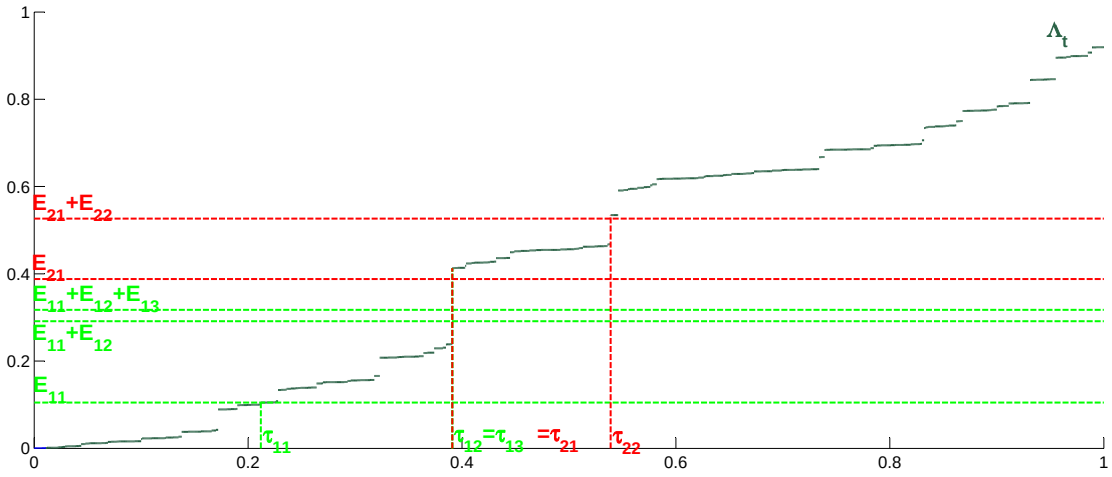
Dissertation

A multivariate Cox process with simultaneous jump arrivals and its application in insurance modelling

Daniela Anna Selch



TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Mathematik
Lehrstuhl für Finanzmathematik

TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Mathematik
Lehrstuhl für Finanzmathematik

A multivariate Cox process with simultaneous jump arrivals and its application in insurance modelling

Daniela Anna Selch

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Univ.-Prof. Dr. Caroline Lasser
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Matthias Scherer
	2.	Prof. Dr. Hansjörg Albrecher Université de Lausanne, Switzerland
	3.	Prof. Dr. Wim Schoutens KU Leuven, Belgium (nur schriftliche Beurteilung)

Die Dissertation wurde am 23.11.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 02.03.2016 angenommen.

Abstract

Recent events like floods, hurricanes, and other environmental catastrophes have shown the importance of accounting for dependence between different types of risks in insurance modelling. Neglecting dependence can lead to severe underestimation of risk in a portfolio context. This thesis presents a realistic, yet mathematically tractable model to describe the joint behaviour of multiple claim arrival processes that are stochastically dependent. The processes are derived from independent Poisson processes by introducing a Lévy subordinator as common stochastic clock. The model supports simultaneous claim arrivals and captures the often observable phenomenon of overdispersion in claim count data. A very efficient simulation routine is available and distributional properties like the Laplace transform, probability mass function, and (mixed) moments are derived in closed form. A convenient approximation for the aggregate claim number distribution in high-dimensional applications is given as well. Furthermore, four methods for estimating the model parameters from historical data are presented and their performance is studied in a Monte Carlo simulation and using a set of real-world claim arrival data. Finally, the effect of the model on pricing and risk management of (re-)insurance products is studied and possible model extensions are discussed.

Zusammenfassung

Die Modellierung von Abhängigkeiten zwischen verschiedenen versicherungstechnischen Risiken hat aufgrund großer Schadensereignisse durch Hurrikane, großflächige Überschwemmungen und andere Naturkatastrophen in jüngster Vergangenheit stark an Bedeutung gewonnen. Werden vorhandene Abhängigkeiten im Modellierungsansatz vernachlässigt, kann dies zu einer deutlichen Unterschätzung das Gesamtrisikos führen. In dieser Arbeit wird ein multivariater Prozess für Schadenszeitpunkte vorgestellt, der eine realistische Abbildung von Abhängigkeiten ermöglicht und zudem mathematisch handhabbar ist. Der Prozess setzt sich aus unabhängigen Poisson Prozessen zusammen, welche gemeinsam in stochastischer Zeit in Gestalt eines Lévy Subordiantors durchlaufen werden. Die Konstruktion erzeugt Abhängigkeit und schafft die Möglichkeit für gebündelt auftretende Schäden, wodurch sich insbesondere Überdispersion in Schadensdaten abbilden lässt, welche in Anwendungen häufig beobachtbar ist. Ein effizienter Algorithmus für die Simulation des Modells wird vorgestellt und wichtige Verteilungskenngrößen werden geschlossen berechnet, darunter die Zähldichte selbst, die zugehörige Laplace transformierte sowie die (gemischten) Momente. Eine Näherung für die Verteilung der Gesamtschadenszahl in hochdimensionalen Anwendungen ist ebenso vorhanden. Darüber hinaus werden vier Methoden zur Schätzung der Modellparameter aus historischen Daten vorgestellt und deren Qualität im Rahmen einer Monte Carlo Simulation und anhand von realen Schadensdaten untersucht. Abschließend werden die Auswirkungen des Modells auf die Bewertung von (Rück-)Versicherungsprodukten untersucht und mögliche Modellerweiterungen diskutiert.

Acknowledgements

I am very grateful to everybody who supported me in any way in the process of writing this thesis; it is impossible to specifically mention you all here.

My sincerest thanks goes to my supervisor Prof. Dr. Matthias Scherer for his continuous support throughout the whole time. His valuable input, the many discussions, and his always open door are only a few of his essential contributions which made the whole project come together in the end.

I would also like to very much thank Prof. Dr. Rudi Zagst for all his support, particularly for giving me a position at his chair for the whole time, accommodating temporary changes, and facilitating my research trips.

Special thanks go to PD Dr. Aleksey Min and Prof. Dr. Kathrin Glau who contributed with their extensive knowledge on stochastic inference and Lévy processe to this thesis through many helpful discussion. Furthermore, my special thanks goes to Prof. Dr. Hansjörg Albrecher who not only shared his extensive knowledge on actuarial mathematics and provided many valuable hints and insights on the dynamics of the model, its applicability, as well as estimation, but who also gave me the opportunity to present my research at his chair at Université de Lausanne and even agreed to referee this thesis.

I would also like to very much thank Prof. Dr. Wim Schoutens for all I learned during his lecture at TUM, which is valuable far beyond the scope of this thesis, and for using his extensive knowledge on Lévy processes to be a referee for this thesis. Furthermore, many thanks go to Prof. Luis Seco for inviting me to present my research at his chair at the University of Toronto and for giving me very valuable feedback on my research. I would also like to take the opportunity to thank Prof. Dr. Caroline Lasser for agreeing to be the chair of the examining committee for this thesis.

In addition, my thanks go to my former colleagues at the Chair of Financial Mathematics as well as all participants in our regular hiking and skiing seminars for all their input and support. Particularly, I would like to thank my roommate Dr. Karl F. Bannör how always cheered me up with great discussions about anything and everything.

Finally, I am deeply grateful to my family and friends who were there for me through all the years.

Contents

1 Introduction

For many applications in the area of non-life insurance, for instance premium and risk reserve calculations, modelling the aggregate claim amount of a portfolio of insurance policies is essential. In the collective risk model, a point process determines the arrival times of claims and a sequence of random variables describes the claim amounts. For its natural interpretation and mathematical tractability, the classical risk model assumes a Poisson process for the claim arrivals and independent and identically distributed (iid) claim sizes which are independent of the claim number process. Claims generated from a Poisson process arrive one after another without the possibility of claim cluster arrivals, an assumption often violated in real-world applications, in particular if a portfolio or even business line outcome is concerned and multiple claims can originate from a single event. Claim cluster arrivals relate to overdispersion in claim count data, i.e. excess variance over the mean, which is often observed in real data but cannot be accommodated by the Poisson process. Hence, since the introduction of the classical risk model, multiple extensions of this modelling approach have been proposed and studied, among these the application of Cox processes for the claim number process.

Furthermore, in traditional actuarial theory individual risks are assumed to be independent. In many real-world situations, however, this is not the case, in particular if risks are subject to the same claim-generating influences, e.g. economic or environmental factors. For instance, claims in a flooding insurance portfolio for risks that are located along the same river or shore are prone to the same flooding events and thus exhibit positive dependence. Besides, a flooding event mostly causes claims to other lines of business such as private health insurance, thus generating dependencies across business sectors as well. In the presence of dependence between risks, the law of large numbers, which is at the core of classical rate making, may no longer hold and the diversification effect from pooling various risks may reduce considerably, hence, making the portfolio outcome significantly riskier than in the case of independence. If the increased riskiness is not reflected in the modelling approach, premiums and risk reserves will be underestimated, putting the solvability and continuity of the business in jeopardy. While also affecting direct insurers, the problem is intensified if non-linear reinsurance contracts come into play. In particular, in the light of regulatory frameworks such as Solvency II and the increasing complexity of (re-)insurance products such as multi-line products, the development of multivariate insurance models has gained importance in recent years. However, while there is only one way to model independence, infinitely many different dependence structures are possible and building multidimensional models that feature realistic dependence structures while

still being mathematically tractable is a matter of ongoing research, see Lindskog and McNeil (2003), Pfeiffer and Nešlehová (2004), or Bäuerle and Grübel (2005), for instance.

Every mathematical model is a simplification of a real-world phenomenon and a careful selection of the properties that have to be captured in the model is necessary. There is always a trade-off between how accurately a model describes the reality and how tractable it remains: fitting the model to available data and the computation of various quantities of actuarial interest have to be possible. In addition, it is often desirable that the model has a natural interpretation which enables an intuitive understanding of the model dynamics and, hence, allows to assess the suitability for the problem at hand beyond the mere fit to the available data.

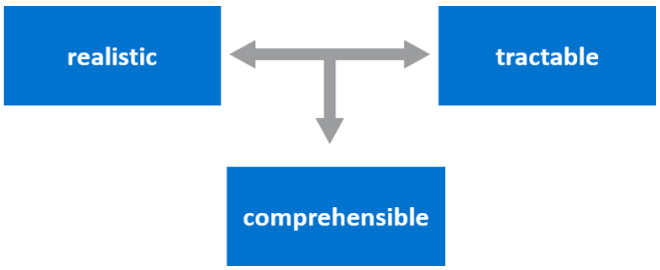


Figure 1.1 Trade-off in modelling real-world phenomena.

Dependence can arise in claim sizes or between claim sizes and claim numbers. In this thesis, however, the focus lies on modelling dependencies in multivariate claim number processes. In the area of credit default modelling, a very elegant approach based on conditionally independent processes has been successfully exploited to generate realistic credit default patterns. This approach is utilized to generate a tractable yet realistic and comprehensible multivariate model for claim arrivals featuring dependence as well as claim clusters.

The multivariate claim number process is built from independent Poisson processes using an independent Lévy subordinator as common time-change, or operational time, resulting in a multivariate Cox process. Due to the jumps of the subordinator, claims in the time-changed process can occur simultaneously and, since the subordinator affects all marginal processes, dependence is generated. This construction is an extension of the Lévy-frailty model introduced in Mai and Scherer (2009b) for modelling credit default times, where only the first arrival time in each component is relevant. Their model has proven to account for various observed default patterns, e.g. simultaneous defaults and positive lower tail dependence, while still being mathematically tractable and many of their results can be recovered for the generalized version studied in this thesis. Even though the

generalized model will be primarily studied in the context of insurance modelling, one application in credit risk modelling will be presented as well: in the well-known CreditRisk$^+$ model Poisson random variables are considered as approximation for Bernoulli default indicators and dependence is generated via gamma distributed mixing variables. Hence, the model studied here is a time-dynamic extension of the static CreditRisk$^+$ model, as will be further discussed in Section 6.2.

In this thesis, an extensive study of the proposed model, the parameter estimation from historical data, and the application to insurance modelling is given. Initial results in the field are collected and generalized in many regards; such a comprehensive treatment of this specific set-up has to the best of our knowledge not been available so far. The distributional properties of the model are investigated in detail and many quantities of interest, e.g. the Laplace transform, probability mass function, and (mixed) moments of the process distribution at any point in time, are available in closed-form. Furthermore, the process has a representation as Poisson cluster process, i.e. as compound Poisson process with non-negative discrete jump size distribution, which is explicitly known and offers a deeper understanding of the model dynamics. An efficient sampling algorithm is available as well as a large portfolio approximation. Thus, the model is highly tractable even in high dimensions. Still, it features dependence as well as overdispersion and cluster arrivals in claim count data and, hence, can account for the occurrence of catastrophic events. As the dependence generating mechanism, i.e. the common stochastic clock, is explicitly known, the model has also a very natural interpretation and its dynamics can be well understood.

Given the detailed study of the process, estimation procedures for the model parameters are formulated and tested – in a simulation study as well as on real-word data. The model has an intensity parameter for each of its components, just as independent Poisson models have. The only additional parameters compared to the standard approach are the Lévy subordinator parameters, the number of which ranges between a few and only a single one. Hence, the parameters can be expected to be determined form historical data reasonably well and without being too sensitive to single outliers.

The impact of the model on insurance premiums and other risk measures is studied as well. Whereas at a fixed point in time the process distribution resembles the one of a comparable mixed Poisson process, the dynamics over time are decisively different due to the Lévy and Poisson cluster properties. Thus, calculating insurance premiums or risk reserves for a one year time horizon does not result in different numbers, but a dynamic model offers a consistent framework over any possible time horizon and allows for the application of ruin theory. In addition, since a dynamic model provides not only information about the total outcome but also about the evolvement of the claim numbers, its suitability for a specific application can be more readily assessed and more information

for the estimation of the model parameters is available.

The basic version of the model is investigated in much detail in this thesis. This way, the properties and dynamics of the process and, hence, also its limitations can be fully understood. However, many actuarial applications may ask for additional stylized facts to be incorporated in the modelling approach. The proposed model lends itself very well to formulate such extensions and a few possibilities like adding a seasonal effect or incorporating multiple Lévy subordinators to generate a more flexible dependence structure are discussed as well. While for these extensions most quantities are no longer available in closed-form, the models are still manageable through simulation.

Whereas this thesis focuses on the application in insurance modelling, Cox and Poisson cluster processes are of interest in a variety of areas, basically wherever counting of any kind is involved. For instance, in inventory control problems they are used to describe incoming demand, see Galliher et al. (1959) and Kemp (1967). In financial modelling, Cox processes turn up in credit risk management to count defaults of credit-risky securities, for instance in Lando (1998). More recently, they have been utilized as order arrival processes for electronic trading environments, see Cont et al. (2010). Operational risk management would be another potential area for application, to name only a few.

The remainder of the thesis is organized as follows. In Chapter 2 the necessary notation and mathematical tools are introduced. In particular, the fundamental properties of the stochastic processes constituting the building blocks of the proposed model, namely the Poisson process and its generalizations as well as Lévy processes in general and subordinators in particular, are reviewed. In Chapter 3 the model is introduced and categorized. Following the model set-up, an efficient sampling algorithm is presented. Furthermore, other dependence generating modelling approaches and the current literature on the topic are reviewed. In particular, the relationship with the Lévy-frailty model for credit risk modelling is discussed. Chapter 4 studies in detail the mathematical properties of the process as well as its compound Poisson representation. Formulas for the Laplace transform, probability mass function, and moments of the distributions are derived and the dependence structure is studied. The matter of implementing the results is addressed as well. In Section 4.3, a large portfolio approximation for the aggregate claim number process in a model with increasing dimensionality is presented. Estimation of the model parameters from historical data is the topic of Chapter 5. Four estimation methods are introduced and their quality is tested in an extensive simulation study. In addition, the procedures are applied to a set of real-world claim number data, namely the Danish fire insurance data. In Chapter 6, applications and extensions of the model are discussed. Section 6.1 incorporates iid claim sizes in the model and examines the effect on important actuarial risk measures, in particular premiums for insurance and reinsurance products. Section 6.2 puts some more thought into the model with iid claim sizes as well as further model extensions, namely with an additional deterministic time transformation and with

multiple subordinators. The latter extension leads to a time-dynamic version of the famous CreditRisk$^+$ model which is briefly discussed as well. Finally, Chapter 7 concludes and suggests some related questions for further research.

2 Mathematical preliminaries

This chapter introduces the fundamental mathematical concepts and tools for the remainder of the thesis. In Section 2.1 the notation is established and fundamental knowledge about random vectors and stochastic processes is reviewed. In this thesis a multivariate claim number process constructed from Poisson processes and a Lévy subordinator is studied. For a rigorous treatment of this process, a sound understanding of the Poisson process and its extensions is essential; the relevant material is covered in Section 2.2 of this chapter. Mostly, claim number processes are studied using the theory and terminology of general point processes, which is also briefly discussed in the second section. The particular process examined in this thesis, however, also belongs to the class of Lévy processes and will be mainly studied from this perspective. Section 2.3 therefore gives a review of the fundamental properties of such processes. Overall, this chapter attempts to give a sound overview of the processes and further mathematical objects relevant for this thesis; for a thorough treatment further references are given within each section.

2.1 Notation and mathematical tools

We start by establishing the notation for the remainder of the thesis. Let $\mathbb{N}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ denote the set of positive and non-negative integers, respectively. For the real numbers we use $\mathbb{R}$, for the non-negative subset $\mathbb{R}_{\geq 0}$, and for the positive subset $\mathbb{R}_{>0}$. Throughout the thesis, $d \in \mathbb{N}$ is some positive integer usually specifying the dimension. We use a dot to indicate that a space is punctured at the origin:

$$\dot{\mathbb{R}}^d_{\geq 0} := \mathbb{R}^d_{\geq 0} \setminus \{\mathbf{0}\},$$

where $\mathbf{0} \in \mathbb{R}^d$ is the vector with d zero entries, the dimension being either clear from the context or specifically stated. In general, vectors are considered as column vectors and mainly written in bold font, that is

$$\boldsymbol{x} := (x_1, \ldots, x_d)' \in \mathbb{R}^d,$$

where $(\cdot)'$ is denoting the transpose of the vector. The i-th unit vector for some $1 \leq i \leq d$ is given by $\boldsymbol{e}_i$ and

$$|\boldsymbol{x}| := |x_1| + \cdots + |x_d|, \quad \|\boldsymbol{x}\| := \sqrt{x_1^2 + \cdots + x_d^2}$$

denote the 1-norm and the Euclidean norm, respectively. For any scalar $y \in \mathbb{R}$ and vector $\boldsymbol{y} \in \mathbb{R}^d$, notations like

$$\boldsymbol{x} \leq y, \quad \boldsymbol{x} \leq \boldsymbol{y} \quad \text{or} \quad \boldsymbol{x} = y, \quad \boldsymbol{x} = \boldsymbol{y}$$

are to be understood component-wise. This applies to $\boldsymbol{x}^k$, $k \in \mathbb{N}_0$, as well:

$$\boldsymbol{x}^k := (x_1^k, \ldots, x_d^k)'.$$

Furthermore, multi-index notation is used, that is for any d-tuple $\boldsymbol{k} \in \mathbb{N}_0^d$:

$$\boldsymbol{x}^{\boldsymbol{k}} := x_1^{k_1} \ldots x_d^{k_d} \quad \text{and} \quad \boldsymbol{k}! := k_1! \ldots k_d!,$$

and for the multinomial coefficient:

$$\binom{|\boldsymbol{k}|}{\boldsymbol{k}} := \frac{|\boldsymbol{k}|!}{\boldsymbol{k}!} = \frac{(k_1 + \cdots + k_d)!}{k_1! \ldots k_d!}.$$

For a sequence of real vectors $\{\boldsymbol{x}_n\}_{n \in \mathbb{N}_0}$ the indexing

$$\boldsymbol{x}_n := (x_{1n}, \ldots, x_{dn})'$$

is used and the difference operator is defined by

$$\Delta \boldsymbol{x}_n := \boldsymbol{x}_n - \boldsymbol{x}_{n-1}, \quad n \in \mathbb{N}.$$

For real-valued functions $f(\boldsymbol{x})$ with domain $\mathbb{R}^d$, higher order partial derivatives are denoted by

$$f^{(\boldsymbol{k})}(\boldsymbol{x}) := \frac{\partial^{|\boldsymbol{k}|} f}{\partial^{k_1} x_1 \dots \partial^{k_d} x_d}(\boldsymbol{x}),$$

and for the evaluation at a certain point $\boldsymbol{x}^* \in \mathbb{R}^d$ the short-hand notation

$$f^{(\boldsymbol{k})}(\boldsymbol{x}^*) := f^{(\boldsymbol{k})}(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*}$$

is used. For any subset $A \subset \mathbb{R}^d$, the cardinality of A is given by $|A|$ and the indicator function of A, that is the function returning 1 for any argument in A and zero otherwise, by $\mathbf{1}_A(\cdot)$.

The Lebesgue measure on the d-dimensional Euclidean space equipped with the corresponding Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ is denoted by $Leb(\cdot)$; the dimension will again be clear from the context. In general, on a measurable space (E, Σ) the Dirac measure in some point $c \in E$ is

$$\delta_c \colon \Sigma \to \{0,1\}, \quad A \mapsto \begin{cases} 1 & \text{if } c \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Throughout this thesis all random objects live on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with sample space Ω, σ-algebra $\mathcal{F}$, and probability measure $\mathbb{P}$. For an $\mathbb{R}^d$-valued random variable $\boldsymbol{X} := (X_1, \dots, X_d)'$, let $\mathbb{P}_{\boldsymbol{X}}$ denote the push-forward measure induced by $\boldsymbol{X}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $F_{\boldsymbol{X}}$ the (cumulative) distribution function. As usual, the expectation and covariance operator are

$$\mathbb{E}[\boldsymbol{X}] := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])', \quad \mathbb{C}\mathrm{ov}[\boldsymbol{X}, \boldsymbol{Y}] := (\mathbb{C}\mathrm{ov}[X_i, Y_j])_{1 \le i,j \le d}.$$

Variance $\mathbb{V}\mathrm{ar}[\boldsymbol{X}]$ and correlation $\mathbb{C}\mathrm{or}[\boldsymbol{X}, \boldsymbol{Y}]$ are defined accordingly. Given some σ-algebra $\mathcal{G} \subset \mathcal{F}$, the conditional expectation of $\boldsymbol{X}$ given the information in $\mathcal{G}$ is written as $\mathbb{E}[\boldsymbol{X}|\mathcal{G}]$.

Let F be a univariate distribution function. The generalized inverse or quantile function of F is

$$F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \ge u\}, \quad u \in [0,1],$$

using the convention $\inf \emptyset := -\infty$. A detailed discussion of the properties of generalized inverse functions is given in Embrechts and Hofert (2013). A copula $C\colon [0,1]^d \to [0,1]$ is the distribution function of a d-dimensional random variable with uniform U$[0,1]$ marginals. Following Sklar's theorem, which was originally stated in Sklar (1959), a copula $C_{\boldsymbol{X}}$ exists for any random vector $\boldsymbol{X}$ on $\mathbb{R}^d$ such that

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = C_{\boldsymbol{X}}(F_{X_1}(x_1), \dots, F_{X_d}(x_d)), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

If the marginal distributions are continuous, the copula is unique. On the other hand, given some copula $C_{\boldsymbol{X}}$ as well as d univariate distribution functions $F_{X_1}, \dots, F_{X_d}$, the function defined by the right-hand side of the above equation defines a proper distribution function. Details on copulas can be found in Nelsen (2006) or Mai and Scherer (2012b). These references also provide more details on the dependence measures mentioned briefly in the following.

Let $\boldsymbol{X} = (X_{11}, X_{21})'$ be a bivariate random variable. Two measures for the degree of monotonic dependence are Spearman's rho and Kendall's tau. Spearman's rho is defined as

$$\rho_S(X_1, X_2) := \mathbb{C}\mathrm{or}[F_{X_1}(X_1), F_{X_2}(X_2)].$$

Kendall's tau measures the difference between the probability of concordance and discordance. Let $\boldsymbol{X}_1 = (X_{11}, X_{21})'$, $\boldsymbol{X}_2 = (X_{12}, X_{22})'$ be independent and identically distributed (iid) copies of $\boldsymbol{X}$, then:

$$\rho_\tau(X_1, X_2) := \mathbb{P}((X_{11} - X_{12})(X_{21} - X_{22}) > 0) - \mathbb{P}((X_{11} - X_{12})(X_{21} - X_{22}) < 0).$$

Both definitions are extended to d-dimensional random vectors by considering the matrix of pairwise measures, comparable to the definition of a correlation matrix. A measure of extremal dependence is given by the coefficients of upper and lower tail dependence. These are defined for a continuously distributed bivariate random vector $\boldsymbol{X}$ as

$$\begin{aligned} UTD(X_1, X_2) &:= \lim_{u \uparrow 1} \mathbb{P}(X_1 > F_{X_1}^{-1}(u) | X_2 > F_{X_2}^{-1}(u)), \\ LTD(X_1, X_2) &:= \lim_{u \downarrow 0} \mathbb{P}(X_1 \leq F_{X_1}^{-1}(u) | X_2 \leq F_{X_2}^{-1}(u)), \end{aligned}$$

given the limits in $[0, 1]$ exist. Note that in this definition X_1 and X_2 are interchangeable. If the coefficients are larger than (equal to) zero, then X_1 and X_2 are called asymptotically dependent (independent) in the upper or lower tail, respectively.

Comparing any two random objects, the notation '$\stackrel{d}{=}$' means that the objects on the left- and right-hand side have the same distribution. For random vectors $\boldsymbol{X}$, $\boldsymbol{Y}$ on $\mathbb{R}^d$ this corresponds to equality of the distribution functions, see (Kallenberg, 2002, Chapter 3, Lemma 3.3, p.48):

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y} \quad \Leftrightarrow \quad F_{\boldsymbol{X}} = F_{\boldsymbol{Y}}.$$

A sequence of random vectors $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ is said to converge in distribution to a random vector $\boldsymbol{X}$ for $n \to \infty$, written $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$, if the induced probability measures $\mathbb{P}_{\boldsymbol{X}^n}$ converge weakly to $\mathbb{P}_{\boldsymbol{X}}$, written $\mathbb{P}_{\boldsymbol{X}_n} \xrightarrow{w} \mathbb{P}_{\boldsymbol{X}}$, that is if

$$\lim_{n\to\infty} \mathbb{E}[f(\boldsymbol{X}_n)] = \mathbb{E}[f(\boldsymbol{X})], \quad \forall f \in \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}),$$

where $C_b(\mathbb{R}^d, \mathbb{R})$ is the set of all continuous and bounded functions $f\colon \mathbb{R}^d \to \mathbb{R}$.

In the univariate case, let $L^p(\Omega, \mathcal{F}, \mathbb{P})$ for some $p > 0$ be the set of all random variables X with finite absolute p-th mean, that is $\mathbb{E}[|X|^p] < \infty$. Convergence of a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ in L^p to a random variable $X \in L^p$ in the p-the mean is defined as:

$$X_n \xrightarrow{L^p} X, \quad n \to \infty \quad :\Leftrightarrow \quad \lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

Furthermore, convergence in distribution is defined as follows:

$$X_n \xrightarrow{\mathbb{P}} X \quad :\Leftrightarrow \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0, \quad \forall \epsilon > 0.$$

The characteristic function of a d-dimensional random vector $\boldsymbol{X}$, which always exists, is denoted by

$$\Phi_{\boldsymbol{X}}(\boldsymbol{v}) := \mathbb{E}[\exp\{\mathrm{i}\boldsymbol{v}'\boldsymbol{X}\}], \quad \boldsymbol{v} \in \mathbb{R}^d,$$

with i being the imaginary unit. This function completely determines the distribution of $\boldsymbol{X}$ and moments – if existent – can be derived from derivatives of the characteristic function evaluated at the origin. For non-negative random vectors $\boldsymbol{Y}$ on $\mathbb{R}^d_{\geq 0}$, it is more convenient to work with the Laplace transform:

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{u}) := \mathbb{E}[\exp\{-\boldsymbol{u}'\boldsymbol{Y}\}], \quad \boldsymbol{u} \in \mathbb{R}^d_{\geq 0}.$$

As we will mainly be concerned with non-negative random objects in the following, the main properties of the Laplace transform are presented in more detail.

Theorem 2.1 (Properties of the Laplace transform)
Let $\boldsymbol{Y}$ be a non-negative random vector on $\mathbb{R}^d_{\geq 0}$ with Laplace transform $\varphi_{\boldsymbol{Y}}(\cdot)$.

(i) Linear transformation: For all $A \in \mathbb{R}^{d\times d}_{\geq 0}$, $\boldsymbol{b} \in \mathbb{R}^d_{\geq 0}$, the Laplace transform of the linear transform $A\boldsymbol{Y} + \boldsymbol{b}$ fulfils:

$$\varphi_{A\boldsymbol{Y}+\boldsymbol{b}}(\boldsymbol{u}) = \exp\{-\boldsymbol{u}'\boldsymbol{b}\}\varphi_{\boldsymbol{Y}}(A'\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^d_{\geq 0}.$$

(ii) Multiplication rule: If $\boldsymbol{Z}$ is a random vector on $\mathbb{R}^d_{\geq 0}$ independent of $\boldsymbol{Y}$, then it holds for the Laplace transform of the convolution $\boldsymbol{Y} + \boldsymbol{Z}$:

$$\varphi_{\boldsymbol{Y}+\boldsymbol{Z}}(\boldsymbol{u}) = \varphi_{\boldsymbol{Y}}(\boldsymbol{u})\varphi_{\boldsymbol{Z}}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^d_{\geq 0}.$$

(iii) Uniqueness: The distribution of $\boldsymbol{Y}$ is uniquely determined by its Laplace transform. Let $\boldsymbol{Z}^*$ be another random vector on $\mathbb{R}^d_{\geq 0}$, then:

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{Z}^* \quad \Leftrightarrow \quad \varphi_{\boldsymbol{Y}}(\boldsymbol{u}) = \varphi_{\boldsymbol{Z}^*}(\boldsymbol{u}), \quad \forall \boldsymbol{u} \in \mathbb{R}^d_{\geq 0}.$$

(iv) Lévy's continuity theorem: Let $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ be a sequence of d-dimensional non-negative random vectors. The sequence converges in distribution to a random vector $\boldsymbol{Y}$ on $\mathbb{R}^d_{\geq 0}$ iff the Laplace transforms converge pointwise:

$$\boldsymbol{Y}_n \xrightarrow{d} \boldsymbol{Y}, \quad n \to \infty \quad \Leftrightarrow \quad \lim_{n\to\infty} \varphi_{\boldsymbol{Y}_n}(\boldsymbol{u}) = \varphi_{\boldsymbol{Y}}(\boldsymbol{u}), \quad \forall \boldsymbol{u} \in \mathbb{R}^d_{\geq 0}.$$

Proof. The properties of the Laplace transform in the univariate case are studied in detail in (Feller, 1971, Chapter XIII.1-4). In the multivariate case, statement (i) can easily be seen from the definition of the Laplace transform and statement (ii) follows by calculation from the univariate case, see also (Kallenberg, 2002, Chapter 5, p.84). Statement (iv) can be concluded from the univariate case using the Cramér–Wold device which states that $\boldsymbol{Y}_n \xrightarrow{d} \boldsymbol{Y}$ iff for all $u \in \mathbb{R}_{\geq 0}$ it holds that $u'\boldsymbol{Y}_n \xrightarrow{d} u'\boldsymbol{Y}$, see (Kallenberg, 2002, Chapter 5, Theorem 5.5, p.87). It follows also from (Kallenberg, 2002, Chapter 5, Theorem 5.3, p.86). Finally, statement (iii) is a consequence of statement (iv). □

Of particular interest in the following will be the derivatives of a Laplace transform; here it will be sufficient to consider the univariate case.

Theorem 2.2 (Derivatives of the Laplace transform)
The Laplace transform $\varphi_Y(\cdot)$ of a non-negative univariate random variable Y is completely monotone, that is derivatives of all order $n \in \mathbb{N}_0$ exist and alternate in sign:

$$(-1)^n \varphi_Y^{(n)}(u) \geq 0, \quad u > 0.$$

Moreover, the derivatives have the representation:

$$(-1)^n \varphi_Y^{(n)}(u) = \mathbb{E}[Y^n \exp\{-uY\}], \quad u > 0.$$

For $u \to 0$, the derivatives converge to a finite or infinite limit. This limit is finite iff the n-th moment of Y exists and is finite; then it holds:

$$\mathbb{E}[Y^n] = (-1)^n \varphi_Y^{(n)}(0).$$

Proof. The results can, for instance, be found in (Feller, 1971, Chapter XIII.2, p.430). □

An even stronger statement regarding complete monotonicity and Laplace transforms can be formulated: a function $\varphi(\cdot)$ on $\mathbb{R}_{\geq 0}$ is the Laplace transform of a probability distribution iff it is completely monotone and $\varphi(0) = 1$. This result goes back to Bernstein (1929) and a proof can be found in (Feller, 1971, Chapter XIII.4, Theorem 1, pp.439). A detailed study of completely monotone functions can be found in Schilling et al. (2012).

For any random vector $\boldsymbol{X}$, the cumulant generating function is defined as the logarithm of the moment generating function:

$$g_{\boldsymbol{X}}(\boldsymbol{v}) := \log \mathbb{E}[\exp\{\boldsymbol{v}'\boldsymbol{X}\}], \quad \boldsymbol{v} \in \mathbb{R}^d.$$

Formally, it can be written as $g_{\boldsymbol{X}}(\boldsymbol{v}) = \log \Phi_{\boldsymbol{X}}(-i\boldsymbol{v})$. Accordingly, for a non-negative random vector $\boldsymbol{Y}$ it can be written $g_{\boldsymbol{Y}}(\boldsymbol{u}) = \log \varphi_{\boldsymbol{Y}}(-\boldsymbol{u})$. For dimension $d = 1$, the so-called cumulants are obtained by evaluating the derivatives of the cumulant generating function at zero:

$$\kappa_n := g_X^{(n)}(0), \quad n \in \mathbb{N}.$$

If existent, (central) moments of the distribution can be derived from the cumulants and vice versa, see (Johnson et al., 1992, Chapter B5 and B6, pp.40) for details.

Following (Rolski et al., 1999, Chapter 2.5, p.49), we call a random variable X heavy-tailed if

$$\mathbb{E}[\exp\{uX\}] = \infty, \quad \forall u > 0.$$

Following this definition, the tails of heavy-tailed distributions have no exponential bound, that is for all $s > 0$:

$$\lim_{u\to\infty} \exp\{su\}\mathbb{P}(X > u) = \infty.$$

Compared to 'well-behaved' distributions where the probability of large values decreases exponentially fast to zero, heavy-tailed distributions are regarded as 'dangerous' in insurance applications as huge values are more likely.

A stochastic process $\boldsymbol{X} = \{\boldsymbol{X}_t\}_{t\geq 0}$ on $\mathbb{R}^d$ is used to model how a stochastic system evolves over time and can be understood either as a family of random vectors $\boldsymbol{X}_t$ or as a function-valued random object, that is a measurable map $\boldsymbol{X}\colon \Omega \to U \subset \mathbb{R}^{d\times\mathbb{R}_{\geq 0}}$. The sample paths are denoted by $\boldsymbol{X}(\omega)\colon t \mapsto \boldsymbol{X}_t(\omega)$. As the time index is given in subscript, the components or marginal processes $X^i := \{X_t^i\}_{t\geq 0}$, $i = 1, \ldots, d$, are indexed in superscript. For a multi-parameter time $\boldsymbol{t} \in \mathbb{R}^d_{\geq 0}$, the multi-parameter process is denoted by

$$\boldsymbol{X}_{\boldsymbol{t}} := (X^1_{t^1}, \ldots, X^d_{t^d})'.$$

This notation is used for a univariate process $X = \{X_t\}_{t\geq 0}$ as well, that is

$$\boldsymbol{X_t} := (X_{t_1}, \ldots, X_{t_d})'.$$

Furthermore, let $\sigma(\boldsymbol{X}) := \sigma(\{\boldsymbol{X}_t : t \geq 0\})$ denote the completed σ-algebra generated by the paths of $\boldsymbol{X}$.

For a second stochastic process $\boldsymbol{Y} = \{\boldsymbol{Y}\}_{t\geq 0}$ on $\mathbb{R}^d$ we write again $\boldsymbol{X} \overset{d}{=} \boldsymbol{Y}$ if the processes are identical in law. The distribution of a process is determined by the system of finite-dimensional distributions, i.e.

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{Y} \quad \Leftrightarrow \quad (\boldsymbol{X}_{t_1}, \ldots, \boldsymbol{X}_{t_k}) \overset{d}{=} (\boldsymbol{Y}_{t_1}, \ldots, \boldsymbol{Y}_{t_k}), \quad \forall \boldsymbol{t} \in \mathbb{R}^k_{\geq 0}, \quad \forall k \in \mathbb{N},$$

see (Kallenberg, 2002, Chapter 3, Proposition 3.2, p.48). A sequence $\{\boldsymbol{X}^n\}_{n\in\mathbb{N}}$ of stochastic processes is said to converge to a stochastic process $\boldsymbol{X}$ in finite dimensional distributions, written $\boldsymbol{X}^n \xrightarrow{fdd} \boldsymbol{X}$, if the finite dimensional distributions converge:

$$(\boldsymbol{X}^n_{t_1}, \ldots, \boldsymbol{X}^n_{t_k}) \xrightarrow{d} (\boldsymbol{X}_{t_1}, \ldots, \boldsymbol{X}_{t_k}), \quad n \to \infty, \quad \forall \boldsymbol{t} \in \mathbb{R}^k_{\geq 0}, \quad \forall k \in \mathbb{N}.$$

Convergence in distribution for stochastic processes is defined accordingly to the definition for random vectors as weak convergence of the respective probability measures, that is

$$\boldsymbol{X}^n \xrightarrow{d} \boldsymbol{X}, \quad n \to \infty \quad :\Leftrightarrow \quad \lim_{n\to\infty} \mathbb{E}[f(\boldsymbol{X}^n)] = \mathbb{E}[f(\boldsymbol{X})], \quad \forall f \in \mathcal{C}_b(U, \mathbb{R}),$$

where $C_b(U, \mathbb{R})$ is the space of continuous and bounded functions on the function space $U \subset \mathbb{R}^{d\times\mathbb{R}_{\geq 0}}$. However, we need to choose a topology on the infinite-dimensional space U to fix the notion of continuity (and convergence).

If U is the space of continuous functions, that is stochastic processes on U have continuous sample paths, the space is usually equipped with the supremum norm which induces the topology of uniform convergence. In the following, the wider class of processes with càdlàg paths is studied, thus for U the space of càdlàg functions, which are right-continuous with existing left limits, are considered, written $\mathcal{D}(\mathbb{R}_{\geq 0}, \mathbb{R}^d)$. This space is usually endowed with the Skorohod J_1-topology that was first introduced in Skorokhod (1956) and is together with the corresponding Borel σ-algebra known as the Skorohod

space. Consider the restriction $\mathcal{D}([0,T],\mathbb{R}^d)$ for some time horizon $T > 0$. Then functions $f_n \in \mathcal{D}([0,T],\mathbb{R}^d)$, $n \in \mathbb{N}$, converge to some function $f \in \mathcal{D}([0,T],\mathbb{R}^d\})$ in the J_1-topology if for all $n \in \mathbb{N}$ there exists some monotone bijection $h_n \colon [0,T] \to [0,T]$ with $h(0) = 0$ and $h(T) = T$ such that

$$\sup_{0\le t\le T} |h_n(s) - s| + \sup_{0\le t\le T} \|f_n(h_n(s)) - f(s)\| \longrightarrow 0, \quad n \to \infty.$$

If $h_n(t) = t$ can be chosen, this definition reduces to the one of uniform convergence. In general, the notion of uniform convergence is too strong to accommodate jumps in càdlàg functions. Using the transformation h_n time is allowed to wiggle a bit – the 'a bit' being specified by the first part of the upper expression – and the corresponding wiggle in space gives some relaxation to the requirement on the convergence of jumps: while they still need to be of almost the same size, they do not need to occur at exactly the same time. More details can be found in Pollard (1984), Billingsley (2009), and Jacod and Shiryaev (2003).

It should be noted that even though the finite dimensional distributions determine the process distribution, convergence of the finite dimensional distributions of processes does in general not lead to convergence in distribution; for this, the sequence needs to fulfil some additional tightness condition, see (Kallenberg, 2002, Chapter 16).

2.2 The Poisson process and its generalizations

In this thesis we are interested in modelling the arrivals of claims over time in actuarial applications. The standard processes for this purpose are still the Poisson process and its generalizations, which are reviewed in this section. The Poisson process is studied in many textbooks on actuarial science and statistics, see e.g. Panjer et al. (1992) and Mikosch (2009). See also Kingman, J. F. C. (1964), Grandell (1997), and Grandell (1991) for more details.

The Poisson process is a stochastic process based on the Poisson distribution. A random variable N is Poisson distributed with intensity $\lambda > 0$, written $N \sim \text{Poi}(\lambda)$, if

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!} \exp\{-\lambda\}, \quad k \in \mathbb{N}_0.$$

The Poisson distribution is the limiting case of the binomial distribution $B(n, \frac{\lambda}{n})$ if the number n of trials or events converges to infinity while the success or occurrence rate shrinks as $\frac{\lambda}{n}$. Due to this property the Poisson distribution is sometimes called the law of small numbers: it is the distribution of the number of occurrences if many events are possible but each happens only rarely.

One important property of the distribution is the equality of mean and variance:

$$\mathbb{E}[N] = \lambda = \mathbb{V}\text{ar}[N].$$

Thus, the distribution has a variance to mean ration – called dispersion index – of one. As the Poisson distribution is the benchmark distribution for count data, other counting distributions are said to be overdispersed or underdispersed if they have a dispersion index greater or smaller than one, respectively.

We start with the general definition of the Poisson process before having a closer look at the homogeneous and inhomogeneous case.

Definition 2.3 (Poisson process)
A Poisson process is a stochastic process $N = \{N_t\}_{t \geq 0}$ with the following properties:

(i) Càdlàg paths: The paths of N are almost surely (a.s.) càdlàg functions, that is right-continuous with existing left limits.

(ii) Start at zero: $N_0 = 0$ a.s.

(iii) Independent increments: For any $0 \leq t_0 < t_1 < \cdots < t_n < \infty$, $n \in \mathbb{N}$, the increments ΔN_{t_i} are mutually independent for $i = 1, \ldots, n$.

(iv) Poisson increments: For a càdlàg function $\mu \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ with $\mu(t) < \infty$ for all $t \geq 0$, the increments have the following distribution:

$$N_t - N_s \sim \mathrm{Poi}(\mu(t) - \mu(s)), \quad 0 \leq s < t < \infty.$$

The function μ is called mean-value function of N.

It comes naturally to call μ the mean-value function as it describes the expectation of the process increments:

$$\mathbb{E}[N_t - N_s] = \mu(t) - \mu(s), \quad 0 \leq s < t < \infty.$$

The classical Poisson process is the homogeneous Poisson process with linear mean-value function,

$$\mu(t) = \lambda t, \quad t \geq 0,$$

for some $\lambda > 0$ called intensity. In this case, the increments are not only independent but also stationary, that is

$$N_t - N_s \overset{d}{=} N_{t-s}, \quad 0 \leq s < t < \infty,$$

and N is a Lévy process, see Section 2.3. In the following, if not explicitly referred to as inhomogeneous, a Poisson process is assumed to be homogeneous.

Though often not flexible enough, the Poisson process still serves as benchmark process for counting claim arrivals in actuarial applications where it was first introduced by Filip Lundberg in 1903. In the context of the collective risk model it was later studied extensively in Harald Cramér's work in the 1930s. As it will serve as starting point for the claim number process studied in this thesis, a few important properties are summarized in the following theorem.

Theorem 2.4 (Properties of the homogeneous Poisson process)
Let $N = \{N_t\}_{t\geq 0}$ be a homogeneous Poisson process with intensity $\lambda > 0$ and let

$$T_j := \inf\{t > 0 : N_t \geq j\}, \quad j \in \mathbb{N},$$

be the jump arrival times. Then the following properties hold:

(i) Inter-arrival times: The inter-arrival times $E_j := \Delta T_j$, $j \in \mathbb{N}$, where $T_0 := 0$, follow an exponential distribution with mean λ^{-1}:

$$E_j \overset{iid}{\sim} \text{Exp}(\lambda), \quad j \in \mathbb{N}.$$

(ii) Order-statistics property: The conditional distribution of $(T_1, \ldots, T_n)$ given $\{N_t = n\}$ for some $n \in \mathbb{N}$ equals the distribution of the order statistics $U_{(1)}, \ldots, U_{(n)}$ of iid random variables $U_i \overset{iid}{\sim} \text{U}[0, t]$ uniformly distributed on the respective time interval.

(iii) Superposition of Poisson processes: Let $\tilde{N}$ be another Poisson process with intensity $\tilde{\lambda} > 0$ independent of N, then $N + \tilde{N}$ is again a Poisson process with intensity $\lambda + \tilde{\lambda}$.

(iv) Early jump arrivals:

$$\mathbb{P}(N_t = k) = \begin{cases} 1 - \lambda t + o(t) & \text{if } k = 0, \\ \lambda t + o(t) & \text{if } k = 1, \\ o(t) & \text{otherwise,} \end{cases}$$

with the limiting behaviour $\frac{o(t)}{t} \to 0$ for $t \downarrow 0$.

Proof. For (i) and (ii) see (Mikosch, 2009, Chapter 2.1.4, Theorem 2.1.6, p.16; Chapter 2.1.6, Theorem 2.1.11, p.25). Property (iii) is discussed in (Cont and Tankov, 2003, Chapter 2.5.3, Proposition 2.13, p.52), and (iv) can be found in (Denuit et al., 2007, Chapter 1.3, p.21). □

The Poisson process can equally be defined as renewal processes based on its independent and exponentially distributed inter-arrival times, i.e.

$$N_t := \sum_{j=1}^{\infty} \mathbb{1}_{\{E_1 + \cdots + E_j \leq t\}},$$

see (Billingsley, 2009, Section 23, pp.316); then the process is automatically càdlàg. From this representation it is obvious that a.s. no jumps larger than size one occur and, hence, that N is a simple counting process (that is right-continuous with jumps in $\{0, 1\}$). Furthermore, independent Poisson processes a.s. never jump together, see (Kingman, J. F. C., 1964, Chapter 2.2, Disjointness Lemma, p.14).

The inhomogeneous Poisson process can be defined through a deterministic time-change of a homogeneous Poisson process. Let N be a homogeneous Poisson precess with intensity $\lambda = 1$ and $\mu(\cdot)$ a valid mean-value function according to Definition 2.3, then the process defined by $N_{\mu(t)}, t \geq 0$, is easily understood to be an inhomogeneous Poisson process with mean-value function μ. Conversely, every inhomogeneous Poisson process has a representation as time-changed Poisson process, see (Mikosch, 2009, Chapter 2.1.3, pp.14). For this reason μ is sometimes called operational time, see e.g. (Bühlmann, 1970, Chapter 2.2.3., pp.49): whereas time runs linearly for a homogeneous Poisson process, it speeds up or slows down according to μ for an inhomogeneous Poisson process.

If μ is continuous and strictly increasing with $\mu(t) \to \infty$ for $t \to \infty$, then the inverse function μ^{-1} exists and the inhomogeneous Poisson process can be converted back to a homogeneous one with intensity 1 by a time-change with μ^{-1}, see (Mikosch, 2009, Chapter 2.1.3, Proposition 2.1.5, p.15). This is often beneficial as working with the homogeneous process is more convenient. If μ is not continuous and μ^{-1} denotes only the generalized inverse, then $\mu \circ \mu^{-1}$ is no longer the identity function, thus switching between the inhomogeneous and homogeneous process no longer works.

In many applications it even is assumed that μ is absolutely continuous, that is a non-negative measurable function $\lambda(\cdot)$ exists such that

$$\mu(t) = \int_0^t \lambda(s)\,\mathrm{d}s, \quad t \geq 0.$$

The function $\lambda(\cdot)$ is called intensity function and for a homogeneous Poisson process with intensity $\lambda > 0$ it is $\lambda(t) = \lambda$, $t \geq 0$.

Like the homogeneous Poisson process, the inhomogeneous Poisson process can be represented by its jump arrival times. However, in any case other than the homogeneous one the inter-arrival times are no longer independent or identically distributed and the process is no longer a renewal process, see (Mikosch, 2009, Chapter 2.1.5, p.22). Furthermore, given discontinuities in the mean-value function, jumps of sizes larger than one are possible and in this case the process is no longer a simple counting process.

(Inhomogeneous) Poisson processes are often studied as general point processes, i.e. as a random measures on the non-negative integers, and referred to as Poisson random measures (PRM). In this case, the definition readily extends to the multivariate case.

Definition 2.5 (Poisson random measure (PRM))
Consider a measurable space $(E, \mathcal{E})$ where $E \subset \mathbb{R}^d$ and $\mathcal{E} := \mathcal{B}(E)$. Let μ be a (non-negative) Radon measure, that is a measure with $\mu(A) < \infty$ for all compact sets $A \in \mathcal{E}$. Then a PRM with mean measure μ is a mapping

$$N \colon \mathcal{E} \times \Omega \to \mathbb{N}_0, \quad (A, \omega) \mapsto N(A, \omega),$$

such that it holds:

(i) For almost all $\omega \in \Omega$, $N(\cdot, \omega)$ is an integer-valued Radon measure and for all measurable sets $A \in \mathcal{E}$, $N(A, \cdot) < \infty$ is an integer-valued random variable.

(ii) For all sets $A \in \mathcal{E}$ it is $N(A) \sim \mathrm{Poi}(\mu(A))$.

(iii) For any pairwise disjoint sets $A_1, \ldots, A_n \in \mathcal{E}$, $n \in \mathbb{N}$, the random variables $N(A_j)$, $j = 1, \ldots, n$, are mutually independent.

If μ is absolutely continuous with respect to the Lebesgue measure, then the non-negative function $\lambda(\cdot)$ with

$$\mu(A) = \int_A \lambda(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}, \quad A \in \mathcal{E},$$

is called the intensity measure. If $\mu(\cdot) = \lambda Leb(\cdot \cap E)$ for the Lebesgue measure $Leb(\cdot)$ and some $\lambda > 0$, then N is called a homogeneous PRM.

For $E = \mathbb{R}$, the PRM N with mean value measure μ corresponds to the Poisson process with mean-value function specified as

$$N_t := N(0, t], \quad \mu(t) := \mu(0, t], \quad t > 0.$$

On the other hand, given a Poisson process N with mean-value function μ, the (random) measures of the PRM representation are specified via the generating sets of $\mathcal{B}(E)$ as:

$$N(s,t] := N_t - N_s, \quad \mu(s,t] := \mu(t) - \mu(s), \quad 0 \leq s < t < \infty,$$

see also (Kingman, J. F. C., 1964, Chapter 2.1, p.14).

For any PRM N exists a random sequence $\{T_j\}_{j\in\mathbb{N}}$ in E such that N has a representation in terms of Dirac measures as

$$N(A) = \sum_{j=1}^{\infty} \delta_{T_j}(A), \quad A \in \mathcal{E},$$

see (Cont and Tankov, 2003, Chapter 2.6.1, p.58). For $E = \mathbb{R}$, these points can be associated to the jump arrival times in the process representation. Given a univariate homogeneous Poisson process, the arrival times can be represented as increasing partial sums of iid exponential random variables, see Theorem 2.4. An introduction to point processes in general and PRMs in particular can be found in Mikosch (2009) or Embrechts et al. (1997). A more rigorous treatment is, for instance, given in Resnick (1987) or Daley and Vere-Jones (2003, 2007).

The Poisson distribution, though very intuitive in the context of insurance modelling, not always fits the observed claim count data. As discussed in the beginning, the Poisson distribution has a variance to mean ratio of one and can therefore often not match the variability observed in claim count data. One common way to create greater flexibility is randomizing the intensity parameter; the resulting distributions are called mixed Poisson distributions. More precisely, a random variable M is said to be mixed Poisson distributed if a non-negative random variable $\Lambda \geq 0$ exists such that the probability mass function of M is given by the expectations of Poisson probabilities with random parameter Λ:

$$\mathbb{P}(M = k) = \mathbb{E}[\frac{\Lambda^k}{k!} \exp\{-\Lambda\}], \quad k \in \mathbb{N}_0.$$

These distributions provide a good model for heterogeneous populations: the claim frequency varies with regard to the random effect Λ and can therefore account for the differences in the individual claim frequencies. By applying the tower rule for conditional expectations it can be found:

$$\mathbb{E}[M] = \mathbb{E}[\Lambda], \quad \mathbb{V}\mathrm{ar}[M] = \mathbb{E}[\Lambda] + \mathbb{V}\mathrm{ar}[\Lambda].$$

It follows that these distributions have a variance to mean ratio larger than one, i.e. they are suitable for overdispersed claim count data.

On a process level, this approach corresponds to randomizing the mean-value function.

Definition 2.6 (Mixed Poisson and Cox process)
Let $N = \{N_t\}_{t\geq 0}$ be a homogeneous Poisson process with intensity $\lambda = 1$. Independently, let $\Lambda = \{\Lambda_t\}_{t\geq 0}$ be a stochastic process that has a.s. non-decreasing càdlàg paths with $\Lambda_0 = 0$ and $\Lambda_t < \infty, t < \infty$. Then the process $\{N_{\Lambda_t}\}_{t\geq 0}$ is called a Cox process directed by Λ. If there exists a non-negative process $\lambda = \{\lambda_t\}_{t\geq 0}$ such that Λ has the representation

$$\Lambda_t \stackrel{d}{=} \int_0^t \lambda_s \, \mathrm{d}s, \quad t \geq 0,$$

then λ is called intensity process. If $\Lambda_t = \Lambda t$, $t \geq 0$, for some non-negative random variable Λ, then the transformed process is called a mixed Poisson process.

Cox processes were introduced in Cox (1955) under the name of doubly-stochastic Poisson processes and further studied, for instance, in Bartlett (1963). Like Poisson processes, they are often studied as general point processes where the directing process corresponds to a random measure given by

$$\Lambda(s,t] := \Lambda_t - \Lambda_s, \quad 0 \leq s < t < \infty.$$

A mixed Poison process can be understood as randomization of a homogeneous Poisson process. Conditioned on the outcome of the mixing variable, a mixed Poisson process becomes a homogeneous Poisson process. In case of a non-degenerate mixing variable Λ, however, a mixed Poisson process has no longer independent increments.

Accordingly, a Cox process can be understood as randomization of an inhomogeneous Poisson process. Even though in general it also loses the independent increment property of the underlying Poisson process, this property can be preserved given a suitable choice of the directing process Λ, as will be seen later on. If the directing process is continuous and converges to infinity over time, the Cox process, like the inhomogeneous Poisson process, can be transformed back to a homogeneous Poisson process with intensity one,

see (Grandell, 1991, Chapter 2.1, p.39). This transformation even works in the multivariate case: using a suitable time-change, a multivariate Cox process can be converted into independent Poisson processes, see Brown and Nair (1988). In this thesis, however, we will study a multivariate Cox process with a Lévy subordinator as directing process. Given the discontinuities of the subordinator, a transformation back to the independent Poisson case is no longer possible.

This section concludes with another (multivariate) extension of the (homogeneous) Poisson process using compounding. Let N be a univariate random variable with values in $\mathbb{N}_0$ and $\boldsymbol{Y}$ be a d-dimensional random vector, then the random variable $\boldsymbol{M}$ defined as

$$\boldsymbol{M} := \sum_{j=1}^{N} \boldsymbol{Y}_j,$$

with $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ iid copies of $\boldsymbol{Y}$ independent of N, is said to have a compound distribution with primary distribution N and secondary distribution $\boldsymbol{Y}$. The classical case, which is of main concern in this thesis, is the compound Poisson case where N follows a Poisson distribution; the corresponding process is the compound Poisson process.

Definition 2.7 (Compound Poisson process)
Let $N = \{N_t\}_{t\geq 0}$ be a Poisson process with intensity $\lambda > 0$. Let $\boldsymbol{Y}$ be a d-dimensional random variable and $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ be iid copies of $\boldsymbol{Y}$ independent of N. Then the process $\boldsymbol{M} = \{\boldsymbol{M}_t\}_{t\geq 0}$ defined as

$$\boldsymbol{M}_t := \sum_{j=1}^{N_t} \boldsymbol{Y}_j, \quad t \geq 0,$$

is called compound Poisson process.

In the classical Cramér–Lundberg model, a univariate compound Poisson process is used to model the evolution of the aggregate claim amount of an insurance portfolio: the Poisson process describes the claim arrivals over time and each claim is assigned an iid random claim size. This approach is readily extended to more general claim number processes like inhomogeneous Poisson or Cox processes.

Compounding can also be used to generate new claim number processes by choosing non-negative integer-valued jump size distributions. In this case, the claims arrive in

clusters with distribution specified by $\boldsymbol{Y} \in \mathbb{N}_0^d$, and the clusters arrive according to N. If N is a Poisson process as in Definition 2.7, the resulting process is called Poisson cluster process. In Galliher et al. (1959) and Kemp (1967), compound Poisson distributions with non-negative integer-valued secondary distribution were studied under the name of stuttering Poisson distribution; hence, the corresponding processes are also be referred to as stuttering Poisson processes.

The compound Poisson process can be understood as a marked PRM on $E = \mathbb{R}_{\geq 0} \times \mathbb{R}^d$, where an independent mark – the jumps size $\boldsymbol{Y}_j$ – is assigned to the jump arrival points T_j of the Poisson process. More precisely, the jump measure associated with the compound Poisson process M is given by

$$M(A) := |\{j \in \mathbb{N} : (T_j, \boldsymbol{Y}_j')' \in A\}|, \quad A \in \mathcal{B}(E).$$

and, hence, M is a Poisson random measure with mean measure $\lambda Leb \times \mathbb{P}_{\boldsymbol{Y}}$, see (Mikosch, 2009, Chapter 7.3.2, Proposition 7.3.3, p.247). Given this observation, the properties of the following remark do not come as much of a surprise.

Remark 2.8 (Decomposition and superposition of the compound Poisson process) According to the space-time decomposition or thinning property of the compound Poisson process, for $A_i \in \mathcal{B}(E)$ disjoint sets, $i = 1, \ldots, d$, the processes

$$\boldsymbol{M}_t^i := \sum_{j=1}^{N_t} \boldsymbol{Y}_j \mathbb{1}_{A_i}((T_j, \boldsymbol{Y}_j')'), \quad t \geq 0,$$

are mutually independent and have themselves a compound Poisson representation, see (Mikosch, 2009, Chapter 3.3.2, Theorem 3.3.6, p.116). In other words, in the compound Poisson model the aggregate claim amount in non-overlapping time intervals and for non-overlapping layers of claim sizes – quantities that are, for instance, relevant in reinsurance applications – are themselves independent compound Poisson processes.

In addition, the superposition of independent compound Poisson processes remains in the same class as well. More precisely, let

$$M_t^i = \sum_{j=1}^{N_t^i} Y_{ij}$$

be independent compound Poisson processes with intensities $\lambda_i > 0$ and independent jump sizes Y_i, $i = 1, \dots, d$. For ease of notation, only the univariate case is considered. Then the aggregate process $\sum_{i=1}^{d} M_t^i$ is a compound Poisson process as well with intensity $\lambda = \sum_{i=1}^{d} \lambda_i$. The jump size distribution Y of the aggregate process is the discrete mixture of the individual jump size distributions, that is for a random variable J on $\{1, \dots, d\}$ with $\mathbb{P}(J = i) = \lambda_i / \lambda$ it holds:

$$Y = \sum_{i=1}^{d} \mathbb{1}_{\{J=i\}} Y_i,$$

see (Mikosch, 2009, Chapter 3.3.1, Proposition 3.3.4, p.113). By the law of total probability it holds for the distribution of Y:

$$\mathbb{P}(Y \in B) = \sum_{i=1}^{d} \frac{\lambda_i}{\lambda} \mathbb{P}(Y_i \in B), \quad B \in \mathcal{B}(\mathbb{R}).$$

▲

2.3 Lévy processes and subordinators

In this thesis, a Cox process directed by a Lévy subordinator is considered, a process which is itself a Lévy subordinator. Hence, the process is mainly study from the perspective and with the methodologies of Lévy processes. For this purpose, an introduction to Lévy processes in general and subordinators in particular is given in this section. Some standard references are Sato (1999), Bertoin (1998), Applebaum (2004), Schoutens (2003), and Cont and Tankov (2003).

Lévy processes are mainly characterized by their independent and stationary increments. They can be understood as a stochastic version of linear functions: whereas linear functions have constant increments, Lévy processes have iid increments. Alternatively, they can be seen as the continuous time extension of discrete time random walks with iid step size distribution. Two important examples of Lévy processes have already been introduced in the previous section: the homogeneous and the compound Poisson process. These two form together with linear functions and Brownian motions the building blocks of general Lévy processes, as will be discussed in more detail after the the formal definition.

Definition 2.9 (Lévy process)
A d-dimensional stochastic process $\boldsymbol{X} = \{\boldsymbol{X}_t\}_{t\geq 0}$ is called a Lévy process if it has the following properties:

(i) Càdlàg paths: The paths of $\boldsymbol{X}$ are almost surely càdlàg functions.

(ii) Start at zero: $\boldsymbol{X}_0 = \boldsymbol{0}$ a.s.

(iii) Independent increments: For any $0 \leq t_0 < t_1 < \cdots < t_n < \infty$, $n \in \mathbb{N}$, the increments $\Delta\boldsymbol{X}_i$, are mutually independent, $i = 1, \ldots, n$.

(iv) Stationary increments: For any $0 \leq s < t < \infty$, it is $\boldsymbol{X}_t - \boldsymbol{X}_s \stackrel{d}{=} \boldsymbol{X}_{t-s}$.

(v) Stochastic continuity: For all $t \geq 0$ and $\epsilon > 0$ it is $\lim_{s\to t} \mathbb{P}(\|\boldsymbol{X}_s - \boldsymbol{X}_t\| > \epsilon) = 0$.

A Lévy processes with component-wise a.s. non-decreasing paths is called a Lévy subordinator.

The structure of the sample paths of a Lévy process is characterized in the famous Lévy–Itô decomposition, which was developed in Lévy (1934) and Itô (1987); a detailed derivation and discussion can be found in (Sato, 1999, Chapter 4, Section 19 and 20, pp.119). It is shown there that a Lévy process is the superposition of a linear function, a diffusion part given by a Brownian motion, and a jump component. The linear part is fully characterized by a real vector specifying the slope and the Brownian motion

is determined by a symmetric positive-semidefinite covariance matrix, called diffusion matrix. Jumps occur at the points of discontinuity of the Lévy process, that is where

$$\nabla \boldsymbol{X}_t := \boldsymbol{X}_t - \lim_{s \uparrow t} \boldsymbol{X}_s > 0.$$

Due to the stochastic continuity of the process, these jumps cannot appear at fixed points in time with a positive probability. Furthermore, due to the càdlàg property of the paths, only countably many jumps can occur on any time interval and the number of jumps with larger norm than some arbitrary barrier $\epsilon > 0$ has to be finite. The number of small jumps, on the other hand, can be infinite, yet has to be still countable, see (Billingsley, 2009, Chapter 3.12, pp.121). More precisely, consider on $\mathbb{R}_{\geq 0} \times \dot{\mathbb{R}}^d$ the jump measure

$$J(A) := |\{t \geq 0 : (t, \nabla \boldsymbol{X}'_t)' \in A\}|, \quad A \in \mathcal{B}(\mathbb{R}_{\geq 0} \times \dot{\mathbb{R}}^d),$$

and define on $\dot{\mathbb{R}}^d$ the so-called Lévy measure as the expected number of jumps with a certain size in a unit time interval:

$$\nu_{\boldsymbol{X}}(B) := \mathbb{E}[J([0,1] \times B)], \quad B \in \mathcal{B}(\dot{\mathbb{R}}^d).$$

The drift vector, diffusion matrix, and Lévy measure together fully determine the distributional properties of the Lévy process and are therefore called the characteristic triplet. As the process has only finitely many big jumps, it holds for any $\epsilon > 0$:

$$\int_{\|\boldsymbol{x}\| \geq \epsilon} \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) < \infty.$$

Thus, the Lévy measure $\nu_{\boldsymbol{X}}$ is a Radon measure on $\dot{\mathbb{R}}^d$, i.e. it is finite on any compact set not containing zero, and the jump measure J can be shown to be a marked PRM with mean measure $Leb \times \nu_{\boldsymbol{X}}$, see (Sato, 1999, Chapter 4, Section 19, Theorem 19.2, p.120). Consequently, jumps with a norm larger than some barrier ϵ – the barrier is usually chosen as $\epsilon := 1$ – can be represented as a compound Poisson process.

For the small jumps some more care has to be taken. The Lévy measure may have a non-integrable singularity at zero due to the possibility of infinitely many small jumps, but it has to at least fulfil the following integrability condition:

$$\int_{\|\boldsymbol{x}\|<\epsilon} \|\boldsymbol{x}\|^2 \, \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) < \infty.$$

The sum over the small jumps not necessarily converges; instead, the sum of the compensated jumps has to be considered and the compensation has to be accounted for in the linear part of the Lévy process. Hence, in general the vector characterizing the linear part in the characteristic triplet depends on the chosen cutting level ϵ for the jump sizes. If the Lévy measure fulfils the additional integrability condition

$$\int_{\|\boldsymbol{x}\|<\epsilon} \|\boldsymbol{x}\| \, \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) < \infty,$$

the small jumps are well behaved and no compensation is necessary. The cutting function for the small jumps can be removed and the whole jump part of the Lévy process can be understood as the superposition of a possibly infinite number of (compound) Poisson processes, see (Sato, 1999, Chapter 2, Section 8, Remark 8.4, pp.38). In this case, the vector characterizing the linear part is called the drift of the Lévy process.

In the following, we are primarily working with Lévy subordinators which always fulfil the stronger integrability condition for the Lévy measure. In addition, they have a non-negative drift, the Lévy measure is concentrated on $\dot{\mathbb{R}}^d_{\geq 0}$ – that is the jump sizes are non-negative – and they have no diffusion part – that is the Gaussian diffusion matrix is zero and the characteristic triplet reduces to a characteristic tuple. Hence, a Lévy subordinator can be seen as a linearly increasing process affected by random upward jumps. This notion will be made precise in Theorem 2.12.

In general, the Lévy processes that have no diffusion part and fulfil the stronger integrability condition for the Lévy measure are exactly the Lévy processes with paths of a.s. finite variation, see (Sato, 1999, Chapter 4, Section 21, Theorem 21.9, pp.140). Furthermore, if the diffusion part is zero and the Lévy measure finite, that is if $\nu_{\boldsymbol{X}}(\dot{\mathbb{R}}^d) < \infty$, the processes are of finite activity and correspond to the compound Poisson processes with drift, see (Sato, 1999, Chapter 4, Section 21, Theorem 12, p.135). In this case, the Lévy measure has the form $\nu_{\boldsymbol{X}} = \lambda \mathbb{P}_{\boldsymbol{Y}}$ with intensity parameter $\lambda = \nu_{\boldsymbol{X}}(\dot{\mathbb{R}}^d)$ and jump size distribution $\mathbb{P}_{\boldsymbol{Y}} = \lambda^{-1}\nu_{\boldsymbol{X}}$. If $\nu_{\boldsymbol{X}}(\dot{\mathbb{R}}^d) = \infty$, the jump times are countable dense in $\mathbb{R}_{\geq 0}$, see (Sato, 1999, Chapter 4, Section 21. Theorem 21.3, p.136).

We now take a closer look at the distribution of Lévy processes at a certain point in time, which is infinitely divisible.

Definition 2.10 (Infinite divisibility)
The distribution of a random vector $\boldsymbol{X}$ is infinitely divisible if for each $n \in \mathbb{N}$ a random vector $\boldsymbol{X}^n$ exists such that for n iid copies $\boldsymbol{X}_k^n$ of $\boldsymbol{X}^n$ it holds

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X}_1^n + \cdots + \boldsymbol{X}_n^n.$$

If $\boldsymbol{X}$ and all $\boldsymbol{X}^n$ are discrete, then the distribution is called discretely infinitely divisible.

Due to the independent and stationary increment property it can easily be seen that the distribution of a Lévy process at any point in time has to be infinitely divisible; in fact, even a one-to-one relationship exists between infinitely divisible distributions and Lévy processes.

Theorem 2.11 (Lévy processes and infinite divisibility)
For a Lévy process $\boldsymbol{X} = \{\boldsymbol{X}_t\}_{t\geq 0}$, the distribution $\mathbb{P}_{\boldsymbol{X}_t}$ is infinitely divisible for any $t \geq 0$ and it holds for the characteristic function:

$$\Phi_{\boldsymbol{X}_t}(\boldsymbol{v}) = \Phi_{\boldsymbol{X}_1}(\boldsymbol{v})^t, \quad \boldsymbol{v} \in \mathbb{R}^d.$$

Furthermore, for any infinitely divisible distribution F exists a Lévy process $\boldsymbol{X} = \{\boldsymbol{X}_t\}_{t\geq 0}$ such that $\boldsymbol{X}_1 \sim F$, and this process is unique up to identity in law.

Proof. See (Sato, 1999, Chapter 2, Section 7, Lemma 7.9 and Theorem 7.10, p.35; Section 11, Corollary 11.6, p.63). □

From the above theorem it directly follows that compound Poisson distributions are infinitely divisible as they correspond to compound Poisson processes which are Lévy processes. In addition, the subset of discretely infinitely divisible distributions coincides with the set of compound Poisson distributions with integer-valued jump-size distribution (and Lévy measure), see (Sato, 1999, Chapter 5, Setion 27, Corollary 27.5, p.176). Furthermore, the non-negative infinitely divisible distributions correspond to Lévy subordinators, a result which is shown for the one-dimensional case in detail in (Sato, 1999, Chapter 5, Section 24, Theorem 24.11, p.153); for the multivariate case see (Barndorff-Nielsen et al., 2001, Proposition 3.1).

The distribution of a Lévy process is determined by its distribution at time $t = 1$, thus this distribution is the only degree of freedom in specifying the process. The famous

Lévy–Khintchine representation, which goes back to Lévy (1934) and Khintchine (1937), completely characterizes this distribution in terms of the Lévy characteristics. The explicit results are given here for the special case of Lévy subordinators, as these results will be of importance later on in this thesis. Whereas in most textbooks Lévy subordinators are only studied in dimension one, for this thesis we particularly consider the multivariate case. Furthermore, due to the non-negativity of the process increments, it is convenient to work with the Laplace transform rather than the characteristic function.

Theorem 2.12 (Lévy–Khintchine representation for Lévy subordinators)
Let $\boldsymbol{X} = \{\boldsymbol{X}_t\}_{t\geq 0}$ be a d-dimensional Lévy subordinator. Then $\boldsymbol{X}$ is fully characterized by a positive drift vector $\boldsymbol{b}_{\boldsymbol{X}} \in \mathbb{R}^d_{\geq 0}$ and a σ-finite Lévy measure $\nu_{\boldsymbol{X}}$ concentrated on $\dot{\mathbb{R}}^d_{\geq 0}$ and satisfying the integrability condition

$$\int_{\mathbb{R}^d_{\geq 0}} \min\{1, \|\boldsymbol{x}\|\}\, \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) < \infty.$$

The Laplace transform $\varphi_{\boldsymbol{X}_t}$ of $\boldsymbol{X}_t$ can be expressed in terms of the so-called Laplace exponent $\Psi_{\boldsymbol{X}} \colon \mathbb{R}^d_{\geq 0} \to \mathbb{R}_{\geq 0}$ by

$$\varphi_{\boldsymbol{X}_t}(\boldsymbol{u}) = \exp\{-t\Psi_{\boldsymbol{X}}(\boldsymbol{u})\}, \quad t \geq 0, \boldsymbol{u} \in \mathbb{R}^d_{\geq 0},$$

and the Laplace exponent is derived from the characteristics of the subordinator as

$$\Psi_{\boldsymbol{X}}(\boldsymbol{u}) = \boldsymbol{u}'\boldsymbol{b}_{\boldsymbol{X}} + \int_{\mathbb{R}^d_{\geq 0}} (1 - \exp\{-\boldsymbol{u}'\boldsymbol{x}\})\, \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}), \quad \boldsymbol{u} \in \mathbb{R}^d_{\geq 0}.$$

Proof. In the case of a univariate subordinator, this follows from the results on general Lévy processes given in (Sato, 1999, Chapter 2, Section 8, pp.37) together with the the results on univariate subordinators given in (Sato, 1999, Chapter 4, Section 21, Theorem 21.5, p.137). The multivariate version is stated in (Barndorff-Nielsen et al., 2001, Proposition 3.1) and was derived from results in (Skorohod, 1991, Chapter 3.3, Theorem 21, pp.156) on stochastically continuous processes with independent increments living on cones, considering that $\mathbb{R}^d_{\geq 0}$ is a particular cone in $\mathbb{R}^d$. □

The Laplace exponent is closely connected to the cumulant generating function of the Lévy process which is linear in time according to the theorem:

$$\Psi_{\boldsymbol{X}}(\boldsymbol{u}) = -\log \varphi_{\boldsymbol{X}_1}(\boldsymbol{u}) = -g_{\boldsymbol{X}_1}(-\boldsymbol{u}), \quad \boldsymbol{u} \in \mathbb{R}^d_{\geq 0}.$$

For a univariate Lévy subordinator $\Lambda = \{\Lambda_t\}_{t\geq 0}$, the Laplace exponent Ψ_Λ is a Bernstein function, i.e. a non-negative function with completely monotone first derivative, fulfilling $\Psi_\Lambda(0) = 0$. More precisely, for all $n \in \mathbb{N}$ it holds:

$$(-1)^{n+1}\Psi_\Lambda^{(n)}(x) \geq 0, \quad x > 0,$$

see (Feller, 1971, Chapter XIII.7, Theorem 1, p.450). For a thorough study of Bernstein functions see Schilling et al. (2012); there it is also shown that any Bernstein function starting in zero has a representation as in Theorem 2.12 and is thus the Laplace exponent of a subordinator, see (Schilling et al., 2012, Theorem 3.2 and Theorem 3.6).

In case of a multivariate Lévy subordinator, the Lévy measure also contains the information about the dependencies between the marginal processes, as is stated in the following theorem.

Theorem 2.13 (Marginals and independence of a Lévy subordinator)
Consider a d-dimensional Lévy subordinator $\boldsymbol{X} = \{\boldsymbol{X}_t\}_{t\geq 0}$ with drift $\boldsymbol{b}_{\boldsymbol{X}} := \boldsymbol{b} \in \mathbb{R}^d_{\geq 0}$ and Lévy measure $\nu_{\boldsymbol{X}}$. Then the characteristic tuple of the marginal processes $X^i = \{X^i\}_{t\geq 0}$ are given w.l.o.g. for $i = 1$:

$$b_{X^1} = b_1, \qquad \nu_{X^1}(B) = \nu_{\boldsymbol{X}}(B \times \mathbb{R}^{d-1}_{\geq 0}), \quad B \in \mathcal{B}(\mathbb{R}_{>0}).$$

Furthermore, the marginal processes of $\boldsymbol{X}$ are independent if and only if they never jump together. In terms of the Lévy measure this translates into having support in $\{\boldsymbol{x} \in \dot{\mathbb{R}}^d_{\geq 0} : x_1 \cdots x_d = 0\}$ and in this case it holds:

$$\nu_{\boldsymbol{X}}(B) = \sum_{i=1}^{d} \nu_{X^i}(B_i), \quad B \in \dot{\mathbb{R}}^d_{\geq 0},$$

where $B_i := \{x \in \mathbb{R} : x\boldsymbol{e}_i \in B\}$, $i = 1, \ldots, d$.

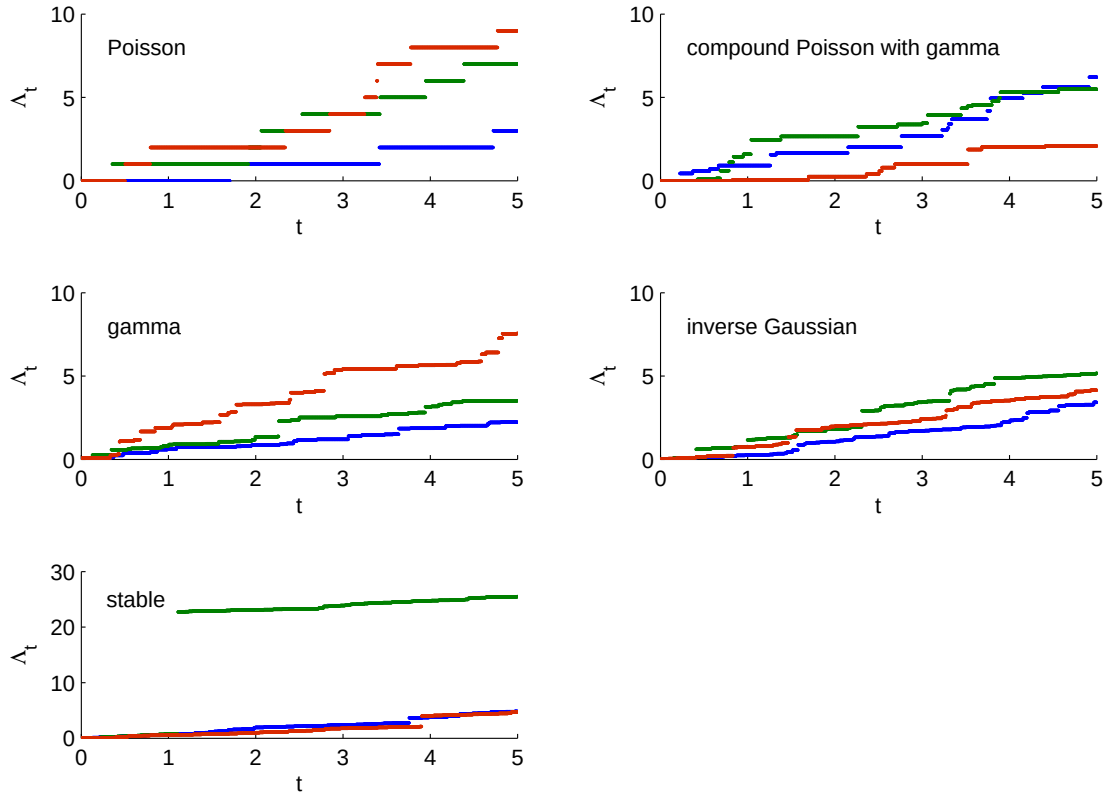


Figure 2.1 Sample paths of Lévy subordinator families: The figure shows three sample paths for each subordinator family introduced in Example 2.14. (Top-left) Poisson process with intensity $\lambda = 1$; (top-right) compound Poisson process with intensity $\lambda = 10/3$ and Gamma$(1.5, 5)$ jump size distribution; (middle-left) gamma subordinator with $\beta = \eta = 2.25$; (middle-right) inverse Gaussian subordinator with $\beta = \eta = 1.5$; (bottom-left) stable subordinator with $\alpha = 0.8$. The parameters of the first four subordinators have been chosen such that mean and variance are identical; the stable subordinator always has infinite mean and variance.

Proof. The result can be found in (Cont and Tankov, 2003, Chapter 5.4, Proposition 5.2 and Proposition 5.3, p.144); even though only the bivariate case is considered there, the multivariate version follows in the same fashion. □

Example 2.14 (Families of Lévy subordinators)
Five common examples of univariate Lévy subordinators $\Lambda = \{\Lambda_t\}_{t\geq 0}$ are introduced in the following; Figure 2.1 shows sample paths for each case.

(i) Poisson process with intensity $\lambda > 0$:
According to Definition 2.3, the infinitely divisible distribution of the Poisson process is the Poisson distribution, that is $\Lambda_t \sim \text{Poi}(\lambda t)$. From the Laplace transform, see (Sato, 1999, Chapter 1, Section 2, Example 2.7, p.10), it can be seen that the Laplace exponent is:

$$\Psi_\Lambda(u) = \lambda(1 - \exp\{-u\}), \quad u \geq 0.$$

Then the Lévy characteristics are:

$$b_\Lambda = 0; \qquad \nu_\Lambda(B) = \lambda\delta_1(B), \quad B \in \mathcal{B}(\mathbb{R}_{>0}).$$

(ii) Compound Poisson process with intensity $\lambda > 0$ and random jump size $Y > 0$:
By definition, the infinitely divisible distribution of the process is a compound Poisson distribution. Again, from the Laplace transform, see (Sato, 1999, Chapter 1, Section 4, pp.18), the Laplace exponent is found to be:

$$\Psi_\Lambda(u) = \lambda(1 - \varphi_Y(u)), \quad u \geq 0,$$

where φ_Y is the Laplace transform of Y. Then the Lévy characteristics are:

$$b_\Lambda = 0; \qquad \nu_\Lambda(B) = \lambda\mathbb{P}(Y \in B), \quad B \in \mathcal{B}(\mathbb{R}_{>0}).$$

(iii) Gamma subordinator with parameters $\beta, \eta > 0$:
The gamma subordinator is the Lévy process corresponding to the gamma distribution, see (Applebaum, 2004, Chapter 1.3, Example 1.3.22, p.52), i.e. for each $t > 0$ it is $\Lambda_t \sim \text{Gamma}(\beta t, \eta)$ with density:

$$f_{\Lambda_t}(x) = \frac{\eta^{\beta t}}{\Gamma(\beta t)} x^{\beta t - 1} \exp\{-\eta x\} \mathbb{1}_{\{\mathbb{R}_{>0}\}}, \quad x \in \mathbb{R},$$

where $\Gamma(\cdot)$ denotes the gamma function. The corresponding Laplace exponent is:

$$\Psi_\Lambda(u) = \beta \log\left(1 + \frac{u}{\eta}\right), \quad u \geq 0.$$

The Lévy measure of the gamma subordinator is absolutely continuous with respect to the Lebesgue measure and the Lévy characteristics are:

$$b_\Lambda = 0; \qquad \nu_\Lambda(\mathrm{d}x) = \beta \exp\{-\eta x\} \frac{1}{x} \mathbb{1}_{\mathbb{R}_{>0}}(x) \mathrm{d}x.$$

(iv) Inverse Gaussian subordinator with parameters $\beta, \eta > 0$:
The inverse Gaussian subordinator stems from the inverse Gaussian distribution, see (Applebaum, 2004, Chapter 1.3, Example 1.3.21, p.51), that is $\Lambda_t \sim \mathrm{IG}(\beta t, \eta)$ for $t > 0$ with density:

$$f_{\Lambda_t}(x) = \frac{\beta t}{\sqrt{2\pi x^3}} \exp\Big\{ -\frac{1}{2x}(\eta x - \beta t)^2 \Big\} \mathbb{1}_{\mathbb{R}_{>0}}(x), \quad x \in \mathbb{R},$$

and Laplace exponent:

$$\Psi_\Lambda(u) = \beta(\sqrt{2u + \eta^2} - \eta), \quad u \geq 0.$$

The Lévy measure of the process is again absolutely continuous with respect to the Lebesgue measure and the Lévy characteristics are, see (Schoutens, 2003, Chapter 5.3.4, p.53):

$$b_\Lambda = 0; \qquad \nu_\Lambda(\mathrm{d}x) = \frac{1}{\sqrt{2\pi}} \beta x^{-\frac{3}{2}} \exp\Big\{ -\frac{1}{2}\eta^2 x \Big\} \mathbb{1}_{\mathbb{R}_{>0}}(x) \mathrm{d}x.$$

(v) Stable subordinator with parameter $0 < \alpha < 1$:
The stable subordinator, see (Applebaum, 2004, Chapter 1.3, Example 1.3.18, p.51), is characterized by its stable distribution, i.e. using the parametrization $\mathrm{S}(\alpha, \beta, \sigma)$ from (Applebaum, 2004, Chapter 1.2, Theorem 1.2.21, p.33):

$$\Lambda_t \sim \mathrm{S}\left(\alpha, 1, \cos\left(\frac{\pi\alpha}{2}t\right)^{\frac{1}{\alpha}}\right).$$

It follows for the Laplace exponent:

$$\Psi_\Lambda(u) = u^\alpha, \quad u \geq 0,$$

and the Lévy characteristics are:

$$b_\Lambda = 0; \qquad \nu_\Lambda(\mathrm{d}x) = \frac{\alpha}{\Gamma(1-\alpha)} x^{-1+\alpha} \mathbb{1}_{\mathbb{R}_{>0}}(x)\mathrm{d}x.$$

Note that the moment generating function does not exist in any point $v > 0$; if it did, it would be given by $\mathbb{E}[\exp\{v\Lambda_t\}] = \exp\{-t\Psi_\Lambda(-v)\}$ which is not a well-defined real function, see (Sato, 1999, Chapter 5, Section 25, Theorem 25.17, p.165). Hence, the distribution is heavy-tailed and moments of all order in $\mathbb{N}$ are infinite.

All five given examples of Lévy subordinators can be extended by a non-negative drift $b > 0$, that is

$$\tilde{\Lambda}_t := bt + \Lambda_t.$$

In this case the drift of the process becomes $b_{\tilde{\Lambda}} = b$, the Lévy measure $\nu_{\tilde{\Lambda}} = \nu_\Lambda$ remains unchanged, and the Lévy exponent has to be adjusted to:

$$\Psi_{\tilde{\Lambda}}(u) = bu + \Psi_\Lambda(u), \quad u \geq 0.$$

3 Model set-up and characterization

Having fixed the notation and covered the mathematical preliminaries, this chapter introduces the model that is studied in this thesis. Section 3.1 presents the set-up and gives a first characterization of the multivariate process. A study of the claim arrival times in Section 3.2 helps to formulate an efficient sampling algorithm for the model, which offers the possibility to numerically evaluate distribution-related quantities which cannot be analytically solved. In Section 3.3, important literature on the topic is reviewed and the relationship with the Lévy-frailty model for credit default times is established.

3.1 Model set-up

Consider $d \in \mathbb{N}$ insurance portfolios, each consisting of a single policy or a collection of policies, and let the univariate processes $L^i := \{L^i_t\}_{t\geq 0}$ count the claims arriving in each portfolio $i = 1, \dots, d$ up to time $t \geq 0$. To describe the joint behaviour of these processes, we not only need a model for each process L^i, but must also specify the dependence structure between them. Thus, a model for the full d-dimensional claim number process $\boldsymbol{L} := (L^1, \dots, L^d)'$ is necessary. Different approaches to this task will be discussed in Section 3.3. In this thesis, we follow a probabilistic construction where $\boldsymbol{L}$ is specified as a multivariate Poisson process that is time-changed by a Lévy subordinator. As we will see, this approach allows for an intuitive understanding of the model properties, which include many typical features of claim count data, while preserving mathematical tractability.

In the traditional modelling approach, claim arrivals in each portfolio i are assumed to follow independent homogeneous Poisson processes $N^i := \{N^i_t\}_{t\geq 0}$ with individual intensities $\lambda_i > 0$. Let $\boldsymbol{N} := (N^1, \dots, N^d)'$ be the corresponding multivariate Poisson process with independent marginals and intensity vector $\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_d)'$. Naturally, all limitations of the univariate Poisson model apply to the multivariate version as well. As was noted in Section 2.2, the dispersion observed in the claim count data often exceeds that predicted by the Poisson distribution, which is one reason why extensions based on mixed Poisson distributions, that is mixed Poisson and Cox processes, have gained in popularity. Furthermore, claims produced from a univariate homogeneous Poisson process always arrive one after another, while in many real-world applications claims can and do arrive simultaneously. This feature directly extends to this multivariate process as well: due to the independence of the marginal processes, simultaneous claim arrivals are

a.s. impossible for each component as well as between different components, as mentioned in Section 2.2. Thus, the multivariate Poisson process with independent marginals fails to account for cluster arrivals or for any dependence between claims arriving in different portfolios.

The basic Poisson model will now be extended such that the new model allows for simultaneous claim arrivals and captures dependence between the individual portfolios. We introduce an additional source of randomness independent of $\boldsymbol{N}$, namely a Lévy subordinator $\Lambda := \{\Lambda_t\}_{t\geq 0}$ with characteristic tuple (b_Λ, ν_Λ), serving as a stochastic clock or operational time. As mentioned in Section 2.3, a Lévy subordinator can be understood as a process that increases linearly with drift b_Λ and is affected by random upward jumps. The Lévy measure ν_Λ encodes all information about the size and timing of the jumps. Like real time the subordinator paths start at zero and are almost surely non-decreasing. But while real time grows linearly, the new time can randomly run slower or faster than real time, depending on the jump activity of the subordinator. High activity periods represent times of high market intensity with many claims coming in, and large subordinator jumps may be directly linked to catastrophic events.

In the following the claim number process $\boldsymbol{L}$ is assumed to be the basic Poisson process $\boldsymbol{N}$ if run through according to the random time Λ:

$$\{\boldsymbol{L}_t\}_{t\geq 0} := \{\boldsymbol{N}_{\Lambda_t}\}_{t\geq 0}. \tag{M}$$

This approach naturally generates dependence between the portfolios in the model as the stochastic factor Λ similarly affects all components. Furthermore, while the underlying Poisson process gives rise to only jumps of size one, the subordinator and therefore time can now jump, leading to simultaneous claim arrivals within and between the individual portfolios in the time-changed process.

