



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2021

TEMPERING AND SEASONALITY IN NON-LIFE INSURANCE MODELING

Araujo Acuna José Carlos

Araujo Acuna José Carlos, 2021, TEMPERING AND SEASONALITY IN NON-LIFE
INSURANCE MODELING

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_2EB2AE54BD9B2

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES
DÉPARTEMENT DE SCIENCES ACTUARIELLES

**TEMPERING AND SEASONALITY IN NON-LIFE
INSURANCE MODELING**

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Actuarielles

par

José Carlos ARAUJO ACUNA

Directeur de thèse
Prof. Hansjoerg Albrecher

Co-directeur de thèse
Prof. Jan Beirlant

Jury

Prof. Felicitas Morhart, présidente
Prof. François Dufresne, expert interne
Prof. Benjamin Avanzi, expert externe

LAUSANNE
2021



UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES
DÉPARTEMENT DE SCIENCES ACTUARIELLES

**TEMPERING AND SEASONALITY IN NON-LIFE
INSURANCE MODELING**

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences Actuarielles

par

José Carlos ARAUJO ACUNA

Directeur de thèse
Prof. Hansjoerg Albrecher

Co-directeur de thèse
Prof. Jan Beirlant

Jury

Prof. Felicitas Morhart, présidente
Prof. François Dufresne, expert interne
Prof. Benjamin Avanzi, expert externe

LAUSANNE
2021

IMPRIMATUR

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur José Carlos ARAUJO ACUNA, titulaire d'un bachelor en Sciences actuarielles de l'Université de Yucatan, et d'un master en Sciences actuarielles de l'Université de Lausanne, en vue de l'obtention du grade de docteur ès Sciences actuarielles.

La thèse est intitulée :

TEMPERING AND SEASONALITY IN NON-LIFE INSURANCE MODELING

Lausanne, le 02 février 2021

Le doyen



Jean-Philippe Bonardi

Members of the Jury

PROF. FELICITAS MORHART

President of the jury,
Department of Marketing, University of Lausanne, Switzerland.

PROF. HANSJÖRG ALBRECHER

Thesis supervisor,
Department of Actuarial Science, University of Lausanne, Switzerland.

PROF. JAN BEIRLANT

Thesis co-supervisor,
Department of Mathematics, Catholic University of Leuven, Belgium.

PROF. FRANÇOIS DUFRESNE

Internal expert,
Department of Actuarial Science, University of Lausanne, Switzerland.

PROF. BENJAMIN AVANZI

External expert,
Department of Economics, University of Melbourne, Australia.

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

José Carlos ARAUJO ACUNA

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:



Date: 01/02/2021

Prof. Hansjoerg ALBRECHER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

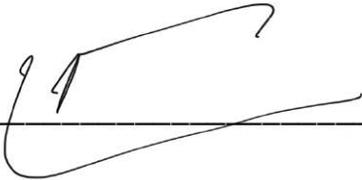
PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

José Carlos ARAUJO ACUNA

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 2/2/2021

Prof. Jan BEIRLANT
Thesis co-supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

José Carlos ARAUJO ACUNA

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  _____

Date: 2 February 2021

Prof. François DUFRESNE
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

José Carlos ARAUJO ACUNA

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: 

Date: 2 February 2021

Prof. Benjamin AVANZI
External member of the doctoral committee

Acknowledgements

First of all, I would like to thank my supervisors Prof. Hansjörg Albrecher and Prof. Jan Beirlant for their enthusiasm, guidance, and kindness. I am very grateful for their valuable input and the interesting discussions we shared. I learned a lot from them and I will always remember their valuable lessons. As a PhD student I can say that anyone who has Prof. Albrecher or Prof. Beirlant as a supervisor is a very lucky person, and I had the fortune of having been supervised by both of them.

I would also like to thank my PhD committee members Prof. François Dufresne and Prof. Benjamin Avanzi. Their helpful advice and insightful suggestions helped me very much improve the quality of my thesis. I particularly would like to thank Prof. Dufresne for his support during my master studies and for giving me the opportunity of becoming a student assistant for the first time. I enjoyed the teaching experience very much and I am very grateful for his trust.

Appreciative thanks to all my friends from the department for the interesting discussions during the coffee breaks and for cheering me up when I needed it the most. Special thanks go to Anca, Maissa, Yves, William, Michel, Martin, Dina, José Miguel, Andrey, and Brandon.

These acknowledgments would not be complete without a mention of my family. First, I would like to thank Angeles, Jürg, Nadhira, and Nick who were always there to encourage me. Stephanie deserves a special word of gratitude for her patience, understanding, and for being my constant source of support. Finally, I would like to thank my parents, Carlos and Lety, to whom this thesis is dedicated, and to my siblings Rebe and Edu. None of these would have been possible without their love, support and continuous inspiration through all these years.

Preface

Insurance is a contract represented by a policy where the insured person or policyholder (individual or entity) pays a fixed amount to safeguard against uncertain losses that could be quite significant financially and where the insurer agrees to cover these losses in case they occur. The underlying risks can be of very different nature, but they are traditionally divided into the areas of life and non-life insurance. This thesis is dedicated to various aspects of the latter and contains results in the area of modeling and assessing risks of insurance portfolios. It is based on three papers that have either been published or submitted for publication in peer-reviewed journals.

Chapter 1 describes some basic concepts and contains classical results that are used in subsequent chapters. Chapter 2 presents a non-stationary extension of the Cox process methodology with Lévy subordinators presented in Selch and Scherer [169], by enabling trends and seasonal behavior in the underlying subordinator. This approach allows to keep all the advantages of the Selch-Scherer model, taking into account non-stationary behavior, and leads to a flexible, yet parsimonious continuous-time model for claim counts that considers both clustering and non-stationary properties. In the context of heavy-tail modeling, Chapter 3 generalizes earlier results of Meerschaert et al. [145] and Raschke [157] by considering the tempering of a Pareto-type distribution with a general Weibull distribution in a peaks-over-threshold approach. The result provides a relevant rather general model for claim sizes where the Pareto behavior only sets in after a certain threshold and where such a behavior does not extend indefinitely due to truncation or tempering effects in the data. Finally, within the setup of a compound Poisson risk model, we consider in Chapter 4 the problem of finding bounds for the ruin probability when only limited information about the claim size distribution is available. We consider random initial surplus, leading to considerably more amenable expressions for the ruin probability. Under that framework, we explore the resulting computational advantages, and we further show how the solution to the problem with deterministic initial surplus can be approximated via random initial surplus.

