} H^{(i)}_{t/n}$, $t \geq 0$, it follows from the strong IDT property that $\{X_t\}_{t\geq 0} \stackrel{d}{=} \{Y_t\}_{t\geq 0}$. Hence,

$$\Big\{ \sum_{i=1}^{n} \sum_{j=1}^{m} a_j \, H^{(i)}_{b_j\, t/n} \Big\}_{t\geq 0} = \Big\{ \sum_{j=1}^{m} a_j \, \sum_{i=1}^{n} H^{(i)}_{b_j\, t/n} \Big\}_{t\geq 0} \stackrel{d}{=} \Big\{ \sum_{j=1}^{m} a_j \, H_{b_j\, t} \Big\}_{t\geq 0},$$

where the equality in distribution follows from $(*)$ above and the first equality is trivial. On the left hand side of the last equation one sees n independent copies of the process $\big\{a_1 \, H_{b_1\, t} + \ldots + a_m \, H_{b_m\, t}\big\}_{t\geq 0}$, time-changed with $t \mapsto t/n$. Hence, the strong IDT property of $\big\{a_1 \, H_{b_1\, t} + \ldots + a_m \, H_{b_m\, t}\big\}_{t\geq 0}$ is established. □

If the stochastic process H in Lemma 3.2 is only weak IDT, then the stochastic process $\big\{a_1 \, H_{b_1\, t} + \ldots + a_m \, H_{b_m\, t}\big\}_{t\geq 0}$ needs not be weak IDT in general, an example is provided in the Appendix. Also this closure property does clearly not hold for the subfamily of Lévy subordinators. So it is a property distinctive to the family of strong IDT processes. The following theorem – together with Lemma 4.4 to be derived in Subsection 4.2 below – explains the importance of IDT processes in the context of multivariate exponential laws.

Theorem 3.3 (Extendible multivariate exponential laws)

(a) If the stochastic process $H = \{H_t\}_{t\geq 0}$ is strong IDT, then the random vector $(X_1, \ldots, X_d)$ in construction (2) has a min-stable multivariate exponential distribution. Moreover, every extendible member of the family MSMVE can be obtained by this construction.

(b) If the stochastic process $H = \{H_t\}_{t\geq 0}$ is weak IDT, then the random vector $(X_1, \ldots, X_d)$ in construction (2) has a distribution with exponential minima. Moreover, every extendible member of the family EM can be obtained by this construction.

Proof

These results are the main contribution of [A4], where the proofs can be found. □

Theorem 3.3 implies that IDT processes play a fundamental role in the context of multivariate exponential distributions. This is surprising, since IDT processes so far have found only little attention in the academic literature, but MSMVEs are very well studied because of their connection to extreme-value theory. The only articles dealing with IDT processes that the author is aware of are [Mansuy (2005), Es-Sebaiy, Ouknine (2008), Hakassou, Ouknine (2012)], and they merely present a couple of examples but no further study such as, e.g., a canonical stochastic construction or a convenient analytical description of their stochastic nature, which is an interesting topic for further research. The results of the author's dissertation and of Theorem 3.3 are summarized in Figure 2.

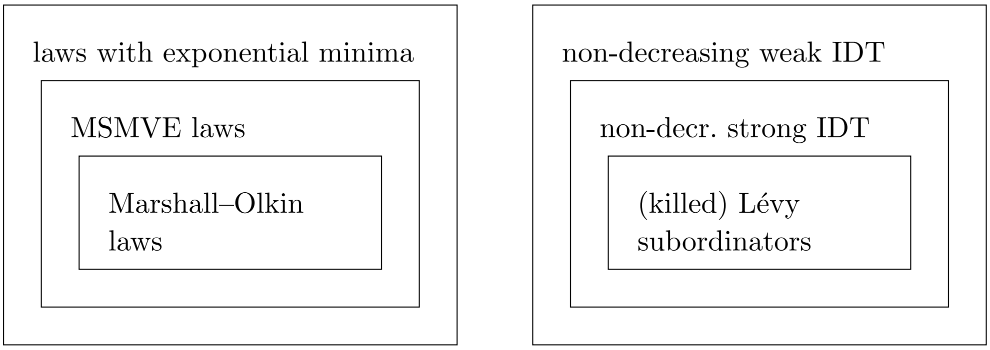


Fig. 2 Subfamilies of EM and weak IDT processes. Whenever a member of a class of processes on the right hand side is applied in construction (2), the result is a vector $(X_1, \ldots, X_d)$ with corresponding distribution from the left hand side of this Venn diagram.

4 A generic recipe for multi-factor models

The previous sections dealt with determining and describing the extendible subclass of a family of multivariate distributions. The author hopes at this point that the reader is convinced that such a study contributes a certain amount of inner-mathematical value, because coherences between analytical concepts such as Bernstein functions on the one hand, and probabilistic concepts such as extendibility and multivariate exponential laws on the other hand, are revealed. Nevertheless, the application-oriented reader might ask whether the conditionally iid nature of extendible distributions is of much practical value at all, because many real world phenomena exhibit a much more complicated dependence structure than that. Firstly, later on in Section 5.3 one specific field of application in Mathematical Finance is indicated, where the conditionally iid nature of these models actually offers an appealing trade-off between realism and practical viability. Secondly, the present section shows how multi-factor models – which go far beyond the cosmos of the conditionally iid setup – can be constructed conveniently from extendible building blocks. By “conveniently” it is meant that all the “hard math” is actually done, when the conditionally iid models are well-developed, because the approach works basically like a tool box that allows to combine extendible building blocks in a very simple way. Compared with the conditionally iid setup, one looses a certain level of analytical viability. However, and this is one decisive point, the presented parametric models are easy to simulate whenever the conditionally iid building blocks can be simulated, and they allow for the number of parameters to be controlled at one’s personal taste. We call multivariate distributions that are constructed like this *h-extendible*, and the generic construction idea is outlined in detail in [A6]. In the sequel, the basic idea is only briefly sketched, and rather two explicit examples are discussed in greater detail (namely h-extendible Archimedean copulas and h-extendible MSMVEs).

The motivation for the notion *h-extendibility* is the synthesis of two desirable properties: (i) the dependence structure is induced by multiple factors which affect the components of the resulting random vector in a hierarchical manner (i.e. the "h" in "h-extendible" stands for "hierarchical"), and (ii) the structure is "dimension-free" in the sense that components can be added or removed from the system without affecting the overall dependence structure, as is the case for extendible models. The following definition aims at formalizing this intuitive idea.

Definition 4.1 (H-extendibility)
A d-dimensional random vector $(X_1, \ldots, X_d)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *h-extendible* with $n \in \{1, \ldots, d\}$ levels of hierarchy if for $n \geq 2$ there exists a σ-algebra $\mathcal{H} \subset \mathcal{F}$ and a partition $d_1 + \ldots + d_J = d$ such that conditioned on $\mathcal{H}$: (a) the random vector $(X_1, \ldots, X_d)$ splits into J independent subvectors according to this partition, (b) each subvector is h-extendible with at most $n-1$ levels of hierarchy, and (c) at least one subvector has $n-1$ levels of hierarchy. For $n = 1$, i.e. at the end of the recursion, h-extendibility with one level of hierarchy corresponds to the usual definition of extendibility.

Definition 4.1 is recursive, but it can alternatively be reformulated iteratively. However, the notation is more involved in this case, see [A6, Remark 2.4]. Explaining the iterative definition in simple terms, an h-extendible random vector $(X_1, \ldots, X_d)$ with n levels of hierarchy has to be thought of as follows: conditioned on a σ-algebra $\mathcal{H}_1$, the random vector splits into J independent subvectors. The σ-algebra $\mathcal{H}_1$ might be thought of as being generated by a stochastic process $H^{(1)} = \{H_t^{(1)}\}$ which affects all groups in the same way. On the second level, there is a σ-algebra $\mathcal{H}_2$ conditioned on which each subgroup again splits into independent subsubgroups. The σ-algebra $\mathcal{H}_2$ can be thought of as being generated by $H^{(1)}$ and J group-specific and independent stochastic factors $H^{(2,1)}, \ldots, H^{(2,J)}$, where $H^{(2,j)} = \{H_t^{(2,j)}\}$, $j = 1, \ldots, J$, affects all components of subgroup $j \in \{1, \ldots, J\}$ in the same way. Subdividing the subsubgroups further, this procedure ends at level n, so that one obtains an increasing sequence of σ-algebras $\mathcal{H}_1 \subsetneq \mathcal{H}_2 \subsetneq \ldots \subsetneq \mathcal{H}_n \subsetneq \mathcal{F}$. The factors inducing the dependence between the components are arranged hierarchically in the sense that the factors entering the construction at level k are included in all σ-algebras $\mathcal{H}_i$ with $i \geq k$.