The parameter set of Model (M) consists of the d intensities λ_i of the underlying Poisson processes as well as the subordinator parameters. As can be seen in Example 2.14, a Poisson process or stable subordinator is characterized by a single parameter, an inverse Gaussian or gamma subordinator by two parameters. For the compound Poisson process, the number of parameters depends on the chosen jump size distribution; for gamma-distributed jumps the compound Poisson process has three parameters. Adding a drift to any of the subordinators increases the number of parameters by one. Depending on the specific application, it may be desirable that the operating time, although randomly running faster or slower, behaves on average like real-time, that is $\mathbb{E}[\Lambda_t] = t$ holds for all $t \geq 0$. Due to the linearity of the first moment of a Lévy process, imposing the following assumption to ensure time-normalization is sufficient:

$$\mathbb{E}[\Lambda_1] = 1. \tag{TN}$$

This will be discussed in more detail in Chapter 4. Time-normalization does not necessarily need to be imposed, but it does improve the interpretation of this subordinator as stochastic time and enables the recovery of the expected claim numbers $\boldsymbol{\lambda}$ in the Poisson model, as will be shown later in Proposition 4.15. To fulfil this condition, one parameter of the subordinator has to be set in advance, reducing the total number of subordinator parameters by one. With or without time-normalization, Model (M) has only a relatively modest number of additional parameters compared to the basic independent Poisson model. Thus, the model offers a better description of many claim arrival patterns while providing a good chance that each additional parameter can be determined reasonably well from the available data. This topic will be discussed further in Chapter 5. Extensions of the time-changed model incorporating an even richer variety of default patterns at the cost of an increased number of parameters will be discussed in Chapter 6.

Figure 3.1 shows an example of how the univariate Poisson claim arrivals are changed if observed while following a stochastic clock[1]. The upper chart shows one realization of an inverse Gaussian subordinator which is one realization of time. The subordinator parameters were chosen such that Assumption (TN) holds. In this particular realization, however, the subordinator path rises well above 10, the level expected in $T = 10$. In the figure's lower chart, the sampled claim arrivals of an independent Poisson process are visualized in grey. If these claims are observed with regard to the subordinator sample path, they arrive in clusters marked in blue. The size of these clusters depends on the jump activity of the time path. In Figure 3.2 the example is extended to dimension $d = 2$. Due to a lower intensity chosen for component 2 compared to component 1, the overall number of jumps is smaller and the cluster sizes tend to be smaller as well. However, as both components are affected by the same time realization, relatively large clusters tend to occur simultaneously in both.

By definition, the time-changed process $\boldsymbol{L}$ is a multivariate Cox process in the sense that each marginal process L^i is a univariate Cox process as in Definition 2.6. The directing processes are all driven by the Lévy subordinator Λ, but each is scaled by the component-specific intensity λ_i, that is the directing processes are $\lambda_i\Lambda$, $i = 1, \ldots, d$. At each point in time $t \geq 0$, the marginals are, therefore, distributed according to a mixed Poisson distribution with the infinitely divisible mixing variable $\lambda_i\Lambda_t$.

Two main factors contribute to the very good mathematical tractability of this multivariate Cox process $\boldsymbol{L}$. First, the dependence between the marginal processes is solely introduced by the common stochastic clock Λ. Given the sigma-algebra $\sigma(\Lambda)$ generated

[1] The data for Figures 3.1 and 3.2 was sampled using Algorithm 3.3.

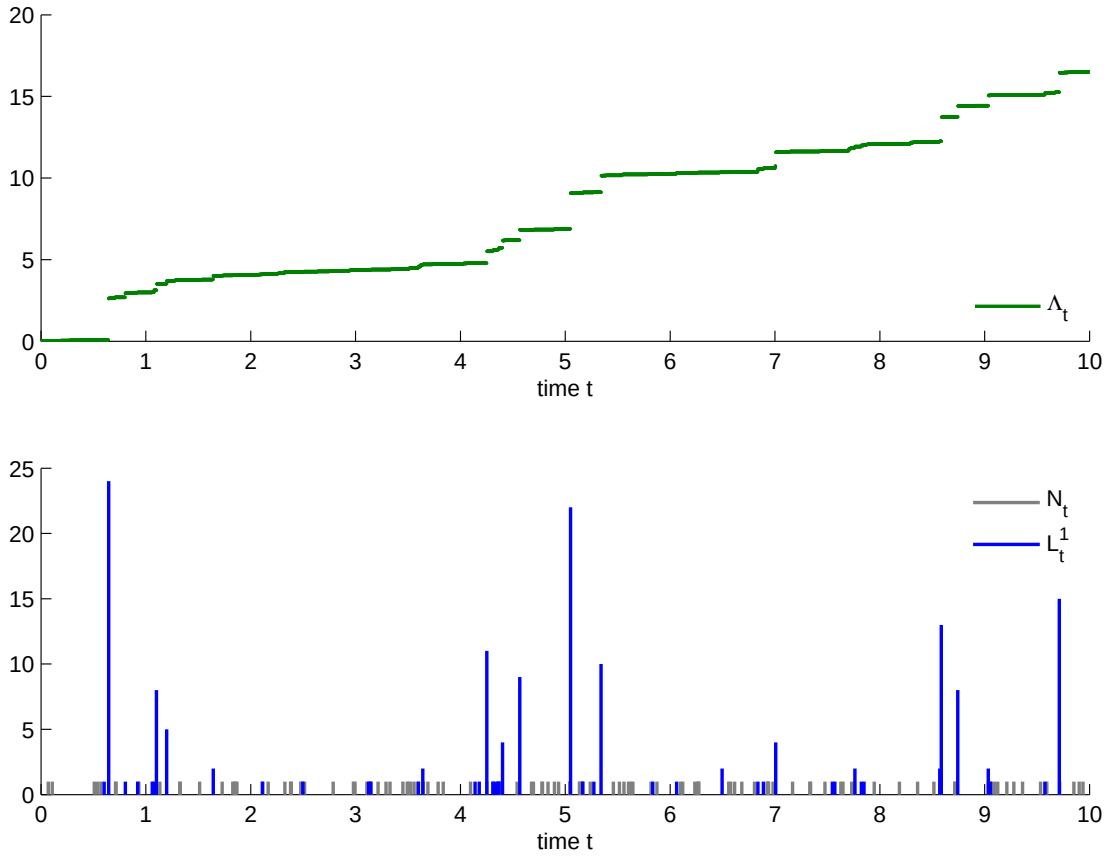


Figure 3.1 Claim arrivals before and after time-change: The figure illustrates the claim arrivals of the underlying Poisson process as well as the time-changed process in dimension $d = 1$. The upper chart shows one sample path of an inverse Gaussian subordinator with no drift ($b = 0$, $\beta = \eta = 1$) and with time horizon $T = 10$. The lower chart illustrates the claim arrivals of a Poisson process with intensity $\lambda_1 = 10$ in grey; the claim arrivals of the Poisson process time-changed by the subordinator sample path are plotted in blue.

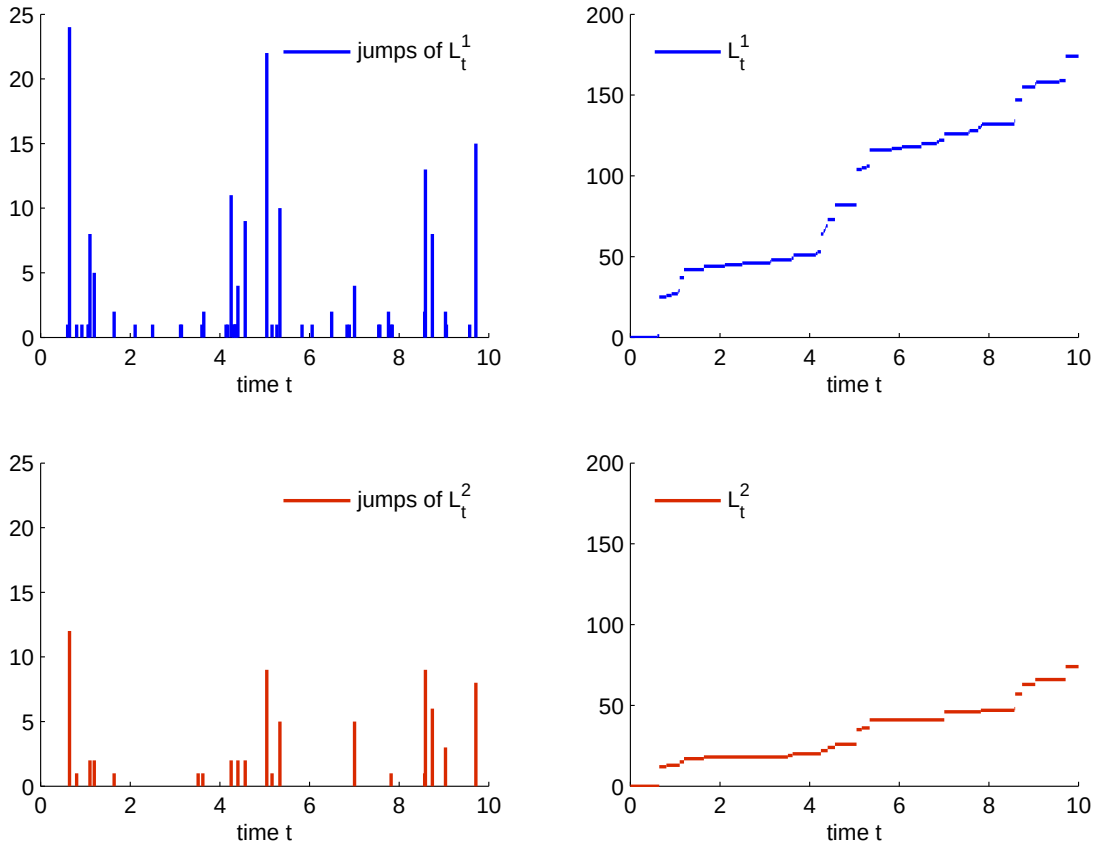


Figure 3.2 Bivariate claim arrivals: The figure illustrates a bivariate extension of Figure 3.1. The upper chart shows the claim arrivals (left) and the claim arrival process (right) of the univariate example of Model (M) from Figure 3.1 ($\lambda_1 = 10$ and inverse Gaussian subordinator with parameters $b = 0$, $\beta = \eta = 1$). For the lower chart, the model is extended by a second component with intensity $\lambda_2 = 5$; again, the claim arrivals and the claim arrival process are plotted.

by the subordinator, the marginal processes $L^1, \dots, L^d$ are independent. More precisely, given the outcome $\Lambda_t = \hat{\Lambda}_t$, $t \geq 0$, the processes are independent inhomogenous Poisson processes with mean value function $\lambda_i \hat{\Lambda}_t$ and the multivariate conditional distribution in $\boldsymbol{k} \in \mathbb{N}_0^d$ is:

$$\begin{aligned}\mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k} | \sigma(\Lambda)) &= \prod_{i=1}^{d} \mathbb{P}(N^i_{\Lambda_t} = k_i | \sigma(\Lambda)) = \prod_{i=1}^{d} \frac{(\lambda_i \Lambda_t)^{k^i}}{k_i!} \exp\{-\lambda_i \Lambda_t\} \\ &= \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!} \Lambda_t^{|\boldsymbol{k}|} \exp\{-|\boldsymbol{\lambda}| \Lambda_t\}.\end{aligned}$$

Due to the subordinator jumps, the paths of Λ are not continuous and hence not differentiable[2]. It follows that, in contrast to many studies of inhomogeneous Poisson and Cox processes, no (random) intensity function exists. Furthermore, the process cannot be converted back to the underlying independent Poisson process by a time-change with the inverse of the directing process (see Section 2.2).

Secondly, as the Poisson process is itself a Lévy subordinator and the family of Lévy subordinators is closed under subordination, the new process $\boldsymbol{L}$ inherits the convenient increment properties of the Lévy subordinator class. Moreover, the only Lévy process with piecewise constant paths is a compound Poisson process, hence $\boldsymbol{L}$ is a Poisson cluster process: the d-dimensional clusters of claims, seen in Figure 3.2, are iid and the arrival times are determined by an independent Poisson process. These properties will be discussed in detail in Chapter 4.

Remark 3.1 (Aggregated claim number process)
The superposition $\bar{N} := \sum_{i=1}^{d} N^i$ of the underlying independent Poisson processes is again a Poisson process with intensity $|\boldsymbol{\lambda}|$ (see Theorem 2.4). Therefore the aggregate claim number process

$$\bar{L} := \sum_{i=1}^{d} L^i$$

in Model (M) is itself a Poisson process time-changed by the subordinator Λ, namely:

$$\{\bar{L}_t\}_{t \geq 0} = \{\bar{N}_{\Lambda_t}\}_{t \geq 0}. \tag{3.1}$$

[2] Except for the degenerate case of $\Lambda_t = b_\Lambda t$ where the time-changed model consists again of homogeneous independent Poisson processes.

Due to this observation all results derived in the following for $\boldsymbol{L}$ can be directly transferred to the aggregate process $\bar{L}$ by choosing dimension $d = 1$ and intensity $|\boldsymbol{\lambda}|$. ▲

3.2 Arrival times and simulation

In this section, a probabilistic construction of the claim arrival times in Model (M) is given. This construction can be exploited to set up an efficient sampling algorithm. As for the Poisson process, a one-to-one correspondence exists between the claim arrival times and the claim number process in the time-changed model. Let τ_{ij} be the arrival time of claim number $j \in \mathbb{N}$ in portfolio $i = 1, \dots, d$, that is:

$$\tau_{ij} := \inf\{t > 0 : L_t^i \geq j\}.$$

The claim number processes $\boldsymbol{L}$ can be recovered from the component-wise sequences of its claim arrival times via:

$$L_t^i = \sum_{j=1}^{\infty} \mathbb{1}_{\{\tau_{ij} \leq t\}}, \quad t \geq 0. \tag{3.2}$$

The following proposition shows how the claim arrival times of $\boldsymbol{L}$ can be constructed as first-passage times of the subordinator over increasing trigger levels with exponentially distributed step sizes.

Proposition 3.2 (Probabilistic construction of claim arrival times)
Independently for all portfolios $i = 1, \dots, d$, let $E_{ij} \overset{iid}{\sim} \mathrm{Exp}(\lambda_i)$ for $j \in \mathbb{N}$ be independent and identically distributed trigger steps following an exponential law with component-specific expectation λ_i^{-1}. Let the claim arrival times be given as

$$\tau_{ij} := \inf\{t > 0 : E_{i1} + \cdots + E_{ij} \leq \Lambda_t\}, \quad i = 1, \dots, d, \quad j \in \mathbb{N}. \tag{3.3}$$

Then the claim number process $\boldsymbol{L}$ resulting from Equation (3.2) can be represented as in Model (M) with the components of $\boldsymbol{N}$ defined by the inter-arrival times E_{ij}:

$$N_t^i := \sum_{j=1}^{\infty} \mathbb{1}_{\{E_{i1} + \cdots + E_{ij} \leq t\}}, \quad t \geq 0.$$

Proof. The processes N_t^i are by definition independent Poisson processes. For any $t \geq 0$ it holds:

$$\{\tau_{ij} \leq t\} = \{\inf\{s > 0 : E_{i1} + \cdots + E_{ij} \leq \Lambda_s\} \leq t\} = \{E_{i1} + \cdots + E_{ij} \leq \Lambda_t\}.$$

It follows for the processes defined by Equation (3.2):

$$L_t^i := \sum_{j=1}^{\infty} \mathbb{1}_{\{\tau_{ij} \leq t\}} = \sum_{j=1}^{\infty} \mathbb{1}_{\{E_{i1}+\cdots+E_{ij} \leq \Lambda_t\}} = N_{\Lambda_t}^i, \quad t \geq 0,$$

As this is the representation in Model (M), the proof is established. □

By exploiting this probabilistic construction and the conditional independence, a simulation routine for the multivariate claim number process $\boldsymbol{L}$ can easily be implemented using the following algorithm. It should be noted that the algorithm equally works for other kinds of directing processes than Lévy subordinators. All that is necessary is a sampling routing for the chosen process.

Algorithm 3.3 (Sampling routine)
Let $T > 0$ be some fixed time horizon. A sample path $\{\hat{\boldsymbol{L}}_t\}_{t\in[0,T]}$ of the time-changed process $\boldsymbol{L}$ can be generated using the following steps:

(1) Simulate a path $\{\hat{\Lambda}_t\}_{t\in[0,T]}$ of the Lévy subordinator Λ on $[0, T]$.

(2) Repeatedly draw independent exponentially distributed trigger steps with parameter λ_i for each portfolio $i = 1, \ldots, d$ until their sum exceeds the level $\hat{\Lambda}_T$ of the subordinator path at time T. Compute the claim arrival times according to Equation (3.3).

(3) Determine the sample path $\{\hat{\boldsymbol{L}}_t\}_{t\in[0,T]}$ from the sampled claim arrival times using Equation (3.2).

As stated in Theorem 2.4, given the total number of jumps of a Poisson process within a certain time interval, the arrival times of these jumps have the distribution of the order statistics of a corresponding number of random variables uniformly distributed on the respective time interval. Using this, Step (2) in Algorithm 3.3 can be replaced by:

(2′) For each portfolio $i = 1, \ldots, d$, draw the total number $\hat{L}_T^i$ of claims at time T from a $\text{Poi}(\lambda_i \hat{\Lambda}_T)$ distribution. Then draw $\hat{L}_T^i$ independent samples $\hat{U}_{ij}$, $j = 1, \ldots, \hat{L}_T^i$, from a uniform distribution on $[0, \hat{\Lambda}_T]$. For the order statistics $U_{i(j)}$, evaluate:

$$\hat{\tau}_{ij} := \inf\{t > 0 : \hat{U}_{i(j)} \leq \hat{\Lambda}_t\}.$$

Using this algorithm, paths of $\boldsymbol{L}$ can be sampled quickly. For instance, it took only 43 seconds on a standard computer (2.4 GHz Intel Core 2 Duo processor, 4.00 GB RAM) to draw the 5000 bivariate samples for $\boldsymbol{\lambda} = (75, 100)'$, time interval $[0, 1]$, and with an inverse Gaussian subordinator generated for Figure 4.3. If certain relevant quantities cannot be calculated analytically, they can, hence, be easily approximated employing a Monte Carlo simulation. Furthermore, increasing the dimension d by adding an additional portfolio to the model is particularly convenient as it requires only one additional run through Steps (2) and (3), without changing the previous sampling results. Figure 3.3 provides an illustration of the algorithm. In the upper left chart, one sample path of an inverse Gaussian subordinator is plotted for a time horizon of $T = 1$. The subordinator parameters have been chosen such that time normalization holds. In the upper right chart, sampled trigger levels for the first portfolio are added to the graph, five in this case. On the left-hand side of the middle row, the first-passage times of the subordinator are evaluated to yield the claim arrival times for portfolio one. As the subordinator overshoots trigger levels two and three with a single jump, these two claims arrive simultaneously. The chart on the middle row on the right-hand side shows the claim (cluster) arrivals of L_t^1. In the bottom row, a sample of a second portfolio is added to the existing results. In this case two trigger levels are sampled before time T and the first happens to be overshot by the same subordinator jump as trigger level one of portfolio one, leading to simultaneous claim arrivals in both portfolios.

If Λ is a finite activity subordinator, i.e. a (compound) Poisson process (with drift), it can be directly sampled in terms of its arrival times similarly to Step (2) or Step (2′), only adding iid jump sizes and accounting for the drift, if applicable. In this case the whole sample path $\{\hat{\Lambda}_t\}_{t\in[0,T]}$ of the subordinator is known, which allows for the time-changed process to be sampled without discretization error. For an infinite activity subordinator, a sample path needs to be evaluated on a discrete time grid. For instance, let $K \in \mathbb{N}$ be the chosen number of simulation steps and $h := T/K$ the resulting step size on an equidistant grid. Then $\{\hat{\Lambda}_{t_k}\}_{k=1,\ldots,K}$ for $t_k := hk$ can be sampled by accumulating iid increment samples of the distribution of Λ_h. By evaluating the first-passage times, however, arrivals within any time step $(t_{k-1}, t_k]$ will be delayed to the next grid point t_k. Hence, the simulated path of the claim number process on non-grid points will always be less than or equal to the real path. Furthermore, the discretization error leads to more claims arriving simultaneously than predicted by the model. Hence, the real distribution of the jump sizes is always stochastically dominated in first-order by the empirical distribution generated from the simulation, that is the empirical cumulative distribution function is

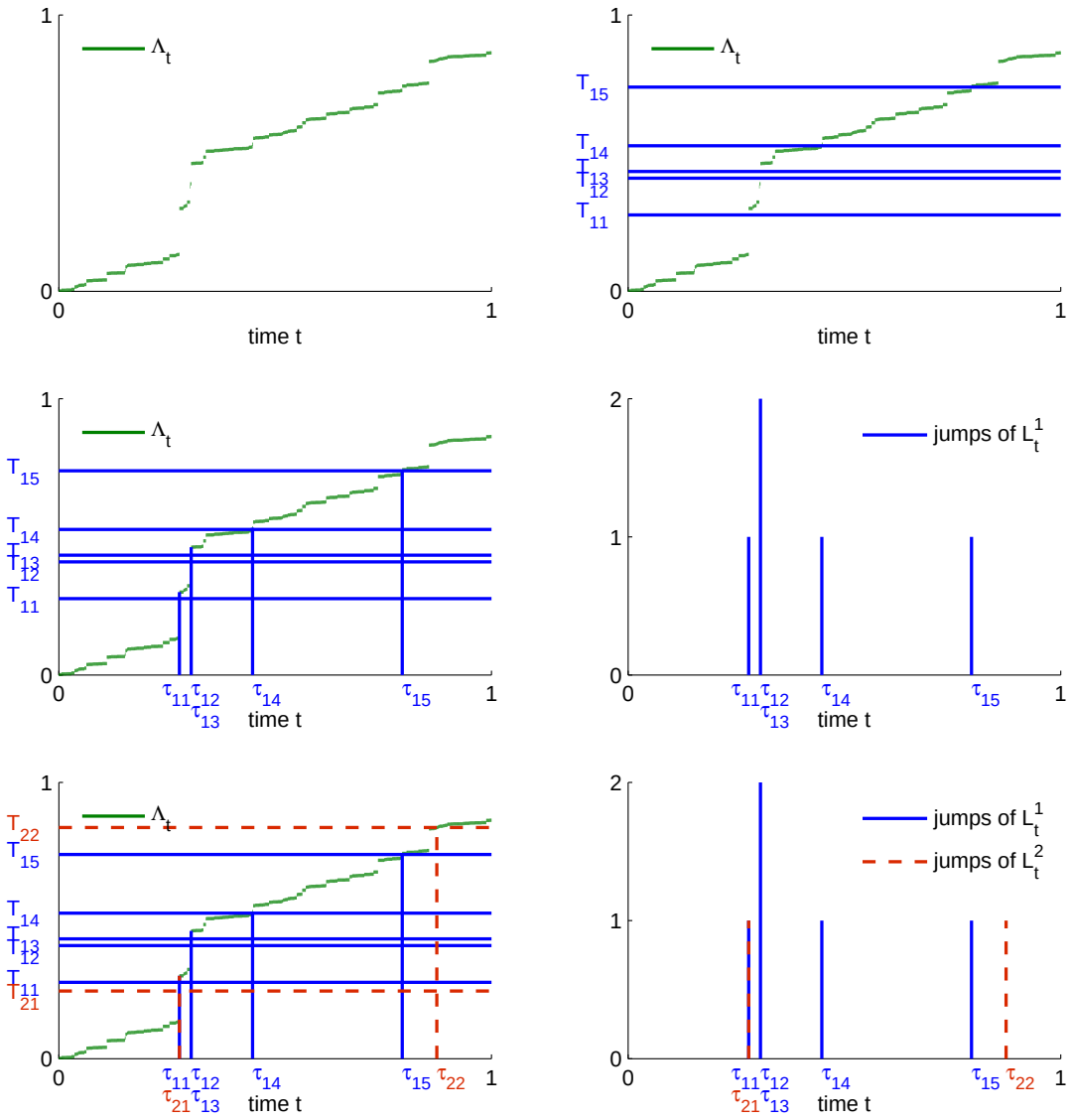


Figure 3.3 Sampling algorithm: The figure illustrates the steps in Algorithm 3.3. The upper left chart shows one sample path of an inverse Gaussian subordinator with no drift ($b = 0$, $\beta = \eta = 3$) and with time horizon $T = 1$. The upper right chart visualizes the trigger steps sampled for the first component with intensity $\lambda_1 = 4$ and the middle left chart adds the evaluated claim arrival times; the chart on the right-hand side shows the claim (cluster) arrivals. In the bottom row charts, a sample of a second portfolio with intensity $\lambda_2 = 2$ is added.

pointwise smaller than the real one. By choosing a sufficiently small step size, however, the discretization error should be small. For more information on how to sample from Lévy processes, see (Cont and Tankov, 2003, Chapter 6) or (Schoutens, 2003, Chapter 8).

The inter-arrival times of the time-changed process are no longer iid, thus the process is no renewal process. The distribution of the arrival and the inter-arrival times will be discussed in Chapter 4. This section concludes with a closer look at the conditional arrival times. Due to the conditional independence property of the model, the sequences of claim arrival times $(\tau_{i1}, \tau_{i2}, \dots)$ are independent for $i = 1, \dots, d$ given the σ-algebra $\sigma(\Lambda)$. From the transformation theorem for densities (see (Czado and Schmidt, 2011, Chapter 1.1, Theorem 1.2, p.5)) it follows for the joint density of the first successive $n \in \mathbb{N}$ trigger levels $(E_{i1}, E_{i1} + E_{i2}, \dots, E_{i1} + \cdots + E_{in})$:

$$f(t_1, \dots, t_n) := \lambda_i^n \exp\{-\lambda_i t_n\} \mathbb{1}_{S_n}(t_1, \dots, t_n), \quad t_1, \dots, t_n \in \mathbb{R},$$

where $S_n := \{s_1, \dots, s_n \in \mathbb{R} : 0 \leq s_1 \leq \cdots \leq s_n\}$. It follows that for the conditional joint survival probability of the first n claim arrival times $(\tau_{i1}, \dots, \tau_{in})$ it holds:

$$\mathbb{P}(\tau_{i1} > t_1, \dots, \tau_{in} > t_n | \sigma(\Lambda)) = \int_{\Lambda_{t_1}}^{\infty} \dots \int_{\Lambda_{t_n}}^{\infty} f(s_1, \dots, s_n) \, \mathrm{d}s_n \cdots \mathrm{d}s_1.$$

Due to jumps in the subordinator paths, no conditional density exists. For any single claim number $j \in \mathbb{N}$, the sum of j independent $\mathrm{Exp}(\lambda_i)$-distributed random variables has an Erlang distribution with shape parameter j and rate λ_i. Thus, given $\sigma(\Lambda)$, the arrival times τ_{ij} are conditionally independent for all $i = 1, \dots, d$ with conditional survival probability:

$$\mathbb{P}(\tau_{ij} > t | \sigma(\Lambda)) = \sum_{k=0}^{j-1} \frac{(\lambda_i \Lambda_t)^k}{k!} \exp\{-\lambda_i \Lambda_t\}, \quad t \geq 0.$$

3.3 Literature review

This sections provides references relevant to the topic of this thesis. Due to the vast amount of work in this field, the overview can by no means be comprehensive and only touches upon several important areas. Furthermore, many more references will be given throughout the thesis for specific aspects where they arise.

At a fixed point in time, the process presented in this thesis follows a multivariate mixed Poisson distribution with joint mixing variable. This class of distributions can be generalized to multivariate mixing distributions, see Karlis and Xekalaki (2005) and references therein for the bivariate case, in particular regarding mixtures with gamma, inverse Gaussian, and Poisson distributions generating bivariate negative binomial (a special case of the negative multinomial distribution as discussed in (Johnson et al., 1997, Chapter 36, pp.93)), Sichel, and Neyman type A distributions (cf. Remark 4.8). For more details, see also Kocherlakota (1988). Dey and Chung (1992) investigate multivariate mixed Poisson distributions, in particular with a gamma-distributed mixing variable, and an estimation approach for the intensities; they also develop a general formula for the multivariate probability mass function (cf. Theorem 4.2). Partrat (1994), for instance, examine compounding with multivariate mixed Poisson distributions for actuarial applications.

Randomizing joint parameters is merely one approach to generate dependency in (discrete) random variables; another route is via superposition of independent random variables in a shock model approach, or more generally using copula functions. In the following review, however, the focus is on time-dynamic processes rather than static random vectors. In particular, an overview over the technique of random time-change in general and with respect to Cox processes in particular is given and different methods discussed in the literature for generating tractable multivariate claim number processes are summarized. Finally, the occurrence of Cox processes in credit risk modelling and in particular the Lévy-frailty model, which is closely related to the model examined in this thesis, will be discussed.

Bochner (1949, 1955) introduced and studied the transformation of a process through a random time-change with an independent Lévy subordinator, called subordination, and Clark (1973) first exploited the method for modelling speculative price series. A more recent survey of the development of Lévy financial asset models through time-changing Brownian motions can be found in Carr and Wu (2004). Prause (1999), Luciano and Schoutens (2006), Luciano and Semeraro (2007), and Semeraro (2008), to name only a few, present multivariate models in this field; the latter two consider subordination employing multivariate subordinators.

The Poisson process (which is itself a Lévy subordinator) is the process most frequently studied under operational time in the class of point processes. In Section 2.2, the univariate inhomogeneous Poisson process was introduced as a Poisson process time-changed with a deterministic mean-value function and the Cox process was presented as a Poisson process directed by a suitable mean-value process. For the latter, several specific choices for the directing process have been studied in the literature, mostly defined via an intensity process. One common example are shot-noise intensities that are defined as

$$\lambda_t := \sum_{j=1}^{N_t} X_j f(t - T_j), \quad t \geq 0,$$

for arrival times T_j, $j \in \mathbb{N}$, commonly specified via a second Poisson process, a sequence of iid random variables X_j, $j \in \mathbb{N}$, independent of the arrival times, and a deterministic function f with $f(t) = 0$ for $t < 0$. In many applications $f(t) := \exp\{-\alpha t\}$ for $\alpha > 0$, i.e. the intensity jumps at times T_j with magnitudes X_j and the effects wear off over time, with exponential decay rate α. Dassios and Jang (2003) propose shot-noise driven Cox processes as claim arrival processes for catastrophic events; the resulting risk process and ruin probabilities are investigated in Albrecher and Asmussen (2006). Moller et al. (1998) examine log-normal intensity processes and also consider a bivariate extension. Actuarial applications, in particular pricing of stop-loss contracts, for shot-noise driven Cox processes are studied in Basu and Dassios (2002). Hellmund et al. (2008) propose a unifying approach where the intensity process is defined as an integral w.r.t. a Lévy basis and Dario, A. De G. and Simonis (2011) focus on affine intensity processes, in particular Feller processes.

However, a Cox process with intensity process differs from the Cox processes presented in this thesis, which is directed by a Lévy subordinator, since the former does not support simultaneous claim arrivals due to the continuity of its mean-value process. Far fewer publications can be found which examine Cox processes with discontinuous time-change processes, even in the univariate case. Kumar et al. (2011) choose inverse Gaussian and stable subordinators as directing processes and derive the difference-differential equations for the probability mass function of the time-changed Poisson process. In Beghin and Macci (2014), fractional Poisson and Poisson cluster processes that result from a time-change with a stable subordinator (or its generalized inverse) and their probability mass functions are investigated, see also Orsingher and Polito (2012). Di Crescenzo et al. (2015) consider the time-change of a compound Poisson process with an independent Poisson process and find expressions for the probability mass function in terms of Bell polynomials (see Section 4.1).

Lee and Whitmore (1993) examine the more general case of time-changing a Markov process through a process with non-negative and independent increments. They also build a multivariate process by time changing independent univariate processes with a common randomized time. Closely related to this thesis, they give special attention to time-changing a Poisson process with so-called Hougaard processes, a subclass of Lévy subordinators including, for instance, the inverse Gaussian, stable, and – as a boundary case – gamma subordinator. The probability generating function as well as the compound Poisson representation (see Section 4.2) are derived in this setting for the univariate case. Furthermore, in Hougaard et al. (1997) a recursive formula for the probability mass function of a univariate mixed Poisson distribution with Hougaard mixing variable is presented. In addition, methods of statistical inference for these kinds of processes are discussed and applied to epileptic seizure count data.

Apart from using a common time-change as proposed in this thesis, other approaches have been employed to build tractable multivariate claim number processes for actuarial applications. The models mentioned in the following, however, exclude multiple simultaneous claim arrivals in any component and generally try to retain (mixed) Poisson marginals. For modelling dependencies in collective risk processes, Pfeiffer and Nešlehová (2004) suggest, among other approaches, a copula-based construction method to generate dependent Poisson processes: at a finite time horizon a copula is imposed on a random vector consisting of independent Poisson processes and the evolvement of the now dependent Poisson processes up to this time horizon is found using the order statistics property (see Theorem 2.4). Lindskog and McNeil (2003) follow a shock model approach where the superposition of independent Poisson processes is used to introduce dependence. More precisely, let $\tilde{N}(\boldsymbol{k}) = \{\tilde{N}_t(\boldsymbol{k})\}_{t\geq 0}$ for $\boldsymbol{k} \in \{0,1\}^d$ be independent Poisson processes with intensity $\tilde{\lambda}(\boldsymbol{k})$ counting the shocks (or claim arrivals) affecting all components where $\boldsymbol{k}$ has entry one. The resulting dependent claim number process $\boldsymbol{N}$ is defined via its components as

$$N_t^i := \sum_{\substack{\boldsymbol{k}\in\{0,1\}^d \\ k_i=1}} \tilde{N}_t^i(\boldsymbol{k}), \quad t \geq 0.$$

If all possible shocks are to be included, $2^d - 1$ processes have to be considered; hence, the number of parameters quickly increases with the dimensionality of the problem. Various extensions and modifications of this approach are investigated in the publication as well. Bäuerle and Grübel (2005) create a multivariate counting process with Poisson marginales by thinning and shifting arrival points of a univariate Poisson process. They investigate the dependence properties of the resulting process and also explore some actuarial applications. Shifting of arrival points generates dependence over time, a property Lévy processes do not support. In Bäuerle and Grübel (2008), a multivariate pure birth

process with dependencies between the marginal birth rates is introduced and its applications to actuarial questions is examined. Zocher (2005) studies multivariate counting processes in general and multivariate mixed Poisson processes with joint mixing variable in particular. While the evolution of the mixed Poisson process over time differs from the multivariate Cox process studied in this thesis, the distribution at a fixed point in time is mixed Poisson in both cases and the probability mass function can be similarly (except for the component-specific intensities) expressed in terms of derivatives of a generating function of the mixing variable, cf. Section 2.2. Zocher also investigates a generalization with a multivariate mixing distribution.

Since (compound) Poisson processes are Lévy processes, a potential approach to generating multivariate processes with marginals of this kind is via Lévy copulas (see Tankov (2003), Kallsen and Tankov (2006), and Section 4.2). As for the static case of random vectors and copulas, Lévy copulas make it possible for multivariate Lévy processes to specify the dependence structure separately from the marginal processes. Hence, this approach allows retaining marginal Poisson processes, if desirable, or choosing compound Poisson or Poisson cluster processes instead. Bäuerle and Blatter (2011) use Lévy copulas to generate multivariate risk processes in actuarial applications and study optimal control problems; Bregman and Klüppelberg (2005) consider the estimation of ruin probabilities, in particular for a Clayton Lévy copula; Esmaeili and Klüppelberg (2010) focus on parameter estimation. Avanzi et al. (2011) compare different Lévy copulas and study their fit to a real-world insurance data set. Closely related, Lévy copulas and compound Poisson marginals are used to model the joint dynamics of operational risks, for instance in Böcker and Klüppelberg (2008).

Cox processes, usually with an existing intensity process, have been successfully applied in credit risk modelling, where the first jump time of the process defines the default time of a credit-risky asset (or a Poisson approximation is used, see Remark 6.9), see Lando (1998). To generate tractable multivariate models for this application, Giesecke and Tomecek (2005) follow a top-down approach by proposing a Poisson process with a continuous time-change as aggregate process and applying random thinning to generate consistent models for the underlying components, see also Giesecke et al. (2011). Multivariate Cox processes with dependent intensity processes were studied as a bottom-up approach in Duffie and Gârleanu (2001), Jarrow and Yu (2001), or Das et al. (2007), among others. In Mai and Scherer (2009b), a time-change with a common Lévy subordinator and, hence, no intensity process is presented. As this construction is directly linked to the process studied in this thesis, it will be presented in more detail in the following subsection. However, in credit risk modelling the focus by nature lies on the first (or – in case of a top-down approach – first few) arrival times of the Cox process, whereas for actuarial applications later arrival times are equally important.

Extension of the Lévy-frailty model

Model (M) is a natural extension of the Lévy-frailty model introduced in Mai and Scherer (2009b) and further studied in Mai and Scherer (2009a). This model was designed for application in credit-risk modelling. Thus, rather than the claim arrival times of a portfolio of insurance policies, the credit default times of a portfolio of d credit-risky assets are considered. While there may be multiple claims on an insurance policy, a loan can default only once, hence, only a model for the vector of random ('first') default times $(\tau_1, \dots, \tau_d) \in \mathbb{R}_+^d$ is necessary. In the Lévy-frailty model, the default times are constructed as first-passage times of a Lévy subordinator $\Lambda = \{\Lambda_t\}_{t \geq 0}$ over iid exponential trigger variables $E_i \overset{iid}{\sim} \text{Exp}(1)$, i.e.

$$\tau_i := \inf\{t > 0 \colon E_i \leq \Lambda_t\}, \quad i = 1, \dots, d.$$

From Proposition 3.2 it naturally follows that the default times in this definition correspond to the first claim arrival times $\tau_{11}, \dots \tau_{d1}$ in Model (M) with unit intensities $\lambda_1 = \cdots = \lambda_d = 1$.

The Lévy-frailty model can account for many observed default patterns while still being mathematically tractable and most of the tractability can be preserved in the extension presented here. The extension of the sampling algorithm, which is available for the Lévy-frailty model, to the process proposed here has already been presented in Algorithm 3.3. Furthermore, the vector of default times in the Lévy-frailty model follows the Marshall–Olkin distribution, which is characterized by its multivariate lack-of-memory property, see Marshall and Olkin (1967). The survival copula can be derived in closed form as follows:

$$\hat{C}(\boldsymbol{u}) = \prod_{i=1}^{d} u_{(i)}^{\{\Psi_\Lambda(i) - \Psi_\Lambda(i-1)\}/\Psi_\Lambda(1)}, \quad \boldsymbol{u} \in [0,1]^d,$$

where $u_{(i)}$ denotes again the order statistics and Ψ_Λ is the Laplace exponent of Λ. The copula has a singular component as $\mathbb{P}(\tau_1 = \cdots = \tau_d) > 0$, thus the model accounts for default clustering. Furthermore, the copula has a positive lower tail dependence:

$$LTD(\tau_i, \tau_j) = 2 - \frac{\Psi_\Lambda(2)}{\Psi_\Lambda(1)}, \quad i \neq j \in \{1, \dots, d\}.$$

The distribution in Model (M) as well as clustering and joint early claim arrivals will be studied in detail in Sections 4.1 and 4.2.

The loss given default of credit-risky assets is often assumed to be a deterministic constant and equal for all assets. Hence, considering as portfolio loss process $\bar{L}_t$ the number of defaults that have occurred up to time t is sufficient:

$$\bar{L}_t := \sum_{i=1}^{d} \mathbb{1}_{\{\tau_i \leq t\}}, \quad t \geq 0.$$

For the pricing of credit derivatives, expectations $\mathbb{E}[f(\bar{L}_t)]$ for potentially non-linear functions $f(\cdot)$ have to be evaluated. Of particular interest in credit applications is the pricing of collateralized debt obligations, which correspond to evaluating collar type functions $f(x) = \min(\max(0, x-l), u-l)$ for some $u > l \geq 0$. Thus, the distribution of the portfolio loss, which is available in closed form in the Lévy-frailty model, is needed. In addition, an approximation for large homogeneous portfolios can be found: for increasing number d of assets, the normalized portfolio loss $\frac{1}{d}\bar{L}_t$ converges uniformly and in L^2 to $1-\exp\{-\Lambda_t\}$. The distribution of the aggregate process in Model (M) together with the distribution of the process itself will be examined in detail in Section 4.1 and a large portfolio approximation will be derived in Section 4.3. Stop-loss reinsurance contracts are the equivalent of collateralized debt obligations in actuarial applications, and these contracts will be investigated in Section 6.1[3]. For insurance modelling, however, the assumption of constant and identical claim sizes is not realistic, so at least iid claim size distributions need to be incorporated in the model. This extension will be discussed in Chapter 6.

In credit risk modelling, the model parameters are usually determined from calibration to market data: the parameters are chosen such that the model-predicted prices closely match the observed market prices for certain liquidly traded credit derivatives. As insurance products are traditionally not liquidly traded, the model parameters are usually estimated from historical portfolio data. Therefore, different estimation procedures for Model (M) will be presented in Chapter 5.

Finally, it is worth mentioning that the Lévy-frailty model can be incorporated in the general framework of conditionally iid default models, see Mai et al. (2013). For a non-decreasing càdlàg process $\{F_t\}_{t\geq 0}$ with $F_0 = 0$, $\lim_{t\to\infty} F_t = 1$ a.s. and iid uniform random variables $U_i \overset{iid}{\sim} \mathrm{U}[0,1]$ independent of the process, the default times are defined as:

[3] It should be mentioned that pricing in financial applications is usually done w.r.t. a risk-neutral pricing measure, whereas in actuarial applications the physical measure is used. An overview of the different pricing approaches can be found in Embrechts (1993).

$$\tau_i := \inf\{t > 0 \colon U_i \leq F_t\}, \quad i = 1, \dots, d.$$

For the Lévy-frailty model $F_t := 1 - \exp\{-\Lambda_t\}$ and most of the results discussed above for this model can be recovered in the generalized setting. Hence, many extensions and variations of the Lévy-frailty model remain tractable, for instance, considering $F_t := 1 - \exp\{-\Lambda_{Mt}\}$ for an independent mixing variable $M > 0$ as in Bernhart et al. (2013) or $F_t := 1 - \exp\{-\Lambda_{\int_0^t \lambda_s \, \mathrm{d}s}\}$ for a suitable positive process $\lambda = \{\lambda_t\}_{t \geq 0}$ as in Mai et al. (2014). Extensions of this kind cannot always be easily transferred to Model (M), as they usually destroy the Lévy property of the process which is fundamental for many of the results derived throughout this thesis. Possible extensions of the model and their limitations will be discussed in Sections 4.1 and 6.2.

4 Distribution and properties

In this chapter, the properties of the proposed model are studied in detail. For the application and estimation of the process it is essential to fully understand how the model behaves and to be able to handle it analytically. Section 4.1 focuses on the the finite-dimensional distribution of the process and develops closed formulas for the probability mass function and many related quantities. In Section 4.2 the Lévy characteristics of the time-changed model are derived, which directly lead to a second stochastic representation of the process as multivariate Poisson cluster process. This representation not only offers a deeper understanding of the model dynamics and its dependence structure, it also enables a robust and speedy computation of the process distribution. In Section 4.3 the convergence of the aggregate process for increasing model dimension is examined and a large portfolio approximation is presented.

4.1 Process distribution

This section contains a thorough analysis of the process distribution. In Algorithm 3.3 an efficient sampling routine for the time-changed process $\boldsymbol{L}$ was given, so the distribution of the process can always be approximated using the empirical distribution generated from a Monte Carlo simulation. For certain applications, however, e.g. estimation procedures as will be presented in Chapter 5, it may be convenient or even necessary to have a faster way of calculating probabilities. In the given model, an analytic expression for the probability mass function is available and an efficient way of quickly evaluating it will be established.

As discussed in Section 3.1, Model (M) defines a multivariate Lévy process. The process is even a multivariate Lévy subordinator, as it has component-wise almost surely non-decreasing paths. We start by deriving the Laplace exponent, which determines the Laplace transform and thus characterizes the distributional properties of the process.

Proposition 4.1 (Laplace exponent of the claim number processes)
The claim number process $\boldsymbol{L}$ is a d-dimensional Lévy subordinator with Laplace exponent

$$\Psi_{\boldsymbol{L}}(\boldsymbol{x}) = \Psi_\Lambda\Big(\sum_{i=1}^{d} \lambda_i(1 - \exp\{-x_i\})\Big), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

Proof. It can be found, for instance, in (Sato, 1999, Chapter 6, Theorem 30.4) that a Lévy subordinator time-changed by an independent Lévy subordinator is itself a Lévy subordinator and the Laplace exponent of the new process is found by composition, that is

$$\Psi_{\boldsymbol{L}}(\boldsymbol{x}) = \Psi_\Lambda(\Psi_{\boldsymbol{N}}(\boldsymbol{x})), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

The claim follows by plugging in the Laplace exponent of the Poisson process $\boldsymbol{N}$, which is, due to independence, the sum of the exponents of the marginal Poisson processes given in Example 2.14:

$$\Psi_{\boldsymbol{N}}(\boldsymbol{x}) = \sum_{i=1}^{d} \Psi_{N^i}(x_i) = \sum_{i=1}^{d} \lambda_i(1 - \exp\{-x_i\}).$$

□

It directly follows from Proposition 4.1 for the marginal processes L^i and, together with Remark 3.1, also for the aggregate claim number process $\bar{L}$:

$$\Psi_{L^i}(x) = \Psi_\Lambda(\lambda_i(1 - \exp\{-x\})), \quad \Psi_{\bar{L}}(x) = \Psi_\Lambda(|\boldsymbol{\lambda}|(1 - \exp\{-x\})), \quad x \geq 0.$$

As a Lévy process, the proposed model has independent and stationary increments. This property, which is largely responsible for the good mathematical tractability of the model, is useful for many applications as it guarantees model consistency over time and exposure size. More precisely, according to the model the distribution of the number of claims over a certain time period like one year does not change over time and does not influence the distribution in any following year. The Lévy property ensures that the type of distribution does not depend on the length of the time interval considered and that the expected claim frequency is proportional to this length. Furthermore, if the business volume is increased by adding additional entities to the portfolio then – all else being equal – the expected number of claims in the model will increase accordingly (see (Klugman et al., 2004, Chapter 4.6.11, pp.108)). However, these assumptions may be too restrictive for certain applications, for instance, if contagion effects are observed, as may be the case in

motor insurance policies where an accident caused by a policyholder increases the likelihood of future accidents, or if a seasonal effect influences claim arrivals that originate from certain types of natural catastrophes. When modelling motor insurance claim numbers, (Denuit et al., 2007, Chapter 2.9, pp.90) argue that independence can be assumed between different policyholders, though assuming independence over time for a single policyholder seems questionable. Chapter 6 sketches some extensions of the model which incorporate such properties by relaxing the Lévy property.

From the results of Proposition 4.1, the distributional properties of the model have already been fully characterized and the probability mass function can be derived from the Laplace transform employing numerical inversion techniques, see for instance Abate and Whitt (1992) and Widder (1952). However, by exploiting the conditional independence of the construction, the distribution of the claim number processes can be obtained explicitly with regard to derivatives of the Laplace transform of the selected subordinator and thus, ultimately, in terms of derivatives of its Laplace exponent, which often can be calculated more easily. The following Theorem 4.2 provides the distribution of the process at one fixed point in time.

Theorem 4.2 (Distribution of the claim number process)
The distribution of the claim number process $\boldsymbol{L}$ *in Model* (M) *at some fixed time* $t \geq 0$ *is given by*

$$\mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k}) = \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \varphi_{\Lambda_t}^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|) \tag{4.1}$$

$$= \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \varphi_{\Lambda_t}(|\boldsymbol{\lambda}|) B_{|\boldsymbol{k}|}(-t\Psi_{\Lambda}^{(1)}(|\boldsymbol{\lambda}|), \ldots, -t\Psi_{\Lambda}^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|)), \quad \boldsymbol{k} \in \mathbb{N}_0^d, \tag{4.2}$$

where $B_{|\boldsymbol{k}|}(\cdot)$ *denotes the* $|\boldsymbol{k}|$*-th complete Bell polynomial.*

Proof. Due to the tower rule for conditional expectations and the conditional independence of the components, it follows for the claim number vector $\boldsymbol{L}_t$:

$$\mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k}) = \mathbb{E}[\mathbb{P}(N_{\Lambda_t}^1 = k_1, \ldots, N_{\Lambda_t}^d = k_d | \sigma(\Lambda))] = \mathbb{E}\left[\prod_{i=1}^d \mathbb{P}(N_{\Lambda_t}^i = k_i | \sigma(\Lambda))\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^d \frac{\lambda_i^{k_i}}{k_i!} \Lambda_t^{k_i} \exp\{-\lambda_i \Lambda_t\}\right] = \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!} \mathbb{E}[\Lambda_t^{|\boldsymbol{k}|} \exp\{-|\boldsymbol{\lambda}|\Lambda_t\}].$$

The last term in the above equation can be expressed in terms of the $|\boldsymbol{k}|$-th derivative of the Laplace transform φ_{Λ_t} of Λ_t. In particular, for all $x > 0$ it holds:

$$\varphi_{\Lambda_t}^{|\boldsymbol{k}|}(x) = (-1)^{|\boldsymbol{k}|}\, \mathbb{E}[\Lambda_t^{|\boldsymbol{k}|} \exp\{-x\Lambda_t\}]. \tag{4.3}$$

This follows by induction considering that differentiation inside the expectation operator is possible as all derivatives exist and are bounded and continuous and thus the differentiation lemma for parameter dependent integrals can be applied, see (Feller, 1971, Chapter XIII.2, p.435), for instance. Together with the calculation before, Equality (4.3) proves Equality (4.1). The second part follows from Faà di Bruno's Formula for multivariate derivatives of composition functions, as the Laplace transform of a subordinator is a composition of the exponential function and the Laplace exponent. In particular, it can be derived from Riordan (1946) or Johnson et al. (1992) that for a smooth function $f(x)$ it holds for $k \in \mathbb{N}_0$:

$$\frac{\mathrm{d}^k}{\mathrm{d}^k\, x} \exp\{f(x)\} = \exp\{f(x)\} B_k(f^{(1)}(x), \ldots, f^{(k)}(x)).$$

Using $f(x) := -t\Psi_\Lambda(x)$, which is smooth as Ψ_Λ is a Bernstein function as mentioned in Section 2.3, proves Equality (4.2).

□

Note that the probabilities in Theorem 4.2 are indeed non-negative, as the Laplace transform φ_{Λ_t} is a completely monotone function, i.e. the derivatives exist and alternate in sign, see Theorem 2.2. The distribution of the marginal processes as well as the aggregate process can again be directly concluded from the multivariate results. In the following we will discuss the evaluation of Equation (4.1) and (4.2). It is, however, worth mentioning that Equation (4.3) offers an alternative approach to calculating the probabilities numerically by means of a Monte Carlo simulation for the subordinator only. Compared to a simulation of the full model as discussed in the beginning of this section, this approach does not require drawing trigger levels for all components and evaluating first-passage times as in step (2) and (3) of Algorithm 3.3. While this approach comes at the additional cost of evaluating Equation (4.1), it will often be more efficient.

Remark 4.3 (Generalization of the directing process)
The proof of Equality (4.1) does not make any use of the special property of Λ being a Lévy subordinator but rather holds for any non-negative directing process as long as it is independent of $\boldsymbol{N}$. If $g_{\Lambda_t}(x)$ is the cumulant generating function of the directing process at time $t \geq 0$ and $\tilde{g}_{\Lambda_t}(x) := g_{\Lambda_t}(-x)$ which exists for all $x \geq 0$, then Equality (4.1) holds as

$$\mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k}) = \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\varphi_{\Lambda_t}(|\boldsymbol{\lambda}|)B_{|\boldsymbol{k}|}(\tilde{g}^{(1)}_{\Lambda_t}(|\boldsymbol{\lambda}|), \ldots, \tilde{g}^{(|\boldsymbol{k}|)}_{\Lambda_t}(|\boldsymbol{\lambda}|)).$$

In the special case of a Lévy subordinator it is $\tilde{g}_{\Lambda_t}(x) = -t\Psi_\Lambda(x)$. Furthermore, Proposition 4.1 can also be extended to this generalized case as it always holds due to the tower rule for conditional expectations for the Laplace transform of the time-changed process:

$$\begin{aligned}\varphi_{\boldsymbol{L}_t}(\boldsymbol{x}) &= \mathbb{E}[\exp\{-\boldsymbol{x}'\boldsymbol{N}_{\Lambda_t}\}] = \mathbb{E}[\mathbb{E}[\exp\{-\boldsymbol{x}'\boldsymbol{N}_{\Lambda_t}\}|\sigma(\Lambda)]] = \mathbb{E}[\varphi_{\boldsymbol{N}_{\Lambda_t}}(\boldsymbol{x})] \\ &= \mathbb{E}[\exp\{-\Lambda_t\Psi_{\boldsymbol{N}}(\boldsymbol{x})\}] = \varphi_{\Lambda_t}(\Psi_{\boldsymbol{N}}(\boldsymbol{x})).\end{aligned}$$

▲

Due to the infinite divisibility property, the distribution of any Lévy process is determined by its distribution at one point in time and thus Theorem 4.2 already fully defines the process distribution. By exploiting the Lévy properties, the finite-dimensional distribution of the process is given explicitly in the following corollary.

Corollary 4.4 (Finite-dimensional distribution of the claim number process)
Let $n \in \mathbb{N}$, $0 := t_0 \leq t_1 \leq \cdots \leq t_n$, and $\mathbf{0} := \boldsymbol{k}_0 \leq \boldsymbol{k}_1 \leq \cdots \leq \boldsymbol{k}_n$ for $\boldsymbol{k}_j \in \mathbb{N}_0^d$, $j = 1, \ldots, n$. Then the finite-dimensional distribution of the claim number process $\boldsymbol{L}$ is given as:

$$\mathbb{P}(\boldsymbol{L}_{t_1} = \boldsymbol{k}_1, \ldots, \boldsymbol{L}_{t_n} = \boldsymbol{k}_n) = (-\boldsymbol{\lambda})^{\boldsymbol{k}_n}\prod_{j=1}^{n}\frac{1}{\Delta\boldsymbol{k}_j!}\varphi^{(|\Delta\boldsymbol{k}_j|)}_{\Lambda_{\Delta t_j}}(|\boldsymbol{\lambda}|).$$

Proof. As was shown in Proposition 4.1, the claim number process $\boldsymbol{L}$ is a Lévy processes. It follows by the independent and stationary increment property:

$$\begin{aligned}\mathbb{P}(\boldsymbol{L}_{t_1} = \boldsymbol{k}_1, \ldots, \boldsymbol{L}_{t_n} = \boldsymbol{k}_n) &= \mathbb{P}(\Delta\boldsymbol{L}_{t_1} = \Delta\boldsymbol{k}_1, \ldots, \Delta\boldsymbol{L}_{t_n} = \Delta\boldsymbol{k}_n) \\ &= \prod_{j=1}^{n}\mathbb{P}(\boldsymbol{L}_{\Delta t_j} = \Delta\boldsymbol{k}_j).\end{aligned}$$

Together with the univariate results of Theorem 4.2 the claim is established. □

Remark 4.5 (Additive directing process)
The results of Corollary 4.4 no longer hold for general time-change processes like in Remark 4.3, as we make use of both Lévy increment properties, independence as well as stationarity. However, if the assumption of stationarity is relaxed and Λ is only assumed to be an additive process with non-decreasing paths, a similar result can still be found. For instance, a Lévy process time-changed by a continuous increasing function $v\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ starting in zero, that is $v(0) = 0$, is an additive process, see (Sato, 1999, Chapter 1, Example 1.7, p.4). Then it still holds by the independent increment property and Remark 4.3:

$$\mathbb{P}(\boldsymbol{L}_{t_1} = \boldsymbol{k}_1, \dots, \boldsymbol{L}_{t_n} = \boldsymbol{k}_n) = (-\boldsymbol{\lambda})^{\boldsymbol{j}_n} \prod_{j=1}^{n} \frac{1}{\Delta \boldsymbol{k}_j!} \varphi_{\Delta\Lambda_{t_j}}^{(\Delta \boldsymbol{k}_j)}(|\boldsymbol{\lambda}|).$$

For evaluating this formula, the Laplace transform of the increments needs to be known. In general, for an additive process $\boldsymbol{X}$ on $\mathbb{R}^d_{\geq 0}$ with non-decreasing paths it can be shown that at each point in time $t \geq 0$, the distribution of $\boldsymbol{X}_t$ is still infinitely-divisible and, hence, a Lévy–Khintchine representation of the Laplace transform can be given, but only in terms of time-dependent characteristics $(\boldsymbol{b}_{\boldsymbol{X}_t}, \nu_{\boldsymbol{X}_t})$, called spot characteristics. More precisely,

$$\varphi_{\boldsymbol{X}_t}(\boldsymbol{x}) = \exp\{-\Psi_{\boldsymbol{X}_t}(\boldsymbol{x})\} \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0},$$

with an exponent that is no longer necessarily linear in time:

$$\Psi_{\boldsymbol{X}_t}(\boldsymbol{x}) = \boldsymbol{x}' \boldsymbol{b}_{\boldsymbol{X}_t} + \int_{\mathbb{R}^d_{\geq 0}} (1 - \exp\{-\boldsymbol{u}'\boldsymbol{x}\})\, \nu_{\boldsymbol{X}_t}(\mathrm{d}\boldsymbol{x}).$$

The spot characteristics uniquely determine the law of the process. For a Lévy subordinator it is $\boldsymbol{b}_{\boldsymbol{X}_t} = t\boldsymbol{b}_{\boldsymbol{X}_1}$ and $\nu_{\boldsymbol{X}_t} = t\nu_{\boldsymbol{X}_1}$. The distribution of the increments $\boldsymbol{X}_t - \boldsymbol{X}_s$, $0 \leq s < t$, is infinitely divisible as well with the so-called forward characteristics $(\boldsymbol{b}_{\boldsymbol{X}_t} - \boldsymbol{b}_{\boldsymbol{X}_s}, \nu_{\boldsymbol{X}_t} - \nu_{\boldsymbol{X}_s})$. An overview over additive processes can be found in (Cont and Tankov, 2003, Chapter 14), for details see (Sato, 1999, Chapter 9). ▲

Following Corollary 4.4, even to deal with the finite-dimensional process distribution it is enough to be able to quickly evaluate Equation (4.1) or (4.2) for the multivariate process distribution at one point in time. The following corollary shows that, in fact, it is sufficient to calculate the distribution in dimension one.

Corollary 4.6 (Multivariate from univariate distribution)
For the distributions of the claim number process $\boldsymbol{L}$ in Model (M) and its aggregate process $\bar{L}$ it holds at any fixed time $t \geq 0$:

$$\mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k}) = \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{|\boldsymbol{\lambda}|^{|\boldsymbol{k}|}} \binom{|\boldsymbol{k}|}{\boldsymbol{k}} \mathbb{P}(\bar{L}_t = |\boldsymbol{k}|), \quad \boldsymbol{k} \in \mathbb{N}_0^d.$$

Proof. By Theorem 4.2 and Remark 3.1, it holds for the aggregate process $\bar{L}$ for all $t \geq 0$:

$$\begin{aligned}\mathbb{P}(\bar{L}_t = k) &= \frac{(-|\boldsymbol{\lambda}|)^k}{k!} \varphi_{\Lambda_t}^{(k)}(|\boldsymbol{\lambda}|) \\ &= \frac{(-|\boldsymbol{\lambda}|)^k}{k!} \varphi_{\Lambda_t}(|\boldsymbol{\lambda}|) B_k(-t\Psi_{\Lambda}^{(1)}(|\boldsymbol{\lambda}|), \ldots, -t\Psi_{\Lambda}^{(k)}(|\boldsymbol{\lambda}|)), \quad k \in \mathbb{N}_0. \end{aligned} \tag{4.4}$$

Comparing this with the result for the multivariate process in Theorem 4.2 establishes the claim. □

According to this corollary, it is sufficient to calculate the probabilities of the univariate model as the probabilities of multivariate models can be deduced by suitable rescaling, something that proves particularly useful for the efficient implementation of the probability mass function. Note that the relationship also holds true for the generalized setting of Remark 4.3. Furthermore, increasing the dimension of the model only changes the evaluation point $\boldsymbol{\lambda}$ of the Laplace transform derivatives and the scaling factor. Hence, rather than the high-dimensional case, the more challenging question is to efficiently calculate the tail of the distribution, as it requires high-dimensional derivatives of the Laplace transform, a problem that will be addressed specifically in the following. Before beginning, it should be noted that implementing the rescaling directly using built-in functions for factorials and multinomial coefficients quickly leads to numerical instabilities, as these quantities grow too explosively when considering the tail of the distribution. For instance, in MATLAB® the factorials from 171 onwards are treated as ∞. Jointly, the rescaling coefficient

$$c(\boldsymbol{k}) := \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{|\boldsymbol{\lambda}|^{|\boldsymbol{k}|}} \binom{|\boldsymbol{k}|}{\boldsymbol{k}} \tag{4.5}$$

is well-behaved, as can be established from the result in Corollary 4.6 due to the boundedness of probability measures. It can also be proven in general using the multinomial theorem which states that

$$|\boldsymbol{\lambda}|^{|\boldsymbol{k}|} = \sum_{|\boldsymbol{l}|=|\boldsymbol{k}|} \binom{|\boldsymbol{k}|}{\boldsymbol{l}} \boldsymbol{\lambda}^{\boldsymbol{l}},$$

where the sum is over all vectors $\boldsymbol{l} \in \mathbb{N}_0^d$ having the same 1-norm as $\boldsymbol{k}$, see (Comtet, 1974, Chapter 1.10, Theorem B, p.28). As all summands are non-negative and one of them corresponds to $\boldsymbol{k}$, it follows immediately that

$$0 \leq \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{|\boldsymbol{\lambda}|^{|\boldsymbol{k}|}} \binom{|\boldsymbol{k}|}{\boldsymbol{k}} \leq 1, \quad \boldsymbol{k} \in \mathbb{N}_0^d.$$

The numerical problems can be avoided by simply calculating the rescaling coefficient stepwise in a loop.

Algorithm 4.7 (Rescaling coefficient)
The rescaling coefficient $c(\boldsymbol{k})$, $\boldsymbol{k} \in \mathbb{N}_0^d$, can be computed iteratively using the following algorithm:

INPUT $\boldsymbol{\lambda}$; $\boldsymbol{k}$; d;
START $\lambda := |\boldsymbol{\lambda}|$; $K := |\boldsymbol{k}|$; $temp := 0$; $c(\boldsymbol{k}) := 1$;
FOR $i = 1, \ldots, d$
  FOR $j = 1, \ldots, k_i$
    $c(\boldsymbol{k}) = c(\boldsymbol{k}) * \frac{\lambda_i}{\lambda} * \frac{temp + j}{j}$;
  END
  $temp = temp + k_i$;
END
RETURN $c(\boldsymbol{k})$;

Of course, if the probability mass function is to be evaluated not only at a single point but rather on a whole grid, it is reasonable to store the results of any iteration to avoid unnecessarily recalculating the same quantities multiple times.

As the univariate distribution essentially governs the full model dynamics, we will take a closer look at it in the following. By definition, the marginal distributions in the general setting of Remark 4.3 are mixed Poisson distributions at each point in time. In the special setting chosen here, with the time-change process a Lévy subordinator, the mixing distribution is an infinitely divisible distribution; more precisely, the mixing distribution is the infinitely divisible distribution of the subordinator scaled by the component-specific intensity. Thus, the marginal distributions of the time-changed process vary only by this scaling factor. Univariate mixed Poisson distributions have been studied extensively and, hence, in this case results like Equation (4.2) have been known for some time, see for

instance (Grandell, 1997, Chapter 2, p.14). These kinds of distributions have been used for a wide range of applications where the Poisson distribution showed a poor fit due to the heterogeneity present in the population or portfolio considered. By multiplying the mean frequency of the Poisson distribution with a positive random variable, the model frequency varies and thus can account for the different individual frequencies observed in the portfolio. Usually, the expectation of the random effect is fixed at one to preserve on average the observed portfolio frequency, in line with our time-normalization Assumption (TN). A good overview of the properties of mixed Poisson distributions with further references is given in Karlis and Xekalaki (2005), see also (Grandell, 1997, Chapter 2, pp.14). One interesting property is given in Shaked's two crossing theorem which was proven in Shaked (1980). It states that there exist two integers $0 \leq k_1 < k_2$ such that the probability mass function of the mixed Poisson distribution is less than or equal to the probability mass function of a Poisson distribution with the same mean for values $k_1 + 1, \ldots, k_2$ and greater or equal otherwise. In our model, this means that, given Assumption (TN), the distribution of the time-changed process has an excess of zeros and a heavier right tail compared to the underlying Poisson process.

Example 4.8 (Families of mixed Poisson distributions)
Some subordinators lead to well-known families of mixed Poisson distributions, cf. (Klugman et al., 2004, Chapter 4.6.10, pp.103). In the following, the univariate claim number distributions resulting from a gamma, inverse Gaussian and Poisson subordinator with no drift are identified.

(i) Gamma subordinator – negative binomial distribution:
Choosing Λ as gamma subordinator with parameters $\beta, \eta > 0$ and with no drift, the Laplace transform φ_{Λ_t} at any point in time $t \geq 0$ is according to Example 2.14:

$$\varphi_{\Lambda_t}(x) = \Big(1 + \frac{x}{\eta}\Big)^{-\beta t}, \quad x \geq 0.$$

In this case the mixing distribution is a gamma distribution and the resulting mixed Poisson distribution is known to be a negative binomial distribution. This can easily be derived from Theorem 4.2. By induction, it follows for the derivatives of the Laplace transform of any order $k \in \mathbb{N}$:

$$\begin{aligned}\varphi^{(k)}_{\Lambda_t}(x) &= -\beta t(-\beta t - 1)\cdots(-\beta t - k + 1)\frac{1}{\eta^k}(1 + \frac{x}{\eta})^{-\beta t - k} \\ &= (-1)^k(\beta t + k - 1)\cdots\beta t\Big(\frac{1}{\eta + x}\Big)^k\Big(\frac{\eta}{\eta + x}\Big)^{\beta t}.\end{aligned}$$

Using Equation (4.1), the distribution of any marginal process L^i_t is found to be:

$$\mathbb{P}(L_t^i = k) = \frac{(-\lambda_i)^k}{k!}\varphi_{\Lambda_t}^{(k)}(\lambda_i) = \binom{k+\beta t-1}{k}\Big(\frac{\lambda_i}{\eta+\lambda_i}\Big)^k\Big(\frac{\eta}{\eta+\lambda_i}\Big)^{\beta t}.$$

This is the probability mass function of the negative binomial distribution:

$$L_t^i \sim \text{NegBin}\left(\beta t, \frac{\lambda_i}{\eta+\lambda_i}\right).$$

The negative binomial distribution is a popular alternative to the Poisson distribution under the presence of overdispersion and very well studied. Some details and many references can be found in (Johnson et al., 1992, Chapter 5, pp.199). An example for the shape of the probability mass function in comparison to the Poisson distribution and other mixed Poisson distributions can be seen in Figure 4.1. There, the parameters of the gamma subordinator are set such that Assumption (TN) holds.

(ii) Inverse Gaussian subordinator – Sichel distribution:
According to Example 2.14, the density f_{Λ_t} of an inverse Gaussian subordinator Λ with parameters $\beta, \eta > 0$ and with no drift is for any $t > 0$ given by

$$f_{\Lambda_t}(x) = \frac{\beta t}{\sqrt{2\pi x^3}}\exp\Big\{-\frac{1}{2x}(\eta x - \beta t)^2\Big\}\mathbb{1}_{\mathbb{R}_{>0}}(x).$$

The scaled process $\lambda_i\Lambda_t$ for any $i = 1, \dots, d$ then has density:

$$f_{\lambda_i\Lambda_t}(x) = f_{\Lambda_t}\Big(\frac{x}{\lambda_i}\Big)\frac{1}{\lambda_i}.$$

This follows easily from integration by substitution for $x > 0$:

$$\mathbb{P}(\lambda_i\Lambda_t \le x) = \mathbb{P}\Big(\Lambda_t \le \frac{x}{\lambda_i}\Big) = \int_0^{\frac{x}{\lambda_i}} f_{\Lambda_t}(s)\,\mathrm{d}s = \int_0^x f_{\Lambda_t}\Big(\frac{r}{\lambda_i}\Big)\frac{1}{\lambda_i}\,\mathrm{d}r.$$

Furthermore, it can be shown by a straightforward calculation that this is the density of an $IG(\tilde{\beta}t, \tilde{\eta})$ distribution with parameters:

$$\tilde{\beta} := \beta\sqrt{\lambda_i} \quad \text{and} \quad \tilde{\eta} := \frac{\eta}{\sqrt{\lambda_i}}.$$

In (Johnson et al., 1992, Chapter 15, pp.455), the distribution of the mixture of a Poisson with an inversion Gaussian random variable is given. There, the inverse Gaussian distribution is parametrized in $\alpha > 0$, $0 < \theta < 1$. By comparison of the density functions it can be found that it has to hold:

$$\tilde{\beta}t = \alpha\sqrt{\frac{\theta}{2}} \quad \text{and} \quad \tilde{\eta} = \sqrt{-2\Big(1 - \frac{1}{\theta}\Big)}.$$

Then it can be concluded for the distribution:

$$\mathbb{P}(L_t^i = k) = \sqrt{\frac{2\alpha}{\pi}}\frac{K_{k-\frac{1}{2}}(\alpha)}{k!}\exp\{\alpha\sqrt{1-\theta}\}\Big(\frac{\alpha\theta}{2}\Big)^k, \quad k \in \mathbb{N}_0,$$

where $K_\cdot(\cdot)$ is the modified Bessel function of the second kind, see (Johnson et al., 1992, Section A12, p.30), and:

$$\alpha = \beta t\sqrt{2\lambda_i\eta^2}, \quad \text{and} \quad \theta = \frac{2\lambda_i}{2\lambda_i + \eta^2}.$$

This distribution is a special two parameter subclass of the three parameter Sichel distribution. Some alternative parametrizations and a recursive formula for the probability mass function are referenced in (Johnson et al., 1992, Chapter 15, pp.455). An example for the shape of the probability mass function can again be found in Figure 4.1. The parameters of the inverse Gaussian subordinator have been set such that Assumption (TN) holds and the variance matches the variance in the gamma subordinator case.

(iii) Poisson process – Neyman type A distribution:
If Λ is a Poisson process with no drift and intensity $\xi > 0$, the mixture distribution is a Neyman type A distribution, see (Johnson et al., 1992, Chapter 6, pp.368) for details and important properties. The probability mass function is given by:

$$\mathbb{P}(L_t^i = k) = \frac{\exp\{-\lambda_i + \lambda_i\exp\{-\xi\}\}\xi^k}{k!}\sum_{j=1}^{k} S(k,j)\lambda_i^j\exp\{-j\xi\},$$

where $S(k,j)$ are the Stirling numbers of the second kind. We will discuss these numbers further after Corollary 4.15. Figure 4.1 again shows exemplarily the shape of the probability mass function. As before, the parameter is set such that Assumption (TN) holds and the variance matches the variance in the gamma and inverse Gaussian case.

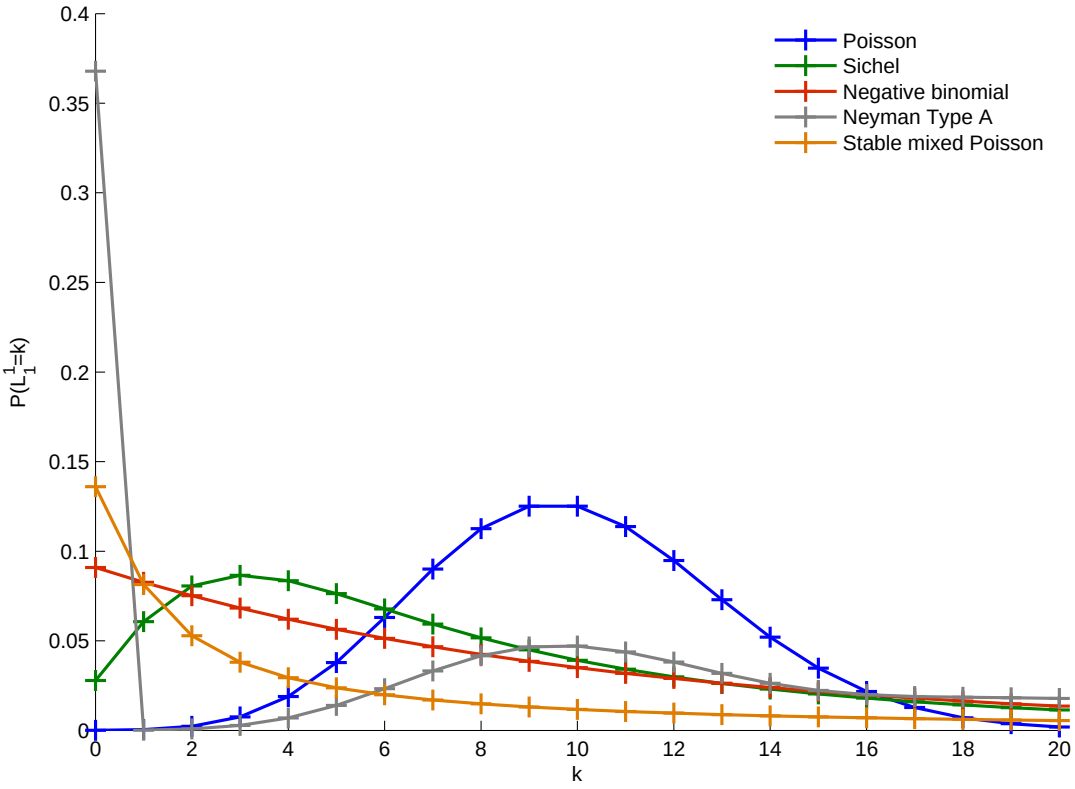


Figure 4.1 Univariate claim number distributions for different subordinators: For dimension $d = 1$, intensity $\lambda_1 = 10$, and time horizon $T = 1$, the figure shows the probability mass function of the underlying Poisson process as well as the time-changed process for different directing Lévy subordinators: inverse Gaussian ($b = 0$, $\beta = \eta = 1$), gamma ($b = 0$, $\beta = \eta = 1$), Poisson ($b = 0$, $\xi = 1$), stable ($b = 0$, $\alpha = 0.3$). The parameters of the inverse Gaussian, gamma, and Poisson subordinator have been set such that mean and variance are both one. For the stable subordinator, mean and variance are infinite. The point masses of the distributions are marked with a cross; the connecting lines are provided only to aid visualization.

In Hofert et al. (2012) maximum likelihood estimation for Archimedean copulas is studied. These copulas are defined by an Archimedean generator $f\colon [0,\infty] \to [0,1]$ as

$$C_f(\boldsymbol{u}) = f(f^{-1}(u_1) + \cdots + f^{-1}(u_k)), \quad \boldsymbol{u} \in [0,1]^k.$$

To avoid confusion we use f for the generator where in the paper they use Ψ. Also, we denote the dimension of the copula by k rather than d as it corresponds in our setting to the tail of the distribution, not the dimension of the model. It was shown in Kimberling (1974) that f defines a proper copula in all dimensions $k \in \mathbb{N}_0$ iff it is completely monotone. Furthermore, we discussed in Section 2.1 that according to Bernstein's Theorem, a function is completely monotone iff it is the Laplace transform of a non-negative random variable. From the families of Archimedean copulas studied in this paper, two are connected to Lévy subordinators, as their generator stems from an infinitely divisible distribution: the Clayton copula corresponding to a gamma distribution and the Gumbel copula corresponding to a stable distribution. For likelihood inference on the copulas, the k-dimensional derivatives of the generators are studied in the paper particularly with the high-dimensional case in mind, that is for large values of k, which in our setting corresponds to the far tail of the distribution. They argue that using computer algebra systems to evaluate these derivatives causes numerical problems quickly and results, if they can be obtained at all, take a long time and are unreliable. They derive algebraic expressions for the derivatives in a slightly reduced setting to ours for the gamma and the stable case which are implemented in the R package nacopula in a numerically robust way. Our more general setting can easily be derived from their results in the following way.

For the Clayton copula, they consider the generator $f_\theta(x) = (1+x)^{-\frac{1}{\theta}}$, which is the Laplace transform of a Gamma$(\frac{1}{\theta}, 1)$ distribution, $\theta > 0$. In our setting it is $\Lambda_t \sim$ Gamma$(\beta t, \eta)$, that is we have the parametrization $\theta := \frac{1}{\beta t}$ and consider a general second parameter $\eta > 0$. It follows for the Laplace transform:

$$\varphi_{\Lambda_t}(x) = \left(1 + \frac{x}{\eta}\right)^{-t\beta} = f_{\frac{1}{\beta t}}\left(\frac{x}{\eta}\right).$$

By the chain rule for differentiation, the derivatives of φ_{Λ_t} can be derived from the ones of f_θ implemented in the nacopula package as

$$\varphi^{(k)}_{\Lambda_t}(x) = f^{(k)}_{\frac{1}{\beta t}}\left(\frac{x}{\eta}\right)\eta^{-k}.$$

The results for the stable subordinator, which corresponds to the Gumbel copula, may be less relevant to practical applications due to the non-existing moments of the stable distribution. This not only prevents the time-normalization condition to hold, but particularly leads to an infinite number of expected claims on any time-interval, an assumption that most likely is not appropriate to any real-world insurance business. Furthermore, the infinite expectation makes simulating the process difficult. However, from the results in Hofert et al. (2012) we can deduce the probability mass function in this case. They consider the generator

$$f_\alpha(x) = \exp\{-x^\alpha\},$$

and find for its derivatives the formula

$$f_\alpha^{(k)}(x) = f_\alpha(x) x^{-k} \sum_{j=1}^{k} x^{\alpha j} \frac{k!}{j!} \sum_{l=1}^{j} (-1)^l \binom{j}{l} \binom{\alpha l}{k}.$$

The generator f_α corresponds to the Laplace transform of Λ in $t = 1$. For general $t \geq 0$, it is:

$$\varphi_{\Lambda_t}(x) = \exp\{-tx^\alpha\} = f_\alpha^{(k)}(t^{\frac{1}{\alpha}} x),$$

and it can again be concluded by the chain rule that

$$\varphi_{\Lambda_t}^{(k)}(x) = f_\alpha^{(k)}(t^{\frac{1}{\alpha}} x) t^{\frac{k}{\alpha}}.$$

Together with Theorem 4.2 the distribution of the time-changed process can be concluded. Figure 4.1 again shows the shape of the probability mass function for $\alpha = 0.3$. Note that mean and variance are infinite and particularly do not match the ones of the other subordinator choices.

In case of the gamma subordinator with no drift, derivatives of the Laplace transform of the subordinator were found easily. For the inverse Gaussian, Poisson, and stable subordinator with no drift, the derivatives have been found as well, but calculations are not as straightforward. For other subordinators or if a drift is added, the calculation may become even more complicated. In many cases, however, the derivatives of the Laplace exponent are easily available. For the important subordinator families introduced in Example 2.14, these derivatives are summarized in Table 4.6. Then, the derivatives

of the Laplace transform and the probability mass function can be computed via the complete Bell polynomials as in Equation (4.2). Table 4.1 shows the Bell polynomials and the corresponding marginal probabilities up to $k = 3$. Table 4.2 presents as example some marginal probabilities jointly for two consecutive points in time along Corollary 4.4. In general, the complete Bell polynomials $B_k := B_k(x_1, \ldots, x_k)$ can be computed as sum over the incomplete Bell polynomials $B_{k,l} := B_{k,l}(x_1, \ldots, x_{k-l+1})$, $l \leq k$, which again can be computed by a recursion relation, see Wheeler (1987):

$$B_{k,l} = \sum_{m=1}^{k-l+1} \binom{k-1}{m-1} x_m B_{k-m,l-1},$$

where $B_{0,0} = 1$, $B_{k,0} = 0$ for $k \geq 1$, $B_{0,l} = 0$ for $l \geq 1$, and

$$B_k = \sum_{l=1}^{k} B_{k,l}, \quad B_0 := 1.$$

The coefficients $B_{k,l}$ can be stored in a lower triangular matrix and then the B_k are evaluated as the column totals. However, the calculations again require binomial coefficients which grow fast in k. For instance, in MATLAB® we start to run into trouble when considering $k \geq 58$, which is not at all unrealistic in a model with high expected claim numbers $\boldsymbol{\lambda}$. Evaluation time for $k = 57$ is around 1.5 seconds. More details on Bell polynomials can be found in (Comtet, 1974, Chapter 3.3, pp.133 and pp.307). There a table of the $B_{k,l}$ for $1 \leq l \leq k \leq 12$ is given. It can be seen that the summation terms grow pretty quickly, for instance for B_{12} the coefficient for $x_1^2 x_2^2 x_3^2$ is $415,800$. In Section 4.2 we will discuss a recursive evaluation scheme to avoid these numerical difficulties and to speed up the evaluation.

Cumulative process distribution and claim arrival times

Calculating the cumulative distribution function or the tail of the distribution can be done simply by summation of the probability mass function. For high claim numbers, this may become cumbersome due to the increasing number of summation terms. For this case it is helpful to observe that the cumulative distribution function is closely linked to a Taylor approximation of the Laplace transform.

Theorem 4.9 (Tails of the claim number process distribution)
For the tail of the distribution of the time-changed process $\boldsymbol{L}_t$ and its aggregate process $\bar{L}_t$ it holds in $t \geq 0$:

	Bell polynomial and process distribution
$k=0$	$B_0 \equiv 1$ $\mathbb{P}(L_t^i = 0) = \exp\{-t\Psi_\Lambda(\lambda_i)\}$
$k=1$	$B_1(x_1) = x_1$ $\mathbb{P}(L_t^i = 1) = \lambda_i t \Psi_\Lambda^{(1)}(\lambda_i) \exp\{-t\Psi_\Lambda(\lambda_i)\}$
$k=2$	$B_2(x_1, x_2) = x_1^2 + x_2$ $\mathbb{P}(L_t^i = 2) = \frac{\lambda_i^2}{2}(t^2\Psi_\Lambda^{(1)}(\lambda_i)^2 - t\Psi_\Lambda^{(2)}(\lambda_i)) \exp\{-t\Psi_\Lambda(\lambda_i)\}$
$k=3$	$B_3(x_1, x_2, x_3) = x_1^3 + 3x_1x_2 + x_3$ $\mathbb{P}(L_t^i = 3) = \frac{\lambda_i^3}{6}(t^3\Psi_\Lambda^{(1)}(\lambda_i)^3 - 3t^2\Psi_\Lambda^{(1)}(\lambda_i)\Psi_\Lambda^{(2)}(\lambda_i) + t\Psi_\Lambda^{(3)}(\lambda_i)) \exp\{-t\Psi_\Lambda(\lambda_i)\}$

Table 4.1 Bell polynomials and univariate claim number distribution: The table gives the Bell polynomials and the probabilities of the univariate marginal claim number process L^i in time $t \geq 0$ up to order $k = 3$.

	$\mathbb{P}(L_{t_1}^i = k_1, L_{t_2}^i = k_2)$
$k_1 = 0, k_2 = 1$	$\lambda_i \Delta t_2 \Psi_\Lambda^{(1)}(\lambda_i) \exp\{-t_2\Psi_\Lambda(\lambda_i)\}$
$k_1 = 0, k_2 = 2$	$\frac{\lambda_i^2}{2}((\Delta t_2)^2\Psi_\Lambda^{(1)}(\lambda_i)^2 - \Delta t_2\Psi_\Lambda^{(2)}(\lambda_i)) \exp\{-t_2\Psi_\Lambda(\lambda_i)\}$
$k_1 = 1, k_2 = 2$	$\lambda_i^2 t_1 \Delta t_2 \Psi_\Lambda^{(1)}(\lambda_i)^2 \exp\{-t_2\Psi_\Lambda(\lambda_i)\}$

Table 4.2 Univariate claim number distribution at two points in time: The table gives selected probabilities of the marginal claim number process L^i jointly at two consecutive points in time $0 \leq t_1 < t_2$.

$$\sum_{k=n+1}^{\infty} \sum_{|\boldsymbol{k}|=k} \mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k}) = \mathbb{P}(\bar{L}_t > n) = R^n \varphi_{\Lambda_t}(0; |\boldsymbol{\lambda}|), \quad n \in \mathbb{N}_0, \tag{4.6}$$

where the last term is the residual of a Taylor expansion of φ_{Λ_t} *in* $|\boldsymbol{\lambda}|$ *evaluated in zero:*

$$R^n \varphi_{\Lambda_t}(0; |\boldsymbol{\lambda}|) = \frac{(-1)^{n+1}}{n!} \int_0^{|\boldsymbol{\lambda}|} x^n \varphi_{\Lambda_t}^{(n+1)}(x) \, \mathrm{d}x = o(|\boldsymbol{\lambda}|^n).$$

The limiting behaviour $o(|\boldsymbol{\lambda}|^n)$ *holds in* $|\boldsymbol{\lambda}| \to 0$*, that is*

$$\lim_{|\boldsymbol{\lambda}| \to 0} \frac{R^n \varphi_{\Lambda_t}(0; |\boldsymbol{\lambda}|)}{|\boldsymbol{\lambda}|^n} = 0.$$

Proof. In general, due to the smoothness of φ_{Λ_t}, the Taylor expansion at point $a \geq 0$ evaluated in $u \geq 0$ is given as

$$\varphi_{\Lambda_t}(u) = \sum_{k=0}^{\infty} \frac{(u-a)^k}{k!} \varphi_{\Lambda_t}^{(k)}(a) = \sum_{k=0}^{n} \frac{(u-a)^k}{k!} \varphi_{\Lambda_t}^{(k)}(a) + R^n \varphi_{\Lambda_t}(u; a),$$

with residual term

$$R^n \varphi_{\Lambda_t}(u; a) := \sum_{k=n+1}^{\infty} \frac{(u-a)^k}{k!} \varphi_{\Lambda_t}^{(k)}(a) = \int_a^u \frac{(u-x)^n}{n!} \varphi_{\Lambda_t}^{(n+1)}(x) \, \mathrm{d}u = o(|u-a|^n),$$

for $u \to a$, see for instance (Apostol, 1962, Chapter 7.5). Setting $a := |\boldsymbol{\lambda}|$ and $u := 0$ gives

$$1 = \varphi_{\Lambda_t}(0) = \sum_{k=0}^{\infty} \frac{(-|\boldsymbol{\lambda}|)^k}{k!} \varphi_{\Lambda_t}^{(k)}(|\boldsymbol{\lambda}|) = \sum_{k=0}^{\infty} \mathbb{P}(\bar{L}_t = k) = \mathbb{P}(\bar{L}_t \leq k) + R^n \varphi_{\Lambda_t}(0; |\boldsymbol{\lambda}|).$$

This proves the second equality in Equation (4.6). By the multinomial theorem, it is

$$\frac{(-|\boldsymbol{\lambda}|)^k}{k!} = \sum_{|\boldsymbol{k}|=k} \frac{(-|\boldsymbol{\lambda}|)^{\boldsymbol{k}}}{\boldsymbol{k}!},$$

so it follows:

$$\begin{aligned}\mathbb{P}(\bar{L}_t \leq k) &= \sum_{k=0}^{k} \frac{(-|\boldsymbol{\lambda}|)^k}{k!} \varphi_{\Lambda_t}^{(k)}(|\boldsymbol{\lambda}|) = \sum_{k=0}^{k} \varphi_{\Lambda_t}^{(k)}(|\boldsymbol{\lambda}|) \sum_{|\boldsymbol{k}|=k} \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \\ &= \sum_{k=0}^{k} \sum_{|\boldsymbol{k}|=k} \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \varphi_{\Lambda_t}^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|).\end{aligned}$$

This proves the first equality in Equation (4.6) and concludes the proof. □

Note that with this theorem we have an explicit representation of the cumulative distribution function of the aggregate claim number process $\bar{L}$, and consequently for the marginals L^i of the multivariate process $\boldsymbol{L}$ as well, but not for the multivariate distribution of $\boldsymbol{L}$. The summation in Equation (4.6) does not cover all $\boldsymbol{k} > \boldsymbol{n}$ for some $\boldsymbol{n} \in \mathbb{N}_0^d$ but rather all $|\boldsymbol{k}| > n$ for $n \in \mathbb{N}_0$. The theorem provides, however, still a means of calculating aggregate and tail quantities more conveniently. For the residual term of a Taylor series expansion, other (approximative) solutions are available, see for instance Apostol (1962). Evaluating the residual in the given integral form analytically may only on rare occasions be possible. It can, however, easily be approximated numerically given the availability of an efficient implementation for the derivatives of the Laplace transform $\varphi_{\Lambda_t}(x)$ or the probabilities in the univariate case for varying intensity x, that is $\mathbb{P}(L_t^x = k)$ for L_t^x from Model (M) with $d = 1$ and $\lambda_1 = x$. For instance, a simple Riemann sum with $m \in \mathbb{N}$ steps and step size $h := \frac{|\boldsymbol{\lambda}|}{m}$ can be calculated:

$$\begin{aligned}\int_0^{|\boldsymbol{\lambda}|} \frac{(-1)^{n+1}}{n!} x^n \varphi_{\Lambda_t}^{(n+1)}(x)\,\mathrm{d}x &\approx h \sum_{j=1}^{m} \frac{(-1)^{n+1}(jh)^n}{n!} \varphi_{\Lambda_t}^{(n+1)}(jh) \\ &= h \sum_{j=1}^{m} \frac{n+1}{jh} \mathbb{P}(L_t^{jh} = n+1).\end{aligned}$$

By using quadrature rules or other methods for numerical integration, some care should by taken when including zero as grid point. While the whole integrand is well-behaved at zero, $\varphi_{\Lambda_t}^{(n+1)}$ itself may not if the $(n+1)$-th moment of the subordinator does not exist.

The results from Theorem 4.9 are also helpful for considering the distribution of claim arrival times, which directly correspond to the cumulative distribution function of the process.