Contents

Acknowledgements	xv
Preface	xvii
1 Introduction	1
1.1 Claim frequency modeling	2
1.1.1 Homogeneous Poisson process	4
1.1.2 Inhomogeneous Poisson process	6
1.1.3 Mixed Poisson process	7
1.1.4 Doubly-stochastic Poisson process	9
1.2 Claim severity modeling	13
1.2.1 Extreme value analysis	13
1.3 The Cramér-Lundberg model	16
1.3.1 An integro-differential equation	18
1.3.2 The Laplace transform and the Pollaczek-Khintchine formula	19
1.3.3 Extensions of the classical risk model	20
1.4 Stochastic orders	23
1.5 Main contributions of this thesis	25
2 Fitting non-stationary Cox processes	27
2.1 Introduction	28
2.2 A motivating example	29
2.3 The non-stationary Cox model	33
2.3.1 Poisson processes directed by a Lévy subordinator	33
2.3.2 Poisson processes directed by an additive process	34
2.4 Parameter estimation	37

2.4.1	Simulation study	39
2.5	Case study: Dutch fire insurance data	41
2.5.1	Model (M_0)	41
2.5.2	Non-stationary models	42
2.5.3	Backtesting	50
2.5.4	Risk assessment based on Model (M_2)	50
2.6	A multivariate extension and case study	54
2.6.1	Case study: Danish fire insurance data.	56
2.7	Conclusion	59
2.8	Appendix A: Model comparison	61
3	Tempered Pareto-type modeling	63
3.1	Introduction	64
3.2	Tempered Pareto-type modeling and estimation	68
3.2.1	Statistical estimation	69
3.2.2	Return periods and extreme quantiles	71
3.2.3	Asymptotic distribution of the maximum likelihood estimators	72
3.3	Simulation results	73
3.4	Insurance cases	75
3.5	Conclusion	76
3.6	Appendix A: Proof of Theorem 3.2.1	77
3.7	Appendix B: Simulation results	80
3.8	Appendix C: Insurance cases	92
4	On the randomized Schmitter problem	99
4.1	Introduction	99
4.2	Preliminaries and previous results	101
4.3	Exponentially distributed initial surplus	105
4.3.1	Numerical illustrations.	110
4.4	Erlang distributed initial surplus	115
4.5	Conclusion	118

Chapter 1

Introduction

An insurance policy is a contract between an insurance company and a policyholder in which the former agrees to compensate the latter for certain unpredictable losses during a period of time against a fee, the premium. The insurance company pools similar risks whose individual insurance claims are unknown at the beginning of the insurance period and hence they need to be treated within the framework of probability theory and statistics. In other words, the insurance risks can be described by sequences of (often independent) random variables representing the random amount of money that the insurance company will have to pay out to indemnify the policyholder for the consequences of the occurrence of the insured risk. Therefore, in order for the insurance company to guarantee its contractual obligations, and its own success, determining an adequate amount for the premium becomes a fundamental necessity.

In non-life insurance, traditional actuarial methods to calculate the premium rely on statistical models using the company's own data on insurance policies and claims for the portfolio. Under the assumption of independence between the number of occurrences (claim frequency) and the size of the insured losses (claim severity), the pure insurance premium is calculated as the product of the claim severity and the claim frequency, and hence two models need to be calibrated. Furthermore, in situations where a large amount of data is available, especially at a policyholder level, it is possible to rate policyholders according to their individual characteristics, or risk factors, using covariate-based methods such as generalized linear models (cf. Ohlsson and Johansson [152]) or even more sophisticated data-driven methods such as trees and neural networks (cf. Denuit et al. [80, 79]). Once the pure premium is set, the actuarial premium can be obtained by applying a safety loading according to some premium principle (cf. Bühlmann [48, Ch.4], Gerber [102, Ch. 5], Dickson [88, Ch. 3], Kaas et al. [124, Ch. 5]) and optimized according to the policyholder, shareholder, and/or market demands (see also Parodi [154] for a survey on insurance pricing from a practitioner's perspective).

The insurance company needs to constantly monitor the insurance portfolio solvency and its regulatory capital requirements. Recently, the focus of solvency and capital requirements has shifted towards risk measures like Value-at-Risk (VaR) and con-

ditional tail expectation (CTE). In the actuarial literature, traditionally the focus was on (finite-time) ruin probabilities which provide a measure of the behavior of an insurance portfolio through time. In particular, it serves as an assessment of the strength of the insurer's premium and claim process in relation to the available capital. A high probability of ruin should trigger suitable risk management actions to avoid undesirable financial outcomes. The event of ruin rarely occurs in practice, so the probability of ruin has only a limited absolute meaning. However, it enables one to compare portfolios concerning the solvency aspect and is a classical object of study in actuarial science.

In this introductory chapter we provide some theoretical foundations in the research areas to which the later chapters of this thesis aim to contribute. In that regard, Section 1.1 will give a brief introduction to claim frequency modeling that will later constitute the backbone of the model presented in Chapter 2. In relation to Chapter 3, Section 1.2 provides a brief introduction to claim severity modeling and in particular to extreme value theory. Furthermore, Section 1.3 cites the basic results of the classic theory of ruin probabilities, and Section 1.4 briefly touches upon stochastic orders, which will also be needed in Chapter 4. Finally, the main contributions of this thesis are outlined in Section 1.5.

This chapter provides references relevant to the research areas of this thesis. Due to the vast amount of work in these fields, the overview can by no means be exhaustive and only touches upon several important aspects. Many further references will be given throughout the later chapters of the thesis.

1.1 Claim frequency modeling

Claim frequency modeling is about estimating the number of claims per time unit. In insurance practice one frequently has annual claim counts available. Hence, if the time unit under consideration and the risk exposure of the portfolio is fixed, one fits discrete positive random variables to these insurance data. The most popular distributions being the binomial, Poisson and negative binomial ones. These distributions not only have properties that are suitable to model insurance data, but also form part of the $(a, b, 0)$ -class of distributions which is based on a simple recursion needed to obtain a recursive algorithm for the calculation of the aggregate claim amount (cf. Panjer [153], Sundt-Jewel, [175], Klugman et al. [130], Johnson and Kotz [120]). Here we present these three distributions.