One might ask what is the use of Definition 4.1. The main motivation is that this definition serves as an umbrella for many hierarchical factor models that can be found in the literature, and therefore provides formalism to compare these stochastic models. In the sequel, some examples of h-extendible structures are introduced.

4.1 H-extendible Archimedean copulas

One of the most popular families of multivariate distribution functions is the family of *Archimedean copulas*. A multivariate distribution function $C_\varphi : [0,1]^d \to [0,1]$ is an

Archimedean copula if it has the functional form

$$C_\varphi(u_1,\ldots,u_d)=\varphi\big(\varphi^{-1}(u_1)+\ldots+\varphi^{-1}(u_d)\big),$$

for a non-increasing function $\varphi:[0,\infty)\to[0,1]$ with $\varphi(0)=1$ and $\lim_{x\to\infty}\varphi(x)=0$, called *(Archimedean) generator*. A result of [Malov (1998)] shows that a necessary and sufficient condition on the generator φ for C_φ to be a proper distribution function is *d-monotonicity*, see also [McNeil, Nešlehová (2009)] who provide a probabilistic interpretation for this notion and derive a probabilistic construction for Archimedean copulas. For the precise definition of d-monotonicity we refer the interested reader to [Malov (1998), McNeil, Nešlehová (2009)]. We denote by Φ_d the set of all d-monotone Archimedean generators. It is known that $\Phi_2\supsetneq\Phi_3\supsetneq\ldots\supsetneq\Phi_\infty$, where Φ_∞ denotes the set of all *completely monotone* generators, i.e. all φ with $\varphi(0)=1$, $\varphi\in C^\infty(0,\infty)$, continuous at zero, and $(-1)^k\,\varphi^{(k)}(x)\geq 0$ for all $x>0$, $k\in\mathbb{N}_0$. By the seminal Bernstein Theorem the set Φ_∞ coincides with the set of Laplace transforms of probability measures on $(0,\infty)$. The original reference is [Bernstein (1929)], see also [Schilling et al. (2010), Theorem 1.4, p. 3]. An Archimedean copula C_φ is extendible if and only if $\varphi\in\Phi_\infty$, and in this case a random vector $(X_1,\ldots,X_d)\sim C_\varphi$ can be constructed canonically as in construction (1) with

$$H_t=\begin{cases}0, & t\leq 0\\ \exp(-M\,\varphi^{-1}(t)), & t\in(0,1)\\ 1, & t\geq 1\end{cases}\;, \tag{4}$$

where M is a positive random variable with Laplace transform $\varphi\in\Phi_\infty$. Since the stochastic process $H=\{H_t\}_{t\in\mathbb{R}}$ is quite trivial in this case, the infimum in the canonical construction (1) can be computed explicitly, yielding the definition $X_k:=\varphi(-\log(U_k)/M)$, $k=1,\ldots,d$. This provides a simple stochastic model which is typically found in the literature, for the first time probably in [Marshall, Olkin (1988)]. The σ-algebra $\mathcal{H}=\sigma(\{H_t\}_{t\in\mathbb{R}})$ is generated by the random variable M, conditioned on which the components of the resulting random vector $(X_1,\ldots,X_d)$ in (1) are iid.

To overcome the exchangeability of Archimedean copulas, the notion of hierarchical Archimedean copulas is introduced in [Joe (1993), Joe, Hu (1996)]. Given a partition $d=d_1+\ldots+d_J$ of the dimension d and a vector $\boldsymbol{u}:=(u_1,\ldots,u_d)\in[0,1]^d$, we introduce the notation $\boldsymbol{u}_j:=(u_{d_1+\ldots+d_{j-1}+1},\ldots,u_{d_1+\ldots+d_{j-1}+d_j})$, $j=1,\ldots,J$, so that $\boldsymbol{u}=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_J)$. With given Archimedean generators $\varphi_0,\varphi_1,\ldots,\varphi_J$, if the function

$$C(\boldsymbol{u}):=C_{\varphi_0}\big(C_{\varphi_1}(\boldsymbol{u}_1),\ldots,C_{\varphi_J}(\boldsymbol{u}_J)\big),\quad \boldsymbol{u}\in[0,1]^d, \tag{5}$$

is a proper distribution function, it is called a hierarchical Archimedean copula (with two levels of hierarchy). Assume for a minute that C is a distribution function and consider a random vector $\boldsymbol{X}=(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_J)\sim C$. The nice thing about such a dependence structure is that within each group $j\in\{1,\ldots,J\}$, the random vector $\boldsymbol{X}_j\sim C_{\varphi_j}$ has a well-known and well-understood Archimedean copula. Moreover, if the indices $1 \leq i_1 < \ldots < i_k \leq d$ are chosen from k distinct groups, the random vector $(X_{i_1},\ldots,X_{i_k}) \sim C_{\varphi_0}$ also has an Archimedean copula. Therefore, such a non-exchangeable dependence model is appealing in the sense that a good understanding can be deduced from simpler, exchangeable building blocks. Unfortunately, the involved Archimedean generators $\varphi_0,\varphi_1,\ldots,\varphi_J$ need to satisfy compatibility conditions in order for the function C in (5) to define a proper distribution function. In particular, it is not sufficient that $\varphi_j \in \Phi_\infty$ for all $j=0,1,\ldots,J$. The only known sufficient condition for compatibility is that the functions $\varphi_0,\varphi_1,\ldots,\varphi_J$ are all in Φ_∞ and, additionally, that the first derivatives of the functions $\varphi_0^{-1}\circ\varphi_j$, $j=1,\ldots,J$, are in Φ_∞ as well, see [McNeil (2008)]. However, it is not straightforward to find two Laplace transforms $\varphi_0,\varphi_1 \in \Phi_\infty$ such that $(\varphi_0^{-1}\circ\varphi_1)' \in \Phi_\infty$ as well. Even though some examples of compatible generators have been found in [McNeil (2008), Hofert (2008)], a comprehensive understanding of the compatibility condition was an open problem for quite a while. In particular, it has been difficult to find examples of two different parametric families of Laplace transforms to satisfy the compatibility condition. This gap is filled by [A5], who derive the set of all compatible Laplace transforms of positive random variables in a convenient form. As a byproduct of the proof, one also obtains a convenient stochastic model for random vectors with distribution function (5), based on Lévy subordinators. These results are summarized in the sequel.

Theorem 4.2 (On the compatibility of Archimedean generators)
Let $\varphi_0,\varphi_1 \in \Phi_\infty$. Then $(\varphi_0^{-1}\circ\varphi_1)' \in \Phi_\infty$ if and only if

$$\varphi_1(x) = \varphi_0\Big(b\,x + \int_0^\infty \big(1-e^{-x\,t}\big)\,\nu(dt)\Big), \quad x \geq 0,$$

where $b \geq 0$, ν is a measure on $(0,\infty)$ satisfying $\int_0^\infty \min\{1,t\}\,\nu(dt) < \infty$, and either $b>0$ or $\nu\big((0,1)\big)=\infty$, or both.

Proof
See [A5, Theorem 2.1]. □

Given an arbitrary Archimedean generator $\varphi_0 \in \Phi_\infty$, Theorem 4.2 parameterizes the set of all generators $\varphi_1 \in \Phi_\infty$ compatible with φ_0 in terms of a constant b and a measure ν. The key to Theorem 4.2 is the fact that the set of functions whose first derivative is in Φ_∞ coincides with the family of Bernstein functions, which appear at various places in this thesis. The Lévy–Khinchin formula for Bernstein functions then provides the given measure representation. The statement of Theorem 4.2 can also be stated as follows: $(\varphi_0^{-1}\circ\varphi_1)' \in \Phi_\infty$ if and only if $\varphi_1 = \varphi_0 \circ \Psi_1$ for a continuous Bernstein function Ψ_1 with $\lim_{x\to\infty}\Psi_1(x)=\infty$. In probabilistic terms, this means that the Laplace transform φ_1 must correspond to a random variable of the form S_M, where $S=\{S_t\}_{t\geq 0}$ is a Lévy subordinator with associated Bernstein function Ψ_1 and M an independent, positive random variable with Laplace transform φ_0. The technical extra conditions given in Theorem 4.2 merely prevent the Lévy subordinator from being of compound Poisson type, which guarantees that S_M is strictly positive with probability one. The connection to Lévy subordinators leads to the following theorem.