Corollary 4.10 (Survival probabilities of the claim arrival times)
For any claim number vector $\boldsymbol{j} \in \mathbb{N}^d$ it holds for the distribution of the vector $\boldsymbol{\tau_j} := (\tau_{1j_1}, \dots, \tau_{dj_d})'$ of the $\boldsymbol{j}$-th arrival times:

$$\mathbb{P}(\boldsymbol{\tau_j} > t) = \sum_{0 \leq \boldsymbol{k} < \boldsymbol{j}} \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \varphi_{\boldsymbol{\Lambda_t}}^{(|\boldsymbol{k}|)}(\boldsymbol{\lambda}), \quad t \geq 0.$$

For the marginals $i = 1, \dots, d$ and some $j \in \mathbb{N}$, this formula simplifies to:

$$\mathbb{P}(\tau_{ij} > t) = \sum_{k=0}^{j-1} \frac{(-\lambda_i)^k}{k!} \varphi_{\Lambda_t}^{(k)}(\lambda_i) = 1 - R^{j-1} \varphi_{\Lambda_t}(0; \lambda_i), \quad t \geq 0. \tag{4.7}$$

Proof. The proof follows from the two Theorems 4.2 and 4.9 together with the observation that the claim arrival times and the claim number process are related via $\{\boldsymbol{\tau_j} > t\} = \{\boldsymbol{L}_t < \boldsymbol{j}\}$. □

From Equation (4.2) we know that the derivatives $\varphi_{\Lambda_t}^{(k)}(x)$, if understood as functions in t, are differentiable in t as they are the product of an exponential and a polynomial. Thus, the arrival times have a continuous distribution and a density exists even though no conditional density exists as was discussed in Section 3.1. The inter-arrival times, however, have a singular component due to the possibility of multiple claims to arrive simultaneously. This will be discussed in more detail in Section 4.2.

Some evaluations of the marginal distributions can be found in Table 4.3. Of course, the results from the above theorem can be extended to the case of finite-dimensional distributions $\mathbb{P}(\tau_{\boldsymbol{j}_1} > t_1, \dots, \tau_{\boldsymbol{j}_n} > t_n)$ for non-decreasing sequences $\{t_k\}_{k=1,\dots,n}$ and $\{\boldsymbol{j}_k\}_{k=1,\dots,n}$, $n \in \mathbb{N}$, by Corollary 4.4. For instance, it can be calculated for the i-th marginal:

$$\begin{aligned}
\mathbb{P}(\tau_{i1} > t_1, \tau_{i2} > t_2) &= \exp\{-t_2 \Psi_\Lambda(\lambda_i)\}(1 + \lambda_i \Delta t_2 \Psi_\Lambda^{(1)}(\lambda_i)), \\
\mathbb{P}(\tau_{i1} > t_1, \tau_{i2} > t_2, \tau_{i3} > t_3) &= \exp\{-t_3 \Psi_\Lambda(\lambda_i)\}[1 + \lambda_i (t_3 - t_1) \Psi_\Lambda^{(1)}(\lambda_i) \\
&\quad + \lambda_i^2 (\frac{(\Delta t_3)^2}{2} + \Delta t_3 \Delta t_2) \Psi_\Lambda^{(1)}(\lambda_i)^2 - \frac{\lambda_i^2}{2} \Delta t_3 \Psi_\Lambda^{(2)}(\lambda_i)].
\end{aligned}$$

	$\mathbb{P}(\tau_{ij} > t)$
$j = 1$	$\exp\{-t\Psi_\Lambda(\lambda_i)\}$
$j = 2$	$(1 + \lambda_i t\Psi_\Lambda^{(1)}(\lambda_i))\exp\{-t\Psi_\Lambda(\lambda_i)\}$
$j = 3$	$(1 + \lambda_i t\Psi_\Lambda^{(1)}(\lambda_i) + \frac{\lambda_i^2}{2}t^2\Psi_\Lambda^{(1)}(\lambda_i)^2 - \frac{\lambda_i^2}{2}t\Psi_\Lambda^{(2)}(\lambda_i))\exp\{-t\Psi_\Lambda(\lambda_i)\}$

Table 4.3 Distribution of claim arrival times: The table shows the survival probabilities in $t \geq 0$ of the individual claim arrival times τ_{ij} of $\boldsymbol{L}$ up to $j = 3$.

Multi-parameter process distribution

It is possible to extend the result of Theorem 4.2 about the distribution of $\boldsymbol{L}_t$ to the multi-parameter process where all marginals of the process are considered at different points in time rather than all at the same time. For this purpose, we introduced for a multi-parameter time $\boldsymbol{t} \in \mathbb{R}^d_{\geq 0}$ the notation:

$$\boldsymbol{L_t} := (L^1_{t_1}, \ldots, L^d_{t_d})' \quad \text{and} \quad \boldsymbol{\Lambda_t} := (\Lambda_{t_1}, \ldots, \Lambda_{t_d})'.$$

For ease of notation, multi-parameter times $\boldsymbol{t}$ with non-decreasing components are considered in the following, that is $0 \leq t_1 \leq \cdots \leq t_d$. The generalization to arbitrary multi-parameter times can always be found by suitable permutation. For instance, the following lemma gives for $\boldsymbol{t}$ with non-decreasing entries the Laplace transform of the vector process $\boldsymbol{\Lambda_t}$, which will be needed in Corollary 4.12. If $\boldsymbol{t}$ is any multi-parameter time and $\pi\colon \{1, \ldots, d\} \to \{1, \ldots, d\}$ a permutation such that $\pi(\boldsymbol{t})$ has non-decreasing entries, then the lemma can be applied to

$$\varphi_{\boldsymbol{\Lambda_t}}(\boldsymbol{x}) = \varphi_{\boldsymbol{\Lambda}_{\pi(\boldsymbol{t})}}(\pi(\boldsymbol{x})).$$

Lemma 4.11 (Laplace transform of the multi-parameter subordinator)
Let $\boldsymbol{t} \in \mathbb{R}^d_{\geq 0}$ be a multi-parameter time with $0 \leq t_1 \leq \cdots \leq t_d$ and set $t_0 := 0$. Then the Laplace transform $\varphi_{\boldsymbol{\Lambda_t}}(\boldsymbol{x})$ of the d-dimensional process $\boldsymbol{\Lambda_t}$ is given by:

$$\varphi_{\boldsymbol{\Lambda_t}}(\boldsymbol{x}) = \exp\Big\{-\sum_{i=1}^{d} \Delta t_i \Psi_\Lambda(x_i + \cdots + x_d)\Big\}, \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

Proof. The Laplace transform of $\boldsymbol{\Lambda_t}$ can be derived in terms of the Laplace exponent Ψ_Λ by exploiting the Lévy property of independent and stationary increments. Using $\Lambda_{t_i} = \Lambda_{t_1} + (\Lambda_{t_2} - \Lambda_{t_1}) + \cdots + (\Lambda_{t_i} - \Lambda_{t_{i-1}})$ it follows:

$$\begin{aligned}\varphi_{\boldsymbol{\Lambda_t}}(\boldsymbol{x}) &= \mathbb{E}\Big[\exp\Big\{-\sum_{i=1}^{d} x_i \Lambda_{t_i}\Big\}\Big] = \mathbb{E}\Big[\exp\Big\{-\sum_{i=1}^{d}(x_i+\cdots+x_d)(\Lambda_{t_i}-\Lambda_{t_{i-1}})\Big\}\Big] \\ &= \prod_{i=1}^{d} \mathbb{E}[\exp\{-(x_i+\cdots+x_d)\Lambda_{\Delta t_i}\}] = \prod_{i=1}^{d} \varphi_{\Lambda_{\Delta t_i}}(x_i+\cdots+x_d) \\ &= \exp\Big\{-\sum_{i=1}^{d} \Delta t_i \Psi_\Lambda(x_i+\cdots+x_d)\Big\}.\end{aligned}$$

□

Corollary 4.12 (Distribution of the multi-parameter claim number process)
Let $\boldsymbol{t} \in \mathbb{R}^d_{\geq 0}$ be a multi-parameter time with $0 \leq t_1 \leq \cdots \leq t_d$ and set $t_0 := 0$. The distribution of the claim number process $\boldsymbol{L_t}$ in Model (M) in time $\boldsymbol{t}$ is given by:

$$\mathbb{P}(\boldsymbol{L_t} = \boldsymbol{k}) = \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \varphi^{(\boldsymbol{k})}_{\boldsymbol{\Lambda_t}}(\boldsymbol{\lambda}), \quad \boldsymbol{k} \in \mathbb{N}^d_0. \tag{4.8}$$

Proof. The proof follows in a similar fashion to the proof of Theorem 4.2. By the tower rule for conditional expectations it holds:

$$\mathbb{P}(\boldsymbol{L_t} = \boldsymbol{k}) = \mathbb{E}\Big[\prod_{i=1}^{d} \mathbb{P}(N^i_{\Lambda_{t_i}} = k_i | \sigma(\Lambda))\Big] = \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!}\, \mathbb{E}[\boldsymbol{\Lambda}^{\boldsymbol{k}}_{\boldsymbol{t}} \exp\{-\boldsymbol{\lambda}' \boldsymbol{\Lambda_t}\}].$$

The last term can be expressed in terms of partial derivatives of the Laplace transform $\varphi_{\boldsymbol{\Lambda_t}}(\boldsymbol{x}) = \mathbb{E}[\exp\{-\boldsymbol{x}'\boldsymbol{\Lambda_t}\}]$ of $\boldsymbol{\Lambda_t}$ as:

$$\varphi^{(\boldsymbol{k})}_{\boldsymbol{\Lambda_t}}(\boldsymbol{x}) = (-1)^{|\boldsymbol{k}|}\, \mathbb{E}[\boldsymbol{\Lambda}^{\boldsymbol{k}}_{\boldsymbol{t}} \exp\{-\boldsymbol{x}'\boldsymbol{\Lambda_t}\}], \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

□

By Lemma 4.11, the Laplace transform $\varphi_{\boldsymbol{\Lambda_t}}(\boldsymbol{x})$ is the product of univariate Laplace transforms which are completely monotone. As complete monotonicity is closed under multiplication, see (Schilling et al., 2012, Chapter 1, Corollary 1.6, p.5), the probability mass function of the multi-parameter process is indeed non-negative. Furthermore, the multivariate derivative $\varphi^{(\boldsymbol{k})}_{\boldsymbol{\Lambda}_t}$ can also be expressed in terms of derivatives of the Laplace exponent Ψ_Λ by means of a multivariate version of Faà di Bruno's formula, see Constantine and Savits (1996) or Hardy (2006). Note also that using Corollary 4.12, the results about the distribution of the claim arrival times from Corollary 4.10 can be extended to the more general case $\mathbb{P}(\boldsymbol{\tau_j} > \boldsymbol{t})$ for some multi-parameter time $\boldsymbol{t} \in \mathbb{R}^d_{\geq 0}$.

Moments of the process distribution

Before we have a closer look at the (mixed) moments of the claim number process, a result about the the heavy-tailedness of the distribution characterized by the moment generating function is given.

Proposition 4.13 (Heavy-tailedness of the claim number process)
The marginal distributions of $\boldsymbol{L}$ are heavy-tailed iff the distribution of the subordinator is heavy-tailed.

Proof. As $\boldsymbol{L}_t$ is a Lévy process, it is sufficient to check for heavy tails at time $t = 1$. Using Proposition 4.1 for the Laplace exponent of $\boldsymbol{L}_1$, it follows:

$$\mathbb{E}[\exp\{xL_1^i\}] = \exp\{-\Psi_{L^i}(-x)\} = \exp\{-\Psi_\Lambda(\lambda_i(1-\exp\{x\}))\}, \quad x > 0.$$

The Poisson exponent $\lambda_i(1-\exp\{x\})$ converges to $-\infty$ for $x \to \infty$ and is zero for $x = 0$. Due to its continuity, it takes on all vales in between and the following equivalence holds:

$$\mathbb{E}[\exp\{xL_1^i\}] = \infty \quad \forall x > 0 \quad \Leftrightarrow \quad \exp\{-\Psi_\Lambda(-x)\} = \infty \quad \forall x > 0.$$

This is the criterion for heavy-tailedness of the subordinator distribution, so the claim is established. □

From the subordinator families introduced in Example 2.14, only the stable and the compound Poisson process with heavy-tailed jump size distribution are heavy-tailed. Following (Cont and Tankov, 2003, Chapter 3.6, Proposition 3.14, p.92), a subordinator has for some $x > 0$ finite exponential moment $\mathbb{E}[\exp\{x\Lambda_t\}]$ iff

$$\int_1^\infty \exp\{x\Lambda_t\}\,\nu_\Lambda(\mathrm{d}t) < \infty;$$

then it holds $\mathbb{E}[\exp\{x\Lambda_t\}] = \varphi_{\Lambda_t}(-x)$.

The moments of the distribution of $\boldsymbol{L}_t$ can be derived from evaluating derivatives of its Laplace transform in zero. Alternatively, they can be concluded from the moments of the selected subordinator. The following remark sums up the properties of the subordinator moments before Theorem 4.15 discusses the implications on the moments of the time-changed process.

Remark 4.14 (Moments of the Lévy subordinator)
The existence of the moments of a Lévy process is not a time-dependent property, that is if a moment exists at some time in time it exists for all $t \geq 0$. In particular, it holds for the subordinator Λ that the n-th moment $\mathbb{E}[\Lambda_t^n]$, $n \in \mathbb{N}$, exists if and only if $\int_0^\infty x^n\,\nu_\Lambda(\mathrm{d}x) < \infty$, as this term determines the cumulants of the distribution. The cumulant generating function is $g_{\Lambda_t}(x) = -t\Psi_\Lambda(-x)$, hence, it follows for the cumulants, if they exist:

$$\kappa_n = g_{\Lambda_t}^{(n)}(0) = (-1)^{n+1} t \Psi_\Lambda^{(n)}(0).$$

Obviously, as the cumulant generating function is linear in time, the cumulants are as well. From the Lévy–Khintchine representation it follows for the derivatives of the exponent considering that differentiation under the integral is permissible:

$$\Psi_\Lambda^{(n)}(x) = b_\Lambda \mathbb{1}_{\{1\}}(n) + \int_0^\infty (-1)^{n+1} t^n \exp\{-xt\}\,\nu_\Lambda(\,\mathrm{d}t), \quad x \geq 0.$$

Evaluating in zero – if finite – gives the following result for the cumulants in terms of the characteristic tuple:

$$\kappa_n = t\Big(b_\Lambda \mathbb{1}_{\{1\}}(n) + \int_0^\infty t^n\,\nu_\Lambda(\mathrm{d}t)\Big),$$

see also (Cont and Tankov, 2003, Chapter 3.6, Proposition 3.13, pp.91). The first cumulant equals the mean, second and third equal the second and third central moment, respectively, see (Johnson et al., 1992, Chapter B6, p.45):

$$\begin{aligned}
\mathbb{E}[\Lambda_t] = \kappa_1 &= \quad t\Psi_\Lambda^{(1)}(0) = t\,\mathbb{E}[\Lambda_1], \\
\mathbb{V}\mathrm{ar}[\Lambda_t] = \kappa_2 &= -t\Psi_\Lambda^{(2)}(0) = t\,\mathbb{V}\mathrm{ar}[\Lambda_1], \\
\mathbb{E}[(\Lambda_t - \mathbb{E}[\Lambda_t])^3] = \kappa_3 &= \quad t\Psi_\Lambda^{(3)}(0) = t\,\mathbb{E}[(\Lambda_1 - \mathbb{E}[\Lambda_1])^3].
\end{aligned}$$

Particularly, these quantities are all linear in time. Note that this relationship between cumulants and (central) moments in general no longer holds for higher orders. For the Lévy subordinators primarily studied in this thesis, these moments are given explicitly in terms of the subordinator parameters in Table 4.4. As is well known or follows from a simple calculation, the second and third general moments can be calculated from the cumulants as:

$$\begin{aligned}
\mathbb{E}[\Lambda_t^2] &= \kappa_2 + \kappa_1^2 = t\,\mathbb{V}\mathrm{ar}[\Lambda_t] + t^2\,\mathbb{E}[\Lambda_1]^2, \\
\mathbb{E}[\Lambda_t^3] &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3.
\end{aligned}$$

In general, moments are connected to cumulants via Bell polynomials, see (Johnson et al., 1992, Chapter B5 and B6, pp.40):

$$\mathbb{E}[\Lambda_t^n] = B_n(\kappa_1, \ldots, \kappa_n).$$

This is also obvious from the argumentation in the proof of Theorem 4.2 given that the exponential of the cumulant generating function is the moment generating function.

For two points in time $0 \leq t_1 \leq t_2$, the mixed moments can by derived by a simple calculation exploiting the independent and stationary increment property:

$$\begin{aligned}
\mathbb{E}[\Lambda_{t_1}\Lambda_{t_2}] &= \mathbb{E}[\Lambda_{t_1}(\Lambda_{t_2} - \Lambda_{t_1} + \Lambda_{t_1})] = \mathbb{E}[\Lambda_{t_1}(\Lambda_{t_2} - \Lambda_{t_1})] + \mathbb{E}[\Lambda_{t_1}^2] \\
&= \mathbb{E}[\Lambda_{t_1}]\,\mathbb{E}[\Lambda_{t_2-t_1}] + \mathbb{V}\mathrm{ar}[\Lambda_{t_1}] + \mathbb{E}[\Lambda_{t_1}]^2 \\
&= t_1(t_2 - t_1)\,\mathbb{E}[\Lambda_1]^2 + t_1\,\mathbb{V}\mathrm{ar}[\Lambda_1] + t_1^2\,\mathbb{E}[\Lambda_1]^2 \\
&= t_1 t_2\,\mathbb{E}[\Lambda_{t_1}]^2 + t_1\,\mathbb{V}\mathrm{ar}[\Lambda_1].
\end{aligned}$$

Then it follows for the covariance of the process between those two points in time:

$$\mathbb{C}\mathrm{ov}[\Lambda_{t_1}, \Lambda_{t_2}] = \mathbb{E}[\Lambda_{t_1}\Lambda_{t_2}] - \mathbb{E}[\Lambda_{t_1}]\,\mathbb{E}[\Lambda_{t_2}] = \mathbb{V}\mathrm{ar}[\Lambda_{t_1}] = t_1\,\mathbb{V}\mathrm{ar}[\Lambda_1].$$

This covariance reduces to the variance of the overlapping time period which is in line with the property of Lévy processes to have no dependence over time. ▲

Theorem 4.15 (Moments and covariance of the claim number processes)
For $n \in \mathbb{N}$, the n-th moment of any component L^i of $\boldsymbol{L}$ exists iff the n-th moment of the subordinator exists. In this case it holds:

$$\mathbb{E}[\boldsymbol{L}_t^n] = \sum_{k=0}^{n} S(n,k)\,\mathbb{E}[\Lambda_t^k]\boldsymbol{\lambda}^k, \quad t \geq 0,$$

where $S(n,k)$ denotes the Stirling numbers of the second kind. In particular, existence assumed, it is

$$\begin{aligned}
\mathbb{E}[\boldsymbol{L}_t] &= t\,\mathbb{E}[\Lambda_1]\boldsymbol{\lambda}, \\
\mathbb{V}\mathrm{ar}[\boldsymbol{L}_t] &= t(\mathbb{E}[\Lambda_1]\boldsymbol{\lambda} + \mathbb{V}\mathrm{ar}[\Lambda_1]\boldsymbol{\lambda}^2), \\
\mathbb{E}[(\boldsymbol{L}_t - \mathbb{E}[\boldsymbol{L}_t])^3] &= t(\mathbb{E}[\Lambda_1]\boldsymbol{\lambda} + 3\,\mathbb{V}\mathrm{ar}[\Lambda_1]\boldsymbol{\lambda}^2 + \mathbb{E}[(\Lambda_1 - \mathbb{E}[\Lambda_1])^3]\boldsymbol{\lambda}^3).
\end{aligned}$$

For the full covariance matrix between two points in time $0 \leq t_1 \leq t_2$, it holds:

$$\mathbb{C}\mathrm{ov}[\boldsymbol{L}_{t_1}, \boldsymbol{L}_{t_2}] = \mathbb{C}\mathrm{ov}[\boldsymbol{L}_{t_1}, \boldsymbol{L}_{t_1}] = t_1(\mathbb{V}\mathrm{ar}[\Lambda_1]\boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbb{E}[\Lambda_1]diag(\boldsymbol{\lambda})),$$

where $diag(\boldsymbol{\lambda})$ denotes the matrix with only diagonal non-zero entries given by $\boldsymbol{\lambda}$.

Proof. The claim about the moments of $\boldsymbol{L}$ follows again from conditioning on $\sigma(\Lambda)$ and using the moments of the underlying Poisson processes that can be found in (Johnson et al., 1992, Chapter 4.3, p.161):

$$\mathbb{E}[\boldsymbol{N}_t^n] = \sum_{k=0}^{n} S(n,k)(t\boldsymbol{\lambda})^k.$$

Mean as well as second and third moment follow by straightforward calculation and applying Remark 4.14. To show that the covariance matrix between different points in time reduces to the covariance for the overlapping time interval follows for the diagonal entries directly from Remark 4.14 and for the off-diagonal entries from a similar calculation. It remains to be shown that the expression for $\mathbb{C}\mathrm{ov}[L_t^i, L_t^j]$, $i \neq j$, is valid. As always, it is

$$\mathbb{C}\mathrm{ov}[L_t^i, L_t^j] = \mathbb{E}[L_t^i L_t^j] - \mathbb{E}[L_t^i]\,\mathbb{E}[L_t^j].$$

The mixed moment can again be calculated by the tower rule:

$$\begin{aligned}\mathbb{E}[L_t^i L_t^j] &= \mathbb{E}[\mathbb{E}[N_{\Lambda_t}^i N_{\Lambda_t}^j | \sigma(\Lambda)]] = \mathbb{E}[\mathbb{E}[N_{\Lambda_t}^i | \sigma(\Lambda)]\, \mathbb{E}[N_{\Lambda_t}^j | \sigma(\Lambda)]] \\ &= \mathbb{E}[\lambda_i \Lambda_t \lambda_j \Lambda_t] = \lambda_i \lambda_j\, \mathbb{E}[\Lambda_t^2].\end{aligned}$$

Together with the formula for the second moment of the subordinator from Remark 4.14 and the result about the mean of the time-changed process from before, this concludes the proof. □

Theorem 4.15 focuses on the higher dimensional moments of the marginal distributions. For a generalization to higher dimensional mixed moments $\mathbb{E}[\boldsymbol{L}_t^{\boldsymbol{k}}]$, $\boldsymbol{k} \in \mathbb{N}_0^d$ (of general processes), see (Johnson et al., 1997, Chapter 2.2, pp.2).

The result about the mean of the time-changed process validates the time-normalization Assumption (TN) introduced in Section 3.1. We can now state the following equivalence:

$$\text{(TN)} \quad \Leftrightarrow \quad \Psi_\Lambda^{(1)}(0) = 1.$$

The implications of this equivalence on the parameters of the selected subordinator are given for the processes introduced in Example 2.14 in Table 4.4.

The Stirling numbers of the second kind $S(n,k)$, which appear in the formula for the higher order moments, are known from combinatorics as they are equal to the number of ways a set of n elements can be partitioned into k indistinguishable and non-empty subsets. For details on Stirling numbers, see (Comtet, 1974, Chapter 5, pp.205). The Stirling numbers are linked to the complete Bell polynomials by

$$S(k,l) = B_{k,l}(1,\dots,1)$$

and can be efficiently computed using the following triangular scheme:

$$\begin{aligned}&S(n,k) = S(n-1,k-1) + kS(n-1,k), \quad 1 \le k \le n, \\ &S(0,0) = 1, \quad S(n,0) = S(0,k) = 0.\end{aligned}$$

subordinator	$\mathbb{E}[\Lambda_1]$	$\mathbb{V}\mathrm{ar}[\Lambda_1]$	$\mathbb{E}[(\Lambda_1 - \mathbb{E}[\Lambda_1])^3]$	(TN)
Poisson + drift $b \geq 0$, $\xi > 0$	$b + \xi$	ξ	ξ	$\xi = 1 - b$
compound Poisson + drift $b \geq 0$, $\xi > 0$, $Y \geq 0$	$b + \xi\,\mathbb{E}[Y]$	$\xi\,\mathbb{E}[Y^2]$	$\xi\,\mathbb{E}[Y^3]$	$\xi = \frac{1-b}{\mathbb{E}[Y]}$
gamma + drift $b \geq 0$, $\beta, \eta > 0$	$b + \frac{\beta}{\eta}$	$\frac{\beta}{\eta^2}$	$2\frac{\beta}{\eta^3}$	$\beta = \eta(1-b)$
inverse Gaussian + drift $b \geq 0$, $\beta, \eta > 0$	$b + \frac{\beta}{\eta}$	$\frac{\beta}{\eta^3}$	$3\frac{\beta}{\eta^5}$	$\beta = \eta(1-b)$
stable + drift $b \geq 0$, $0 < \alpha < 1$	∞	∞	∞	–

Table 4.4 Moments of subordinator families: The table summarizes mean, variance, and third central moment in time $t = 1$ for the subordinator families introduced in Example 2.14. For other points in time $t \geq 0$ the given formulas have to be multiplied with t.

	$k=1$	$k=2$	$k=3$	$k=4$
$n=1$	1			
$n=2$	1	1		
$n=3$	1	3	1	
$n=4$	1	7	6	1

Table 4.5 Stirling numbers: The table summarizes the Stirling numbers of the second kind $S(n,k)$ for $1 \leq k \leq n$ up to $k = n = 4$.

For $k > n$, the Stirling numbers are zero which allows for the results to be stored in a lower triangular matrix. The values for $1 \leq k \leq n \leq 4$ can be found in Table 4.5; a table up to order 27 is given in (Comtet, 1974, pp.310).

Often in real-world application overdispersion in claim count data is observed, cf. Hougaard et al. (1997), that is the index of dispersion defined as variance to mean ratio is larger than one, the dispersion index of the benchmark Poisson distribution. The presence of overdispersion points towards concentrated or clustered claim occurrences. For the time-changed processes L_t^i studied here, the index is $\mathrm{DI}(L_t^i)$ – existence assumed:

$$\mathrm{DI}(L_t^i) = \frac{\mathbb{V}\mathrm{ar}[L_t^i]}{\mathbb{E}[L_t^i]} = 1 + \lambda_i \frac{\mathbb{V}\mathrm{ar}[\Lambda_1]}{\mathbb{E}[\Lambda_1]} = \mathrm{DI}(N_1^i) + \lambda_i \,\mathrm{DI}(\Lambda_1).$$

The dispersion index for Lévy processes is not a time-dependent quantity but remains constant over time. To account for this, we simply skip the time parameter in the notation when appropriate and write, for example, $\mathrm{DI}(L^i)$. The dispersion index here is always greater than one, the value of the underlying Poisson process, as we have to add the dispersion of the subordinator scaled by the intensity of the Poisson process. Hence, the process features overdispersion as is a well-known fact for any mixed Poisson distribution. This property can also be understood intuitively as the process at hand explicitly supports the arrival of claim clusters. Thus, the construction is explicitly suitable for application to overdispersed data and the degree of overdispersion can be adjusted by choosing a suitable subordinator as a stochastic clock. By imposing the time-normalization Assumption (TN), the mean of the subordinator in $t = 1$ is one and the dispersion index is reduced to the variance of the subordinator.

Dependence structure of the process distribution

From Theorem 4.15, and if the second moment of the subordinator exists, the correlation between marginals $i \neq j$ can be concluded to be:

$$\begin{aligned}\mathbb{C}\mathrm{or}[L^i, L^j] &= \frac{\mathbb{C}\mathrm{ov}[L_1^i, L_1^j]}{\sqrt{\mathbb{V}\mathrm{ar}[L_1^i]\,\mathbb{V}\mathrm{ar}[L_1^j]}} = \frac{\lambda_i \lambda_j \,\mathbb{V}\mathrm{ar}[\Lambda_1]}{\sqrt{(\lambda_i \,\mathbb{E}[\Lambda_1] + \lambda_i^2 \,\mathbb{V}\mathrm{ar}[\Lambda_1])(\lambda_j \,\mathbb{E}[\Lambda_1] + \lambda_j^2 \,\mathbb{V}\mathrm{ar}[\Lambda_1])}} \\ &= \frac{1}{\sqrt{(1 + (\lambda_i \,\mathrm{DI}(\Lambda))^{-1})(1 + (\lambda_j \,\mathrm{DI}(\Lambda))^{-1})}}.\end{aligned}$$

Note that the correlation is, again, not a time-dependent property as time will always cancel out in the calculation. This property is accounted for in the notation $\mathbb{C}\mathrm{or}[L^i, L^j]$

as it was for the dispersion index before. Since the subordinator affects all components similarly, the dependence generated is naturally non-negative. This formula makes explicit that the correlation always takes values in $[0, 1]$ – which is not surprising as Sundt (2000) has already shown that discretely infinitely divisible distributions only allow for non-negative correlations – and depends on the dispersion index of the subordinator (or, in case of time-normalization, the variance) as well as the intensities involved. For fixed intensities $\lambda_i, \lambda_j > 0$, the correlation grows with an increasing dispersion index and the following holds:

$$\begin{aligned} \mathrm{DI}[\Lambda] \to 0 \quad &\Rightarrow \quad \mathbb{C}\mathrm{or}[L^i, L^j] \to 0, \\ \mathrm{DI}[\Lambda] \to \infty \quad &\Rightarrow \quad \mathbb{C}\mathrm{or}[L^i, L^j] \to 1. \end{aligned}$$

On the other hand, given a certain positive dispersion index $\mathrm{DI}[\Lambda] > 0$, the correlation grows with increasing intensities (assume $\lambda_j > 0$ fixed):

$$\begin{aligned} \lambda_i \to 0 \quad &\Rightarrow \quad \mathbb{C}\mathrm{or}[L^i, L^j] \to 0, \\ \lambda_i \to \infty \quad &\Rightarrow \quad \mathbb{C}\mathrm{or}[L^i, L^j] \to 1. \end{aligned}$$

It should also be stressed that the correlation between any pair of components is fundamentally governed by the Lévy subordinator and only varies by a certain scaling with the component-specific intensities, which, obviously, limits the possible range of correlation patterns, as the correlation between any pairs will be similar. To create greater flexibility, additional subordinators can be introduced in the model, an approach that will be discussed in Chapter 6.

Correlation, though the most widely used measure, is just one measure of stochastic dependence. Since correlation is a moment-based measure, it can often be easily calculated. However, the distributions need to have finite second moments or else the correlation is undefined. In many insurance applications heavy-tailed distributions are of importance, thus this restriction often limits the usefulness of the correlation as a dependence measure in these cases. Furthermore, correlation is the canonical measure for the multivariate normal distribution. In this case mean and correlation fully characterize the distribution and, hence, the dependence. In general, however, such full characterization is not the case. Correlation only measures the degree of linear association between pairs of random variables and thus may lead to a severe underestimation of the true dependence. Most obviously, and in contrast to the normal case, the correlation can be zero while the random variables are not independent. In general, thus, two models with the same correlation may have a significantly different dependence structure. For instance, extreme values may show a much higher tendency to occur together in one case than in another, though the correlation is the same. If the model neglects this and only accounts

for the observed correlation, the result may be to overestimate the diversification effect and severely underestimate the risks involved for an insurer. For example, the worst obtainable value at risk (VaR) for a portfolio with given marginal distributions does not necessarily correspond to the highest possible correlation. This stems from the fact that the VaR is not subadditive and situations may exist where the VaR of the sum even exceeds the sum of VaRs of the individuals. Much research regarding bounds for the distribution of a sum with given marginals but unspecified dependence structure has been carried out, see for instance see Frank et al. (1987).

Many other dependence measures are proposed as alternatives to the correlation coefficient (in statistics it is called the Pearson's correlation coefficient). Among the most popular are Spearman's rho (statistics: Spearman's rank correlation coefficient) and Kendall's tau (statistics: Kendall's rank correlation coefficient). While the former measures linear correlation between ranks, the latter measures linear correlation between concordances. Lower and upper tail dependence coefficients are particularly concerned with extreme values and seek to measure asymptotical dependence in the lower and upper tail, respectively. These measures are, in the case of continuous distributions, copula-based in that they do not depend on the marginal distributions. For the definition of these measures, see Section 2.1. For more details on the fallacies in using correlation as a dependence measure and the rationale behind using copulas and copula-associated dependence measures, see Embrechts et al. (2002).

From Sklar's theorem introduced in Sklar (1959) it is known that the dependence structure can be separated from the marginal distributions by a copula function. In case of a continuous distribution, the copula is unique and can be derived in terms of the distribution function and the generalized inverse functions of the marginals, that is for a random vector $\boldsymbol{X}$ on $\mathbb{R}^d$ it is

$$C_{\boldsymbol{X}}(u_1, \dots, u_d) = F_{\boldsymbol{X}}(F_{X_1}^{-1}(u_1), \dots, F_{X_d}^{-1}(u_d)), \quad \boldsymbol{u} \in [0,1]^d. \tag{4.9}$$

This uniqueness only holds if all marginal distributions are continuous, so designating the copula function 'the dependence structure' may only be justified in this case. In particular, continuity is not given for the multivariate discrete claim number distributions studied here, where the copula is following Equation (4.9) only specified at the non-equidistant grid points $F_{X_1}(\mathbb{N}_0) \times \cdots \times F_{X_d}(\mathbb{N}_0)$. Simply extending the definition to any $u_i \in [0,1]$ does not in general even define a proper copula. There can only certain bounds, called Carly bounds, be given within which the copula varies, see Carley (2002). This lack of identifiability also translates to any measure of concordance, see Scarsini (1984). Using the Carly bounds, only ranges for measures like Kendall's tau and Spearman's rho can be given, which may indeed be pretty broad. As both these measures are rank-based, ties have to be taken into account in the discrete case for the calculation.

Different options are discussed like neglecting ties or splitting them in a reasonable way; however, in any case the measures are no longer independent of the marginal distributions. The coefficients of tail dependence may no longer be well defined or even exist at all. For a detailed discussion of copulas and dependence measures in the discrete case, see Genest and Nešlehová (2007).

Later in Chapter 6 we will look at an alternative approach for measuring the dependence in the given model. For now, while keeping the limitations in mind, we will still look not only at the multivariate distribution but also at the copula of $\boldsymbol{L}_t$, $t \geq 0$, for some examples to gain a better understanding of the dependence structure generated in the model. While the multivariate distribution function can be calculated explicitly, an analytic formula for the copula – where it is defined – is not available. It can, however, be calculated from Equation (4.9).

In the following examples, we look at a time horizon of $T = 1$ and select the subordinator parameters such that the time-normalization condition holds. In Figure 4.2, a scatter plot and a contour plot of the copula generated from an inverse Gaussian subordinator is shown. The intensities have been set rather low as $\boldsymbol{\lambda} = (5, 5)'$ and the subordinator parameters have been selected such that the correlation of the time-changed process is roughly 81%. For the scatter plot, the marker size corresponds to the number of equal observations. The points being symmetrically grouped around the bisecting line corresponds to symmetric dependence generated in the model. Positive dependence in the sense of concordance can be observed from the many concordant pairs. As discussed earlier, values only exist at certain grid points and, due to the low intensities, the grid points in the lower right corner are rather far apart while the number of equal observation points is high. This complicates an interpretation of the plot, so we increase the intensities to achieve a more detailed study of the dependence in the following.

Let the intensities be $\boldsymbol{\lambda} = (75, 100)'$. Note that to keep the same level of correlation of approximately 81% after increasing the intensities, the subordinator parameters have to be adjusted in order to decrease the subordinator variance accordingly. Figure 4.3 offers an overview of the resulting bivariate distribution. The upper left plot shows the probability mass functions of the two marginal distributions. While similar in shape, the masses for the second component are shifted to higher values due to the higher intensity of the second component. They also spread out more, as the intensity factors into the variance. The upper right plot shows the bivariate probability mass function, while the middle row shows the functional copula and its contour plot. In the bottom row, the left-hand side shows a scatter plot of 5000 sampled values from the model using Algorithm 3.3, while the right-hand side shows the respective scatter plot of the copula. Compared to the low intensity case before, the copula scatter plot is now much more filled out and many fewer observation points lie on top of each other, which makes it easier to interpret. As before, the points group symmetrically around the bisecting line pointing towards symmetric

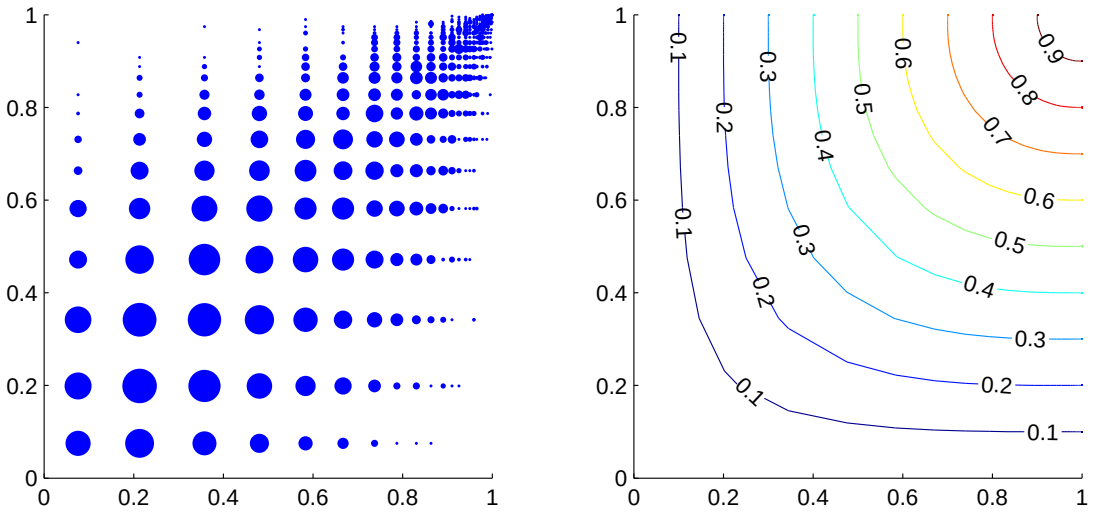


Figure 4.2 Bivariate claim number copula for an inverse Gaussian subordinator and low intensities: For dimension $d = 2$, time $T = 1$, intensity $\boldsymbol{\lambda} = (5, 5)'$, and an inverse Gaussian subordinator ($b = 0$, $\beta = \eta = 1.1$), the figure shows a scatter plot (left) and a contour plot (right) of the claim number copula. The parameters have been set such that time-normalization holds and the correlation of the time-changed process is approximately 81%. The scatter plot is generated from 5000 samples and the marker size corresponds to the number of respective observations.

dependence. The non-uniform intensities, in particular, do not affect the symmetry of the dependence structure. It is now more obvious that the dependence is stronger in extreme scenarios, pointing towards positive tail dependence, a property often observed in real-world applications. As discussed in Section 3.1, the model can feature the simultaneous arrival of a large quantity of claims which could be interpreted as the result of a catastrophic event. Here we now see that bivariate dependence in case of extreme outcomes is particularly high. Thus, the model lends itself well to modelling the joint outcome of portfolios which show exposure to the same kind of catastrophic risk. From the sample, we calculated the empirical Kendall's tau as 0.597 and Spearman's rho as 0.782 using the respective build-in MATLAB® functions[1].

For comparison, Figure 4.4 shows the results in a setting with lower correlation of roughly 29%. The scatter plot of the copula shows only a slight concentration of observation points around the bisecting line and the corners are much less pronounced. Overall, the copula is much closer to the independence copula. The empirical Kendall's tau is given by 0.193, Spearman's rho by 0.281.

In Figure 4.5, the copulas generated from an inverse Gaussian subordinator (top), a gamma subordinator (middle), and a compound Poisson process with gamma-distributed jump sizes are compared using scatter and contour plots. The parameters of the three subordinators have been set to yield about the same correlation in the time-changed process. The level of correlation is increased, compared to the examples before, to 97.5 (Kendall's tau/Spearman's rho: inverse Gaussian– 0.828/0.956; gamma – 0.854/0.956; compound Poisson – 0.875/0.977). Nevertheless, the difference between the plots is not particularly pronounced. Due to the comparably high intensities, the variance of the subordinator has to be rather low (0.44), even for this high level of correlation. Hence, it is not too surprising that the effect of different subordinators is not very strong. Note, however, that the gamma subordinator and the compound Poisson process create a somewhat stronger dependence in the lower tail compared to the inverse Gaussian subordinator, while the inverse Gaussian subordinator tends to dominate in the upper tail. Furthermore, the compound Poisson process shows a pronounced marker at the origin. Compared to the infinite-activity subordinators, the compound Poisson process has a positive probability of not jumping at all until $T = 1$ – particularly in this setting as the intensity is only $\xi = 10/3$ – in which case no claims occur. This property of the finite activity subordinators was also perceptible in the marginal Neyman Type A distribution for a Poisson process in Figure 4.1.

So far, we have only considered subordinators with no drift. Assuming (TN), $b \in [0, 1]$ has to hold for any drift. If the drift is one, the subordinator reduces to a linear function

[1] The ties in the calculation of Spearman's rho are accounted for by averaging the values of the affected ranks. For Kendall's tau, the tie-adjusted version tau-b is calculated. For details we refer to the MATLAB® documentation and the references given there.

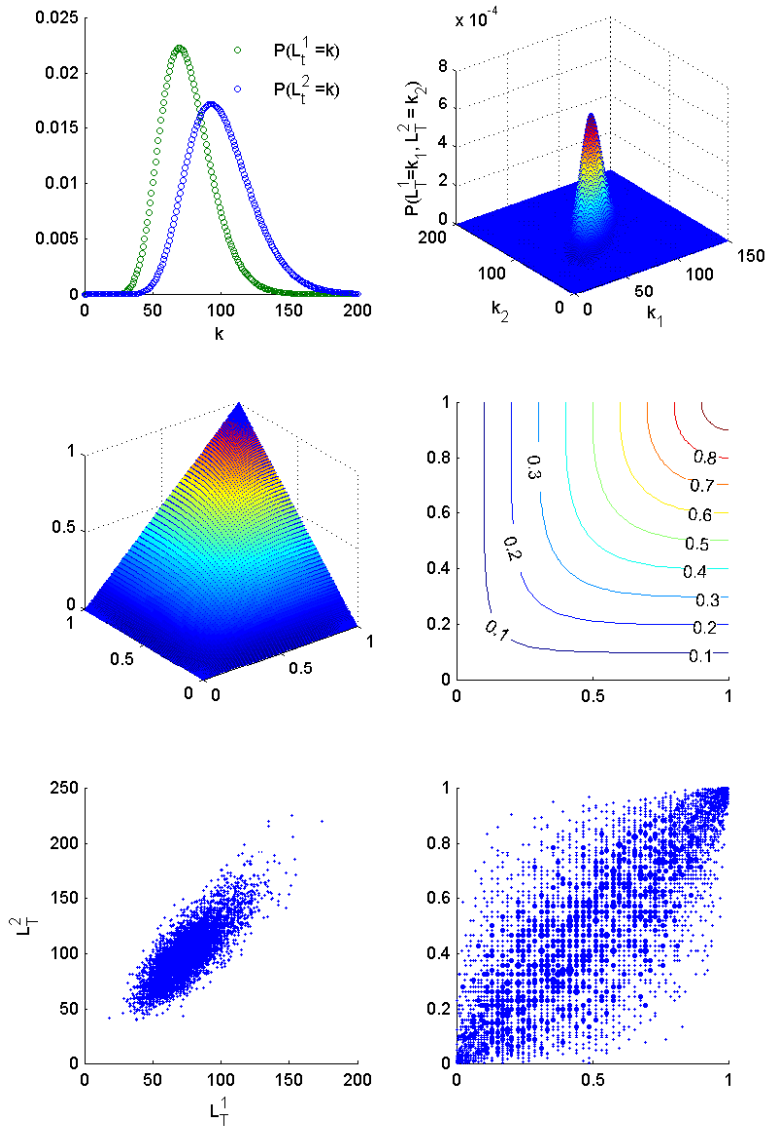


Figure 4.3 Bivariate claim number distribution for an inverse Gaussian subordinator and high correlation: For dimension $d = 2$, time $T = 1$, intensity $\boldsymbol{\lambda} = (75, 100)'$, and an inverse Gaussian subordinator ($b = 0$, $\beta = \eta = 4.5$), the figure shows: (top-left) probability mass functions of marginals; (top-right) bivariate probability mass function; (middle-left) functional copula; (middle-right) contour plot of copula; (bottom-left) scatter of process values; (bottom-right) scatter of copula. The parameters have been set such that time-normalization holds and the correlation of the time-changed process is approximately 81%. The scatter plots are generated from 5000 samples and the marker size corresponds to the number of respective observations.

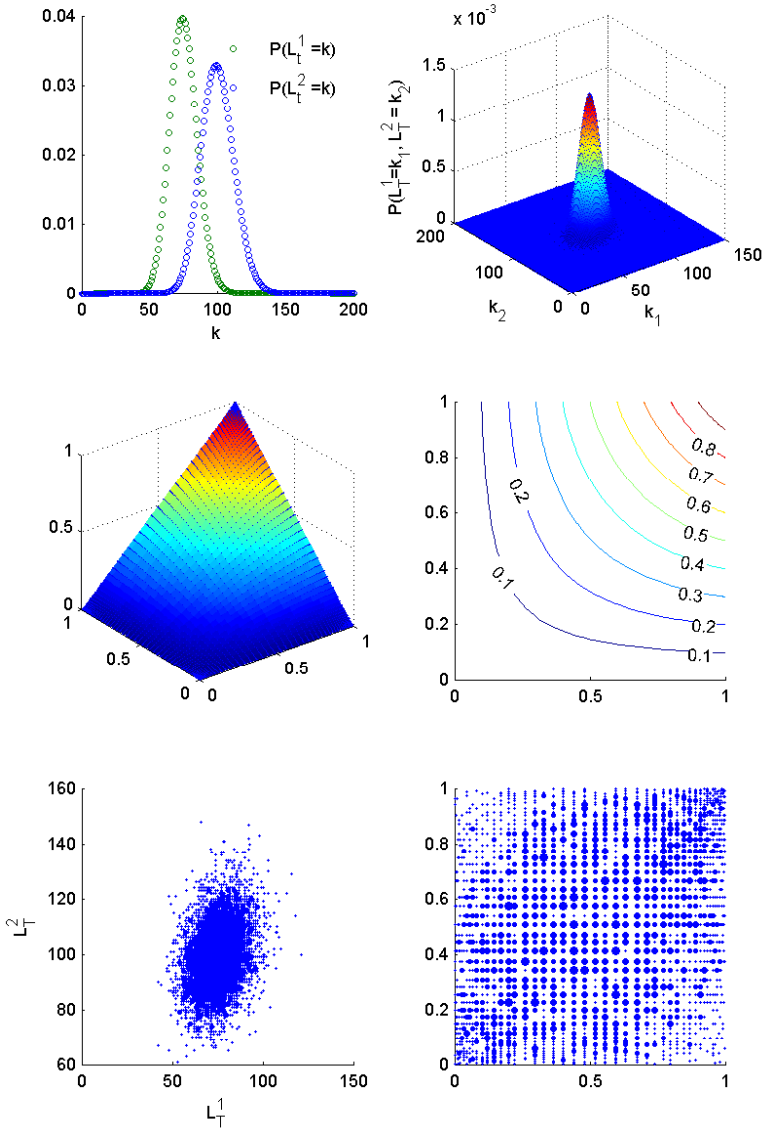


Figure 4.4 Bivariate claim number distribution for an inverse Gaussian subordinator and low correlation: For dimension $d = 2$, time $T = 1$, intensity $\boldsymbol{\lambda} = (75, 100)'$, and an inverse Gaussian subordinator ($b = 0$, $\beta = \eta = 14.5$), the figure shows: (top-left) probability mass functions of marginals; (top-right) bivariate probability mass function; (middle-left) functional copula; (middle-right) contour plot of copula; (bottom-left) scatter of process values; (bottom-right) scatter of copula. The parameters have been set such that time-normalization holds and the correlation of the time-changed process is approximately 29%. The scatter plots are generated from 5000 samples and the marker size corresponds to the number of respective observations.

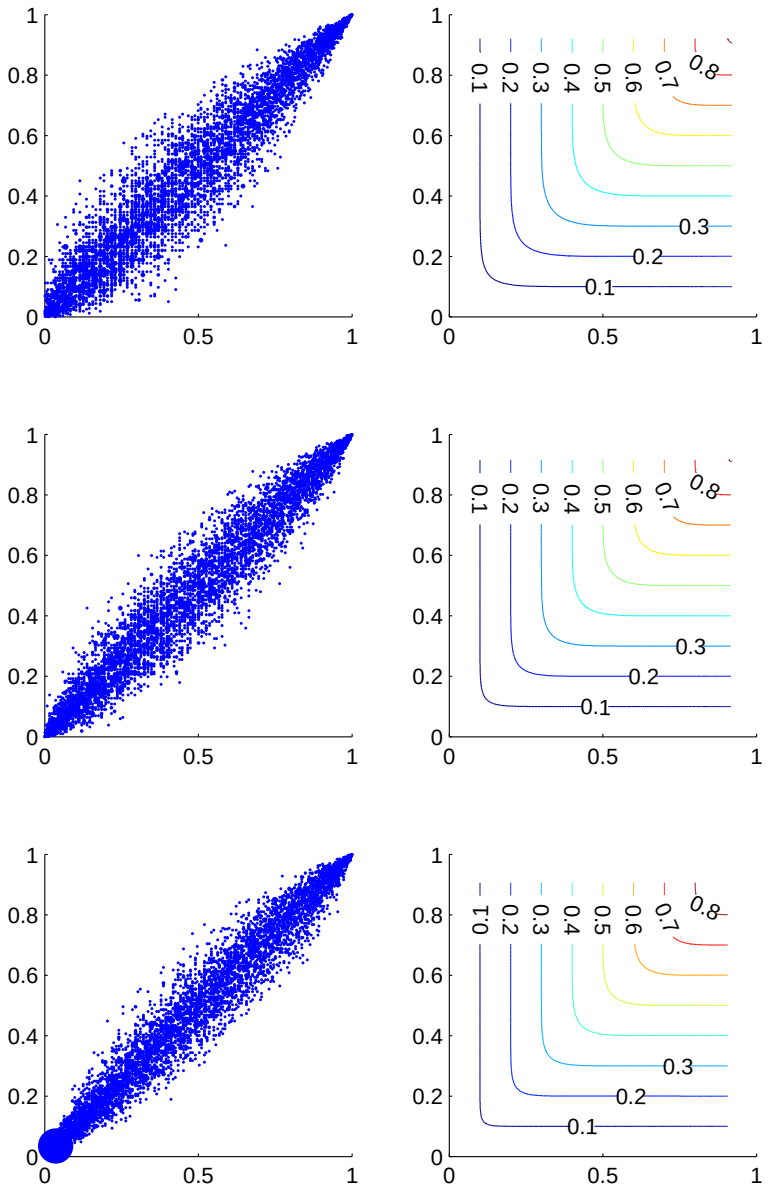


Figure 4.5 Bivariate copula of the claim number distribution for different subordinators: For dimension $d = 2$, time $T = 1$, and intensity $\boldsymbol{\lambda} = (75, 100)'$, the figure shows scatter plots (left) and contour plots (right) of the claim number copulas for an inverse Gaussian (top; $b = 0$, $\beta = \eta = 1.5$), gamma (middle; $b = 0$, $\beta = \eta = 2.25$), and compound Poisson subordinator with gamma distributed jumps (bottom; $b = 0$, $\xi = 10/3$, $\beta = 1.5$, $\eta = 5$). The parameters have been set such that time-normalization holds and the correlation of the time-changed process is approximately 97.5%. The scatter plots are generated from 5000 samples and the marker size corresponds to the number of respective observations.

and no dependence is generated. In general, a drift decreases the overall subordinator variance, and hence the dependence in the model, but the drift can be counteracted by increasing the variance of the stochastic part of the subordinator (given it is finite). For instance, for the parameters of an inverse Gaussian subordinator with time-normalization it holds $\beta = (1-b)\eta$ and the variance is

$$\mathbb{V}\mathrm{ar}[\Lambda_1] = \frac{1-b}{\eta^2}.$$

Thus, for any value of b less than one, the variance can take values from zero to ∞ by adjusting η and, consequently, the correlation in the model can vary between zero and one. However, even given the same correlation, the distribution structure is different if generated with different levels of subordinator drift. Figure 4.6 compares the model for an inverse Gaussian subordinator and a correlation of about 81% in cases with no drift (blue) and with a high drift of $b = 0.8$ (green). The upper left plot shows one sample path of the subordinator for each setting. It can be seen that these paths look fundamentally different – the one with a strong drift and few high jumps, the other with many very small jumps. The middle and bottom plots on the left-hand side compare the marginal distributions. While the center of the distributions remains around the same values, the distribution in the no-drift setting has a much more pronounced peak and a fatter right tail. The upper plot on the right-hand side compares the contour plots of the copulas and the middle and lower plots show a scatter plot of the copula in the two settings. These scatter plots look different, particularly in the tails. In the setting with drift, dependence in the upper tail is stronger compared to the setting with no drift due to the higher probability of large subordinator jumps, while the dependence in the lower tail is smaller due to the continuous drift.

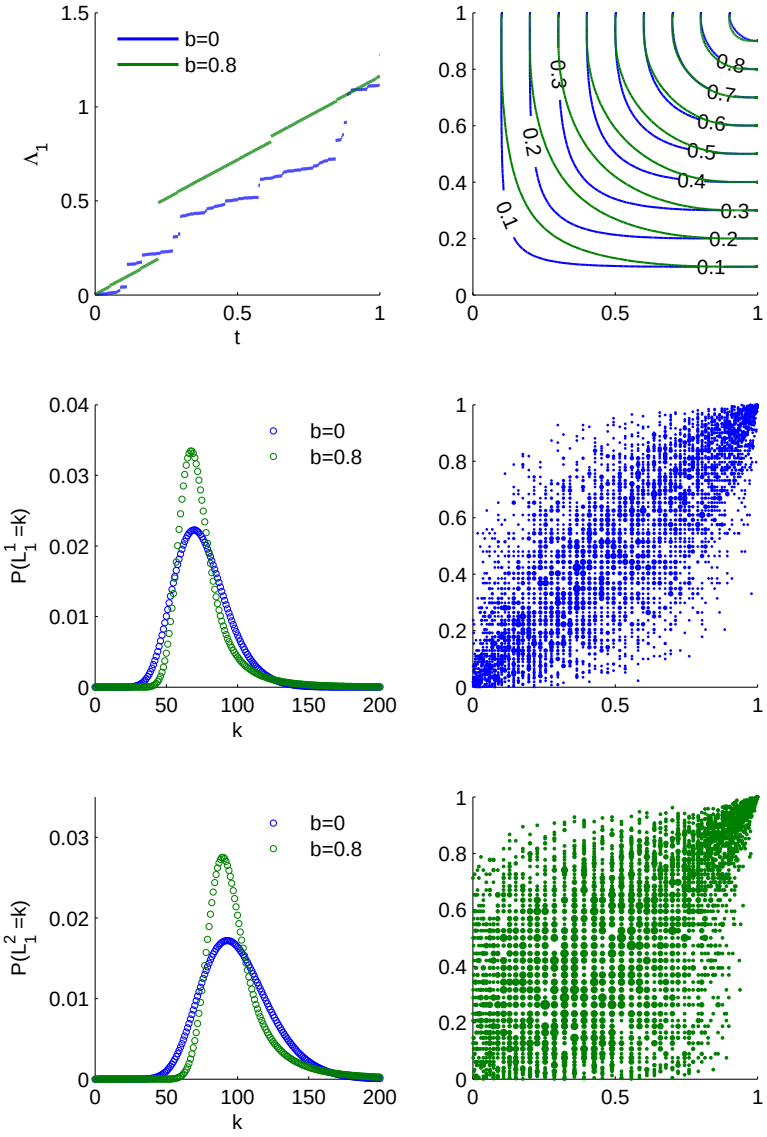


Figure 4.6 Effect of a subordinator drift on the bivariate claim number distribution: For dimension $d = 2$, time $T = 1$, intensity $\boldsymbol{\lambda} = (75, 100)'$, and an inverse Gaussian subordinator with no drift (blue: $b = 0$, $\beta = \eta = 4.5$) and with drift (green: $b = 0.8$, $\beta = 0.4$, $\eta = 2$), the figure shows: (top-left) one sample path of the subordinator for each setting; (middle-left) comparison of marginal distribution of first component; (bottom-left) comparison of marginal distribution of second component; (top-right) comparison of copula contour plot; (middle-right) scatter of copula in setting with no drift; (bottom-right) scatter of copula in setting with drift. The parameters have been set such that time-normalization holds and the correlation of the time-changed process is around 81%. The scatter plots are generated from 5000 samples and the marker size corresponds to the number of respective observations.

4.2 Lévy and compound Poisson characterization

In the previous section, the distribution of the time-changed process $\boldsymbol{L}$ was studied in detail. Many of the results given there can be recovered with only slight modifications for more general time-change processes than a Lévy subordinator, see Remarks 4.3 and 4.5. In this section, the Lévy properties are fully exploited to gain an even more thorough understanding of the nature of the process $\boldsymbol{L}$ and its dependence structure. As a result, a stable and efficient recursive evaluation method for the process distribution is formulated.

As discussed in Section 4.1, subordination of a Lévy process results in another Lévy process and subordination of a Lévy subordinator – which is the case for the claim number process proposed in Model (M) – generates a Lévy subordinator. In the previous section, the multivariate distribution and in particular the copula of the process $\boldsymbol{L}$ were studied to obtain a first idea about the dependence structure introduced in the model by the common time-change. However, copulas are static objects and represent the dependence structure (if it may be called this at all in the case of a discrete distribution) at a certain point in time. While the distribution of a Lévy process is completely determined by its distribution at one point in time, the copula of the coordinates of the process at one point in time can in general not describe the evolution of the dependence structure over time, see (Cont and Tankov, 2003, Chapter 5.4, p.143).

The dependence structure of a Lévy process is of a special kind: due to the Lévy properties, no dependence over time exists. Dependencies between components can be generated in a general Lévy process by correlating the Brownian motion component, see Section 2.3. As a Lévy subordinator has no diffusion component, dependence is solely created by simultaneous jumps, see Theorem 2.13. Therefore, for a deeper understanding of the dependence structure, the Lévy measure, which encodes all information about size and timing of the jumps, has to be examined.

For subordination of Lévy processes, the Lévy characteristics of the new process can be derived from those of the original process and the subordinator. The result can be found in many textbooks on Lévy processes for the subordination of a general Lévy process, i.e. in terms of a characteristic triplet that depends on a certain cutting function for the small jumps. As we are concerned with subordination of a Lévy subordinator, the transformation of the characteristics for this particular setting, i.e. in terms of a characteristic tuple, is given in the following Lemma.

Lemma 4.16 (Subordination of a Lévy subordinator)
Let $\boldsymbol{X} = \{\boldsymbol{X}_t\}_{t\geq 0}$ be a d-dimensional Lévy subordinator with characteristics $\boldsymbol{b_X}$, $\nu_{\boldsymbol{X}}$ and let $\Lambda = \{\Lambda_t\}_{t\geq 0}$ be an independent univariate Lévy subordinator with characteristics

b_Λ, ν_Λ. Then the process $\boldsymbol{Y} := \{\boldsymbol{X}_{\Lambda_t}\}_{t\geq 0}$ is a Lévy subordinator with the following characteristics:

$$\begin{aligned} b_{\boldsymbol{Y}} &= b_\Lambda \boldsymbol{b}_{\boldsymbol{X}}, \\ \nu_{\boldsymbol{Y}}(B) &= b_\Lambda \nu_{\boldsymbol{X}}(B) + \int_0^\infty \mathbb{P}(\boldsymbol{X}_t \in B)\, \nu_\Lambda(\mathrm{d}t), \quad B \in \mathcal{B}(\dot{\mathbb{R}}^d_{\geq 0}). \end{aligned}$$

Proof. The general result in terms of the characteristic triplet $(\boldsymbol{\gamma}_{\boldsymbol{X}}, A_{\boldsymbol{X}}, \nu_{\boldsymbol{X}})$ where $\boldsymbol{\gamma}_{\boldsymbol{X}}$ corresponds to the truncation function $\mathbb{1}_{\{\|\boldsymbol{x}\|\leq 1\}}$ for small jumps is stated in (Sato, 1999, Chapter 6, Section 30, pp.197):

$$\begin{aligned} A_{\boldsymbol{Y}} &= b_\Lambda A_{\boldsymbol{X}}, \\ \nu_{\boldsymbol{Y}}(B) &= b_\Lambda \nu_{\boldsymbol{X}}(B) + \int_0^\infty \mathbb{P}(\boldsymbol{X}_t \in B)\, \nu_\Lambda(\mathrm{d}t), \quad B \in \mathcal{B}(\dot{\mathbb{R}}^d_{\geq 0}), \\ \boldsymbol{\gamma}_{\boldsymbol{Y}} &= b_\Lambda \boldsymbol{\gamma}_{\boldsymbol{X}} + \int_0^\infty \int_{\|\boldsymbol{x}\|\leq 1} \boldsymbol{x}\, \mathbb{P}(\boldsymbol{X}_t \in \mathrm{d}\boldsymbol{x})\, \nu_\Lambda(\mathrm{d}t). \end{aligned}$$

A Lévy subordinator has no diffusion part, thus $A_{\boldsymbol{X}}$ as well as $A_{\boldsymbol{Y}}$ are zero matrices. As the Lévy measure is not affected by the truncation of small jumps, it is the same in both characterizations, which proofs the representation of $\nu_{\boldsymbol{Y}}$ given in the Lemma. The drift $\boldsymbol{b}_{\boldsymbol{X}}$ corresponds to $\boldsymbol{\gamma}_{\boldsymbol{X}}$ by

$$\boldsymbol{b}_{\boldsymbol{X}} = \boldsymbol{\gamma}_{\boldsymbol{X}} - \int_{\|\boldsymbol{x}\|\leq 1} \boldsymbol{x}\, \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}),$$

see (Sato, 1999, Chapter 2, Section 8, Remark 8.4, pp.38). As this relationship holds for $\boldsymbol{b}_{\boldsymbol{Y}}$ and $\boldsymbol{\gamma}_{\boldsymbol{Y}}$ as well, it follows for the drift:

$$\begin{aligned} \boldsymbol{b}_{\boldsymbol{Y}} =& \boldsymbol{\gamma}_{\boldsymbol{Y}} - \int_{\|\boldsymbol{x}\|\leq 1} \boldsymbol{x}\, \nu_{\boldsymbol{Y}}(\mathrm{d}\boldsymbol{x}) \\ =& b_\Lambda \boldsymbol{\gamma}_{\boldsymbol{X}} + \int_0^\infty \int_{\|\boldsymbol{x}\|\leq 1} \boldsymbol{x}\, \mathbb{P}(\boldsymbol{X}_t \in \mathrm{d}\boldsymbol{x})\, \nu_\Lambda(\mathrm{d}t) \\ & - \int_{\|\boldsymbol{x}\|\leq 1} \boldsymbol{x} \Big\{ b_\Lambda\, \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) + \int_0^\infty \mathbb{P}(\boldsymbol{X}_t \in \mathrm{d}\boldsymbol{x})\, \nu_\Lambda(\mathrm{d}t) \Big\} = \\ =& b_\Lambda \Big\{ \boldsymbol{\gamma}_{\boldsymbol{X}} - \int_{\|\boldsymbol{x}\|\leq 1} \boldsymbol{x}\, \nu_{\boldsymbol{X}}(\mathrm{d}\boldsymbol{x}) \Big\} = b_\Lambda \boldsymbol{b}_{\boldsymbol{X}}. \end{aligned}$$

□

Using Lemma 4.16, the characteristics of the time-changed process $\boldsymbol{L}$ are stated in the following Theorem.

Theorem 4.17 (Lévy characteristics of the claim number processes)
The Lévy subordinator $\boldsymbol{L}$ specified in Model (M) by a time-change with the Lévy subordinator Λ has no drift and discrete Lévy measure:

$$\boldsymbol{b_L} = \mathbf{0}; \qquad \nu_{\boldsymbol{L}}(\boldsymbol{k}) = -\frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\Psi_{\Lambda}^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|), \quad \boldsymbol{k} \in \dot{\mathbb{N}}_0^d.$$

The Lévy measure is finite with total mass $\nu_{\boldsymbol{L}}(\dot{\mathbb{N}}_0^d) = \Psi_{\Lambda}(|\boldsymbol{\lambda}|)$.

Proof. According to Example 2.14, the marginal Poisson processes N^i of the underlying process $\boldsymbol{N}$ in Model (M) have the following characteristics:

$$b_{N^i} = 0; \qquad \nu_{N^i}(B) = \lambda_i \delta_1(B), \quad B \in \mathcal{B}(\mathbb{R}_{>0}).$$

As the Poisson process only allows for jumps of size one, its Lévy measure only has a point mass in one. In particular, it is a discrete measure and can be written as

$$\nu_{N^i}(k) = \lambda_i \mathbb{1}_{\{1\}}(k), \quad k \in \mathbb{N}.$$

From the multiplicative relation of the Laplace transform of a vector with independent components, namely $\varphi_{\boldsymbol{N}}(\boldsymbol{x}) = \prod_{i=1}^d \varphi_{N^i}(x_i)$, see Theorem 2.1, together with the Lévy–Khintchine representation in Theorem 2.12 it can be concluded for the characteristics of the process $\boldsymbol{N}$:

$$\boldsymbol{b_N} = (b_{N^1}, \ldots, b_{N^d})', \quad \nu_{\boldsymbol{N}}(B) = \sum_{i=1}^d \nu_{N^i}(B_i), \quad B \in \mathcal{B}(\dot{\mathbb{R}}_{\geq 0}^d),$$

where $B_i := \{x \in \mathbb{R} : x\boldsymbol{e}_i \in B\}$; see also Theorem 2.13 and (Barndorff-Nielsen et al., 2001, after Lemma 3.2, p.167). Given the specifics of $\boldsymbol{N}$, it follows:

$$\boldsymbol{b_N} = \mathbf{0}; \qquad \nu_{\boldsymbol{N}}(\boldsymbol{k}) = \sum_{i=1}^d \lambda_i \mathbb{1}_{\{\boldsymbol{e}_i\}}(\boldsymbol{k}) = \boldsymbol{\lambda}^{\boldsymbol{k}} \mathbb{1}_{\{1\}}(|\boldsymbol{k}|), \quad \boldsymbol{k} \in \dot{\mathbb{N}}_0^d.$$

From Lemma 4.16 it can be concluded that $b_{\boldsymbol{L}} = b_\Lambda \boldsymbol{b_N} = \boldsymbol{0}$ holds for the drift of $\boldsymbol{L}$. Furthermore, due to the discrete distribution of the Poisson process, the Lévy measure of $\boldsymbol{L}$ is discrete as well and it holds:

$$\begin{aligned}\nu_{\boldsymbol{L}}(\boldsymbol{k}) &= b_\Lambda \nu_{\boldsymbol{N}}(\boldsymbol{k}) + \int_0^\infty \mathbb{P}(\boldsymbol{N}_t = \boldsymbol{k})\, \nu_\Lambda(\mathrm{d}t) \\ &= b_\Lambda \boldsymbol{\lambda}^{\boldsymbol{k}} \mathbb{1}_{\{1\}}(|\boldsymbol{k}|) + \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!} \int_0^\infty t^{|\boldsymbol{k}|} \exp\{-|\boldsymbol{\lambda}|t\}\, \nu_\Lambda(\mathrm{d}t), \quad \boldsymbol{k} \in \dot{\mathbb{N}}_0^d.\end{aligned}$$

According to the Lévy–Khintchine representation in Theorem 2.12, the Laplace exponents Ψ_Λ of Λ can be calculated from the characteristics by

$$\Psi_\Lambda(x) = b_\Lambda x + \int_0^\infty (1 - \exp\{-tx\})\, \nu_\Lambda(\mathrm{d}t), \quad x \geq 0.$$

As was done in the proof of Theorem 4.2 for the Laplace transform, it can be concluded from the differentiation lemma for parameter dependent integrals by induction that it holds for the k-th derivative of the Laplace exponent:

$$\Psi_\Lambda^{(k)}(x) = b_\Lambda \mathbb{1}_{\{1\}}(k) + (-1)^{k+1} \int_0^\infty t^k \exp\{-tx\}\, \nu_\Lambda(\mathrm{d}t), \quad x > 0.$$

Together, it follows for the Lévy measure of $\boldsymbol{L}$:

$$\begin{aligned}\nu_{\boldsymbol{L}}(\boldsymbol{k}) &= b_\Lambda \boldsymbol{\lambda}^{\boldsymbol{k}} \mathbb{1}_{\{1\}}(|\boldsymbol{k}|) + (-1)^{|\boldsymbol{k}|+1} \frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!} (\Psi^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|) - b_\Lambda \mathbb{1}_{\{1\}}(|\boldsymbol{k}|)) \\ &= -\frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \Psi_\Lambda^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|).\end{aligned}$$

The total mass of the Lévy measure follows from the Taylor series expansion of Ψ_Λ at point $|\boldsymbol{\lambda}|$ evaluated in zero, together with the multinomial theorem:

$$\begin{aligned}\nu_{\boldsymbol{L}}(\dot{\mathbb{N}}_0^d) &= -\sum_{\boldsymbol{k} \in \dot{\mathbb{N}}_0^d} \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \Psi_\Lambda^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|) = -\sum_{k=1}^\infty \Psi_\Lambda^{(k)}(|\boldsymbol{\lambda}|) \sum_{|\boldsymbol{k}|=k} \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \\ &= -\sum_{k=1}^\infty \Psi_\Lambda^{(k)}(|\boldsymbol{\lambda}|) \frac{(-|\boldsymbol{\lambda}|)^k}{k!} = -\Psi_\Lambda(0) + \Psi_\Lambda(|\boldsymbol{\lambda}|) = \Psi_\Lambda(|\boldsymbol{\lambda}|).\end{aligned}$$

□

Note that the Laplace exponent Ψ_Λ is a Bernstein function, i.e. its derivatives of all order $k \in \mathbb{N}$ exist and alternate in sign:

$$(-1)^{k+1}\Psi^{(k)}(x) \geq 0, \quad x > 0,$$

see Section 2.3. Thus the Lévy measure in Theorem 4.17 is indeed non-negative. In contrast to the derivatives of the Laplace transform, the derivatives of the Laplace exponent of the subordinator families introduced in Example 2.14 can easily be calculated in closed form; these derivatives as well as the induced Lévy measure are summarized in Table 4.6. Some considerations regarding a stable and efficient implementation will be given later on.

Remark 4.18 (Lévy measure of the claim number process)
As discussed in Section 2.3, the jump measure J of a Lévy process $\boldsymbol{X}$,

$$J(A) := |\{t \geq 0 : (t, \nabla \boldsymbol{X}'_t)' \in A\}|, \quad A \in \mathcal{B}(\mathbb{R}_{\geq 0} \times \dot{\mathbb{R}}^d),$$

is a marked PRM and its mean measure is the product measure $Leb \times \nu_{\boldsymbol{X}}$ of the Lebesgue measure and the Lévy measure of the process. Considering the discreteness of the time-changed process $\boldsymbol{L}$ and Definition 2.5 of a PRM, it follows for $\boldsymbol{L}$ that multivariate jumps of a certain size arrive according to independent homogeneous Poisson processes, where the intensity is determined by the mass assigned by the Lévy measure to the respective jump size. More precisely, let $N(\boldsymbol{k}) := \{N_t(\boldsymbol{k})\}_{t \geq 0}$ for $\boldsymbol{k} \in \dot{\mathbb{N}}_0^d$ be the process counting the jumps of $\boldsymbol{L}$ of size $\boldsymbol{k}$:

$$N_t(\boldsymbol{k}) := |\{0 \leq s \leq t : \nabla \boldsymbol{L}_s = \boldsymbol{k}\}|, \quad t \geq 0.$$

Then $N(\boldsymbol{k})$ is a Poisson process with intensity $\nu_{\boldsymbol{L}}(\boldsymbol{k})$, and for different $\boldsymbol{k}$ the processes are independent. In particular, the number of (simultaneous) claims of size $\boldsymbol{k}$ in any given time interval follows a Poisson distribution with mean

$$\mathbb{E}[|\{0 \leq s \leq t : \nabla \boldsymbol{L}_s = \boldsymbol{k}\}|] = t\nu_{\boldsymbol{L}}(\boldsymbol{k}), \quad t \geq 0.$$

To construct the process $\boldsymbol{L}$ from a shock model approach as, for instance, in McNeil et al. (2005), where dependence is generated by the superposition of Poisson shock processes affecting certain components (cf. Section 3.3), one would need according to the representation

Λ Poisson process with drift: $b \geq 0, \xi > 0$	
$\Psi_\Lambda(x)$	$bx + \xi(1 - \exp\{-x\})$
$\Psi_\Lambda^{(k)}(x)$	$b\delta_1(k) + (-1)^{k+1}\xi \exp\{-x\}$
$\nu_{\boldsymbol{L}}(\boldsymbol{k})$	$\frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!}(b\delta_1(\lvert\boldsymbol{k}\rvert) + \xi \exp\{-\lvert\boldsymbol{\lambda}\rvert\}$
Λ compound Poisson process with drift: $b \geq 0$, $\xi > 0$, $Y \geq 0$	
$\Psi_\Lambda(x)$	$bx + \xi(1 - \varphi_Y(x))$
$\Psi_\Lambda^{(k)}(x)$	$b\delta_1(k) - \xi\varphi_Y^{(k)}(x)$
$\nu_{\boldsymbol{L}}(\boldsymbol{k})$	$\frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!}(b\delta_1(\lvert\boldsymbol{k}\rvert) + (-1)^{\lvert\boldsymbol{k}\rvert}\xi\varphi_Y^{(k)}(\lvert\boldsymbol{\lambda}\rvert))$
Λ gamma subordinator with drift: $b \geq 0, \beta > 0, \eta > 0$	
$\Psi_\Lambda(x)$	$bx + \beta \log(1 + \frac{x}{\eta})$
$\Psi_\Lambda^{(k)}(x)$	$b\delta_1(k) + (-1)^{k-1}(k-1)!\beta(\eta + x)^{-k}$
$\nu_{\boldsymbol{L}}(\boldsymbol{k})$	$\frac{\boldsymbol{\lambda}^{\boldsymbol{k}}(\lvert\boldsymbol{k}\rvert-1)!}{\boldsymbol{k}!}(b\delta_1(\lvert\boldsymbol{k}\rvert) + \beta(\eta + \lvert\boldsymbol{\lambda}\rvert)^{-\lvert\boldsymbol{k}\rvert})$
Λ inverse Gaussian subordinator with drift: $b \geq 0, \beta > 0, \eta > 0$	
$\Psi_\Lambda(x)$	$bx + \beta(\sqrt{2x + \eta^2} - \eta)$
$\Psi_\Lambda^{(k)}(x)$	$b\delta_1(k) + (-1)^{k-1}\beta(2x + \eta^2)^{\frac{1}{2}-k} \prod_{j=1}^{k-1}(2j-1)$
$\nu_{\boldsymbol{L}}(\boldsymbol{k})$	$\frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!}(b\delta_1(\lvert\boldsymbol{k}\rvert) + \beta(2\lvert\boldsymbol{\lambda}\rvert + \eta^2)^{\frac{1}{2}-\lvert\boldsymbol{k}\rvert} \prod_{j=1}^{\lvert\boldsymbol{k}\rvert-1}(2j-1))$
Λ stable subordinator with drift: $b \geq 0, 0 < \alpha < 1$	
$\Psi_\Lambda(x)$	$bx + x^\alpha$
$\Psi_\Lambda^{(k)}(x)$	$b\delta_1(k) + (-1)^{k-1}x^{\alpha-k}\alpha \prod_{j=1}^{k-1}(j-\alpha)$
$\nu_{\boldsymbol{L}}(\boldsymbol{k})$	$\frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!}(b\delta_1(\lvert\boldsymbol{k}\rvert) + \lvert\boldsymbol{\lambda}\rvert^{\alpha-\lvert\boldsymbol{k}\rvert}\alpha \prod_{j=1}^{\lvert\boldsymbol{k}\rvert-1}(j-\alpha))$

Table 4.6 Derivatives of Laplace exponent for different subordinators and Lévy measure of $\boldsymbol{L}$: The table shows the Laplace exponent and its derivatives of order $k \in \mathbb{N}$ as well as the Lévy measure of the time-changed process $\boldsymbol{L}$ in $\boldsymbol{k} \in \dot{\mathbb{N}}_0^d$ for the families of Lévy subordinators introduced in Example 2.14. The empty product is to be understood as one.

$$\{\boldsymbol{L}_t\}_{t\geq 0} = \Big\{ \sum_{\boldsymbol{k}\in\dot{\mathbb{N}}_0^d} N_t(\boldsymbol{k})\boldsymbol{k} \Big\}_{t\geq 0}.$$

an infinite number of shocks – one for each possible jump size. Hence, in a shock model approach an infinite number of Poisson intensities would need to be specified, whereas in the given approach based on a common time-change only d intensities and the subordinator parameters are necessary. ▲

The behaviour of the (joint) jumps of $\boldsymbol{L}$ is characterized in Remark 4.18. As discussed at the beginning of this section, the joint jumps carry information about the dependence structure of the process. For multivariate distributions, copulas are used to separate the dependence structure from the marginal distributions. A comparable concept is available for Lévy processes, so called Lévy copulas that separate the marginals of the Lévy measure from its dependence structure. In contrast to a probability measure with values in $[0,1]$ and total mass one, a Lévy measure may have infinite mass with a non-integrable singularity at zero, see the discussion in Section 2.3. Therefore, the theory of copulas is not directly transferable to the Lévy measure. Lévy copulas were introduced in Tankov (2003) and Kallsen and Tankov (2006) to provide the necessary adjustments, see also (Cont and Tankov, 2003, Chaper 5.5 and 5.6, pp.145) for an introduction. The case is simplified if the Lévy process has only non-negative jumps, in particular for Lévy subordinators, where the Lévy copula is comparable to a survival copula of a probability distribution. Let $\nu_{\boldsymbol{X}}$ be the Lévy measure of a d-dimensional Lévy subordinator $\boldsymbol{X}$ and let the so-called tail integral be denoted by $G(\cdot)$:

$$G(\boldsymbol{x}) := \nu_{\boldsymbol{X}}([x_1,\infty)\times\cdots\times[x_d,\infty)), \quad \boldsymbol{x}\in\mathbb{R}_{>0}^d.$$

Furthermore, let $G_i(x) := G(x\boldsymbol{e}_i)$, $x>0$, be the marginal tail integrals. Then the Lévy copula is a function $C_{\boldsymbol{X}}^L\colon \mathbb{R}_{>0}^d \to \mathbb{R}_{\geq 0}$ (which exits) such that

$$G(\boldsymbol{x}) = C_{\boldsymbol{X}}^L(G_1(x_1),\ldots,G_d(x_d)).$$

Tail integrals are used in the definition as these quantities are always finite – the origin is by definition not included in the calculation. In the special case of the claim number process $\boldsymbol{L}$ studied here, the Lévy measure $\nu_{\boldsymbol{L}}$ is finite and can be normalized by its total mass to a probability measure $\Psi_\Lambda(|\boldsymbol{\lambda}|)^{-1}\nu_{\boldsymbol{L}}$. Then the Lévy copula is indeed only a scaled version of the survival copula and we can proceed to study the distribution function and the copula of $\Psi_\Lambda(|\boldsymbol{\lambda}|)^{-1}\nu_{\boldsymbol{L}}$ rather than the Lévy measure $\nu_{\boldsymbol{L}}$ and its Lévy

copula. Of course, the distribution is again discrete, so the limitations of the copula function discussed at the end of the previous section have to be kept in mind.

As discussed in Section 2.3, Lévy processes with finite activity, i.e. without Gaussian part and with finite Lévy measure, are compound Poisson processes (with drift). Since $\boldsymbol{L}$ has in addition piecewise constant paths with values in $\mathbb{N}_0^d$, it has to be a Poisson cluster process and $\Psi_\Lambda(|\boldsymbol{\lambda}|)^{-1}\nu_{\boldsymbol{L}}$ is the cluster size distribution; this observation is made precise in the following corollary.

Corollary 4.19 (Poisson cluster process representation of the claim number process) *The claim number process $\boldsymbol{L}$ has a representation as d-dimensional Poisson cluster process. Let $M = \{M_t\}_{t\geq 0}$ be a univariate Poisson process with intensity $\lambda^M := \Psi_\Lambda(|\boldsymbol{\lambda}|)$ and let $\boldsymbol{Y} \sim \Psi_\Lambda(|\boldsymbol{\lambda}|)^{-1}\nu_{\boldsymbol{L}}$ be a random vector on $\dot{\mathbb{N}}_0^d$, that is*

$$\mathbb{P}(\boldsymbol{Y} = \boldsymbol{k}) = \Psi_\Lambda(|\boldsymbol{\lambda}|)^{-1}\nu_{\boldsymbol{L}}(\boldsymbol{k}) = -\frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\frac{\Psi_\Lambda^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)}, \quad \boldsymbol{k} \in \dot{\mathbb{N}}_0^d.$$

Then, using iid copies $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ of $\boldsymbol{Y}$ independent of M, the process $\boldsymbol{L}$ has the representation

$$\{\boldsymbol{L}_t\}_{t\geq 0} \stackrel{d}{=} \Big\{\sum_{j=1}^{M_t} \boldsymbol{Y}_j\Big\}_{t\geq 0},$$

and the cluster size distribution $\boldsymbol{Y}$ has the Laplace transform

$$\varphi_{\boldsymbol{Y}}(\boldsymbol{x}) = 1 - \frac{\Psi_{\boldsymbol{L}}(\boldsymbol{x})}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} = 1 - \frac{\Psi_\Lambda\Big(|\boldsymbol{\lambda}| - \sum_{i=1}^d \lambda_i \exp\{-x_i\}\Big)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)}, \quad \boldsymbol{x} \in \mathbb{R}_{\geq 0}^d.$$

Proof. As stated in Section 2.3, a Lévy process with piecewise constant paths is a compound Poisson process (or a zero process), see (Sato, 1999, Chapter 4, Section 21, Theorem 12, p.135) and, due to its non-negativity and discreteness, $\boldsymbol{L}$ even has to be a Poisson cluster process. Furthermore, it was mentioned in Section 2.3 that $Leb \times \nu_{\boldsymbol{L}}$ is the mean measure of the marked PRM counting the jumps of $\boldsymbol{L}$. According to the discussion after Definition 2.7, the mean measure of a compound Poisson process is $\lambda^M Leb \times \mathbb{P}_{\boldsymbol{Y}}$; this result is also explicitly stated in (Cont and Tankov, 2003, Chapter 33, Proposition 3.5, p.75). Hence, it is $\nu_{\boldsymbol{L}} = \lambda^M \mathbb{P}_{\boldsymbol{Y}}$ and according to Theorem 4.17:

$$\Psi_\Lambda(|\boldsymbol{\lambda}|) = \nu_{\boldsymbol{L}}(\dot{\mathbb{R}}^d_{\geq 0}) = \lambda^M \mathbb{P}(\boldsymbol{Y} \in \dot{\mathbb{R}}^d_{\geq 0}) = \lambda^M.$$

The Laplace exponent of a compound Poisson process in the univariate case was introduced in Example 2.14; the multivariate version can be found in (Sato, 1999, Chapter 1, Section 4, pp.18):

$$\Psi_{\boldsymbol{L}}(\boldsymbol{x}) = \lambda^M(1 - \varphi_{\boldsymbol{Y}}(\boldsymbol{x})), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

According to Proposition 4.1, the Laplace exponent of $\boldsymbol{L}$ also has the following representation:

$$\Psi_{\boldsymbol{L}}(\boldsymbol{x}) = \Psi_\Lambda\Big(\sum_{i=1}^d \lambda_i(1 - \exp\{-x_i\})\Big) = \Psi_\Lambda\Big(|\boldsymbol{\lambda}| - \sum_{i=1}^d \lambda_i \exp\{-x_i\}\Big), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

Equating the two expressions and solving for $\varphi_{\boldsymbol{Y}}$ shows the formula given in the corollary for the Laplace transform of $\boldsymbol{Y}$ and concludes the proof. □

According to the above corollary, the claims in Model (M) arrive in clusters of iid random size $\boldsymbol{Y}$, and the cluster arrivals are determined by the Poisson process M with intensity $\Psi_\Lambda(|\boldsymbol{\lambda}|)$. In comparison, in the underlying model $\boldsymbol{N}$ consisting of independent Poisson processes, the claims arrive one after another, that is with jump size in $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$, and with intensity $|\boldsymbol{\lambda}|$. Furthermore, given the compound Poisson representation of $\boldsymbol{L}$, the observations in Remark 4.18 simply reflect the space-time decomposition property of the compound Poisson process given in Remark 2.8.