Binomial distribution. Given that the number of claims in an insurance portfolio can be thought of as each policyholder being subject to a random experiment with a binary outcome (claim, no claim), the binomial distribution arises naturally in claim number modeling. Moreover, it is particularly useful in situations where the insurance data show underdispersion (sample variance is smaller than the sample mean).

We say that N has a binomial distribution, $N \sim \text{Bin}(n, p)$, if

$$\mathbb{P}(N = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 < p < 1, \quad k = 0, 1, 2, \dots, n.$$

The first two moments of the binomial distribution are given by

$$\mathbb{E}(N) = np, \quad \text{Var}(N) = np(1-p).$$

Poisson distribution. The Poisson distribution is frequently used to count the number of events that occur in a certain time interval or spatial area. It was first introduced by the French mathematician Siméon Denis Poisson in 1838. We say that N has a Poisson distribution, $N \sim \text{Poisson}(\lambda)$, if

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda > 0, \quad k = 0, 1, 2, \dots$$

The first two moments of the Poisson distribution are given by

$$\mathbb{E}(N) = \lambda, \quad \text{Var}(N) = \lambda,$$

so that it is equidispersed (the variance equals the mean). It arises as the limit of the binomial distribution for n large, p small and $np := \lambda$ fixed. The latter is sometimes referred to as the law of small numbers. In an insurance portfolio the insurer will pool a sufficiently large number of risks in order to balance the aggregate risk. Each individual risk is likely to have a small probability to occur and under independence the Poisson distribution can then be a good approximation. Furthermore, the Poisson distribution is a good model for the number of events under the assumption that the time in between claim arrivals is exponentially distributed. This distribution is the backbone of the Poisson process introduced later and further extended in Chapter 2.

Negative binomial distribution. We say that N has a negative binomial distribution, $N \sim \text{NBin}(r, p)$, if

$$\mathbb{P}(N = k) = \binom{r+k-1}{k} p^k (1-p)^r, \quad 0 < p < 1, \quad r > 0, \quad k = 0, 1, 2, \dots$$

It can be obtained as an extension of the Poisson distribution when the formerly constant parameter λ is assumed to be a Gamma random variable (i.e., it is a particular mixed Poisson distribution). The first two moments of the negative binomial distribution are given by

$$\mathbb{E}(N) = \frac{rp}{1-p}, \quad \text{Var}(N) = \frac{rp}{(1-p)^2},$$

so that – as any mixed Poisson distribution – it is overdispersed (the variance is larger than the mean).

While the above distributions are convenient and simple models, they do not capture how the number of claims evolves over time. In this thesis we focus on continuous

time modeling of claim occurrences, which can be a more flexible and tractable tool that allows the implementation of stylized features in a transparent and elegant way. To provide the necessary background, we now give some basic definitions of counting processes and stochastic processes based on Poisson processes. Standard references, where more mathematical details and derivations can be found, are Grandell [109], Mikosch [146], Rolski et al. [161], Denuit et al. [87] and Albrecher et al. [7, Ch.5].

Definition 1.1.1. *A counting process $\{N(t); t \geq 0\}$ is a stochastic process that counts the number of claims up to particular point in time t and satisfies the following properties:*

- i) It is non-negative.*
- ii) It is integer-valued.*
- iii) It is non-decreasing, i.e., if $s < t$, then $N(s) \leq N(t)$.*
- iv) For $s < t$, $N(t) - N(s)$ denotes the number of claims in the interval $(s, t]$.*

1.1.1 Homogeneous Poisson process

When modeling count data using counting processes, the homogeneous Poisson process is the benchmark model. Though often not flexible enough, its attractive theoretical properties make it the most popular claim number process in the actuarial literature. A formal definition of a homogeneous Poisson process is now given.

Definition 1.1.2. *A homogeneous Poisson processes $\{\tilde{N}_\lambda(t); t \geq 0\}$ with intensity $\lambda > 0$ is a stochastic process with the following properties:*

- i) Starts at zero, i.e. $\tilde{N}_\lambda(0) = 0$ a.s.*
- ii) Independent increments: for any $0 \leq t_1 < t_2 < \dots < t_n < \infty$, $n \in \mathbb{N}$, the increments $\tilde{N}_\lambda(t_i) - \tilde{N}_\lambda(t_{i-1})$ are mutually independent for $i = 1, \dots, n$.*
- iii) Poisson increments: for any $0 \leq s < t < \infty$*

$$\tilde{N}_\lambda(t) - \tilde{N}_\lambda(s) \sim \text{Poisson}(\lambda(t - s)).$$

- iv) \tilde{N}_λ has càdlàg paths a.s.*

From Property iii) it follows that the increments are stationary. In other words, the intensity of the process depends only upon the length of the time interval and not upon its position: $\tilde{N}_\lambda(t) - \tilde{N}_\lambda(s) \stackrel{d}{=} \tilde{N}_\lambda(t - s)$ for any $0 \leq s < t < \infty$. Furthermore,

$$\mathbb{E}(\tilde{N}_\lambda(t - s)) = \text{Var}(\tilde{N}_\lambda(t - s)) = \lambda(t - s) \quad 0 \leq s < t < \infty.$$

The homogeneous Poisson processes has several desirable properties, some of them are summarized in the following. The technical details of this can be found in Mikosch [146, Ch. 2.1.4 to 2.1.6] and Rolski et al. [161, Ch. 5.2.1]. Define the claim arrival times

$$T_j := \inf\{t > 0 : \tilde{N}_\lambda(t) \geq j\}, \quad j \in \mathbb{N}.$$

- i) The inter-arrival times $W_j := T_j - T_{j-1}, j \in \mathbb{N}$ where $T_0 := 0$, are independent and identically distributed (i.i.d.) exponential random variables with mean $1/\lambda$:

$$F_{W_j}(x) = 1 - \exp(-\lambda x), \quad j \in \mathbb{N}^+, \quad x \geq 0.$$

From this property it follows the memoryless property. Namely, the distribution of the time until the next arrival is independent of the time t we have already been waiting for that arrival:

$$\mathbb{P}(W_j > t + y \mid W_j > t) = \mathbb{P}(W_j > y) \quad \text{for all } y, t \geq 0.$$

- ii) Order statistics property: the conditional distribution of (T_1, \dots, T_n) given $\{\tilde{N}_\lambda(t) = n\}$ for some $n \in \mathbb{N}$ equals the distribution of the order statistics $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$ of independent uniform $(0, t)$ distributed random variables.
- iii) Jump sizes: as $t \downarrow 0$

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

where $o(t)$ is a function that tends to 0 faster than the identity, i.e., $\lim_{t \rightarrow 0} o(t)/t = 0$. Intuitively, this means that in a sufficiently small time interval, the probability of two or more claim arrivals is negligible when compared to the probability of zero or one arrival. Hence, at any point in time, no more than one claim can occur with positive probability.