Theorem 4.3 (Construction of h-extendible Archimedean copulas)
Let $\varphi_0 \in \Phi_\infty$ and $\Psi_1, \ldots, \Psi_J$ be continuous, unbounded Bernstein functions, and consider a partition $d = d_1 + \ldots + d_J$ of the integer d. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting d iid exponential random variables $\epsilon_1, \ldots, \epsilon_d$ with unit mean, J independent Lévy subordinators $S^{(j)} = \{S_t^{(j)}\}_{t \geq 0}$, $j = 1, \ldots, J$, with associated Bernstein functions Ψ_j, and an independent random variable M with Laplace transform φ_0. The random vector

$$(X_1, \ldots, X_d) := \Bigg(\varphi_0 \circ \Psi_1\Big(\frac{\epsilon_1}{S_M^{(1)}}\Big), \ldots, \varphi_0 \circ \Psi_1\Big(\frac{\epsilon_{d_1}}{S_M^{(1)}}\Big), \varphi_0 \circ \Psi_2\Big(\frac{\epsilon_{d_1+1}}{S_M^{(2)}}\Big), \ldots$$
$$\ldots\ldots, \varphi_0 \circ \Psi_J\Big(\frac{\epsilon_{d_1+\ldots+d_{J-1}+1}}{S_M^{(J)}}\Big), \ldots, \varphi_0 \circ \Psi_J\Big(\frac{\epsilon_{d_1+\ldots+d_J}}{S_M^{(J)}}\Big) \Bigg)$$

has distribution function (5) with $\varphi_j := \varphi_0 \circ \Psi_j$, $j = 1, \ldots, J$.

Proof
See [A5, Theorem 3.1]. □

Together with Theorem 4.2 this result provides a canonical construction for all hierarchical Archimedean copulas of the form (5) whose completely monotone generators satisfy the known compatibility condition. A convenient simulation algorithm is immediate from this construction and the interested reader is referred to [Mai, Scherer (2012), Chapter 2, p. 91–93] for details. Clearly, the random vector constructed in Theorem 4.3 is h-extendible with two levels, the respective σ-algebras being $\mathcal{H}_1 := \sigma(M) \subsetneq \mathcal{H}_2 := \sigma(M, S^{(1)}, \ldots, S^{(J)}) \subsetneq \mathcal{F}$. Deeper levels of hierarchy can be constructed easily in an analogous manner, see [A6] for details.

4.2 H-extendible exponential distributions

We have seen in Subsection 3.2 how to construct extendible exponential distributions from weak IDT processes. This construction can be extended to overcome exchangeability, as the following (new) lemma shows.

Lemma 4.4 (Multi-factor exponential distributions)
Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting $n+1 \in \mathbb{N}$ independent, non-decreasing weak IDT processes $\tilde{H}^{(i)} = \{\tilde{H}_t^{(i)}\}_{t \geq 0}$, $i = 0, \ldots, n$, and an independent iid sequence $\epsilon_1, \ldots, \epsilon_d$ of exponential random variables with unit mean. Moreover, let $A = (a_{i,j}) \in \mathbb{R}^{d \times (n+1)}$ be an arbitrary matrix with non-negative entries, and at least one positive entry per row. We define the vector-valued stochastic process

$$\boldsymbol{H}_t = \begin{pmatrix} H_t^{(1)} \\ H_t^{(2)} \\ \vdots \\ H_t^{(d)} \end{pmatrix} := A \cdot \begin{pmatrix} \tilde{H}_t^{(0)} \\ \tilde{H}_t^{(1)} \\ \vdots \\ \tilde{H}_t^{(n)} \end{pmatrix} = \begin{pmatrix} a_{1,0}\,\tilde{H}_t^{(0)} + \ldots + a_{1,n}\,\tilde{H}_t^{(n)} \\ a_{2,0}\,\tilde{H}_t^{(0)} + \ldots + a_{2,n}\,\tilde{H}_t^{(n)} \\ \vdots \qquad\qquad \vdots \\ a_{d,0}\,\tilde{H}_t^{(0)} + \ldots + a_{d,n}\,\tilde{H}_t^{(n)} \end{pmatrix},$$

whose component processes are all weak IDT processes. The random vector $(X_1, \ldots, X_d)$ defined via

$$X_k := \inf\{t > 0 \,:\, H_t^{(k)} > \epsilon_k\}, \quad k = 1, \ldots, d,$$

has an EM law. Moreover, if all processes $\tilde{H}^{(i)}$, $i = 0, \ldots, n$, are strong IDT, then $(X_1, \ldots, X_d)$ has an MSMVE law, and if they are all (possibly killed) Lévy subordinators, then $(X_1, \ldots, X_d)$ has an MO law.

Proof
First, notice that the component processes of $\boldsymbol{H}$ are all weak IDT processes by [A4, Lemma 4.1(b)]. Fix $t > 0$, $k \in \{1, \ldots, d\}$, a subset of indices $1 \leq i_1 < \ldots < i_k \leq d$, and let $c_1, \ldots, c_k > 0$ be constants. We observe that

$$\begin{aligned}\mathbb{P}\big(\min\{c_1\, X_{i_1}, \ldots, c_k\, X_{i_k}\} > t\big) &= \mathbb{P}\big(H_{t/c_j}^{(i_j)} < \epsilon_{i_j},\, j = 1, \ldots, k\big) = \mathbb{E}\Big[\exp\Big(-\sum_{j=1}^{k} H_{t/c_j}^{(i_j)}\Big)\Big] \\ &= \mathbb{E}\Big[\exp\Big(-\sum_{\ell=0}^{n}\sum_{j=1}^{k} a_{i_j,\ell}\, \tilde{H}_{t/c_j}^{(\ell)}\Big)\Big].\end{aligned}$$

First assume $c_1 = \ldots = c_k = 1$. Then $Y := \Big\{\sum_{\ell=0}^{n}\Big(\sum_{j=1}^{k} a_{i_j,\ell}\Big)\tilde{H}_t^{(\ell)}\Big\}_{t\geq 0}$ equals a weighted sum of n independent weak IDT processes, which itself is a weak IDT process by [A4, Lemma 4.1(b)]. Hence there exists a Bernstein function Ψ such that

$$\mathbb{P}\big(\min\{X_{i_1}, \ldots, X_{i_k}\} > t\big) = \mathbb{E}\Big[e^{-Y_t}\Big] = e^{-t\,\Psi(1)},$$

implying that $\min\{X_{i_1}, \ldots, X_{i_k}\}$ is exponential with rate $\Psi(1)$. Hence, $(X_1, \ldots, X_d)$ has an EM law. Next let $c_1, \ldots, c_k$ arbitrary and assume that all involved processes are actually strong IDT. In this case it follows from Lemma 3.2 that the processes $\Big\{\sum_{j=1}^{k} a_{i_j,\ell}\,\tilde{H}_{t/c_j}^{(\ell)}\Big\}_{t\geq 0}$ are independent, strong IDT processes, $\ell = 0, \ldots, n$. Hence, their sum is a weak IDT process by [A4, Lemma 4.1(b)] and again it follows the existence of a Bernstein function Ψ such that

$$\mathbb{P}\big(\min\{c_1\, X_{i_1}, \ldots, c_k\, X_{i_k}\} > t\big) = e^{-t\,\Psi(1)},$$

implying that $\min\{c_1\, X_{i_1}, \ldots, c_k\, X_{i_k}\}$ is exponential with rate $\Psi(1)$, and $(X_1, \ldots, X_d)$ has an MSMVE law. Finally, if all involved processes are (possibly killed) Lévy subordinators then the resulting random vector $(X_1, \ldots, X_d)$ has an MO law, which can be derived analogously to [Mai (2010), Lemma 5.2.5, p. 133]. □

All members of the family MO can be obtained by the stochastic construction in Lemma 4.4, see, e.g., [Sun et al. (2012), Theorem 4.2]. In particular, the construction of Lemma 4.4 goes far beyond the cosmos of conditionally iid models. However, it is an interesting open question how far-reaching the subclass of MSMVE is which is obtained via the construction in Lemma 4.4 with strong IDT processes.