Remark 4.20 (Infinitely divisible mixed Poisson vs. compound Poisson distributions)
In the univariate case, (Feller, 1968, Chapter 12.2, p.290) show that the non-negative, discretely infinitely divisible distributions coincide with the compound Poisson distributions on the non-negative integers. The multivariate version was discussed more recently in Sundt (2000). This connection was already given in Section 2.3, as it is directly linked to Lévy process theory. Theorem 2.11 states that the infinitely divisible distributions directly correspond to Lévy processes and, following the discussion about the Lévy–Itô decomposition, the only Lévy processes with paths in $\mathbb{N}_0^d$ are Poisson cluster processes. Furthermore, a univariate mixed Poisson distribution is discretely infinitely divisible – and hence a compound Poisson distribution with support $\mathbb{N}_0$ – if its non-negative mixing variable is infinitely divisible, see (Grandell, 1997, Chapter 2, Proposition 2.5, p.27). The

opposite of the statement is not generally true, but counterexamples are rather complicated to construct, see the discussion in (Grandell, 1997, Chapter 2, pp.28). Given the one-to-one relationship of infinitely divisible distributions and Lévy processes – and in the non-negative case Lévy subordinators – 'nearly all' univariate Poisson cluster processes can be generated using the construction in Model (M), i.e. through time-changing a Poisson process with an independent Lévy subordinator. In the multivariate case, however, the construction is more limited. The marginal processes cannot be selected freely but are closely connected by the common subordinator. In addition, the dependence between the marginal processes is generated by the subordinator, therefore not every dependence structure can be imposed. ▲

Remark 4.21 (Marginal vs. univariate jump size distribution)
Applying Corollary 4.19 to the univariate case it is easily concluded that each marginal process L_i of $\boldsymbol{L}$ can be represented as a Poisson cluster process with Poisson process $\tilde{M}^i = \{\tilde{M}^i_t\}_{t\geq 0}$ with intensity $\lambda^{\tilde{M}^i} = \Psi_\Lambda(\lambda_i)$ and cluster size distribution $\tilde{Y}_i \sim \Psi_\Lambda(\lambda_i)^{-1}\nu_{L^i}(\lambda_i)$, that is

$$\mathbb{P}(\tilde{Y}_i = k) = -\frac{(-\lambda_i)^k}{k!}\frac{\Psi^{(k)}_\Lambda(\lambda_i)}{\Psi_\Lambda(\lambda_i)}, \quad k \in \mathbb{N}.$$

It should be noted, however, that $\tilde{Y}_i$ does not coincide with the marginal Y_i of the multivariate cluster size distribution $\boldsymbol{Y}$. The marginal Y_i corresponds to the cluster size distribution of the compound Poisson representation of L^i with claim arrival intensity $\lambda^M = \Psi_\Lambda(|\boldsymbol{\lambda}|)$. Whereas $\tilde{Y}_i$ is concentrated on $\mathbb{N}$, Y_i has a positive mass at zero as in the multivariate version a jump is recorded if at least one component – not necessarily component i – jumps. The distribution function of Y_i could be calculated directly from the distribution of $\boldsymbol{Y}$ as (w.l.o.g. for $i = 1$):

$$\mathbb{P}(Y_1 = k) = \begin{cases} \sum_{\boldsymbol{k}\in\dot{\mathbb{N}}^{d-1}} \mathbb{P}(\boldsymbol{Y} = (0, \boldsymbol{k}')') & \text{if } k = 0, \\ \sum_{\boldsymbol{k}\in\mathbb{N}^{d-1}} \mathbb{P}(\boldsymbol{Y} = (k, \boldsymbol{k}')') & \text{if } k \in \mathbb{N}. \end{cases}$$

It is easier to observe that the two compound Poisson representations are connected by zero-inflation of the jump size distribution. Consider a general Poisson cluster process

$$\sum_{j=1}^{K_t} Z_j, \quad t \geq 0,$$

where $K = \{K_t\}_{t\geq 0}$ is a Poisson process with intensity $\zeta > 0$ and Z is a jump size distribution concentrated on the positive integers $\mathbb{N}$. Let J be a Bernoulli random variable with

$$\mathbb{P}(J = 0) = 1 - \mathbb{P}(J = 1) =: p.$$

Then the zero-inflation $\tilde{Z}$ of Z is defined as

$$\tilde{Z} := 0\mathbb{1}_{\{J=0\}} + Z\mathbb{1}_{\{J=1\}} = JZ$$

and has the distribution

$$\mathbb{P}(\tilde{Z} = k) = \begin{cases} p & \text{if } k = 0, \\ (1-p)\mathbb{P}(Z = k) & \text{if } k \in \mathbb{N}. \end{cases}$$

Let now $\tilde{K} = \{\tilde{K}_t\}_{t\geq 0}$ be an independent Poisson process with intensity $\tilde{\zeta} > 0$ and consider the Poisson cluster process

$$\sum_{j=1}^{\tilde{K}_t} \tilde{Z}_j, \quad t \geq 0.$$

By the space-time decomposition property of compound Poisson processes, see Remark 2.8, jumps of size zero arrive according to a Poisson process with intensity $p\tilde{\zeta}$. Independently, jumps of sizes in $\mathbb{N}$ arrive according to a Poisson process with intensity $(1-p)\tilde{\zeta}$. Choosing $\tilde{\zeta} := \frac{\zeta}{1-p}$ therefore establishes the equality

$$\Big\{\sum_{j=1}^{K_t} Z_j\Big\}_{t\geq 0} \stackrel{d}{=} \Big\{\sum_{j=1}^{\tilde{K}_t} \tilde{Z}_j\Big\}_{t\geq 0}.$$

Now back to the distribution of the jump sizes Y_i of the claim number process L^i. Comparing the jump intensities it has to hold:

$$p = 1 - \frac{\lambda^{\tilde{M}^i}}{\lambda^M} = 1 - \frac{\Psi_\Lambda(\lambda_i)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)}.$$

Then it can be concluded for the marginal distribution Y_i of $\boldsymbol{Y}$:

$$\mathbb{P}(Y_i = k) = \begin{cases} 1 - \frac{\Psi_\Lambda(\lambda_i)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} & \text{if } k = 0, \\ \frac{\Psi_\Lambda(\lambda_i)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} \mathbb{P}(\tilde{Y}_i = k) = -\frac{(-\lambda_i)^k}{k!} \frac{\Psi_\Lambda^{(k)}(\lambda_i)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} & \text{if } k \in \mathbb{N}. \end{cases}$$

All considerations are readily extended to the multivariate case, that is for some $1 \leq r \leq d$ the r-dimensional marginals, w.l.o.g. $(Y_1, \ldots, Y_r)'$, have the following distribution:

$$\mathbb{P}((Y_1, \ldots, Y_r)' = \boldsymbol{k}) = \begin{cases} 1 - \frac{\Psi_\Lambda(\lambda_1 + \ldots + \lambda_r)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} & \text{if } k = \mathbf{0} \in \mathbb{N}_0^r, \\ -\frac{(-\lambda_1)^{k_1} \ldots (-\lambda_r)^{k_r}}{\boldsymbol{k}!} \frac{\Psi_\Lambda^{(|\boldsymbol{k}|)}(\lambda_1 + \ldots + \lambda_r)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} & \text{if } \boldsymbol{k} \in \dot{\mathbb{N}}_0^r. \end{cases}$$

By rescaling with the total mass of the Lévy measure it follows for the r-dimensional marginals $\nu_{\boldsymbol{L}}^{1,\ldots,r}$ of the Lévy measure $\nu_{\boldsymbol{L}}$:

$$\nu_{\boldsymbol{L}}^{1,\ldots,r}(\boldsymbol{k}) = \begin{cases} \Psi_\Lambda(|\boldsymbol{\lambda}|) - \Psi_\Lambda(\lambda_1 + \ldots + \lambda_r) & \text{if } k = \mathbf{0} \in \mathbb{N}_0^r, \\ -\frac{(-\lambda_1)^{k_1} \ldots (-\lambda_r)^{k_r}}{\boldsymbol{k}!} \Psi_\Lambda^{(|\boldsymbol{k}|)}(\lambda_1 + \ldots + \lambda_r) & \text{if } \boldsymbol{k} \in \dot{\mathbb{N}}_0^r. \end{cases}$$

Hence, the marginal measure $\nu_{\boldsymbol{L}}^{1,\ldots,r}$ of the Lévy measure $\nu_{\boldsymbol{L}}$ coincides on $\dot{\mathbb{N}}_0^r$ with the Lévy measure of $(L^1, \ldots, L^r)'$ but has additional mass $\Psi_\Lambda(|\boldsymbol{\lambda}|) - \Psi_\Lambda(\lambda_1 + \ldots + \lambda_r)$ in zero. This information is useful, as it tells us how often certain components jump alone or together. For instance, jumps only in component 1 occur if $\boldsymbol{L}$ jumps but none of the components $L^2, \ldots, L^d$ jumps, hence with intensity $\nu_{\boldsymbol{L}}^{2,\ldots,d}(\mathbf{0}) = \Psi_\Lambda(|\boldsymbol{\lambda}|) - \Psi_\Lambda(\lambda_2 + \cdots + \lambda_d)$. ▲

Remark 4.22 (Compound Poisson decomposition)
As discussed in Remark 4.18, to construct $\boldsymbol{L}$ from a shock model approach, an infinite number of shocks have to be considered – one for each possible jump size – and consequently an infinite number of parameters needs to be specified. Using the space-time decomposition property, the process also has a representation as the superposition of independent Poisson cluster processes, where each of these cluster processes considers simultaneous jumps affecting exclusively certain components. Let for all elements $I \in \mathcal{P}(\{1, \ldots, d\})$, the power set of $\{1, \ldots, d\}$, the independent Poisson processes $\tilde{N}(I)$ count the number of simultaneous claim arrivals exclusively to components in I, that is:

$$\tilde{N}_t(I) := |\{t > 0 : \nabla L_t^i > 0 \text{ for } i \in I \text{ and } \nabla L_t^i = 0 \text{ for } i \notin I\}|.$$

Following this approach, $2^d - 1$ intensities and as many (multivariate) jump size distributions need to be specified, more precisely $\binom{d}{i}$ i-dimensional jump size distributions for $i = 1, \dots, d$. In the given model in dimension $d = 2$ and for a non-trivial subordinator[2] the following decomposition holds:

$$\{\boldsymbol{L}_t\}_{t\geq 0} \stackrel{d}{=} \Big\{\Big(\sum_{j=1}^{\tilde{N}_t(\boldsymbol{e}_1)} Z_{1j}, \sum_{j=1}^{\tilde{N}_t(\boldsymbol{e}_2)} Z_{2j}\Big)' + \sum_{j=1}^{\tilde{N}_t(\mathbf{1})} \boldsymbol{X}_j\Big\}_{t\geq 0},$$

where $\tilde{N}(\boldsymbol{e}_1), \tilde{N}(\boldsymbol{e}_2), \tilde{N}(\mathbf{1})$ are independent Poisson processes with intensity $\Psi_\Lambda(\lambda_1+\lambda_2)-\Psi_\Lambda(\lambda_2)$, $\Psi_\Lambda(\lambda_1+\lambda_2)-\Psi_\Lambda(\lambda_1)$, and $\Psi_\Lambda(\lambda_1)+\Psi_\Lambda(\lambda_2)-\Psi_\Lambda(\lambda_1+\lambda_2)$, respectively. The jump size distributions $Z_1, Z_2, \boldsymbol{X}$ are independent and concentrated on $\mathbb{N}$, $\mathbb{N}$, and $\mathbb{N}^2$, respectively. In particular, the distribution of the jumps exclusively affecting component one is for $k \in \mathbb{N}$:

$$\begin{aligned}\mathbb{P}(Z_1 = k) = \mathbb{P}(Y_1 = k | Y_1 > 0, Y_2 = 0) &= \frac{\mathbb{P}(Y_1 = k, Y_2 = 0)}{\mathbb{P}(Y_2 = 0)} \\ &= \frac{\nu_{\boldsymbol{L}}(k, 0)}{\Psi_\Lambda(\lambda_2 + \lambda_1) - \Psi_\Lambda(\lambda_2)}.\end{aligned}$$

The distribution of Z_2 follows accordingly. The joint jump size distribution $\boldsymbol{X}$ has dependent components and for $k, l \in \mathbb{N}$ it holds:

$$\mathbb{P}(X_1 = k, X_2 = l) = \frac{\nu_{\boldsymbol{L}}(k, l)}{\Psi_\Lambda(\lambda_1) + \Psi_\Lambda(\lambda_2) - \Psi_\Lambda(\lambda_2 + \lambda_1)}.$$

Due to the concavity of the Laplace exponent (and excluding the case $\Psi_\Lambda(x) = ax$ for some $a > 0$ corresponding to independence of the processes L_t^1 and L_t^2), the denominator is indeed non-negative in both cases. ▲

Remark 4.23 (Multivariate from univariate jump size distribution)
As for the process distribution, the multivariate jump distribution can be calculated from the univariate distribution of the aggregate jumps by rescaling. More precisely, it holds similarly to Corollary 4.6:

$$\mathbb{P}(\boldsymbol{Y} = \boldsymbol{k}) = c(\boldsymbol{k})\mathbb{P}(\bar{Y} = \boldsymbol{k}), \quad \boldsymbol{k} \in \dot{\mathbb{N}}_0^d,$$

[2] If $\Lambda_t = at$ for some $a > 0$, then the components L_t^1 and L_t^2 remain independent and a.s. never jump together.

where $c(\boldsymbol{k})$ is the rescaling coefficient from Equation (4.5), which can be implemented using Algorithm 4.7, and $\bar{Y} := \mathbf{1}'\boldsymbol{Y}$ is the aggregate claim size distribution, that is

$$\mathbb{P}(\bar{Y} = k) = -\frac{(-|\boldsymbol{\lambda}|)^k}{k!} \frac{\Psi_\Lambda^{(k)}(|\boldsymbol{\lambda}|)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)}, \quad k \in \mathbb{N}^d.$$

Of course, a similar relationship holds for the Lévy measure. ▲

It was already mentioned after Theorem 4.17 that the derivatives of the Laplace exponent as well as the Lévy measure – or the jump size distribution – can be calculated analytically in the subordinator families introduced in Example 2.14; the results are presented in Table 4.6. For a numerically stable implementation for high values of $k \in \mathbb{N}$, an iterative calculation is often preferable due to similar considerations as those for the implementation of the rescaling coefficient. Consider for instance an inverse Gaussian subordinator with parameters $\beta, \eta > 0$ and with no drift. Then it holds:

$$\Psi_\Lambda^{(1)}(x) = \beta(2x + \eta^2)^{-\frac{1}{2}}; \qquad \Psi_\Lambda^{(k)} = -\frac{2k-3}{2x+\eta^2}\Psi_\Lambda^{(k-1)}(x), \quad k = 2, 3, \ldots.$$

Adding a drift $b > 0$ to the subordinator changes the Laplace exponent to $\Psi_{\tilde{\Lambda}}(x) = \Psi_\Lambda(x) + bx$ and accordingly

$$\Psi_{\tilde{\Lambda}}^{(k)}(x) = b\delta_1(k) + \Psi_\Lambda^{(k)}(x), \quad x \geq 0.$$

In particular, adding a drift increases the number of solitary jumps, that is jumps of size one in only a single component. Together, the Lévy measure can be computed using the following algorithm.

Algorithm 4.24 (Univariate Lévy measure for inverse Gaussian subordinator)
For dimension $d = 1$ in Model (M) with some intensity $\lambda > 0$ and an inverse Gaussian subordinator with drift, the Lévy measure on $(1, \ldots, k)$, $k \in \mathbb{N}$, can be implemented as follows:

$$
\begin{aligned}
&\textit{INPUT } \lambda;\quad b;\quad \beta;\quad \eta;\quad k;\\
&\textit{START } \nu := \mathbf{0} \in \mathbb{R}^k;\\
&\qquad \nu(1) := \lambda * \beta * (2 * \lambda + \eta^2)^{-1/2};\\
&\textit{FOR } j = 2, \ldots, k\\
&\qquad \nu(j) = \frac{\lambda}{j} * \nu(j-1) * \frac{j * 2 - 3}{2 * \lambda + \eta^2};\\
&\textit{END}\\
&\nu(1) = \nu(1) + b * \lambda;\\
&\textit{RETURN } \nu;
\end{aligned}
$$

For calculating the univariate jump size distribution, a final step for rescaling ν with $\Psi_{\tilde{\Lambda}}(\lambda)^{-1}$ has to be added. According to Remark 4.21, for the i-th marginal Lévy measure in a d-dimensional model with intensity $\boldsymbol{\lambda}$, the algorithm has to be called using $\lambda = \lambda_i$ and an additional entry for zero has to be added:

$$\nu(0) = \Psi_{\tilde{\Lambda}}(|\boldsymbol{\lambda}|) - \Psi_{\tilde{\Lambda}}(\lambda_i).$$

For the marginal jump size distribution, rescaling has to be done with $\Psi_{\tilde{\Lambda}}(|\boldsymbol{\lambda}|)^{-1}$. Considering another subordinator, the initialization step $\nu(1)$ and the iteration step $\nu(j)$ have to be adjusted accordingly; a summary can be found in Table 4.7. Note that for the gamma subordinator an iterative calculation is not necessary for numerical stability as the factorial terms cancel out:

$$\nu(k) = \frac{\beta}{k}\Big(1 + \frac{\eta}{\lambda}\Big)^{-k}, \quad k \in \mathbb{N}.$$

Iterative calculation is, however, useful for calculating all quantities up to a high order. Ideally, though, scaling with k^{-1} is then done as a final step after the iteration.

Furthermore, for efficiency in calculating quantities at higher dimensions, implementing the calculation directly is better than rescaling the univariate quantities in most subordinator cases. The jump size distribution can also be approximated by the empirical distribution generated in a Monte Carlo simulation of the model employing Algorithm 3.3. As mentioned previously, for an infinite activity subordinator Λ some bias – though small if choosing a fine grid for sampling the subordinator paths – towards overestimating the probabilities of larger jump sizes compared to smaller ones should be expected.

subordinator	$\nu(1)$	$\nu(j) = \nu(j-1) * ...$
Poisson $\xi > 0$	$\xi\lambda \exp(-\lambda)$	$\frac{\lambda}{j}$
compound Poisson with gamma $\xi, \beta, \eta > 0$	$\frac{\xi\beta\lambda}{\eta}\left(1+\frac{\lambda}{\eta}\right)^{-\beta-1}$	$\frac{\beta+j-1}{j}\frac{\lambda}{\eta+\lambda}$
gamma $\beta, \eta > 0$	$\frac{\lambda}{\eta+\lambda}\beta$	$\frac{k-1}{k}\frac{\lambda}{\eta+\lambda}$
inverse Gaussian $\beta, \eta > 0$	$\lambda\beta(2\lambda+\eta^2)^{-1/2}$	$\frac{2j-3}{j}\frac{\lambda}{2\lambda+\eta^2}$
stable $\alpha > 0$	$\alpha\lambda^{\alpha}$	$\frac{j-1-\alpha}{j}$

Table 4.7 Iterative computation of univariate Lévy measure in the time-changed model for different subordinators: The table summarizes the initialization $\nu(1)$ and iteration step $\nu(j)$, $j = 2, 3, \ldots$, for calculating the Lévy measure of the time-changed process in Model (M) in dimension one with intensity $\lambda > 0$. The different subordinator families have been introduced in Example 2.14.

Example 4.25 (Families of jump size distributions)
In Example 4.8, the univariate claim number distributions resulting from a gamma, inverse Gaussian and Poisson subordinator with no drift were identified. In these cases, also the jump size distributions are known distribution families, see (Klugman et al., 2004, Chapter 4.6.10, pp.103). Let in the following Y be the jump size in Model (M) with dimension $d = 1$ and intensity $\lambda > 0$. Figure 4.7 illustrates an example of the jump size distribution in each case for the same parameter setting as in Figure 4.1.

(i) Gamma subordinator - logarithmic jump size distribution:
Choosing a gamma subordinator with parameters $\beta, \eta > 0$ it holds for the jump size distribution in the univariate time-changed model according to Table 4.6:

$$\mathbb{P}(Y = k) = \frac{1}{k \log\left(1 + \frac{\lambda}{\eta}\right)} \left(\frac{\lambda}{\eta + \lambda}\right)^k, \quad k \in \mathbb{N}.$$

This is the distribution of a logarithmic random variable with parameter $\theta := \frac{\lambda}{\eta+\lambda} \in [0, 1]$, see (Johnson et al., 1992, Chapter 7, Section 1, p.285).

(ii) Inverse Gaussian subordinator - extended truncated negative binomial jumps size distribution:
According to Table 4.6, the jump size distribution resulting from an inverse Gaussian subordinator with parameters $\beta, \eta > 0$ in the univariate case is:

$$\mathbb{P}(Y = k) = \frac{1}{k!} \left(\frac{\lambda}{2\lambda + \eta^2}\right)^k \frac{\sqrt{2\lambda + \eta^2}}{\sqrt{2\lambda + \eta^2} - \eta} \prod_{j=1}^{k-1} (2j - 1), \quad k \in \mathbb{N}.$$

A random variable Z has a negative binomial distribution, if it holds:

$$\mathbb{P}(Z = k) = \binom{k + r - 1}{k} p^k (1 - p)^r, \quad k \in \mathbb{N}_0,$$

where $r > 0$ and $0 < p < 1$. The extended negative binomial distribution allows for $-1 < r < 0$, see (Klugman et al., 2004, Chapter 4.6.6, p.87). For the zero-truncation $\tilde{Z}$ it is according to Remark 4.21:

$$\mathbb{P}(\tilde{Z} = k) = \frac{\mathbb{P}(Z = k)}{1 - \mathbb{P}(Z = 0)} = \binom{k + r - 1}{k} p^k \frac{(1 - p)^r}{1 - (1 - p)^r}, \quad k \in \mathbb{N}.$$

By some calculation and observing that

$$-\frac{1}{2^k k!}\prod_{j=1}^{k-1}(2j-1) = \binom{k+\frac{3}{2}}{k},$$

it follows that the jump size Y is extended negative binomial distributed with parameters

$$r := -\frac{1}{2}, \quad p := \frac{2\lambda}{2\lambda+\eta^2}.$$

Note that according to Remark 4.21 one could also consider the un-truncated negative binomial distribution if the jump intensity is adjusted from $\Psi_\Lambda(\lambda)$ to $\frac{\Psi_\Lambda(\lambda)}{1-(1-p)^r}$.

(iii) Poisson process - truncated Poisson jump size distribution:
From Table 4.6 it follows for the jump size distribution corresponding to a time-change with a Poisson process with intensity $\xi > 0$:

$$\mathbb{P}(Y=k) = \frac{\lambda^k}{k!}\frac{\exp\{-\lambda\}}{1-\exp\{-\lambda\}}, \quad k \in \mathbb{N}.$$

Note that the jump size distribution does not depend on the subordinator parameter at all. The subordinator does, however, affect the jump intensity $\Psi_\Lambda(\lambda)$. For $Z \sim \text{Poi}(\lambda)$ it follows that

$$\mathbb{P}(Y=k) = \frac{\mathbb{P}(Z=k)}{1-\mathbb{P}(Z=0)}, \quad k \in \mathbb{N},$$

hence, Y is distributed according to a zero-truncated Poisson distribution with intensity λ, i.e. it follows the distribution of the underlying Poisson process in Model (M) in $t=1$ except for the truncation of zeros.

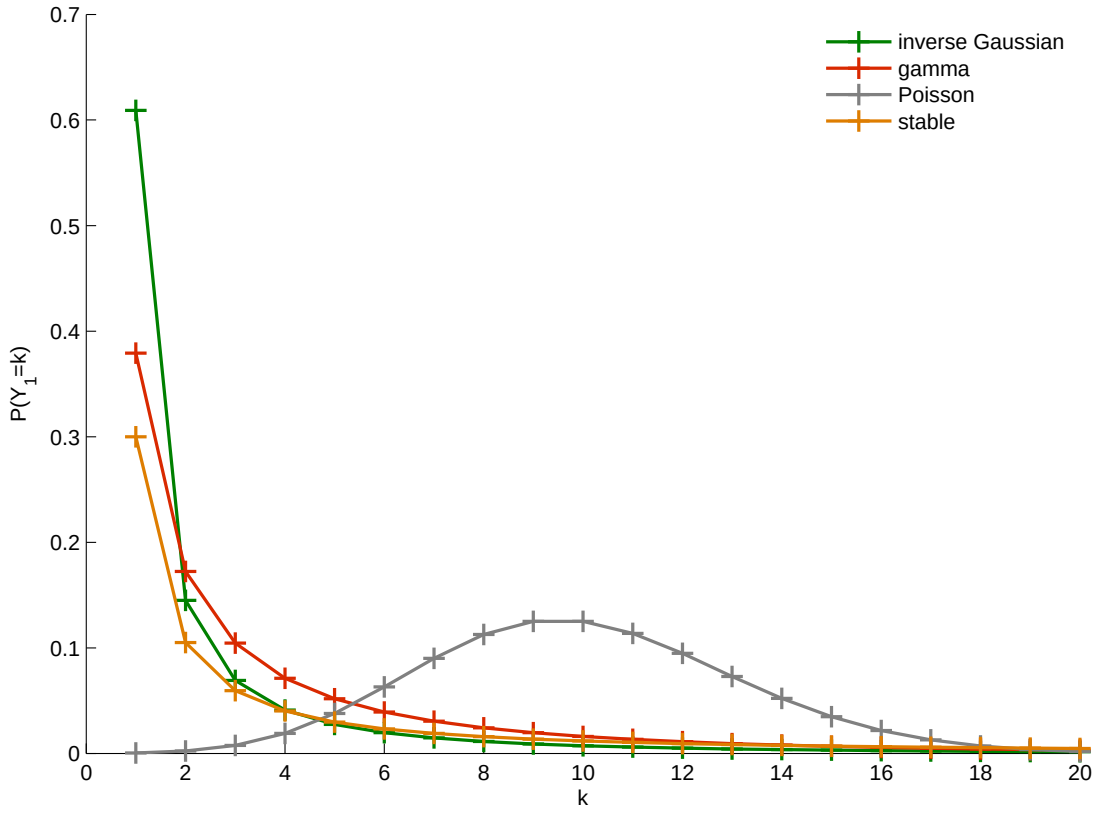


Figure 4.7 Univariate jump size distributions for different subordinators: The figure shows the univariate jump size distributions corresponding to the setting in Figure 4.1 ($d = 1$, $\lambda_1 = 10$) for an inverse Gaussian ($b = 0$, $\beta = \eta = 1$), gamma ($b = 0$, $\beta = \eta = 1$), Poisson ($b = 0$, $\xi = 1$), and stable subordinator ($b = 0$, $\alpha = 0.3$). The parameters of the inverse Gaussian, gamma, and Poisson subordinator have been set such that mean and variance are both one. For the stable subordinator, mean and variance are infinite. The point masses of the distributions are marked with a cross; the connecting lines are provided only to aid visualization.

Tails and moments of the cluster size distribution

Tails of the Lévy measure and the cluster size distribution can – similarly to the process distribution – be calculated by the residual of a Taylor expansion of Ψ_Λ. The following corollary gives the result for the Lévy measure.

Corollary 4.26 (Tails of the Lévy measure of the claim number process)
For the tails of the Lévy measure of the time-changed process $\boldsymbol{L}$ and the aggregate process $\bar{L}$ it holds:

$$\sum_{k=n+1}^{\infty} \sum_{|\boldsymbol{k}|=k} \nu_{\boldsymbol{L}}(\boldsymbol{k}) = \sum_{k=n+1}^{\infty} \nu_{\bar{L}}(k) = \Psi_\Lambda(|\boldsymbol{\lambda}|) - \sum_{k=1}^{n} \nu_{\bar{L}}(k) = -R^n \Psi_\Lambda(0; |\boldsymbol{\lambda}|), \quad n \in \dot{\mathbb{N}}_0,$$

where $R^n \Psi_\Lambda(0; |\boldsymbol{\lambda}|)$ is defined as in Theorem 4.9 as residual of the Taylor expansion of Ψ_Λ in $|\boldsymbol{\lambda}|$ evaluated in zero:

$$R^n \Psi_\Lambda(0; |\boldsymbol{\lambda}|) = \frac{(-1)^{n+1}}{n!} \int_0^{|\boldsymbol{\lambda}|} x^n \Psi_\Lambda^{(n+1)}(x) \, \mathrm{d}x = o(|\boldsymbol{\lambda}|^n).$$

Proof. The Taylor expansion of Ψ_Λ has been given in the proof of Theorem 4.17. The remainder of the proof follows in the same fashion as the proof of Theorem 4.9. □

Following the above corollary, the lower tail of the multivariate Lévy measure can, of course, be calculated as well considering the total mass $\Psi_\Lambda(|\boldsymbol{\lambda}|)$, that is

$$\sum_{k=1}^{n} \sum_{|\boldsymbol{k}|=k} \nu_{\boldsymbol{L}}(\boldsymbol{k}) = \Psi_\Lambda(|\boldsymbol{\lambda}|) - \sum_{k=n+1}^{\infty} \sum_{|\boldsymbol{k}|=k} \nu_{\boldsymbol{L}}(\boldsymbol{k}) = \Psi_\Lambda(|\boldsymbol{\lambda}|) + R^n \Psi_\Lambda(0; |\boldsymbol{\lambda}|).$$

The corollary also states the results for the univariate aggregate process $\bar{L}$, which can be transferred to any univariate model specification; for the marginal measures ν_{L^i}, it follows with Remark 4.21:

$$-R^n \Psi_\Lambda(0; |\lambda|) = \sum_{k=n+1}^{\infty} \nu_{L^i}(k) = \Psi_\Lambda(\lambda_i) - \sum_{k=1}^{n} \nu_{L^i}(k) = \Psi_\Lambda(|\boldsymbol{\lambda}|) - \sum_{k=0}^{n} \nu_{L^i}(k).$$

For a stable subordinator, the computation of the residual can be done analytically with the results in Table 4.6:

$$\begin{aligned} -R^n\Psi_\Lambda(0;x) &= \frac{(-1)^n}{n!}\int_0^x y^n(-1)^n y^{\alpha-n-1}\alpha\prod_{j=1}^n(j-\alpha)\,\mathrm{d}y \\ &= \frac{\alpha}{n!}\prod_{j=1}^n(j-\alpha)\int_0^x y^{\alpha-1}\,\mathrm{d}y = \frac{x^\alpha}{n!}\prod_{j=1}^n(j-\alpha). \end{aligned}$$

In general, the residual can be approximated numerically as discussed after Theorem 4.9 for the residual of φ_{Λ_t}.

The (mixed) moments of the jump size distribution can be derived from its Laplace transform. Heavy-tailedness is, of course, inherited from the process distribution which again inherits it from the subordinator.

Theorem 4.27 (Moments and heavy-tailedness of the jumps size distribution)
If existent, mean vector and covariance matrix of the jump size distribution $\boldsymbol{Y}$ of the claim number process $\boldsymbol{L}$ in Model (M) are:

$$\begin{aligned} \mathbb{E}[\boldsymbol{Y}] &= \frac{\mathbb{E}[\Lambda_1]}{\Psi_\Lambda(|\boldsymbol{\lambda}|)}\boldsymbol{\lambda} = \frac{1}{\lambda_M}\,\mathbb{E}[\boldsymbol{L}_1], \\ \mathbb{C}\mathrm{ov}[\boldsymbol{Y},\boldsymbol{Y}] &= \Big(\frac{\mathbb{V}\mathrm{ar}[\Lambda_1]}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} - \frac{\mathbb{E}[\Lambda_1]^2}{\Psi_\Lambda(|\boldsymbol{\lambda}|)^2}\Big)\boldsymbol{\lambda}\boldsymbol{\lambda}' + \frac{\mathbb{E}[\Lambda_1]}{\Psi_\Lambda(|\boldsymbol{\lambda}|)}diag(\boldsymbol{\lambda}) \\ &= \frac{1}{\lambda_M}\,\mathbb{C}\mathrm{ov}[\boldsymbol{L}_1,\boldsymbol{L}_1] - \frac{1}{\lambda_M^2}\,\mathbb{E}[\boldsymbol{L}_1]\,\mathbb{E}[\boldsymbol{L}_1]'. \end{aligned}$$

The marginal jump size distributions are heavy-tailed iff the distribution of the subordinator is heavy-tailed.

Proof. First and second moment of $\boldsymbol{Y}$ can be derived from evaluating derivatives of the Laplace transform $\varphi_{\boldsymbol{Y}}$ in zero, that is:

$$\mathbb{E}[Y_i] = \varphi_{\boldsymbol{Y}}^{(\boldsymbol{e}_i)}(\boldsymbol{0}), \quad \mathbb{E}[Y_iY_j] = -\varphi_{\boldsymbol{Y}}^{(\boldsymbol{e}_i+\boldsymbol{e}_j)}(\boldsymbol{0}), \quad i,j=1,\dots,d.$$

For the univariate case, this was mentioned in Theorem 2.2; the multivariate result follows, for instance, from (Johnson et al., 1997, Chapter 2.2, pp.3). The derivatives can be found as:

$$\begin{aligned}
\varphi_{\boldsymbol{Y}}^{(\boldsymbol{e}_i)}(\boldsymbol{x}) &= -\frac{\Psi_\Lambda^{(1)}\Big(|\boldsymbol{\lambda}| - \sum_{k=1}^d \lambda_k \exp\{-x_k\}\Big)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} \lambda_i \exp\{-x_i\} \\
\varphi_{\boldsymbol{Y}}^{(\boldsymbol{e}_i+\boldsymbol{e}_j)}(\boldsymbol{x}) &= -\frac{\Psi_\Lambda^{(2)}\Big(|\boldsymbol{\lambda}| - \sum_{k=1}^d \lambda_k \exp\{-x_k\}\Big)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} \lambda_i\lambda_j \exp\{-x_i - x_j\} \\
&\quad + \mathbb{1}_{\{i\}}(j) \frac{\Psi_\Lambda^{(1)}\Big(|\boldsymbol{\lambda}| - \sum_{k=1}^d \lambda_k \exp\{-x_k\}\Big)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)} \lambda_i \exp\{-x_i\}.
\end{aligned}$$

Evaluating these expressions in zero, using Remark 4.14 about the subordinator moments and Theorem 4.15 about the moments of the claim number process, and calculating the (co-)variance as $\mathbb{C}\mathrm{ov}[Y_i, Y_j] = \mathbb{E}[Y_iY_j] - \mathbb{E}[Y_i]\,\mathbb{E}[Y_j]$ gives the results formulated in the proposition. The statement on heavy-tailedness is shown similarly to the corresponding statement about the process distribution given in Proposition 4.13. By definition, the distribution of Y_i is heavy-tailed iff $\varphi_{Y_i}(-x) = \infty$ for all $x > 0$. According to Theorem 2.1 and Corollary 4.19 it is

$$\varphi_{Y_i}(-x) = \varphi_{\boldsymbol{Y}}(-x\boldsymbol{e}_i) = 1 - \frac{\Psi_\Lambda(\lambda_i(1 - \exp\{x\}))}{\Psi_\Lambda(|\boldsymbol{\lambda}|)}.$$

Hence the following equivalence holds:

$$\begin{aligned}
\varphi_{Y_i}(-x) = \infty \quad \forall x > 0 \quad &\Leftrightarrow \quad \Psi_\Lambda(-x) = -\infty \quad \forall x > 0 \\
&\Leftrightarrow \quad \exp\{-\Psi_\Lambda(-x)\} = \infty \quad \forall x > 0,
\end{aligned}$$

where the last statement is the definition of heavy-tailedness of Λ. □

Dependence structure of the cluster size distribution

In Figure 4.8, the bivariate copulas of the jump size distributions in the setting of Figure 4.5 for an inverse Gaussian, gamma, and compound Poisson process are illustrated. While the copulas of the process distributions looked rather similar in Figure 4.5, the copulas of the jump sizes do look different. The scatter plot of the copula in the inverse Gaussian case has many fewer distinct observation points with much greater intensity than the one in the compound Poisson case; the gamma case ranges somewhere in

between. This is not surprising considering the univariate distributions in Figure 4.7, where in the inverse Gaussian case more mass is allocated at small jump sizes, while in the Poisson case jump sizes tend to be larger. Table 4.8 summarizes mean, standard deviation, and correlation of the jump sizes as well as the jump intensity in the three cases: for the inverse Gaussian subordinator, the jump intensity is by far the highest of the three, while expected jump size and standard deviation are low in comparison. For the compound Poisson subordinator, the relation is exactly the other way round, and for the gamma subordinator the values range between the other ones. Consequently, while expectation and variance of the claim number distribution at any point in time are the same in all three cases according to the selected parameters, the evolution of the claim number process over time is very different: in the inverse Gaussian case, many small claim clusters are accumulated whereas in the compound Poisson case fewer clusters of larger size occur. The dependence structure, however, may not be as different as a first glance at the pictures suggests. The correlation of the jump sizes, in any case, are very similar in the three examples. However, the discreteness of the distributions together with the differences in the marginals makes comparing the copulas difficult, see the discussion at the end of Section 4.1.

subordinator	λ^M	$\mathbb{E}[\boldsymbol{Y}]$	$\sqrt{\mathbb{V}\mathrm{ar}[\boldsymbol{Y}]}$	$\mathbb{C}\mathrm{or}[Y_1, Y_2]$
inverse Gaussian $b = 0$, $\beta = \eta = 1.5$	25.90	$(2.90, 3.86)'$	$(9.54, 12.67)'$	0.7530
gamma $b = 0$, $\beta = \eta = 2.25$	9.83	$(7.63, 10.18)'$	$(14.28, 18.95)'$	0.7535
compound Poisson + gamma $b = 0$, $\xi = \frac{10}{3}$, $\beta = 1.5$, $\eta = 5$	3.32	$(22.61, 30.14)'$	$(18.96, 25.07)'$	0.7560

Table 4.8 Characteristics of jump sizes for different subordinators: The table summarizes for the subordinator and parameter choices of Figure 4.8 mean, standard deviation, and correlation of the jump size distribution as well as jump intensity of the time-changed process.

Distribution of the claim arrival times

The Poisson cluster process representation of the claim number process can also be used to get a better understanding of the claim inter-arrival times τ_{ij}. The distribution of the arrival times was already discussed in Section 4.1. For the inter-arrival times it was mentioned in Section 3.1 that they are – as for any non-trivial inhomogeneous Poisson or Cox process, see Section 2.2 – no longer iid. In addition, the distribution has a singularity in zero due to the possibility of simultaneous jumps. These observations can be made explicit by exploiting the Poisson cluster representation of the process.

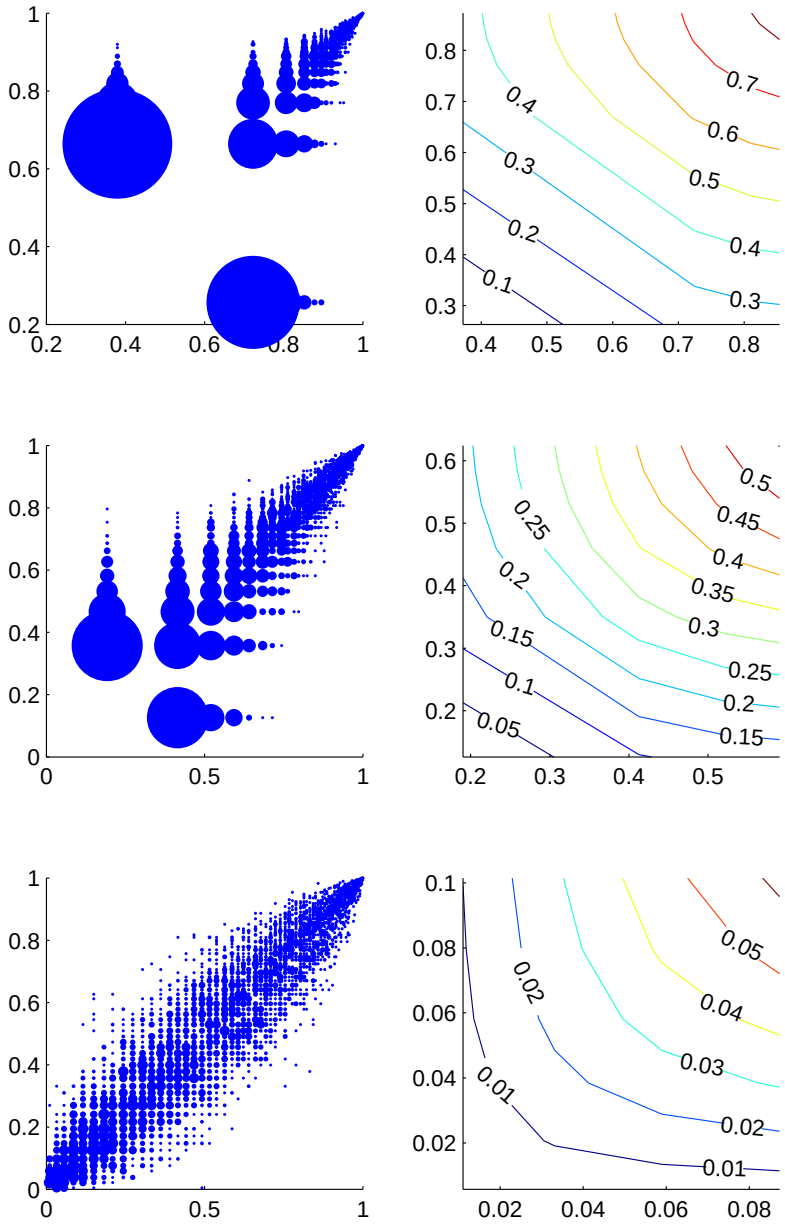


Figure 4.8 Bivariate copula of the jump size distribution for different subordinators: Corresponding to the setting in Figure 4.5 ($d = 2$, $\boldsymbol{\lambda} = (75, 100)'$), the figure shows scatter plots (left) and contour plots (right) of the jump size copulas stemming from an inverse Gaussian (top; $b = 0$, $\beta = \eta = 1.5$), gamma (middle; $b = 0$, $\beta = \eta = 2.25$), and compound Poisson subordinator with gamma distributed jumps ($b = 0$, $\xi = 10/3$, $\beta = 1.5$, $\eta = 5$). Time-normalization holds and the correlation of the time-changed process is around 97.5%. The scatter plots are generated from 5000 samples and the marker size corresponds to the number of respective observations.

The arrival time of the first claim in each component can be concluded from Corollary 4.10 or from observing that it coincides with the first arrival time of the Poisson process counting the cluster arrivals in the respective univariate model:

$$\tau_{i1} \sim \mathrm{Exp}(\Psi_\Lambda(\lambda_i)), \quad i = 1, \dots, d.$$

From there on, successive claims have a positive probability of arriving in the same cluster and hence having an inter-arrival time of zero. Let $\tilde{Y}_i$ be the zero-deflated cluster size distribution of component i. If for some $j = 2, 3, ...$ the inter-arrival time $\Delta\tau_{ij}$ is zero, the claims number $j-1$ and j need to have arrived in the same cluster. Due to the zero-deflation, at least one claim arrives in each cluster, hence, a maximum of $j-1$ clusters can have arrived. Together, it follows from the law of total probability for $j = 2, 3, ...$:

$$\begin{aligned}
\mathbb{P}(\Delta\tau_{ij} = 0) &= \sum_{n=1}^{j-1} \mathbb{P}(\text{claim no. } j \text{ and } j-1 \text{ of component } i \text{ arrive in cluster no. } n) \\
&= \mathbb{P}(\tilde{Y}_{i1} \geq j) + \sum_{n=1}^{j-2} \mathbb{P}(\tilde{Y}_{i1} + \ldots + \tilde{Y}_{in} < j-1, \tilde{Y}_{i1} + \cdots + \tilde{Y}_{i(n+1)} \geq j) \\
&= \mathbb{P}(\tilde{Y}_i \geq j) + \sum_{n=1}^{j-2} \sum_{k=n}^{j-2} \mathbb{P}(\tilde{Y}_{i1} + \cdots + \tilde{Y}_{in} = k, \tilde{Y}_{i(n+1)} \geq j-k) \\
&= \mathbb{P}(\tilde{Y}_i \geq j) + \sum_{n=1}^{j-2} \sum_{k=n}^{j-2} \mathbb{P}(\tilde{Y}_{i1} + \cdots + \tilde{Y}_{in} = k) \mathbb{P}(\tilde{Y}_i \geq j-k).
\end{aligned}$$

As this formula depends on j, the inter-arrival times are not identically distributed. Furthermore, it is

$$\mathbb{P}(\Delta\tau_{i3} = 0, \Delta\tau_{i2} = 0) = \mathbb{P}(\tilde{Y}_i \geq 3),$$

which is unequal to:

$$\mathbb{P}(\Delta\tau_{i3} = 0)\mathbb{P}(\Delta\tau_{i2} = 0) = (\mathbb{P}(\tilde{Y}_i \geq 3) + \mathbb{P}(\tilde{Y}_i = 1)\mathbb{P}(\tilde{Y}_i \geq 2))\mathbb{P}(\tilde{Y}_i \geq 2).$$

This confirms that the inter-arrival times are indeed no longer independent. Except for the point mass at zero, however, the distribution is absolutely continuous as, given that

successive claims do not arrive in the same cluster, they arrive in successive clusters – which have exponentially distributed inter-arrival times:

$$\begin{aligned}\mathbb{P}(\Delta\tau_{ij} > t) &= \mathbb{P}(\Delta\tau_{ij} > t|\Delta\tau_{ij} > 0)\mathbb{P}(\Delta\tau_{ij} > 0) + \mathbb{P}(\Delta\tau_{ij} > t|\Delta\tau_{ij} = 0)\mathbb{P}(\Delta\tau_{ij} = 0)\\ &= \mathbb{P}(\Delta\tau_{ij} > t|\Delta\tau_{ij} > 0)\mathbb{P}(\Delta\tau_{ij} > 0)\\ &= \exp\{-\Psi_\Lambda(\lambda_i)t\}(1 - \mathbb{P}(\Delta\tau_{ij} = 0)), \quad t > 0.\end{aligned}$$

As a measure of extremal dependence, the probability of joint early claim arrivals is given in the following proposition.

Corollary 4.28 (Joint early claim arrivals)
For the d-dimensional claim number process $\boldsymbol{L}$ in Model (M), consider two different components $1 \leq i, k \leq d$, $i \neq k$, and let ν_j, ν_k be either the respective marginal Lévy measures or the Lévy measures in corresponding univariate models; accordingly, let ν_{ik} be either the bivariate marginal Lévy measure or the Lévy measure in a corresponding bivariate model. Then the probability of the joint early occurrence of the j-th claim to portfolio i and the l-th claim to portfolio k for some $j, l \in \mathbb{N}$ is given as:

$$\begin{aligned}\lim_{t\downarrow 0}\mathbb{P}(\tau_{ij} \leq t|\tau_{kl} \leq t) =& \frac{\sum_{m_1=j}^{\infty}\sum_{m_2=l}^{\infty}\nu_{ik}(m_1, m_2)}{\sum_{m=l}^{\infty}\nu_{ik}(m)}\\ =& 1 - \frac{1}{\Psi_\Lambda(\lambda_k) - \sum_{m=1}^{l-1}\nu_k(m)}\Big(\Psi_\Lambda(\lambda_i + \lambda_k) - \Psi_\Lambda(\lambda_i)\\ &+ \sum_{m=1}^{j-1}\nu_i(m) - \sum_{m_1=0}^{j-1}\sum_{\substack{m_2=0\\ m_1+m_2>0}}^{l-1}\nu_{ik}(m_1, m_2)\Big).\end{aligned}$$

For a single component and claim numbers $j > l$ it is

$$\lim_{t\downarrow 0}\mathbb{P}(\tau_{ij} \leq t|\tau_{il} \leq t) = 1 - \frac{\sum_{m=l}^{j-1}\nu_i(m)}{\Psi_\Lambda(\lambda_i) - \sum_{m=1}^{l-1}\nu_i(m)}.$$

Proof. Consider the Poisson cluster process representation of $\boldsymbol{L}$ where the claim clusters arrive according to a Poisson process M with intensity $\lambda^M = \Psi_\Lambda(|\boldsymbol{\lambda}|)$ and with d-dimensional random jump size $\boldsymbol{Y}$. Then it holds by definition of the conditional probability and the law of total probability:

$$\begin{aligned}\mathbb{P}(\tau_{ij} \leq t | \tau_{kl} \leq t) &= \frac{\mathbb{P}(\tau_{ij} \leq t, \tau_{kl} \leq t)}{\mathbb{P}(\tau_{kl} \leq t)} = \frac{\mathbb{P}(L_t^i \geq j, L_t^k \geq l)}{\mathbb{P}(L_t^k \geq l)} \\ &= \frac{\sum_{m=0}^{\infty} \mathbb{P}(M_t = m)\mathbb{P}(\sum_{r=1}^{m} Y_{ir} \geq j, \sum_{r=1}^{m} Y_{kr} \geq l)}{\sum_{m=0}^{\infty} \mathbb{P}(M_t = m)\mathbb{P}(\sum_{r=1}^{m} Y_{kr} \geq l)}.\end{aligned}$$

According to Theorem 2.4, the limiting behaviour of $\mathbb{P}(M_t = m)$ for $t \downarrow 0$ is $\lambda^M t + o(t)$ for $m = 1$ and $o(t)$ for $m \geq 2$, thus it follows:

$$\begin{aligned}\lim_{t \downarrow 0} \mathbb{P}(\tau_{ij} \leq t | \tau_{kl} \leq t) &= \lim_{t \downarrow 0} \frac{\Psi_\Lambda(|\boldsymbol{\lambda}|)t\mathbb{P}(Y_{i1} \geq j, Y_{k1} \geq l) + o(t)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)t\mathbb{P}(Y_{k1} \geq l) + o(t)} \\ &= \lim_{t \downarrow 0} \frac{\Psi_\Lambda(|\boldsymbol{\lambda}|)\mathbb{P}(Y_{i1} \geq j, Y_{k1} \geq l) + o(1)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)\mathbb{P}(Y_{k1} \geq l) + o(1)} \\ &= \frac{\Psi_\Lambda(|\boldsymbol{\lambda}|)\mathbb{P}(Y_{i1} \geq j, Y_{k1} \geq l)}{\Psi_\Lambda(|\boldsymbol{\lambda}|)\mathbb{P}(Y_{k1} \geq l)}.\end{aligned}$$

Omitting the index 1 in the jumps size notation and applying the principle of inclusion and exclusion, it holds:

$$\begin{aligned}\mathbb{P}(Y_i \geq j, Y_k \geq l) =& 1 - \mathbb{P}(Y_i < j) - \mathbb{P}(Y_k < l) + \mathbb{P}(Y_i < j, Y_k < l) \\ =& 1 - \sum_{m=0}^{j-1} \mathbb{P}(Y_i = m) - \sum_{m=0}^{l-1} \mathbb{P}(Y_k = m) \\ &+ \sum_{m_1=0}^{j-1} \sum_{m_2=0}^{l-1} \mathbb{P}(Y_i = m_1, Y_k = m_2).\end{aligned}$$

To get a representation which does not depend on whether the marginals or a respective lower-dimensional version of the model is considered, cf. Remark 4.21, we switch to the Lévy measure and treat the zero case separately from the rest of the sums. Then we get for the nominator:

$$\begin{aligned}\Psi_\Lambda(|\boldsymbol{\lambda}|)\mathbb{P}(Y_i \geq j, Y_k \geq l) =& \Psi_\Lambda(|\boldsymbol{\lambda}|)(1 - \mathbb{P}(Y_i = 0) - \mathbb{P}(Y_k = 0) + \mathbb{P}(Y_i = 0, Y_k = 0)) \\ & - \sum_{m=1}^{j-1} \nu_i(m) - \sum_{m=1}^{l-1} \nu_k(m) + \sum_{m_1=0}^{j-1} \sum_{\substack{m_2=0 \\ m_1+m_2>0}}^{l-1} \nu_{ik}(m_1, m_2) \\ =& \Psi_\Lambda(\lambda_i) + \Psi_\Lambda(\lambda_k) - \Psi_\Lambda(\lambda_i + \lambda_k) - \sum_{m=1}^{j-1} \nu_i(m) \\ & - \sum_{m=1}^{l-1} \nu_k(m) + \sum_{m_1=0}^{j-1} \sum_{\substack{m_2=0 \\ m_1+m_2>0}}^{l-1} \nu_{ik}(m_1, m_2).\end{aligned}$$

For the denominator it holds accordingly:

$$\Psi_\Lambda(|\boldsymbol{\lambda}|)\mathbb{P}(Y_k \geq l) = \Psi_\Lambda(|\boldsymbol{\lambda}|)(1 - \mathbb{P}(Y_k < l)) = \Psi_\Lambda(\lambda_k) - \sum_{m=1}^{l-1} \nu_k(m).$$

Together this establishes the claim on the arrival times in two different components. For the claim numbers $j > l$ of the same component L^i, it is

$$\mathbb{P}(\tau_{ij} \leq t | \tau_{il} \leq t) = \frac{\mathbb{P}(L_t^i \geq j, L_t^i \geq l)}{\mathbb{P}(L_t^i \geq l)} = \frac{\mathbb{P}(L_t^i \geq j)}{\mathbb{P}(L_t^i \geq l)},$$

thus the claim follows from the previous considerations. □

The probability of joint early claim arrivals for a single component is indeed a number in $[0, 1]$ as it can be re-written:

$$\lim_{t \downarrow 0} \mathbb{P}(\tau_{ij} \leq t | \tau_{il} \leq t) = 1 - \frac{\sum_{m=l}^{j-1} \nu_i(m)}{\sum_{m=l}^{\infty} \nu_i(m)}.$$

Then the denominator has positive summation terms from l to ∞, while in the nominator the summation stops at $j - 1$.

As mentioned in the proof, the probability of joint early claim arrivals corresponds to the claim number process as

$$\lim_{t\downarrow 0} \mathbb{P}(\tau_{ij} \leq t | \tau_{kl} \leq t) = \lim_{t\downarrow 0} \mathbb{P}(L_t^i \geq j | L_t^k \geq l).$$

Hence, the quantity is equally a measures of extremal dependence of the claim number processes. It is for $i \neq k$ closely related to the lower tail dependence coefficient of the claim arrival times, which is according to Section 2.1:

$$LTD(\tau_{ij}, \tau_{kl}) = \lim_{u\downarrow 0} \mathbb{P}(\tau_{ij} \leq F_{\tau_{ij}}^{-1}(u) | \tau_{kl} \leq F_{\tau_{kl}}^{-1}(u)).$$

If $j = l$ and $\lambda_i = \lambda_k$, the two quantities coincide as the marginal distributions and their inverse are the same. Then it holds for all $j \in \mathbb{N}$:

$$LTD(\tau_{ij}, \tau_{kj}) = 2 - \frac{\Psi_\Lambda(2\lambda_i) - \sum_{\substack{m_1,m_2=0 \\ m_1+m_2>0}}^{j-1} \nu_{ik}(m_1, m_2)}{\Psi_\Lambda(\lambda_i) - \sum_{m=1}^{j-1} \nu_i(m)}.$$

In particular, for $j = 1$ and $\lambda_i = 1$ the lower tail dependence coefficient of the Lévy-frailty model is recovered, cf. Section 3.3:

$$LTD(\tau_{i1}, \tau_{k1}) = 2 - \frac{\Psi_\Lambda(2)}{\Psi_\Lambda(1)}.$$

For $i \neq k$ and arbitrary λ_i, λ_k, it is

$$\lim_{t\downarrow 0} \mathbb{P}(\tau_{i1} \leq t | \tau_{k1} \leq t) = 1 - \frac{\Psi_\Lambda(\lambda_i + \lambda_k) - \Psi_\Lambda(\lambda_i)}{\Psi_\Lambda(\lambda_k)}.$$

Remember that $\Psi_\Lambda(\lambda_i)$, $\Psi_\Lambda(\lambda_k)$, and $\Psi_\Lambda(\lambda_i+\lambda_k)$ are the intensities of non-zero jumps in L^i, L^k, and $(L_t^i, L_t^k)'$, respectively, and $\Psi_\Lambda(\lambda_i + \lambda_k) - \Psi_\Lambda(\lambda_i)$ is the intensity of non-zero jumps only in k, not i. Hence, if the processes rarely jump together – i.e. dependence is small – then $\Psi_\Lambda(\lambda_i + \lambda_k) \approx \Psi_\Lambda(\lambda_i) + \Psi_\Lambda(\lambda_k)$ and the probability above is close to zero. If however the processes mostly jump together – i.e. dependence is strong - then $\Psi_\Lambda(\lambda_i + \lambda_k) \approx \Psi_\Lambda(\lambda_i) \approx \Psi_\Lambda(\lambda_k)$ and the probability is close to one.

Alternative calculation of the claim number process distribution

We conclude this section by taking another look at the distribution of the claim number process $\boldsymbol{L}$. In Section 4.1 different ways of calculating this distribution were discussed. Numerically, the distribution can be approximated by the empirical distribution generated in a Monte Carlo simulation of the time-changed model using Algorithm 3.3; alternatively, subordinator paths can be sampled to compute a Monte Carlo estimate using Equation (4.3). However, often neither of these approaches is fast enough, particularly concerning estimation methods directly based on the distribution function as will be discussed in Chapter 5. Analytically, the distribution can be calculated using derivatives of the Laplace transform of the directing subordinator. The crucial point is not computation in high dimension d – multivariate probabilities can be derived from univariate ones by rescaling – but calculating probabilities far in the tail which corresponds to evaluating high order derivatives of the Laplace transform. In some cases the derivatives can be calculated in closed form (see Example 4.8), although the derivations were lengthy in most cases and not straightforwardly implemented in a robust way. The derivatives of the Laplace transform can also be calculated by means of Bell polynomials from the derivatives of the Laplace exponent which are readily available for the subordinators introduced in Example 2.14. However, evaluating high order Bell polynomials is numerically challenging as well.

In addition to Lévy subordinators, the given methods for calculating the process distribution at one point in time are also available for more general directing processes in Model (M). For the numerical methods, all that is necessary is a way of sampling the directing process. For the analytical methods, the applicability depends on the availability of derivatives of the Laplace transform or the cumulant generating function. In the following, an efficient and robust recursive evaluation method is introduced that makes particular use of the Lévy nature of the process and therefore can not be applied to other directing processes.

The time-changed process $\boldsymbol{L}$ only has a representation as a Poisson cluster process in the Lévy case. A classical approach in the actuarial literature for evaluating compound distributions is Panjer's recursion scheme. This method was first introduced in Panjer (1981) and has since been much discussed. While it generally delivers an approximation of the compound distribution, it provides a precise solution if the primary distribution is either Poisson, binomial or negative binomial, and if in addition the secondary distribution has lattice support – conditions which are both fulfilled by a Poisson cluster process. The results for the univariate case are given in the following corollary; the multivariate quantities follow from rescaling according to Corollary 4.6.

Corollary 4.29 (Panjer's recursion for the distribution of the claim number process)

Let Y be the jump size in Model (M) with dimension $d = 1$ and intensity $\lambda > 0$. Then the distribution of the process L at some point in time $t > 0$ can be computed recursively using the following relation:

$$\mathbb{P}(L_t = k) = \frac{t\Psi_\Lambda(\lambda)}{k} \sum_{j=1}^{k} j\mathbb{P}(Y = j)\mathbb{P}(L_t = k - j), \quad k \in \mathbb{N},$$

together with the initial value $\mathbb{P}(L_t = 0) = \exp\{-t\Psi_\Lambda(\lambda)\}$.

Proof. According to (Mikosch, 2009, Chapter 3.3, Theorem 3.3.9, p.122), given any primary distribution $\tilde{N}$ fulfilling for some $a, b \in \mathbb{R}$ the relation

$$\mathbb{P}(\tilde{N} = k) = \Big(a + \frac{b}{n}\Big)\mathbb{P}(\tilde{N} = k - 1), \quad k \in \mathbb{N},$$

and a secondary distribution X with values in $\mathbb{N}_0$, it holds for the compound distribution $S := \sum_{j=1}^{\tilde{N}} X_j$:

$$\mathbb{P}(S = k) = \frac{1}{1 - a\mathbb{P}(X = 0)} \sum_{j=1}^{k} \Big(a + \frac{bj}{k}\Big)\mathbb{P}(X = j)\mathbb{P}(S = k - j), \quad k \in \mathbb{N}.$$

For a Poisson distribution it is $a = 0$ and b equals the intensity. Hence, the recursion presented in the corollary follows from the Poisson cluster process representation of L_t from Corollary 4.19 with Poisson process M_t with intensity $\lambda^M = \Psi_\Lambda(\lambda)$ and jump size distribution Y concentrated on $\mathbb{N}$. Furthermore, given $\mathbb{P}(Y = 0) = 0$, the starting value for the recursion is found to be:

$$\mathbb{P}(L_t = 0) = \mathbb{P}(M_t = 0) = \exp\{-t\Psi_\Lambda(\lambda)\}.$$

□

Multivariate versions of Panjer's recursion are available (see for instance Sundt (1999)); however, they are rather complicated and therefore often of less practical use in applications. For the model here, a multivariate version is unnecessary, as one can deduce the multivariate distribution from the univariate case, which is a big advantage. As discussed in Section 4.1, the evaluation time for Bell polynomials – approximately the runtime for a single evaluation point of the process distribution – in MATLAB® on a

standard computer (2.4 GHz Intel Core 2 Duo processor, 4.00 GB RAM) is around 1.5 seconds for order $k = 57$, while for higher orders the results are unreliable. For comparison, in the bivariate setting of Figures 4.3 and 4.4, the evaluation time for the full distribution at values $\{1, \ldots, 150\} \times \{1, \ldots, 200\}$ using Panjer's recursion was only 0.05 and 0.01 seconds, respectively.

At the beginning of Section 4.1, an additional method for calculating the process distribution was mentioned: inverting the Laplace transform or the characteristic function. This method even works in case of a more general time-change process and is also very fast. It delivers, however, a numerical approximation while Panjer's recursion provides precise results in the given setting. For more about the advantages and disadvantages of using Panjer's recursion or transform inversion methods, see Embrechts and Frei (2009).

4.3 Large portfolio approximation

In this section, the convergence of the aggregate claim number process in Model (M) is studied for increasing dimension $d \to \infty$. It will be shown that the process, normalized by the dimension, converges to – and can hence be approximated by – the directing Lévy subordinator scaled with the average intensity. It was discussed in Section 4.1 that in calculating the distribution of the process, the dimension of the model only enters into the evaluation point of the derivatives of the Laplace transform and the scaling factor. The most effort goes into calculating the derivatives of high order, that is evaluating the tail of the distribution. However, an increase in dimension leads to a shift of the mass of the distribution to higher values and, hence, it becomes necessary to calculate the distribution further in the tail. For this reason, an approximation may be useful.

Consider the set-up of Model (M) for increasing dimension $d \to \infty$. For this purpose, let $N^1 = \{N_t^1\}_{t\geq 0}, N^2 = \{N_t^2\}_{t\geq 0}, \dots$ be a sequence of independent univariate Poisson processes with intensities $\lambda_1, \lambda_2, \dots > 0$. The average of the intensities is assumed to converge to some constant $\bar{\lambda} > 0$, that is:

$$\lim_{d\to\infty} \frac{1}{d} \sum_{i=1}^{d} \lambda_i = \bar{\lambda}. \tag{A}$$

Furthermore, let $\Lambda = \{\Lambda_t\}_{t\geq 0}$ be an independent Lévy subordinator and consider the sequence of time-changed processes $L^i := \{N^i_{\Lambda_t}\}_{t\geq 0}$. The normalized aggregate time-changed process up to component $d \in \mathbb{N}$ is denoted by $d^{-1}\bar{L}^d$ and the average aggregate intensity by $d^{-1}\bar{\lambda}_d$:

$$\frac{1}{d}\bar{L}^d := \Big\{\frac{1}{d}\sum_{i=1}^{d} N^i_{\Lambda_t}\Big\}_{t\geq 0}, \quad \frac{1}{d}\bar{\lambda}_d := \frac{1}{d}\sum_{i=1}^{d}\lambda_i, \quad d \in \mathbb{N}.$$

The following theorem states that the normalized aggregate process $\frac{1}{d}\bar{L}^d$ converges to the scaled directing subordinator $\bar{\lambda}\Lambda$ for increasing dimension d.

Theorem 4.30 (Large portfolio approximation)
Considering Model (M) for increasing dimension $d \in \mathbb{N}$, where the intensities fulfil Assumption (A), the normalized aggregate process converges in finite-dimensional distributions to the scaled subordinator:

$$\frac{1}{d}\bar{L}^d = \Big\{\frac{1}{d}\sum_{i=1}^{d} N^i_{\Lambda_t}\Big\}_{t\geq 0} \xrightarrow{fdd} \bar{\lambda}\Lambda, \quad d\to\infty.$$

Furthermore, if the subordinator has finite k-th moment for some $k \in \mathbb{N}$, then the convergence holds in the k-th mean:

$$\lim_{d\to\infty}\int_0^T \mathbb{E}\left[|\frac{1}{d}\bar{L}^d_t - \bar{\lambda}\Lambda_t|^k\right]\mathrm{d}t = 0, \quad T>0.$$

Proof. For any time $t>0$, convergence in distribution of $\frac{1}{d}\bar{L}^d_t$ to $\bar{\lambda}\Lambda_t$ is shown by pointwise convergence of the Laplace transforms according to Theorem 2.1. All processes involved are Lévy processes, hence the Laplace transforms have a representation via the respective Laplace exponents, i.e. $\varphi_{\Lambda_t}(x) = \exp\{-t\Psi_\Lambda(x)\}$. As the exponential function is continuous, it is sufficient to show convergence of the Laplace exponents. The Laplace exponent of the aggregate process follows from Proposition 4.1 and Remark 3.1:

$$\Psi_{\bar{L}^d}(x) = \Psi_\Lambda\Big(\sum_{i=1}^{d}\lambda_i(1-\exp\{-x\})\Big) = \Psi_\Lambda(\bar{\lambda}_d(1-\exp\{-x\})), \quad x>0.$$

Furthermore, following Theorem 2.1 it is $\varphi_{\frac{1}{d}\bar{L}_t}(x) = \varphi_{\bar{L}_t}(\frac{x}{d})$ and hence

$$\Psi_{\frac{1}{d}\bar{L}^d}(x) = \Psi_{\bar{L}^d}\Big(\frac{x}{d}\Big) = \Psi_\Lambda\Big(\bar{\lambda}_d\Big(1-\exp\Big\{-\frac{x}{d}\Big\}\Big)\Big), \quad x>0.$$

Accordingly, the Laplace exponent of the scaled subordinator is $\Psi_{\bar{\lambda}\Lambda}(x) = \Psi_\Lambda(\bar{\lambda}x)$. The Laplace exponent is continuous as well, so it is sufficient to show convergence of the function arguments. For the exponential function it holds:

$$1-\exp\Big\{-\frac{x}{d}\Big\} = \frac{x}{d} + o\Big(\frac{1}{d}\Big), \quad d\to\infty, \quad x>0.$$

Together with Assumption (A), it follows:

$$\lim_{d\to\infty}\bar{\lambda}^d\Big(1-\exp\Big\{-\frac{x}{d}\Big\}\Big) = \lim_{d\to\infty}\frac{\bar{\lambda}_d}{d}(x+o(1)) = \bar{\lambda}x, \quad x>0.$$

Together with the fact that $\Psi_{\frac{1}{d}\bar{L}^d}(0) = \Psi_{\bar{\lambda}\Lambda}(0) = 0$ for all $d \in \mathbb{N}$, pointwise convergence of the Laplace exponent – and the Laplace transform – is shown for all $x \geq 0$ and hence $d^{-1}\bar{L}_t^d \xrightarrow{d} \bar{\lambda}\Lambda_t$ is proven. Convergence of the finite dimensional distributions follows from the Lévy property of the processes. Consider for any $n \in \mathbb{N}$ w.l.o.g. consecutive points in time $0 := t_0 \leq t_1 \leq \cdots \leq t_n$ and let $\boldsymbol{t} := (t_1, \ldots, t_n)'$. Following Lemma 4.11, the Laplace transform of the random vector $\bar{\lambda}\boldsymbol{\Lambda_t}$ is:

$$\begin{aligned}\varphi_{\bar{\lambda}\boldsymbol{\Lambda_t}}(\boldsymbol{x}) &= \exp\Big\{-\sum_{i=1}^{n} \Delta t_i \Psi_{\bar{\lambda}\Lambda}(x_i + \cdots + x_n)\Big\} \\ &= \prod_{i=1}^{n} \varphi_{\bar{\lambda}\Lambda_{\Delta t_i}}(x_i + \cdots + x_n), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.\end{aligned}$$

Accordingly, the Laplace transform of the random vector $d^{-1}\bar{\boldsymbol{L}}_{\boldsymbol{t}}^d$ is:

$$\varphi_{\frac{1}{d}\bar{\boldsymbol{L}}_{\boldsymbol{t}}^d}(\boldsymbol{x}) = \prod_{i=1}^{n} \varphi_{\frac{1}{d}\bar{L}^d_{\Delta t_i}}(x_i + \cdots + x_d), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

The product of the Laplace transforms converges due to convergence of the n factors, thus $d^{-1}\bar{\boldsymbol{L}}_{\boldsymbol{t}}^d \xrightarrow{d} \bar{\lambda}\boldsymbol{\Lambda_t}$ holds and convergence of the processes in finite dimensional distributions is established.

For the second mode of convergence, we start by showing that convergence in the k-th mean holds for any time $t > 0$ given the k-th moment of Λ exists. Let for now $k \in \mathbb{N}$ be even. Using the multinomial theorem and the formula for the moments of the Poisson distribution as in the proof of Theorem 4.15, it follows:

$$\begin{aligned}\mathbb{E}\left[\left|\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right|^k\right] &= \mathbb{E}\left[\left(\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right)^k\right] = \sum_{j=0}^{k}\binom{k}{j}\frac{(-\bar{\lambda})^{k-j}}{d^j}\,\mathbb{E}[(\bar{L}_t^d)^j\Lambda_t^{k-j}] \\ &= \sum_{j=0}^{k}\binom{k}{j}\frac{(-\bar{\lambda})^{k-j}}{d^j}\,\mathbb{E}[\Lambda_t^{k-j}\,\mathbb{E}[(\bar{L}_t^d)^j|\sigma(\Lambda)]] \\ &= \sum_{j=0}^{k}\binom{k}{j}\frac{(-\bar{\lambda})^{k-j}}{d^j}\,\mathbb{E}[\Lambda_t^{k-j}\sum_{l=0}^{j}S(j,l)\bar{\lambda}_d^l\Lambda_t^l] \\ &= \sum_{j=0}^{k}\binom{k}{j}\frac{(-\bar{\lambda})^{k-j}}{d^j}\sum_{l=0}^{j}S(j,l)\bar{\lambda}_d^l\,\mathbb{E}[\Lambda_t^{k-j+l}].\end{aligned}$$

The number of summation terms is fixed for given k and the only quantities dependent on d are the factor d^{-j} in the outer summation and $\bar{\lambda}_d^l$ in the inner summation. For indices $0 \leq l < j \leq k$, the summation term is scaled with

$$\frac{\bar{\lambda}_d^l}{d^j} = \left(\frac{\bar{\lambda}_d}{d}\right)^l \frac{1}{d^{j-l}}.$$

While the fist factor converges to $\bar{\lambda}^l$ according to Assumption (A), the second factor converges to zero as $l - j > 0$. Thus, all these summation terms vanish for $d \to \infty$ and only the behaviour of the terms corresponding to $0 \leq l = j \leq k$ remains to be examined:

$$\begin{aligned}
\lim_{d\to\infty} \mathbb{E}\left[\left|\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right|^k\right] &= \lim_{d\to\infty} \sum_{j=0}^{k} \binom{k}{j} \frac{(-\bar{\lambda})^{k-j}}{d^j} S(j,j) \bar{\lambda}_d^j \mathbb{E}[\Lambda_t^k] \\
&= \mathbb{E}[\Lambda_t^k] \lim_{d\to\infty} \sum_{j=0}^{k} \binom{k}{j} (-\bar{\lambda})^{k-j} \left(\frac{\bar{\lambda}_d}{d}\right)^j \\
&= \mathbb{E}[\Lambda_t^k] \sum_{j=0}^{k} \binom{k}{j} (-\bar{\lambda})^{k-j} \bar{\lambda}^j = \mathbb{E}[\Lambda_t^k](-\bar{\lambda})^k \sum_{j=0}^{k} \binom{k}{j} (-1)^j \\
&= \mathbb{E}[\Lambda_t^k](-\bar{\lambda})^k (1-1)^k = 0.
\end{aligned}$$

This proves $d^{-1}\bar{L}_t^d \xrightarrow{L^k} \bar{\lambda}\Lambda$ for $k \in \mathbb{N}$ even. Convergence of the integral w.r.t. time up to some horizon $T > 0$ follows as the only factors dependent on t are the subordinator moments $\mathbb{E}[\Lambda_t^j]$. According to Remark 4.14, these moments are found by evaluating the j-th complete Bell polynomial in the first j cumulants of Λ_t. The cumulants are linear in t, hence the moments are a polynomial in t and the integral is finite given the k-th moment of the subordinator is finite. Following this, the convergence in d is not affected by the integral and

$$\lim_{d\to\infty} \int_0^T \mathbb{E}\left[\left|\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right|^k\right] \mathrm{d}t = 0$$

holds for k even. For $k+1$ odd, convergence is shown similarly to the even case by using the inequality

$$\left|\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right|^{k+1} = \left|\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right|\left(\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right)^k \leq \left(\frac{1}{d}\bar{L}_t^d + \bar{\lambda}\Lambda_t\right)\left(\frac{1}{d}\bar{L}_t^d - \bar{\lambda}\Lambda_t\right)^k.$$

□

The result about the convergence of the aggregate process is not too surprising considering that the infinitely divisible distributions are the limit sequences of compound Poisson distributions (see (Feller, 1971, Chapter XVII.1, Theorem 3, p.557)) and every Lévy process is the limit of a possibly infinite number of compound Poisson processes, as was discussed in Section 2.3. While in the Lévy–Itô representation the convergence is considered specifically regarding the small jumps, here the subordinator is approximated by compound Poisson processes on the increasingly fine grid $d^{-1}\mathbb{N}$.

Whereas from convergence in the k-th mean follows convergence in probability, from which again follows convergence in distribution (see (Kallenberg, 2002, Chapter 4, Proposition 4.12, p.68, and Lemma 4.7, p.66)), existence of the k-th subordinator moment has to be assumed in this case. In the above theorem, convergence in distribution is shown regardless of the existence of subordinator moments. Furthermore, according to (Kallenberg, 2002, Chapter 1, Lemma 1.29, p.15), from convergence in the k-th mean follows convergence of the absolute k-th moments. As in the given setting the processes are non-negative, it follows in case of existing k-th subordinator moment:

$$\lim_{d\to\infty} \frac{1}{d^k}\,\mathbb{E}[(\bar{L}_t^d)^k] = \bar{\lambda}^k\,\mathbb{E}[\Lambda_t^k].$$

The result about convergence in distribution of the aggregate process can be exploited for approximations in high-dimensional applications. This may be particularly relevant for the pricing of certain reinsurance contracts, cf. Section 6.1. According to the definition of convergence in distribution – see Section 2.1 – it holds for all continuous and bounded functions $f \in C_b(\mathbb{R},\mathbb{R})$:

$$\lim_{d\to\infty} \mathbb{E}\left[f\left(\frac{1}{d}\bar{L}_t^d\right)\right] = \mathbb{E}[f(\bar{\lambda}\Lambda_t)], \quad t \geq 0.$$

For instance, calculating the expected number of claims within a specified layer, that is for some $0 < a < b$ the expectation

$$\mathbb{E}[\min\{\max\{\bar{L}_t^d - a, 0\}, b\}],$$

corresponds to the continuous and bounded function $f(x) = \min\{\max\{x - a, 0\}, b\}$. Thus, the following approximation is valid for large $d \in \mathbb{N}$:

$$\mathbb{E}[\min\{\max\{\bar{L}_t^d - a, 0\}, b\}] \approx \mathbb{E}[\min\{\max\{d\bar{\lambda}\Lambda_t - a, 0\}, b\}].$$

For comparison, the normalized aggregate process $d^{-1}\sum_{i=1}^{d} N^i$ in the underlying Poisson model converges in finite dimensional distributions to a linear function with slope $\bar{\lambda}$. This can be shown by convergence of the Laplace exponent similarly to the above proof:

$$\lim_{d\to\infty} \Psi_{\frac{1}{d}\sum_{i=1}^{d} N^i}(x) = \lim_{d\to\infty} \bar{\lambda}_d\Big(1 - \exp\Big\{-\frac{x}{d}\Big\}\Big) = \bar{\lambda}x, \quad d \to \infty, \quad x > 0.$$

It also follows from the law of large numbers for independent (but not necessarily identically distributed) random variables, for instance in the version given in (Feller, 1968, Chapter X.5, pp.253). Versions with different assumptions can be found, for instance, in (Feller, 1971, Chapter VII.8, Theorem 3, p.239) and (Czado and Schmidt, 2011, Chapter 1.4, Theorem 1.29, p.26). Model (M) can also be compared to the model with the same marginal processes but without dependence. For this purpose, let $\Lambda^1 = \{\Lambda_t^1\}_{t\geq 0}, \Lambda^2 = \{\Lambda_t^2\}_{t\geq 0}, \ldots$ be independent copies of the subordinator $\Lambda = \{\Lambda_t\}_{t\geq 0}$ – the copies are assumed to be independent of the Poisson processes as well – and consider the process

$$\{(N^1_{\Lambda_t^1}, \ldots, N^d_{\Lambda_t^d})'\}_{t\geq 0}.$$

The behaviour of the normalized aggregate process $d^{-1}\sum_{i=1}^{d} N^i_{\Lambda_t^i}$ for increasing dimension can also be determined from the law of large numbers due to the independence of the components. Assuming suitable conditions are met, it holds:

$$\frac{1}{d}\sum_{i=1}^{d}(N^i_{\Lambda_t^i} - \mathbb{E}[N^i_{\Lambda_t^i}]) \longrightarrow 0, \quad d \to \infty.$$

The convergence holds almost surely or at least in probability, thus it also holds in distribution. Given Assumption (A), it follows:

$$\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \mathbb{E}[N^i_{\Lambda_t^i}] = t\,\mathbb{E}[\Lambda_1] \lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \lambda_i = \bar{\lambda} t\,\mathbb{E}[\Lambda_1].$$

Thus, the process convergences in distribution to $\bar{\lambda}\,\mathbb{E}[\Lambda_1]t$ for any point in time $t \geq 0$ and due to the Lévy property, the process converges in finite-dimensional distributions to the linear function with slope $\bar{\lambda}\,\mathbb{E}[\Lambda_1]$. If Assumption (TN) for time-normalization holds, the limit is the same as in the independent Poisson case.

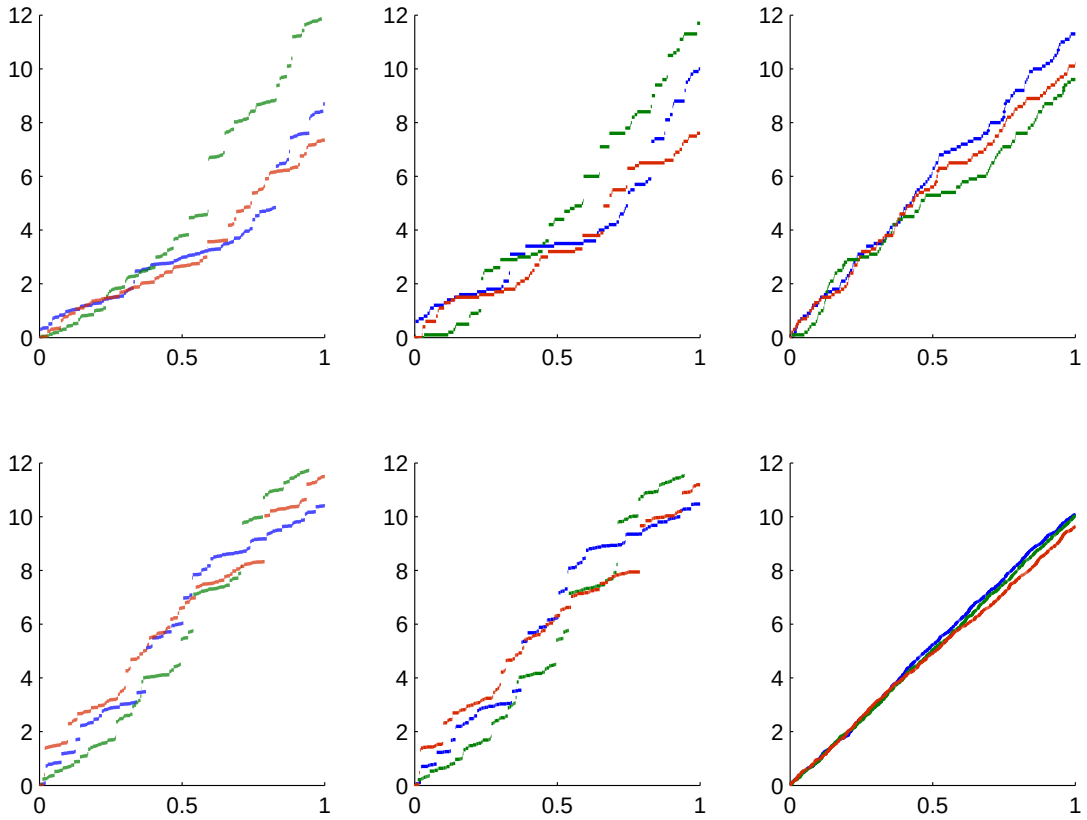


Figure 4.9 Large portfolio approximation: For dimension $d = 10$ (upper row) and $d = 100$ (lower row), intensity $\lambda_i = 10$ for $i = 1, \dots, d$, time horizon $T = 1$, and an inverse Gaussian subordinator ($b = 0$, $\beta = \eta = 4.5$), the figure shows three sample paths each of the scaled subordinator $10\Lambda_t$ (left), of the normalized aggregate underlying Poisson process $d^{-1} \sum_{i=1}^{d} N_t^i$ (right), and of the corresponding normalized aggregate time-changed process $d^{-1} \bar{L}_t^d$ (middle).

In Figure 4.9 the aforementioned convergence results are visualized. On the right-hand side, three paths of a normalized aggregate Poisson process corresponding to dimension $d = 10$ (upper chart) and $d = 100$ (lower chart) are plotted. The intensities are assumed to be homogeneous. For $d = 100$, the paths are already almost linear. On the left-hand side, three paths of an inverse Gaussian subordinator scaled with the intensity are plotted in the upper as well as the lower chart. In the middle column, the normalized aggregate process in the time-changed model corresponding to the subordinator paths on the left and the Poisson paths on the right are shown. While in the low-dimensional case the paths look still different from the subordinator paths, in the high-dimensional case the paths closely resemble each other. In general, convergence of finite-dimensional distributions does not necessarily mean that the paths of the processes look alike or even similar – a counterexample can be found in (Billingsley, 2012, Chapter 4.23, pp. 327) – but for Lévy processes, this is the case. The following corollary gives the result on pathwise approximation specifically for the process at hand, but holds in fact for Lévy processes in general.

Corollary 4.31 (Pathwise approximation)
In the setting of Theorem 4.30, a sequence of processes $X^1, X^2, \dots$ exists such that $X^d \overset{d}{=} d^{-1}\bar{L}^d$ for all $d \in \mathbb{N}$ and

$$\sup_{0 \le s \le t} \left| X_s^d - \frac{1}{d}\bar{L}_s^d \right| \xrightarrow{\mathbb{P}} 0, \quad d \to \infty, \quad t \ge 0.$$

Furthermore, the convergence of the processes holds in distribution in the Skorohod J_1-topology.

Proof. Given the convergence of the processes in distribution at one point in time, which was proven in Theorem 4.30, the first statement follows from (Kallenberg, 2002, Chapter 15, Theorem 15.17, p.298) and the second statement follows from (Jacod and Shiryaev, 2003, Chapter VII.3, Corollary 3.6, p.415). □

Finally, convergence in distribution for Lévy processes also means convergence of the characteristics. The results given in (Sato, 1999, Chapter 2.8, Theorem 8.7, pp.41) for general Lévy processes can be adapted to the specific case of the Lévy subordinator studied here similarly to the proof of Theorem 4.17. According to Remark 3.1 and Theorem 4.17, the aggregate process $\bar{L}^d$ has zero drift $b_{\bar{L}^d} = 0$ and its Lévy measure is concentrated on $\mathbb{N}$:

$$\nu_{\bar{L}^d}(k) = -\frac{(-\bar{\lambda}_d)^k}{k!}\Psi_\Lambda^{(k)}(\bar{\lambda}_d), \quad k \in \mathbb{N}.$$

The normalized process $d^{-1}\bar{L}^d$ has zero drift $b_{\frac{1}{d}\bar{L}^d} = 0$ as well and the jumps are shifted from $k \in \mathbb{N}$ to k/d, hence the Lévy measure is $\nu_{\frac{1}{d}\bar{L}^d}(l) = \nu_{\bar{L}^d}(ld)$ for $l \in d^{-1}\mathbb{N}$. The Lévy measure converges vaguely on $\mathbb{R}_{>0}$ to $\nu_{\bar{\lambda}\Lambda}(\cdot) = \nu_\Lambda(\bar{\lambda}^{-1}\cdot)$, that is for all continuous and bounded functions $f \in C_b(\mathbb{R}_{>0}, \mathbb{R})$ it holds:

$$\lim_{d\to\infty} \int_{(0,\infty]} f(x)\, \nu_{\frac{1}{d}\bar{L}^d}(\mathrm{d}x) = \lim_{d\to\infty} \sum_{k=1}^{\infty} f\Big(\frac{k}{d}\Big) \nu_{\bar{L}^d}(k) = \int_{(0,\infty]} f(x)\, \nu_{\bar{\lambda}\Lambda}(\mathrm{d}x).$$

The drift $b_{\frac{1}{d}\bar{L}^d} = 0$ of $d^{-1}\bar{L}^d$ does not directly converge to the drift $b_{\bar{\lambda}\Lambda} = \bar{\lambda} b_\Lambda$ of $\bar{\lambda}\Lambda$. This is obvious as the subordinator may be selected with non-zero drift. Instead, the following convergence holds:

$$\lim_{d\to\infty} \Big(b_{\frac{1}{d}\bar{L}^d} + \int_{(0,1]} x\, \nu_{\frac{1}{d}\bar{L}^d}(\mathrm{d}x) \Big) = \lim_{d\to\infty} \sum_{k=1}^{d} \frac{k}{d} \nu_{\bar{L}^d}(k) = b_{\bar{\lambda}\Lambda} + \int_{(0,1]} x\, \nu_{\bar{\lambda}\Lambda}(\mathrm{d}x).$$

5 Estimation

This chapter treats the problem of estimating the model parameters from historical claim count data. In general, statistical inference for multivariate processes is far from trivial and often computationally expensive. In Section 5.1 four estimation procedures are developed which essentially try to fit a multivariate distribution, either the infinitely divisible process distribution or the jump size distribution (and jump intensity). Since in Chapter 4 the methods for an efficient evaluation of these distributions were established, the estimation procedures can be implemented without much difficulty, as will be discussed in Section 5.2. An extensive simulation study will then be carried out to examine the quality and asymptotical properties of the estimators. In Section 5.3, the methods are applied to the set of Danish fire insurance data in order to test the effectiveness of the estimation procedures as well as the model itself in a real-world setting.

5.1 Estimation procedures

In this section, four estimation methods will be presented for the parameters of Model (M), two for discrete observation points of the process exploiting the Lévy property and the results about the infinitely divisible distribution of the process, and another two based on the knowledge of the realized process path up to some finite time horizon using the compound Poisson representation. For an introduction to statistical inference, see Casella and Berger (2002) and Davison (2003). An overview specifically for actuarial applications is given in Klugman et al. (2004). Estimation methods for stochastic processes are discussed in Basawa and Prakasa Rao (1980). Estimation of multivariate distributions and processes with multiple parameters is often computationally difficult. The results derived in Chapter 4 for the process and jump-size distribution, however, enable an efficient implementation of the presented methods.

Separating the marginals from the dependence structure – for the multivariate distribution of the process at one point in time via a copula function or for the dynamic process via a Lévy copula – and following a two-step estimation approach (e.g. the method of inference function for margins (IFM), see Joe and Xu (1996) and Joe (2005)) is difficult in the given setting: neither of the (Lévy) copula functions is available in closed form and, more importantly, the influence of the model parameters cannot be isolated to either the marginals or the dependence structure. Estimation based on dependence measures like Kendall's tau or Spearman's rho face the same problem. In addition, the difficulties in

using copulas and copula-based dependence measures for count data have been discussed in Section 4.1; see also Genest and Nešlehová (2007).

The parameter set that has to be estimated consists of the d intensities in the vector $\boldsymbol{\lambda}$ and the subordinator parameters, which will be denoted by vector $\boldsymbol{\theta}$ in the following. Throughout this chapter, the dependence of quantities on the parameters is often emphasised in the notation, e.g. it will be written

$$\mathbb{E}[\boldsymbol{L}_t; \boldsymbol{\lambda}, \boldsymbol{\theta}] \quad \text{or} \quad \mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k}; \boldsymbol{\lambda}, \boldsymbol{\theta}),$$

for $\mathbb{E}[\boldsymbol{L}_t]$ or $\mathbb{P}(\boldsymbol{L}_t = \boldsymbol{k})$, respectively. The parameter space for $\boldsymbol{\lambda}$ is $\mathbb{R}^d_{>0}$, the parameter space for $\boldsymbol{\theta}$ depends on the selected subordinator and will be denoted by Θ. For an inverse Gaussian subordinator with drift, for instance, it is $\boldsymbol{\theta} := (b, \beta, \eta)' \in \Theta := [0, 1) \times \mathbb{R}^2_{>0}$, i.e. the total number of parameters in the model $d + 3$. Without drift and if the time-normalization Assumption (TN) is imposed, only one subordinator parameter $\beta = \eta > 0$ needs to be estimated in addition to $\boldsymbol{\lambda}$ (cf. Table 4.4). Since all estimators discussed in the following will be specified via minimization (or maximization) problems, optimization methods considering constraints will be necessary. As the number of subordinator parameters is relatively modest in most cases, the number of intensities will dominate in high dimensional applications. If high dimensionality renders the joint optimization infeasible in the time-changed model, time-normalization may be imposed and the estimator $\hat{\boldsymbol{\lambda}}$ for the intensities may be selected upfront as the sample mean of the process distribution after one time unit, i.e. for the observation $\hat{\boldsymbol{L}}_T$ of $\boldsymbol{L}$ at the finite time-horizon $T > 0$:

$$\hat{\boldsymbol{\lambda}} := \frac{1}{T} \hat{\boldsymbol{L}}_T. \tag{5.1}$$

Hence, (5.1) is the moment and maximum likelihood estimator for the underlying independent Poisson process $\boldsymbol{N}$. Given the estimated intensities $\hat{\boldsymbol{\lambda}}$, only the remaining subordinator parameters need to be determined from the methods presented in the following. Note, however, that this two-step procedure does not correspond to a separation of marginals and dependence structure as discussed above, since the subordinator parameters influence both, the marginal distribution as well as the dependence structure.

For each additional component in the model, one additional parameter, namely the intensity of the new component, is introduced; the number of subordinator parameters remains the same. Observations of the new component, however, reveal not only information about its intensity but also about the subordinator. Hence, the quality of the estimators for θ can be expected to increase if the dimensionality of the model increases. Overall, the number of parameters of the model relative to its dimension and, hence, the dimension of the observations is low enough to give rise to the hope that the parameters can be determined reasonably well from the following estimation methods.

Discrete monitoring

Assume that the process $\boldsymbol{L}$ is observed up to a finite time horizon $T > 0$ on a discrete time grid $0 := t_0 < t_1 < \ldots < t_n := T$ with $n \in \mathbb{N}$ monitoring points. For convenience, we assume an equidistant grid with step size $h := T/n$, i.e. $t_j := hj$ for $j = 0, \ldots, n$. Denote the observations of $\boldsymbol{L}$ at times t_j as

$$\hat{\boldsymbol{l}}_j = (\hat{l}_{1j}, \ldots, \hat{l}_{dj})' := \hat{\boldsymbol{L}}_{t_j}, \quad j = 1, \ldots n,$$

and set $\hat{\boldsymbol{l}}_0 := \mathbf{0}$. From a process perspective, these panel data are only a single partial observation $\{\hat{\boldsymbol{l}}_j = \hat{\boldsymbol{L}}_{t_j}\}_{j=1,\ldots,n}$ of $\boldsymbol{L}$, i.e. one path of the process is observed at n grid points. Following the Lévy property of the process $\boldsymbol{L}$, however, the increments $\Delta\hat{\boldsymbol{l}}_j = \hat{\boldsymbol{l}}_j - \hat{\boldsymbol{l}}_{j-1}$, $j = 1, \ldots, n$, are iid observations of the infinitely divisible distribution $\boldsymbol{L}_h$ of the process $\boldsymbol{L}$ at the step size h. Hence, estimation methods for the process $\boldsymbol{L}$ with only one partial observation can be reduced to estimation methods for the multivariate distribution $\boldsymbol{L}_h$ with n observations $\Delta\hat{\boldsymbol{l}}_j$. In the following, two approaches will be discussed, one based on the method of moments and the other using maximum likelihood estimation. For general Lévy processes, these methods are discussed in (Cont and Tankov, 2003, Chapter 7, pp.207).

For univariate distributions the method of moments intends to match the empirical sample moments with the model predicted moments. If the same number of moments are taken into consideration as there are parameters to be estimated, the result is a system of equations where the number of unknowns equals the number of equations and, ideally, there should be a unique solution. In reality, however, no solution may exist or none within the admissible parameter range. In such cases, the method (or the model) may be rejected or the parameters minimizing the error, for instance in a least-square sense, within the feasible set may be selected instead. The minimum error parameters are also typically chosen if more moments than parameters are included or a generalized method of moments with more advanced moment conditions is applied.

In the multivariate setting here, matching the expectation gives d equations. Including the second (mixed) moments adds another $d(d+1)/2$ equations. The total number of parameters in the model is d intensities plus the subordinator parameters. Hence, in most applications the system of equations will already be overdetermined if the first and second moments are included, thus we restrict our considerations to this case. Otherwise, additional higher moments may be added as necessary.