Property iii) gives the homogeneous Poisson process a special role among all claim number processes and may be seen as one of the main reasons for its popularity from a modeling perspective. Furthermore, the homogeneous Poisson process can be constructed based on the arrivals of its independent and exponentially distributed inter-arrival times, that is,

$$\tilde{N}_\lambda(t) = \sum_{j=1}^{\infty} 1_{\{W_1 + \dots + W_j \leq t\}}.$$

Note that the latter construction together with Property i) provides the basis to simulate sample paths of a Poisson process. Alternatively, one can take advantage of Property iii). For a discussion on simulating Poisson processes see for example Korn et al. [131, Ch. 6.2] or Schoutens [166, Ch. 8.1.2].

The homogeneous Poisson process plays a major role in ruin theory also. In the classical Cramér-Lundberg model it is assumed that the claims in the portfolio arrive according an homogeneous Poisson processes with rate λ . Hence, the evolution of the aggregate claims amount of the insurance portfolio is modeled by means of a compound Poisson process. A definition of the latter follows.

Definition 1.1.3. Let $\{\tilde{N}_\lambda(t); t \geq 0\}$ be a homogeneous Poisson process with intensity $\lambda > 0$. Let X be a random variable and X_1, X_2, \dots be i.i.d. copies of X independent of $\tilde{N}_\lambda(t)$. Then the process

$$S(t) = \sum_{i=1}^{\tilde{N}_\lambda(t)} X_i, \quad t \geq 0,$$

is called a compound Poisson process.

Section 1.3 will provide further insights about its role in the surplus process of the insurance company.

1.1.2 Inhomogeneous Poisson process

One restriction of the homogeneous Poisson process is its stationary property. One can define a more general Poisson process to deal with the latter:

Definition 1.1.4. A Poisson processes $\{N_\mu(t); t \geq 0\}$ with mean-value function μ is a stochastic process defined by the following properties:

- i) Starts at zero, i.e. $N_\mu(0) = 0$ a.s.
- ii) Independent increments: for any $0 \leq t_1 < t_2 < \dots < t_n < \infty$, $n \in \mathbb{N}$, the increments $N_\mu(t_i) - N_\mu(t_{i-1})$ are mutually independent for $i = 1, \dots, n$.
- iii) Poisson increments: for a non-decreasing càdlàg function $\mu : [0, \infty) \rightarrow [0, \infty)$ with $\mu(t) < \infty$ for all $0 \leq s < t < \infty$

$$N_\mu(t) - N_\mu(s) \sim \text{Poisson}(\mu(t) - \mu(s)).$$

- iv) N_μ has càdlàg paths a.s.

Note that a Poisson process with linear mean value function $\mu(t) = \lambda t$, $t \geq 0$ for some $\lambda > 0$ is a homogeneous Poisson process. Moreover, from the properties of the Poisson distribution one gets

$$\mathbb{E}(N_\mu(t) - N_\mu(s)) = \text{Var}(N_\mu(t) - N_\mu(s)) = \mu(t) - \mu(s), \quad 0 \leq s < t < \infty.$$

In fact, there is a strong connection between both processes. Namely, an inhomogeneous Poisson processes can be defined through a deterministic time-change of a homogeneous process. Let μ be a mean-value function according to the definition above and \tilde{N}_1 a standard homogeneous Poisson process with intensity parameter $\lambda = 1$. Then, the process defined by $\{\tilde{N}_1(\mu(t)); t \geq 0\}$ is an inhomogeneous Poisson process with mean-value function μ . Moreover, the process $\{N_\mu(\mu(t)); t \geq 0\}$ is a standard homogeneous Poisson process. For this reason, μ can be considered as the operational time of the counting process $N(t)$: whereas time runs linearly for a homogeneous Poisson process, it accelerates or decelerates according to μ for an

inhomogeneous Poisson process (cf. Bühlmann [48, Ch. 2.2.3], see Albrecher [2] for a general account of the concept of operational time).

If the inverse function μ^{-1} exists, i.e. μ is continuous, strictly increasing and $\lim_{t \rightarrow \infty} \mu(t) = \infty$, then the inhomogeneous Poisson process can be converted back to a homogeneous Poisson process with intensity 1 by the time change μ^{-1} (cf. Mikosch [146, Ch. 2.1.3]).

In many applications it is assumed that μ is absolutely continuous and has an intensity function λ , i.e., there exists a nonnegative measurable function $\lambda(\cdot)$ such that for any $s < t$ the increment $\mu(t) - \mu(s)$ has representation

$$\mu(t) - \mu(s) = \int_s^t \lambda(z) dz, \quad s < t.$$

Although the inhomogeneous Poisson process is no longer stationary, given that the Poisson distribution has a mean to variance ratio of 1, it fails to explain the variability often present in insurance data sets. Also, the deterministic nature of the intensity function is not always suitable when modeling counting phenomena. Thus, one way to create greater flexibility is randomizing the intensity parameter (or the mean-value function).

1.1.3 Mixed Poisson process

The mixed Poisson process was introduced to actuaries by Dubourdieu [89] and plays an important role in actuarial practice given its connection with the negative binomial distribution. Mixed Poisson processes generalize the Poisson process when replacing the intensity parameter λ by a positive random variable Λ with distribution function F_Λ .

The randomization of the intensity function increases the flexibility of the model, and at the same time preserves some nice properties of the homogeneous Poisson process, for example the order statistics property. Namely, conditional on $\tilde{N}_\Lambda(t) = n$, the occurrence times of the n claim events are uniformly distributed on $(0, t)$ (cf. Grandell [109]). Frequently, the randomization of the intensity function approach is motivated by addressing heterogeneity within the insured population. For example, in a portfolio of car insurance policies we can think of Λ representing different factors of influence on the claim arrivals of the portfolio such as policyholders' driving behaviors, driving skills, age, etc.