The random vectors constructed in Lemma 4.4 are h-extendible if the matrix A is of a special form. For example, assume a partition $d_1 + \ldots + d_J$ of the dimension into J groups, and set $n := J$. We interpret the factor $\tilde{H}^{(0)}$ as the global factor and the factors $\tilde{H}^{(j)}$, $j = 1, \ldots, J$, as group-specific factors. If $k \in \{1, \ldots, d\}$ is an index of group j, then the k-th row of A must be defined as

$$(a_{k,0}, \ldots, a_{k,J}) = (1, \underbrace{0, \ldots, 0}_{j-1}, 1, 0, \ldots, 0),$$

i.e. a loading vector which loads the global factor $\tilde{H}^{(0)}$ and the j-th group-specific factor $\tilde{H}^{(j)}$. This produces an h-extendible structure with two levels of hierarchy. Deeper levels can be produced in a similar manner. For the sake of notational simplicity, let us proceed with the example of two levels. The vector-valued process $\boldsymbol{H}$ looks as follows in this case:

$$\boldsymbol{H}_t = \begin{pmatrix} \tilde{H}_t^{(1)} + \tilde{H}_t^{(0)} \\ \vdots \\ \tilde{H}_t^{(1)} + \tilde{H}_t^{(0)} \\ \tilde{H}_t^{(2)} + \tilde{H}_t^{(0)} \\ \vdots \\ \vdots \\ \tilde{H}_t^{(J)} + \tilde{H}_t^{(0)} \\ \vdots \\ \tilde{H}_t^{(J)} + \tilde{H}_t^{(0)} \end{pmatrix} = \begin{pmatrix} \tilde{H}_t^{(0)} \\ \vdots \\ \tilde{H}_t^{(0)} \\ \tilde{H}_t^{(0)} \\ \vdots \\ \vdots \\ \tilde{H}_t^{(0)} \\ \vdots \\ \tilde{H}_t^{(0)} \end{pmatrix} + \begin{pmatrix} \tilde{H}_t^{(1)} \\ \vdots \\ \tilde{H}_t^{(1)} \\ 0 \\ \vdots \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \ldots + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ \vdots \\ \tilde{H}_t^{(J)} \\ \vdots \\ \tilde{H}_t^{(J)} \end{pmatrix}.$$

In particular, it equals the sum of independent processes, say $\tilde{\boldsymbol{H}}^{(0)}, \tilde{\boldsymbol{H}}^{(1)}, \ldots, \tilde{\boldsymbol{H}}^{(J)}$, whose non-zero components are identical. One can construct independent random vectors $(Y_1^{(0)}, \ldots, Y_d^{(0)})$, $(Y_1^{(1)}, \ldots, Y_d^{(1)})$, ..., $(Y_1^{(J)}, \ldots, Y_d^{(J)})$ from the stochastic processes $\tilde{\boldsymbol{H}}^{(0)}, \tilde{\boldsymbol{H}}^{(1)}, \ldots, \tilde{\boldsymbol{H}}^{(J)}$ as follows:

$$Y_k^{(0)} := \inf\{t > 0 \,:\, \tilde{H}_t^{(0)} > \epsilon_k^{(0)}\}, \quad k = 1, \ldots, d,$$
$$Y_k^{(j)} := \begin{cases} \infty, & \text{if } \{\tilde{H}_t^{(j)}\}_{t \geq 0} \neq 0 \\ \inf\{t > 0 \,:\, \tilde{H}_t^{(j)} > \epsilon_k^{(j)}\}, & \text{else} \end{cases},$$

where $\epsilon_k^{(j)}$ are iid unit exponentials, $j = 0, 1, \ldots, J$, $k = 1, \ldots, d$. It then follows that

$$(X_1, \ldots, X_d) \stackrel{d}{=} \Big(\min_{j=0,\ldots,J} \{Y_1^{(j)}\}, \ldots, \min_{j=0,\ldots,J} \{Y_d^{(j)}\} \Big), \tag{6}$$

a proof is postponed to the Appendix. The finite components of the random vectors $(Y_1^{(j)}, \ldots, Y_d^{(j)})$ are conditionally iid. Hence, the stochastic model (6) relies solely on the computation of componentwise minima of otherwise conditionally iid random vectors. We have seen earlier that each conditionally iid random vector arises from an IDT process, which is $\{\tilde{H}_t^{(j)}\}_{t \geq 0}$ for the j-th group. In words, the operation "addition of independent IDT processes" in the canonical construction corresponds to the operation "componentwise minima" on the level of the associated multivariate exponential laws. This indicates why h-extendibility is a nice feature, in particular with regards to simulation. Sampling these models is accomplished easily by the minimum operation, once the extendible building blocks can be simulated.

4.3 H-extendible scale mixtures of Marshall–Olkin distributions

The reference [A7] studies an extendible family of distributions called *scale mixtures of Marshall–Olkin copulas (SMMO)*, as well as h-extendible generalizations thereof. The family SMMO is a superclass of both the extendible Archimedean copula family as well as the extendible Marshall–Olkin survival copula family. Therefore, the main motivation for studying this class of distributions lies in combining the distinct properties of the two building blocks. This example is therefore perfect in order to illustrate how the concept of h-extendibility allows to combine different families of distributions to obtain richer families. Interestingly, in the bivariate case the family SMMO is a subclass of a family of copulas called *Archimax copulas*, which was introduced by [Capéràa et al. (2000)]. The interested reader is referred to [A7] for a precise definition, a thorough investigation of dependence properties, and an application to the pricing of portfolio credit derivatives.

5 Applications and related results

The present section deals with diverse applications related to the distributions presented in earlier sections and is organized as follows. Subsection 5.1 tackles parameter estimation for (high-dimensional) Marshall–Olkin distributions with a conditionally iid structure. Subsection 5.2 shows how to compute densities for a large family of (one-dimensional) infinitely divisible distributions. This might be necessary when the stochastic model (2) is applied with a weak IDT process which is parameterized in terms of its associated Bernstein function and its density is required. One such application – the pricing of portfolio credit derivatives – is illustrated in Subsection 5.3.

5.1 Parameter estimation for extendible Marshall–Olkin distributions

One of the classical statistical problems is the parametric estimation of distributions from observed data. The family of d-dimensional Marshall–Olkin distribution functions is parameterized by $2^d - 1$ parameters in general. It is known from the author's dissertation

[Mai (2010)] that the extendible subfamily can be constructed from Lévy subordinators, and parameterized in terms of their associated Bernstein functions. Consequently, every parametric family of Bernstein functions $\Psi = \Psi_{\boldsymbol{\theta}}$ induces a parametric (sub)family of Marshall–Olkin distributions. In other words, the $2^d - 1$ parameters of the general Marshall–Olkin law are given as functions of the parameter (vector) $\boldsymbol{\theta}$. The article [A8] investigates parametric estimation in such a model, and the main finding is sketched in the sequel.

The probability law of a d-dimensional random vector $(X_1, \ldots, X_d)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ belongs to the extendible subfamily of MO if and only if there is a Bernstein function $\Psi : [0, \infty) \to [0, \infty)$ such that

$$\bar{F}(t_1, \ldots, t_d) := \mathbb{P}(X_1 > t_1, \ldots, X_d > t_d) = \exp\Big(- \sum_{k=1}^{d} t_{[d+1-k]} \big(\Psi(k) - \Psi(k-1)\big)\Big),$$

where $0 \leq t_{[1]} \leq \ldots \leq t_{[d]}$ denotes the ordered list of the arguments $t_1, \ldots, t_d \geq 0$. In the sequel, we assume that $\Psi = \Psi_{\boldsymbol{\theta}}$ is chosen from a parametric family for a parameter (vector) $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. It is furthermore assumed that the p-dimensional parameter space Θ is open and $\Psi_{\boldsymbol{\theta}}(1) = 1$. The latter condition normalizes the univariate marginal laws in the sense that $X_k \sim \mathcal{E}(1)$ for all components $k = 1, \ldots, d$, which implies that the estimator derived below focuses solely on the dependence structure. Regarding the considered family of multivariate distributions, there is one interesting aspect indicating that parameter estimation is difficult in general: considering a subvector $(X_{i_1}, \ldots, X_{i_k})$ of length $2 \leq k \leq d$, its distribution function is completely determined by the numbers $\Psi(2), \ldots, \Psi(k)$, even though the law of the full vector $(X_1, \ldots, X_d)$ is only determined by the full sequence $\Psi(2), \ldots, \Psi(d)$. This means that a statistical estimator based on proper subvectors in general cannot be a sufficient statistic for the d parameters $\Psi(1),\ldots,\Psi(d)$ of the extendible Marshall-Olkin law. This makes parameter estimation difficult and is a fundamental difference compared with, e.g., the multivariate normal distribution, where bivariate subvectors determine the overall distribution. Another difficulty is the fact that the Marshall–Olkin distribution does not have a density with respect to Lebesgue measure, rendering standard Maximum Likelihood techniques infeasible.