Instead of pure moments, we match relative central moments to account for scale differences of the various quantities. To counteract dependence between the empirical moments, data-based and non-diagonal weighting matrices for the vector of moments can

be considered, see Hansen (1982). The distance is measured in a least-square sense, i.e. using the (squared) Euclidean norm for the expectation based vector part and the (squared) Frobenius norm for the matrix based covariance part. We formulate the first estimation method using the results from Theorem 4.15. For the sake of readability, we suppress the dependence of the objective function on the observations in the notation; we will do so as well for the other methods presented in the remainder of this section.

Estimation 5.1 (Moment matching (M1))
Given equidistant observations $\hat{\boldsymbol{l}}_0 := \mathbf{0} \leq \hat{\boldsymbol{l}}_1 \leq \cdots \leq \hat{\boldsymbol{l}}_n$ with step size h, let $\hat{\boldsymbol{m}}$ denote the sample mean vector:

$$\hat{\boldsymbol{m}} := \hat{\mathbb{E}}[\boldsymbol{L}_h] := \frac{1}{n}\hat{\boldsymbol{l}}_n.$$

Furthermore, let $\hat{Q}$ denote the unbiased sample covariance matrix with components

$$\hat{Q}_{ik} := \hat{\mathbb{C}}\mathrm{ov}[L_h^i, L_h^k] := \frac{1}{(n-1)} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} (\Delta \hat{l}_{ij_1} - m_i)(\Delta \hat{l}_{kj_2} - m_k), \quad i,k = 1,\ldots,d.$$

Then the estimators $\hat{\boldsymbol{\lambda}}^{M1}$ of the intensities $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\theta}}^{M1}$ of the subordinator parameters $\boldsymbol{\theta}$ in Model (M) are defined via the optimization problem

$$\{\hat{\boldsymbol{\lambda}}^{M1}, \hat{\boldsymbol{\theta}}^{M1}\} := \underset{\boldsymbol{\lambda}\in\mathbb{R}^d_{>0}, \boldsymbol{\theta}\in\Theta}{\arg\min}\ f^{M1}(\boldsymbol{\lambda}, \boldsymbol{\theta}), \tag{M1}$$

with objective function

$$\begin{aligned} f^{M1}(\boldsymbol{\lambda}, \boldsymbol{\theta}) :=& \sum_{i=1}^{d} \left(\frac{\mathbb{E}[L_h^i; \lambda_i, \boldsymbol{\theta}] - \hat{m}_i}{\mathbb{E}[L_h^i; \lambda_i, \boldsymbol{\theta}]} \right)^2 + \sum_{i=1}^{d} \sum_{k=1}^{d} \left(\frac{\mathbb{C}\mathrm{ov}[L_h^i, L_h^k; \lambda_i, \lambda_k, \boldsymbol{\theta}] - \hat{Q}_{ik}}{\mathbb{C}\mathrm{ov}[L_h^i, L_h^k; \lambda_i, \lambda_k, \boldsymbol{\theta}]} \right)^2 \\ =& \sum_{i=1}^{d} \left(1 - \frac{\hat{m}_i}{h\lambda_i \, \mathbb{E}[\Lambda_1; \boldsymbol{\theta}]} \right)^2 + \sum_{i=1}^{d} \left(1 - \frac{\hat{Q}_{ii}}{h(\mathbb{V}\mathrm{ar}[\Lambda_1; \boldsymbol{\theta}]\lambda_i^2 + \mathbb{E}[\Lambda_1; \boldsymbol{\theta}]\lambda_i)} \right)^2 \\ &+ 2 \sum_{i=1}^{d-1} \sum_{k=i+1}^{d} \left(1 - \frac{\hat{Q}_{ik}}{h \, \mathbb{V}\mathrm{ar}[\Lambda_1; \boldsymbol{\theta}]\lambda_i\lambda_k} \right)^2 \end{aligned}$$

As discussed at the beginning of this section, if the time-normalization Assumption (TN) is imposed and $\mathbb{E}[\boldsymbol{L}_h] = h\boldsymbol{\lambda}$ holds, the intensities may be set upfront via the sample mean:

$$\hat{\boldsymbol{\lambda}} := \frac{1}{h}\hat{\boldsymbol{m}} = \frac{1}{T}\hat{\boldsymbol{l}}_n.$$

Since the subordinator only affects the covariance part in this case, the approach may be reasonable for this method. The estimation problem then reduces to:

$$\hat{\boldsymbol{\theta}}^{M1} := \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{d}\sum_{k=1}^{d} \Big(\frac{\mathbb{C}\mathrm{ov}[L_h^i, L_h^k; \hat{\lambda}_i, \hat{\lambda}_k, \boldsymbol{\theta}] - \hat{Q}_{ik}}{\mathbb{C}\mathrm{ov}[L_h^i, L_h^k; \hat{\lambda}_i, \hat{\lambda}_k, \boldsymbol{\theta}]} \Big)^2.$$

Either way, often the optimization cannot be solved analytically by equating the first partial derivatives to zero and testing the second derivative (particularly considering the restrictions on the parameter space) and numerical optimization routines are necessary. The runtime, however, is expected to be short as mean and variance are known in closed form for many subordinators. Of course, this estimation approach considers by definition only parts of the full distribution of $\boldsymbol{L}_h$. In particular, only linear dependence between the components is included and any further dependence is neglected, compare the discussion at the end of Section 4.1. For very high-dimensional problems or insufficient data that render other methods intractable, this simple approach may prove useful. In addition, the estimators can be used as starting values for more advanced methods. Though often not unbiased, the generalized method of moment produces consistent and asymptotically normal estimators in many cases, see Hansen (1982). These properties will be investigated in a simulation study in Section 5.2.

In contrast to the moment matching Method (M1), maximum likelihood estimation is based on the full distribution function. The aim is to choose the parameters that make the observations most likely to occur. Using the results from Theorem 4.2 and Corollary 4.4, the log-likelihood function for the given observations is:

$$\begin{aligned} l^{M2}(\boldsymbol{\lambda}, \boldsymbol{\theta}) &:= \ln\Big\{ \prod_{j=1}^{n} \mathbb{P}(\boldsymbol{L}_h = \Delta\hat{\boldsymbol{l}}_j; \boldsymbol{\lambda}, \boldsymbol{\theta}) \Big\} = \sum_{j=1}^{n} \ln\Big\{ \frac{(-\boldsymbol{\lambda})^{\Delta\hat{\boldsymbol{l}}_j}}{\Delta\hat{\boldsymbol{l}}_j!} \varphi_{\Lambda_h}^{(|\Delta\hat{\boldsymbol{l}}_j|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta}) \Big\} \\ &= \sum_{i=1}^{d} \hat{l}_{in} \ln\{\lambda_i\} - \sum_{i=1}^{d}\sum_{j=1}^{n} \ln\{\Delta\hat{l}_{ij}!\} + \sum_{j=1}^{n} \ln\{(-1)^{|\Delta\hat{\boldsymbol{l}}_j|} \varphi_{\Lambda_h}^{(|\Delta\hat{\boldsymbol{l}}_j|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\}. \end{aligned}$$

Note that the second term is independent of the parameters and, thus, can be neglected in the optimization. Together, the following maximum likelihood estimation method can be formulated.

Estimation 5.2 (Maximum likelihood for increments (M2))
Given equidistant observations $\hat{\boldsymbol{l}}_0 := \mathbf{0} \leq \hat{\boldsymbol{l}}_1 \leq \cdots \leq \hat{\boldsymbol{l}}_n$ with step size h, the estimators $\hat{\boldsymbol{\lambda}}^{M2}$ of the intensities $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\theta}}^{M2}$ of the subordinator parameters $\boldsymbol{\theta}$ in Model (M) are defined via the optimization problem

$$\{\hat{\boldsymbol{\lambda}}^{M2}, \hat{\boldsymbol{\theta}}^{M2}\} := \underset{\boldsymbol{\lambda}\in\mathbb{R}^d_{>0},\, \boldsymbol{\theta}\in\Theta}{\arg\max}\; f^{M2}(\boldsymbol{\lambda}, \boldsymbol{\theta}), \tag{M2}$$

with objective function

$$f^{M2}(\boldsymbol{\lambda}, \boldsymbol{\theta}) := \sum_{i=1}^{d} \hat{l}_{in} \ln\{\lambda_i\} + \sum_{j=1}^{n} \ln\{(-1)^{|\Delta\hat{\boldsymbol{l}}_j|} \varphi_{\Lambda_h}^{(|\Delta\hat{\boldsymbol{l}}_j|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\}.$$

Analytical solutions for the maximum likelihood estimators can in general not be found, even in the univariate case (cf. (Klugman et al., 2004, Chapter 12.5, pp.383)); see also Stein et al. (1987) for the multivariate extension of the Sichel distribution and (Johnson et al., 1997, Chapter 36.5, pp.102) for the negative binomial distribution. As for Method (M1), the optimization problem has to be solved numerically. Hence, the log-likelihood function needs to be evaluated repeatedly in multiple optimization steps and thus the tractability of this method strongly depends on the availability of quick and stable evaluation routines for the probabilities of $\boldsymbol{L}$, or derivatives of the Laplace transform of Λ – a problem discussed at length in Section 4.1 and 4.2. In particular, the results of Corollary 4.29 enable an efficient and reliable implementation of these quantities such that the optimization can be carried out in a reasonable amount of time, as will be seen in Section 5.2. Maximum likelihood estimators are under certain regularity conditions consistent and asymptotically normal, see (Serfling, 2002, Chapter 4.2, pp.143). These properties will be investigated heuristically in the following section.

Finally, it is worth mentioning that, given only discrete observation points, it is unknown whether the increments are the result of many small or few large jumps of $\boldsymbol{L}$. Hence, it may come to identifiability problems regarding the parameters for any estimation method based solely on such data, particularly if a subordinator drift is included in the modelling approach, see the discussion along Figure 4.6 in Section 4.1. Decreasing the step size of the grid increases the number of observations for the two estimation methods. If no clusters arrive within some time interval, intermediate steps reveal no additional information, but an increasingly fine grid isolates the cluster arrivals. Taking the limit, the process is continuously monitored and the precise arrival times and sizes of all clusters are recorded. Two estimation methods specifically for his setting will be discussed in the remainder of this section.

Continuous monitoring

Ideally, the claim number process $\boldsymbol{L}$ is monitored continuously up to a finite time horizon $T > 0$, providing one observation $\{\hat{\boldsymbol{L}}_t\}_{0 \le t \le T}$ of $\boldsymbol{L}$ on $[0, T]$. As $\boldsymbol{L}$ is a Poison cluster process, the arrival times and jump sizes of each cluster of claims are recorded in this scenario. Let $\hat{m} \in \mathbb{N}_0$ be the total number of incoming clusters in the time period $[0, T]$, $0 := \hat{w}_0 < \hat{w}_1 < \ldots < \hat{w}_{\hat{m}}$ the successive cluster arrival times, and $\hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_{\hat{m}}$ the corresponding cluster sizes. From these data, a discrete observation of $\boldsymbol{L}$ can be extracted and the previously discussed Methods (M1) and (M2) can be applied. To exploit all available information, two more estimation methods are proposed in the following based on the Poisson cluster process representation of $\boldsymbol{L}$.

Firstly, maximum likelihood estimation is used jointly for the iid cluster inter-arrival times and the iid cluster sizes, cf. the approach in (Basawa and Prakasa Rao, 1980, Chapter 6.4, pp.105) for univariate compound Poisson processes. The maximum likelihood method requires no iid samples, it is sufficient that the joint likelihood function of the observations is available, i.e. the probability of observing the given data in the model, which reduces for point observations of continuous distributions to the mass of the density function at the observation point. For a general compound Poisson process, where the jump size distribution is specified independently of the Poisson arrival process, maximum likelihood estimation can be performed separately for both parts. In the given setting, however, the parameters affect both, jump size and jump intensity, making a joint optimization necessary.

Theorem 5.3 (Loglikelihood function of compound Poisson process)
Given an observation period $[0, T]$ *with* $\hat{m} \in \mathbb{N}_0$ *cluster arrivals at times* $\hat{w}_0 := 0 < \hat{w}_1 < \ldots < \hat{w}_{\hat{m}}$ *and with sizes* $\hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_{\hat{m}}$*, the log-likelihood function of the observations from Model* (M) *is*

$$l^{M3}(\boldsymbol{\lambda}, \boldsymbol{\theta}) := -T\Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta}) + \sum_{i=1}^{d} \hat{c}_i \ln\{\lambda_i\} - \sum_{i=1}^{d}\sum_{j=1}^{\hat{m}} \ln\{\hat{y}_{ij}!\}$$
$$+ \sum_{j=1}^{\hat{m}} \ln\{(-1)^{|\hat{\boldsymbol{y}}_j|+1} \Psi_\Lambda^{(|\hat{\boldsymbol{y}}_j|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\},$$

where $\hat{c}_i := \sum_{j=1}^{\hat{m}} \hat{y}_{ij}$ *is the sum over all jumps of component* $i = 1, \ldots, d$.

Proof. Following Corollary 4.19, the cluster arrival times W_j in Model (M) stem from a Poisson process with intensity $\Psi_\Lambda(|\boldsymbol{\lambda}|)$, hence, the inter-arrival times ΔW_j are iid

$\mathrm{Exp}(\Psi_\Lambda(|\boldsymbol{\lambda}|))$-distributed random variables. For $\hat{m}$ arrivals before T at times $\hat{w}_1, \dots, \hat{w}_{\hat{m}}$, the joint likelihood is:

$$\begin{aligned} f^1(\boldsymbol{\lambda}, \boldsymbol{\theta}) :=& \mathbb{P}(\Delta W_{\hat{m}+1} > T - \hat{w}_{\hat{m}}; \boldsymbol{\lambda}, \boldsymbol{\theta}) \prod_{j=1}^{\hat{m}} \Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta}) \exp\{-\Delta \hat{w}_j \Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\} \\ =& \exp\{-(T - \hat{w}_m)\Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\} \Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})^{\hat{m}} \exp\{-\hat{w}_{\hat{m}} \Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\} \\ =& \exp\{-T\Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\} \Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})^{\hat{m}}. \end{aligned}$$

The likelihood function of the iid cluster sizes follows also from Corollary 4.19:

$$\begin{aligned} f^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) :=& \prod_{j=1}^{\hat{m}} \mathbb{P}(\boldsymbol{Y} = \hat{\boldsymbol{y}}_j; \boldsymbol{\lambda}, \boldsymbol{\theta}) = \prod_{j=1}^{\hat{m}} -\frac{(-\boldsymbol{\lambda})^{\hat{\boldsymbol{y}}_j}}{\hat{\boldsymbol{y}}_j!} \frac{\Psi_\Lambda^{(|\hat{\boldsymbol{y}}_j|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta})}{\Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})} \\ =& \Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta})^{-\hat{m}} \frac{\boldsymbol{\lambda}^{\sum_{j=1}^{\hat{m}} \hat{\boldsymbol{y}}_j}}{\prod_{j=1}^{\hat{m}} \hat{\boldsymbol{y}}_j!} \prod_{j=1}^{\hat{m}} (-1)^{|\hat{\boldsymbol{y}}_j|+1} \Psi_{\boldsymbol{\Lambda}}^{(|\hat{\boldsymbol{y}}_j|)}(|\boldsymbol{\lambda}|). \end{aligned}$$

Due to the independence of the cluster arrival process and the cluster sizes, the joint log-likelihood function is the logarithm of the product of the two likelihood functions:

$$\begin{aligned} l^{M3}(\boldsymbol{\lambda}, \boldsymbol{\theta}) :=& \ln\{f^1(\boldsymbol{\lambda}, \boldsymbol{\theta}) f^2(\boldsymbol{\lambda}, \boldsymbol{\theta})\} = -T\Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta}) + \sum_{i=1}^{d} \underbrace{\sum_{j=1}^{\hat{m}} \hat{y}_{ij}}_{=\hat{c}_i} \ln\{\lambda_i\} \\ & - \sum_{i=1}^{d} \sum_{j=1}^{\hat{m}} \ln\{\hat{y}_{ij}!\} + \sum_{j=1}^{[} \hat{m} \ln\{(-1)^{|\hat{\boldsymbol{y}}_j|+1} \Psi_\Lambda^{(|\hat{\boldsymbol{y}}_j|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\}. \end{aligned}$$

□

The third term in the log-likelihood function l^{M3} does not depend on the parameters and can be neglected in the optimization.

Estimation 5.4 (Maximum likelihood for compound Poisson process (M3))
Given an observation period $[0, T]$ with $\hat{m} \in \mathbb{N}_0$ cluster arrivals with sizes $\hat{\boldsymbol{y}}_1, \dots, \hat{\boldsymbol{y}}_{\hat{m}}$, the estimators $\hat{\boldsymbol{\lambda}}^{M3}$ of the intensities $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\theta}}^{M3}$ of the subordinator parameters $\boldsymbol{\theta}$ in Model (M) are defined via the optimization problem

$$\{\hat{\boldsymbol{\lambda}}^{M3}, \hat{\boldsymbol{\theta}}^{M3}\} := \underset{\boldsymbol{\lambda}\in\mathbb{R}^d_{>0}, \boldsymbol{\theta}\in\Theta}{\arg\max}\ f^{M3}(\boldsymbol{\lambda}, \boldsymbol{\theta}), \tag{M3}$$

with objective function

$$f^{M3}(\boldsymbol{\lambda}, \boldsymbol{\theta}) := -T\Psi_\Lambda(|\boldsymbol{\lambda}|; \boldsymbol{\theta}) + \sum_{i=1}^{d} \hat{c}_i \ln\{\lambda_i\} + \sum_{j=1}^{\hat{m}} \ln\{(-1)^{|\hat{\boldsymbol{y}}_j|+1} \Psi_\Lambda^{(|\hat{\boldsymbol{y}}_j|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta})\}.$$

Solving this optimization numerically is more convenient than in case of the maximum likelihood estimation in Method (M3), as the objective function depends on the derivatives of the Laplace exponent of the subordinator rather than the Laplace transform, which are mostly available in closed form (cf. Section 4.2).

The second approach we study for continuous monitoring is based on the Lévy measure $\nu_{\boldsymbol{L}}$ of $\boldsymbol{L}$. It follows from the space-time decomposition property of the compound Poisson process (see Remark 2.8 and Remark 4.18) that clusters of all sizes $\boldsymbol{k} \in \dot{\mathbb{N}}_0^d$ arrive as independent Poisson processes $N(\boldsymbol{k})$ with intensity $\nu_{\boldsymbol{L}}(\boldsymbol{k}; \boldsymbol{\lambda}, \boldsymbol{\theta})$. The moment and maximum likelihood estimator for the intensity parameter of a Poisson process is the final process level normalized with the length of the observation period, i.e. for a cluster of size $\boldsymbol{k}$ the estimator is

$$\frac{\hat{N}_T(\boldsymbol{k})}{T} = \frac{|\{j = 1, \ldots, \hat{m} : \hat{\boldsymbol{y}}_j = \boldsymbol{k}\}|}{T},$$

cf. the proof of Theorem 5.3 or see, for instance, (Davison, 2003, Chapter 6.5.1, p.277)). Hence, we may estimate the parameters of Model (M) such that the distance between observed and model predicted intensities of cluster arrivals of all sizes is minimized in a least-square sense, which equates to the minimum distance estimator for the Lévy measure:

$$\underset{\boldsymbol{\lambda}\in\mathbb{R}^d_{>0}, \boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min} \sum_{\boldsymbol{k}\in\dot{\mathbb{N}}_0^d} \left(\nu_{\boldsymbol{L}}(\boldsymbol{k}; \boldsymbol{\lambda}, \boldsymbol{\theta}) - \frac{|\{j = 1, \ldots, \hat{m} : \hat{\boldsymbol{y}}_j = \boldsymbol{k}\}|}{T}\right)^2.$$

The infinite summation needs to be truncated at some point. Using the information about the tail of the Lévy measure, see Corollary 4.26, a maximum level $K \in \mathbb{N}$ for considering the cluster sizes $|\boldsymbol{k}| \leq K$ individually can be selected and a final term for the tail can be added. The approach is made explicit in the following using Theorem 4.17 and Corollary 4.26.

Estimation 5.5 (Intensity-based method (M4))
Given an observation period $[0,T]$ with $\hat{m} \in \mathbb{N}_0$ cluster arrivals of sizes $\hat{\boldsymbol{y}}_1, \dots, \hat{\boldsymbol{y}}_{\hat{m}}$, the estimators $\hat{\boldsymbol{\lambda}}^{M4}$ of the intensities $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\theta}}^{M4}$ of the subordinator parameters $\boldsymbol{\theta}$ in Model (M) are defined via the optimization problem

$$\{\hat{\boldsymbol{\lambda}}^{M4}, \hat{\boldsymbol{\theta}}^{M4}\} := \underset{\boldsymbol{\lambda}\in\mathbb{R}^d_{>0},\, \boldsymbol{\theta}\in\Theta}{\arg\min}\; f^{M4}(\boldsymbol{\lambda}, \boldsymbol{\theta}), \tag{M4}$$

with objective function

$$\begin{aligned} f^{M4}(\boldsymbol{\lambda}, \boldsymbol{\theta}) := \sum_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d \\ |\boldsymbol{k}|\le K}} \Big(- \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!} \Psi_\Lambda^{(|\boldsymbol{k}|)}(|\boldsymbol{\lambda}|; \boldsymbol{\theta}) - \frac{|j = 1, \dots, \hat{m} : \{\hat{\boldsymbol{y}}_j = \boldsymbol{k}\}|}{T} \Big)^2 \\ + \Big(- R^n \Psi_\Lambda(0; |\boldsymbol{\lambda}|; \boldsymbol{\theta}) - \frac{|\{j = 1, \dots, \hat{m} : |\hat{\boldsymbol{y}}_j| > K\}|}{T} \Big)^2, \end{aligned}$$

where $R^n\Psi_\Lambda(0; |\boldsymbol{\lambda}|; \boldsymbol{\theta})$ is the residual of a Taylor expansion of Ψ_Λ in $|\boldsymbol{\lambda}|$ evaluated in zero.

As for all methods discussed so far, the optimization has to be performed numerically. This method quickly becomes computationally expensive for increasing dimension (and expected claim numbers) due to the rising number of tuples $\boldsymbol{k}$ that have to be considered. Thus, the threshold K needs to be selected such as to achieve a reasonable trade-off between fitting the tail and guaranteeing tractability of the method. A possible approach that will be explored in Section 5.2, is to set K as the aggregate sample mean plus one standard deviation of the observed cluster sizes:

$$K := \Big\lfloor \frac{1}{\hat{m}} \hat{\boldsymbol{L}}_T + \Big(\frac{1}{\hat{m} - 1} \sum_{j=1}^{\hat{m}} (\hat{\boldsymbol{y}}_j - \hat{\boldsymbol{L}}_T)^2 \Big)^{\frac{1}{2}} \Big\rfloor, \tag{5.2}$$

where $\hat{\boldsymbol{L}}_T = \hat{m}^{-1} \sum_{j=1}^{\hat{m}} \hat{\boldsymbol{y}}_j$ is the process level at the end of the observation period and $\lfloor \cdot \rfloor$ denotes the floor function. If K is still high, a boundary for the left tail may be considered as well; it should, however, be made sure that solitary jumps of individual components are still included as these contain valuable information about the dependence structure. The approach can also be modified to consider only selected cluster sizes which are of most importance for the specific application, while disregarding the rest in the estimation. If the observation period is short, the number of observations for individual cluster sizes will be small; particularly in high dimensional applications with many

expected claims, where a huge number of cluster sizes are possible but appear with low intensities, many cluster sizes will not occur at all and their estimated intensities will be zero. Thus, a sufficiently long observation period may be crucial for the performance of this method.

Finally, it is worth mentioning that both Methods (M3) and (M4) depend only on the length of the observation period T, the number of cluster arrivals $\hat{m}$, and the observed cluster sizes $\hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_{\hat{m}}$; the cluster arrival times $\hat{w}_1, \ldots, \hat{w}_{\hat{m}}$ are not explicitly necessary for the calculation of these estimators.

5.2 Simulation study

The performance of the estimation procedures introduced in the previous section will now be investigated in a simulation study. For different specifications of the model in dimension $d = 3$, $m = 500$ sample paths of the claim number process $\boldsymbol{L}$ are generated using Algorithm 3.3. Estimators $\hat{\boldsymbol{\lambda}}_j$ and $\hat{\boldsymbol{\theta}}_j$ from all Methods (M1)–(M4) are calculated in each scenario $j = 1, \dots, m$ and Monte Carlo estimates for expectation, standard deviation (std), relative bias (rbias), and root-mean-square error (rmse) of the distribution of the estimators are computed as sample means, e.g. for $\hat{\lambda}_1$:

$$\hat{\mathbb{E}}[\hat{\lambda}_1] := \frac{1}{m}\sum_{j=1}^{m} \lambda_{1j}, \qquad \hat{\mathrm{std}}[\hat{\lambda}_1] := \Big(\frac{1}{m-1}\sum_{j=1}^{m}(\hat{\lambda}_{1j} - \hat{\mathbb{E}}[\hat{\lambda}_1])^2\Big)^{\frac{1}{2}},$$
$$\hat{\mathrm{rbias}}[\hat{\lambda}_1] := \frac{\hat{\mathbb{E}}[\hat{\lambda}_1] - \lambda_1}{\lambda_1}, \qquad \hat{\mathrm{rmse}}[\hat{\lambda}_1] := \Big(\frac{1}{m}\sum_{j=1}^{m}(\hat{\lambda}_{1j} - \lambda_1)^2\Big)^{\frac{1}{2}}.$$

To gain insight into the asymptotical properties of the estimators, two time horizons $T \in \{1, 10\}$ are compared for each setting. In particular, we want to find out if the estimators appear to be consistent, i.e. converge to the true parameters, and asymptotically normal, which would enable the construction of confidence intervals.

Two different Lévy subordinators are chosen for the time-change, an inverse Gaussian and a gamma subordinator, as the resulting Sichel and inverse Gaussian distributions are popular choices for insurance count data with overdispersion. Since the results are quite similar, the inverse Gaussian case is presented in detail here and the data for the gamma case are collected in Appendix A. As was discussed after Algorithm 3.3, the inverse Gaussian and the gamma subordinator both are of infinite activity and need to be sampled on a discrete time-grid, which leads to some bias towards overestimating the jump sizes of the time-changed process. Thus, some (small) portion of the error of the estimators from Methods (M3) and (M4) has to be attributed to the simulation routine rather than the estimation procedures. However, the influence of the discretization error should be low, since a very fine grid with one million steps per time unit was chosen for sampling the subordinator paths. For Methods (M1) and (M2), the sampled paths of the time-changed process are discretized with 365 steps per time unit, corresponding to daily observations of the process.

For both subordinator choices, two distinct settings are investigated, one with high and another with low dependence and variance in the time-changed model. More precisely, the parameters of the subordinators are set such that the subordinator variance is approximately $\mathbb{V}\mathrm{ar}[\Lambda_1] \approx 0.05$ and $\mathbb{V}\mathrm{ar}[\Lambda_1] \approx 0.005$, respectively; then it follows from Theorem 4.15 in the first case (low dependence and variance):

$$\mathbb{V}\mathrm{ar}[\boldsymbol{L}_1] \approx (63, 103, 150)', \quad \mathbb{C}\mathrm{or}[\boldsymbol{L}_1] \approx \begin{pmatrix} 1 & 0.23 & 0.26 \\ 0.23 & 1 & 0.30 \\ 0.26 & 0.30 & 1 \end{pmatrix},$$

and in the second case (strong dependence and high variance):

$$\mathbb{V}\mathrm{ar}[\boldsymbol{L}_1] \approx (175, 356, 600)', \quad \mathbb{C}\mathrm{or}[\boldsymbol{L}_1] \approx \begin{pmatrix} 1 & 0.75 & 0.77 \\ 0.75 & 1 & 0.81 \\ 0.77 & 0.81 & 1 \end{pmatrix}.$$

In addition, for both subordinator types and subordinator variance settings, a model with no subordinator drift and another with drift $b = 0.4$ is examined. Assumption (TN) is always imposed by fixing one subordinator parameter, i.e.

$$\eta := \beta \quad \text{or} \quad \eta := \frac{\beta}{1-b}$$

in case of no drift or with drift b, respectively, for both subordinator types (cf. Table 4.4). Thus, the total number of parameters to be estimated is four in the models without drift and five otherwise, with subordinator parameter vectors $\boldsymbol{\theta} = \beta$ and $\boldsymbol{\theta} = (b, \beta)'$, respectively. The admissible set Θ in the former case is $\mathbb{R}_{>0}$, and in the latter due to time-normalization ($\eta = \beta/(1-b) > 0$ needs to hold) $[0, 1) \times \mathbb{R}_{>0}$. The intensity vector is set to $\boldsymbol{\lambda} := (50, 75, 100)'$. Since time-normalization is assumed, the intensities correspond to the expected number of claims per time unit. All subordinator parameter settings as well as the corresponding jump intensities and expected jump sizes in the time-changed model are summarized in Table 5.1. In the case of low subordinator variance, the jump intensities are higher and the jump sizes on average smaller than in the high variance case, leading to more observation points for Methods (M3) and (M4). Adding a drift has the same effect, i.e. the jump intensities increase and the average jump sizes decrease.

Simulation and optimization are implemented in MATLAB® and performed on a standard computer (3.33 GHz Intel Core i5 processor, 4.00 GB RAM). Multidimensional constrained optimization is a highly non-trivial problem and topic of ongoing research, but a discussion of optimization algorithms and their convergence properties is beyond the scope of this thesis and the study relies on the build-in 'fmincon' function with 'sqp solver' for constrained non-linear optimization problems. The termination settings for the solver are left at their default values, e.g. the termination tolerance for the parameters as well as the objective function is 10^{-6}. The initial intensity values for the solver are set to the sample mean; for the subordinator parameters, the initial values are calculated via moment matching for the aggregate claim number process, as discussed in the following remark.

subordinator	drift	$\mathbb{V}\mathrm{ar}$	b	β	η	$\mathbb{E}[\boldsymbol{Y};\boldsymbol{\theta}]$	$\Psi_\Lambda(\lvert\boldsymbol{\lambda}\rvert;\boldsymbol{\theta})$
inverse Gaussian	no	low	0	14.5	14.5	$(0.31, 0.46, 0.62)'$	162.33
		high	0	4.5	4.5	$(0.65, 0.97, 1.29)'$	77.33
	yes	low	0.4	6.9	11.5	$(0.28, 0.42, 0.56)'$	177.15
		high	0.4	2.1	3.5	$(0.39, 0.59, 0.78)'$	127.80
gamma	no	low	0	210	210	$(0.33, 0.49, 0.65)'$	152.93
		high	0	21	21	$(0.97, 1.45, 1.94)'$	51.68
	yes	low	0.4	78	130	$(0.30, 0.45, 0.59)'$	168.36
		high	0.4	7.5	12.5	$(0.45, 0.67, 0.89)'$	112.08

Table 5.1 Parameter settings for the simulation study: The table summarizes the parameters for the inverse Gaussian and gamma subordinator in the settings with and without drift as well as with high and low subordinator variance. In addition, the expected jump sizes and jump intensities in each setting are given.

Remark 5.6 (Initial parameters for the optimization algorithm)
Let $\hat{\boldsymbol{l}}_0 := \mathbf{0} \leq \hat{\boldsymbol{l}}_1 \leq \cdots \leq \hat{\boldsymbol{l}}_n$ for $n := 365T$ be the discrete observations of $\boldsymbol{L}$ in any single Monte Carlo run. Then for all four estimation methods, the initial intensity parameters in the optimization algorithm are set to the sample mean:

$$\hat{\boldsymbol{\lambda}}_0 := \frac{1}{T}\hat{\boldsymbol{l}}_n.$$

The initial subordinator parameters $\hat{\boldsymbol{\theta}}_0$ are determined from central moment matching of the aggregate process. The models with no drift have only a one subordinator parameter, hence, the initial value is found as the solution to the variance condition:

$$\mathbb{V}\mathrm{ar}[\bar{L}_1; \hat{\boldsymbol{\lambda}}_0, \boldsymbol{\theta}] \overset{!}{=} \frac{1}{n}\sum_{j=1}^{n}(|\hat{\boldsymbol{l}}_j| - |\hat{\boldsymbol{\lambda}}_0|)^2 =: zm_2.$$

Since the models with drift feature two subordinator parameters, an additional equation for the third central moment is added:

$$\mathbb{E}[(\bar{L}_1 - \mathbb{E}[\bar{L}_1])^3; \hat{\boldsymbol{\lambda}}_0, \boldsymbol{\theta}] \overset{!}{=} \frac{1}{n}\sum_{j=1}^{n}(|\hat{\boldsymbol{l}}_j| - |\hat{\boldsymbol{\lambda}}_0|)^3 =: zm_3.$$

The equations can be solved analytically using Theorem 4.15 and Table 4.4; the results are summarized in Table 5.2. However, the existence of a solution is not guaranteed and,

in addition, the condition $\hat{\boldsymbol{\theta}}_0 \in \Theta$ needs to be fulfilled for the initial values to be valid. If the calculated solutions are not admissible, moment matching for the sum of any two out of the three components of $\boldsymbol{L}$, or even any individual component, may be given a try. If neither leads to a valid initial parameter vector, some default values need to be chosen. In the model without drift, if the solution for β is invalid, the initial value is set to $\hat{\beta}_0 = 10$ in the following study. In the model with drift, if the estimator for the drift is invalid, the default is $\hat{b}_0 = 0.5$, and if the estimator for β is invalid, the default is $\hat{\beta}_0 = 5$. ▲

		inverse Gaussian	gamma
no drift	$\hat{\beta}_0$	$\lvert\hat{\boldsymbol{\lambda}}_0\rvert\left(\frac{h}{zm_2 - h\lvert\hat{\boldsymbol{\lambda}}_0\rvert}\right)^{\frac{1}{2}}$	$\frac{h\lvert\hat{\boldsymbol{\lambda}}_0\rvert^2}{zm_2 - h\lvert\hat{\boldsymbol{\lambda}}_0\rvert}$
with drift	$\hat{b}_0$	$1 - \frac{3(zm_2 - h\lvert\hat{\boldsymbol{\lambda}}_0\rvert)^2}{h\lvert\hat{\boldsymbol{\lambda}}_0\rvert(zm_3 + 2h\lvert\hat{\boldsymbol{\lambda}}_0\rvert - 3zm_2)}$	$1 - \frac{2(zm_2 - h\lvert\hat{\boldsymbol{\lambda}}_0\rvert)^2}{h\lvert\hat{\boldsymbol{\lambda}}_0\rvert(zm_3 + 2h\lvert\hat{\boldsymbol{\lambda}}_0\rvert - 3zm_2)}$
	$\hat{\beta}_0$	$\frac{3^{\frac{3}{2}}(zm_2 - h\lvert\hat{\boldsymbol{\lambda}}_0\rvert)^{\frac{5}{2}}}{h\lvert\hat{\boldsymbol{\lambda}}_0\rvert^{\frac{1}{2}}(zm_3 + 2h\lvert\hat{\boldsymbol{\lambda}}_0\rvert - 3zm_2)^{\frac{3}{2}}}$	$\frac{4(zm_2 - h\lvert\hat{\boldsymbol{\lambda}}_0\rvert)^3}{h(zm_3 + 2h\lvert\hat{\boldsymbol{\lambda}}_0\rvert - 3zm_2)^2}$

Table 5.2 Initial values of subordinator parameters for the optimization: For an inverse Gaussian and gamma subordinator, both with and without drift, and assuming time-normalization, the table presents the initial subordinator parameters for the optimization algorithm that are generated from a moment estimation approach for the aggregate process $\bar{L}$.

Finally, the moments of the process distribution for Method (M1) as well as the jump size distribution and Lévy measure for Methods (M3) and (M4) are directly implemented using the results derived for the two chosen subordinators in Chapter 4. For Method (M4), the threshold K is selected according to Equation (5.2). The process distribution for Method (M2) is implemented using Panjer's recursion as derived in Corollary 4.29.

Setting: inverse Gaussian subordinator (no drift)

The initial parameters calculated as discussed in Remark 5.6 are valid input parameters for the optimization in all Monte Carlo runs – for the short and long time-horizon as well as for the low and high variance setting. Table 5.3 summarizes the runtimes of the optimization procedures and the objective function values (ofv) for all four estimation methods. The ofv for the estimators is always better (i.e. smaller for Methods (M1) and (M4), larger for Methods (M2) and (M3)) than for the true parameters and improves in the order of ten with increasing time horizon of the same order. As to be expected, Method (M1) has by far the shortest runtime and Method (M3) is always considerably faster than Method (M2). For increasing time horizon, the computation times for Methods (M1) and (M4) remain approximately the same, whereas for Methods (M2)

and (M3) computation takes considerably longer due to the increasing number of terms in the objective function. Hence, while Method (M2) is quicker than Method (M4) for a short time-horizon, Method (M4) is much quicker in the long run. In the case of high subordinator variance, runtimes for Methods (M1) and (M2) are almost unchanged compared to the low variance case (even slightly faster); Method (M3), however, gets faster due to the on average lower number of observations and Method (M4) slows down as the expected jump size and, thus, the number of terms in the objective function increases.

		(M1)		(M2)		(M3)		(M4)	
T		1	10	1	10	1	10	1	10
runtime	low	0.456	0.423	39.5	227	9.45	60.8	75.1	73.8
	high	0.439	0.413	34.3	211	2.95	30.8	127	118
ofv	low	0.268	0.0234	-574	-5750	337	3340	1.78	0.164
	high	0.0559	0.00617	-413	-4180	13.9	125	0.257	0.0201

Table 5.3 Runtime and ofv for inverse Gaussian subordinator (no drift): The table summarizes runtimes and objective function values (ofv) for the estimation Methods (M1)–(M4) in the setting with an inverse Gaussian subordinator (without drift) with low and high variance and for short and long time horizon. The runtime is expressed in minutes and all quantities are given with 3 significant digits.

The estimation results for the low and high variance setting are summarized in Tables 5.4 and 5.5, respectively. Boxplots can be found in Figures A.1–A.4 in the Appendix A. Overall, the estimators are on average close to the true parameter values, i.e. the relative bias is small. The highest deviations occur for the estimators of the subordinator parameter β for the short time horizon, in particular for Method (M1). The rbias for the subordinator parameter increases in the high variance case for Methods (M1)–(M3), whereas for the intensities the deviations tend to decrease. For Method (M4) it is often the other way round.

The estimated parameters from Methods (M2) and (M3) are very close to each other, in particular for the intensities. The root-mean-squared difference between the estimators from these two methods are given in Table 5.6 (see Table A.1 in Section A for the other methods). There it can be seen that the estimators for the intensities are almost identical; more precisely, they are both very close to the moment estimators of the intensities that were chosen as initial values. For the subordinator parameter β, the difference is more pronounced, in particular for the short time horizon in the low variance setting. In terms of the rmse, $\hat{\beta}^{M3}$ tends to be closer to the real parameter than $\hat{\beta}^{M2}$.

The rmse of all estimates is lowest for Method (M3). In the low variance setting,

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 14.5$	
T		1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	51.899	50.270	78.803	75.434	104.908	100.654	15.524	14.615
	(M2)	50.224	50.196	75.868	75.030	100.386	100.184	14.786	14.539
	(M3)	50.224	50.197	75.868	75.030	100.386	100.184	14.699	14.510
	(M4)	49.993	50.252	75.745	75.036	99.983	100.175	14.743	14.508
$\hat{\text{std}}$	(M1)	10.769	3.300	14.608	4.450	22.020	6.134	2.894	0.891
	(M2)	7.759	2.540	9.559	3.104	12.032	3.871	1.992	0.659
	(M3)	7.759	2.540	9.559	3.104	12.032	3.871	1.598	0.525
	(M4)	9.720	3.069	12.150	3.782	13.693	4.531	1.775	0.562
$\hat{\text{rbias}}$ (%)	(M1)	3.798	0.540	5.071	0.578	4.908	0.654	7.062	0.794
	(M2)	0.448	0.393	1.157	0.040	0.386	0.184	1.971	0.268
	(M3)	0.448	0.393	1.157	0.040	0.386	0.184	1.375	0.068
	(M4)	−0.015	0.503	0.994	0.049	−0.017	0.175	1.679	0.054
$\hat{\text{rmse}}$	(M1)	10.935	3.311	15.095	4.472	22.561	6.168	3.070	0.898
	(M2)	7.763	2.548	9.599	3.104	12.038	3.875	2.013	0.660
	(M3)	7.762	2.548	9.598	3.104	12.038	3.875	1.610	0.525
	(M4)	9.720	3.079	12.173	3.782	13.693	4.535	1.792	0.562

Table 5.4 Estimation results for low variance inverse Gaussian subordinator (no drift): The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the low variance inverse Gaussian subordinator without drift for short and long time horizon.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 4.5$	
T		1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	50.221	50.116	76.385	75.259	101.602	100.001	5.151	4.601
	(M2)	49.945	50.063	75.368	75.028	99.778	99.833	4.645	4.520
	(M3)	49.944	50.066	75.367	75.032	99.777	99.838	4.616	4.514
	(M4)	53.110	50.113	79.952	75.334	105.965	100.479	4.555	4.497
$\hat{\text{std}}$	(M1)	14.357	4.551	21.878	6.360	26.817	8.304	1.110	0.414
	(M2)	13.840	4.135	20.295	5.827	25.385	7.871	0.811	0.246
	(M3)	13.839	4.136	20.295	5.828	25.385	7.872	0.767	0.236
	(M4)	24.928	6.168	34.717	8.277	44.161	11.034	1.041	0.291
$\hat{\text{rbias}}$ (%)	(M1)	0.441	0.233	1.847	0.346	1.602	0.001	14.475	2.234
	(M2)	−0.110	0.127	0.491	0.038	−0.222	−0.167	3.222	0.441
	(M3)	−0.111	0.132	0.490	0.043	−0.223	−0.162	2.580	0.312
	(M4)	6.220	0.226	6.602	0.446	5.965	0.479	1.230	−0.073
$\hat{\text{rmse}}$	(M1)	14.359	4.552	21.922	6.365	26.864	8.304	1.287	0.426
	(M2)	13.840	4.136	20.298	5.827	25.385	7.873	0.823	0.247
	(M3)	13.840	4.137	20.298	5.828	25.386	7.874	0.776	0.236
	(M4)	25.122	6.169	35.068	8.284	44.562	11.044	1.042	0.291

Table 5.5 Estimation results for high variance inverse Gaussian subordinator (no drift): The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the high variance inverse Gaussian subordinator without drift for short and long time horizon.

Method (M1) consistently performs worst; in the high variance setting, however, the Method (M4) has higher rmse for the intensities. Method (M2) outperforms Methods (M1) and (M4) where the intensities are concerned; for the subordinator parameter, Method (M2) does better than Method (M1), but Method (M4) has a lower rmse in case of low subordinator variance. For increasing time horizon from one to ten, the rmse overall tends to decrease roughly as $\sqrt{10}$, i.e. it increases in square-root w.r.t. time. For the intensity estimators of Method (M4) in the high variance setting, which are rather off for the short time horizon, the improvement is stronger. Finally, the rmse for the intensities is smaller in the low variance setting, but for the subordinator parameter all methods perform better in the high variance case. For an additional comparison in terms of the Pitman closeness criterion, i.e. the probability that the absolute deviation for one method is smaller than for another, see Table A.2.

	λ_1		λ_2		λ_3		β	
T	1	10	1	10	1	10	1	10
low	0.000797	0.00209	0.00120	0.00262	0.00165	0.00314	1.21	0.384
high	0.00302	0.00478	0.00478	0.00708	0.00595	0.00926	0.222	0.0638

Table 5.6 Difference between estimators from Methods (M2) and (M3) for inverse Gaussian subordinator (no drift): The table presents the root-mean- squared difference between the estimators from the two Methods (M2) and (M3) in case of an inverse Gaussian subordinator without drift in the setting with low and high variance as well as for short and long time horizon (3 significant digits).

The subordinator parameter is mainly responsible for the degree of dependence in the model and, in general, strong positive dependence often aids the estimation of the respective parameters. In (Joe, 2014, Chapter 5.7, pp.234), the Kullback–Leibler divergence, which measures the difference between two densities, is employed to determine the sample size necessary to distinguish between different models with a given high probability. If applied to copula models, it is shown that for weak positive dependence this sample size is higher, i.e. it is harder to discriminate between models. In addition, it is also pointed out that discrete distributions overall require a much larger sample size than continuous ones.

Table 5.7 presents the p-values of a chi-squared goodness-of-fit test for normality of the estimators. For the short time-horizon $T = 1$, the null hypothesis of normality has to be rejected for a significance level $\alpha = 5\%$ in many cases. For low subordinator variance, the intensity estimators of Method (M4) and two of the intensity estimators of Methods (M2) and (M3) test positive. In the high variance case, the hypothesis of normality has to be rejected in all cases but for $\hat{\beta}^{M2}$. For the longer time-horizon $T = 10$, however, the null hypothesis cannot be rejected for any estimator. Figures 5.1–5.4 show qq-plots for all estimators to illustrate how closely their distributions resemble the normal distri-

bution: for $T = 10$, the estimator quantiles closely match the normal quantiles, whereas for $T = 1$ deviations still appear in the tails, in particular in the high variance setting. These deviations are not surprising, considering that the parameter space is restricted to the positive line: the variance of the estimators needs to be small in order for the distribution to be close to the normal distribution.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 14.5/4.5$	
T		1	10	1	10	1	10	1	10
(M1)	low	0.00	0.28	0.03	0.07	0.00	0.66	0.00	0.13
	high	0.00	0.33	0.00	0.24	0.00	0.59	0.00	0.75
(M2)	low	0.03	0.42	0.35	0.54	0.83	0.66	0.00	0.41
	high	0.00	0.33	0.00	0.26	0.00	0.62	0.07	0.69
(M3)	low	0.03	0.42	0.35	0.54	0.83	0.66	0.00	0.43
	high	0.00	0.34	0.00	0.22	0.00	0.54	0.00	0.59
(M4)	low	0.17	0.25	0.33	0.54	0.76	0.18	0.00	0.15
	high	0.00	0.14	0.00	0.15	0.00	0.36	0.01	0.22

Table 5.7 Chi-squared test for inverse Gaussian subordinator (no drift): The table presents the p-values of a chi-squared goodness-of-fit test for normality of the estimators in case of the inverse Gaussian subordinator without drift in the setting with low and high variance as well as for short and long time horizon.

Setting: inverse Gaussian subordinator with drift

In the case of the model specified with an inverse Gaussian subordinator with drift, the initial parameters calculated as discussed in Remark 5.6 turn out to be invalid in about 20% (low variance) and 40% (high variance) of the simulation runs for $T = 1$, mainly due to the requirements on the drift. For $T = 10$, the parameters only fail once and ten times in the 500 runs, respectively. However, the ofv after the optimization is again in all scenarios and for all estimation methods better than for the true parameters, also in case of default initial values. Compared to the setting without drift, the runtime of the optimization routines increases, but the qualitative behaviour remains the same. The values can be found in Table 5.8.

Estimated mean, std, rmse, and bias of the estimators are summarized in Tables 5.9 and 5.10. For boxplots, see Figures A.5–A.8. In the low variance case, the quality of the intensity estimators is not too different from the setting without drift; for the subordinator parameter β, however, rbias and rmse mostly increase. In particular, the rbias for Methods (M2)–(M4) for $T = 1$ is considerably higher, roughly of order 10 or more, and

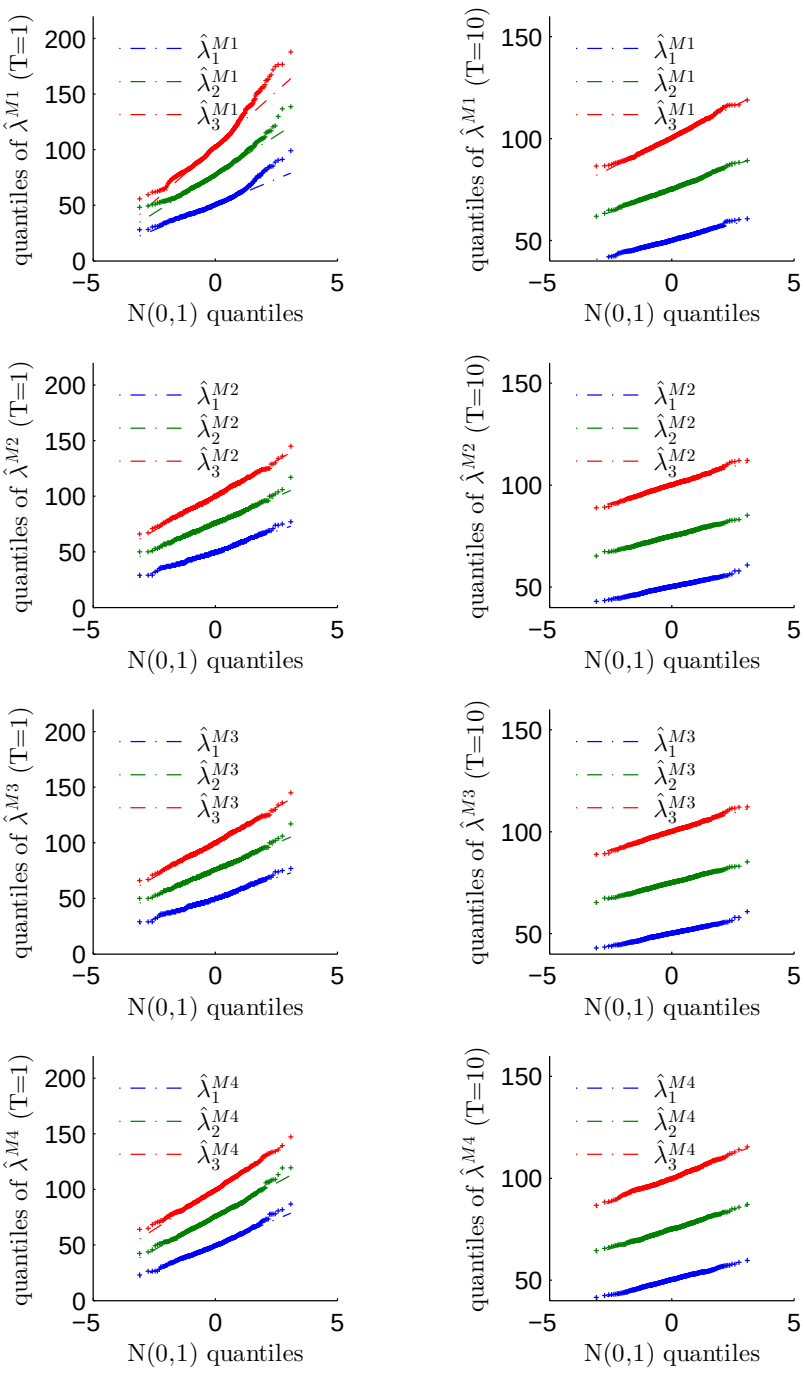


Figure 5.1 qq-plots of intensity estimators for low variance inverse Gaussian subordinator (no drift): The figure illustrates qq-plots of the intensity estimators in case of the low variance inverse Gaussian subordinator without drift for short and long time horizon.

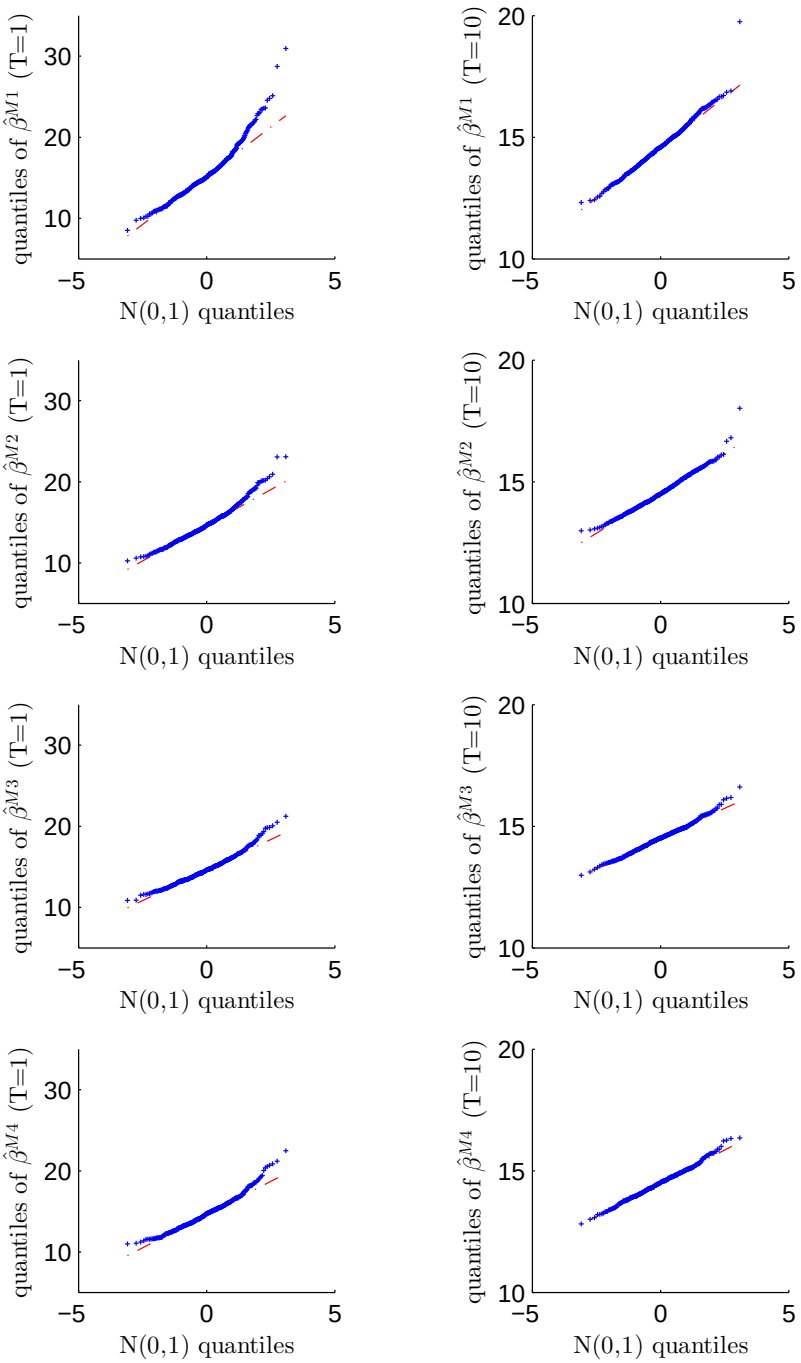


Figure 5.2 qq-plots of subordinator estimators for low variance inverse Gaussian subordinator (no drift): The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the low variance inverse Gaussian subordinator without drift for short and long time horizon.

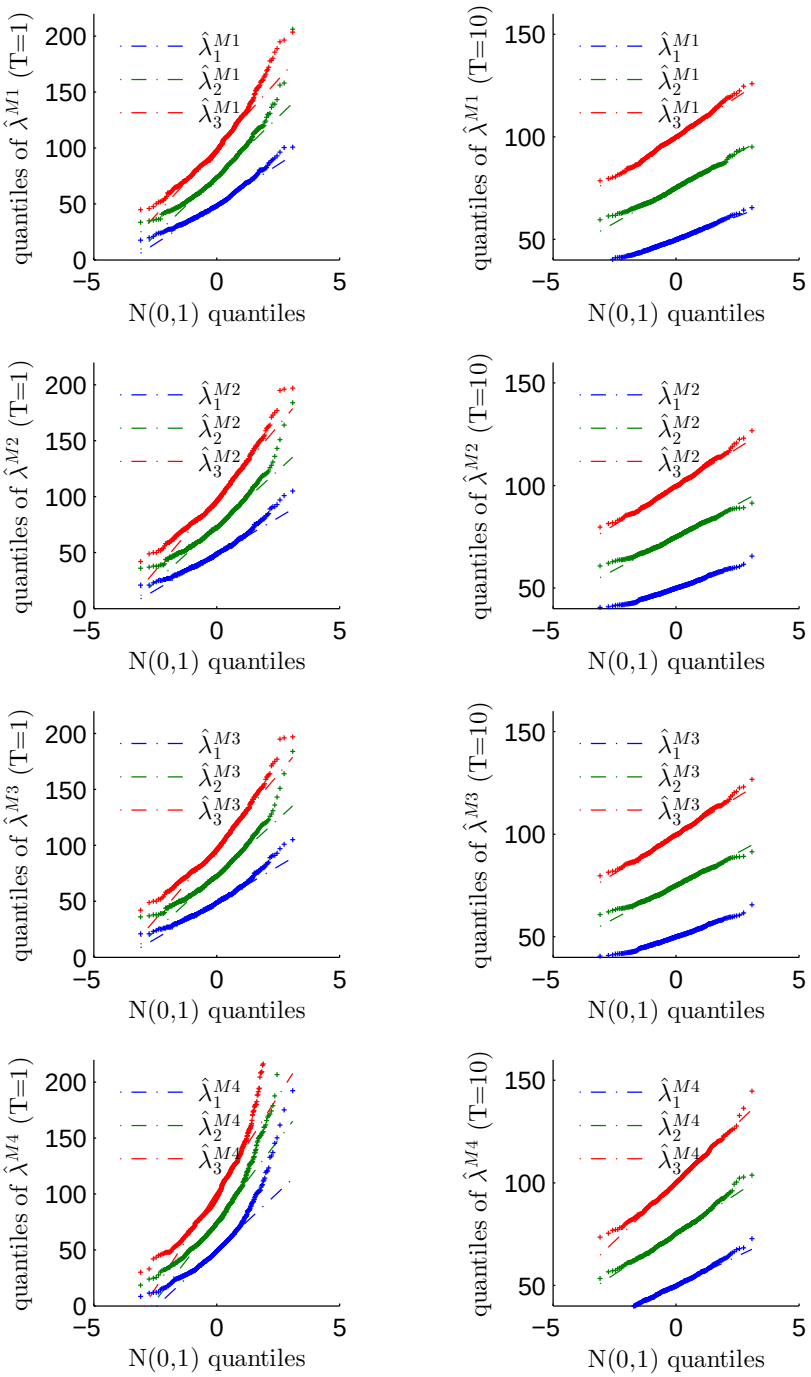


Figure 5.3 qq-plots of intensity estimators for high variance inverse Gaussian subordinator (no drift): The figure illustrates qq-plots of the intensity estimators in case of the high variance inverse Gaussian subordinator without drift for short and long time horizon.

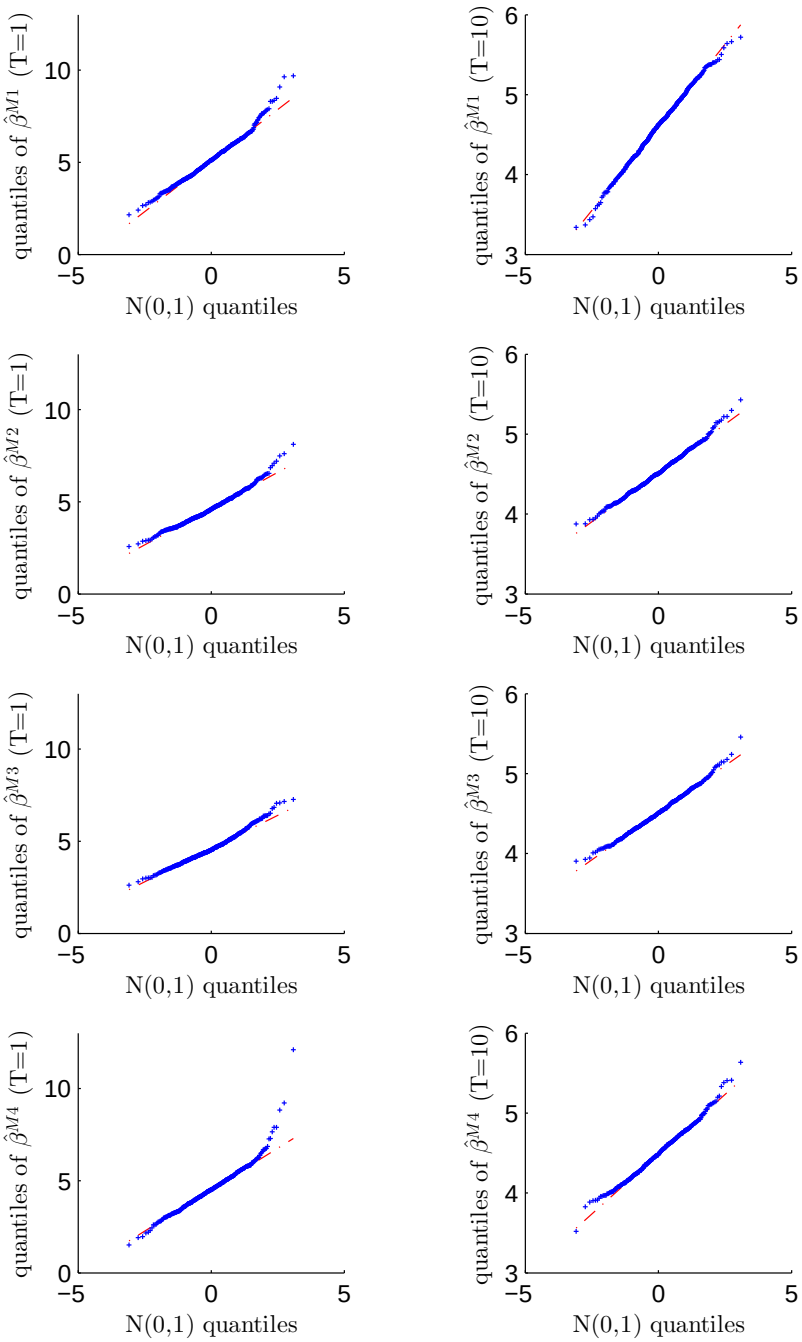


Figure 5.4 qq-plots of subordinator estimators for high variance inverse Gaussian subordinator (no drift): The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the high variance inverse Gaussian subordinator without drift for short and long time horizon.

		(M1)		(M2)		(M3)		(M4)	
T		1	10	1	10	1	10	1	10
runtime	low	0.709	0.526	70.1	415	17.4	118	116	114
	high	0.727	0.526	75.2	597	13	104	209	154
ofv	low	0.385	0.0273	-575	-5780	430	4290	0.872	0.0849
	high	0.128	0.00893	-482	-4860	246	2460	0.195	0.0144

Table 5.8 Runtime and ofv for inverse Gaussian subordinator with drift: The table summarizes runtimes and objective function values (ofv) for the estimation Methods (M1)–(M4) in the setting with an inverse Gaussian subordinator with drift, low as well as high variance and for short and long time horizon. The runtime is expressed in minutes and all quantities are given with 3 significant digits.

for Method (M1) the bias even increases for the longer time horizon. For the drift, the rbias of Methods (M1) and (M2) is rather high (approximately 14% in the former and -15% in the latter case) for $T = 1$; for Method (M2) it strongly improves for the longer time horizon, but for Method (M1) it changes to approximately -12%. The rmse of the drift, however, is low for all estimation methods.

In the setting with high variance, it is particularly obvious that Method (M1) struggles to accurately capture the subordinator parameters. Conversely, Method (M4) works well for the subordinator parameters but is far off for the intensities in the short run, though it strongly improves for the longer time horizon. This is mainly due to outliers, as can be see in the accompanying boxplot in Figure A.7. Calculating the median of the Monte Carlo samples as a robust estimator rather than the mean, the values improve considerably, see Table 5.11. A robust estimator for the rmse can be computed as follows:

$$\hat{\text{rmse}}^{rob}(\hat{\lambda}_i) := \left[(\hat{\text{median}}(\hat{\lambda}_i) - \lambda_i)^2 + \left(\frac{\hat{\text{median}}(|\hat{\text{median}}(\hat{\lambda}_i) - \hat{\lambda}_i|)}{\Phi^{-1}(\frac{3}{4})}\right)^2\right]^{\frac{1}{2}},$$

where $\Phi(\cdot)$ denotes the cumulative normal distribution function. The $\hat{\text{rmse}}^{rob}$ for the intensity estimators from Method (M4) are also given in Table 5.11. Though these errors are still highest among all methods, they are much lower than before. For a detailed discussion of robust statistics see Huber and Ronchetti (2009). Methods (M2) and (M3) are again almost indistinguishable for the intensities and close for the subordinator parameters with a slight edge for Method (M3), see also Table A.3 containing the *rmse* for the differences between the estimators from all methods. The results of the Pitman closeness criterion are summarized in Table A.4.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 6.9$	
T		1	10	1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	51.687	50.068	78.463	75.803	106.385	100.704	0.455	0.351	7.262	7.976
	(M2)	49.550	49.880	74.572	75.230	100.358	100.248	0.341	0.394	8.791	7.058
	(M3)	49.550	49.880	74.572	75.230	100.358	100.248	0.370	0.395	7.950	7.002
	(M4)	50.405	49.925	76.450	75.334	102.895	100.380	0.374	0.397	7.680	6.938
$\hat{\text{std}}$	(M1)	11.772	3.522	18.639	5.016	21.939	6.587	0.282	0.107	4.936	2.159
	(M2)	7.374	2.549	10.207	2.994	11.524	3.899	0.191	0.059	4.646	1.194
	(M3)	7.374	2.550	10.207	2.994	11.524	3.899	0.116	0.035	3.135	0.820
	(M4)	9.774	2.921	14.706	3.413	18.459	4.492	0.137	0.038	4.192	0.980
$\hat{\text{rbias}}$ (%)	(M1)	3.374	0.136	4.618	1.071	6.385	0.704	13.740	−12.168	5.245	15.597
	(M2)	−0.900	−0.240	−0.571	0.306	0.358	0.248	−14.861	−1.602	27.404	2.284
	(M3)	−0.900	−0.239	−0.571	0.307	0.358	0.248	−7.385	−1.179	15.212	1.475
	(M4)	0.809	−0.150	1.933	0.446	2.895	0.380	−6.390	−0.724	11.299	0.551
$\hat{\text{rmse}}$	(M1)	11.892	3.522	18.958	5.080	22.849	6.625	0.287	0.117	4.950	2.413
	(M2)	7.388	2.552	10.216	3.003	11.529	3.907	0.200	0.060	5.016	1.205
	(M3)	7.387	2.552	10.216	3.003	11.529	3.907	0.119	0.035	3.306	0.826
	(M4)	9.782	2.922	14.777	3.430	18.684	4.508	0.140	0.038	4.264	0.980

Table 5.9 Estimation results for low variance inverse Gaussian subordinator with drift: The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the low variance inverse Gaussian subordinator with drift for short and long time horizon.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 2.1$	
T		1	10	1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	50.653	50.120	77.278	75.478	101.462	100.405	0.409	0.271	3.106	3.018
	(M2)	49.982	49.969	74.944	75.215	98.969	100.200	0.395	0.398	2.516	2.139
	(M3)	49.983	49.971	74.944	75.216	98.969	100.203	0.409	0.400	2.332	2.120
	(M4)	78.216	51.697	117.093	77.563	155.561	103.679	0.340	0.391	2.110	2.088
$\hat{\text{std}}$	(M1)	13.775	4.807	20.099	6.791	26.869	8.720	0.284	0.109	2.245	0.879
	(M2)	12.637	4.133	17.819	5.992	24.270	7.856	0.106	0.036	1.105	0.247
	(M3)	12.638	4.133	17.819	5.992	24.271	7.856	0.095	0.032	0.790	0.195
	(M4)	63.145	7.364	95.720	11.105	124.334	14.572	0.145	0.045	1.275	0.357
$\hat{\text{rbias}}$ (%)	(M1)	1.305	0.241	3.037	0.638	1.462	0.405	2.300	−32.192	47.887	43.735
	(M2)	−0.035	−0.062	−0.075	0.286	−1.031	0.200	−1.197	−0.404	19.789	1.869
	(M3)	−0.034	−0.059	−0.074	0.289	−1.031	0.203	2.221	0.048	11.027	0.956
	(M4)	56.433	3.393	56.124	3.417	55.561	3.679	−14.985	−2.143	0.458	−0.579
$\hat{\text{rmse}}$	(M1)	13.790	4.808	20.228	6.808	26.909	8.730	0.284	0.169	2.460	1.272
	(M2)	12.637	4.133	17.819	5.996	24.292	7.859	0.106	0.036	1.181	0.250
	(M3)	12.638	4.133	17.820	5.996	24.293	7.859	0.096	0.032	0.824	0.196
	(M4)	69.162	7.557	104.566	11.397	136.183	15.029	0.157	0.046	1.275	0.357

Table 5.10 Estimation results for high variance inverse Gaussian subordinator with drift: The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the high variance inverse Gaussian subordinator with drift for short and long time horizon.

The p-values of a chi-squared goodness-of-fit test for normality of the estimators are summarized in Table 5.12. For the short time horizon, the null hypothesis has to be rejected in almost all cases. For $T = 10$, the p-values mostly increase and in many cases normality can no longer be rejected for the significance level $\alpha = 5\%$. The results are confirmed in the qq-plots in Figures 5.5–5.10. However, in the low variance setting the estimators of the first two intensities from Method (M1) as well as the β estimators from Methods (M1)–(M3) still significantly deviate from the normal distribution; in the high variance setting, normality has to be rejected for the intensity estimators of Method (M4) and the β estimator of Method (M1).

	λ_1		λ_2		λ_3	
T	1	10	1	10	1	10
$\hat{\text{median}}$	56.114	51.008	80.384	75.463	107.321	101.772
$\hat{\text{rmse}}^{rob}$	26.592	6.077	35.298	9.911	44.462	12.612

Table 5.11 Estimated median and robust rmse for intensity estimators from Method (M4): The table presents the estimated median and the robustly estimated rmse for the intensity estimators from Method (M4) in the case of an inverse Gaussian subordinator with drift and high variance.

parameter		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 6.9/2.1$	
T		1	10	1	10	1	10	1	10
(M1)	low	0.00	0.03	0.00	0.03	0.00	0.16	0.01	0.02
	high	0.00	0.33	0.00	0.23	0.00	0.52	0.00	0.00
(M2)	low	0.24	0.78	0.01	0.67	0.05	0.92	0.00	0.02
	high	0.00	0.13	0.00	0.12	0.00	0.76	0.00	0.07
(M3)	low	0.20	0.89	0.00	0.67	0.04	0.93	0.00	0.00
	high	0.00	0.19	0.00	0.12	0.00	0.75	0.00	0.40
(M4)	low	0.00	0.34	0.00	0.89	0.00	0.47	0.00	0.54
	high	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13

Table 5.12 Chi-squared test for inverse Gaussian subordinator with drift: The table presents the p-values of a chi-squared goodness-of-fit test for normality of the estimators in case of the inverse Gaussian subordinator with drift in the setting with low and high variance as well as for short and long time horizon.

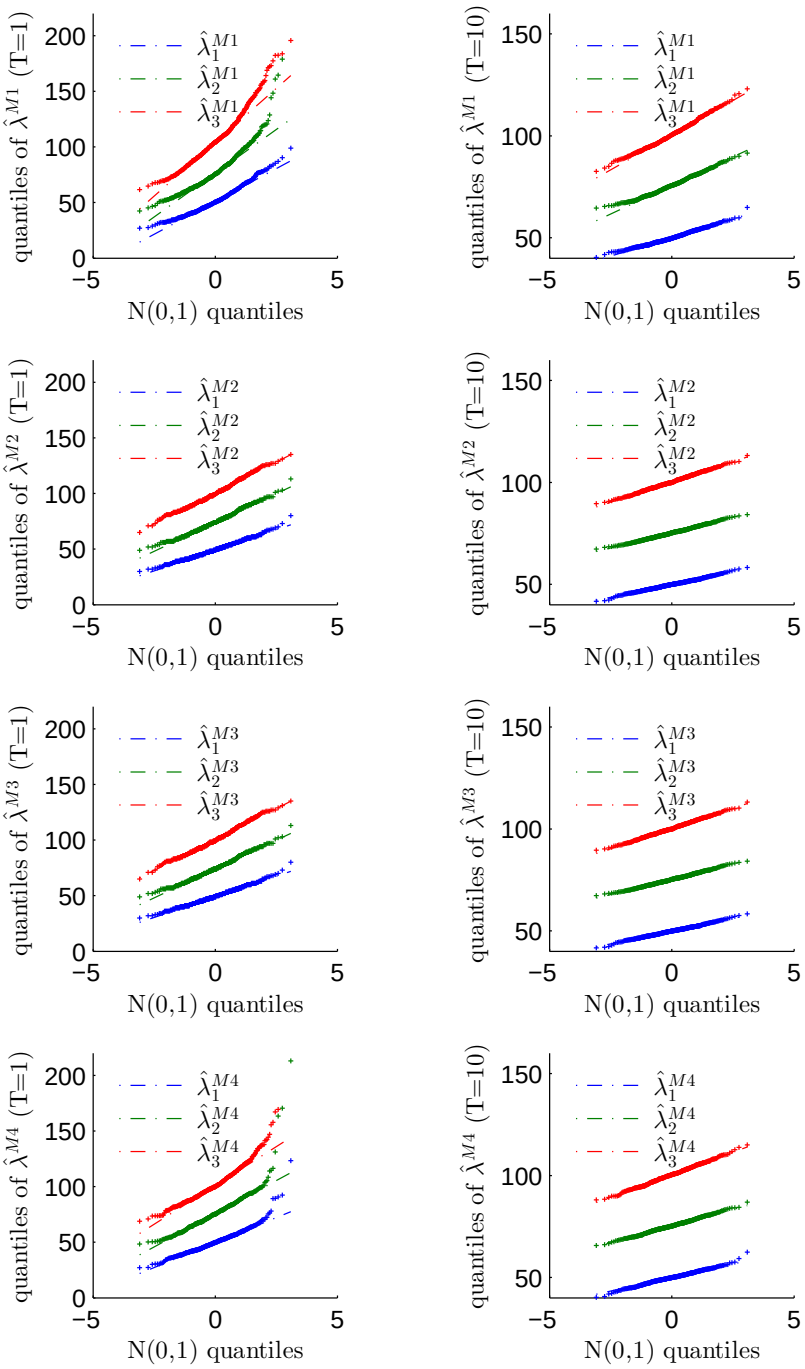


Figure 5.5 qq-plots of intensity estimators for low variance inverse Gaussian subordinator with drift: The figure illustrates qq-plots of the intensity estimators in case of the low variance inverse Gaussian subordinator with drift for short and long time horizon.

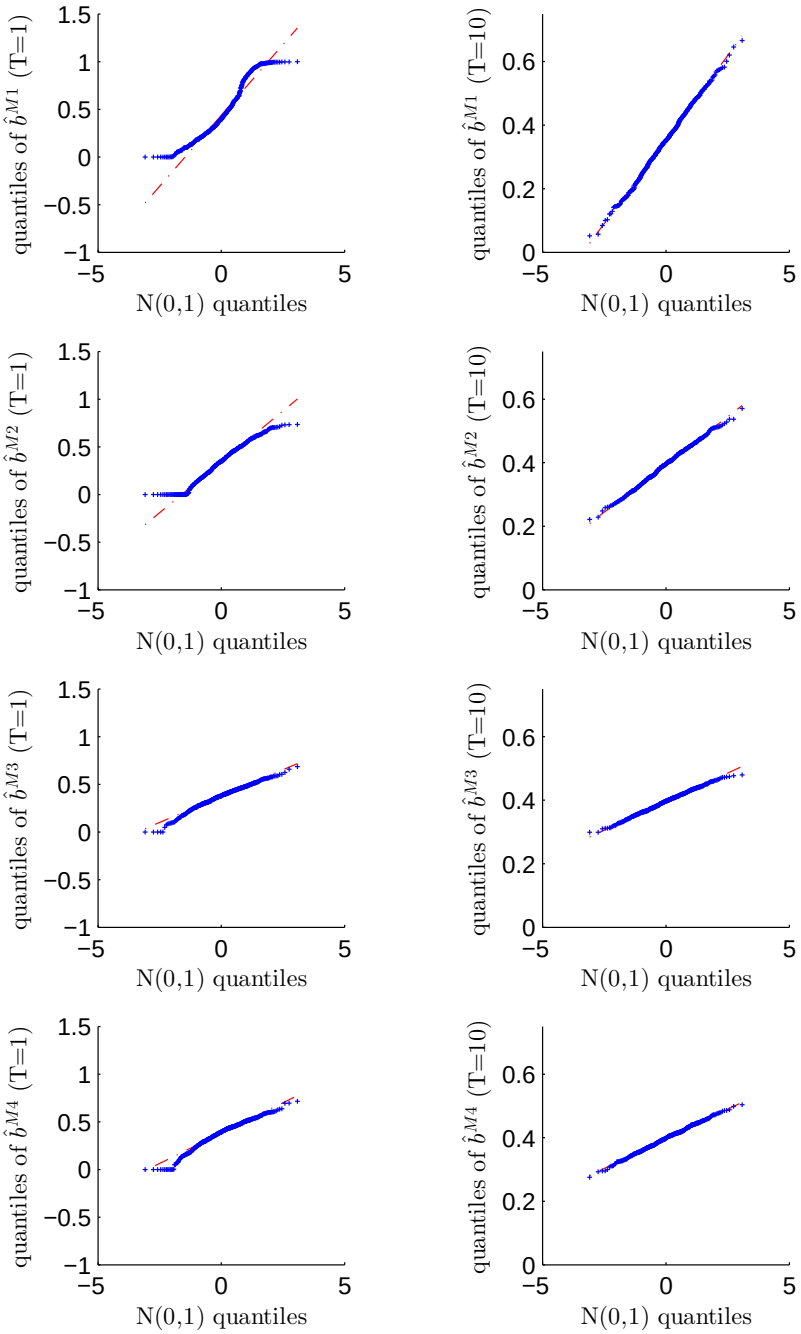


Figure 5.6 qq-plots of subordinator drift estimators for low variance inverse Gaussian subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator drift b in case of the low variance inverse Gaussian subordinator with drift for short and long time horizon.

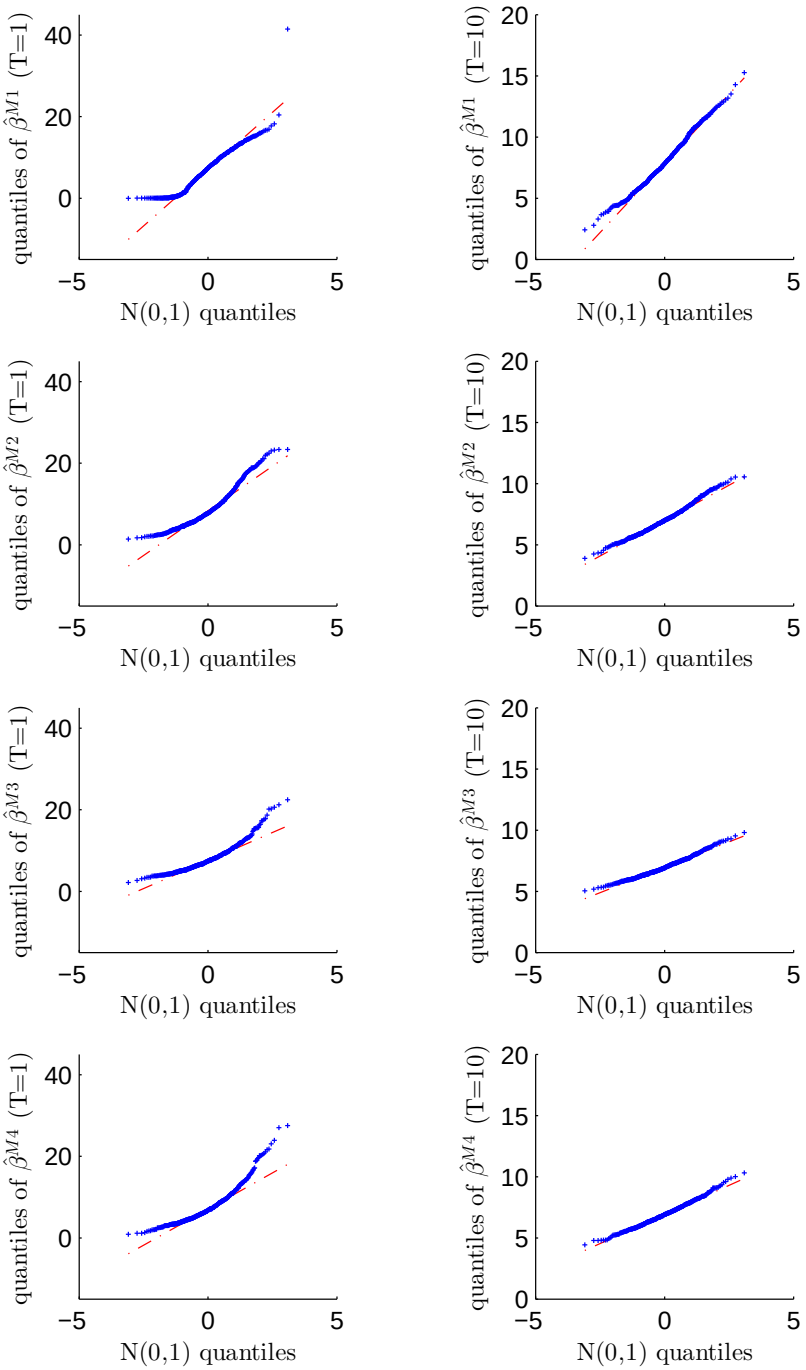


Figure 5.7 qq-plots of subordinator β estimators for low variance inverse Gaussian subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the low variance inverse Gaussian subordinator with drift for short and long time horizon.

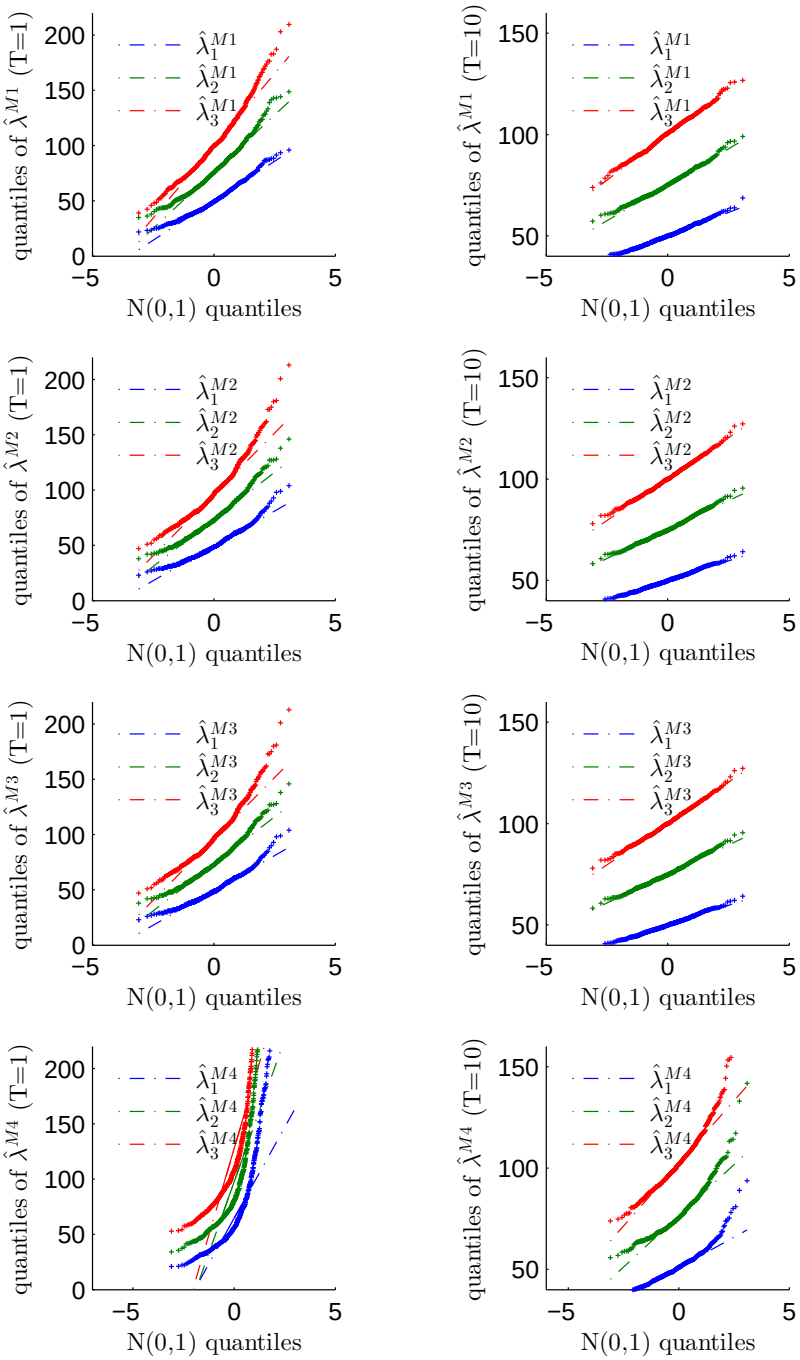


Figure 5.8 qq-plots of intensity estimators for high variance inverse Gaussian subordinator with drift: The figure illustrates qq-plots of the intensity estimators in case of the high variance inverse Gaussian subordinator with drift for short and long time horizon.

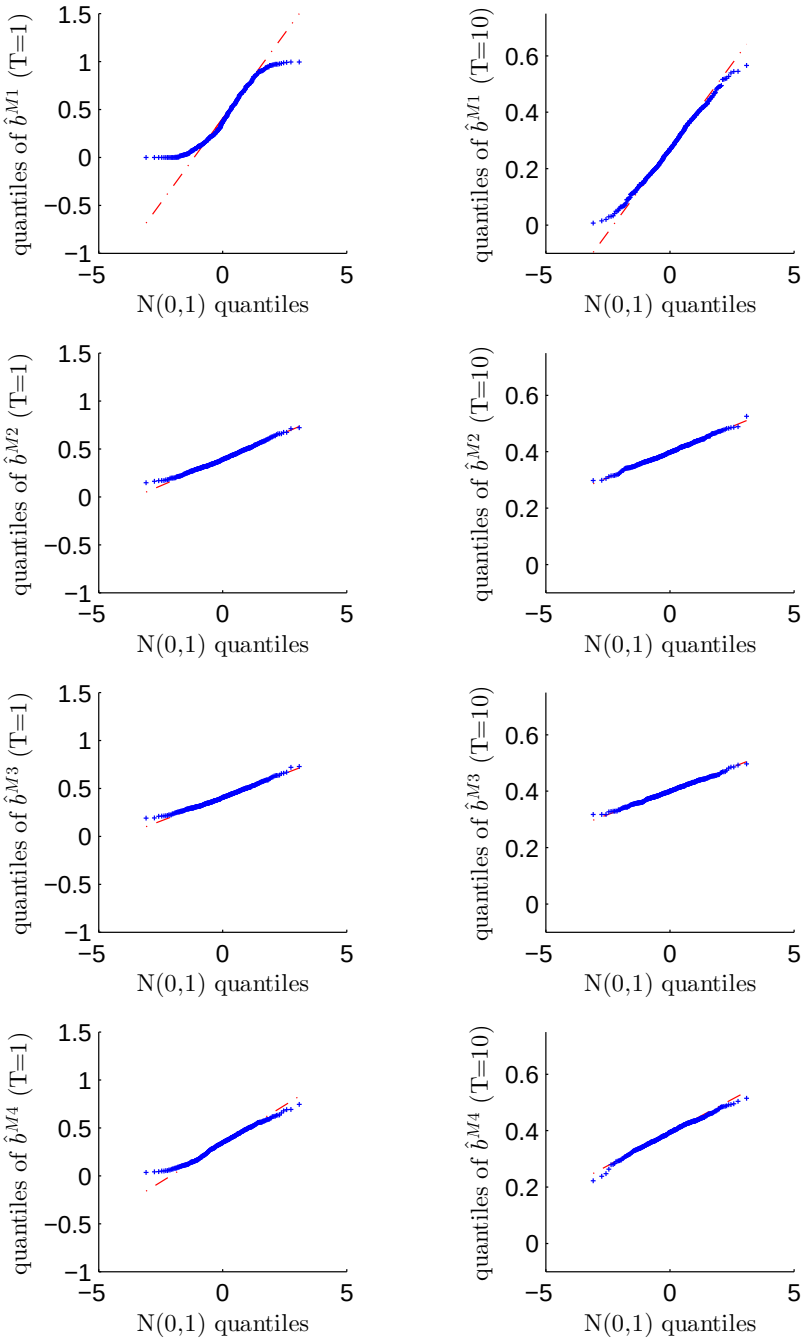


Figure 5.9 qq-plots of subordinator drift estimators for high variance inverse Gaussian subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator drift b in case of the high variance inverse Gaussian subordinator with drift for short and long time horizon.