A general definition of mixed Poisson processes is given in the following:

Definition 1.1.5. *Let $\{\tilde{N}_1(t); t \geq 0\}$ be a standard homogeneous Poisson process and μ a valid mean value function of a Poisson process. Furthermore, let $\Lambda > 0$ a.s. be a random variable independent of $\tilde{N}_1(t)$. Then the process*

$$\tilde{N}_\Lambda(t) = \{\tilde{N}_1(\Lambda \cdot \mu(t)); t \geq 0\},$$

is said to be a mixed Poisson process with mixing variable Λ .

The mean value function $\mu(t)$ plays a central role in insurance applications. While the random variable Λ will explain the heterogeneity, it is through the mean value function $\mu(t)$ that one captures the non-stationary behavior often present in the data. For example, it might describe the evolution of the number of insured persons, number of policies, or number of risks through time.

In this work (as at many other places in the literature), we restrict ourselves to the case $\mu(t) = t$. For any $s > 0$ and $0 \leq t < \infty$, the probability that k events occur in the time interval $(t, t + s]$ is given by

$$\mathbb{P}(\tilde{N}_\Lambda(t + s) - \tilde{N}_\Lambda(t) = k) = \int_0^\infty \frac{(\lambda s)^k}{k!} \exp(-\lambda s) dF_\Lambda(\lambda).$$

Therefore, for a non-degenerated mixing variable, a mixed Poisson process no longer has independent increments (but conditionally independent ones). However, conditional to the outcome of the mixing variable, a mixed Poisson process becomes a homogeneous Poisson process.

From the tower property of the conditional expectation and the law of total variance, it follows that

$$\mathbb{E}(\tilde{N}_\Lambda(t)) = t\mathbb{E}(\Lambda) \quad \text{and} \quad \text{Var}(\tilde{N}_\Lambda(t)) = t\mathbb{E}(\Lambda) + t^2\text{Var}(\Lambda),$$

(cf. Grandell [109, Proposition 2.1]). Hence, unless Λ is a degenerate random variable, the variance exceeds the mean, i.e., the process is overdispersed. It is immediate to see that the index of dispersion equals

$$I_{\tilde{N}_\Lambda(t)} = 1 + t \frac{\mathbb{E}(\Lambda)}{\text{Var}(\Lambda)} > 1.$$

When the mixing distribution is a Gamma random variable with probability density function

$$f_\Lambda(\lambda) = \frac{\eta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\eta\lambda}, \quad \lambda > 0, \alpha, \eta > 0,$$

the resulting mixed Poisson process is referred to as a *Pólya process*. Elementary calculations show that the resulting distribution of the number of claims up to time t is a negative binomial distribution with probability mass function

$$\mathbb{P}(\tilde{N}_\Lambda(t) = k) = \binom{\alpha + k - 1}{k} \left(\frac{\eta}{t + \eta} \right)^\alpha \left(\frac{t}{t + \eta} \right)^k.$$

As one has two parameters here, there is greater modeling flexibility than for the homogeneous Poisson process, and this model is used a lot when modeling overdispersed count data.

We do not discuss the mixed Poisson process further, but we refer to the reader to Grandell [109] for a thorough treatment. Further references in the actuarial field are Klugman et al. [129, Ch. 7.2], Denuit et al. [87, Ch. 1.4] and Albrecher et al. [7, Ch. 5.2.3], where other particular cases relevant for actuarial practice are described.

1.1.4 Doubly-stochastic Poisson process

A richer class of Poisson processes can be constructed by randomizing the entire mean-value function of an inhomogeneous Poisson process. In that regard, Cox [57] introduced doubly-stochastic Poisson processes, which can be defined as follows:

Definition 1.1.6. *Let $\{\Lambda(t); t \geq 0\}$ a stochastic process with $\Lambda(0) = 0$ a.s., $\Lambda(t) < \infty$ for each $t < \infty$ and non-decreasing càdlàg paths. Independently, let $\{\tilde{N}_1(t); t \geq 0\}$ be a standard homogeneous Poisson process. Then the counting process*

$$N_\Lambda(t) = \{\tilde{N}_1(\Lambda(t)); t \geq 0\},$$

is said to be a doubly-stochastic Poisson process, or Cox process¹.

If a non-negative stochastic process $\{\Theta(s); t \geq 0\}$ exists such that $\{\Lambda(t); t \geq 0\}$ has the representation

$$\Lambda(t) =_d \int_0^t \Theta(s) ds, \quad t \geq 0,$$

then $\Theta(s)$ is called the intensity process. As remarked in Grandell [108], in some applications it is more natural to define a Cox process by specifying $\Theta(\cdot)$ rather than $\Lambda(t)$. Moreover, the stochastic intensity function can, for example, be interpreted as variations in the environment of the modeled phenomena over time.

From Definition 1.1.5 we can see that a mixed Poisson process is a particular case of a Cox process with $\Lambda(t) \equiv \Lambda \cdot t$.

Given a sample path, say, $\lambda(t)$ of $\Lambda(t)$, for $s < t$ we get

$$\mathbb{P}(N_\Lambda(t) - N_\Lambda(s) = k \mid \Lambda(t) = \lambda(t), \Lambda(s) = \lambda(s)) = \frac{(\lambda(t) - \lambda(s))^k}{k!} \exp(-(\lambda(t) - \lambda(s))).$$

In other words, conditional on a sample path of the intensity process, a Cox process is an inhomogeneous Poisson process. Furthermore, we do not need information concerning the full path of Λ between s and t , but rather the values at the boundaries. This is different when the conditioning is in terms of the intensity processes $\{\Theta(t); t \geq 0\}$. Namely,

$$\mathbb{P}(N_\Lambda(t) - N_\Lambda(s) = k \mid \Theta(u) = \theta(u), s \leq u \leq t) = \frac{1}{k!} \left(\int_s^t \theta(u) du \right)^k \exp \left(- \int_s^t \theta(u) du \right).$$

Using the two properties above, it follows that

$$\begin{aligned} \mathbb{E}(N_\Lambda(t) - N_\Lambda(s) \mid \Theta(u) = \theta(u), s \leq u \leq t) &= \int_s^t \theta(u) du, \\ \text{Var}(N_\Lambda(t) - N_\Lambda(s) \mid \Theta(u) = \theta(u), s \leq u \leq t) &= \int_s^t \theta(u) du, \end{aligned}$$

¹To avoid confusion among definitions we reserve the \tilde{N}_Λ notation for mixed Poisson processes, while we reserve N_Λ for Cox processes. In the former case Λ denotes the mixing random variable, whereas in the latter it denotes the stochastic process $\{\Lambda(t); t \geq 0\}$.