The main result of [A8] is summarized in Theorem 5.1 below. The estimation procedure pursues the following two-step algorithm, based on n iid observations $(X_1^{(i)}, \ldots, X_d^{(i)})$, $i = 1, \ldots, n$, sharing the extendible Marshall–Olkin distribution of $(X_1, \ldots, X_d)$ with true parameter vector $\boldsymbol{\theta}_0 \in \Theta$:

(i) Unbiased and strongly consistent estimators $\hat{b}_{k,n}$, $k = 2, \ldots, d$, are derived for the sequence $b_k(\boldsymbol{\theta}) := 1/(\Psi_{\boldsymbol{\theta}}(k) + 1)$, $k = 2, \ldots, d$.

(ii) It is shown that the minimum-distance estimator

$$\hat{\boldsymbol{\theta}}_n := \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \sum_{k=2}^{d} \big(\hat{b}_{k,n} - b_k(\boldsymbol{\theta})\big)^2 \tag{7}$$

is well-defined for sufficiently large n, strongly consistent, and asymptotically normal.

In order to formulate the estimator, required are certain constants that have to be defined recursively. We denote $\kappa_{n,n} := 1$ and

$$\kappa_{n-k,n} := -\sum_{i=1}^{k} S(n, n-i)\, \kappa_{n-k,n-i}, \quad k = 1, \ldots, n-1,$$

where

$$S(k,l) := \frac{(-1)^l}{l!} \sum_{j=0}^{l} (-1)^j \binom{l}{j} j^k, \quad l, k \in \mathbb{N}_0,$$

are the so-called *Stirling numbers of the second kind.* Furthermore, the following battery of technical assumptions on the parametric family $\{\Psi_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ of Bernstein functions is required:

(IC) If $\{\boldsymbol{\theta}_n\}_{n\in\mathbb{N}}$ is a sequence of parameters with $\lim_{n\to\infty} \Psi_{\boldsymbol{\theta}_n}(k) = \Psi_{\boldsymbol{\theta}_0}(k)$, $k = 2, \ldots, d$, then $\lim_{n\to\infty} \boldsymbol{\theta}_n = \boldsymbol{\theta}_0$. Moreover, $\boldsymbol{\theta} \mapsto \Psi_{\boldsymbol{\theta}}(k)$ is continuous for every fixed $k = 2, \ldots, d$.

(SC1) The partial derivative $\frac{\partial^2}{\partial\theta_i\,\partial\theta_j} b_k(\boldsymbol{\theta})$ exists and is continuous for every $1 \leq i, j \leq p$, $2 \leq k \leq d$.

(SC2) The Hessian matrix of $\Phi(\boldsymbol{\theta}) := \sum_{k=2}^{d} \big(b_k(\boldsymbol{\theta}_0) - b_k(\boldsymbol{\theta})\big)^2$ is invertible at $\boldsymbol{\theta}_0$.

The identifiability condition (IC), as well as both smoothness conditions (SC1), (SC2), are satisfied for typical parametric families of Bernstein functions. The condition (IC) implies in particular that two different parameter vectors $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ imply two different Marshall–Olkin distributions, which is a reasonable assumption for any parametric model.

Theorem 5.1 (Parameter estimation for extendible MO)
For $k = 2, \ldots, d$, define the statistics

$$\hat{b}_{k,n} := \frac{1}{n} \sum_{i=1}^{n} L_k\Big(\exp\big(-X_1^{(i)}\big), \ldots, \exp\big(-X_d^{(i)}\big)\Big),$$

$$L_k(u_1, \ldots, u_d) := \frac{(d-k)!}{d!} \sum_{j=1}^{k} \kappa_{j,k} \sum_{l=1}^{d} l^j\, (u_{[l+1]} - u_{[l]}),$$

where $0 \leq u_{[1]} \leq \ldots \leq u_{[d]} \leq u_{[d+1]} := 1$ denotes the ordered list of $u_1, \ldots, u_d \in [0,1]$ with the convention $u_{[d+1]} := 1$. Moreover, define the estimator $\hat{\boldsymbol{\theta}}_n$ as in (7). The following statements are valid:

(a) $(\hat{b}_{1,n}, \ldots, \hat{b}_{d,n})$ tends almost surely to $(b_2(\boldsymbol{\theta}_0), \ldots, b_d(\boldsymbol{\theta}_0))$, as $n \to \infty$.

(b) If (IC) holds, $\hat{\boldsymbol{\theta}}_n$ is well-defined for almost all n and tends almost surely to $\boldsymbol{\theta}_0$ as $n \to \infty$.

(c) If additionally (SC1) and (SC2) hold, then $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ tends in distribution to a multivariate normal distribution with zero mean vector.

Proof
See [A8]. □

The asymptotic normality statement (c) implies the strong consistency statement (b), and therefore requires more technical assumptions. It is obvious that the evaluation of the estimator $\hat{\boldsymbol{\theta}}_n$ itself is a numerically burdensome task, because a p-dimensional minimization has to be carried out. From a practical point of view this prevents the dimension p of the parameter space Θ from being large. Typical examples are such that $p \ll d$, for example $p = 2$. Interestingly, the empirical example carried out in [A8] indicates that the accuracy of the estimator improves not only with the sample size (which is clear by consistency), but also massively with the dimension d. Finally, the asymptotic covariance matrix of the estimator can be given in closed form.

5.2 Evaluating the density of distributions from the Bondesson class

As we have seen, the family of univariate probability laws μ on $[0, \infty]$ which are infinitely divisible, denoted by $ID[0, \infty]$ in the sequel, is of paramount interest in the theory of multivariate exponential distributions. It is well-known that there is a one-to-one relationship between this family of probability laws and the family of Bernstein functions, denoted BF in the sequel. For each $\mu \in ID[0, \infty]$ there exists a unique $\Psi \in BF$ such that the Laplace transform of μ equals $\exp(-\Psi)$. Conversely, for every $\Psi \in BF$ the function $\exp(-\Psi)$ is the Laplace transform of some uniquely determined $\mu \in ID[0, \infty]$. Clearly, there are many applications (e.g. Maximum Likelihood estimation), where it is convenient to be able to compute the density (if existent) of $\mu \in ID[0, \infty]$ efficiently. There exist some $\mu \in ID[0, \infty]$ for which the density is known analytically, but the Bernstein function is not explicitly known, a prominent example being the lognormal distribution. However, the reverse situation occurs much more frequently: in the literature one can find numerous Bernstein functions with nice algebraic form for which the associated density is known to exist but an algebraic expression is unknown and/or difficult to evaluate numerically. The most prominent example is probably the stable distribution with Bernstein function $\Psi(x) = x^\alpha$ for $\alpha \in (0, 1)$. In such a situation, deriving the density numerically via Laplace inversion algorithms appears to be natural. The article [A9] provides a convenient Laplace inversion algorithm for distributions of the so-called *Bondesson class* $BO[0, \infty] \subsetneq ID[0, \infty]$, which is a large subfamily of $ID[0, \infty]$.

Assume that a probability law $\mu \in ID[0, \infty]$ has Bernstein function Ψ and density f_μ. By the well-known Bromwich inversion formula, under mild conditions f_μ can be retrieved from Ψ via the formula

$$f_\mu(x) = \lim_{R\to\infty} \frac{1}{2\,\pi\,i} \int_{a-i\,R}^{a+i\,R} e^{x\,z}\, e^{-\Psi(z)}\, dz, \quad x > 0, \tag{8}$$

with an arbitrary parameter $a > 0$. Conditions ensuring the validity of (8) are existence, bounded variation and continuity of the density f_μ. Unfortunately, these conditions can often not be checked, because f_μ is not known and we only know Ψ – that's why we consider Laplace inversion at all. However, in our situation [Sato (1999), Proposition 28.1] states that $f_\mu \in C^1$ if $\int_{\mathbb{R}} |\exp(-\Psi(-i\,s))|\, ds < \infty$, constituting a sufficient condition for (8) to be valid that can be checked solely from Ψ. By elementary manipulations, the integral in (8) can be simplified to

$$f_\mu(x) = \lim_{R\to\infty} \frac{1}{\pi}\, \mathrm{Im}\Big(\int_a^{a+i\,R} e^{x\,z}\, e^{-\Psi(z)}\, ds \Big), \quad x > 0,$$

yielding only a one-sided improper integral. When applying this formula for numerically retrieving the density there are two numerical issues: (i) one faces a truncation error because $R \gg 0$ must be fixed, and (ii) irrespectively of the Bernstein function the integrand is naturally oscillating due to the term $\exp(x\,(a + i\,s)) = \exp(x\,a)\,(\cos(x\,s) + i\,\sin(x\,s))$, $s \in (0, R)$. One standard approach to tackle the second issue is by resorting to Cauchy's Theorem and change the path of integration to a different, more convenient contour. That is, instead of evaluating the integrand along the Bromwich contour $a+i\,s$, $s \in (0, R)$, the integrand is evaluated along an alternative contour in the complex plane that avoids the regions of high oscillations. Such a strategy is pursued below.