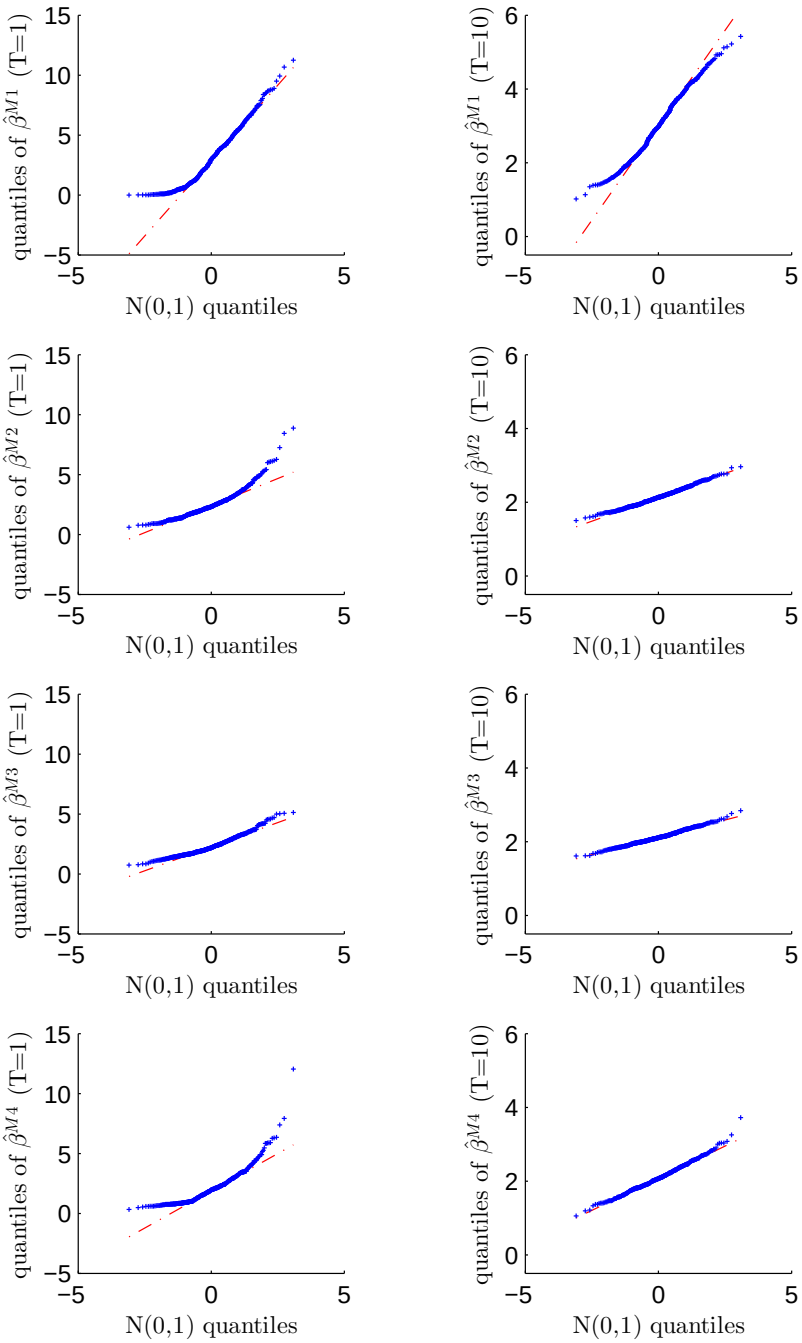


Figure 5.10 qq-plots of subordinator β estimators for high variance inverse Gaussian subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the high variance inverse Gaussian subordinator with drift for short and long time horizon.

Summary

Though no estimation method consistently outperforms all the others, the maximum likelihood-based Methods (M2) and (M3) have in this study produced the overall most reliable estimates. Whereas the intensity estimators from these two methods were in all settings almost equal, Method (M3) performed slightly better for the subordinator parameters and, in addition, is considerably less computationally expensive. However, Method (M3) requires that the cluster sizes are precisely monitored (the arrival times are not necessary), i.e. it is necessary to differentiate exactly which claims belong to the same clusters, a requirement that cannot always be fulfilled in real-world applications, as will be seen, for instance, in the following section. Method (M1) is the fastest of all methods, but had often difficulties producing the correct subordinator parameters. The usefulness of Method (M4) strongly depended on a sufficiently long observation period; otherwise the method struggled with outliers for the intensity estimators.

Finally, if time-normalization is assumed, it is reasonable to set the intensities upfront to the sample mean, which produced good estimates, and determine only the subordinator parameters from the presented estimation methods. This way the dimensionality of the optimization problems is greatly reduced, particularly in high-dimensional applications, which leads not only to a significant reduction in computation time, but most likely also to more robustness of the optimization routine. The intensity estimators from Methods (M2) and (M3) were always close to the sample mean anyway and Methods (M1) and (M4) performed worse rather than better. Without time-normalization, however, the full optimization problem needs to be solved.

5.3 Real-world data example

In this section, the estimation methods developed and tested in the previous sections are applied to real-world claim arrival data, namely the well-known Danish fire insurance data (see www.macs.hw.ac.uk/ mcneil/data.html). For the period 01/1980 to 12/1990, the day of occurrence and the total losses in the categories building, content, and profit of 2167 fire events are recorded. To fit Model (M), $d = 3$, to the fire data, Methods (M1) and (M2) can be straightforwardly employed using a daily observation grid. Methods (M3) and (M4) can only be applied under the assumption that the simultaneous occurrences of claims were precisely recorded as such. The claim cluster sizes in the categories building, content, and profit vary between zero and five for the first two categories, and between zero and three for the latter. For the aggregate process, the cluster sizes lie between one and eleven and the total number of cluster arrivals is 1645. Figure 5.11 illustrates the three claim number processes as well as their cluster arrivals. For the categories buildings and content, the claim number processes appear almost linear which fits the Lévy nature of the proposed model. Furthermore, large clusters tend to appear simultaneously, as severe fire events mostly cause high damages in all categories. Hence, a common source of randomness represented through the Lévy subordinator seems like a reasonable modelling approach. In the profits category, the observations appear to suggest a trend towards higher cluster sizes over time. Since such a trend cannot be covered with the presented model, a suitable model extension, as will be discussed in Section 6.2, may be necessary.

Figure 5.12 shows the monthly moving average of the daily process increments as well as the sample autocorrelation function up to a one-year lag for all categories. The plot confirms that no strong trend is apparent for the categories building and content, but a slight upward trend is present in the profit data. Furthermore, no serial correlation is revealed for any of the three categories, which supports the Lévy assumption. According to the model, the multivariate cluster arrival times are generated by a Poisson process, thus the observed inter-arrival times should be iid exponentially distributed for the model to be applicable. A histogram of the observations together with the estimated density of an exponential distribution is plotted in Figure 5.13. The fit is not very convincing, but this was to be expected since the arrival times are reported on a daily basis rather than precisely.

In the presented modelling approach, the time-normalization Assumption (TN) is imposed and an inverse Gaussian and gamma subordinator are chosen for the time-change. The data was first fitted to the models including a drift component for both subordinator choices. Since only Method (M1) detected a non-zero drift, which was of order 10^{-6}, only the estimation results for the models with no subordinator drift are presented in the following.

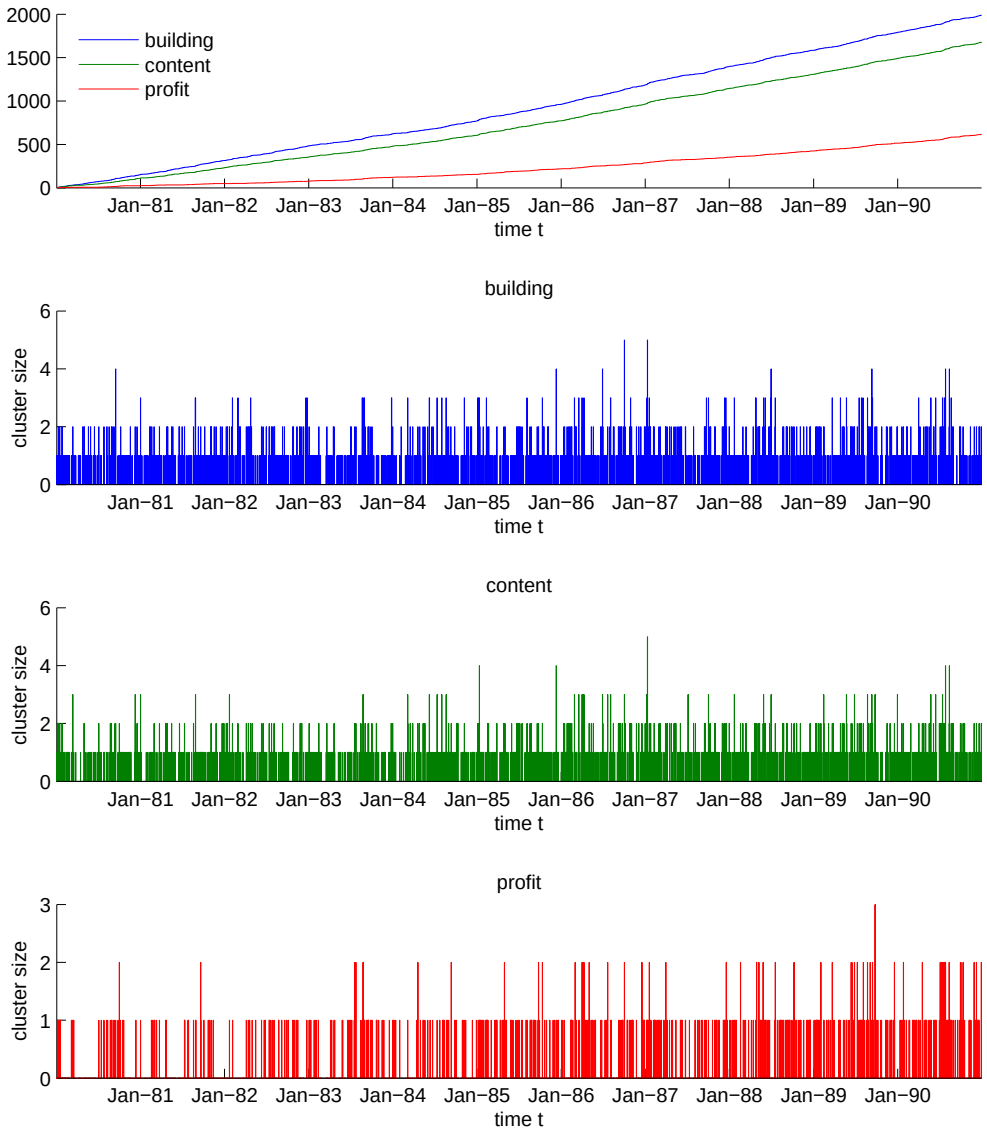


Figure 5.11 Danish fire insurance data: The figure illustrates the claim arrivals of the Danish fire insurance data. The top charts shows the claim number processes for the three categories building, content, and profit; the bottom three charts present the cluster arrivals in each category.

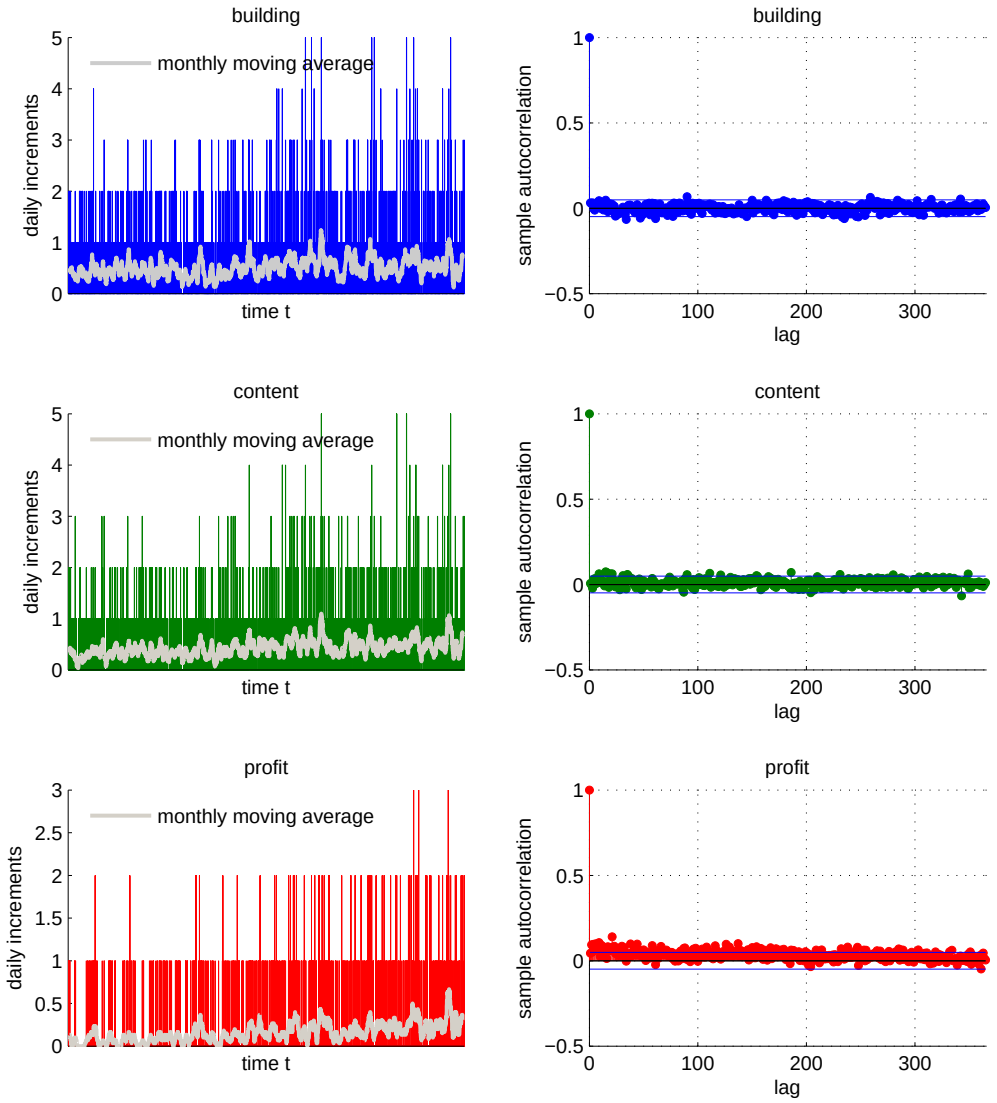


Figure 5.12 Moving average and autocorrelation of Danish fire insurance data: The figure illustrates the daily increments of the claim number processes including the monthly moving average (left) and the autocorrelation function up to a one-year lag (right) for the three categories building, content, and profit.

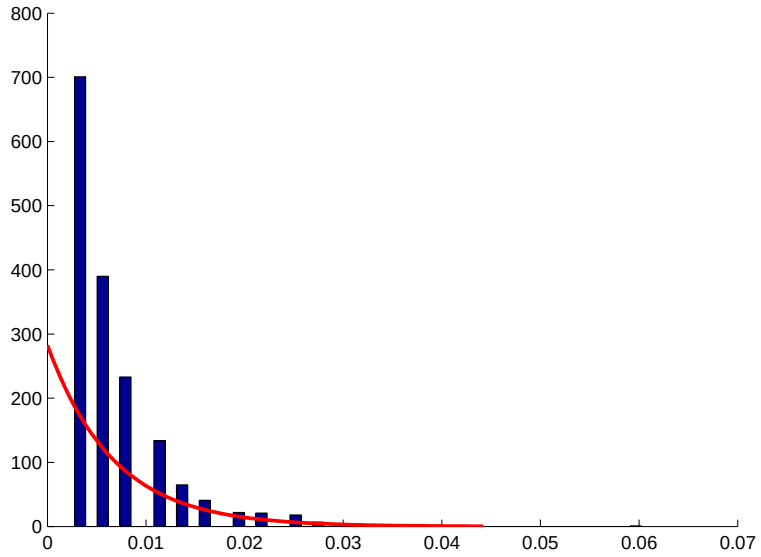


Figure 5.13 Histogram of cluster inter-arrival times of Danish fire insurance data: The figure presents a histogram of the inter-arrival times of (multivariate) claim clusters in the Danish fire insurance data. The estimated density of an exponential distribution is added to the plot in red.

The initial parameter determined as suggested in Remark 5.6 are valid in both settings; the initial intensities are

$$\hat{\boldsymbol{\lambda}}_0 = (180.909, 152.636, 56.000)',$$

and the initial subordinator parameters for the inverse Gaussian and gamma case, respectively, are

$$\hat{\beta}_0^{IG} = 15.163, \quad \hat{\beta}_0^{gamma} = 229.910.$$

The estimation results for both subordinator choices and all four estimation methods are summarized in Table 5.13. The intensity estimators for Methods (M1)–(M3) are, as they should be, almost identical for the inverse Gaussian and the gamma setting. Furthermore, the intensity estimators from the two maximum likelihood methods are again close to the initial values, i.e. the sample mean, while the estimated intensities from Method (M1) are a bit lower. Method (M4), however, does not produce reasonable estimators for λ_1 and λ_3 in the case of an inverse Gaussian subordinator. Hence, we restrict our attention to the comparison of the results from Methods (M1)–(M3) in the following.

For the subordinator parameter, the estimators in the inverse Gaussian and gamma case naturally differ. In order to amount to the same subordinator variance, $(\beta^{IG})^2 = \beta^{gamma}$ needs to hold (cf. Table 4.4). This equality is approximately fulfilled for the estimators of Method (M1), while for the two maximum likelihood methods there is some deviation. Furthermore, in each subordinator setting, the β estimators of Methods (M1) and (M2) are quite close, whereas the estimators from Method (M3) are considerably smaller. For both subordinator choices, a smaller β value corresponds to a higher subordinator variance and, hence, larger expected cluster sizes (with lower intensities). As was mentioned in the beginning of this section, Method (M3) (and Method (M4)) should only be applied if all claim clusters are recorded separately; due to the daily data, however, it may be the case that clusters occurring on the same day are aggregated and recorded as a single cluster instead. If this was the case, it would lead to overestimating the cluster sizes and could partly explain the discrepancy in the estimated subordinator parameters.

		λ_1	λ_2	λ_3	β
inverse Gaussian	(M1)	149.178	143.371	55.394	11.497
	(M2)	180.907	152.636	55.999	11.243
	(M3)	180.909	152.636	56.000	6.826
	(M4)	338.588	196.590	0.566	2.546
gamma	(M1)	149.179	143.374	55.395	132.193
	(M2)	180.906	152.632	55.999	148.969
	(M3)	180.911	152.639	56.001	88.812
	(M4)	178.634	119.212	8.942	53.997

Table 5.13 Estimation results Danish fire insurance data: The table presents the estimators from Methods (M1)–(M4) for the Danish fire claim arrival data for Model (M) with an inverse Gaussian and a gamma subordinator (without drift) as time-change process.

Figure 5.14 illustrates the observed claim number processes in each category together with sample paths from the models with an inverse Gaussian and a gamma subordinator using the estimated parameters of Methods (M1)–(M3). The models based on the gamma subordinator produce smoother paths compared to the inverse Gaussian case, which better resembles the observed data. The paths for the categories building and content that are sampled using the estimators from Method (M1) are very close to each other for both subordinator choices, which is not in line with the observed relationship between these two categories. As to be expected, being based on Lévy processes, no setting can account for the observed non-linearity of the profit claim number process. Overall, the sample paths from the gamma subordinator setting based on the estimated parameters from Methods (M2) and (M3) seem to provide the best fit.

Studying the aggregate cluster sizes of the observed process as well as a sample of each

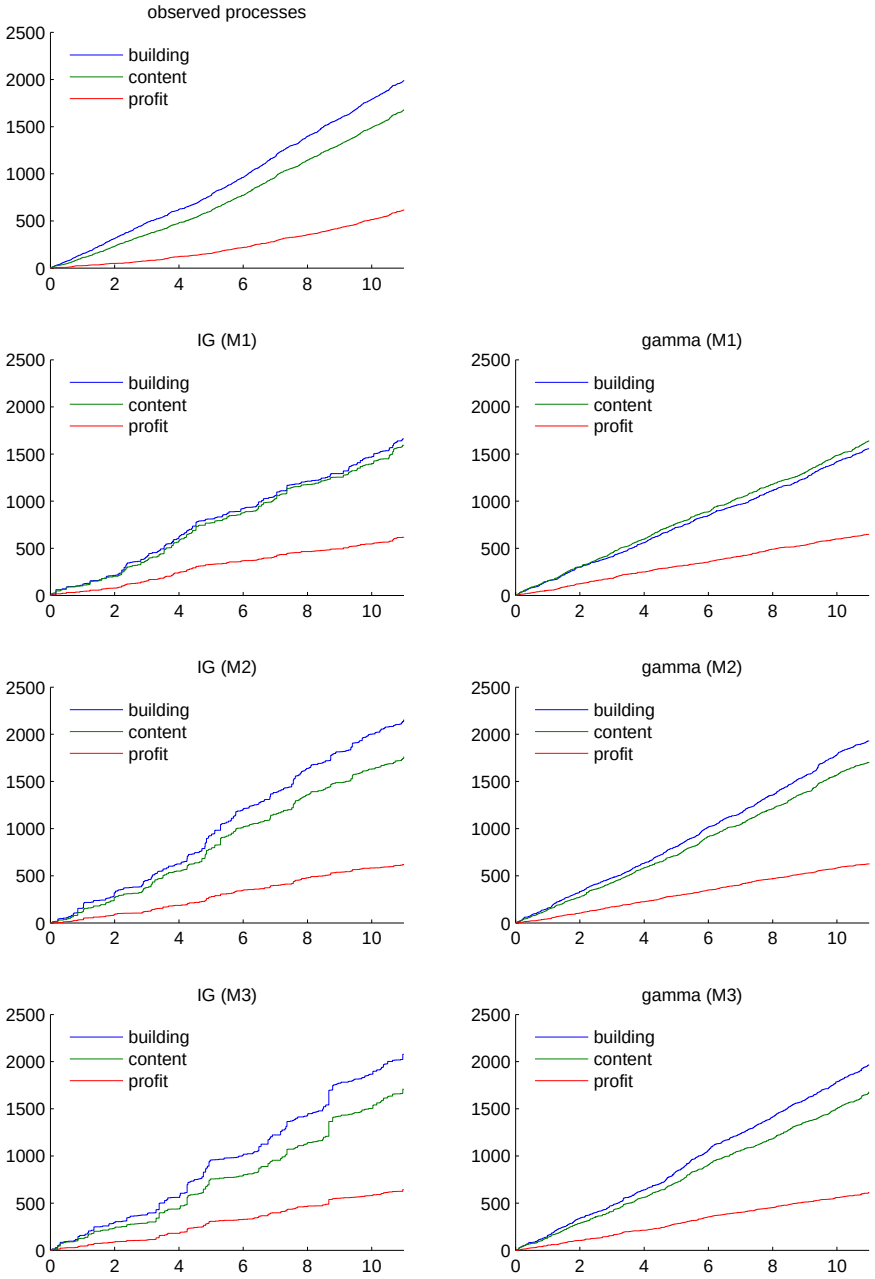


Figure 5.14 Sample paths of Danish fire insurance data and fitted models: The top chart shows the observed sample paths of the claim number processes from the Danish fire insurance data in the three categories building, content, and profit. The charts in the bottom three rows present sample paths of the fitted models based on an inverse Gaussian (left) and a gamma (right) subordinator using the estimated parameters from Methods (M1)–(M3). (Method (M4) is excluded due to the in part unreliable estimation results.)

fitted model illustrated in Figure 5.15, all models show a tendency to allow for higher jump sizes than the ones observed, in the inverse Gaussian setting even more so than in the gamma setting, which is in line with the less smooth sample paths. The effect is most pronounced in the cluster sizes corresponding to the estimators from Method (M3). The same observation holds true for the component-wise clusters presented in Figures A.27-A.29 in the Appendix. Of course, higher jump sizes correspond to lower jump intensities, as can be seen in Table 5.14. The individual jump intensities of each category are considerably lower in all models compared to the observed values, in particular so for the estimators from Method (M3). On the other hand, the joint jump intensities corresponding to Methods (M1) and (M2) are much higher then they should be, whereas for Method (M3) they fit almost perfectly. As the joint jumps are responsible for the dependence between the components, it follows that the models based on the estimators from Method (M3) more adequately capture the dependence structure in the data compared to the other estimators, but for the price of a reduced fit to the marginal processes. This observation is confirmed when looking at the observed and theoretical component-wise standard deviations and pairwise correlations presented in Table 5.15. All models overestimate the standard deviations, but for the estimators from Method (M3) the deviation is highest. The correlations, however, are always underestimated but this time Method (M3) outperforms the other two methods.

Overall, the gamma subordinator seems to offer a better fit than the inverse Gaussian subordinator, but in its basic form no model specification is flexible enough to provide a perfect fit to the Danish fire insurance data. As the subordinator influences both, the variability of the marginal processes as well as the dependence structure, limitations exist for the outcomes that can be generated in the model. Choosing the estimated parameters from Methods (M1) or (M2) gives preference to a better fit of the marginal processes compared to the dependence structure, whereas the estimators from Method (M3) should be chosen if capturing the dependence structure is given higher priority compared to matching the marginal processes. Introducing additional subordinators in the modelling approach, as will be discussed in Section 6.2, may provide the increased flexibility necessary to overcome this trade-off and describe the Danish fire insurance data more comprehensively.

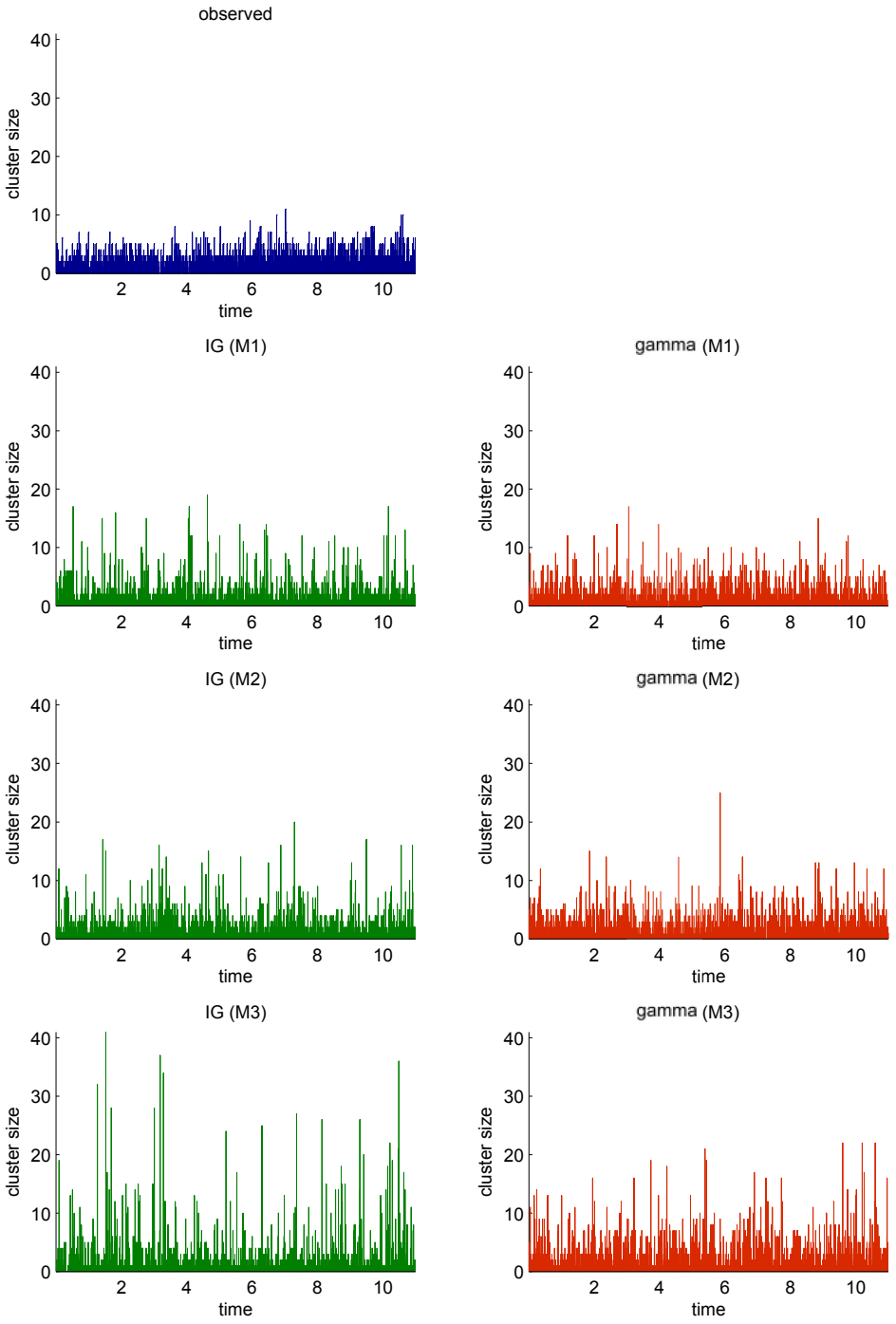


Figure 5.15 Cluster sizes of aggregate Danish fire insurance data and fitted models: The top chart shows the observed cluster sizes of the aggregate claim number processes from the Danish fire insurance data. The charts in the bottom three rows present sampled aggregate cluster sizes of the fitted models based on an inverse Gaussian (left) and gamma (right) subordinator using the estimated parameters from Methods (M1)–(M3). (Method (M4) is excluded due to the in parts unreliable estimation results.)

		L^1	L^2	L^3	$\bar{L}$
observed		140.091	123.909	51.000	149.546
inverse Gaussian	(M1)	123.354	108.285	47.475	214.886
	(M2)	122.019	107.190	47.191	211.916
	(M3)	91.352	81.447	39.368	149.546
gamma	(M1)	113.986	101.475	46.691	181.488
	(M2)	118.428	105.080	47.5391	191.437
	(M3)	98.658	88.824	43.422	149.545

Table 5.14 Jump intensities of Danish fire insurance data and fitted models: The table summarizes the observed jump intensities of the individual claim number processes from the Danish fire insurance data in the three categories building, content, and profit as well as for the joint/aggregate claim number process. In addition, the theoretical jump intensities of the fitted models with inverse Gaussian and gamma subordinator using the estimated parameters from Methods (M1)–(M3) are presented. (Method (M4) is excluded due to the in part unreliable estimation results.)

		std(.)			$\mathbb{C}$or(.,.) in %		
		L_1^1	L_1^2	L_1^3	L_1^1, L_1^2	L_1^1, L_1^3	L_1^2, L_1^3
observed		14.774	13.335	7.826	85.2	53.9	63.3
inverse Gaussian	(M1)	17.819	17.288	8.866	52.5	39.6	39.2
	(M2)	20.972	18.356	8.989	56.7	42.5	41.0
	(M3)	29.721	25.548	11.104	78.1	65.9	64.7
gamma	(M1)	17.819	17.288	8.866	52.5	39.6	39.2
	(M2)	20.015	17.579	8.778	52.7	38.7	37.2
	(M3)	23.440	20.371	9.556	65.1	50.9	49.4

Table 5.15 Standard deviation and correlation of Danish fire insurance data and fitted models: The table summarizes the empirical standard deviations and correlations of the individual claim number processes from the Danish fire insurance data in the three categories building, content, and profit. In addition, the theoretical quantities of the fitted models with inverse Gaussian and gamma subordinator using the estimated parameters from Methods (M1)–(M3) are presented. (Method (M4) is excluded due to the in part unreliable estimation results.)

6 Applications and extensions

In Section 6.1 of this chapter, actuarial applications of the presented model are examined in more detail. For this purpose, iid claim sizes are introduced and premium principles and other actuarial risk measures are evaluated. The results are compared to those stemming from a claim number process with independent Poisson marginals and, in addition, to those of a claim number process with the same Poisson cluster marginals as the proposed model, but independence between the components. In Section 6.2, the model extension with iid claim sizes is investigated in more detail and two more extensions are introduced, namely a deterministic time-change and multivariate subordination. A deterministic time-change breaks up the stationarity of the process increments and offers, for instance, the possibility to incorporate a seasonal effect in the model. Multivariate subordination provides the flexibility to generate inhomogeneous marginals (beyond a mere difference in the intensity parameters) and a wider variety of dependence structures. In addition, a time-dynamic version of the well-known CreditRisk$^+$ model can be formulated. All extensions, however, reduce the mathematical tractability of the model and lead to an increased reliance on numerical evaluation methods.

6.1 Premium calculation and dependence ordering

Traditionally, the core business of insurance and reinsurance companies has been to sell risk coverage. Pooling multiple risks, the insurer expects to profit from the diversification effect that makes the aggregate outcome of the portfolio less risky or volatile and, hence, more manageable than the individual risks. Let random variables $X_i \geq 0$, $i \in \mathbb{N}$, describe the individual risks and assume for simplicity that the risks are iid with mean μ and variance σ^2. Then the strong law of large numbers predicts that the normalized portfolio outcome $d^{-1}\sum_{i=1}^{d} X_i$ converges to μ for $d \to \infty$. In particular, in the limit the variance $\mathbb{V}\mathrm{ar}[d^{-1}\sum_{i=1}^{d} X_i] = d^{-1}\sigma^2$ is zero and the outcome becomes almost surely deterministic. With dependencies between the risks, however, the law of large numbers in general no longer holds and the portfolio may be significantly riskier than in the independent case: convergence may be slower and the limit may no longer be deterministic (cf. also Section 4.3). More precisely, in general it holds for the portfolio variance:

$$\mathbb{V}\mathrm{ar}\left[\frac{1}{d}\sum_{i=1}^{d} X_i\right] = \frac{\sigma^2}{d} + \frac{2}{d^2}\sum_{i=1}^{d}\sum_{j=i+1}^{d} \mathbb{C}\mathrm{ov}[X_i, X_j].$$

If the risks are positively correlated with $\mathbb{C}\text{or}[X_i, X_j] = \rho > 0$ for all $i \neq j$, it follows:

$$\mathbb{V}\text{ar}\left[\frac{1}{d}\sum_{i=1}^{d} X_i\right] = \frac{\sigma^2}{d} + \frac{d(d-1)}{d^2}\rho\sigma^2 = \frac{\sigma^2}{d} + \Big(1 - \frac{1}{d}\Big)\rho\sigma^2 \longrightarrow \rho\sigma^2, \quad d \to \infty.$$

Thus, neglecting dependence in the modelling approach leads to underestimating the risks involved. In this thesis, a multivariate model for claim numbers has been presented which incorporates arrivals of claim clusters as well as positive dependence. Hence, the model predicts larger portfolio variance than independent Poisson models due to the increased variance of each individual component and the positive correlation between the different components. Of course, variance (or correlation) is only one possible measure of risk and does not account for the full dependence structure (see the discussion in Section 4.1). In this section, the effect of the model on various actuarial risk measures and in particular premium principles is studied and a comparison is made firstly with the basic independent Poisson model, excluding both dependence and cluster arrivals. Secondly, a comparison is made with the process that features the same marginals, i.e. incorporates component-wise cluster arrivals, but assumes independence and, hence, excludes the possibility of simultaneous claim arrivals in multiple components.

Let a random variable $X \geq 0$ represent an actuarial risk. A premium principle Π is a mapping that assigns X a non-negative value $\Pi(X)$. The premium is thought of as the financial compensation the insurer receives in return for taking on the risk X (excluding any operational costs) and should be low enough to attract customers but high enough to guarantee solvability and profitability of the insurer. In particular, premium principles constitute risk measures and, hence, should assign higher values to perceived greater risks. The net or actuarial premium is defined as the expectation $\mathbb{E}[X]$ of X (under the physical measure $\mathbb{P}$). Obviously, as the insurer has to cover the incoming claims obligations, the actuarial premium has to be a lower bound on the premium charged to the customers. Due to the uncertainty of X, the insurer needs to add a safety loading $\Pi(X) - \mathbb{E}[X] > 0$ to create a reserve for adverse outcomes and avoid ruin. If two risks have the same expectation, the risk that has a thicker tailed distribution will be considered less attractive since extreme outcomes are more likely. In general, a premium principle is a function of the distribution function F_X of X. For the various methods of developing premium principles and a discussion of their respective properties, see (Rolski et al., 1999, Chapter 3, pp.79), Goovaerts et al. (1984), and Bühlmann (1970). Some common principles derived from the actuarial premium are:

$$\begin{aligned}
&\text{expectation principle:} && \Pi^e(X) = \mathbb{E}[X] + \theta\,\mathbb{E}[X],\\
&\text{variance principle:} && \Pi^v(X) = \mathbb{E}[X] + \theta\,\mathbb{V}\text{ar}[X],\\
&\text{standard deviation principle:} && \Pi^s(X) = \mathbb{E}[X] + \theta\sqrt{\mathbb{V}\text{ar}[X]},
\end{aligned}$$

where a suitable scaling parameter $\theta > 0$ is chosen, for instance, from ruin probabilities or solvency capital requirements. Often, premium principles are generated using utility theory. Let $w > 0$ be the initial wealth and $u(\cdot)$ a utility function, i.e. a non-decreasing and – given that risk aversion is plausible in this setting – concave function. Then the so-called zero-utility premium is the solution of the following equation:

$$u(w - \Pi(X)) = \mathbb{E}[u(w - X)].$$

The expectation principle can be embedded in this framework using a linear utility function and the exponential principle is defined via the utility $u(x) = -\exp\{-\theta x\}$ as:

$$\text{exponential principle:} \quad \Pi^{ex}(X) = \frac{1}{\theta} \log \mathbb{E}[\exp\{\theta X\}].$$

The quantile principle directly considers the tail of the risk:

$$\text{quantile principle:} \quad \Pi^{q}(X) = F_X^{-1}(1-\epsilon) = VaR_X(1-\epsilon),$$

where $0 < \epsilon < 1$. The $(1-\epsilon)$-quantile of X, or value at risk with level $p := 1-\epsilon$, is the amount that is not exceeded with (high) probability p. This risk measure is used in the Solvency II regulatory framework to determine the required capital an insurer has to hold in order to be able to pay incoming claims with high probability.

We start with considering a direct insurer facing actuarial risks $X_1, \dots, X_d \geq 0$. Each risk may consist of a single policy, a portfolio of policies, or even a business line outcome and is represented through a collective risk model. The claim sizes $Z_{i1}, Z_{i2}, \dots \overset{d}{=} Z_i \geq 0$ are assumed to be iid for each component $i = 1, \dots, d$ and independent for different components. For the random vector consisting of the independent jump sizes, the notation $\boldsymbol{Z} := (Z_1, \dots, Z_d)'$ is employed. As d-dimensional claim number process, three cases are compared: the time-changed Poisson process $\boldsymbol{L}_t = (N^1_{\Lambda_t}, \dots, N^d_{\Lambda_t})'$ which was introduced in Model (M), the independent Poisson process $\boldsymbol{N}_t = (N^1_t, \dots, N^d_t)'$ used as building block for Model (M), and the process $\boldsymbol{M}_t := (N^1_{\Lambda^1_t}, \dots, N^d_{\Lambda^d_t})'$ with iid copies $\Lambda^1, \Lambda^2, \dots$ of Λ which has the same marginals as L but no dependence, cf. Section 4.3. All three processes, as well as the respective aggregate processes $\bar{L}_t = \sum_{i=1}^d N^i_{\Lambda_t}$, $\bar{N}_t = \sum_{i=1}^d N^i_t$, and $\bar{M}_t = \sum_{i=1}^d N^i_{\Lambda^i_t}$, have compound Poisson representations, which can be found in 6.1. The table also contains the probability mass functions together with mean and variance of the process distributions. The results for $\boldsymbol{L}$ and the univariate marginals M^i of $\boldsymbol{M}$ follow from Chapter 4 and the results for $\boldsymbol{N}$ and the vector $\boldsymbol{M}$ are a consequence of the space-time decomposition property of the compound Poisson process, see Remark 2.8. It

should be noted that closed form expressions of the multivariate process and jump-size distributions could be given in all three cases and the jump size distributions of the aggregate processes are available in closed form as well. However, the distributions of the aggregate processes can only be given for the claim number processes $\boldsymbol{N}$ and $\boldsymbol{L}$. In case of $\boldsymbol{M}$, the distribution needs to be calculated, for instance, using Panjer's recursion – which delivers precise results in this setting. However, as was discussed in Sections 4.1 and 4.2, it is often preferable to evaluate the formulas for the distributions of $\boldsymbol{L}_t$ and $\bar{L}_t$ using Panjer's recursion as well.

If Assumption (TN) holds, i.e. the first moment of the subordinator Λ exists and fulfils $\mathbb{E}[\Lambda_1] = 1$, the expected number of claims is identical for all three claim number processes. The component-wise variance – assuming existence – of the Poisson cluster processes $\boldsymbol{L}$ and $\boldsymbol{M}$ are also identical and, given time-normalization, exceed the variance of the Poisson process $\boldsymbol{N}$. However, as the components of $\boldsymbol{L}$ positively correlate whereas the components of $\boldsymbol{M}$ are independent, the variance of the aggregate process $\bar{L}$ exceeds the variance of $\bar{M}$, w.l.o.g. in $t = 1$:

$$\mathbb{V}\mathrm{ar}[\bar{L}_1] = |\boldsymbol{\lambda}| + \mathbb{V}\mathrm{ar}[\Lambda_1]|\boldsymbol{\lambda}|^2 \geq \mathbb{V}\mathrm{ar}[\bar{M}_1] = |\boldsymbol{\lambda}| + \mathbb{V}\mathrm{ar}[\Lambda_1]\boldsymbol{\lambda}'\boldsymbol{\lambda} \geq \mathbb{V}\mathrm{ar}[\bar{N}_1] = |\boldsymbol{\lambda}|.$$

The first equality holds as the square of the sum $|\boldsymbol{\lambda}|^2$ is larger than (or equal to in dimension one) the sum of the squares $\boldsymbol{\lambda}'\boldsymbol{\lambda}$.

	$\boldsymbol{L}_t$	$\boldsymbol{M}_t$	$\boldsymbol{N}_t$
distribution in $\boldsymbol{k} \in \mathbb{N}_0^d$	$\frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\varphi_{\Lambda_t}^{(\lvert\boldsymbol{k}\rvert)}(\lvert\boldsymbol{\lambda}\rvert)$	$\frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\prod_{i=1}^d \varphi_{\Lambda_t}^{(k_i)}(\lambda_i)$	$\frac{\boldsymbol{\lambda}^{\boldsymbol{k}}}{\boldsymbol{k}!}\exp\{-\lvert\boldsymbol{\lambda}\rvert t\}$
agg. distribution in $k \in \mathbb{N}_0$	$\frac{(-\lvert\boldsymbol{\lambda}\rvert)^k}{k!}\varphi_{\Lambda_t}^{(k)}(\lvert\boldsymbol{\lambda}\rvert)$	Panjer	$\frac{\lvert\boldsymbol{\lambda}\rvert^k}{k!}\exp\{-\lvert\boldsymbol{\lambda}\rvert t\}$
jump intensity	$\Psi_\Lambda(\lvert\boldsymbol{\lambda}\rvert)$	$\sum_{i=1}^d \Psi_\Lambda(\lambda_i)$	$\lvert\boldsymbol{\lambda}\rvert$
jump distribution in $\boldsymbol{k} \in \mathbb{N}^d$	$\frac{-(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\frac{\Psi_\Lambda^{(\lvert\boldsymbol{k}\rvert)}(\lvert\boldsymbol{\lambda}\rvert)}{\Psi_\Lambda(\lvert\boldsymbol{\lambda}\rvert)}$	$\sum_{i=1}^d \mathbb{1}_{\{k_i\boldsymbol{e}_i\}}(\boldsymbol{k})\frac{-(-\lambda_i)^{k_i}}{k_i!}\frac{\Psi_\Lambda^{(k_i)}(\lambda_i)}{\sum_{j=1}^d \Psi_\Lambda(\lambda_j)}$	$\sum_{i=1}^d \mathbb{1}_{\{\boldsymbol{e}_i\}}(\boldsymbol{k})\frac{\lambda_i}{\lvert\boldsymbol{\lambda}\rvert}$
agg. jump distribution in $k \in \mathbb{N}$	$\frac{-(-\lvert\boldsymbol{\lambda}\rvert)^k}{k!}\frac{\Psi_\Lambda^{(k)}(\lvert\boldsymbol{\lambda}\rvert)}{\Psi_\Lambda(\lvert\boldsymbol{\lambda}\rvert)}$	$\sum_{i=1}^d \frac{-(-\lambda_i)^k}{k!}\frac{\Psi_\Lambda^{(k)}(\lambda_i)}{\sum_{j=1}^d \Psi_\Lambda(\lambda_j)}$	$\mathbb{1}_{\{1\}}(k)$
mean	$t\,\mathbb{E}[\Lambda_1]\boldsymbol{\lambda}$	$t\,\mathbb{E}[\Lambda_1]\boldsymbol{\lambda}$	$t\boldsymbol{\lambda}$
agg. mean	$t\,\mathbb{E}[\Lambda_1]\lvert\boldsymbol{\lambda}\rvert$	$t\,\mathbb{E}[\Lambda_1]\lvert\boldsymbol{\lambda}\rvert$	$t\lvert\boldsymbol{\lambda}\rvert$
variance	$t(\mathbb{E}[\Lambda_1]\boldsymbol{\lambda} + \mathbb{V}\mathrm{ar}[\Lambda_1]\boldsymbol{\lambda}^2)$	$t(\mathbb{E}[\Lambda_1]\boldsymbol{\lambda} + \mathbb{V}\mathrm{ar}[\Lambda_1]\boldsymbol{\lambda}^2)$	$t\boldsymbol{\lambda}$
agg. variance	$t(\mathbb{E}[\Lambda_1]\lvert\boldsymbol{\lambda}\rvert + \mathbb{V}\mathrm{ar}[\Lambda_1]\lvert\boldsymbol{\lambda}\rvert^2)$	$t(\mathbb{E}[\Lambda_1]\lvert\boldsymbol{\lambda}\rvert + \mathbb{V}\mathrm{ar}[\Lambda_1]\boldsymbol{\lambda}'\boldsymbol{\lambda})$	$t\lvert\boldsymbol{\lambda}\rvert$

Table 6.1 Comparison of characteristics of claim number processes $\boldsymbol{L}$, $\boldsymbol{M}$, and $\boldsymbol{N}$: The table summarizes the characteristics of the time-changed claim number process $\boldsymbol{L}_t$, the process $\boldsymbol{M}_t$ with the same marginal distributions but independence, and the process $\boldsymbol{N}_t$ with independent Poisson components, as well as the corresponding aggregate processes $\bar{L}_t$, $\bar{M}_t$, and $\bar{N}_t$, $t \geq 0$.

The compound vector processes with secondary distributions $Z_1, \dots, Z_d$ describes the component-wise total claim amount, for instance in terms of $\boldsymbol{L}$:

$$\Big(\sum_{j=1}^{L_t^1} Z_{1j}, \dots, \sum_{j=1}^{L_t^d} Z_{dj}\Big)'. \tag{6.1}$$

Then the total portfolio outcome is the aggregate compound process $\sum_{i=1}^{d}\sum_{j=1}^{L_t^i} Z_{ij}$. The following theorem summarizes mean and (co-)variance of the compound vector and aggregate process given the claim number process $\boldsymbol{L}$.

Theorem 6.1 (Mean and variance of the compound process)
Let $\boldsymbol{L}$ be the d-dimensional claim number process specified in Model (M) and let $\boldsymbol{Z}$ be a d-dimensional random vector with independent, non-negative components. Then it holds for mean and variance of the i-th compound process in Equation (6.1), $i = 1, \dots, d$, given the first and second moment of the subordinator and claim size distribution exist:

$$\mathbb{E}\Big[\sum_{j=1}^{L_t^i} Z_{ij}\Big] = t\lambda_i\,\mathbb{E}[Z_i]\,\mathbb{E}[\Lambda_1], \quad \mathbb{V}\mathrm{ar}\Big[\sum_{j=1}^{L_t^i} Z_{ij}\Big] = t(\lambda_i\,\mathbb{E}[Z_i^2]\,\mathbb{E}[\Lambda_1] + \lambda_i^2\,\mathbb{E}[Z_i]^2\,\mathbb{V}\mathrm{ar}[\Lambda_1]).$$

The covariance between different components $i \neq k$ is:

$$\mathbb{C}\mathrm{ov}\Big[\sum_{j=1}^{L_t^i} Z_{ij}, \sum_{j=1}^{L_t^k} Z_{kj}\Big] = t\,\mathbb{E}[Z_i]\,\mathbb{E}[Z_k]\lambda_i\lambda_k\,\mathbb{V}\mathrm{ar}[\Lambda_1].$$

Mean and variance of the aggregate process are:

$$\mathbb{E}\Big[\sum_{i=1}^{d}\sum_{j=1}^{L_t^i} Z_{ij}\Big] = t\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}]\,\mathbb{E}[\Lambda_1],$$

$$\mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^{d}\sum_{j=1}^{L_t^i} Z_{ij}\Big] = t(\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}^2]\,\mathbb{E}[\Lambda_1] + \mathbb{V}\mathrm{ar}[\Lambda_1](\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}])^2).$$

Proof. Mean and variance of a compound sum exist if the first and second moment of the primary and the secondary distribution exist. Then it holds for each component i:

$$\mathbb{E}\left[\sum_{j=1}^{L_t^i} Z_{ij}\right] = \mathbb{E}[L_t^i]\,\mathbb{E}[Z_i], \quad \mathbb{V}\mathrm{ar}\left[\sum_{j=1}^{L_t^i} Z_{ij}\right] = \mathbb{E}[L_t^i]\,\mathbb{V}\mathrm{ar}[Z_i] + \mathbb{V}\mathrm{ar}[L_t^i]\,\mathbb{E}[Z_i]^2.$$

These results can be found using the tower rule for conditional expectations with conditioning on the outcome of the claim number process L_t^1 as shown, for instance, in (Mikosch, 2009, Chapter 3.1.1, p.73). It follows for the expectation using Table 6.1:

$$\mathbb{E}\left[\sum_{j=1}^{L_t^i} Z_{ij}\right] = t\lambda_i\,\mathbb{E}[Z_i]\,\mathbb{E}[\Lambda_1],$$

and for the variance:

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\sum_{j=1}^{L_t^i} Z_{ij}\right] &= t\lambda_i\,\mathbb{E}[\Lambda_1]\,\mathbb{V}\mathrm{ar}[Z_i] + t(\lambda_i\,\mathbb{E}[\Lambda_1] + \lambda_i^2\,\mathbb{V}\mathrm{ar}[\Lambda_1])\,\mathbb{E}[Z_i]^2 \\ &= t(\lambda_i\,\mathbb{E}[Z_i^2]\,\mathbb{E}[\Lambda_1] + \lambda_i^2\,\mathbb{E}[Z_i]^2\,\mathbb{V}\mathrm{ar}[\Lambda_1]).\end{aligned}$$

Using the tower rule and exploiting the conditional independence of the marginal processes, we compute for the mixed moment of two components $i \neq k$:

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^{L_t^i} Z_{ij}\sum_{j=1}^{L_t^k} Z_{kj}\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{L_t^i} Z_{ij}\sum_{j=1}^{L_t^k} Z_{kj}|\sigma(L^i, L^k)\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{j=1}^{L_t^i} Z_{ij}|\sigma(L^i)\right]\mathbb{E}\left[\sum_{j=1}^{L_t^k} Z_{kj}|\sigma(L^k)\right]\right] \\ &= \mathbb{E}[L_t^i\,\mathbb{E}[Z_i]L_t^k\,\mathbb{E}[Z_k]] = \mathbb{E}[Z_i]\,\mathbb{E}[Z_k]\,\mathbb{E}[L_t^iL_t^k].\end{aligned}$$

Then it follows for the covariance:

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left[\sum_{j=1}^{L_t^i} Z_{ij}, \sum_{j=1}^{L_t^k} Z_{kj}\right] &= \mathbb{E}\left[\sum_{j=1}^{L_t^i} Z_{ij}\sum_{j=1}^{L_t^k} Z_{kj}\right] - \mathbb{E}\left[\sum_{j=1}^{L_t^i} Z_{ij}\right]\mathbb{E}\left[\sum_{j=1}^{L_t^k} Z_{kj}\right] \\ &= \mathbb{E}[Z_i]\,\mathbb{E}[Z_k](\mathbb{E}[L_t^iL_t^k] - \mathbb{E}[L_t^i]\,\mathbb{E}[L_t^k]) \\ &= \mathbb{E}[Z_i]\,\mathbb{E}[Z_k]\,\mathbb{C}\mathrm{ov}[L_t^i, L_t^k] = t\,\mathbb{E}[Z_i]\,\mathbb{E}[Z_k]\lambda_i\lambda_k\,\mathbb{V}\mathrm{ar}[\Lambda_1].\end{aligned}$$

Finally, the expectation of the aggregate compound Poisson process follows from summation of the component-wise expectations, and for the variance it holds:

$$\begin{aligned}\mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^{d}\sum_{j=1}^{L_t^i} Z_{ij}\Big] &= \sum_{i=1}^{d}\sum_{k=1}^{d}\mathbb{C}\mathrm{ov}\Big[\sum_{j=1}^{L_t^i} Z_{ij}, \sum_{j=1}^{L_t^k} Z_{kj}\Big] \\ &= t(\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}^2]\,\mathbb{E}[\Lambda_1] + \mathbb{V}\mathrm{ar}[\Lambda_1]\sum_{i=1}^{d}\sum_{k=1}^{d}\lambda_i\lambda_k\,\mathbb{E}[Z_i]\,\mathbb{E}[Z_k]) \\ &= t(\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}^2]\,\mathbb{E}[\Lambda_1] + \mathbb{V}\mathrm{ar}[\Lambda_1](\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}])^2).\end{aligned}$$

□

The mean and variance of the compound vector process w.r.t $\boldsymbol{M}$ are identical to the respective values of the compound vector process w.r.t $\boldsymbol{L}$. For the Poisson process $\boldsymbol{N}$ it holds:

$$\mathbb{E}\Big[\sum_{j=1}^{N_t^i} Z_{ij}\Big] = \mathbb{V}\mathrm{ar}\Big[\sum_{j=1}^{N_t^i} Z_{ij}\Big] = \lambda_i\,\mathbb{E}[Z_i].$$

The expectation of the aggregate compound process is found from the summation of the component-wise expectations and is, given time-normalization, identical in all three cases. For both claim number processes $\boldsymbol{N}$ and $\boldsymbol{M}$, the components of the compound vector process are again independent. Hence, the covariance between the components is zero and the variance of the aggregate compound process is the sum of the component-wise variances. For $\boldsymbol{N}$ that is:

$$\mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^{d}\sum_{j=1}^{N_t^i} Z_{ij}\Big] = t\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}^2],$$

and for $\boldsymbol{M}$ that is:

$$\mathbb{V}\mathrm{ar}\Big[\sum_{i=1}^{d}\sum_{j=1}^{M_t^i} Z_{ij}\Big] = t(\boldsymbol{\lambda}'\,\mathbb{E}[\boldsymbol{Z}^2]\,\mathbb{E}[\Lambda_1] + (\boldsymbol{\lambda}^2)'\,\mathbb{E}[\boldsymbol{Z}]^2\,\mathbb{V}\mathrm{ar}[\Lambda_1]).$$

In particular, it follows analogously to the case without compounding that the variance in the model using the claim number process $\boldsymbol{L}$ exceeds the variance present in case of

using $\boldsymbol{M}$ which in turn exceeds, given time-normalization (or $\mathbb{E}[\Lambda_1] \geq 1$), the variance in case of $\boldsymbol{N}$.

If the risk of the direct insurer is the total portfolio outcome, then given time-normalization the actuarial premium as well as the premium due to the expected value principle coincide in all three cases as they are unaffected by variance or dependence. If, however, the risks are compared using the variance or standard deviation principle, the premium in the presence of dependence and cluster arrivals (i.e. for $\boldsymbol{L}$) exceeds the premium in case of independence and cluster arrivals (i.e. for $\boldsymbol{M}$) which in turn exceeds the premium if neither dependence nor cluster arrivals are accounted for (i.e. for $\boldsymbol{N}$). Putting it differently, if dependence and cluster arrivals are present in the claim count data but neglected in the modelling approach, the premium to be charged will likely be underestimated. The variance and standard deviation principles are only sensitive to the variance and covariance of the compound process. Hence, choosing a model with the same marginal distributions and the same correlations as $\boldsymbol{L}$ will result in the same premium. For the exponential and quantile principle, this is in general no longer true. We will later see that for the exponential premium as well as any zero-utility premium the same ordering holds as for the variance and covariance principles, though generally not for the quantile principle.

If the risk is a non-linear function of the aggregate portfolio outcome, even the actuarial premium is no longer unaffected by the dependence structure, which is particularly true for non-proportional reinsurance contracts. Reinsurance represents a risk transfer from a direct insurer (cedant) to another insurer. The direct insurer thereby converts uncertain payments to its customers due to claims into fixed premium payments to the reinsurer. For an introduction to reinsurance and its economic function see Liebwein (2009). In proportional reinsurance contracts, the risk is split proportionally between the direct insurer and the reinsurer. Here, the actuarial premium for the reinsurer is the respective portion of the actuarial premium of the direct insurer; hence, we are basically back in the case discussed previously for the direct insurer. For non-proportional reinsurance, the two most prominent examples are *per risk excess-of-loss* and *stop-loss contracts*. In a *per risk excess-of-loss contract*, the reinsurer covers of each incoming claim the portion exceeding a specified retention or priority $r > 0$ for the direct insurer. In a *stop-loss contract*, the reinsurer covers the excess amount of the annual aggregate claims over a retention. With these non-proportional contracts, the actuarial premium can no longer be directly derived from the premium calculated by the direct insurer and a self-contained calculation for the reinsurer is necessary. In experience rating, the reinsurer uses its previous claim experience with a particular contract. Typically, extreme value theory comes into play and the estimation is based on only a few observations. In exposure rating, the calculation of the reinsurer is based on the direct insurer's portfolio. Assuming a collective approach as before, the risk for the reinsurer in case of a *per risk excess-of-loss contract* is then:

$$\sum_{i=1}^{d}\sum_{j=1}^{L_t^i}(Z_{ij}-r)^+,$$

where $(Z_{ij}-r)^+ := \max\{Z_{ij}-r,0\}$ and $r>0$ is the retention level. Compared to the previous discussion regarding the portfolio of the direct insurer, only the claim sizes have changed (or maybe an independent thinning of L_t^i has to be considered in addition). Hence, the considerations regarding the premium calculations apply. For a *stop-loss contract* with retention $r>0$, the risk for the reinsurer is:

$$\Big(\sum_{i=1}^{d}\sum_{j=1}^{L_t^i}Z_{ij}-r\Big)^+.$$

The expectation or actuarial premium of this quantity is called the stop-loss premium of the aggregate portfolio outcome. In general, the expected exceedance of a random variable X over a variable threshold $r\in\mathbb{R}$ is called the stop-loss transform $\Pi_X(r)$ of X:

$$\Pi_X(r) := \mathbb{E}[(X-r)^+], \quad r\in\mathbb{R}.$$

Stop-loss transforms can also be used to order risks. For instance, Kaas (1993) argues that comparing risks in actuarial applications using stop-loss premiums is preferable to directly comparing the distribution functions, since changes in the premiums allow for an immediate interpretation whereas the effects of changes in the distributions are not so obvious. A random variable Y is said to exceed X in stop-loss order if the following condition holds:

$$X\le_{sl}Y \quad :\Leftrightarrow \quad \Pi_X(r)\le\Pi_Y(r), \quad \forall r\in\mathbb{R}.$$

In this case the upper tail of X is uniformly smaller than the upper tail of Y and, hence, X is more attractive than Y. For risks $X,Y\ge 0$, considering a positive retention $r\ge 0$ is sufficient. Furthermore, Y is said to exceed X in convex order if, in addition, the expectations of both random variables exist and are identical:

$$X\le_{cx}Y \quad :\Leftrightarrow \quad X\le_{sl}Y \text{ and } \mathbb{E}[X]=\mathbb{E}[Y].$$

In this case, the actuarial premiums are identical but the stop-loss premiums of X are always less than or equal to the stop-loss premiums of Y with the same retention.

The effect of dependence on stop-loss premiums has, for instance, been investigated in Wang and Dhaene (1998) and Dhaene et al. (2000); the effect on general risk measures was examined in Dhaene et al. (2006). An introduction to dependence ordering in general and the stop-loss order in particular is provided in (Kaas et al., 2008, Chapter 7, pp.149), and detailed discussions can be found in Denuit et al. (2005), Shaked and Shanthikumar (2007), and Müller and Stoyan (2002). Some interesting properties of the stop-loss transform and stop-loss order are summarized in the following remark.

Remark 6.2 (Properties of the stop-loss transform and stop-loss order)
The stop-loss transform of a random variable X can be understood as the weight of the upper tail of the distribution, since it holds following integration by parts:

$$\Pi_X(r) = \int_r^\infty (1 - F_X(s))\,\mathrm{d}s.$$

If $\mathbb{E}[|X|] < \infty$, the transform is decreasing and convex in r and converges to 0 for $r \to \infty$, see (Denuit et al., 2005, Chapter 1.7, Property 1.7.2, p.29). The stop-loss order can equally be defined based on non-decreasing, convex functions, see (Denuit et al., 2005, Chapter 3.4, Proposition 3.4.6, p.152):

$$X \leq_{sl} Y \quad \Leftrightarrow \quad \mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)] \quad \forall f\colon \mathbb{R} \to \mathbb{R} \text{ non-decreasing and convex such that the expectations exist.}$$

Due to this characterization the order is also called increasing convex order. Comparing two risks $X, Y \geq 0$ using expected utility, X is preferred to Y (considering X, Y describe losses) if

$$-\mathbb{E}[u(-X)] \leq -\mathbb{E}[u(-Y)],$$

where for a risk-averse decision-maker u is a non-decreasing and concave function. Setting $v(x) := -u(-x)$, v is non-decreasing and convex. It follows that X is smaller than Y in stop-loss order iff any risk-averse decision-maker would prefer X to Y. Furthermore, if $X \leq_{sl} Y$ then for any zero-utility premium like the exponential premium the same ordering applies. According to (Denuit et al., 2005, Chapter 1.7, Property 1.7.2, p.29) it also holds:

$$X \leq_{sl} Y \quad \Leftrightarrow \quad f(X) \leq_{sl} f(Y) \quad \forall f\colon \mathbb{R} \to \mathbb{R} \text{ non-decreasing and convex,}$$

and for independent sequences $X_1, \ldots, X_d$ and $Y_1, \ldots, Y_d$ with $X_i \leq_{sl} Y_i$, $i = 1, \ldots, d$, and a non-decreasing and component-wise convex function $f\colon \mathbb{R}^d \to \mathbb{R}$ it can be shown that $f(X_1, \ldots, X_d) \leq_{sl} f(Y_1, \ldots, Y_d)$. In particular, the stop-loss order is closed under convolution.

The convex order also has a characterization in terms of convex functions, see (Denuit et al., 2005, Chapter 3.4, Proposition 3.4.3, p.150):

$$X \leq_{cx} Y \quad \Leftrightarrow \quad \mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)] \quad \forall f\colon \mathbb{R} \to \mathbb{R} \text{ convex such that the expectations exist.}$$

Since $\mathbb{E}[(X-r)^+] - \mathbb{E}[(r-X)^+] = \mathbb{E}[X] - r$, it follows from $X \leq_{cx} Y$ that $\mathbb{E}[(r-X)^+] \leq \mathbb{E}[(r-Y)^+]$ for all $r \in \mathbb{R}$. Thus, in addition to uniformly heavier upper tails, convex order also indicates uniformly heavier lower tails of the risks. ▲

While the stop-loss and convex order compare univariate distributions, for the comparison of multivariate distributions the supermodular stochastic order is used. In Dhaene and Goovaerts (1995), the effect of the dependence of two individual risks on the portfolio outcome is investigated and the correlation order for bivariate random vectors is introduced. In Müller (1997) the results are extended to the multivariate case based on supermodular ordering, see also Bäuerle and Müller (1998). The definition as well as some interesting properties are summarized in the following remark.

Remark 6.3 (Supermodular order)
A function $f\colon \mathbb{R}^d \to \mathbb{R}$ is said to be supermodular if it fulfils:

$$f(\boldsymbol{x}) + f(\boldsymbol{y}) \leq f(\boldsymbol{x} \wedge \boldsymbol{y}) + f(\boldsymbol{x} \vee \boldsymbol{y}), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d,$$

where $\wedge$ and $\vee$ denote the component-wise minimum and maximum operator, respectively. The definition says that the increase in the function due to an increase of the function arguments is higher the larger the function arguments are. If f is twice differentiable, the definition is equivalent to the mixed second order partial derivatives being non-negative, see (Denuit et al., 2005, Chapter 3.4, Property 3.4.61, p.179).

If $\boldsymbol{X}, \boldsymbol{Y}$ are d-dimensional random vectors, than $\boldsymbol{X}$ is said to be smaller than $\boldsymbol{Y}$ in supermodular order if the following condition holds:

$$\boldsymbol{X} \leq_{sm} \boldsymbol{Y} \quad :\Leftrightarrow \quad \mathbb{E}[f(\boldsymbol{X})] \leq \mathbb{E}[f(\boldsymbol{Y})] \quad \forall f\colon \mathbb{R}^d \to \mathbb{R} \text{ supermodular such that the expectations exist.}$$

From supermodular order follows that the marginals have the same distribution, see (Denuit et al., 2005, Chapter 6.3.3, p.296), and the ordering is therefore a way of comparing the dependence structure. If $\boldsymbol{X} \leq_{sm} \boldsymbol{Y}$, then the same ordering applies to the pairwise correlation, Kendall's tau, and Spearman's rho (see (Denuit et al., 2005, Chapter 6.3.6, p.297)). Furthermore, it holds according to (Denuit et al., 2005, Chapter 6.3, Proposition 6.3.9 and Corollary 6.3.10, p.299) for non-decreasing and supermodular functions $f\colon \mathbb{R}^d \to \mathbb{R}$:

$$\boldsymbol{X} \leq_{sm} \boldsymbol{Y} \quad \Rightarrow \quad f(\boldsymbol{X}) \leq_{sl} f(\boldsymbol{Y}),$$

and for non-decreasing functions $f_i\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$, $i = 1 \ldots, d$:

$$\boldsymbol{X} \leq_{sm} \boldsymbol{Y} \quad \Rightarrow \quad \sum_{i=1}^{d} f_i(X_i) \leq_{sl} \sum_{i=1}^{d} f_i(Y_i).$$

In particular, if two random vectors are ordered w.r.t. the supermodular order, then the sums of the components are ordered w.r.t. the stop-loss order. ▲

We start with the stop-loss transforms and related orderings of the claim number processes $\boldsymbol{L}$, $\boldsymbol{M}$, and $\boldsymbol{N}$. Even though the stop-loss transforms of the claim numbers without compounding not directly correspond to certain (re-)insurance contracts, they aid the understanding of the impact of cluster arrivals and dependence in the models. Furthermore, the ordering can be employed to compare different subordinator choices and therefore helps to overcome the restrictions on the use of copulas and copula-based dependence measures discussed in Section 4.1. The ordering in the compound models will later be derived from the results found for the claim number processes.

Proposition 6.4 (Stop-loss transform of the claim number process)
Let $\boldsymbol{L}$ be the claim number process proposed in Model (M). Then it holds for the stop-loss transform of each marginal L_t^i at any point in time $t > 0$:

$$\begin{aligned}\mathbb{E}[\big(L_t^i - n\big)^+] &= \mathbb{E}[L_t^i] - n + \sum_{k=0}^{n-1}(n-k)\mathbb{P}(L_t^i = k) \\ &= \lambda_i R^{n-1}\varphi_{\Lambda_t}^{(1)}(0;\lambda_i) - nR^n\varphi_{\Lambda_t}(0;\lambda_i), \quad n \in \mathbb{N}.\end{aligned}$$

Proof. The first equality follows using straightforward calculations from

$$\mathbb{E}[\big(L_t^i - n\big)^+] = \sum_{k=0}^{\infty}(k-n)\mathbb{P}(L_t^i = k) - \sum_{k=0}^{n-1}(n-k)\mathbb{P}(L_t^i = k).$$

For the second equality, consider the summation terms separately:

$$\mathbb{E}[\big(L_t^i - n\big)^+] = \sum_{k=n+1}^{\infty} k\mathbb{P}(L_t^i = k) - n\sum_{k=n+1}^{\infty}\mathbb{P}(L_t^i = k).$$

The second term equals $R^n\varphi_{\Lambda_t}(0;\lambda_i)$ according to Theorem 4.9 and for the first summation term it follows in a similar fashion:

$$\begin{aligned}\sum_{k=n+1}^{\infty} k\mathbb{P}(L_t^i = k) &= \sum_{k=n+1}^{\infty} k\frac{(-\lambda_i)^k}{k!}\varphi_{\Lambda_t}^{(k)}(\lambda_i) = -\lambda_i\sum_{k=n}^{\infty}\frac{(-\lambda_i)^k}{k!}(\varphi_{\Lambda_t}^{(1)})^{(k)}(\lambda_i) \\ &= -\lambda_i R^{n-1}\varphi_{\Lambda_t}^{(1)}(0;\lambda_i).\end{aligned}$$

□

Of course, the above proposition can equally be applied to the aggregate process $\bar{L}$ and the first equality, since it holds for any process on $\mathbb{N}$, also applies to the marginal and aggregate processes of $\boldsymbol{N}$ and $\boldsymbol{M}$. In addition to the representation given in the proposition, the stop-loss transforms can be computed using Panjer's recursion, see (Rolski et al., 1999, Chapter 4.4, p.120). From Remark 6.2 it follows:

$$\mathbb{E}[\big(L_t^i - n\big)^+] = \sum_{k=n}^{\infty}(1 - \mathbb{P}(L_t^i \leq k)) = \mathbb{E}[\big(L_t^i - (n-1)\big)^+] - 1 + \mathbb{P}(L_t^i \leq n-1).$$

Hence, the initial value is $\mathbb{E}[(L_t^i - 0)^+] = \mathbb{E}[L_t^i]$ and the update requires an increasing number of probabilities which follow from Panjer's recursion.

The following theorem summarizes the results on the ordering of the claim number processes $\boldsymbol{N}, \boldsymbol{M}$, and $\boldsymbol{L}$.

Theorem 6.5 (Ordering of the claim number processes)
Let $\boldsymbol{L}$ be the claim number process specified in Model (M), $\boldsymbol{N}$ the corresponding Poisson process with independent components, and $\boldsymbol{M}$ the processes with the same marginals as $\boldsymbol{L}$ but independence in between. Furthermore, assume Assumption (TN) to hold. Then for each component $i = 1, \dots, d$, the Poisson process is smaller in convex order than the respective components of $\boldsymbol{L}$ and $\boldsymbol{M}$:

$$N_t^i \leq_{cx} M_t^i \overset{d}{=} L_t^i, \quad t \geq 0.$$

Furthermore, the d-dimensional process $\boldsymbol{M}$ is smaller than $\boldsymbol{L}$ in supermodular order:

$$\boldsymbol{M}_t \leq_{sm} \boldsymbol{L}_t, \quad t \geq 0,$$

and the aggregate processes $\bar{N}, \bar{M}$, and $\bar{L}$ are increasing in convex order:

$$\bar{N}_t \leq_{cx} \bar{M}_t \leq_{cx} \bar{L}_t, \quad t \geq 0.$$

Proof. In Section 4.1 Shaked's two crossing theorem for mixed Poisson distributions was discussed. The theorem states that the probability mass functions of a Poisson distribution and a mixed Poisson distribution with the same mean cross two times and the mixed distribution has the heavier lower and upper tails while the Poisson distribution has more mass at intermediate values. It follows, then, that the corresponding cumulative distribution functions cross only once, see (Kaas et al., 2008, Chapter 7.3, Theorem 7.3.3, p.155). More precisely, a $c \in \mathbb{N}_0$ exists such that $F_{N_t^i}(k) \leq F_{M_t^i}(k) = F_{L_t^i}(k)$ for $k < c$ and $F_{N_t^i}(k) \geq F_{M_t^i}(k) = F_{L_t^i}(k)$ for $k > c$. Due to the one-cut criterion, see (Rolski et al., 1999, Chapter 3.2.3, Theorem 3.2.4, p.89), one-time crossing of the cumulative distribution functions is a sufficient condition for stop-loss ordering of the random variables involved. Given that the expectations are the same in all cases, this proves the first claim in the theorem. The supermodular ordering between $\boldsymbol{M}_t$ and $\boldsymbol{L}_t$ follows from (Bäuerle and Müller, 1998, Corollary 3.5), which states that if a random vector has conditionally

independent components and the conditional distribution of the components stochastically increases in the mixing variable, then the vector exceeds in supermodular order the one with the same marginal distributions but independence in between. In the given case, the distribution of L_t^i given $\Lambda_t = \hat{\Lambda}_t$ is $\text{Poi}(\lambda_i \hat{\Lambda}_t)$ and the Poisson distribution with higher intensity stochastically dominates the one with lower intensity, thus proving the second claim of the theorem. The ordering of the aggregate processes $\bar{N}_t$ and $\bar{M}_t$ follows due to the independence of the increments from the univariate result according to Remark 6.2 which says that the stop-loss order – and then the convex order as well – is closed under convolution. Finally, the ordering of $\bar{M}_t$ and $\bar{L}_t$ follows from the supermodular order of $\boldsymbol{M}_t$ and $\boldsymbol{L}_t$, see Remark 6.3. □

All relationships between $\boldsymbol{M}$ and $\boldsymbol{L}$ hold regardless of the time-normalization assumption. The results for the ordering w.r.t. $\boldsymbol{N}$ remain valid under the softer condition $\mathbb{E}[\Lambda_1] \geq 1$ if considering stop-loss instead of convex order. It is also interesting to note that from supermodular order follows positive quadrant dependence order, that is (see (Shaked and Shanthikumar, 2007, Chapter 9.A.4, pp.395)):

$$\mathbb{P}(\boldsymbol{M}_t \leq \boldsymbol{k}) \leq \mathbb{P}(\boldsymbol{L}_t \leq \boldsymbol{k}) \quad \text{and} \quad \mathbb{P}(\boldsymbol{M}_t > \boldsymbol{k}) \leq \mathbb{P}(\boldsymbol{L}_t > \boldsymbol{k}) \quad \forall \boldsymbol{k} \in \mathbb{N}_0^d.$$

Considering the difference of the stop-loss transforms, for instance, of $\bar{N}_t$ and $\bar{L}_t$, it holds given time-normalization for retention zero:

$$\Pi_{\bar{L}_t}(0) - \Pi_{\bar{N}_t}(0) = \mathbb{E}[\bar{L}_t] - \mathbb{E}[\bar{N}_t] = 0.$$

As discussed in Remark 6.2, each stop-loss transform converges to zero for increasing retention level, hence, the difference converges to zero as well. In between it holds:

$$\begin{aligned} 0 \leq \Pi_{\bar{L}_t}(r) - \Pi_{\bar{N}_t}(r) &= \int_r^\infty (1 - F_{\bar{L}_t}(s))\,\mathrm{d}s - \int_r^\infty (1 - F_{\bar{N}_t}(s))\,\mathrm{d}s \\ &= \int_r^\infty (F_{\bar{N}_t}(s) - F_{\bar{L}_t}(s))\,\mathrm{d}s. \end{aligned}$$

From the proof of Theorem 6.5 we know that the distribution functions $F_{\bar{N}_t}$ and $F_{\bar{L}_t}$ cross only once and $F_{\bar{L}_t}$ exceeds $F_{\bar{N}_t}$ in the lower range. Consequently, the difference in the stop-loss premium starts at zero, then increases until the retention reaches the level where the distribution functions cross, and then decreases again back to zero. Kaas (1993) provides a formula for the integrated difference in stop-loss transforms of convex ordered random variables $X \leq_{cx} Y$:

$$\int_0^\infty \Pi_Y(r) - \Pi_X(r)\,\mathrm{d}r = \frac{1}{2}(\mathbb{V}\mathrm{ar}[Y] - \mathbb{V}\mathrm{ar}[X]).$$

In addition, a rule of thumb for the ratio of the stop-loss transforms is given for retentions exceeding the mean, i.e. $r > \mathbb{E}[X]$:

$$\frac{\Pi_Y(r)}{\Pi_X(r)} \approx \frac{\mathbb{V}\mathrm{ar}[Y]}{\mathbb{V}\mathrm{ar}[X]}.$$

In the particular case of $\bar{L}_t$, $\bar{M}_t$, and $\bar{N}_t$ it follows:

$$\sum_{k=0}^{\infty}(\Pi_{\bar{L}_t}(k) - \Pi_{\bar{N}_t}(k)) = \frac{|\boldsymbol{\lambda}|^2 t}{2}\,\mathbb{V}\mathrm{ar}[\Lambda_1],$$
$$\sum_{k=0}^{\infty}(\Pi_{\bar{M}_t}(k) - \Pi_{\bar{N}_t}(k)) = \frac{|\boldsymbol{\lambda}^2| t}{2}\,\mathbb{V}\mathrm{ar}[\Lambda_1],$$

and for $k > |\boldsymbol{\lambda}|t$:

$$\frac{\Pi_{\bar{L}_t}(k)}{\Pi_{\bar{N}_t}(k)} \approx 1 + |\boldsymbol{\lambda}|\,\mathbb{V}\mathrm{ar}[\Lambda_1], \quad \frac{\Pi_{\bar{M}_t}(k)}{\Pi_{\bar{N}_t}(k)} \approx 1 + \frac{|\boldsymbol{\lambda}^2|}{|\boldsymbol{\lambda}|}\,\mathbb{V}\mathrm{ar}[\Lambda_1].$$

Figure 6.1 illustrates the differences between the stop-loss transforms of the aggregate processes $\bar{N}, \bar{M}$, and $\bar{L}$ for increasing retention using a two-dimensional example with inverse Gaussian subordinator. On the left-hand side, a parameter setting was chosen where the claim number distribution has variance 4 in both components and correlation 25%; on the right-hand side the variance is 12 and the correlation is 75%. The charts in the top row show the stop-loss transforms in all three cases. As stated in Theorem 6.5, the values for $\bar{N}$ are lower than for $\bar{M}$, and $\bar{L}$ has the highest values. Naturally, the deviations in the high variance and correlation case are much more pronounced: due to the higher variance in the components, $\bar{M}$ deviates more from $\bar{N}$, and the higher correlation increases the distance between $\bar{L}$ and $\bar{M}$. The total deviation between $\Pi_{\bar{L}_t}$ and $\Pi_{\bar{N}_t}$ is 18 and 2 in the high and low variance/correlation case, respectively, and the absolute distance to $\Pi_{\bar{M}_t}$ is split in half between $\Pi_{\bar{L}_t}$ and $\Pi_{\bar{N}_t}$. The charts in the middle row show the pairwise differences between the three processes. As discussed before, the difference starts in zero, converges to zero, and is non-negative in between. In accordance with the ordering, $\Pi_{\bar{L}_t}(k) - \Pi_{\bar{N}_t}(k) \geq \Pi_{\bar{M}_t}(k) - \Pi_{\bar{N}_t}(k)$ holds. The difference between $\Pi_{\bar{L}_t}(k)$ and $\Pi_{\bar{M}_t}(k)$ is smaller than the difference between $\Pi_{\bar{M}_t}(k)$ and $\Pi_{\bar{N}_t}(k)$ at first, but higher

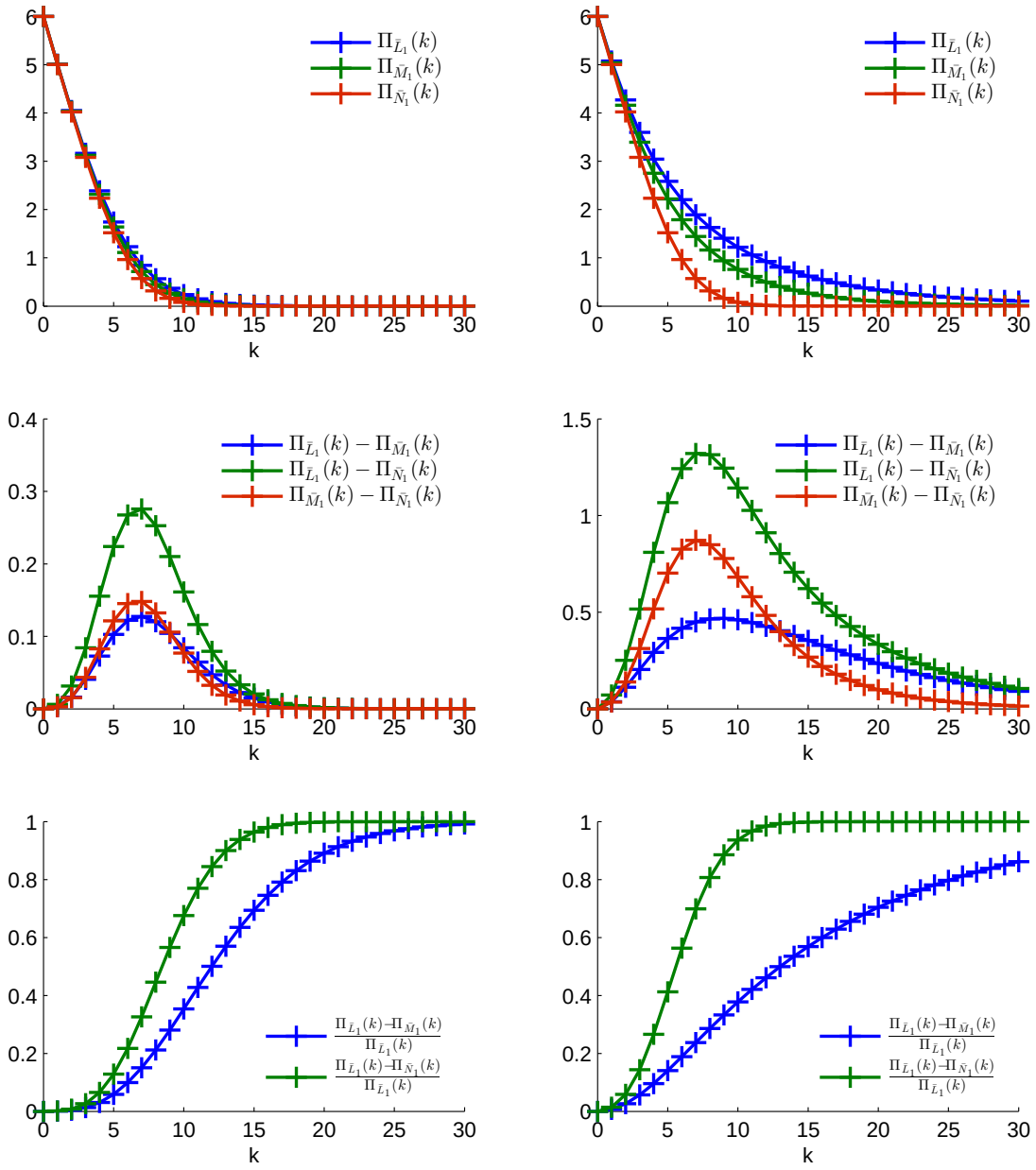


Figure 6.1 Difference of stop-loss transforms of $\bar{L}$, $\bar{M}$, and $\bar{N}$: For dimension $d = 2$, time $T = 1$, intensity $\boldsymbol{\lambda} = (3, 3)$, and an inverse Gaussian subordinator, the figure shows the stop-loss transforms of $\bar{L}$, $\bar{M}$, and $\bar{N}$ (top row), the absolute deviations (middle row), and the relative deviations (bottom row). For the charts on the left-hand-side the subordinator parameters are $b = 0$, $\beta = \eta = 3$, which corresponds to $\mathbb{V}\mathrm{ar}[L_1^1] = \mathbb{V}\mathrm{ar}[L_1^2] = 4$ and $\mathbb{C}\mathrm{or}[L_1^1, L_1^2] = 25\%$. For the charts on the right-hand-side, the subordinator parameters are $b = 0$, $\beta = \eta = 1$, and $\mathbb{V}\mathrm{ar}[L_1^1] = \mathbb{V}\mathrm{ar}[L_1^2] = 12$, $\mathbb{C}\mathrm{or}[L_1^1, L_1^2] = 75\%$. The discrete evaluation points marked with a cross; the connecting lines are provided only to aid visualization.

in the tail. Hence, in the lower range the increase in the variance of the components has a stronger impact, while in the long run the influence of the dependence dominates. The charts in the bottom row show the relative deviation of $\Pi_{\bar{L}_t}$ to both other cases. Whereas the absolute deviation decreases to zero for increasing retention, the relative deviation considering $\bar{N}$ moves quickly to one; if using $\bar{M}$, the increase is slower but heading towards one as well.

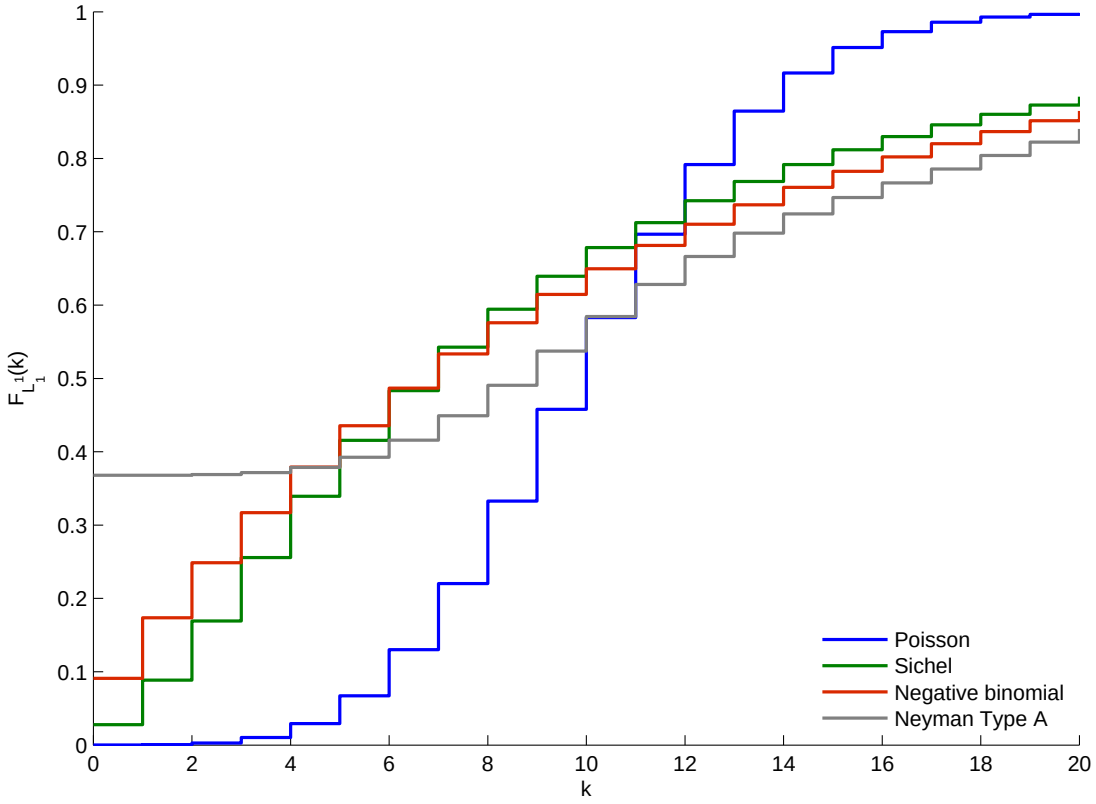


Figure 6.2 Cdf of claim numbers for different subordinator choices: For dimension $d = 1$, intensity $\lambda_1 = 10$, and time horizon $T = 1$, the figure shows the cumulative distribution function of the underlying Poisson process as well as the time-changed process for different choices of the directing Lévy subordinator: inverse Gaussian ($b = 0$, $\beta = \eta = 1$), gamma ($b = 0$, $\beta = \eta = 1$), and Poisson ($b = 0$, $\xi = 1$). The parameters have been chosen such that mean and variance of the three subordinators are one.

Figure 6.2 shows the cumulative distribution functions of the claim numbers in the examples (with time-normalization) given in Figure 4.1 – except for the distribution resulting from a stable subordinator since here time-normalization is not possible. As stated in

the proof of Theorem 6.5, the distribution functions cross only once with the one of the underlying Poisson distribution; they also cross with each other only once. It follows that in the given parameter settings with identical subordinator mean and variance, the claim number distribution resulting from a time-change with the inverse Gaussian subordinator exceeds in convex order the distribution resulting from the gamma subordinator, which exceeds the one from the Poisson subordinator (which exceeds the one without time-change, i.e. the Poisson distribution). The general result of ordering $\boldsymbol{L}$ – component-wise and aggregate – for different subordinator choices is given in the following proposition.

Proposition 6.6 (Ordering for different subordinator choices)
Let $\boldsymbol{L}_t := \boldsymbol{N}_{\Lambda_t}$ be the process proposed in Model (M). Furthermore, let $\boldsymbol{N}^*$ be an independent copy of $\boldsymbol{N}$, Λ^* another Lévy subordinator independent of all other processes, and define $L_t^* := \boldsymbol{N}^*_{\Lambda_t^*}$. Then it holds:

$$\Lambda_t \leq_{sl} \Lambda_t^* \quad \Rightarrow \quad \bar{L}_t \leq_{sl} \bar{L}_t^* \quad \text{and} \quad L_t^i \leq_{sl} L_t^{*i}, \quad i = 1, \ldots, d.$$

If $\mathbb{E}[\Lambda_1] = \mathbb{E}[\Lambda_1^*]$, then the statement holds in convex order.

Proof. Following (Kaas et al., 2008, Chapter 7.4.6, pp.169), stop-loss ordering of the mixing variable leads to stop-loss ordering of the respective mixed Poisson distributions. □

Finally, the ordering of the claim number processes are inherited by the respective compound processes. The results are summarized in the following corollary.