as expected from the conditional properties of the Cox processes. Moreover, using the tower property of the conditional expectation and the law of total variance

$$\begin{aligned}\mathbb{E}(N_\Lambda(t) - N_\Lambda(s)) &= \int_s^t \mathbb{E}(\Theta(u))du, \\ \text{Var}(N_\Lambda(t) - N_\Lambda(s)) &= \int_s^t \mathbb{E}(\Theta(u))du + \text{Var}\left(\int_s^t \Theta(u)du\right),\end{aligned}$$

from which the overdispersion for Cox processes follows (cf. Cox and Isham [58]). Furthermore, Grandell [107, Ch. 1.6] showed that for $t > 0$, $h \geq 0$,

$$\text{Cov}(N_\Lambda(t+h), N_\Lambda(t)) = \mathbb{E}\left(\int_t^{t+h} \Theta(u)du\right) + \text{Cov}\left(\int_0^{t+h} \Theta(u)du, \int_0^t \Theta(u)du\right),$$

from which the autocorrelation function can be obtained.

Cox processes find several fields of application such as financial time series (Chavez-Demoulin et al. [53]), credit risk modeling (Lando [136]), risk theory (Björk and Grandell [39]; Asmussen [21]; Albrecher and Asmussen [6]), catastrophe modeling (Dassios and Jang [59]; Schmidt [164]), insurance claims reserving (Avanzi et al. [28, 27, 26]; Badescu et al. [31, 30]); reinsurance pricing (Dassios and Jang [60]) and operational risk modeling (Fung et al. [100]). In the following we describe three types of Cox processes that are particularly relevant in the actuarial literature.

Doubly-stochastic Poisson process directed by Lévy subordinators

Chapter 3 of this work will build upon Cox processes directed by a Lévy subordinator. This model is a classic tool in mathematical finance (cf. Schoutens [166], Cont and Tankov [55] and Kyprianou [133]). As a model of insurance claim counts it has first been suggested by Selch and Scherer [169].

Let us start with the definition, some basic facts, and properties about Lévy processes that are needed later.

Definition 1.1.7. *A càdlàg stochastic process $X = \{X(t); t \geq 0\}$ is called a Lévy process, if it satisfies the following properties:*

- (i) $X(0) = 0$ a.s.
- (ii) *Independent increments: for any $0 \leq t_0 < t_1 < \dots < t_n < \infty$, the increments $\Delta X(t_i) := X(t_i) - X(t_{i-1})$ are mutually independent for $i = 1, \dots, n$.*
- (iii) *Stationary increments: for any $0 \leq s < t$ and $h \geq 0$, the increments satisfy $X(t) - X(s) =_d X(t+h) - X(s+h)$.*
- (iv) *Stochastic continuity: for all $t \geq 0$ and $\varepsilon > 0$, one has $\lim_{s \rightarrow t} \mathbb{P}(|X(t) - X(s)| > \varepsilon) = 0$.*

General references on mathematical details of Lévy processes are for example in Sato [162], Bertoin [37], Applebaum [20], Kyprianou [133]. Property iii) allows one to define a new version of the process $X(t)$ with càdlàg sample paths (cf. Sato [162, Ch. 2]). Therefore, one can assume the càdlàg property. One prime example of a Lévy process is the Brownian motion, which in addition to properties ii) and iii) has normally distributed increments. However, here we focus our attention on increasing Lévy processes, i.e., with almost surely non-decreasing paths (like the homogeneous Poisson process and the compound Poisson process). These processes are also called subordinators because they can be used as time changes for other Lévy process. In this thesis Lévy subordinators are denoted by $L(t)$. A formal definition of Cox process directed by a Lévy subordinator follows.

Definition 1.1.8. *Let $\{\tilde{N}_\lambda(t); t \geq 0\}$ be a homogeneous Poisson process with intensity $\lambda > 0$ and $\{L(t); t \geq 0\}$ a Lévy subordinator. The counting process $N_L(t) := \{\tilde{N}_\lambda(L(t)); t \geq 0\}$ is a Cox process directed by a Lévy subordinator.*

The advantage of this counting process is that the Lévy subordinator acts as a random operational time, also referred to as a stochastic clock. Then, the random time jumps produced by the subordinator $L(t)$ extend the basic Poisson model to allow for simultaneous claim arrivals in physical time. However, due to the subordinator jumps, the paths are not continuous and hence no random intensity function $\Theta(\cdot)$ exists for such processes.

Classical examples of subordinators are the homogeneous Poisson process, the Gamma process and the inverse Gaussian process (cf. [169, Ch. 2]). In particular, when $L(t)$ is a Gamma process, one gets that the increments of $N_L(t)$ are independent and negative binomially distributed, which is a popular assumption for claim counts in insurance practice. This shows how the Gamma process embeds time into the negative binomial count model in a natural way.

One advantage of Cox processes, and in particular Lévy subordinated ones, is that they lead to natural multivariate models, where the dependence between different lines of business is introduced by sharing the same realization of the subordinator for otherwise independent Poisson processes (cf. Scherer and Selch [169] for details).

To that end, let $\{N_j(t); t \geq 0\}$ count the claims arriving in each line of business (or portfolio) $j = 1, \dots, d$ up to time $t \geq 0$. The multivariate Cox process $\mathbf{N} = (N_1, \dots, N_d)$ is defined through

$$\begin{aligned} \mathbf{N} &= \widetilde{\mathbf{N}}_\lambda(L(t)) \\ &= \left\{ (\tilde{N}_{1,\lambda_1}(L(t)), \dots, \tilde{N}_{d,\lambda_d}(L(t))); t \geq 0 \right\}. \end{aligned}$$

Note that we allow for d different marginal intensities λ_j , while the common $L(t)$ introduces dependence between the d lines of business. In particular, this construction can now lead to simultaneous claim arrivals within and between the individual portfolios in the time-changed process. Selch and Scherer [169, Sec. 4.2.3.] provide other alternatives to introduce dependency between components, one of them being

constructing multivariate Lévy subordinators from univariate subordinators using Lévy copulas.

We refer the reader to the excellent monograph by Selch and Scherer [169] for extensive details about Cox processes directed by Lévy subordinators. In order to allow for non-stationarity, but still maintaining the analytical tractability of the model, in Chapter 2 of this thesis we replace the Lévy subordinator $L(t)$ by an additive process $M(t)$ (see for instance Sato [162] and Cont et al. [55]). That allows to consider simultaneous claim arrivals and to incorporate other features such as trends and seasonal behavior in the underlying subordinator.