A probability measure $\mu \in ID[0,\infty]$ is said to be in the Bondesson class $BO[0,\infty]$, which has been introduced in [Bondesson (1981)] under the name *g.c.m.e.d.* distributions, if its associated Bernstein function Ψ is *complete*, i.e. the associated Lévy measure has a completely monotone derivative with respect to the Lebesgue measure. The family $BO[0,\infty]$ can alternatively be introduced as the smallest class of distributions closed under convergence and convolution containing mixtures of a special family of distributions, see [Sato (1999), Definition 51.9, p. 389]. Using the complete monotonicity of the Lévy density together with Bernstein's Theorem, the associated complete Bernstein function Ψ for $\mu \in BO[0,\infty]$ can be written as

$$\Psi(x) = \mu\,x + \int_0^\infty \frac{x}{x+t}\, \sigma(dt), \quad x \geq 0, \tag{9}$$

with σ a measure on $(0,\infty)$ satisfying $\int_{(0,\infty)} 1/(1+t)\,\sigma(dt) < \infty$, called the *Stieltjes measure*, see [Schilling et al. (2010), Theorem 6.2(ii), p. 49]. Compared with the classical Lévy-Khinchin formula, which holds for arbitrary Bernstein functions, the representation (9) for complete Bernstein functions implies the existence of a holomorphic extension of Ψ from the domain $[0,\infty)$ to the sliced complex plane $\mathbb{C} \setminus (-\infty, 0)$, see [Schilling et al. (2010), Theorem 6.2, p. 49]. This is an essential observation which allows to consider integration contours ending in the left half-plane $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\}$. We can now state the main result of [A9].

Theorem 5.2 (Laplace inversion for distributions from the Bondesson class)
If μ is a distribution from $BO[0,\infty]$ with associated complete Bernstein function Ψ and density f_μ such that (8) holds, then for $x > 0$ one has

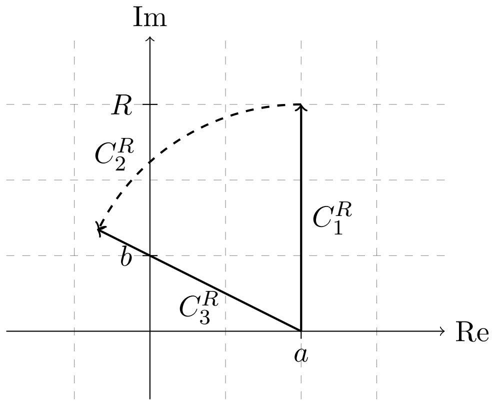
$$f_\mu(x) = \frac{M\,e^{x\,a}}{\pi} \int_0^1 \mathrm{Im}\Big(e^{x\,M\,\log(v)\,(b\,i-a)}\,e^{-\Psi(a-M\,\log(v)\,(b\,i-a))}\,(b\,i-a)\Big)\,\frac{dv}{v}$$

with arbitrary parameters $a, b > 0$ and $M > 2/(a\,x)$. This integral is a proper Riemannian integral as one can show that the integrand vanishes for $v \downarrow 0$.

Proof
See [A9, Theorem 3.1]. $\square$

The heuristic idea for the proof of Theorem 5.2 is explained best in the following picture.

$$C_1^R : a + i\,u, \quad 0 \le u \le R$$
$$C_2^R : a + R\,e^{i\,u}, \quad \frac{\pi}{2} \le u \le \pi - \arctan(b/a)$$
$$C_3^R : a + u\,(b\,i - a), \quad 0 \le u \le \frac{R}{\sqrt{a^2+b^2}}$$



The classical Bromwich contour evaluates the integrand along the contour C_1^R. Since this contour is quite unfavorable due to the high oscillations along lines that are parallel to the imaginary axis, one would rather like to evaluate the integrand along the contour C_3^R, which ends in the left half-plane, where the oscillations disappear rapidly due to the exponential decay of the term $\exp(x\,z)$ in (8) as $\mathrm{Re}(z) \to -\infty$. Cauchy's Theorem implies that the integral along the closed contour $C_1^R + C_2^R - C_3^R$ is zero. It can be shown that the integral along the connecting path C_2^R tends to zero as $R \to \infty$, which is the major technical step in the proof of Theorem 5.2. Consequently, the integral along the Bromwich contour C_1^R equals the integral along the alternative contour C_2^R, as $R \to \infty$.

Regarding practical applications, [A9] discuss some recommendations regarding the choice of the free parameters a, b, M. As a prominent example, the case of the stable distribution is investigated in great detail, and the resulting Laplace inversion algorithm is shown to be able to keep up in terms of efficiency and accuracy with an alternative formula that is specifically designed for the stable law in [Nolan (1997)]. This is surprising given the generality of Theorem 5.2, which is by far not restricted to stable distributions. Moreover, as a corollary to Theorem 5.2 – with almost identical proof – a convenient Laplace inversion algorithm for computing the distribution function of $\mu \in BO[0,\infty]$ is derived as well.

5.3 Pricing portfolio credit derivatives

The author's initial motivation to study random vectors with conditionally iid dependence structure stems from an application in Mathematical Finance: the pricing of portfolio credit derivatives. Before the default of the investment bank Lehman Brothers in 2008, the market for so-called collateralized debt obligations (CDOs) has experienced a period of steady growth and was one of the paramount topics in the banking industry. CDOs are credit derivatives whose market value depends critically on the creditworthiness of an underlying basket of credit-risky assets. From a mathematical viewpoint, there are two fundamental challenges that make this kind of products interesting also for theorists: (i) the dependence structure between the underlying assets has a strong effect on the market value of CDOs, and (ii) the number d of underlying assets is quite large, e.g. $d = 125$ is a standard assumption in many baskets. Therefore, one has to build a high-dimensional stochastic model with an intuitive dependence structure which is still simple enough to guarantee a high level of practical viability – a Herculean task.

Now what is required for such a model to be viable? Extracting the essential mathematical issues related to CDO pricing, required is a stochastic model for the random vector $(X_1, \ldots, X_d)$ of default times of the d underlying assets such that the probability distribution of the stochastic process $\{L_t^{(d)}\}_{t \geq 0}$, defined by

$$L_t^{(d)} := \frac{1}{d} \sum_{k=1}^{d} 1_{\{X_d \leq t\}} = \text{relative portfolio loss until time } t,$$

is given in convenient form. Since the default times $X_1, \ldots, X_d$ have to be modeled dependently, the distribution of $L_t^{(d)}$ is far from trivial in general. However, in the conditionally iid setup of construction (2) the Theorem of Glivenko–Cantelli implies the almost sure convergence

$$L_t^{(d)} \longrightarrow 1 - e^{-H_t}, \quad d \to \infty,$$

uniformly in $t \geq 0$, see [Mai et al. (2013)]. For continuous (and hence bounded) functions $f : [0, 1] \to \mathbb{R}$ this justifies approximations such as

$$\mathbb{E}\big[f(L_t^{(d)})\big] \approx \mathbb{E}\Big[f\Big(1 - e^{-H_t}\Big)\Big], \quad t > 0,$$

which render several pricing formulas for CDOs numerically tractable. A survey of numerous models in this spirit is provided in [Mai et al. (2013)]. Generally speaking, pricing CDOs in a conditionally iid setup relies on homogeneity assumptions, which are required in order to make the aforementioned approximation technique feasible, but are unrealistic in general. Nevertheless, extendible models are typically the ones that are used in the industry to track observed market prices in front office systems, because their computational efficiency is indispensable.