Corollary 6.7 (Ordering of the compound processes)
Let $\boldsymbol{N}$, $\boldsymbol{M}$, and $\boldsymbol{L}$ be as in Theorem 6.5 and let Assumption (TN) hold. Furthermore, let $\boldsymbol{Z}$ be a d-dimensional random vector with independent, non-negative components. Then it holds for the i-th components of the compound processes, $i = 1, \ldots, d$:

$$\sum_{j=1}^{N_t^i} Z_{ij} \leq_{cx} \sum_{j=1}^{M_t^i} Z_{ij} \overset{d}{=} \sum_{j=1}^{L_t^i} Z_{ij}, \quad t \geq 0.$$

The vector of compounds w.r.t. $\boldsymbol{M}_t$ and $\boldsymbol{L}_t$ fulfils in supermodular order:

$$\Big(\sum_{j=1}^{M_t^1} Z_{1j}, \ldots, \sum_{j=1}^{M_t^d} Z_{dj}\Big)' \leq_{sm} \Big(\sum_{j=1}^{L_t^1} Z_{1j}, \ldots, \sum_{j=1}^{L_t^d} Z_{dj}\Big)', \quad t \geq 0.$$

Finally, it holds for the aggregate compound processes:

$$\sum_{i=1}^{d}\sum_{j=1}^{N_t^i} Z_{ij} \leq_{cx} \sum_{i=1}^{d}\sum_{j=1}^{M_t^i} Z_{ij} \leq_{cx} \sum_{i=1}^{d}\sum_{j=1}^{L_t^i} Z_{ij}, \quad t \geq 0.$$

Proof. Stop-loss order is closed under compounding according to (Denuit et al., 2005, Chapter 3.4, Property 3.4.39, p.169), which proves the result for the marginal processes. Following (Denuit et al., 2002, Proposition 2), supermodular order is closed under compounding as well, thus the second statement holds. The results for the aggregate processes follow as in the proof of Theorem 6.5. □

The results regarding the integrated difference of stop-loss transforms presented after Theorem 6.5 can be applied to the compound processes as well. It follows from the above corollary that the actuarial premium for a *stop-loss contract* in the model with claim number process $\boldsymbol{L}$ featuring simultaneous claim arrivals within and between individual components exceeds the premium in the other cases. If such claim arrival patterns are present in the data but neglected in the modelling approach, the premium charged or solvency capital reserved will not adequately cover the risk involved in the contract. Together with Remark 6.2 on the properties of the stop-loss and convex order, it also follows that the same ordering applies to any risk which is a convex function of the aggregate portfolio outcome. In particular, any risk-averse decision-maker will deem the risk, given the simultaneous claim arrivals, more severe and any premium principle stemming from a zero-utility approach, e.g. the expected utility principle, will reflect this ordering – for the direct insurer considering the total portfolio outcome as a risk or for a reinsurer considering any convex function of it as in case of a *stop-loss contract*.

One more premium principle (or risk measure) was mentioned in the beginning and has not yet been covered in the results: the quantile principle for $0 < \epsilon < 1$ (or VaR for $p := 1 - \epsilon$). Here the ordering does not apply regardless of the level p. It was pointed out previously that for $\bar{N}$ and $\bar{L}$ the distribution functions cross once. Hence, neither distribution stochastically dominates the other - in the lower tail the distribution function of $\bar{N}$ exceeds the distribution function of $\bar{L}$, whereas in the upper tail the positions switch. Accordingly, the generalized inverse functions cross once as well and for low levels p the VaR of $\bar{N}$ exceeds the VaR of $\bar{L}$. For high levels, however, the VaR of $\bar{L}$ dominates again.

Stop-loss order is, however, consistent with the tail value at risk which is defined for a random variable X as:

$$TVaR_X(p) := \frac{1}{1-p}\int_p^\infty VaR_X(r)\,\mathrm{d}r, \quad 0 < p < 1.$$

More precisely, it holds for two random variables X, Y according to (Dhaene et al., 2006, Theorem 3.2):

$$X \leq_{sl} Y \quad \Leftrightarrow \quad TVaR_X(p) \leq TVaR_Y(p) \quad \forall 0 < p < 1.$$

The consistency of risk measures with stop-loss order is studied in more detail in Bäuerle and Müller (2006). They show that under certain regularity conditions on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ stop-loss order of random variables leads to the corresponding order in any monotone and law-invariant risk measure.

This section concludes with another example from reinsurance modelling. In a *per event excess-of-loss contract* the reinsurer covers the aggregate claim amount due to a single event exceeding a retention $r > 0$. If the contract coverage is extended the aggregate claim amount in different branches due to a single event exceeding a retention, it is called an *umbrella cover contract*. Considering the Poisson process $\boldsymbol{N}$ for the claim arrivals, which claims stem from the same event is unclear. Given the Poisson cluster process representation of $\boldsymbol{L}_t = \sum_{j=1}^{M_t} \boldsymbol{Y}_j$, however, a natural assumption is that claims originate from the same event if they arrive in the same cluster $\boldsymbol{Y}_j$. Then the risk of a *per event excess-of-loss* or *umbrella cover contract* in the compound model is given as:

$$\sum_{j=1}^{M_t}\Big(\sum_{i=1}^{d}\sum_{k=1}^{Y_{ij}} Z_{ijk} - r\Big)^+,$$

where Z_{ijk} are iid copies of Z_i. Considering $\boldsymbol{M}$ as the claim number process, simultaneous claim arrivals are possible for each individual risk but not for different individual risks. Hence, the modelling approach is unrealistic in the face of an *umbrella cover contract*. Even for a *per event excess-of-loss contract* its applicability is limited; after all, if no simultaneous claim arrivals for different individual risks are expected, then the *per event excess-of-loss coverage* of the portfolio is reduced to a *per event excess-of-loss coverage* of each individual risk.

Ignoring claim amounts, the following proposition provides the expectation of the number of claims exceeding a certain retention due to a single event for the time-changed model.

Proposition 6.8 (Per event excess claim number)
Let $\boldsymbol{L}_t = \sum_{j=1}^{M_t} \boldsymbol{Y}_j$ be the claim number process defined in Model (M) in the Poisson cluster process representation of Corollary 4.19. Then it holds for any retention $n \in \mathbb{N}$:

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^{M_t}(|\boldsymbol{Y}_j| - n)^+\right] &= \mathbb{E}[M_t]\,\mathbb{E}[(|\boldsymbol{Y}| - n)^+] = \mathbb{E}[\bar{L}_t] - tn\Psi_\Lambda(|\boldsymbol{\lambda}|) + t\sum_{k=1}^{n-1}(n-k)\nu_{\bar{L}}(k) \\ &= t(|\boldsymbol{\lambda}|R^{n-1}\Psi_\Lambda^{(1)}(0;|\boldsymbol{\lambda}|) + nR^n\Psi_\Lambda(0;|\boldsymbol{\lambda}|)).\end{aligned}$$

Proof. The proof follows similarly to Proposition 6.4 and Corollary 4.26. □

6.2 Model extensions

As discussed in Chapter 4 and encountered in the real-world data example in Section 5.3, the model presented in this thesis, though it already incorporates many interesting properties of claim arrivals, may prove too restrictive to provide a good fit for some actuarial applications. Three extensions of the model are outlined in this section. While in all cases the price for the greater flexibility is a loss of tractability, a numerical treatment remains possible.

Compound model

For actuarial applications including claim sizes in addition to claim arrivals in the modelling approach is essential. In Section 6.1, compounding of the claim number process $\boldsymbol{L}$ of Model (M) was discussed, but some more thought is given to this important extension in this paragraph. The focus lies again on independent claim sizes that are identically distributed for each component and represented through a non-negative random vector $\boldsymbol{Z}$. Let $\boldsymbol{S} := \{\boldsymbol{S}_t\}_{t\geq 0}$ denote the process consisting of the component-wise aggregate claim amounts w.r.t. the claim arrivals generated from the basic Poisson process $\boldsymbol{N}$ with independent components, that is:

$$\boldsymbol{S}_t := \Big(\sum_{j=1}^{N_t^1} Z_{1j}, \ldots, \sum_{j=1}^{N_t^d} Z_{dj}\Big)', \quad t \geq 0.$$

The components of $\boldsymbol{S}$ are independent, but deriving the distribution of the compound process in closed form is rarely possible, even in the univariate case. The law of total probability tells us that

$$F_{S_t^i}(x) = \sum_{n=0}^{\infty} \mathbb{P}(N_t^i = n) F_{Z_i}^{*n}(x), \quad t \geq 0, \tag{6.2}$$

where $F_{Z_i}^{*n}(x)$ denotes the distribution function of the n-fold convolution of Z_i. The convolution as well as the aggregate distribution have to be mostly dealt with using numerical methods like transform inversion or Panjer's recursion. If a multivariate version with dependence between claim number processes is considered, the complexity of the problem increases accordingly.

With $\boldsymbol{L}_t = \boldsymbol{N}_{\Lambda_t}$ as claim number process, the aggregate process corresponds to the process S if time-changed with the subordinator Λ:

$$\boldsymbol{S}_{\Lambda_t} = \Big(\sum_{j=1}^{N^1_{\Lambda_t}} Z_{1j}, \dots, \sum_{j=1}^{N^d_{\Lambda_t}} Z_{dj}\Big)' = \Big(\sum_{j=1}^{L^1_t} Z_{1j}, \dots, \sum_{j=1}^{L^d_t} Z_{dj}\Big)', \quad t \geq 0.$$

Equation (6.2) holds for the distribution of $S^i_{\Lambda_t}$ if N^i_t is replaced with L^i_t. The process $\boldsymbol{S}_{\Lambda_t}$ is again a compound Poisson process and, using the Poisson cluster process representation $\boldsymbol{L}_t = \sum_{j=0}^{M_t} \boldsymbol{Y}_j$, it can be written:

$$\boldsymbol{S}_{\Lambda_t} = \sum_{j=0}^{M_t} (\sum_{k=1}^{Y_{1j}} Z_{1jk}, \dots, \sum_{k=1}^{Y_{dj}} Z_{djk})', \quad t \geq 0,$$

where $Z_{ijk} \overset{iid}{\sim} Z_i$ are independent copies of the claim sizes. Hence, the component-wise secondary distributions are random convolutions of the claim size distributions, e.g. for the i-th marginal:

$$F_{\sum_{k=1}^{Y_i} Z_{ik}}(x) = \sum_{n=0}^{\infty} \mathbb{P}(Y_i = n) F^{*n}_{Z_i}(x), \quad x \geq 0.$$

In most cases neither representation – as compound process with $\boldsymbol{L}$ or as compound Poisson process with the jump size distribution specified as random convolution – leads to an analytically tractable formula for the distribution of $\boldsymbol{S}_{\Lambda_t}$. Extending Algorithm 3.3 for the independent claim sizes, however, the process $\boldsymbol{S}_{\Lambda_t}$ can be studied using Monte Carlo simulation. In Theorem 6.1 mean and variance of the multivariate and aggregate compound process have been given. Furthermore, the Laplace exponent of the process is known, cf. Proposition 4.1:

$$\Psi_{\boldsymbol{S}_\Lambda}(\boldsymbol{x}) = \Psi_\Lambda(\Psi_{\boldsymbol{S}}(\boldsymbol{x})), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

The exponent of $\boldsymbol{S}$ follows from the univariate result in Example 2.14 and the independence of the components:

$$\Psi_{\boldsymbol{S}_\Lambda}(\boldsymbol{x}) = \sum_{i=1}^{d} \lambda_i (1 - \varphi_{Z_i}(x_i)), \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0},$$

where φ_{Z_i} is the Laplace transform of the claim size Z_i. For the aggregate process $\bar{S}_{\Lambda_t} := \sum_{i=1}^{d} S^i_{\Lambda_t}$ it holds following Theorem 2.1:

$$\Psi_{\bar{S}_\Lambda}(x) = \Psi_{\boldsymbol{S}_\Lambda}(x\mathbf{1}') = \sum_{i=1}^{d} \lambda_i(1 - \varphi_{Z_i}(x)), \quad x \geq 0.$$

Thus, the process can be managed using Laplace inversion techniques or recursions. For instance, Sundt and Vernic (2004) consider recursions for multivariate compound distributions with mixed Poisson primary distributions. Furthermore, Rudolph (2014) examines Panjer's recursion in the static version of the setting studied here and even considers multiple mixing variables (an extension which will be discussed later) and dependence in the claim sizes.

Deterministic time-change

For many applications, assuming stationarity of the claim arrivals may be too restrictive. Adding a deterministic time-change to the model is an easy way to account for predictable fluctuations in claim arrivals, for instance, due to seasonality. Let $\mu\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a non-decreasing and continuous function and consider the process $\boldsymbol{L}^*_t := \boldsymbol{L}_{\mu(t)}$. While $\boldsymbol{L}$ results from a time-change of the homogeneous Poisson process $\boldsymbol{N}$ with Λ, the process $\boldsymbol{L}^*$ corresponds to a time-change of the inhomogeneous Poisson process $\boldsymbol{N}_{\mu(t)}$. Hence, for all but a linear transformation μ the process $\boldsymbol{L}^*_t$ is no longer a Lévy process. More precisely, the increments are no longer stationary, whereas independence is preserved. According to Remark 4.5, $\boldsymbol{L}^*$ is an additive process and the finite dimensional distributions of the process as well as the Laplace transforms can still be calculated.

Often the mean-value function μ is assumed to have an intensity function $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$, that is $g(t) = \int_0^t \lambda(s)\,\mathrm{d}s$, and a model for the intensity rather than the mean-value function is proposed. For instance, for a cyclic Poisson process in Lewis, P. A. W. (1970) and Vere-Jones (1982) intensities of the form

$$\lambda(t) = A \exp\{\rho \cos(\omega_0 t + \Phi)\}$$

for $A > 0$, $\rho > 0$, $\omega_0 > 0$, and $0 < \theta < 2\pi$ are proposed and the estimation of the parameters is studied. Of course, additional sine and cosine waves may be added, if necessary. The exponential is chosen to ensure function positivity. Given (equidistant) grid data $\hat{L}^*_{t_j}$, $j = 1, \ldots, n$, the intensity can be estimated from the fluctuations in the increments $\Delta\hat{L}^*_{t_j}$. Afterwards, as the inverse μ^{-1} exists in this setting, the sample can be transformed to a (no longer equidistant) sample of the process $\boldsymbol{L}$ via $\hat{L}_{s_j} := \hat{L}^*_{\mu^{-1}(s_j)}$

on the grid $s_j := \mu(t_j)$, $j = 1, \ldots, n$. Then a maximum likelihood estimation for the remaining parameters can be carried out as discussed in Chapter 5.

Allowing for a component-specific deterministic time-change $\mu_i \colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ to consider $(L^1_{\mu_1(t)}, \ldots, L^d_{\mu_d(t)})'$ complicates matters. At a fixed point in time, the distribution can still be derived as in Remark 4.3. Over time, however, the new process is only component-wise an additive process; the full vector process is no longer additive: the increments of the transformed process may correspond to increments of the process $\boldsymbol{L}$ that overlap in time for the d components and are, hence, not independent. Consequently, the finite dimensional distributions can no longer be easily derived and the maximum likelihood procedure cannot be applied in the way previously discussed.

Multivariate subordination

As explored in Section 4.1, the dependence of the time-changed process $\boldsymbol{L}$ is governed by the subordinator Λ. In particular, the dependence between any two components is of the same kind. Given a huge cluster arrival in one component (relative to its intensity λ_i) at some point in time, the subordinator most likely has a jump of high magnitude as well. The other components – being affected by the very same subordinator jump – are then very likely to have cluster arrivals of large sizes (relative to their respective intensities) as well. To offer a more flexible dependence structure, multivariate subordination can be considered.

Let $\boldsymbol{\Lambda} := \{\boldsymbol{\Lambda}_t\}_{t \geq 0}$ be a d-dimensional Lévy subordinator with possibly dependent components and consider the component-wise time-change

$$\boldsymbol{N}_{\boldsymbol{\Lambda}_t} := (N^1_{\Lambda^1_t}, \ldots, N^d_{\Lambda^d_t})', \quad t \geq 0.$$

The process $\boldsymbol{N}_{\boldsymbol{\Lambda}} := \{\boldsymbol{N}_{\boldsymbol{\Lambda}_t}\}_{t \geq 0}$ is again a Lévy subordinator and, in particular, a Poisson cluster process (note that due to the independence of the components of $\boldsymbol{N}$, no problems arise for overlapping time periods in contrast to the component-wise deterministic time-change of $\boldsymbol{L}$ investigated previously). Multivariate subordination was studied in Barndorff-Nielsen et al. (2001), and in Semeraro (2008) and Luciano and Semeraro (2007) multivariate subordination was examined in the context of Lévy driven asset models. Mai and Scherer (2012a) introduced a hierarchical extension of the Lévy-frailty model (see Section 3.3) based on multivariate subordination.

The Laplace transform of the process $\boldsymbol{N}_{\boldsymbol{\Lambda}}$ can be derived using the tower rule for conditional expectations:

$$\varphi_{\boldsymbol{N}_{\boldsymbol{\Lambda}_t}}(\boldsymbol{x}) = \mathbb{E}\left[\mathbb{E}\left[\exp\{-\sum_{i=1}^{d} x_i N^i_{\Lambda^i_t}\}|\sigma(\boldsymbol{\Lambda})\right]\right] = \mathbb{E}\left[\prod_{i=1}^{d}\mathbb{E}[\exp\{-x_i N^i_{\Lambda^i_t}\}|\sigma(\Lambda^i)]\right]$$
$$= \mathbb{E}\left[\prod_{i=1}^{d}\exp\{-\Lambda^i_t \Psi_{N^i}(x_i)\}\right] = \exp\{-t\Psi_{\boldsymbol{\Lambda}}((\Psi_{N^1}(x_1),\dots,\Psi_{N^d}(x_d))')\}.$$

Hence, the Laplace exponent is $\Psi_{\boldsymbol{N}_{\boldsymbol{\Lambda}}}(\boldsymbol{x}) = \Psi_{\boldsymbol{\Lambda}}((\Psi_{N^1}(x_1),\dots,\Psi_{N^d}(x_d))')$. The expression for the distribution of the process at a fixed point in time can also be recovered considering partial derivatives of the multivariate Laplace transform $\varphi_{\boldsymbol{\Lambda}_t}$ of $\boldsymbol{\Lambda}_t$, see Corollary 4.12:

$$\mathbb{P}(\boldsymbol{N}_{\boldsymbol{\Lambda}_t} = \boldsymbol{k}) = \frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\varphi^{(\boldsymbol{k})}_{\boldsymbol{\Lambda}_t}(\boldsymbol{\lambda}), \quad \boldsymbol{k} \in \mathbb{N}_0^d.$$

As the process is a Lévy process, the finite dimensional distributions follow similarly to Corollary 4.4. The transformation of the Lévy characteristics in case of multivariate subordination was established in Barndorff-Nielsen et al. (2001). Using their results, the drift of $\boldsymbol{N}_{\boldsymbol{\Lambda}}$ is zero and the Lévy measure $\nu_{\boldsymbol{N}_{\boldsymbol{\Lambda}}}$ has – as to be expected – a representation in terms of partial derivatives of the Laplace exponent $\Psi_{\boldsymbol{\Lambda}}$ of $\boldsymbol{\Lambda}$:

$$\nu_{\boldsymbol{N}_{\boldsymbol{\Lambda}}}(\boldsymbol{k}) = -\frac{(-\boldsymbol{\lambda})^{\boldsymbol{k}}}{\boldsymbol{k}!}\Psi^{(\boldsymbol{k})}_{\boldsymbol{\Lambda}}(\boldsymbol{\lambda}), \quad \boldsymbol{k} \in \dot{\mathbb{N}}_0^d.$$

Given these results, the Poisson cluster process representation of $\boldsymbol{N}_{\boldsymbol{\Lambda}}$ can be derived as in Corollary 4.19.

In general, multivariate Lévy subordinators can be constructed from independent univariate subordinators using Lévy copulas, but a factor approach is more common due to its tractability and comprehensibility. For this purpose, let $\tilde{\boldsymbol{\Lambda}} = (\tilde{\Lambda}^1,\dots,\tilde{\Lambda}^r)'$ be a Lévy subordinator with independent marginal processes representing $r \in \mathbb{N}$ risk factors. The degree to which each component of the portfolio is affected by the factors is specified through some non-negative weight matrix $W \in \mathbb{R}^{d\times r}_{\geq 0}$. Using this approach, the number of parameters in the model can increase significantly, so a careful selection of the number of risk factors and the structure of the weight matrix is necessary. For example, Luciano and Semeraro (2007) consider $d+1$ risk factors where d factors are idiosyncratic for each component and one common factor affects all components. To obtain time-normalization for the factor approach, it is sufficient to assume that all risk factors fulfil Assumption (TN), i.e. $\mathbb{E}[\tilde{\Lambda}^k_1] = 1$ for $k = 1,\dots,r$, and the sum of the component-wise weights is one, i.e. $\sum_{k=1}^{r} W_{ik} = 1$ for $i = 1,\dots d$. The multivariate subordinator for the time-change is then

defined as the linear transformation $\boldsymbol{\Lambda} := W\tilde{\boldsymbol{\Lambda}}$. The so-defined process $\boldsymbol{\Lambda}$ is indeed a Lévy subordinator and its Laplace transform follows from Theorem 2.1:

$$\varphi_{\boldsymbol{\Lambda}_t}(\boldsymbol{x}) = \varphi_{\tilde{\boldsymbol{\Lambda}}_t}(W'\boldsymbol{x}) = \exp\Big\{ - t \sum_{k=1}^{r} \Psi_{\tilde{\Lambda}^k}\Big(\sum_{i=1}^{d} W_{ik} x_i \Big) \Big\}, \quad \boldsymbol{x} \in \mathbb{R}^d_{\geq 0}.$$

Hence, the Laplace exponent is $\Psi_{\boldsymbol{\Lambda}}(\boldsymbol{x}) = \sum_{k=1}^{r} \Psi_{\tilde{\Lambda}^k}\Big(\sum_{i=1}^{d} W_{ik} x_i \Big)$ and he Lévy characteristics can be concluded from (Cont and Tankov, 2003, Chapter 4.2, Theorem 4.1, p.105):

$$b_{\boldsymbol{\Lambda}} = W(b_{\tilde{\Lambda}^1}, \ldots, b_{\tilde{\Lambda}^r})', \quad \nu_{\boldsymbol{\Lambda}}(B) = \nu_{\tilde{\boldsymbol{\Lambda}}}(\{\boldsymbol{x} \in \dot{\mathbb{R}}^r_{\geq 0} \colon W\boldsymbol{x} \in B\}), \quad B \in \mathcal{B}(\dot{\mathbb{R}}^d_{\geq 0}).$$

The extension of the presented model using multivariate subordination allows to formulate a time-dynamic extension of the widely used CreditRisk$^+$ model, as discussed in the following remark.

Remark 6.9 (Dynamic extension of the CreditRisk$^+$ framework)
In Suisse (1997), Credit Swiss Financial Products introduced the CreditRisk$^+$ model for credit-risky portfolios. It is a static Poisson mixture model and, hence, techniques from actuarial mathematics can be employed to calculate the portfolio loss distribution. A brief overview of the model is provided in the following to point out the resemblance with the process studied in this thesis. For details about the original set-up and many extensions that have been studied since, see Gundlach and Lehrbass (2004).

Considering a credit-risky portfolio with d counterparties or assets, let J_i, $i = 1, \ldots, d$, be default indicators following a Bernoulli distribution with default probability

$$\lambda_i := \mathbb{P}(J_i = 1) \geq 0.$$

For simplicity, the loss given default of each asset is, as in Section 3.3, assumed to be a deterministic constant equal for all assets. Hence, it is again sufficient to consider as portfolio loss the sum

$$\bar{L} := \sum_{i=1}^{d} J_i.$$

The CreditRisk$^+$ approach utilizes the Poisson approximation of the Bernoulli variables, i.e. the default indicators are replaced with Poisson variables $N_i \sim \mathrm{Poi}(\lambda_i)$, hence, allowing the asset to default 'more than once'. For the Poisson approximation to be justified, the expected default probabilities λ_i need to be small making successive defaults unlikely. Then the default probability fulfils

$$\mathbb{P}(N_i > 0) = 1 - \exp\{-\lambda_i\} \approx \lambda_i, \quad i = 1, \dots, d.$$

Note that the expected loss is unaffected by the switch from Bernoulli to Poisson variables with the same mean.

Dependence in this setting is introduced using Poisson mixtures with a factor approach for the mixing variables. The $r \in \mathbb{N}$ independent risk factors are assumed to follow a gamma distribution:

$$\Lambda_k \sim \mathrm{Gamma}(\beta_k, \beta_k), \quad k = 1, \dots, r,$$

where $\beta_k > 0$. Hence, $\mathbb{E}[\Lambda_k] = 1$ and $\mathbb{V}\mathrm{ar}[\Lambda_k] = 1/\beta_k$. The influence of the risk factors on each component is specified using a weight matrix $W \in \mathbb{R}^{d \times r}$ such that $\sum_{k=1}^r W_{ik} = 1$. The variables N_i are then supposed to be conditionally independent and Poisson distributed with intensity $\lambda_i \sum_{k=1}^r W_{ik} \Lambda_k$, $i = 1, \dots, d$. Due to the chosen parameter setting, the approximation of the expected default intensities from the independent Poisson case remains valid for the mixtures.

For $r = 1$ the portfolio loss $\bar{L}$ has a negative binomial distribution. For arbitrary r, the distribution can be computed using Panjer's recursion after observing that $\bar{L}$ is equal in distribution to a sum of r independent negative binomial random variables. This approach is studied in a more general setting in Rudolph (2014).

Obviously, the CreditRisk$^+$ approach corresponds to a static version of the model studied in this thesis if multivariate subordination is considered. More precisely, if the processes $\tilde{\Lambda}^k = \{\tilde{\Lambda}^k_t\}_{t \geq 0}$ are independent gamma subordinators with no drift and $\beta_k = \eta_k$, $k = 1, \dots, r$, and the directing process is $W\tilde{\boldsymbol{\Lambda}}$, then the time-changed process $\boldsymbol{N}_{\boldsymbol{\Lambda}}$ is in $t = 1$ equivalent to the CreditRisk$^+$ model. Following this observation it is convenient to switch from the gamma distribution to other infinitely divisible distributions for the risk factors. Moreover, this time-dynamic extension of the CreditRisk$^+$ model enables a consistent and arbitrage-free valuation of credit derivatives, such as collateralized debt obligations, with arbitrary maturities, and for the calibration of the model the full term structure of prices can be used. In addition, dynamic hedging of derivative positions is possible in this setting. ▲

7 Conclusion and outlook

In this thesis, a multivariate model for dependent claim arrivals was presented and investigated in detail. The appeal of the model lies in its ability to balance the need for a plausible, yet tractable set-up with dynamics that are well understood and easy to verify. Starting point of the construction was the classical modelling approach consisting of independent Poisson processes; by introducing a Lévy subordinator as common stochastic time, dependence between the components and the possibility of claim cluster arrivals were generated. The marginals of the time-changed process can be categorized as Cox as well as Poisson cluster processes, and for the claim number distribution at one point in time classical distributions such as the negative binomial and Sichel distribution were recovered. Through the possibility of large cluster arrivals as well as the increasing degree of dependence in case of extreme scenarios, the model is able to account for catastrophic events. Since the source of the dependence and claim clusters is explicitly known and governed by the subordinator parameters, which are only few, the dynamics of the model are very transparent and the applicability of the model to actuarial applications can be readily assessed.

Due to the specific choice of a Lévy subordinator as joint operational time, the model remains highly tractable. Many distribution-related quantities of both, the Lévy as well as the compound Poisson representation, were found in closed form and approximations for tails and large portfolio distributions were presented as well. In addition, an efficient sampling algorithm was discussed, which offers the possibility to numerically evaluate expressions that cannot be analytically solved. Four estimation methods for the model parameters were introduced and the quality was examined in an extensive simulation study. These methods were also used to evaluate the applicability of the model to the Danish fire insurance data set. Furthermore, the impact of the model on premium calculations for insurance and reinsurance products as well as other actuarial risk measures was examined and it was shown that ignoring positive dependence and cluster arrivals in the modelling approach, though present in the data, will lead to underestimation of the risks and the adequate compensation for the insurer.

The basic time-changed model was thoroughly analysed to provide an in-depth understanding of the model properties and limitations. Naturally, for some actuarial applications the model flexibility is not yet rich enough to account for all the relevant properties of the claim count data. In particular, since the subordinator directs the dependence between all components as well as the variability in each component and the Lévy nature

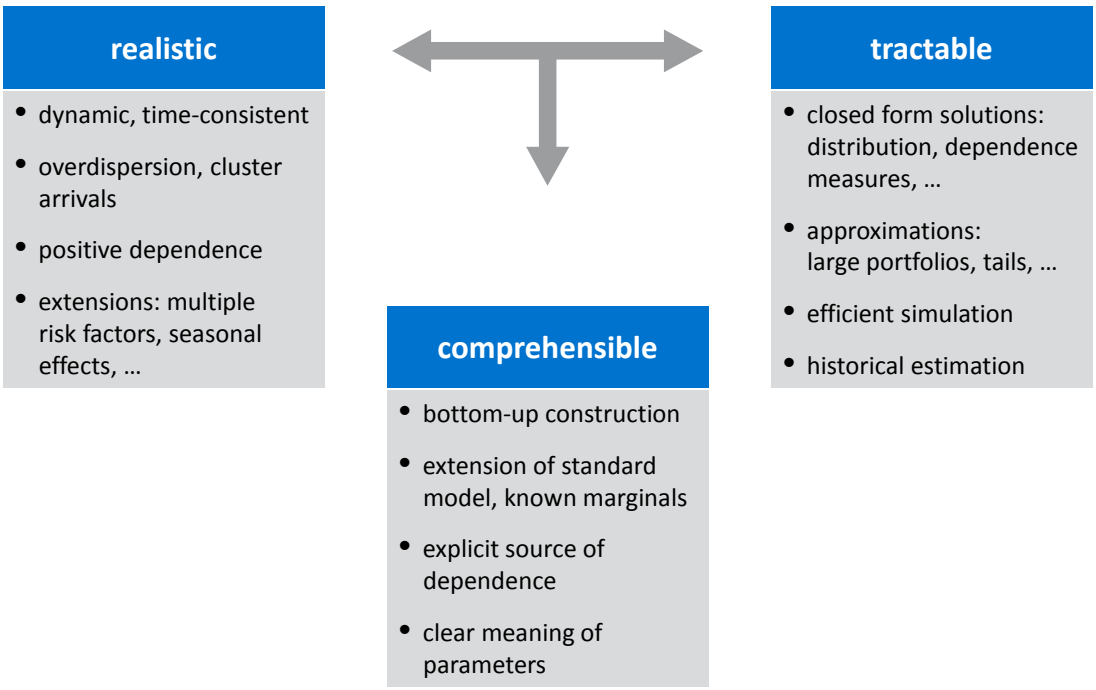


Figure 7.1 Overview of model features.

of the model calls for independent and stationary increments of the claim numbers, the model may prove too restrictive for some applications, as could be seen in the Danish fire insurance data. However, the constructive set-up of the model provides the opportunity to easily formulate extensions for more advanced applications. As an example, the incorporation of iid claim sizes, a deterministic time-change, as well as the introduction of multiple subordinators were outlined in this thesis. Though these extensions lose some of the analytical tractability, a numerical treatment using the sampling algorithm remains possible.

For further research, the proposed extensions could be examined in more detail and the problem of estimating the parameters needs to be addressed. In general, it may be worth exploring further actuarial questions such as how the model affects the optimal retention levels in reinsurance contracts. The important topic of ruin theory has also not been covered in this thesis. The process normalized by time converges, as every subordinator, to the expected value after one unit of time (see (Sato, 1999, Chapter 36, pp.245)), i.e.

$$\frac{1}{t}\boldsymbol{L}_t \xrightarrow{a.s.} \mathbb{E}[\Lambda_1]\boldsymbol{\lambda}, \quad t \to \infty.$$

However, this result needs to be supplemented by the asymptotic distribution in order to calculate ruin probabilities and solvency capital requirements. The asymptotic distribution of general Cox processes is, for instance, discussed in Bening and Korolev (2002). The model may also prove useful for pricing and managing alternative risk transfer products such as insurance-linked bonds and derivatives. For this purpose, arbitrage-free pricing rather than premium calculation becomes necessary and calibration to market prices replaces historical estimation for determining the model parameters.

In this thesis, the focus was primarily on using the model for actuarial problems. One application in the area of credit risk modelling, however, was given as well, namely the dynamic extension of the CreditRisk$^+$ framework. This application should be studied more thoroughly, thereby treating similar questions as for the alternative risk transfer products. Furthermore, it may be worth examining the usefulness of the model in other areas like electronic order book modelling, operational risk management, and inventory control problems, to name only a few.

A Additional output from the estimation study

Setting: inverse Gaussian subordinator (no drift)

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 14.5/4.5$	
T		1	10	1	10	1	10	1	10
1-2	low	7.3	2.1	12	3.3	18	4.7	2.1	0.57
	high	4.7	1.8	7.2	2.5	9.3	3	0.89	0.34
1-3	low	7.3	2.1	12	3.3	18	4.7	2.4	0.71
	high	4.7	1.8	7.2	2.5	9.3	3	0.92	0.35
1-4	low	11	3.3	15	4.5	22	5.9	2.7	0.82
	high	21	5.4	30	7	36	8.3	1.5	0.48
2-3	low	0.0008	0.0021	0.0012	0.0026	0.0016	0.0031	1.2	0.38
	high	0.003	0.0048	0.0048	0.0071	0.0059	0.0093	0.22	0.064
2-4	low	6.1	1.8	6.8	2.1	7.5	2.3	1.5	0.47
	high	19	4.5	27	5.9	33	7.1	0.78	0.19
3-4	low	6.1	1.8	6.8	2.1	7.5	2.3	0.73	0.21
	high	19	4.5	27	5.9	33	7.1	0.71	0.17

Table A.1 Differences between estimators from all methods for inverse Gaussian subordinator (no drift): The table summarizes the root-mean-squared differences between the estimators from two Methods (Mi) and (Mj) (noted $i-j$) in case of an inverse Gaussian subordinator without drift in the setting with low and high variance as well as for short and long time horizon (2 significant digits).

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 14.5/4.5$	
T		1	10	1	10	1	10	1	10
1-2	low	0.40	0.36	0.33	0.37	0.33	0.33	0.39	0.36
	high	0.45	0.41	0.40	0.45	0.50	0.42	0.31	0.29
1-3	low	0.40	0.36	0.33	0.37	0.33	0.33	0.32	0.29
	high	0.45	0.41	0.40	0.45	0.50	0.42	0.30	0.29
1-4	low	0.50	0.45	0.43	0.47	0.38	0.40	0.40	0.35
	high	0.66	0.60	0.61	0.56	0.64	0.64	0.39	0.37
2-3	low	0.47	0.52	0.46	0.46	0.51	0.52	0.39	0.37
	high	0.47	0.50	0.49	0.54	0.46	0.55	0.41	0.47
2-4	low	0.59	0.61	0.66	0.59	0.58	0.62	0.45	0.43
	high	0.69	0.66	0.65	0.62	0.69	0.67	0.57	0.60
3-4	low	0.59	0.61	0.66	0.59	0.58	0.62	0.61	0.59
	high	0.69	0.66	0.65	0.62	0.69	0.67	0.60	0.62

Table A.2 Pitman closeness criterion for estimators from different methods for inverse Gaussian subordinator (no drift): The table summarizes the empirical probabilities that the absolute deviations of estimators from Method (Mi) are less then for another Method (Mj) (noted $i - j$) in the setting of an inverse Gaussian subordinator without drift, for low and high variance as well as for short and long time horizon (2 significant digits).

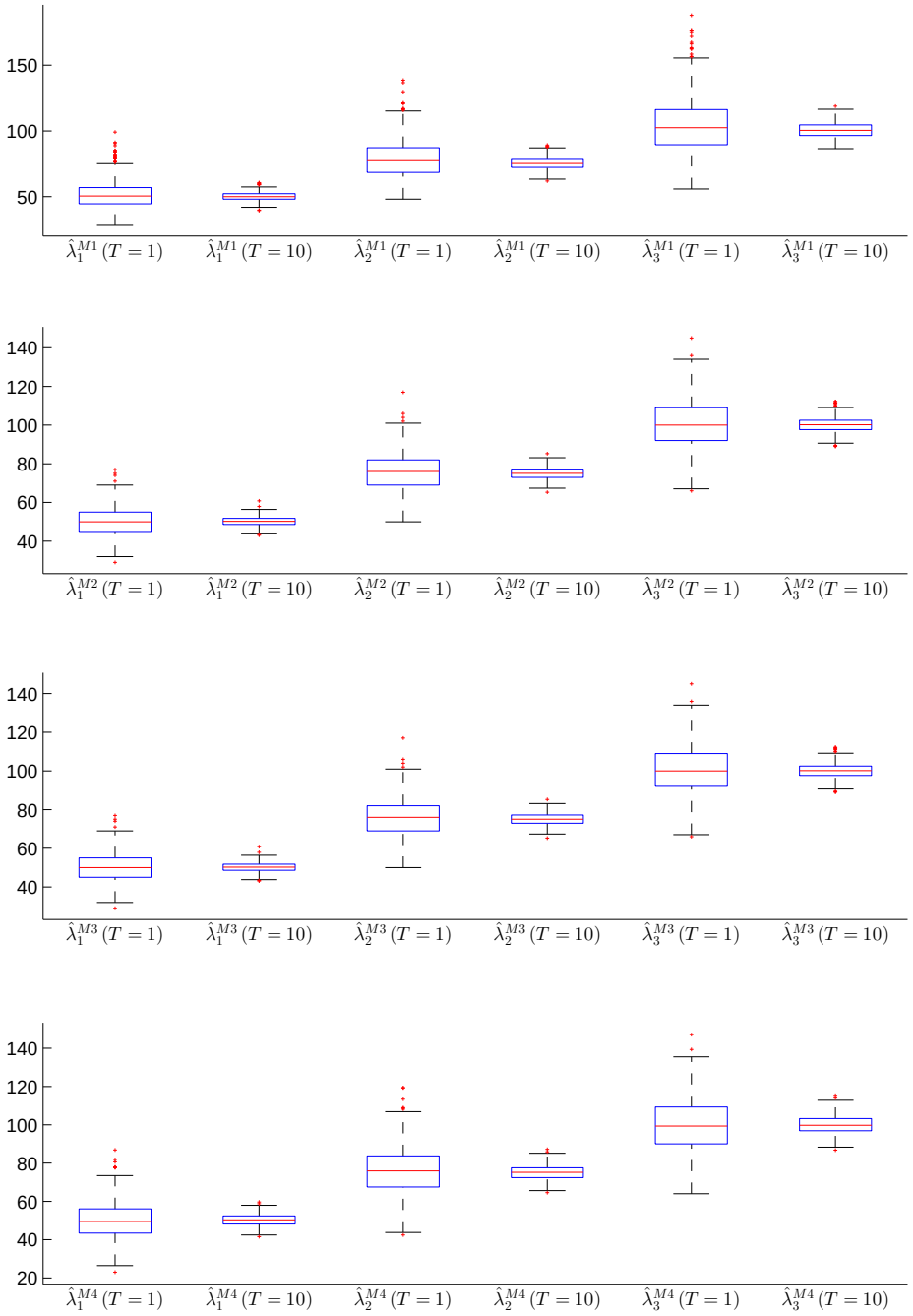


Figure A.1 Boxplots of intensity estimators for low variance inverse Gaussian subordinator (no drift): The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a low variance Inverse Gaussian subordinator (no drift) for short and long time horizon.

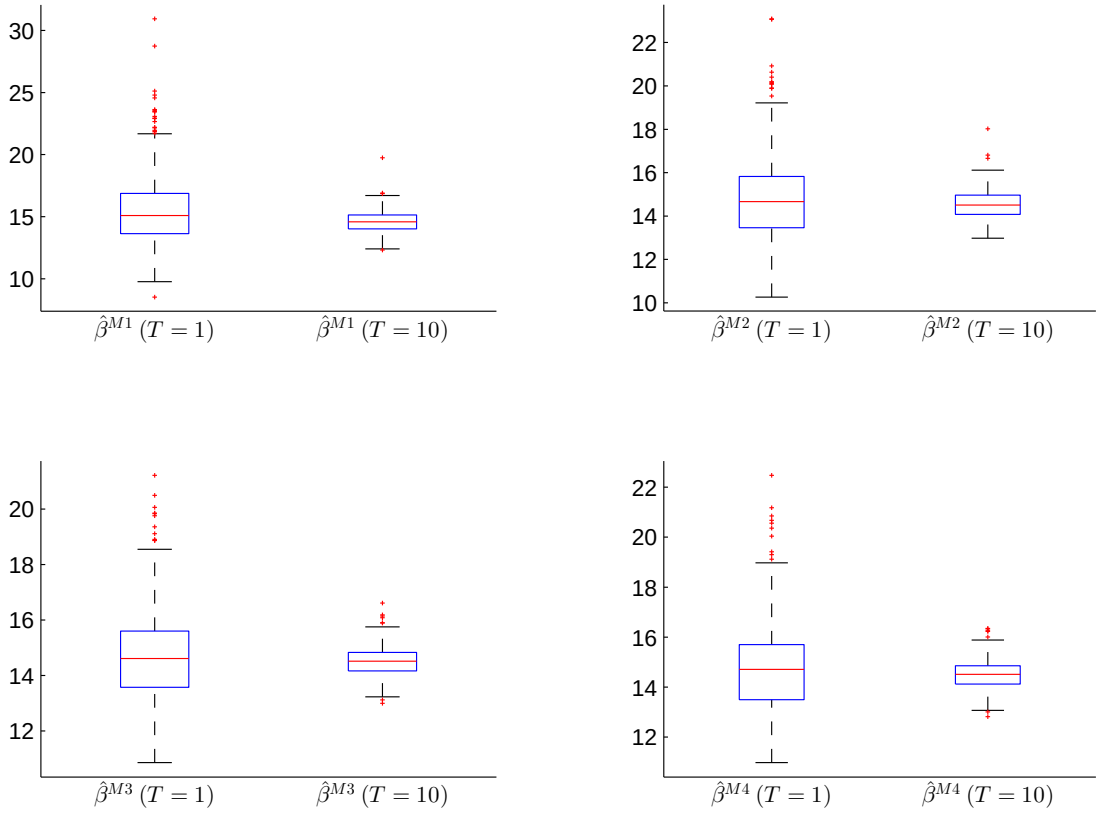


Figure A.2 Boxplots of subordinator parameter estimators for low variance inverse Gaussian subordinator (no drift): The figure shows boxplots of the subordinator parameter estimators from all estimation methods in the setting with a low variance Inverse Gaussian subordinator (no drift) for short and long time horizon.

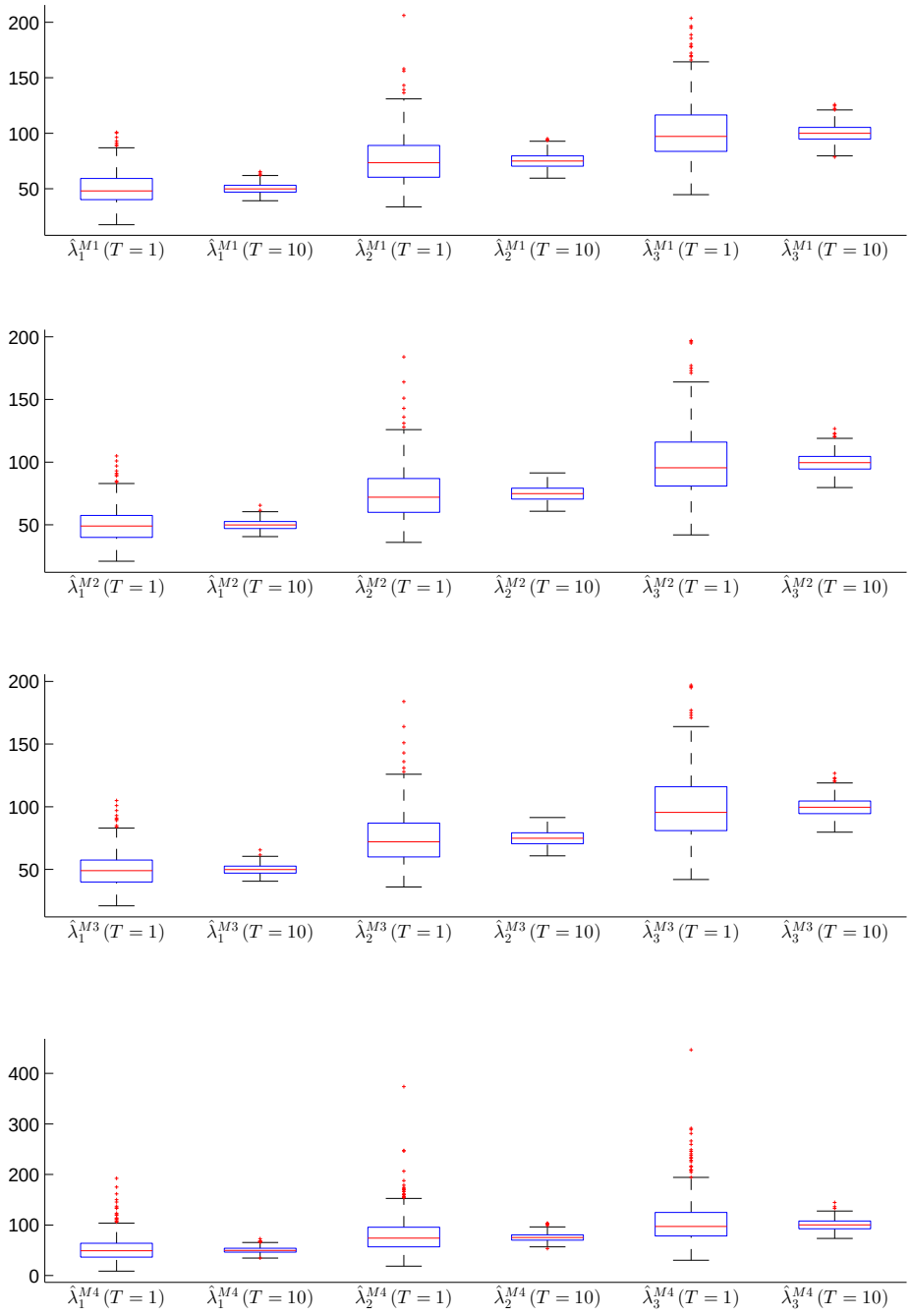


Figure A.3 Boxplots of intensity estimators for high variance inverse Gaussian subordinator (no drift): The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a high variance Inverse Gaussian subordinator (no drift) for short and long time horizon.

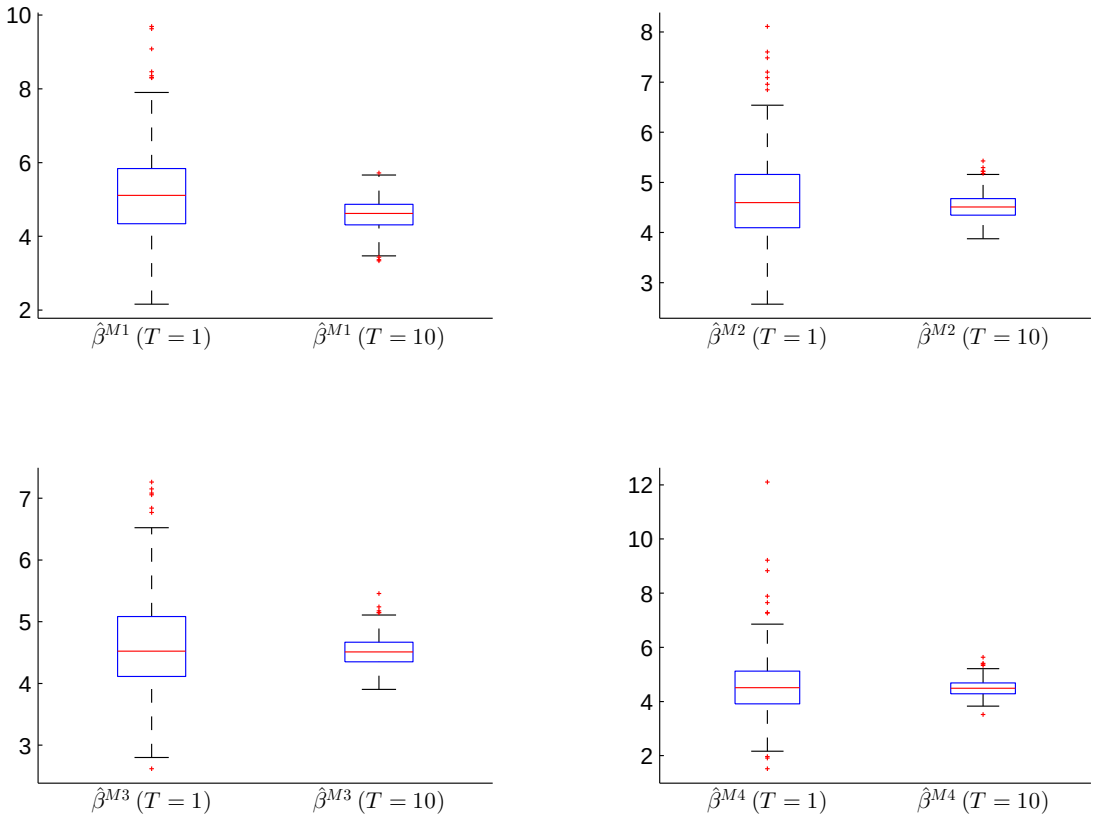


Figure A.4 Boxplots of subordinator parameter estimators for high variance inverse Gaussian subordinator (no drift): The figure shows boxplots of the subordinator parameter estimators from all estimation methods in the setting with a high variance Inverse Gaussian subordinator (no drift) for short and long time horizon.

Setting: inverse Gaussian subordinator with drift

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 6.9/2.1$	
T		1	10	1	10	1	10	1	10	1	10
1-2	low	8.9	2.5	16	3.8	20	5.4	0.34	0.098	6.2	2
	high	6.6	2.3	9.8	3.1	13	4.3	0.3	0.16	2.2	1.2
1-3	low	8.9	2.5	16	3.8	20	5.4	0.31	0.11	5.5	2.2
	high	6.6	2.3	9.8	3.1	13	4.3	0.29	0.17	2.3	1.2
1-4	low	13	3.5	21	4.9	26	6.5	0.33	0.12	6.5	2.5
	high	65	6.9	98	9.9	130	13	0.32	0.16	2.7	1.4
2-3	low	0.00069	0.0035	0.00094	0.0047	0.0014	0.0055	0.18	0.05	4.1	0.94
	high	0.011	0.0057	0.015	0.0084	0.022	0.011	0.066	0.016	0.8	0.15
2-4	low	7.2	1.6	11	1.8	15	2.1	0.21	0.059	5.7	1.3
	high	65	6.1	98	9	130	12	0.13	0.033	1.6	0.4
3-4	low	7.2	1.6	11	1.8	15	2.1	0.082	0.02	2.8	0.58
	high	65	6.1	98	9	130	12	0.13	0.033	1	0.29

Table A.3 Differences between estimators from all methods for inverse Gaussian subordinator with drift: The table summarizes the root-mean-squared Differences between the estimators from two Methods (Mi) and (Mj) (noted $i - j$) in case of an inverse Gaussian subordinator with drift in the setting with low and high variance as well as for short and long time horizon (2 significant digits).

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 6.9/2.1$	
T		1	10	1	10	1	10	1	10	1	10
1-2	low	0.36	0.35	0.34	0.31	0.31	0.28	0.42	0.25	0.46	0.27
	high	0.43	0.43	0.40	0.41	0.43	0.41	0.18	0.12	0.20	0.12
1-3	low	0.36	0.35	0.34	0.31	0.31	0.28	0.23	0.15	0.31	0.22
	high	0.43	0.44	0.40	0.41	0.43	0.41	0.17	0.11	0.17	0.11
1-4	low	0.45	0.40	0.42	0.35	0.38	0.35	0.26	0.20	0.37	0.26
	high	0.65	0.60	0.64	0.63	0.66	0.62	0.24	0.15	0.25	0.19
2-3	low	0.44	0.51	0.48	0.52	0.52	0.56	0.28	0.31	0.35	0.36
	high	0.50	0.46	0.49	0.47	0.49	0.49	0.41	0.42	0.41	0.39
2-4	low	0.57	0.56	0.60	0.57	0.57	0.60	0.36	0.36	0.46	0.45
	high	0.71	0.70	0.71	0.69	0.67	0.66	0.61	0.60	0.57	0.61
3-4	low	0.57	0.56	0.60	0.57	0.57	0.60	0.60	0.58	0.63	0.60
	high	0.71	0.70	0.71	0.69	0.67	0.66	0.67	0.65	0.65	0.73

Table A.4 Pitman closeness criterion for estimators from different methods for inverse Gaussian subordinator with drift: The table summarizes the empirical probabilities that the absolute deviations of estimators from Method (Mi) are less then for another Method (Mj) (noted $i-j$) in the setting of an inverse Gaussian subordinator with with drift, for low and high variance as well as for short and long time horizon (2 significant digits).

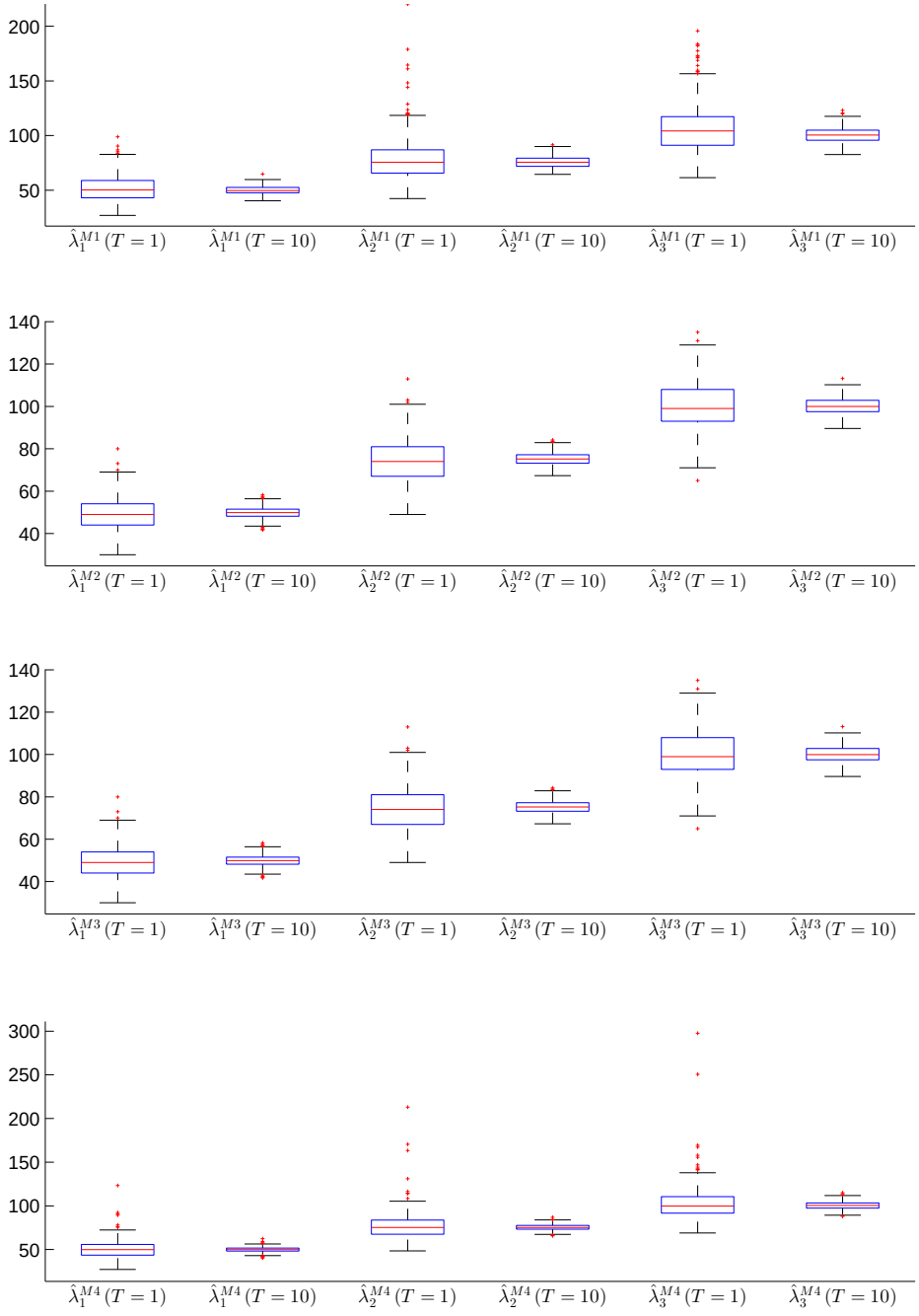


Figure A.5 Boxplots of intensity estimators for low variance inverse Gaussian subordinator with drift: The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a low variance Inverse Gaussian subordinator with drift for short and long time horizon.

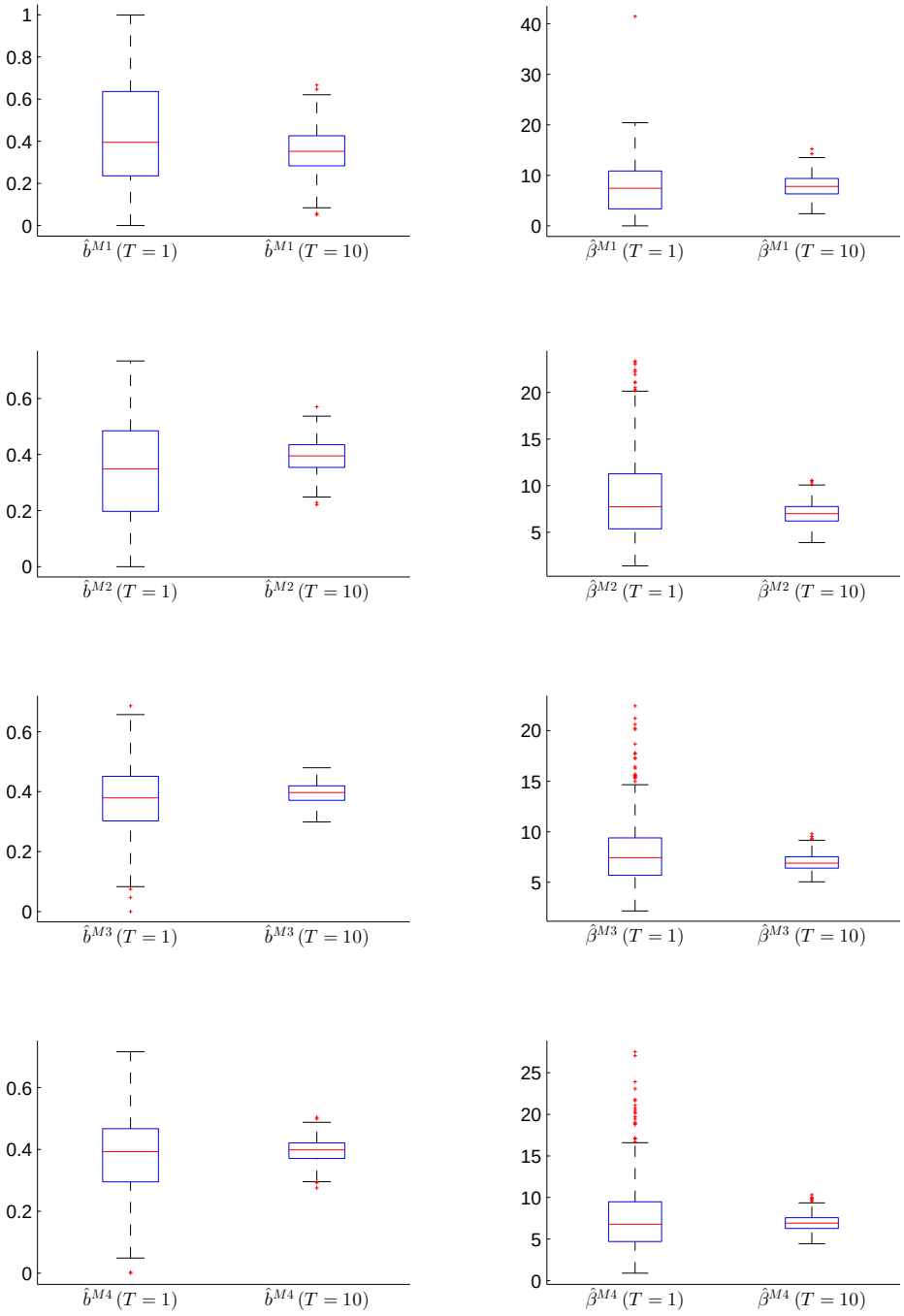


Figure A.6 Boxplots of subordinator parameter estimators for low variance inverse Gaussian subordinator with drift: The figure shows boxplots of the estimators for the subordinator parameters from all estimation methods in the setting with a low variance Inverse Gaussian subordinator with drift for short and long time horizon.

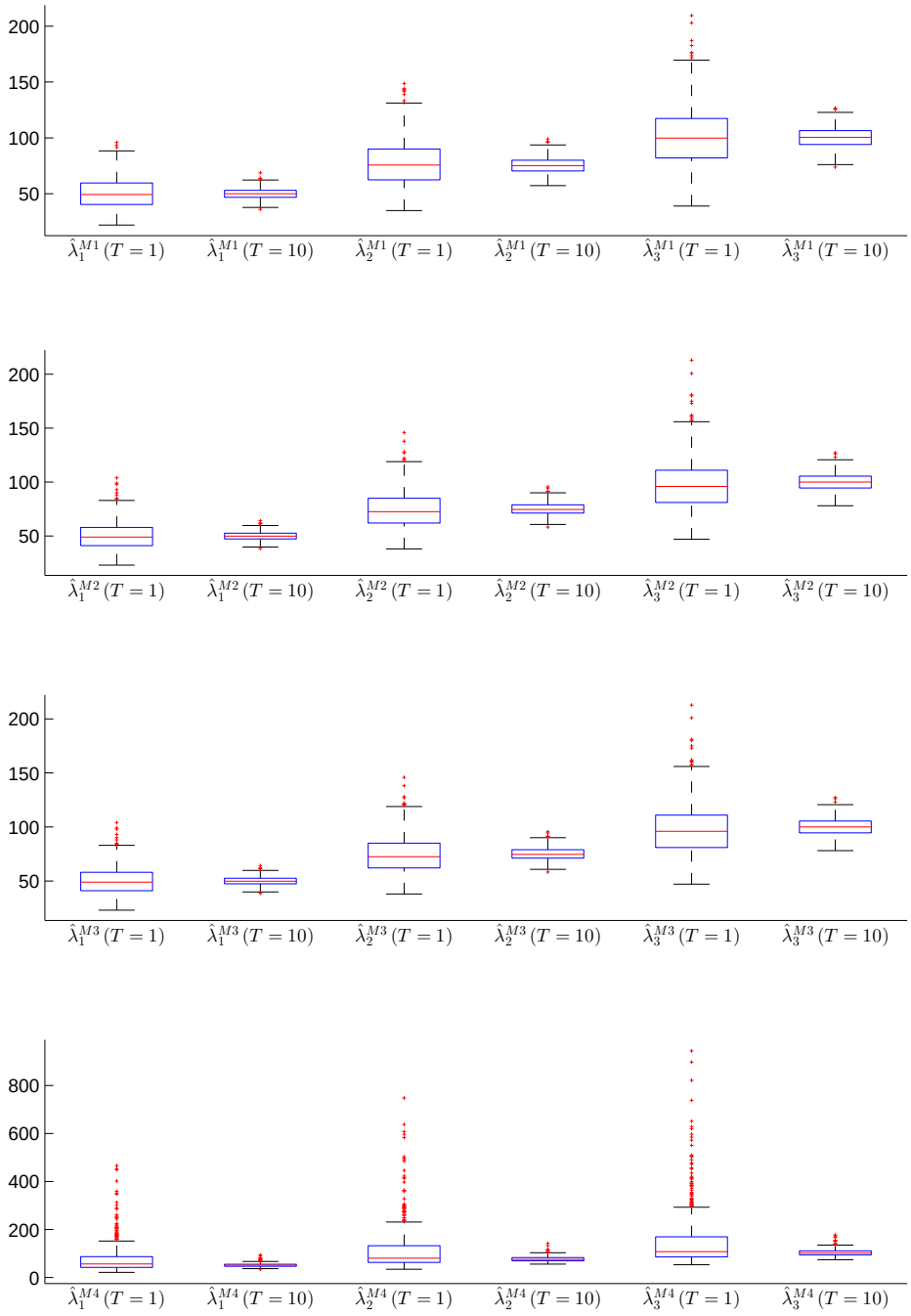


Figure A.7 Boxplots of intensity estimators for high variance inverse Gaussian subordinator with drift: The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a high variance Inverse Gaussian subordinator with drift for short and long time horizon.

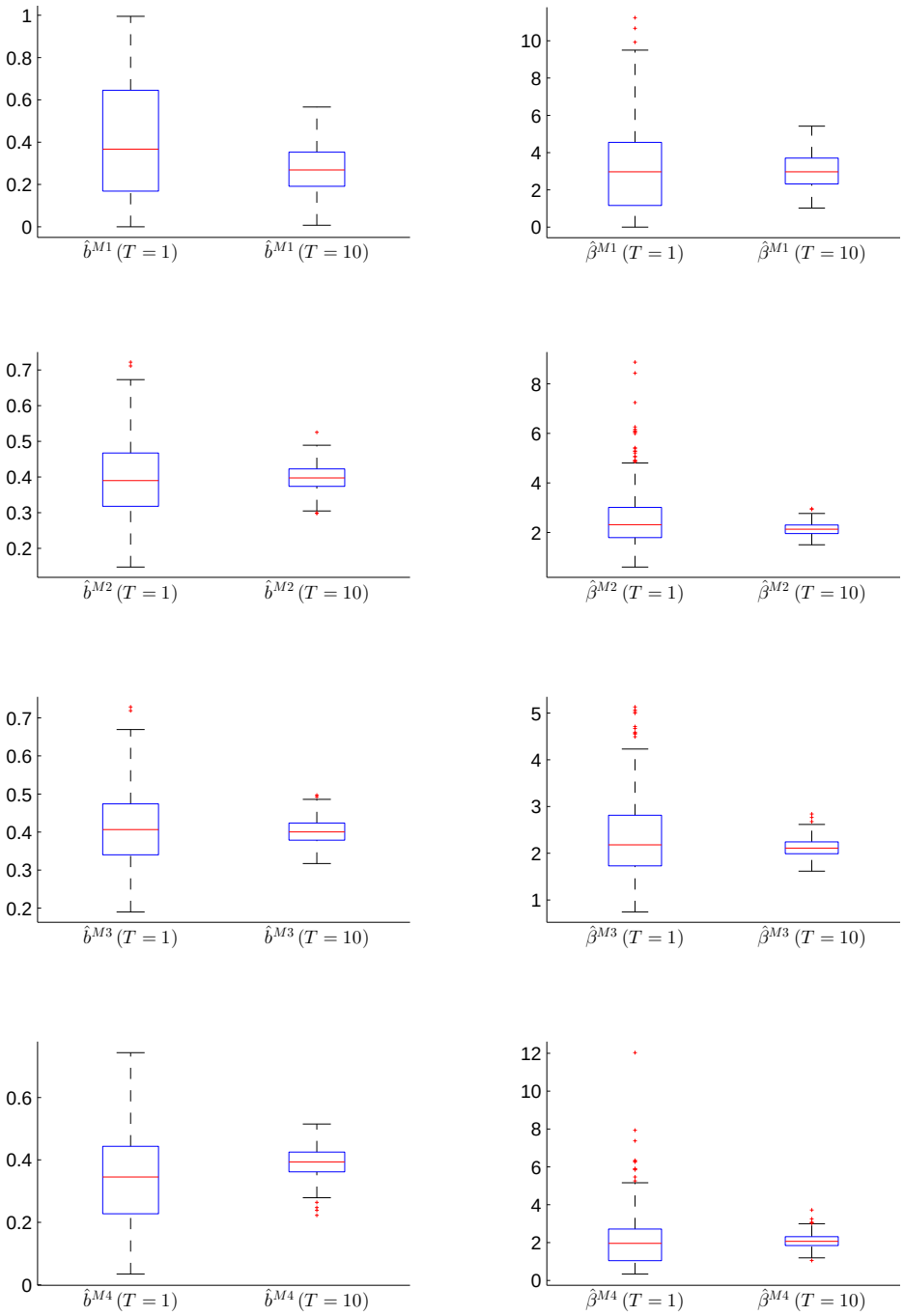


Figure A.8 Boxplots of subordinator parameter estimators for high variance inverse Gaussian subordinator with drift: The figure shows boxplots of the estimators for the subordinator parameters from all estimation methods in the setting with a high variance Inverse Gaussian subordinator with drift for short and long time horizon.

Setting: gamma subordinator (no drift)

		(M1)		(M2)		(M3)		(M4)	
T		1	10	1	10	1	10	1	10
runtime	low	0.544	0.514	37.9	426	6.64	100	94.8	71
	high	0.451	0.425	14.9	148	1.18	9.66	171	169
ofv	low	0.233	0.0195	-568	-5700	276	2770	2.33	0.233
	high	0.0401	0.00433	-353	-3550	-65.8	-674	0.163	0.0133

Table A.5 Runtime and ofv for gamma subordinator (no drift): The table summarizes runtimes and objective function values (ofv) for the estimation Methods (M1)-(M4) in the setting with a gamma subordinator (without drift) with low and high variance and for short and long time horizon. The runtime is expressed in minutes and all quantities are given with 3 significant digits.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 210$	
T		1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	51.385	50.015	77.891	75.229	104.077	100.591	240.595	215.117
	(M2)	49.716	49.898	75.774	75.108	99.258	99.887	216.553	211.438
	(M3)	49.716	49.898	75.774	75.108	99.258	99.887	211.954	210.098
	(M4)	49.057	49.882	74.920	74.945	97.986	99.565	214.655	210.372
$\hat{\text{std}}$	(M1)	10.488	2.993	14.940	4.171	20.356	5.791	80.358	20.679
	(M2)	7.891	2.395	9.677	2.952	11.859	3.774	47.700	13.590
	(M3)	7.891	2.395	9.677	2.952	11.859	3.774	32.878	9.903
	(M4)	10.108	3.125	12.022	3.885	13.993	4.425	38.327	10.525
$\hat{\text{rbias}}$ (%)	(M1)	2.770	0.030	3.855	0.306	4.077	0.591	14.569	2.437
	(M2)	−0.568	−0.204	1.032	0.144	−0.742	−0.113	3.121	0.685
	(M3)	−0.568	−0.204	1.032	0.144	−0.742	−0.113	0.931	0.047
	(M4)	−1.885	−0.236	−0.106	−0.073	−2.014	−0.435	2.216	0.177
$\hat{\text{rmse}}$	(M1)	10.579	2.993	15.217	4.178	20.760	5.821	85.985	21.303
	(M2)	7.896	2.398	9.708	2.954	11.883	3.776	48.148	13.666
	(M3)	7.896	2.398	9.708	2.954	11.883	3.776	32.936	9.904
	(M4)	10.152	3.127	12.022	3.885	14.137	4.446	38.609	10.531

Table A.6 Estimation results for low variance gamma subordinator (no drift): The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the low variance gamma subordinator without drift for short and long time horizon.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 21$	
T		1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	50.225	49.999	75.482	74.936	100.451	100.226	24.466	21.501
	(M2)	49.882	49.977	74.824	74.905	100.214	100.108	21.594	21.067
	(M3)	49.882	49.977	74.824	74.905	100.214	100.108	21.522	21.042
	(M4)	52.040	50.142	76.896	75.164	103.201	100.671	21.249	20.959
$\hat{\text{std}}$	(M1)	13.714	4.150	19.488	6.098	25.027	7.943	6.396	2.293
	(M2)	12.957	3.966	18.679	5.703	24.378	7.599	3.833	1.208
	(M3)	12.957	3.966	18.679	5.703	24.378	7.599	3.563	1.137
	(M4)	21.826	6.465	30.042	8.885	34.736	10.907	4.822	1.450
$\hat{\text{rbias}}$ (%)	(M1)	0.449	−0.002	0.643	−0.085	0.451	0.226	16.504	2.384
	(M2)	−0.236	−0.046	−0.235	−0.127	0.214	0.108	2.827	0.317
	(M3)	−0.236	−0.046	−0.235	−0.127	0.214	0.108	2.486	0.198
	(M4)	4.080	0.285	2.528	0.219	3.201	0.671	1.184	−0.196
$\hat{\text{rmse}}$	(M1)	13.716	4.150	19.494	6.099	25.031	7.946	7.274	2.347
	(M2)	12.958	3.966	18.679	5.704	24.379	7.600	3.879	1.210
	(M3)	12.958	3.966	18.679	5.704	24.379	7.600	3.601	1.138
	(M4)	21.921	6.466	30.102	8.887	34.884	10.928	4.829	1.450

Table A.7 Estimation results for high variance gamma subordinator (no drift): The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the high variance gamma subordinator without drift for short and long time horizon.

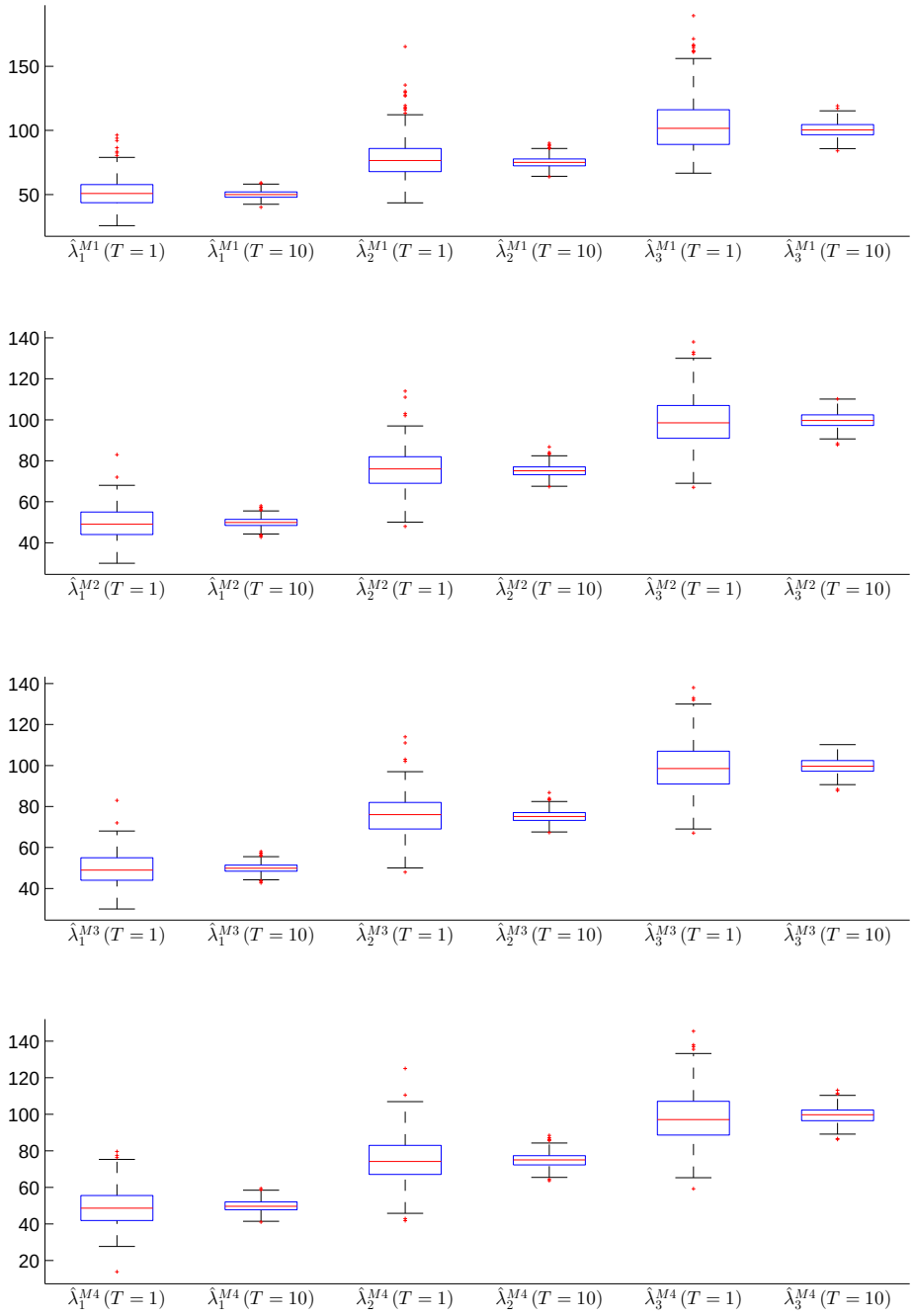


Figure A.9 Boxplots of intensity estimators for low variance gamma subordinator (no drift): The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a low variance gamma subordinator (no drift) for short and long time horizon.

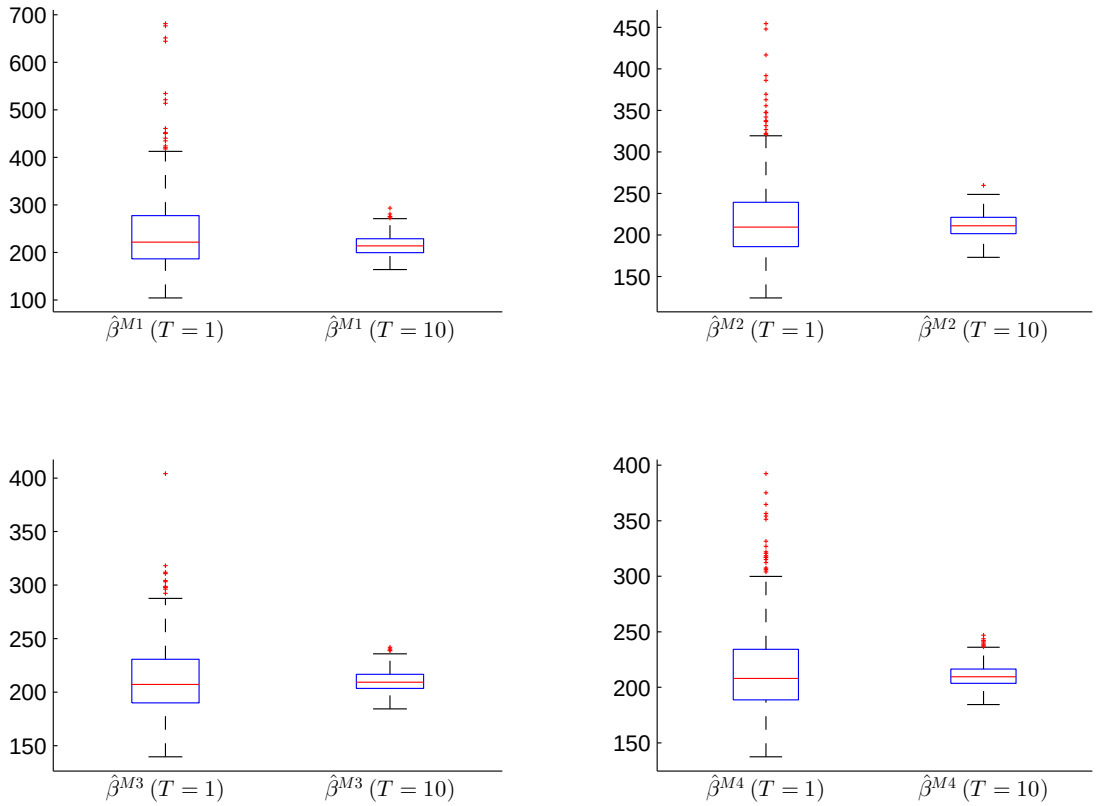


Figure A.10 Boxplots of subordinator parameter estimators for low variance gamma subordinator (no drift): The figure shows boxplots of the subordinator parameter estimators from all estimation methods in the setting with a low variance gamma subordinator (no drift) for short and long time horizon.

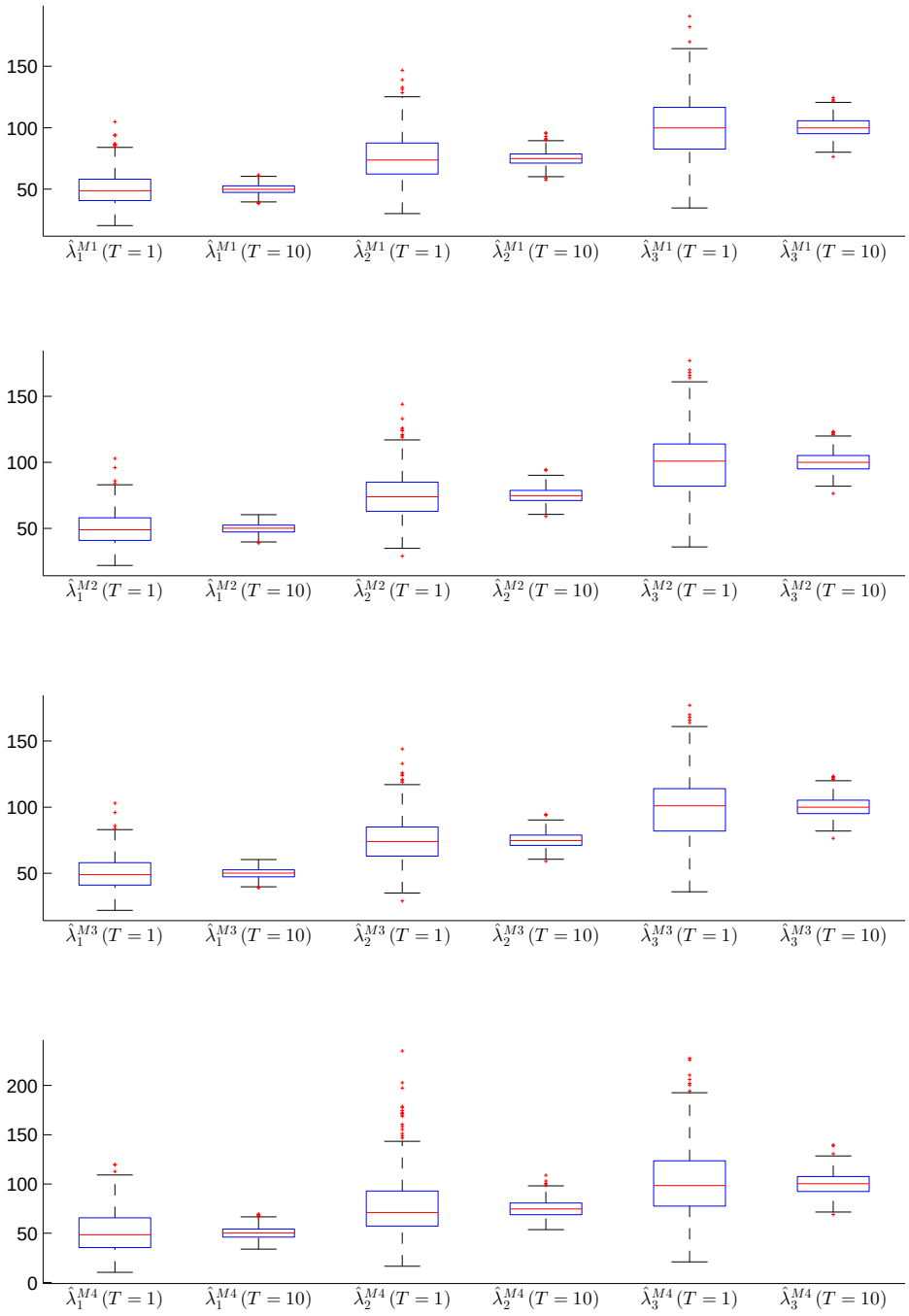


Figure A.11 Boxplots of intensity estimators for high variance gamma subordinator (no drift): The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a high variance gamma subordinator (no drift) for short and long time horizon.

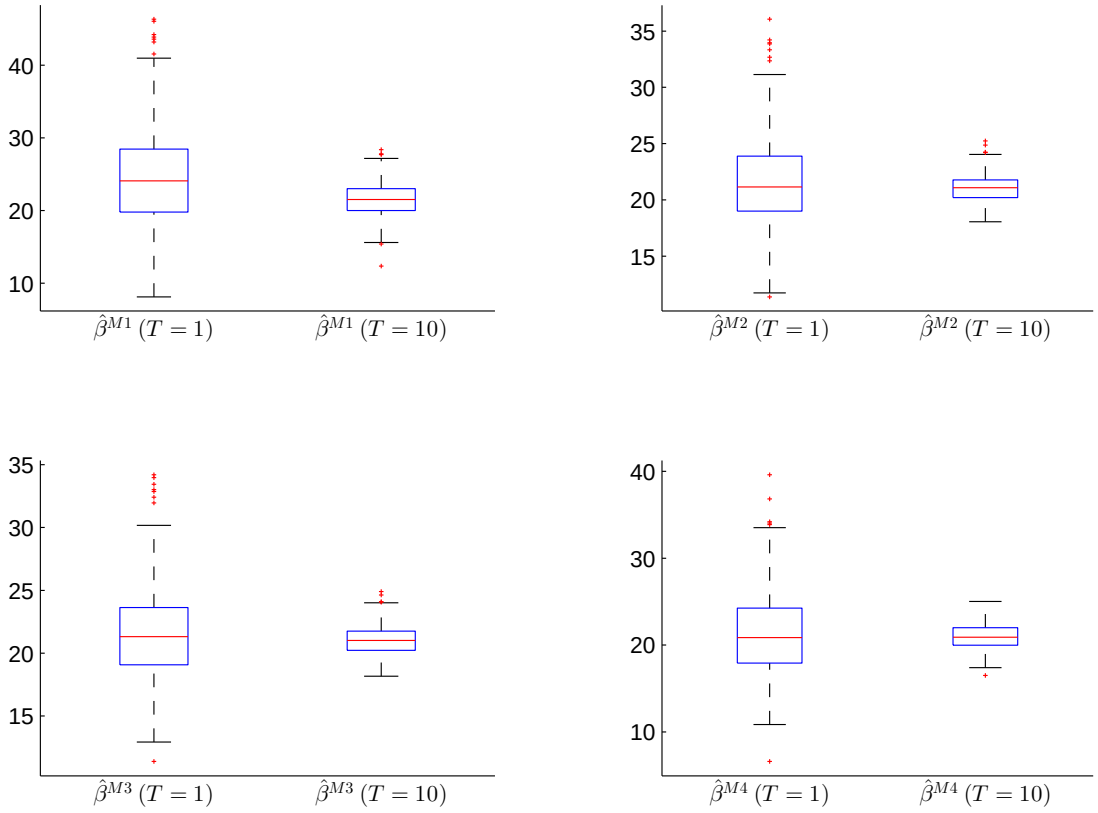


Figure A.12 Boxplots of subordinator parameter estimators for high variance gamma subordinator (no drift): The figure shows boxplots of the subordinator parameter estimators from all estimation methods in the setting with a high variance gamma subordinator (no drift) for short and long time horizon.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 210/21$	
T		1	10	1	10	1	10	1	10
1-2	low	6.7	1.7	11	2.9	16	4.1	67	15
	high	3.8	1.3	5.1	1.8	6.5	2.5	5.8	1.9
1-3	low	6.7	1.7	11	2.9	16	4.1	78	19
	high	3.8	1.3	5.1	1.8	6.5	2.5	6	2
1-4	low	11	3.1	16	4.4	20	5.5	83	20
	high	19	5.7	25	7.6	28	8.8	8	2.5
2-3	low	0.00088	0.002	0.0013	0.0026	0.0014	0.0039	33	9
	high	5.4e-05	0.00061	3.7e-05	0.00057	2.9e-05	0.00051	1.3	0.39
2-4	low	6.5	1.9	7.4	2.4	7.9	2.4	39	10
	high	17	5.1	23	6.9	25	7.9	3.5	1
3-4	low	6.5	1.9	7.4	2.4	7.9	2.4	18	3.9
	high	17	5.1	23	6.9	25	7.9	3	0.87

Table A.8 Differences between estimators from all methods for gamma subordinator (no drift): The table summarizes the root-mean-squared Differences between the estimators from two Methods (Mi) and (Mj) (noted $i - j$) in case of a gamma subordinator without drift in the setting with low and high variance as well as for short and long time horizon (2 significant digits).

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 210/21$	
T		1	10	1	10	1	10	1	10
1-2	low	0.41	0.35	0.34	0.36	0.31	0.31	0.35	0.34
	high	0.45	0.46	0.42	0.44	0.46	0.43	0.28	0.24
1-3	low	0.41	0.35	0.34	0.36	0.31	0.31	0.31	0.25
	high	0.45	0.46	0.42	0.44	0.46	0.43	0.29	0.24
1-4	low	0.51	0.50	0.47	0.49	0.39	0.42	0.32	0.29
	high	0.68	0.64	0.66	0.67	0.62	0.63	0.39	0.35
2-3	low	0.51	0.51	0.51	0.48	0.47	0.50	0.39	0.36
	high	0.45	0.47	0.49	0.49	0.48	0.48	0.45	0.45
2-4	low	0.62	0.65	0.61	0.63	0.60	0.58	0.43	0.37
	high	0.72	0.68	0.68	0.68	0.66	0.65	0.61	0.61
3-4	low	0.62	0.65	0.61	0.63	0.60	0.58	0.57	0.55
	high	0.72	0.68	0.68	0.68	0.66	0.65	0.65	0.66

Table A.9 Pitman closeness criterion for estimators from different methods for gamma subordinator (no drift): The table summarizes the empirical probabilities that the absolute deviations of estimators from Method (Mi) are less then for another Method (Mj) (noted $i-j$) in the setting of a gamma subordinator without drift, for low and high variance as well as for short and long time horizon (2 significant digits).

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 210/21$	
T		1	10	1	10	1	10	1	10
(M1)	low	0.01	0.74	0.00	0.21	0.00	0.94	0.00	0.33
	high	0.04	0.01	0.00	0.26	0.00	0.24	0.02	0.04
(M2)	low	0.00	0.79	0.33	0.88	0.14	0.67	0.00	0.68
	high	0.36	0.12	0.00	0.86	0.00	0.71	0.00	0.10
(M3)	low	0.00	0.79	0.35	0.86	0.14	0.43	0.00	0.38
	high	0.36	0.12	0.00	0.86	0.00	0.71	0.02	0.03
(M4)	low	0.01	0.47	0.58	0.87	0.52	0.06	0.00	0.01
	high	0.00	0.35	0.00	0.00	0.00	0.11	0.31	0.35

Table A.10 Chi-squared test for gamma subordinator (no drift): The table presents the p-values of a chi-squared goodness-of-fit test for normality of the estimators in case of the gamma subordinator without drift in the setting with low and high variance as well as for short and long time horizon.