Cox processes with shot noise intensity

Let the intensity process be of the form

$$\Theta(t) = \kappa + \nu(t) + \sum_{j=1}^{\tilde{N}_\rho(t)} h(t - T_j, Y_j), \quad t \geq 0,$$

where $\kappa > 0$ is assumed to be constant; $\nu(t) \geq 0$ represents initial conditions of the process, independent of $\sum_{j=1}^{\infty} h(t - T_j, Y_j)$; and T_1, T_2, \dots represent the arrival times of the shots resulting from a homogeneous Poisson process $\tilde{N}_\rho(t)$ with intensity ρ . Furthermore Y_1, Y_2, \dots is a sequence of i.i.d. positive random variables independent of the Poisson process, and h is some decay function with $h(t, y) = 0$ for $t < 0$. Dassios and Jang [59] give the following interpretation: In addition to the occurrence of claims described by the homogeneous Poisson process with rate κ , there are additional claims independently triggered by external events (e.g., natural catastrophes) occurring at times T_1, T_2, \dots . The model captures the effect that these events lead to a significant increase of the number of claims. Due to reporting lags of the claims that originate from such external events, the resulting increase in intensity will develop according to the function h .

One possible parametric specification of the function h that has been used in the literature is $h(t, y) = y \cdot \exp(-rt)$ (see Avanzi et al. [28]), leading to

$$\Theta(t) = \Theta(0)e^{-rt} + \sum_{j=1}^{\tilde{N}_\rho(t)} Y_j e^{-r(t-T_j)}.$$

In this setup, the intensity jumps at times T_j with magnitudes Y_j and the effects decay exponentially over time at rate r , and the resulting claim number process in this case stays in fact Markovian. Liu [140] provides a nice overview of shot noise Cox processes. Moreover, Avanzi et al. [27] provide a multivariate extension of this concept, where the dependency structure is introduced via a multivariate shot noise intensity process with the help of a Lévy copula. See also Jang and Oh [118] for a recent review of the distributional properties of Cox, Hawkes and dynamic contagion processes and their compound processes.

Markov-modulated Poisson process

This type of process belongs to the general family of Hidden Markov models. In this case, the intensity process $\Theta(t)$ is controlled by a continuous-time Markov chain $X(t)$, which is typically assumed to have a finite state space. This modeling framework has applications in a variety of fields. In queuing theory, the intensity of service may fluctuate due to random breakdowns in components, in credit risk modeling the intensity of defaults may fluctuate due to changes in business cycles, or in rainfall data environmental weather conditions that generate different levels of precipitation. Intuitively, the “hidden” states of the underlying Markov model provide a representation for unobservable circumstances allowing the model to transit through different intensity levels.

We write the intensity as $\Lambda(t) = \int_0^t g(X(s))ds$, where the function $g(\cdot)$ determines the intensity structure depending on the continuous-time Markov chain $X(t)$. For example, Avanzi et al. [26] considered $g = \lambda_{X(t)} \cdot \gamma(t)$, where $\lambda_{X(t)}$ is a constant intensity conditional on the state of the underlying Markov chain at time t , and $\gamma(t)$ represents the known exogenous volume or exposure process. See also Guillou et al. [110, 111] and Badescu et al. [31, 30] for further insurance applications and additional variants of this model.

1.2 Claim severity modeling

Due to the nature of the insured losses the focus in claim severity modeling is to consider risks using positive skew random variables such as the exponential distribution or its generalization the gamma distribution. Often particular attention has to be given to events that are rare in occurrence, but can have tremendous (financial) consequences. The resulting probability distributions have “heavier” tails than an exponentially decaying tail. Given their rare nature, not many observations of such events might be available and hence there is a need for modeling techniques that can differ substantially from classic statistical methods.

Extreme value theory (EVT) provides a solid theoretical basis and statistical framework to deal with such extreme events and leads to estimators for relevant quantities in insurance practice, such as the Value-at-Risk. In this section, we briefly summarize some facts from extreme value analysis (EVA) with the aim of defining the Pareto-type models needed for Chapter 3 of this thesis. Standard references in the field are Embrechts et al. [94], Coles [54], Beirlant et al. [35], de Haan and Ferreira [64] and Reiss and Thomas [158].

1.2.1 Extreme value analysis

The prime interest in extreme value theory is the study of rare/extreme events, and to find general rules and connections that apply not only to one specific model but rather to entire classes of models. A fundamental result in EVT is the one describing

the limit behavior of the normalized partial maxima of i.i.d. random variables. More specifically, to identify a non-degenerate distribution function G for which there exist sequences of constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that for all real x

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) \rightarrow G(x), \quad (1.1)$$

as $n \rightarrow \infty$. This is provided by the so-called Fisher-Tippett theorem (cf. Fisher and Tippett [96], Gnedenko [104] and later de Haan [63]).

Theorem 1.2.1. (cf. [94, Theorem 3.2.3]) *Let X_1, X_2, \dots be a sequence of i.i.d. random variables. If there exists some non-degenerate distribution function G and some constants $a_n > 0, b_n \in \mathbb{R}$ such that $a_n^{-1}(X_{n,n} - b_n) \xrightarrow{d} G$, then G has to be of one of the following types:*

$$\begin{aligned} \text{Weibull: } \Psi_\alpha(x) &= \begin{cases} \exp(-(-x)^\alpha), & x \leq 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0. \\ \text{Gumbel: } \Lambda(x) &= \exp(-e^{-x}), \quad x \in \mathbb{R}. \\ \text{Fréchet-Pareto: } \Phi_\alpha(x) &= \begin{cases} 0, & x \leq 0 \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} \quad \alpha > 0. \end{aligned}$$

In other words, when the sample maxima can be stabilized by suitable constants, the corresponding normalized variable converges in distribution to a random variable having a distribution function of one of the three families above. This theorem can be restated in terms of the so-called generalized extreme value distribution (cf. von Mises [180] and Jenkinson [119]), by combining these three parametric families into a single family of models having a distribution function of the form

$$G_\gamma(x) = \exp(-(1 + \gamma x)^{-1/\gamma}) \quad \text{for } 1 + \gamma x > 0,$$

and $\gamma \in \mathbb{R}$. The real parameter γ is called the extreme value index (EVI), which is the key quantity in extreme value analysis. Furthermore, if there exist sequences $a_n > 0$ and b_n such that the rescaled sample maxima converge in distribution to G_γ , we say that the distribution F of the random sample X_1, X_2, \dots , is in the max-domain of attraction of G_γ (MDA(γ)).