In this context, an interesting analysis is to explore precisely which stylized facts can be explained via conditionally iid models, and which cannot. Many CDO pricing models can be subsumed under the umbrella of so-called *(one-factor) copula models*. The term "one-factor" is added intentionally in brackets because what people often mean implicitly when referring to "copula models" is an approach which combines marginal distributions with a copula function inherited from an extendible distribution – predominantly a one-factor Gaussian copula, a one-factor t-copula, an Archimedean copula, or extensions of a Gaussian one-factor copula which replace the underlying normal distribution of the latent factor by another distribution. All these copula families are static in the sense that the underlying stochastic factor inducing dependence is a single random variable. This fact provides solid ground for criticism – of which the academic literature is packed – because all these models cannot explain changing developments over time. It is one of the major concerns of the present author to point out that using a one-factor structure is not equivalent to static modeling. As we have seen, in general a latent factor in a conditionally iid model is a stochastic process $H = \{H_t\}$. The aforementioned popular copula families are all such that $H = \{H_t\}$ is a function of a single random variable and time t, e.g., like in (4) for the case of Archimedean copulas. Clearly, it is impossible to extract a reasonable information flow (i.e. filtration) from such a (trivial) stochastic process. However, we have also seen that there exist popular multivariate distributions arising from stochastic processes $H = \{H_t\}$ with non-trivial natural filtration, like Lévy subordinators, or strong and weak IDT processes. Stemming from a similar motivation, the so-called Cox processes (or doubly-stochastic processes) have found their way into credit risk modeling, precisely due to the fact that the market's opinion about future defaults changes continuously and this information flow needs to be modeled. Studying the connection between stochastic "drivers of information flow" on the one hand and the associated law of static, future event times on the other hand is another perspective that might be taken on the investigations carried out in the present thesis. Indeed, the formalism of a canonical construction like (2) can help to compare apples ("static" one-factor copula models) and oranges ("dynamic" top down models for the portfolio loss process), see [Mai et al. (2013)].

The article [A10] presents a new conditionally iid model which is based on two independent stochastic building blocks: a Lévy subordinator and a Brownian motion. On the one hand, the jumps of the Lévy subordinator can account for cataclysmic events which are required in order to explain market quotes for CDOs. A huge proportion of these observed prices can be attributed to the market's fear of dramatic downturns. Moreover, the connection between Lévy subordinators and Marshall–Olkin distributions has been studied intensively in the author's dissertation, where it is also shown that CDO market prices are explained well. However, observed fluctuations of market prices over time cannot be captured by the Lévy subordinator alone. This is due to the lack-of-memory property of the Marshall–Olkin distribution, implying constant credit spreads between observed defaults. On the other hand, it is well-known that multivariate default models built solely on the "doubly-stochastic" idea from Brownian drivers fail to explain observed market prices of CDOs, because cataclysmic events do not happen with high probability, see, e.g. [Das et al. (2007)]. However, these models can explain fluctuations of market prices, as desired. In the combined model proposed in [A10], the two building blocks are put together in such a way that their synthesis inherits the desirable features, but not too much analytical tractability is lost. In particular, efficient pricing algorithms based on Laplace inversion algorithms are feasible.

6 Conclusion

A constructive approach to the modeling of high-dimensional random vectors was presented. The major idea was to first study conditionally iid models in depth, and then construct multi-factor models from conditionally iid building blocks. Particular focus has been put on multivariate exponential distributions. It has been shown that the underlying, dependence-inducing factors are stochastic processes which are infinitely divisible with respect to time. Finally, applications of the introduced concepts to parametric estimation of Marshall-Olkin distributions, to the numerical density evaluation for certain infinitely divisible laws, and to the pricing of portfolio credit derivatives have been demonstrated.

Appendix

Proof (that Lemma 3.2 does not hold for weak IDT processes)
Let $\alpha \in (0,1)$ and consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting the following three independent objects: an α-stable Lévy subordinator $S = \{S_t\}_{t \geq 0}$, a positive random variable M with Laplace transform $\mathbb{E}[\exp(-x\,M)] = \exp(-x^{\alpha})$, $x \geq 0$, and a Bernoulli variable Z with success probability $1/2$. The stochastic process $\{H_t\}_{t \geq 0}$, defined by

$$H_t := Z\,S_t + (1-Z)\,M\,t^{1/\alpha}, \quad t \geq 0,$$

is weak IDT but not strong IDT, as shown in [A4, Example 5.1]. We show in the sequel that the process $\{H_t + H_{2\,t}\}_{t \geq 0}$ is not weak IDT, showing that the closure property of Lemma 3.2 is distinctive to strong IDT processes.
We fix $t > 0$ and $x > 0$ and compute

$$\begin{aligned}\mathbb{E}\Big[e^{-x\,(H_t+H_{2\,t})}\Big] &= \frac{1}{2}\,\mathbb{E}\Big[e^{-x\,(S_t+S_{2\,t})}\Big] + \frac{1}{2}\,\mathbb{E}\Big[e^{-x\,M\,t^{1/\alpha}\,(1+2^{1/\alpha})}\Big]\\ &= \frac{1}{2}\Big(e^{-t\,x^{\alpha}(2^{\alpha}+1)} + e^{-t\,x^{\alpha}\,(1+2^{1/\alpha})^{\alpha}}\Big).\end{aligned}$$

If the process $\{H_t + H_{2\,t}\}_{t \geq 0}$ were weak IDT, then the logarithm of its Laplace transforms were linear in t by [A4, Theorem 1.1]. However, it can be checked that the logarithm of the last expression is not linear in t, because $2^{\alpha} + 1 \neq (1 + 2^{1/\alpha})^{\alpha}$ (even though it looks "almost" linear when visualized numerically). This implies the claim. $\square$

Proof (of Equation (6))
Let $s_1, \ldots, s_d \geq 0$. For the sake of notational simplicity we denote

$$I(1) := \{1, \ldots, d_1\}, \quad I(j) := \{d_1 + \ldots + d_{j-1} + 1, \ldots, d_1 + \ldots + d_j\}, \quad j \geq 2,$$

the set of indices corresponding to group $j = 1, \ldots, J$, and compute

$$\mathbb{P}\Big(X_1 > s_1, \ldots, X_d > s_d\Big) = \mathbb{P}\Big(\bigcap_{j=1}^{J} \bigcap_{k \in I(j)} \Big\{\inf\{t > 0 \,:\, \tilde{H}_t^{(j)} + \tilde{H}_t^{(0)} > \epsilon_k\} > s_k\Big\}\Big)$$
$$= \mathbb{P}\Big(\bigcap_{j=1}^{J} \bigcap_{k \in I(j)} \Big\{\tilde{H}_{s_k}^{(j)} + \tilde{H}_{s_k}^{(0)} \leq \epsilon_k\Big\}\Big) = \mathbb{E}\Big[\prod_{j=1}^{J} \prod_{k \in I(j)} e^{-(\tilde{H}_{s_k}^{(j)} + \tilde{H}_{s_k}^{(0)})}\Big]$$
$$= \mathbb{E}\Big[\prod_{k=1}^{d} e^{-\tilde{H}_{s_k}^{(0)}}\Big] \prod_{j=1}^{J} \mathbb{E}\Big[\prod_{k \in I(j)} e^{-\tilde{H}_{s_k}^{(j)}}\Big] = \mathbb{P}\Big(\bigcap_{k=1}^{d} \Big\{\tilde{H}_{s_k}^{(0)} \leq \epsilon_k^{(0)}\Big\}\Big) \prod_{j=1}^{J} \mathbb{P}\Big(\bigcap_{k \in I(j)} \Big\{\tilde{H}_{s_k}^{(j)} \leq \epsilon_k^{(j)}\Big\}\Big)$$
$$= \mathbb{P}\Big(\Big(\bigcap_{j=1}^{J} \bigcap_{k \in I(j)} \Big\{\tilde{H}_{s_k}^{(j)} \leq \epsilon_k^{(j)}\Big\}\Big) \cap \Big(\bigcap_{k=1}^{d} \Big\{\tilde{H}_{s_k}^{(0)} \leq \epsilon_k^{(0)}\Big\}\Big)\Big)$$
$$= \mathbb{P}\Big(\Big(\bigcap_{j=1}^{J} \bigcap_{k \in I(j)} \Big\{\inf\{t : \tilde{H}_t^{(j)} > \epsilon_k^{(j)}\} > s_k\Big\}\Big) \cap \Big(\bigcap_{k=1}^{d} \Big\{\inf\{t : \tilde{H}_t^{(0)} > \epsilon_k^{(0)}\} > s_k\Big\}\Big)\Big)$$
$$= \mathbb{P}\Big(\bigcap_{j=1}^{J} \bigcap_{k \in I(j)} \Big\{\min\Big\{\inf\{t : \tilde{H}_t^{(j)} > \epsilon_k^{(j)}\}, \inf\{t : \tilde{H}_t^{(0)} > \epsilon_k^{(0)}\}\Big\} > s_k\Big\}\Big)$$
$$= \mathbb{P}\Big(\min_{j=0,\ldots,J} \big\{Y_1^{(j)}\big\} > s_1, \ldots, \min_{j=0,\ldots,J} \big\{Y_d^{(j)}\big\} > s_d\Big). \qquad \square$$

Acknowledgements

I would like to thank all my co-authors German Bernhart, Marcos Escobar-Anel, Christian Hering, Marius Hofert, Pablo Olivares, Steffen Schenk, and Matthias Scherer for the great collaboration and the many vivid and fruitful “math discussions” related to our joint projects during the last years.