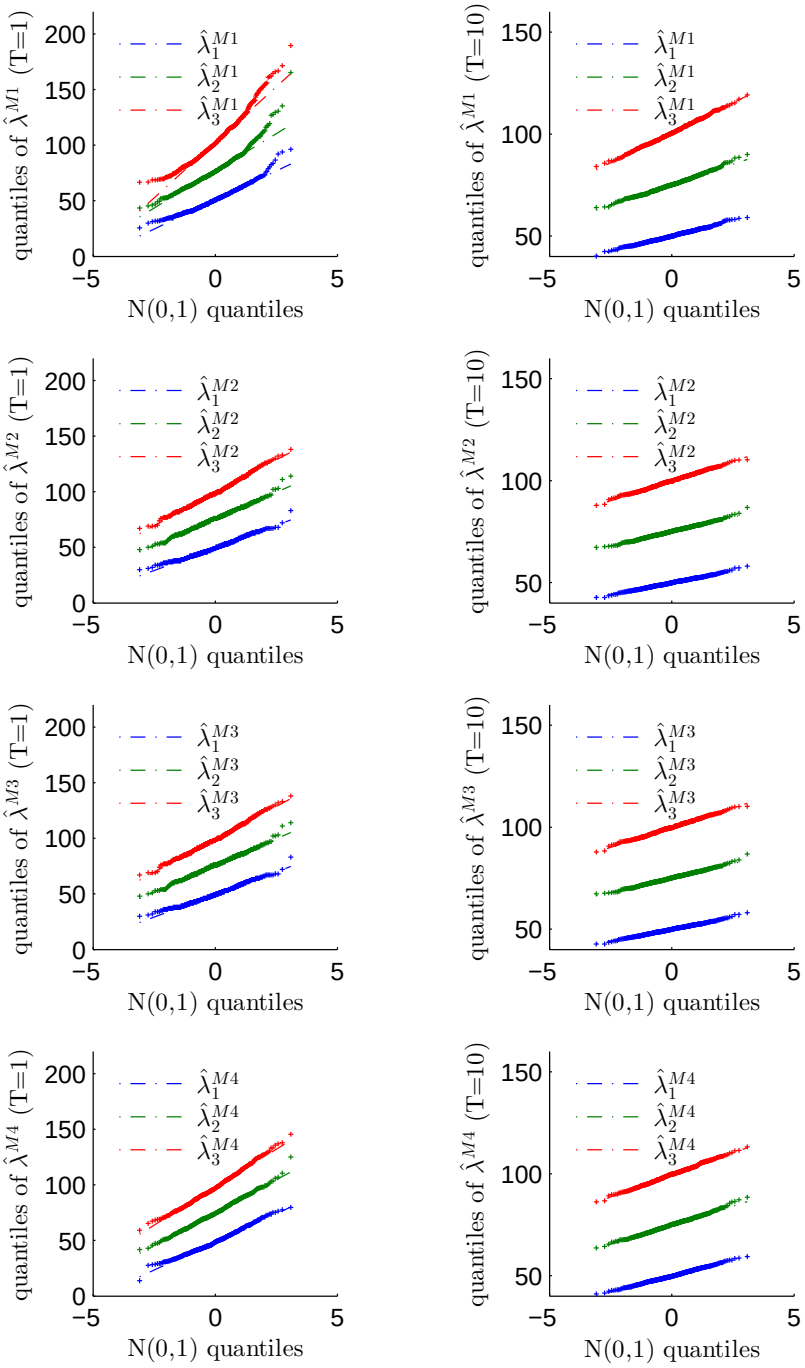


Figure A.13 qq-plots of intensity estimators for low variance gamma subordinator (no drift): The figure illustrates qq-plots of the intensity estimators in case of the low variance gamma subordinator without drift for short and long time horizon.

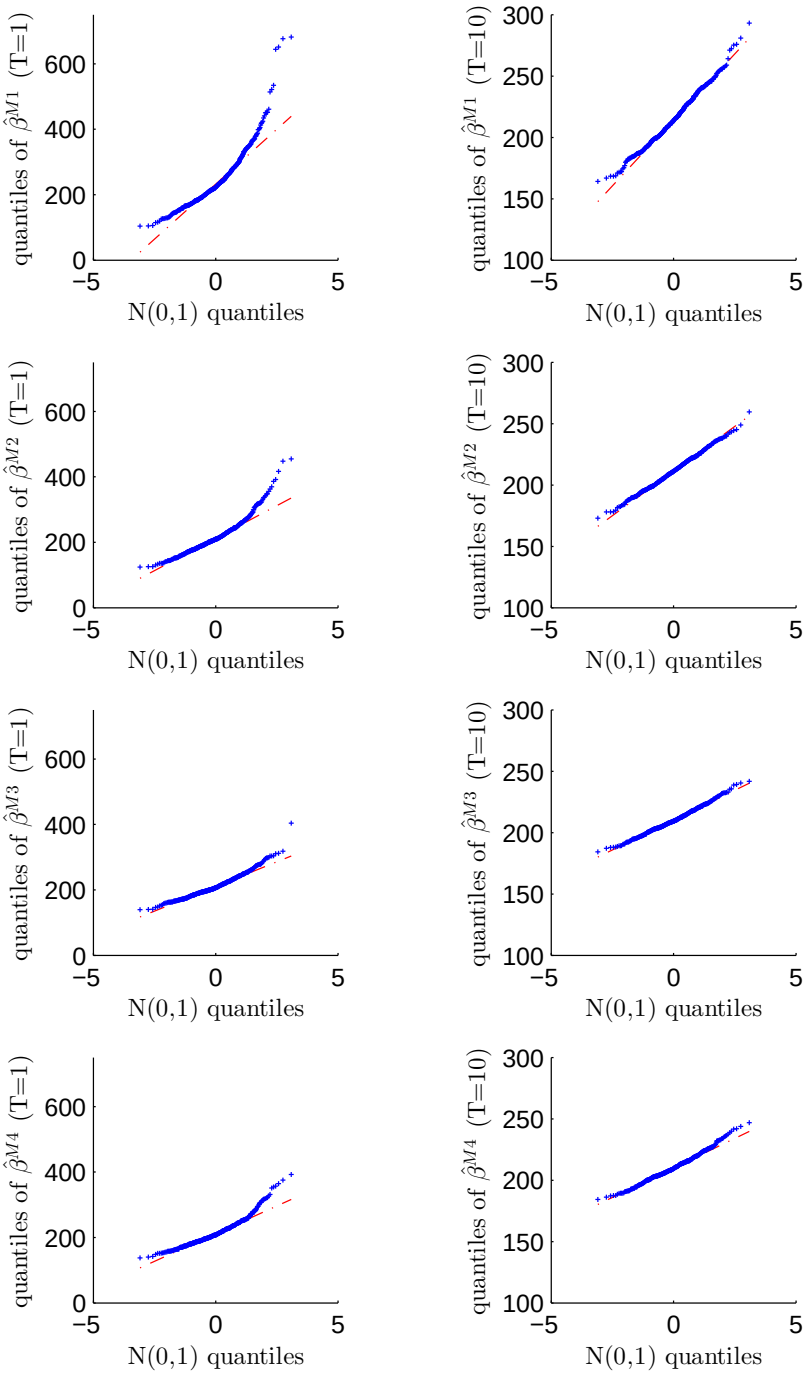


Figure A.14 qq-plots of subordinator estimators for low variance gamma subordinator (no drift): The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the low variance gamma subordinator without drift for short and long time horizon.

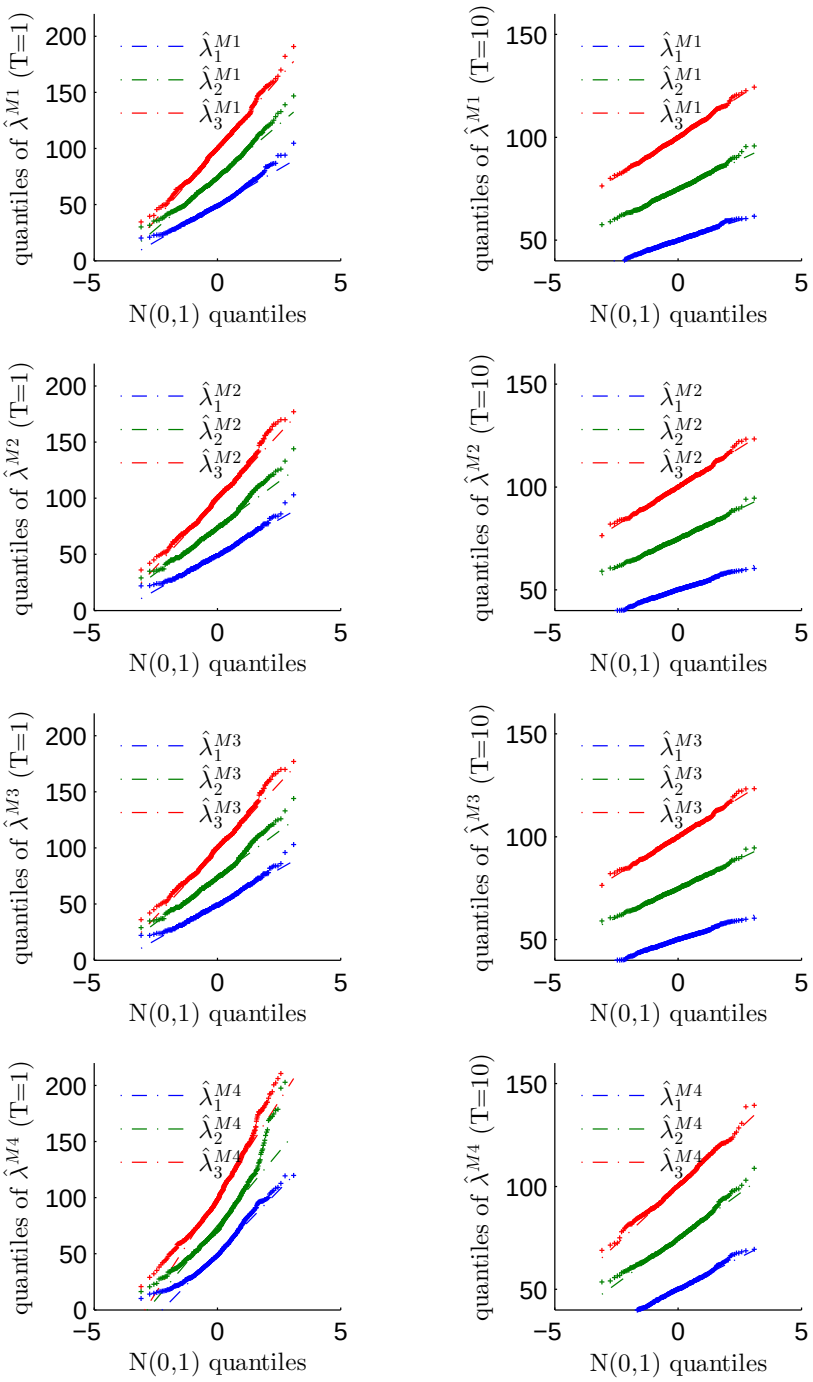


Figure A.15 qq-plots of intensity estimators for high variance gamma subordinator (no drift): The figure illustrates qq-plots of the intensity estimators in case of the high variance gamma subordinator without drift for short and long time horizon.

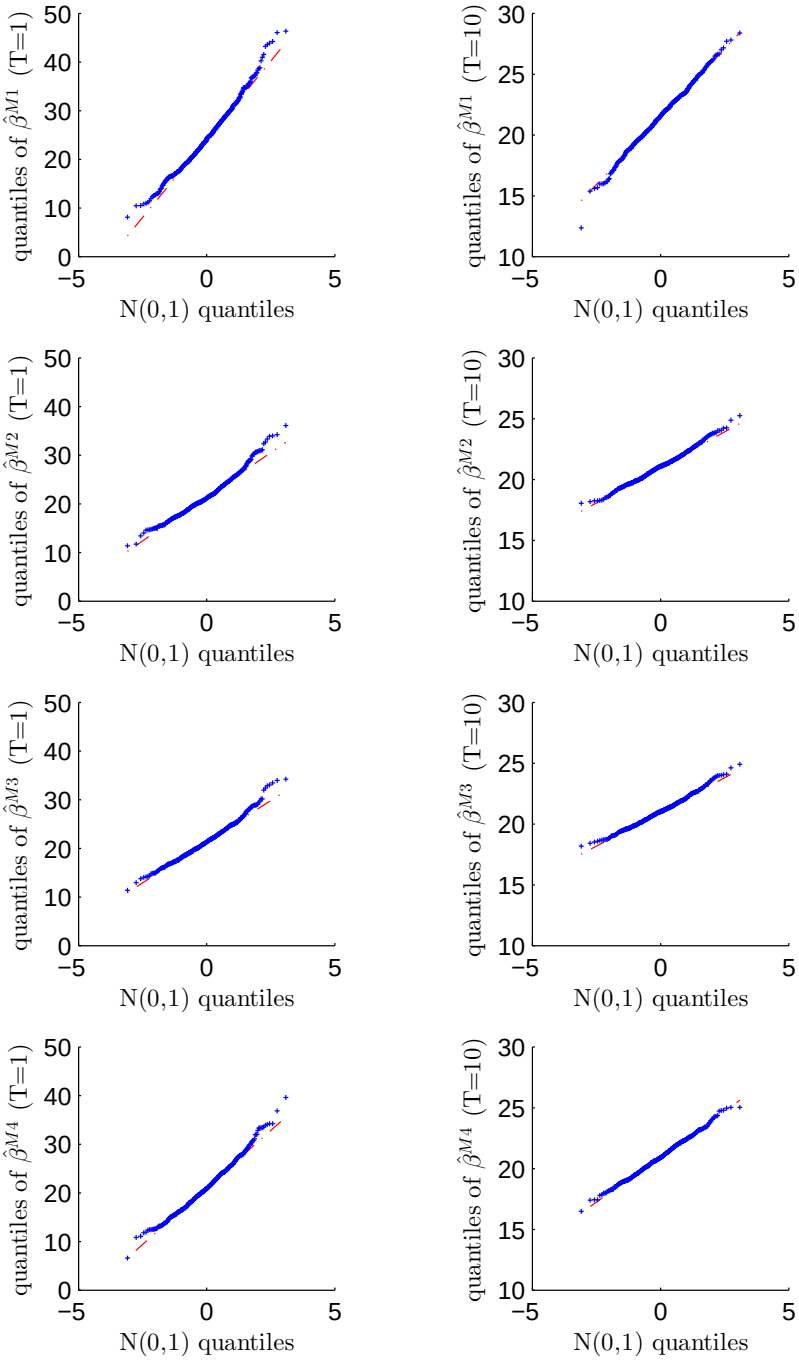


Figure A.16 qq-plots of subordinator estimators for high variance gamma subordinator (no drift): The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the high variance gamma subordinator without drift for short and long time horizon.

Setting: gamma subordinator with drift

		(M1)		(M2)		(M3)		(M4)	
T		1	10	1	10	1	10	1	10
runtime	low	0.559	0.532	76.3	566	22.2	183	156	150
	high	0.56	0.527	78.2	536	14.7	114	221	180
ofv	low	0.305	0.025	-574	-5750	375	3730	1.07	0.106
	high	0.077	0.00713	-454	-4570	188	1860	0.124	0.0102

Table A.11 Runtime and ofv for gamma subordinator with drift: The table summarizes runtimes and objective function values (ofv) for the estimation Methods (M1)-(M4) in the setting with a gamma subordinator with drift, low as well as high variance and for short and long time horizon. The runtime is expressed in minutes and all quantities are given with 3 significant digits.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 78$	
T		1	10	1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	52.256	50.255	79.220	75.618	105.430	100.027	0.306	0.373	133.164	89.214
	(M2)	50.388	50.052	75.138	75.060	99.942	99.862	0.366	0.396	103.058	80.456
	(M3)	50.388	50.051	75.138	75.060	99.942	99.862	0.383	0.398	91.056	79.354
	(M4)	50.259	49.986	75.027	75.002	100.224	99.837	0.390	0.398	89.634	79.647
$\hat{\text{std}}$	(M1)	11.620	3.162	16.881	4.711	21.798	5.971	0.166	0.088	75.523	28.874
	(M2)	7.889	2.345	9.889	3.131	11.203	3.599	0.155	0.048	74.015	16.814
	(M3)	7.889	2.345	9.888	3.131	11.203	3.599	0.092	0.027	42.962	10.643
	(M4)	9.446	2.871	11.413	3.549	13.358	4.168	0.107	0.031	52.949	12.835
$\hat{\text{rbias}}$ (%)	(M1)	4.512	0.510	5.626	0.824	5.430	0.027	−23.406	−6.772	70.723	14.377
	(M2)	0.776	0.103	0.184	0.081	−0.058	−0.138	−8.591	−1.037	32.125	3.149
	(M3)	0.776	0.103	0.184	0.080	−0.058	−0.138	−4.219	−0.588	16.738	1.735
	(M4)	0.518	−0.027	0.035	0.003	0.224	−0.163	−2.393	−0.550	14.916	2.111
$\hat{\text{rmse}}$	(M1)	11.837	3.173	17.400	4.751	22.464	5.971	0.190	0.092	93.524	30.975
	(M2)	7.898	2.346	9.890	3.132	11.203	3.601	0.158	0.048	78.141	16.993
	(M3)	7.898	2.345	9.889	3.132	11.203	3.601	0.093	0.027	44.902	10.728
	(M4)	9.450	2.871	11.413	3.549	13.360	4.172	0.108	0.031	54.212	12.940

Table A.12 Estimation results for low variance gamma subordinator with drift: The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the low variance gamma subordinator with drift for short and long time horizon.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 7.5$	
T		1	10	1	10	1	10	1	10	1	10
$\hat{\mathbb{E}}$	(M1)	51.542	50.097	76.760	75.338	100.413	100.185	0.260	0.354	17.801	9.506
	(M2)	50.354	50.042	75.298	75.073	99.392	99.966	0.412	0.402	8.525	7.613
	(M3)	50.354	50.042	75.298	75.073	99.392	99.966	0.415	0.402	8.197	7.600
	(M4)	59.824	50.556	90.702	75.825	121.597	100.787	0.377	0.399	7.988	7.628
$\hat{\text{std}}$	(M1)	13.753	4.298	19.765	6.538	26.766	8.530	0.161	0.093	11.523	3.461
	(M2)	12.717	3.973	18.125	5.896	24.240	7.690	0.096	0.032	3.940	0.985
	(M3)	12.717	3.973	18.125	5.897	24.240	7.690	0.092	0.030	2.945	0.805
	(M4)	29.743	5.802	43.808	8.276	60.389	10.716	0.122	0.038	4.467	1.236
$\hat{\text{rbias}}$ (%)	(M1)	3.084	0.194	2.346	0.450	0.413	0.185	−34.880	−11.620	137.344	26.747
	(M2)	0.708	0.084	0.397	0.097	−0.608	−0.034	2.969	0.409	13.669	1.509
	(M3)	0.708	0.084	0.397	0.098	−0.608	−0.034	3.763	0.418	9.292	1.336
	(M4)	19.648	1.112	20.936	1.100	21.597	0.787	−5.799	−0.145	6.513	1.712
$\hat{\text{rmse}}$	(M1)	13.839	4.299	19.843	6.547	26.770	8.532	0.213	0.104	15.456	4.001
	(M2)	12.722	3.973	18.128	5.896	24.248	7.690	0.097	0.032	4.072	0.991
	(M3)	12.722	3.973	18.127	5.897	24.248	7.690	0.093	0.031	3.027	0.811
	(M4)	31.323	5.829	46.537	8.317	64.135	10.744	0.124	0.038	4.494	1.242

Table A.13 Estimation results for high variance gamma subordinator with drift: The table presents the Monte Carlo estimates for mean, standard deviation, relative bias, and root-mean-square error of all parameter estimators in case of the high variance gamma subordinator with drift for short and long time horizon.

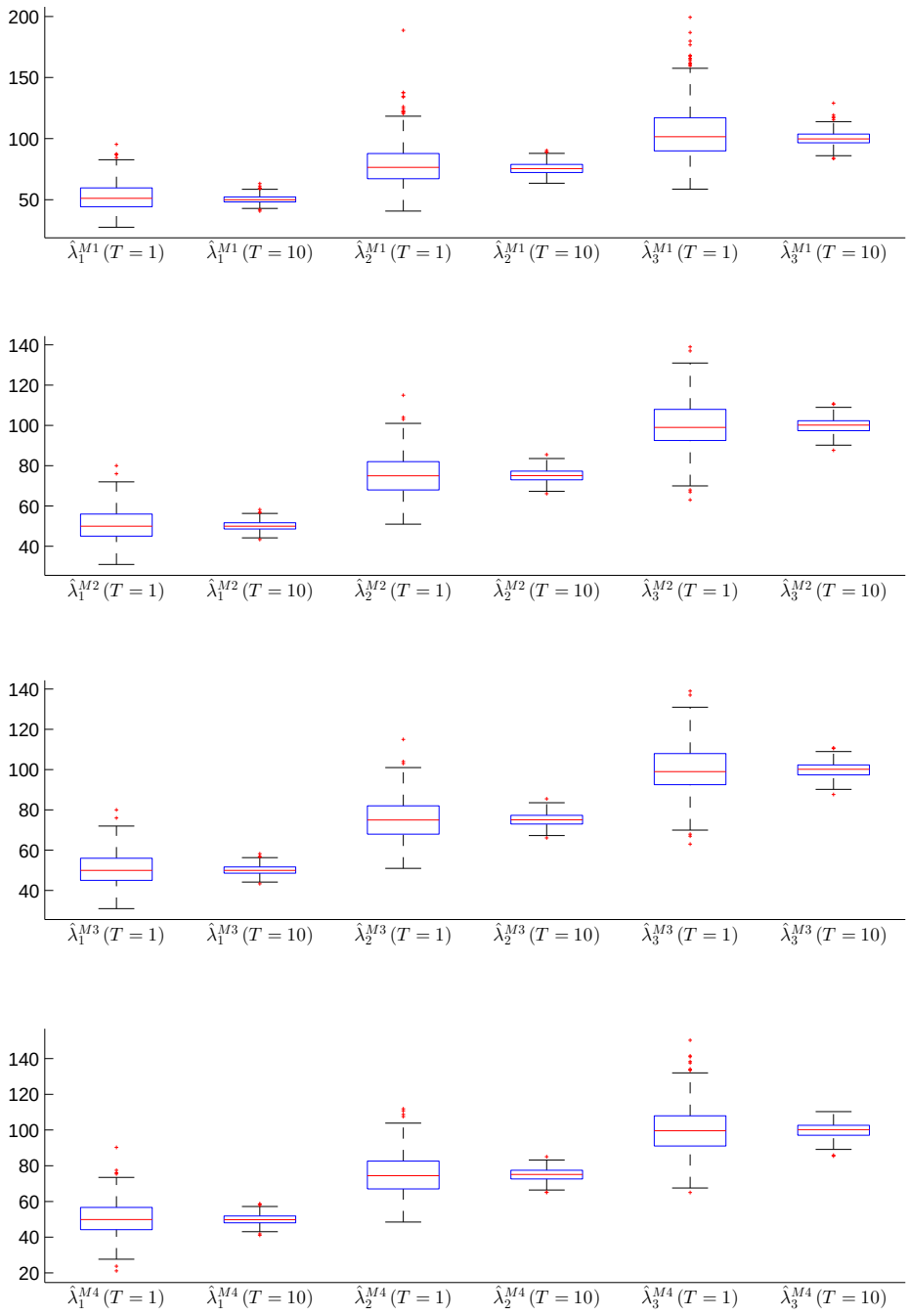


Figure A.17 Boxplots of intensity estimators for low variance gamma subordinator with drift: The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a low variance gamma subordinator with drift for short and long time horizon.

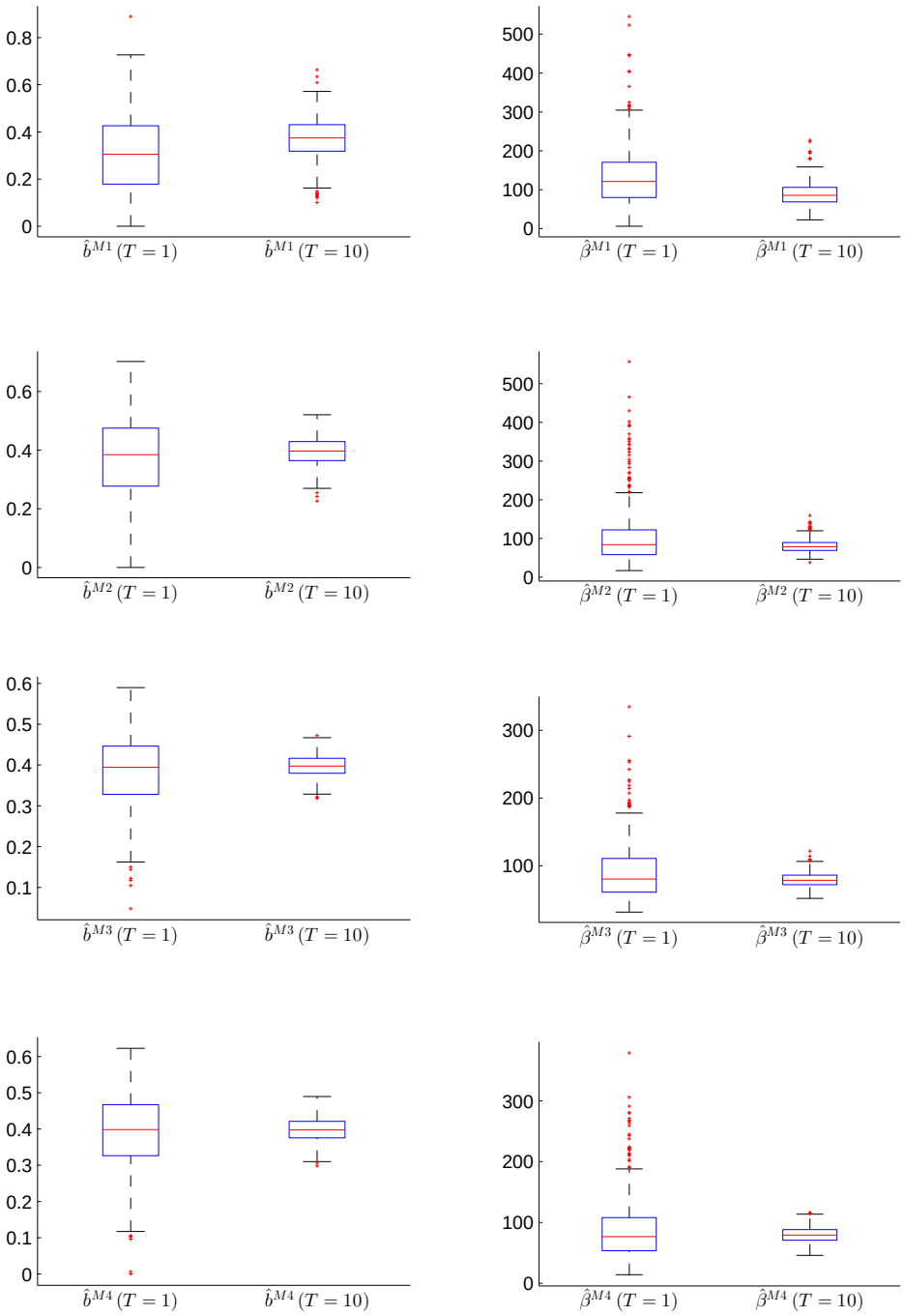


Figure A.18 Boxplots of subordinator parameter estimators for low variance gamma subordinator with drift: The figure shows boxplots of the estimators for the subordinator parameters from all estimation methods in the setting with a low variance gamma subordinator with drift for short and long time horizon.

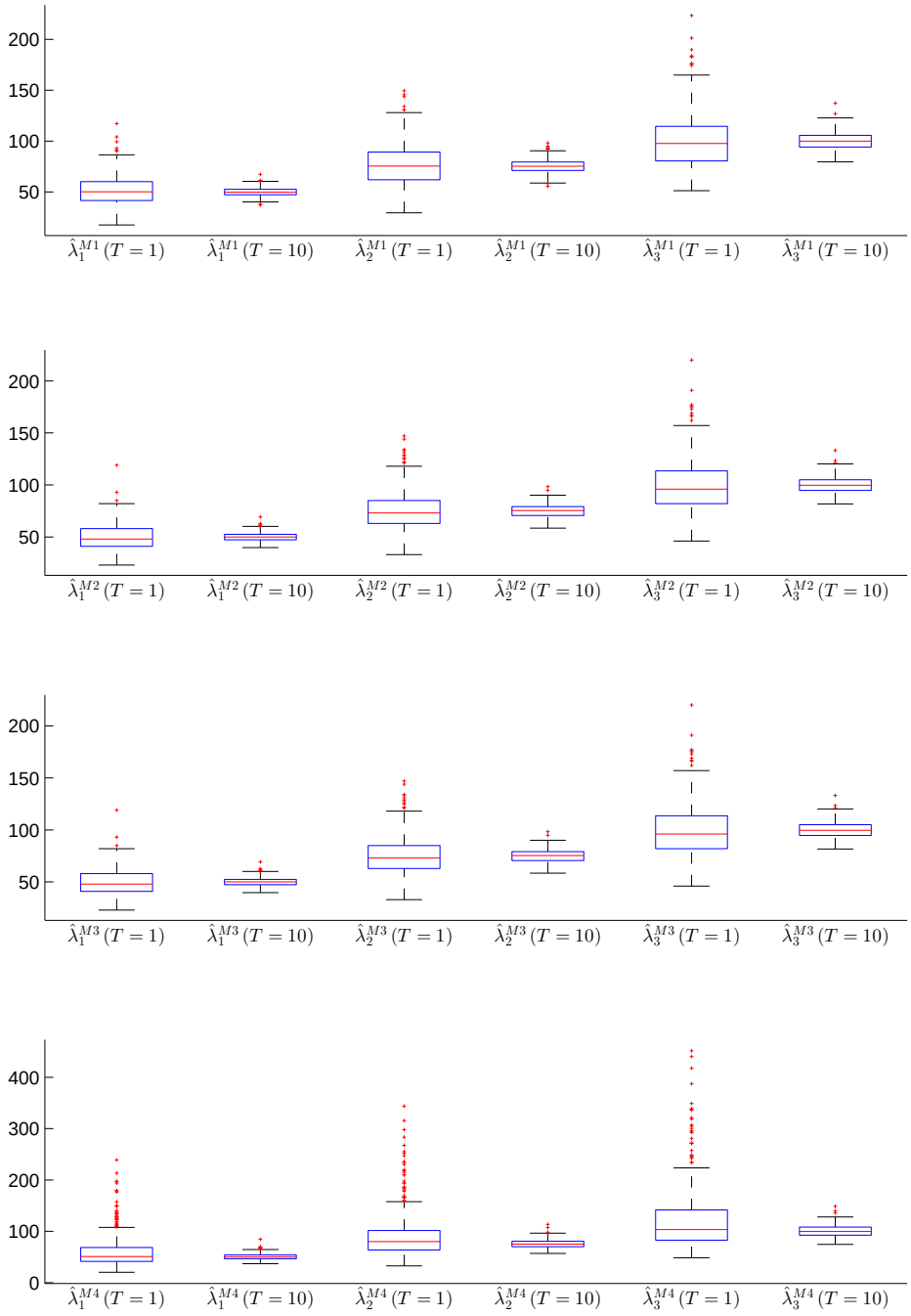


Figure A.19 Boxplots of intensity estimators for high variance gamma subordinator with drift: The figure shows boxplots of the intensity estimators from all estimation methods in the setting with a high variancegamma subordinator with drift for short and long time horizon.

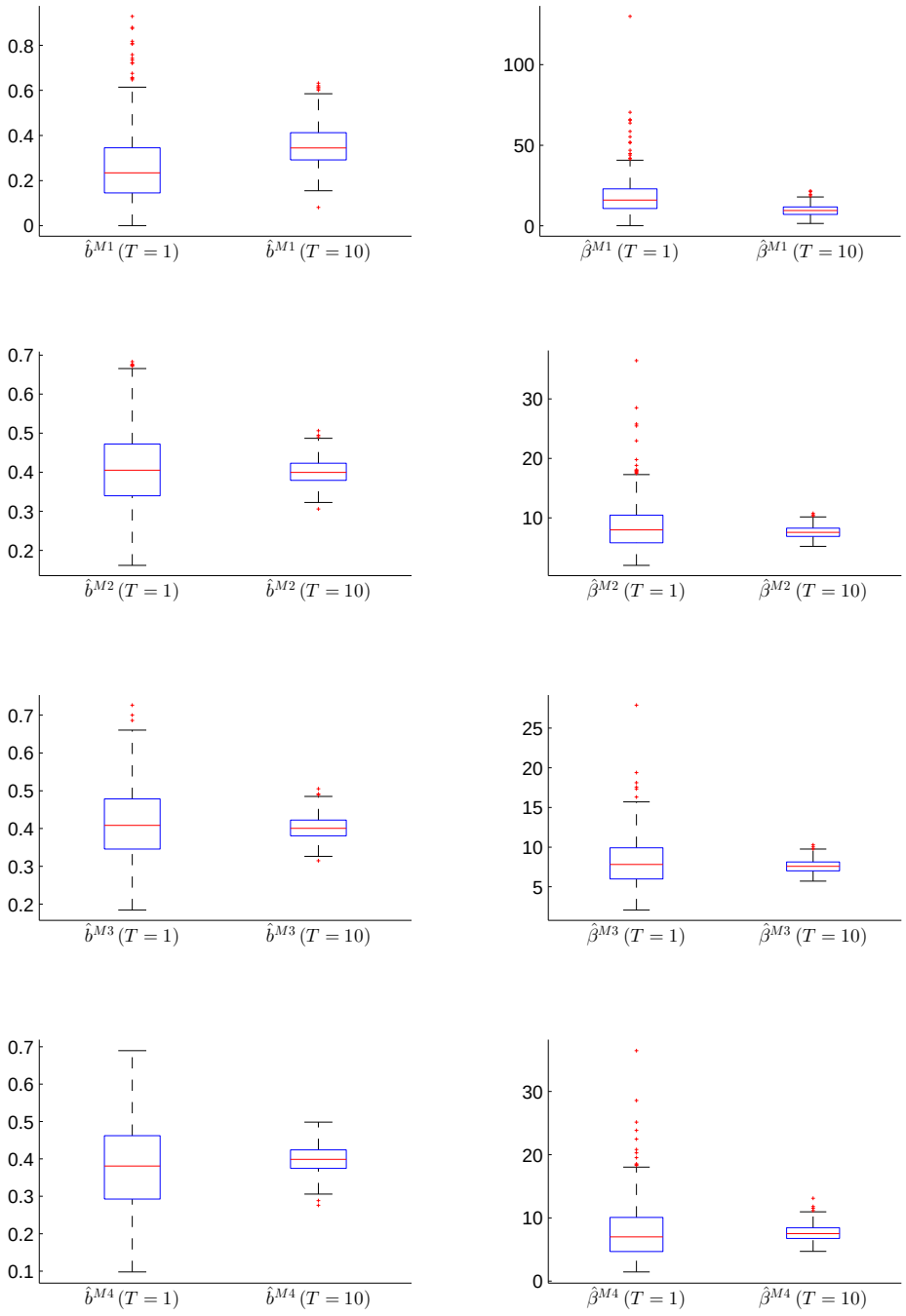


Figure A.20 Boxplots of subordinator parameter estimators for high variance gamma subordinator with drift: The figure shows boxplots of the estimators for the subordinator parameters from all estimation methods in the setting with a high variance gamma subordinator with drift for short and long time horizon.

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 78/7.5$	
T		1	10	1	10	1	10	1	10	1	10
1-2	low	7.9	2.2	14	3.3	20	5.1	0.17	0.077	65	24
	high	5.8	2	8.2	2.8	12	3.4	0.21	0.099	13	3.7
1-3	low	7.9	2.2	14	3.3	20	5.1	0.18	0.088	79	28
	high	5.8	2	8.2	2.8	12	3.4	0.22	0.099	14	3.8
1-4	low	12	3.4	17	4.5	23	6.2	0.2	0.093	92	32
	high	29	4.9	42	7.3	59	8.5	0.2	0.096	15	4.1
2-3	low	0.00082	0.0032	0.0013	0.0044	0.0017	0.0058	0.13	0.041	60	13
	high	0.0021	0.0065	0.0032	0.0097	0.0042	0.013	0.034	0.0093	2.4	0.56
2-4	low	5.4	1.7	5.9	1.9	6.8	2.1	0.16	0.048	73	17
	high	27	3.9	41	5.8	57	7.1	0.092	0.022	5.2	1.4
3-4	low	5.4	1.7	5.9	1.9	6.8	2.1	0.053	0.015	29	7.1
	high	27	3.9	41	5.8	57	7.1	0.096	0.023	3.5	0.94

Table A.14 Differences between estimators from all methods for gamma subordinator with drift: The table summarizes the root-mean-squared Differences between the estimators from two Methods (Mi) and (Mj) (noted $i - j$) in case of a gamma subordinator with drift in the setting with low and high variance as well as for short and long time horizon (2 significant digits).

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$d = 0.4$		$\beta = 78/7.5$	
T		1	10	1	10	1	10	1	10	1	10
1-2	low	0.33	0.38	0.34	0.33	0.33	0.35	0.42	0.30	0.36	0.31
	high	0.43	0.46	0.41	0.43	0.40	0.42	0.23	0.17	0.16	0.13
1-3	low	0.33	0.38	0.34	0.33	0.33	0.35	0.26	0.18	0.25	0.23
	high	0.43	0.46	0.41	0.43	0.40	0.42	0.23	0.17	0.13	0.10
1-4	low	0.43	0.49	0.42	0.41	0.37	0.40	0.31	0.21	0.32	0.27
	high	0.58	0.61	0.60	0.58	0.65	0.60	0.28	0.17	0.22	0.17
2-3	low	0.51	0.48	0.47	0.50	0.50	0.50	0.32	0.33	0.39	0.35
	high	0.48	0.55	0.47	0.56	0.46	0.51	0.47	0.43	0.39	0.39
2-4	low	0.57	0.59	0.61	0.59	0.59	0.59	0.39	0.37	0.45	0.46
	high	0.61	0.66	0.67	0.62	0.69	0.64	0.63	0.60	0.57	0.56
3-4	low	0.57	0.59	0.61	0.59	0.59	0.59	0.59	0.61	0.62	0.62
	high	0.61	0.66	0.67	0.62	0.69	0.64	0.62	0.63	0.67	0.66

Table A.15 Pitman closeness criterion for estimators from different methods for gamma subordinator with drift: The table summarizes the empirical probabilities that the absolute deviations of estimators from Method (Mi) are less then for another Method (Mj) (noted $i - j$) in the setting of a gamma subordinator with with drift, for low and high variance as well as for short and long time horizon (2 significant digits).

		$\lambda_1 = 50$		$\lambda_2 = 75$		$\lambda_3 = 100$		$\beta = 78/7.5$	
T		1	10	1	10	1	10	1	10
(M1)	low	0.04	0.01	0.00	0.59	0.00	0.19	0.00	0.00
	high	0.01	0.01	0.00	0.14	0.00	0.89	0.00	0.73
(M2)	low	0.88	0.92	0.00	0.16	0.11	0.31	0.00	0.00
	high	0.00	0.16	0.00	0.01	0.00	0.11	0.00	0.19
(M3)	low	0.88	0.92	0.00	0.12	0.16	0.31	0.00	0.02
	high	0.00	0.17	0.00	0.01	0.00	0.09	0.00	0.32
(M4)	low	0.68	0.30	0.44	0.18	0.10	0.19	0.00	0.12
	high	0.00	0.15	0.00	0.03	0.00	0.00	0.00	0.17

Table A.16 Chi-squared test for gamma subordinator with drift: The table presents the p-values of a chi-squared goodness-of-fit test for normality of the estimators in case of the gamma subordinator with drift in the setting with low and high variance as well as for short and long time horizon.

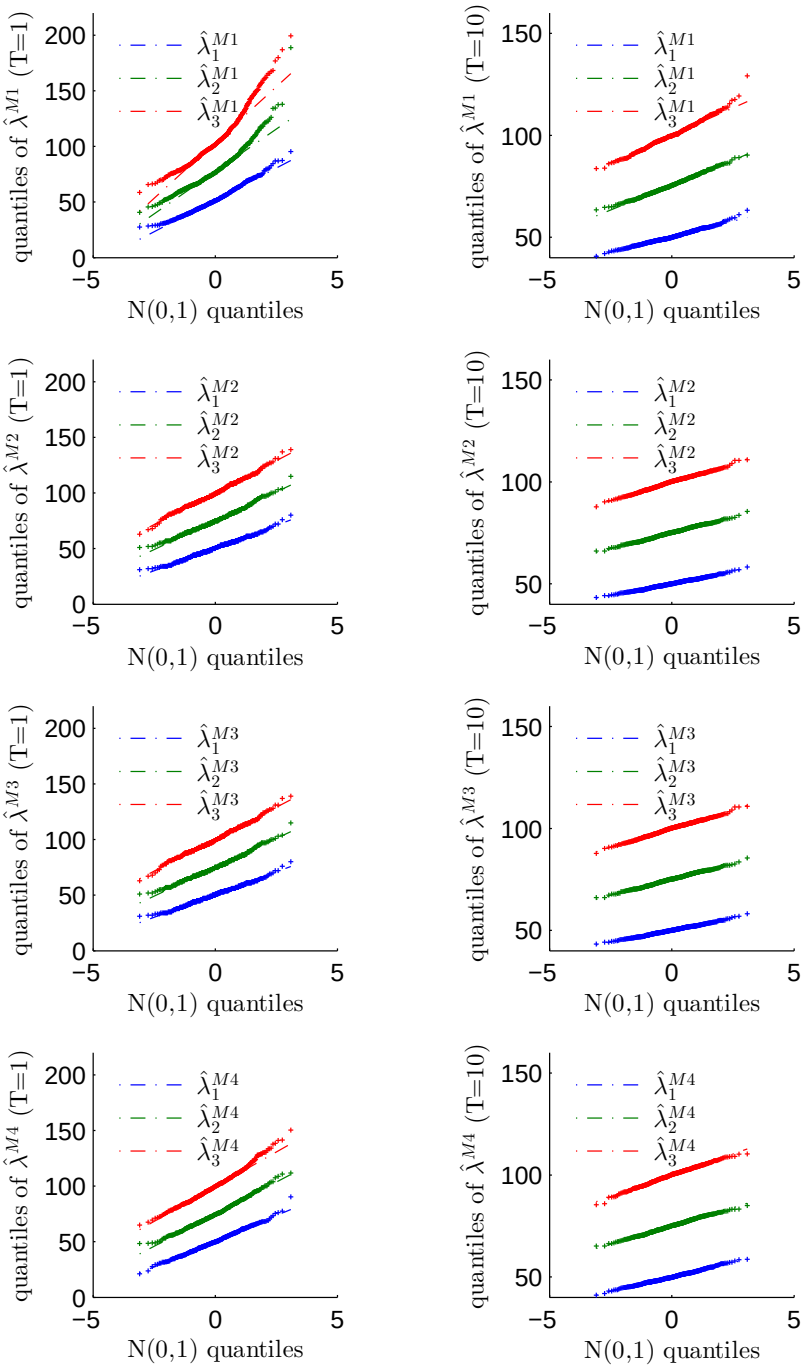


Figure A.21 qq-plots of intensity estimators for low variance gamma subordinator with drift: The figure illustrates qq-plots of the intensity estimators in case of the low variance gamma subordinator with drift for short and long time horizon.

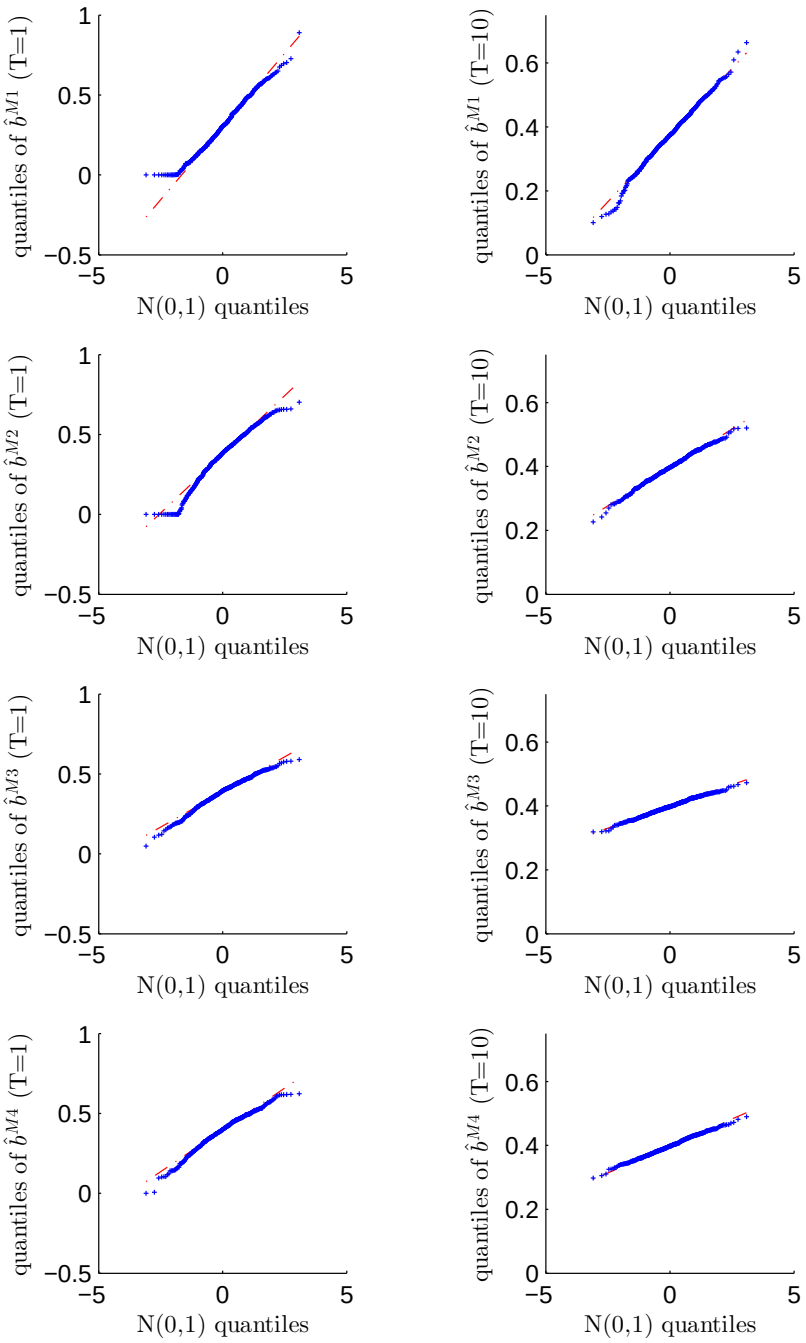


Figure A.22 qq-plots of subordinator drift estimators for low variance gamma subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator drift b in case of the low variance gamma subordinator with drift for short and long time horizon.

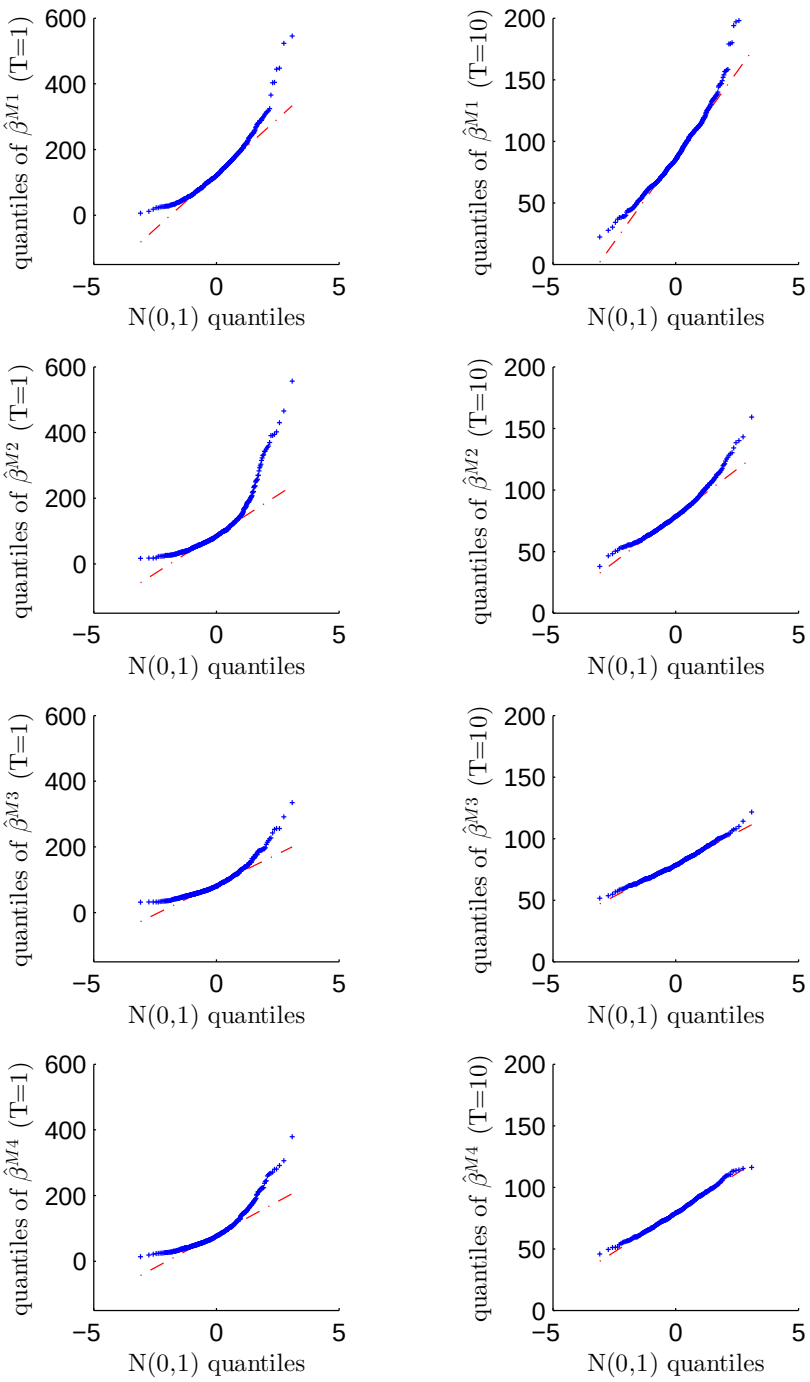


Figure A.23 qq-plots of subordinator β estimators for low variance gamma subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the low variance gamma subordinator with drift for short and long time horizon.

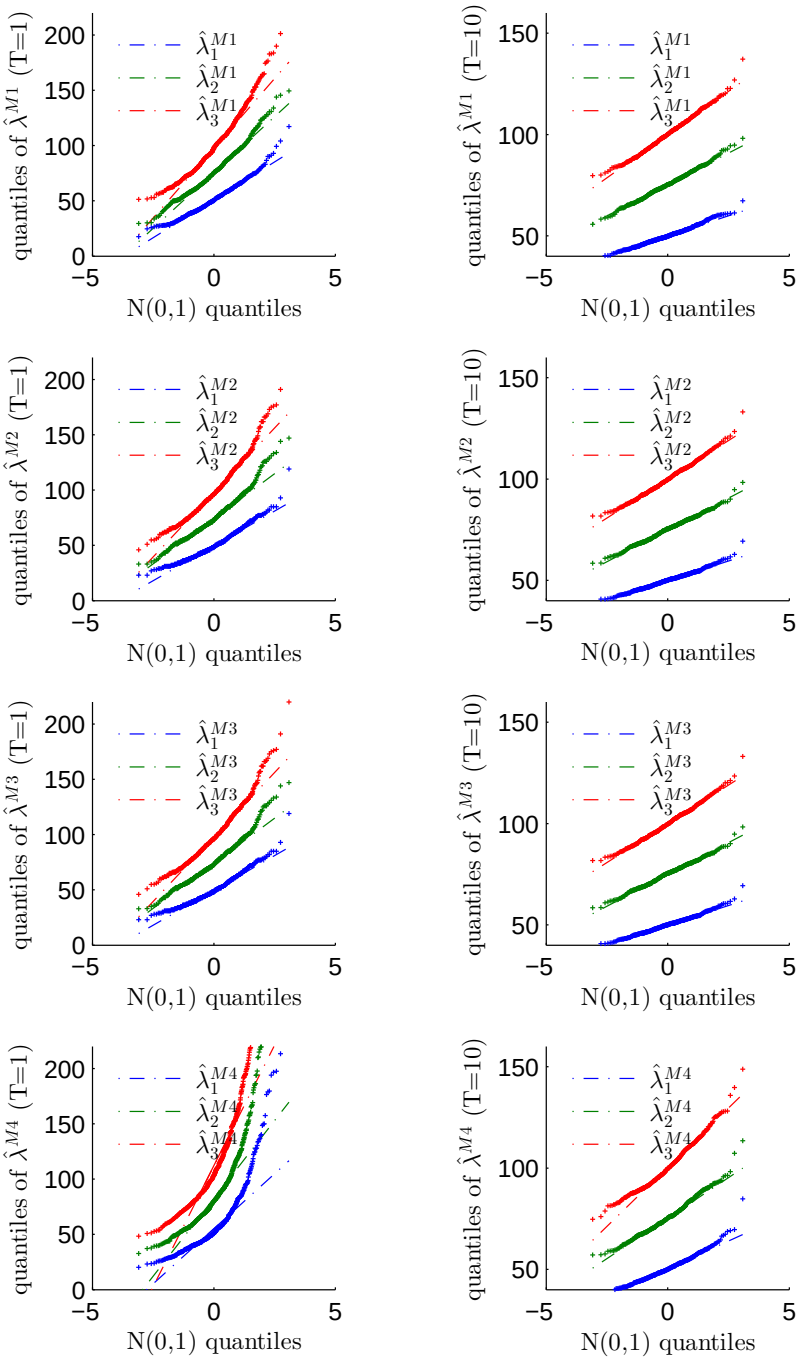


Figure A.24 qq-plots of intensity estimators for high variance gamma subordinator with drift: The figure illustrates qq-plots of the intensity estimators in case of the high variance gamma subordinator with drift for short and long time horizon.

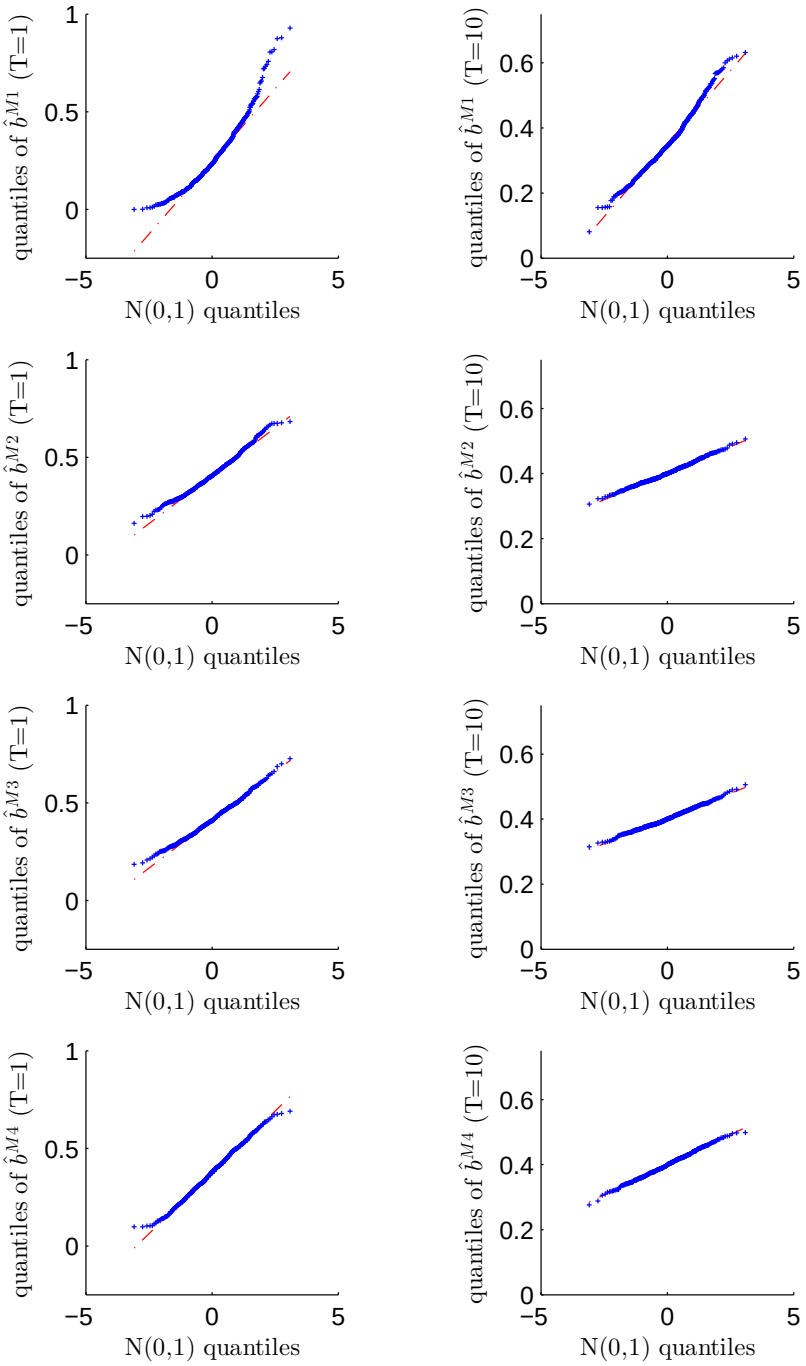


Figure A.25 qq-plots of subordinator drift estimators for high variance gamma subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator drift b in case of the high variance gamma subordinator with drift for short and long time horizon.

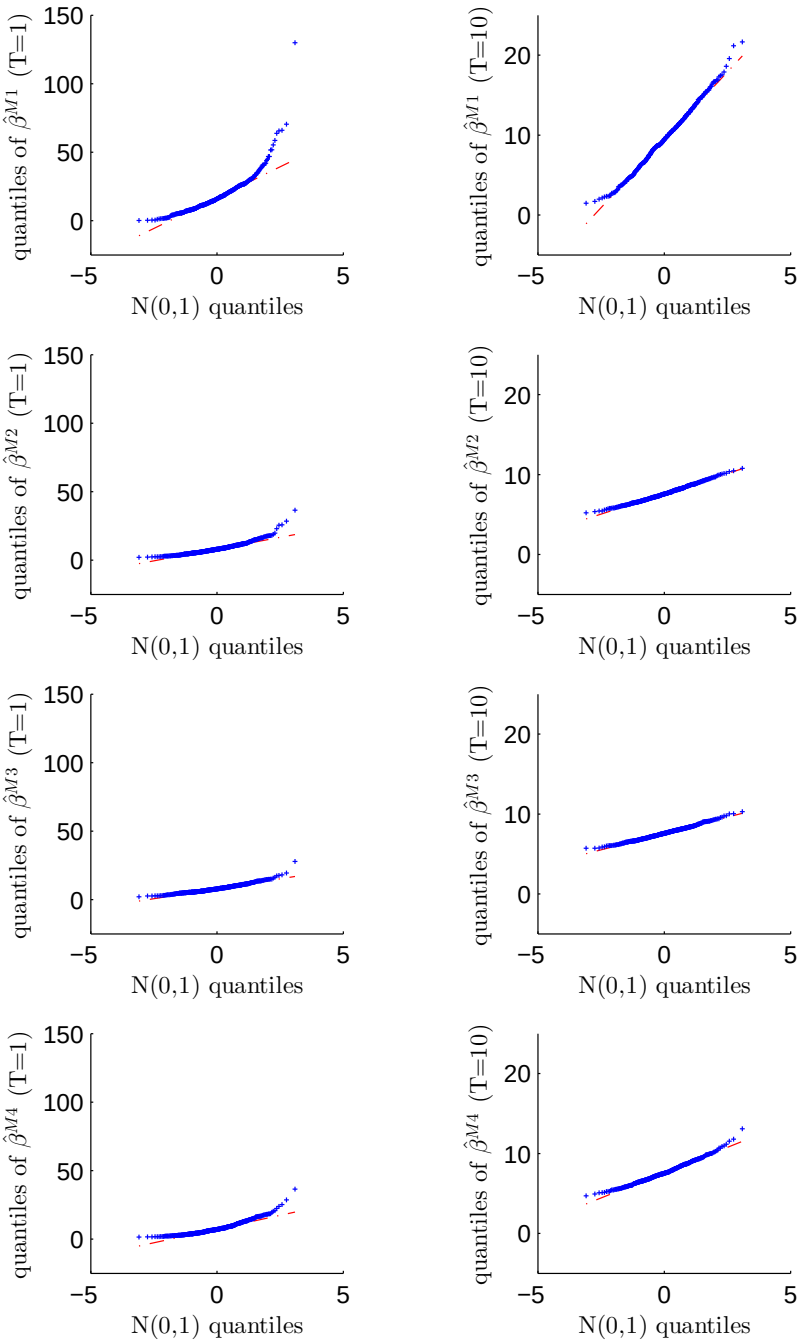


Figure A.26 qq-plots of subordinator β estimators for high variance gamma subordinator with drift: The figure illustrates qq-plots of the estimators for the subordinator parameter β in case of the high variance gamma subordinator with drift for short and long time horizon.

Danish fire insurance data

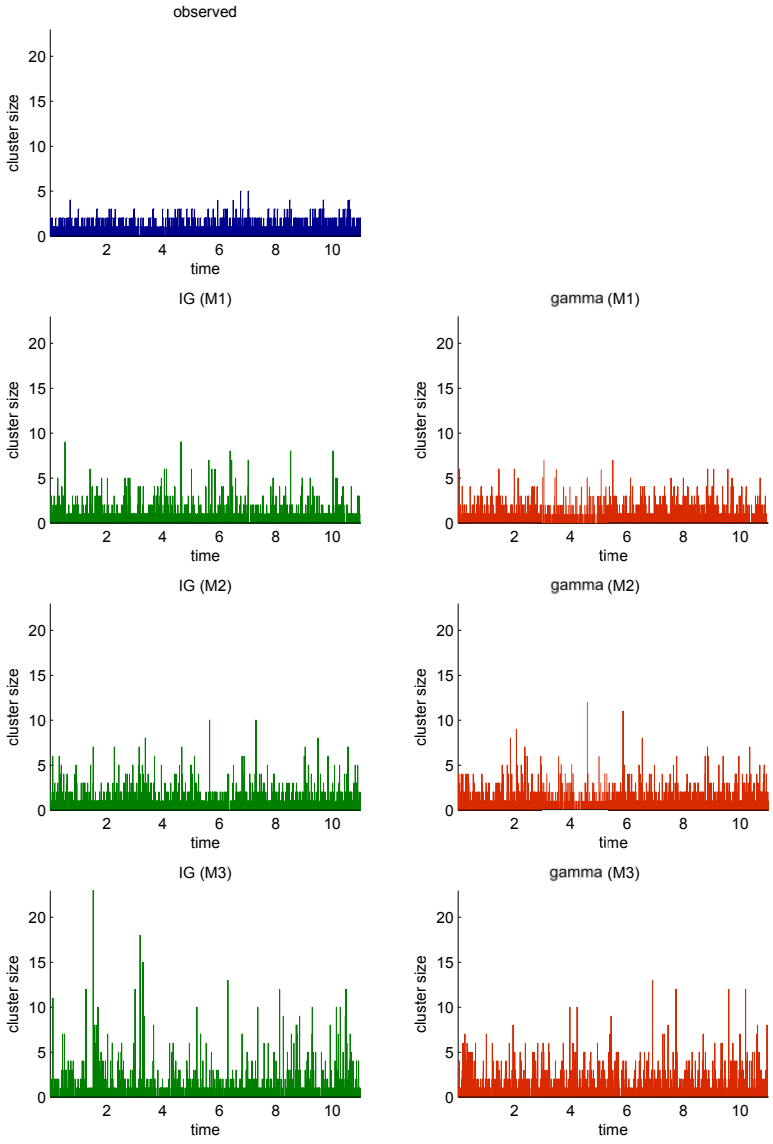


Figure A.27 Cluster sizes of building Danish fire insurance data and fitted models: The top chart shows the observed cluster sizes of the building claim number processes from the Danish fire insurance data. The charts in the bottom three rows present corresponding sampled cluster sizes of the fitted models based on an inverse Gaussian (left) and gamma (right) subordinator using the estimated parameters from Methods (M1)–(M3). (Method (M4) is excluded due to the in parts unreliable estimation results.)

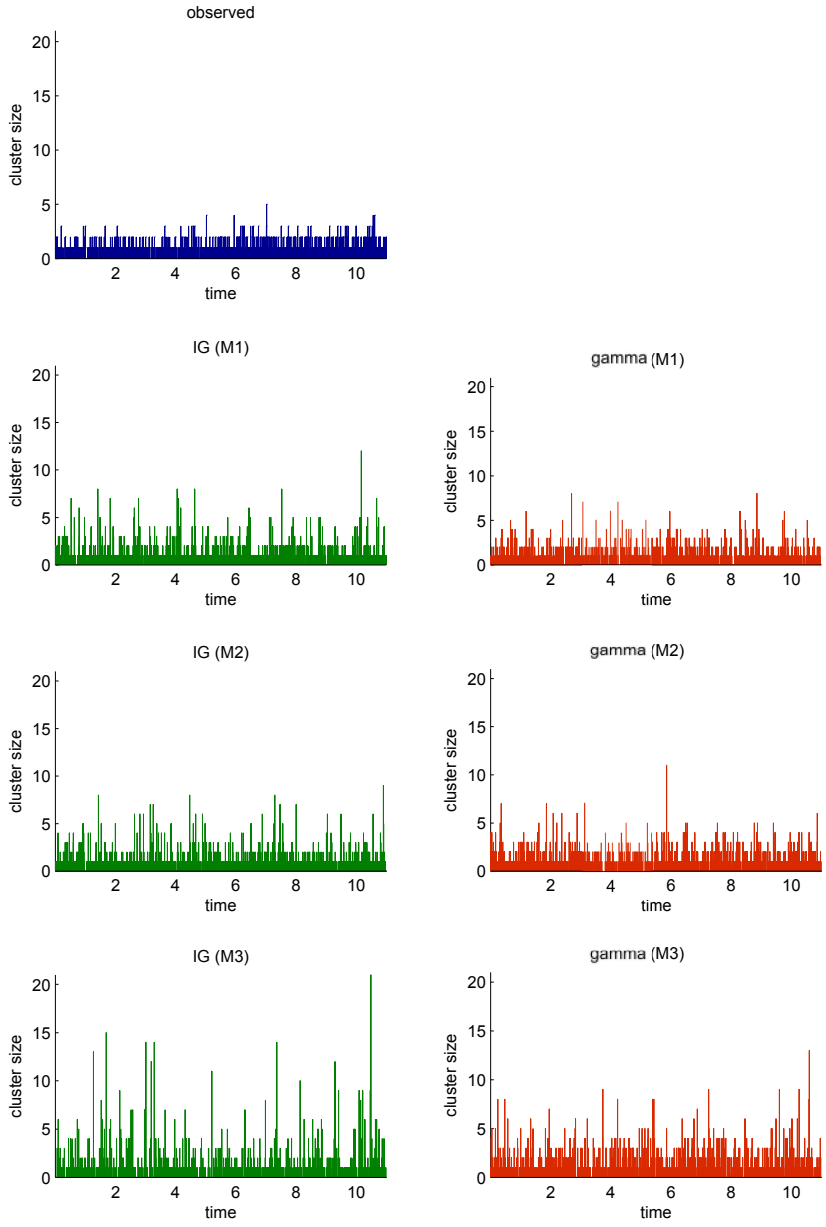


Figure A.28 Cluster sizes of content Danish fire insurance data and fitted models: The top chart shows the observed cluster sizes of the content claim number processes from the Danish fire insurance data. The charts in the bottom three rows present corresponding sampled cluster sizes of the fitted models based on an inverse Gaussian (left) and gamma (right) subordinator using the estimated parameters from Methods (M1)–(M3). (Method (M4) is excluded due to the in parts unreliable estimation results.)

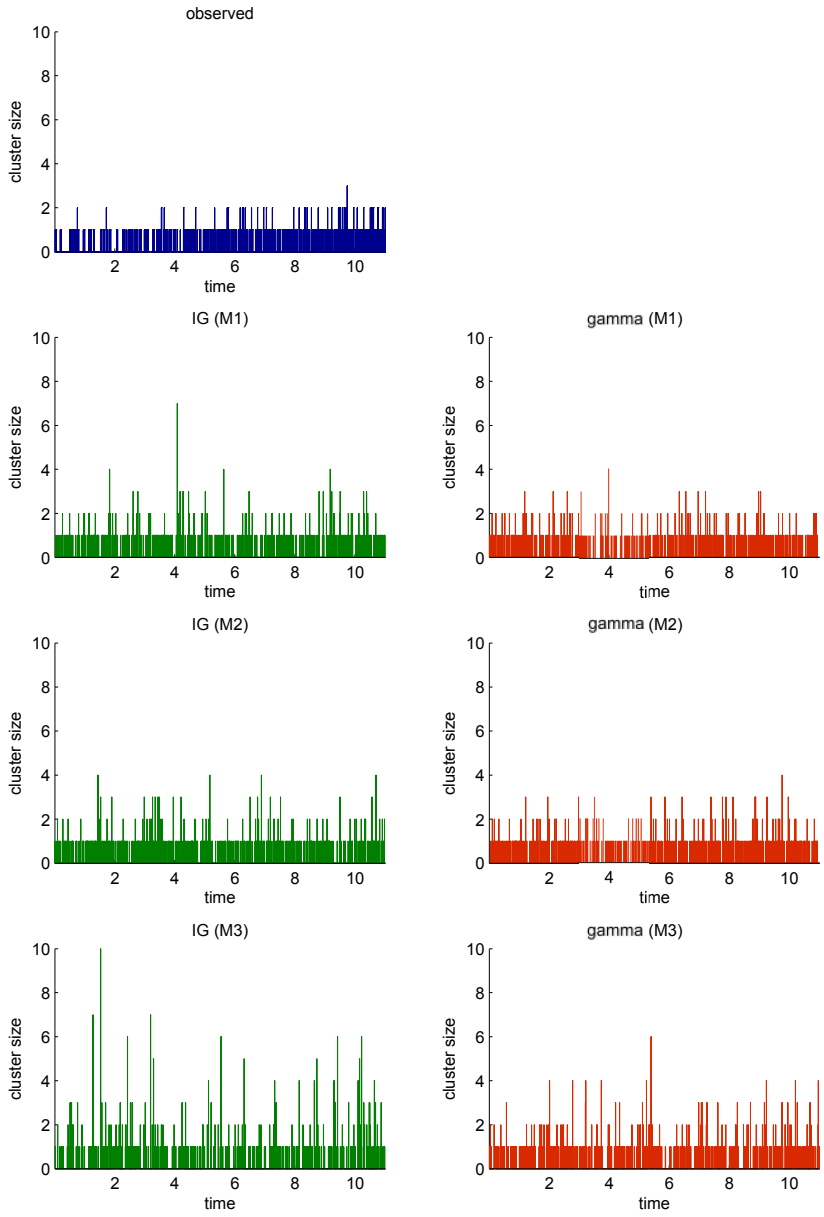


Figure A.29 Cluster sizes of profit Danish fire insurance data and fitted models: The top chart shows the observed cluster sizes of the profit claim number processes from the Danish fire insurance data. The charts in the bottom three rows present corresponding sampled cluster sizes of the fitted models based on an inverse Gaussian (left) and gamma (right) subordinator using the estimated parameters from Methods (M1)–(M3). (Method (M4) is excluded due to the in parts unreliable estimation results.)

Bibliography

Abate, J. and Whitt, W. (1992). Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, 10(1-2):5–87.

Albrecher, H. and Asmussen, S. (2006). Ruin probabilities and aggregrate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal*, 2006(2):86–110.

Apostol, T. M. (1962). *Calculus*. Blaisdell mathematics series. Blaisdell Publishing Company.

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

Avanzi, B., Cassar, L. C., and Wong, B. (2011). Modelling dependence in insurance claims processes with Lévy copulas. *Astin Bulletin*, 41(02):575–609.

Barndorff-Nielsen, O. E., Pedersen, J., and Sato, K.-I. (2001). Multivariate subordination, self-decomposability and stability. *Advances in Applied Probability*, 33(1):160–187.

Bartlett, M. S. (1963). The spectral analysis of point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 264–296.

Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical inference for stochastic processes*. Probability and mathematical statistics. Academic Press, London and New York.

Basu, S. and Dassios, A. (2002). A Cox process with log-normal intensity. *Insurance: Mathematics and Economics*, 31(2):297–302.

Bäuerle, N. and Blatter, A. (2011). Optimal control and dependence modeling of insurance portfolios with Lévy dynamics. *Insurance: Mathematics and Economics*, 48(3):398–405.

Bäuerle, N. and Grübel, R. (2005). Multivariate counting processes: copulas and beyond. *ASTIN Bulletin*, 35(2):379–408.

Bäuerle, N. and Grübel, R. (2008). Multivariate risk processes with interacting intensities. *Advances in Applied Probability*, 40(2):578–601.

Bäuerle, N. and Müller, A. (1998). Modeling and comparing dependencies in multivariate risk portfolios. *ASTIN Bulletin*, 28(1):59–76.

Bäuerle, N. and Müller, A. (2006). Stochastic orders and risk measures: consistency and bounds. *Insurance: Mathematics and Economics*, 38(1):132–148.

Beghin, L. and Macci, C. (2014). Fractional discrete processes: compound and mixed Poisson representations. *Journal of Applied Probability*, 51(1):19–36.

Bening, V. E. and Korolev, V. Y. (2002). *Generalized Poisson models and their applications in insurance and finance*. Walter de Gruyter.

Bernhart, G., Escobar Anel, M., Mai, J.-F., and Scherer, M. (2013). Default models based on scale mixtures of Marshall–Olkin copulas: properties and applications. *Metrika*, 76(2):179–203.

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

Bertoin, J. (1998). *Lévy processes*. Cambridge University Press.

Billingsley, P. (2009). *Convergence of probability measures*. John Wiley & Sons.

Billingsley, P. (2012). *Probability and measure*. Wiley Series in Probability and Statistics. Wiley, Hoboken N.J., anniversary edition.

Bochner, S. (1949). Diffusion equation and stochastic processes. *Proceedings of the National Academy of Sciences of the United States of America*, 35(7):368–370.

Bochner, S. (1955). *Harmonic analysis and the theory of probability*. University of California press Berkeley and Los Angeles.

Böcker, K. and Klüppelberg, C. (2008). Modelling and measuring multivariate operational risk with Lévy copulas. *Journal of Operational Risk*, 3(2):3–28.

Bregman, Y. and Klüppelberg, C. (2005). Ruin estimation in multivariate models with Clayton dependence structure. *Scandinavian Actuarial Journal*, 2005(6):462–480.

Brown, T. C. and Nair, M. G. (1988). A simple proof of the multivariate random time change theorem for point processes. *Journal of Applied Probability*, 25(1):210–214.

Bühlmann, H. (1970). *Mathematical methods in risk theory*, volume 172. Springer Science & Business Media.

Carley, H. (2002). Maximum and minimum extensions of finite subcopulas. *Communications in Statistics-Theory and Methods*, 31(12):2151–2166.

Carr, P. and Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics*, 71(1):113–141.

Casella, G. and Berger, R. L. (2002). *Statistical inference*. Duxbury advanced series. Thomson Learning, Australia Pacific Grove CA, 2nd edition.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, pages 135–155.

Comtet, L. (1974). *Advanced combinatorics: the art of finite and infinite expansions*. Springer.

Constantine, G. M. and Savits, T. H. (1996). A multivariate Faà di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348(2):503–520.

Cont, R., Stoikov, S., and Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563.

Cont, R. and Tankov, P. (2003). *Financial modelling with jump processes*. Taylor & Francis.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–164.

Czado, C. and Schmidt, T. (2011). *Mathematische Statistik*. Springer-Verlag.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes, volume I: elementary theory and methods of probability and its applications*. Springer New York.

Daley, D. J. and Vere-Jones, D. (2007). *An introduction to the theory of point processes, volume II: general theory and structure*. Springer New York.

Dario, A. De G. and Simonis, A. (2011). Properties of doubly stochastic Poisson process with affine intensity. *preprint arXiv:1109.2884*.

Das, S. R., Duffie, D., Kapadia, N., and Saita, L. (2007). Common failings: how sorporate defaults are correlated. *The Journal of Finance*, 62(1):93–117.

Dassios, A. and Jang, J. W. (2003). Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastics*, 7(1):73–95.

Davison, A. C. (2003). *Statistical models*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press.

Denuit, M., Dhaene, J., Goovaerts, M. J., and Kaas, R. (2005). *Actuarial theory for dependent risks*. Wiley Online Library.

Denuit, M., Genest, C., and Marceau, É. (2002). Criteria for the stochastic ordering of random sums, with actuarial applications. *Scandinavian Actuarial Journal*, 2002(1):3–16.

Denuit, M., Maréchal, X., Pitrebois, S., and Walhin, J. F. (2007). *Actuarial modelling of claim counts*. Wiley Online Library.

Dey, D. K. and Chung, Y. (1992). Compound Poisson distributions: properties and estimation. *Communications in Statistics - Theory and Methods*, 21(11):3097–3121.

Dhaene, J. and Goovaerts, M. J. (1995). Dependency of risks and stop-loss order. *DTEW Research Report 9545*, pages 1–18.

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q., and Vyncke, D. (2006). Risk measures and comonotonicity: a review. *Stochastic Models*, 22(4):573–606.

Dhaene, J., Wang, S., Young, V. R., and Goovaerts, M. (2000). Comonotonicity and maximal stop-loss premiums. *Bulletin of the Swiss Association of Actuaries*, 2:99–113.

Di Crescenzo, A., Martinucci, B., and Zacks, S. (2015). Compound Poisson process with Poisson subordinator. *Journal of Applied Probability*, 52(2).

Duffie, D. and Gârleanu, N. (2001). Risk and valuation of collateralized debt obligations. *Financial Analysts Journal*, 57(1):41–59.

Embrechts, P. (1993). Actuarial versus financial pricing of insurance. *The Journal of Risk Finance*, 1(4):17–26.

Embrechts, P. and Frei, M. (2009). Panjer recursion versus FFT for compound distributions. *Mathematical Methods of Operations Research*, 69(3):497–508.

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer.

Embrechts, P., McNeil, A. J., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, pages 176–223.

Esmaeili, H. and Klüppelberg, C. (2010). Parameter estimation of a bivariate compound Poisson process. *Insurance: Mathematics and Economics*, 47(2):224–233.

Feller, W. (1968). *An Introduction to probability theory and its applications: volume 1*, volume 1. John Wiley & Sons, 3rd edition.

Feller, W. (1971). *An introduction to probability theory and its applications: volume 2*. John Wiley & Sons, 2nd edition.

Frank, M. J., Nelsen, R. B., and Schweizer, B. (1987). Best-possible bounds for the distribution of a sum –a problem of Kolmogorov. *Probability Theory and Related Fields*, 74(2):199–211.

Galliher, H. P., Morse, P. M., and Simond, M. (1959). Dynamics of two classes of continuous-review inventory systems. *Operations Research*, 7(3):362–384.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *ASTIN Bulletin*, 37(2):475–515.

Giesecke, K., Goldberg, L. R., and Ding, X. (2011). A top-down approach to multiname credit. *Operations Research*, 59(2):283–300.

Giesecke, K. and Tomecek, P. I. (2005). Dependent events and changes of time. *Cornell University*.

Goovaerts, M., Vylder, F. E. d., and Haezendonck, J. (1984). Insurance premiums. *North-Holland: Amsterdam*, pages 1–398.

Grandell, J. (1991). *Aspects of risk theory*. Springer.

Grandell, J. (1997). *The mixed Poisson process*. Chapman & Hall.

Gundlach, M. and Lehrbass, F. (2004). *CreditRisk+ in the banking industry*. Springer Finance, Berlin and Heidelberg.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029.

Hardy, M. (2006). Combinatorics of partial derivatives. *The Electronic Journal of Combinatorics*, 13(1).

Hellmund, G., Prokešová, M., Jensen, E., and Vedel Jensen, E. B. (2008). Lévy-based Cox point processes. *Advances in Applied Probability*, pages 603–629.

Hofert, M., Mächler, M., and McNeil, A. J. (2012). Likelihood inference for Archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 110:133–150.

Hougaard, P., Lee, M.-L. T., and Whitmore, G. A. (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics*, 53(4):p 1225–1238.

Huber, P. J. and Ronchetti, E. (2009). *Robust statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken N.J., 2nd edition.

Itô, K. (1942, Reprinted in Kiyosi Itô Selected Papers, Springer: New York, 1987). On stochastic processes 1 (infinitely divisible laws of probability). *Japanese Journal of Methematics*, 18:261–301.

Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*. Springer, Berlin and New York, 2nd edition.

Jarrow, R. A. and Yu, F. (2001). Counterparty risk and the pricing of defaultable securities. *The Journal of Finance*, 56(5):1765–1799.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.

Joe, H. (2014). *Dependence modeling with copulas*, volume 134 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. *Technical Report no. 166, Department of Statistics, University of British Columbia*.

Johnson, N. L., Kemp, A. W., and Kotz, S. (1992). *Univariate discrete distributions*. Wiley, New York, 2nd edition.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete multivariate distributions*. Wiley, New York.

Kaas, R. (1993). How to (and how not to) compute stop-loss premiums in practice. *Insurance: Mathematics and Economics*, 13(3):241–254.

Kaas, R., Goovaerts, M. J., Dhaene, J., and Denuit, M. (2008). *Modern actuarial risk theory: using R*. Springer.

Kallenberg, O. (2002). *Foundations of modern probability*. Springer.

Kallsen, J. and Tankov, P. (2006). Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *Journal of Multivariate Analysis*, 97(7):1551–1572.

Karlis, D. and Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review*, 73(1):35–58.

Kemp, C. D. (1967). Stuttering-Poisson distributions. *Journal of the Statistical and Social Inquiry Society of Ireland*, 21(5):151–157.

Khintchine, A. (1937). Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. *Математический сборник*, 2(1):79–119.

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

Kingman, J. F. C. (1964). On doubly stochastic poisson processes. In *Mathematical proceedings of the Cambridge Philosophical Society*, volume 60, pages 923–930. Cambridge University Press.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2004). *Loss models: from data to decisions*. Wiley Series in Probability and Statistics. Wiley Interscience, Hoboken N.J., 2nd edition.

Kocherlakota, S. (1988). On the compounded bivariate Poisson distribution: a unified treatment. *Annals of the Institute of Statistical Mathematics*, 40(1):61–76.

Kumar, A., Nane, E., and Vellaisamy, P. (2011). Time-changed Poisson processes. *Statistics & Probability Letters*, 81(12):1899–1910.

Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives Research*, 2(2-3):99–120.

Lee, M.-L. T. and Whitmore, G. A. (1993). Stochastic processes directed by randomized time. *Journal of Applied Probability*, 30(2):p 302–314.

Lévy, P. (1934). Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 3(3-4):337–366.

Lewis, P. A. W. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. *Journal of Sound and Vibration*, 12(3):353–375.

Liebwein, P. (2009). *Klassische und moderne Formen der Rückversicherung*. Verlag Versicherungswirtschaft.

Lindskog, F. and McNeil, A. J. (2003). Common Poisson shock models: applications to insurance and credit risk modelling. *ASTIN Bulletin*, 33(2):209–238.

Luciano, E. and Schoutens, W. (2006). A multivariate jump-driven financial asset model. *Quantitative Finance*, 6(5):385–402.

Luciano, E. and Semeraro, P. (2007). Extending time-changed Lévy asset models through multivariate subordinators. *SSRN Electronic Journal*.

Mai, J.-F., Olivares, P., Schenk, S., and Scherer, M. (2014). A multivariate default model with spread and event risk. *Applied Mathematical Finance*, 21(1):51–83.

Mai, J.-F. and Scherer, M. (2009a). Lévy-frailty copulas. *Journal of Multivariate Analysis*, 100(7):1567–1585.

Mai, J.-F. and Scherer, M. (2009b). A tractable multivariate default model based on a stochastic time-change. *International Journal of Theoretical and Applied Finance*, 12(2):227–249.

Mai, J.-F. and Scherer, M. (2012a). H-extendible copulas. *Journal of Multivariate Analysis*, 110:151–160.

Mai, J.-F. and Scherer, M. (2012b). *Simulating copulas: stochastic models, sampling algorithms, and applications*, volume 4. World Scientific.

Mai, J.-F., Scherer, M., and Zagst, R. (2013). CIID frailty models and implied copulas. In *Copulae in mathematical and quantitative finance*, pages 201–230. Springer.

Marshall, A. W. and Olkin, I. (1967). A generalized bivariate exponential distribution. *Journal of Applied Probability*, pages 291–302.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative risk management*. Princeton Series in Finance. Princeton University Press.

Mikosch, T. (2009). *Non-life insurance mathematics*. Springer, Berlin and Heidelberg.

Moller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482.

Müller, A. (1997). Stop-loss order for portfolios of dependent risks. *Insurance: Mathematics and Economics*, 21(3):219–223.

Müller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models & risks*. Wiley Series in Probability and Statistics. Wiley, Chichester.

Nelsen, R. (2006). *An introduction to copulas*. Springer Series in Statistics. Springer, 2nd edition.

Orsingher, E. and Polito, F. (2012). The space-fractional Poisson process. *Statistics & Probability Letters*, 82(4):852–858.

Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. *Astin Bulletin*, 12(01):22–26.

Panjer, H. H., Willmot, G. E., and Education, A. (1992). *Insurance risk models*, volume 479. Society of Actuaries Schaumburg, Illinois.

Partrat, C. (1994). Compound model for two dependent kinds of claim. *Insurance: Mathematics and Economics*, 15(2):219–231.

Pfeiffer, D. and Nešlehová, J. (2004). Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin*, 34(2):333–360.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer.

Prause, K. (1999). *The generalized hyperbolic model: estimation, financial derivatives, and risk measures*. PhD thesis, University of Freiburg.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer.

Riordan, J. (1946). Derivatives of composite functions. *Bulletin of the American Mathematical Society*, 52:664–667.

Rolski, T., Schmidli, H., Schmidt, V., and Teugels, J. (1999). *Stochastic processes for insurance and finance*. Wiley.

Rudolph, C. (2014). *A Generalization of Panjer's recursion for dependent claim numbers and an approximation of Poisson mixture models*. PhD thesis, Technische Universität Wien, Vienna.

Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Scarsini, M. (1984). On measures of concordance. *Stochastica: revista de matemática pura y aplicada*, 8(3):201–218.

Schilling, R. L., Song, R., and Vondraček, Z. (2012). *Bernstein functions: theory and applications*. De Gruyter, Berlin.

Schoutens, W. (2003). *Lévy processes in finance: pricing financial derivatives*. Wiley, Chichester and West Sussex and New York.

Semeraro, P. (2008). A multivariate variance Gamma model for financial applications. *International Journal of Theoretical and Applied Finance*, 11(01):1–18.

Serfling, R. J. (2002). *Approximation theorems of mathematical statistics*. Wiley series in probability and mathematical statistics. Wiley, New York.

Shaked, M. (1980). On mixtures from exponential families. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–198.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. Springer Science & Business Media.

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

Skorohod, A. V. (1991). *Random processes with independent increments*. Kluwer Academic Publishers, Dordrecht and Boston.

Skorokhod, A. V. (1956). Limit theorems for stochastic processes. *Theory of Probability & Its Applications*, 1(3):261–290.

Stein, G. Z., Zucchini, W., and Juritz, J. M. (1987). Parameter estimation for the Sichel distribution and its multivariate extension. *Journal of the American Statistical Association*, 82(399):938.

Suisse, C. (1997). CreditRisk+: a credit risk management framework. *Credit Suisse Financial Products*.

Sundt, B. (1999). On multivariate Panjer recursions. *Astin Bulletin*, 29(01):29–45.

Sundt, B. (2000). Multivariate compound Poisson distributions and infinite divisibility. *Astin Bulletin*, 30(02):305–308.

Sundt, B. and Vernic, R. (2004). Recursions for compound mixed multivariate poisson distributions. *Blätter der DGVFM*, 26(4):665–691.

Tankov, P. (2003). Dependence structure of spectrally positive multidimensional Lévy processes. *Unpublished manuscript, http://www.proba.jussieu.fr/pageperso/tankov/levycopulas.ps*.

Vere-Jones, D. (1982). On the estimation of frequency in point-process data. *Journal of Applied Probability*, 19:383–394.

Wang, S. and Dhaene, J. (1998). Comonotonicity, correlation order and premium principles. *Insurance: Mathematics and Economics*, 22(3):235–242.

Wheeler, F. S. (1987). Bell polynomials. *ACM SIGSAM Bulletin*, 21(3):44–53.

Widder, D. V. (1952). *The Laplace transform*, volume 6 of *Princeton mathematical series*. Princeton University Press, 3rd edition.

Zocher, M. (2005). *Multivariate mixed Poisson processes*. PhD thesis, Technische Universität Dresden.

List of Tables

List of Figures