References

[A1] J.-F. Mai, M. Scherer, The Pickands representation of survival Marshall–Olkin copulas, *Statistics and Probability Letters* **80** (2010) pp. 357–360.

[A2] J.-F. Mai, M. Scherer, Bivariate extreme-value copulas with discrete Pickands dependence measure, *Extremes* **14** (2011) pp. 311–324.

[A3] J.-F. Mai, M. Scherer, Sampling exchangeable and hierarchical Marshall–Olkin distributions, *Communications in Statistics Theory and Methods* **42** (2013) pp. 619–632.

[A4] J.-F. Mai, M. Scherer, Characterization of extendible distributions with exponential minima via processes that are infinitely divisible with respect to time, *Extremes* **17** (2014) pp. 77–95.

[A5] C. Hering, M. Hofert, J.-F. Mai, M. Scherer, Constructing hierarchical Archimedean copulas with Lévy subordinators, *Journal of Multivariate Analysis* **101:6** pp. 1428–1433 (2010).

[A6] J.-F. Mai, M. Scherer, H-extendible copulas, *Journal of Multivariate Analysis* **110** (2012) pp. , 151–160.

[A7] G. Bernhart, M. Escobar-Anel, J.-F. Mai, M. Scherer, Default models based on scale mixtures of Marshall–Olkin copulas: properties and applications, *Metrika* **76** (2013) pp. 179–203.

[A8] C. Hering, J.-F. Mai, Moment-based estimation of extendible Marshall–Olkin distributions, *Metrika* **75** (2012) pp. 601–620.

[A9] G. Bernhart, J.-F. Mai, S. Schenk, M. Scherer, The density for distributions from the Bondesson class, *Journal of Computational Finance, forthcoming* (2013).

[A10] J.-F. Mai, P. Olivares, S. Schenk, M. Scherer, A multivariate default model with spread and event risk, *Applied Mathematical Finance* **21** (2014) pp. 51–83.

[Applebaum (2004)] D. Applebaum, Lévy processes and stochastic calculus, *Cambridge University Press* (2004).

[Arnold (1975)] B.C. Arnold, A characterization of the exponential distribution by multivariate geometric compounding, *Sankhya: The Indian Journal of Statistics* **37** (1975) pp. 164–173.

[Berg et al. (1984)] C. Berg, J.P.R. Christensen, P. Ressel, Harmonic analysis on semigroups, *Springer, Berlin* (1984).

[Bernstein (1929)] S. Bernstein, Sur les fonctions absolument monotones, *Acta Mathematica* **52:1** 1–66 (1929).

[Billingsley (1995)] P. Billingsley, Probability and Measure, *Wiley Series in Probability and Statistics, Wiley, New York* (1995).

[Bondesson (1981)] L. Bondesson, Classes of infinitely divisible distributions and densities, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **57:1** (1981) pp. 39–71.

[Capéràa et al. (2000)] P. Capéràa, A.-L. Fougères, C. Genest, Bivariate distributions with given extreme value attractor, *Journal of Multivariate Analysis* **72** (2000) pp. 30–49.

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

[De Finetti (1931)] B. De Finetti, Funzione caratteristica di un fenomeno allatorio, *Atti della R. Accademia Nazionale dei Lincii Ser. 6, Memorie, Classe di Scienze, Fisiche, Matematiche e Naturali* **4** (1931) pp. 251–299.

[De Finetti (1937)] B. De Finetti, La prévision: ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincaré* **7** (1937) pp. 1–68.

[De Haan, Resnick (1977)] L. De Haan, S.I. Resnick, Limit theory for multivariate sample extremes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **40** (1977) pp. 317–337.

[Esary, Marshall (1974)] J.D. Esary, A.W. Marshall, Multivariate distributions with exponential minimums, *Annals of Statistics* **2** (1974) pp. 84–98.

[Es-Sebaiy, Ouknine (2008)] K. Es-Sebaiy, Y. Ouknine, How rich is the class of processes which are infinitely divisible with respect to time, *Statistics and Probability Letters* **78** (2008) pp. 537–547.

[Hakassou, Ouknine (2012)] A. Hakassou, Y. Ouknine, A contribution to the study of IDT processes, *Working paper* (2012).

[Hausdorff (1921)] F. Hausdorff, Summationsmethoden und Momentfolgen I, *Mathematische Zeitschrift* **9** (1921) pp. 74–109.

[Hausdorff (1923)] F. Hausdorff, Momentenproblem fr ein endliches Intervall, *Mathematische Zeitschrift* **16** (1923) pp. 220–248.

[Hewitt, Savage (1955)] E. Hewitt, L.J. Savage, Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.* **80** (1955) pp. 470–501.

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

[Joe (1993)] H. Joe, Parametric families of multivariate distributions with given margins, *Journal of Multivariate Analysis* **46** pp. 262–282 (1993).

[Joe (1997)] H. Joe, Multivariate models and dependence concepts, *Chapman & Hall/CRC* (1997).

[Joe, Hu (1996)] H. Joe, T. Hu, Multivariate distributions from mixtures of max-infinitely divisible distributions, *Journal of Multivariate Analysis* **57** pp. 240–265 (1996).

[Ressel (1985)] P. Ressel, De Finetti type theorems: an analytical approach, *Annals of Probability* **13** (1985) pp. 898–922.

[Mai (2010)] J.-F. Mai, Extendibility of Marshall–Olkin distributions via Lévy subordinators and an application to portfolio credit risk, *Dissertation Technische Universität München* (2010).

[Mai, Scherer (2012)] J.-F. Mai, M. Scherer, Simulating copulas, *Series in Quantitative Finance, Vol. 4, Imperial College Press, London* (2012).

[Mai, Scherer (2013)] J.-F. Mai, M. Scherer, Extendibility of Marshall-Olkin distributions and inverse Pascal triangles, *Brazilian Journal of Probability and Statistics* **27** (2013) pp. 310–321.

[Mai et al. (2013)] J.-F. Mai, M. Scherer, R. Zagst, CIID frailty models and implied copulas, *Copulae in Mathematical and Quantitative Finance (Proceedings of the 2nd Workshop on Copula Theory and its Applications, Krakau, Poland), edited by P. Jaworski et al., Springer-Verlag* (2013) pp. 201–230.

[Malov (1998)] S.V. Malov, On finite-dimensional Archimedean copulas, *in Asymptotic Methods in Probability and Statistics with Applications, in Stat. Ind. Technol. Birkhäuser, St. Petersburg, Boston, MA* (1998) pp. 19–35.

[Mansuy (2005)] R. Mansuy, On processes which are infinitely divisible with respect to time, *Working paper* (2005).

[Marshall, Olkin (1967)] A.W. Marshall, I. Olkin, A multivariate exponential distribution, *Journal of the American Statistical Association* **62** (1967) pp. 30–44.

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

[Marshall, Olkin (1995)] A.W. Marshall, I. Olkin, Multivariate exponential and geometric distributions with limited memory, *Journal of Multivariate Analysis* **53** (1995) pp. 110–125.

[McNeil (2008)] A.J. McNeil, Sampling nested Archimedean copulas, *Journal of Statistical Computation and Simulation* **78:6** pp. 567–581 (2008).

[McNeil, Nešlehová (2009)] A.J. McNeil, J. Nešlehová, Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions, *Annals of Statistics* **37:5B** (2009) pp. 3059–3097.

[Nolan (1997)] J.P. Nolan, Numerical calculation of stable densities and distribution functions, *Communications in Statistics – Stochastic Models* **13** (1997) pp. 759–774. (2009).

[Ressel (2013)] P. Ressel, Homogeneous distributions and a spectral representation of classical mean values and stable tail dependence functions, *Journal of Multivariate Analysis* **117** (2013) pp. 246–256.

[Sato (1999)] K.-I. Sato, Lévy processes and infinitely divisible distributions, *Cambridge University Press, Cambridge* (1999).

[Schilling et al. (2010)] R. Schilling, R. Song, Z. Vondracek, Bernstein functions, *De Gruyter* (2010).

[Sun et al. (2012)] Y. Sun, R. Mendoza–Arriaga, V. Linetsky, Marshall–Olkin multivariate exponential distributions, multidimensional Lévy subordinators, efficient simulation, and applications to credit risk, *Working paper* (2012).