The sign of the EVI governs the right-tail of the distribution and one distinguishes between three cases.

- The (extremal) Weibull case ($\gamma < 0$): The distributions in the max-domain of attraction are bounded above (have a finite right endpoint). Examples include the uniform, the Beta and the reverse Burr distribution.
- The Gumbel case ($\gamma = 0$): This class is rather extensive and contains a great amount of light-tailed distributions, such as the exponential and the Gamma distribution. Also, the Weibull and the lognormal distribution are contained in this class.

- The Fréchet-Pareto case ($\gamma > 0$): This class contains heavy-tailed distributions like Pareto, Fréchet and Burr distributions.

Beirlant et al. [36, Ch. 3] and Albrecher et al. [7, Ch. 3.3] provide a nice summary of distributions in each class. In Chapter 3 of this thesis, we will deal with the Fréchet-Pareto extreme value distributions which correspond to the Pareto-type distributions.

So far, we gave examples of distributions of each case. However there is still the question of how to characterize the distributions F satisfying (1.1) for a given G_γ , which constitutes the domain of attraction problem. The class of distributions in the maximum domain of attraction (MDA) can be determined in terms of the tail quantile function. For a given distribution function F , the quantile function and tail quantile function are defined as

$$Q(p) := \inf\{x : F(x) \geq p\}, \quad p \in (0, 1) \quad \text{and} \quad U(t) = Q\left(1 - \frac{1}{t}\right), \quad t > 1.$$

It can be shown that if F is a distribution function with tail quantile function $U(t)$, the distribution F belongs to the $\text{MDA}(\gamma)$ if there is a positive function a such that for $x > 0$,

$$\lim_{x \rightarrow \infty} \frac{U(xu) - U(x)}{a(x)} =: h_\gamma(u) = \frac{u^\gamma - 1}{\gamma}, \quad (1.2)$$

where for $\gamma = 0$ the expression on the right reads $\log(u)$. Moreover, the weak convergence (1.1) holds with $b_n = U(n)$ and $a_n = a(n)$ (cf. [64, Theorem 1.1.6]). If (1.2) holds, then it is often said in the literature that the underlying distribution F satisfies the extreme value condition $\mathcal{C}_\gamma(a)$ with the auxiliary function a .

In the specific case $\gamma > 0$, i.e. for the Pareto-Fréchet domain of attraction, the extreme value condition \mathcal{C}_γ corresponds to asking for U to be regularly varying, which means $U(x) = x^\gamma \ell(x)$, where ℓ is a slowly varying function. The latter is defined as a measurable and ultimately positive function that satisfies

$$\frac{\ell(tx)}{\ell(t)} \rightarrow 1 \quad \text{as } x \rightarrow \infty,$$

for every $t > 0$ (cf. [35, Ch. 2.3]). An extensive treatment of regularly and slowly varying functions is given in Bingham et al. [38].

In de Haan [63] it was shown that the extreme value condition \mathcal{C}_γ can be equivalently stated in terms of F , namely that there exists a function h such that

$$\frac{1 - F(t + uh(t))}{1 - F(t)} \rightarrow (1 + \gamma u)^{-\gamma^{-1}}, \quad \text{as } t \rightarrow x_+, \quad (1.3)$$

for all u for which $1 + \gamma u > 0$, where $x_+ = \sup\{x : F(x) < 1\}$. For $\gamma = 0$ the expression in the middle equals e^{-u} . Also, it holds with $h(t) = a(1/(1 - F(t)))$ (cf.

[64, Theorem 1.1.6]). In the case $\gamma > 0$, the latter condition also tells us that the class \mathcal{C}_γ equals the class of Pareto-type distributions defined by

$$1 - F(x) \sim x^{-\alpha}\ell(x), \quad (1.4)$$

as $x \rightarrow \infty$. Here \sim denotes asymptotic equivalence, i.e., the ratio between the two expressions is 1 in the limit $x \rightarrow \infty$. Furthermore, $\alpha = 1/\gamma$ and ℓ is slowly varying as above. Then $1 - F$ is regularly varying with index $-\alpha$. In the exact Pareto case $\ell(x) = 1$.

Condition (1.4) is equivalent to

$$\mathbb{P}\left(\frac{X}{t} > u \mid X > t\right) \rightarrow u^{-\alpha}, \quad (1.5)$$

as $t \rightarrow \infty$, for every $u > 1$. The latter shows that after a certain threshold t , Pareto-type models exhibit an exact Pareto behavior. The latter expression is relevant for the estimation of the tail index in a peaks-over-threshold (POT) approach.

It is clear from (1.5) that the accuracy of the right-hand side used as an approximation for a given t depends on the slowly varying function ℓ . In the literature the following sub-class of the Pareto-type distributions is often considered:

$$1 - F(x) = Ax^{-\alpha}(1 + Dx^\rho(1 + o(1))), \quad \text{as } x \rightarrow \infty,$$

where $A > 0$ is the scale parameter, $\rho < 0$ and D are the second-order shape and scale parameters. The notation $o(1)$ here means $f(x) \rightarrow 0$ as $x \rightarrow \infty$ (this class was introduced by Hall [112], see also [35, Ch. 3.3]). This assumption allows to derive specific approximations for the bias and variance of the tail-index estimators.

In Chapter 3 of this thesis, we will study the tempering of Pareto-type tails by a general Weibull distribution. This is motivated by the fact that often the power-law behavior does not extend indefinitely due to some truncation or tapering effects. In fact, in the context of truncation (see Beirlant et al. [33, 34]), tempering can be considered as an interpolation between pure Pareto tails and their truncated counterparts. We will also discuss the statistical estimation of the tail index α and other relevant parameters using a POT approach.

1.3 The Cramér-Lundberg model

In the actuarial literature, a criterion to analyze the solvency of an insurance portfolio is to consider the insurer's surplus process $C(t)$ as a function of time. Classically, one starts by decomposing $C(t)$ into three main components. First, assume that the insurer starts at time zero with a positive working capital, or initial surplus, $C(0) = u$. Then, the insurer collects premiums according to a certain premium process during the lifetime of the portfolio. Finally, the company will pay the policyholders' losses. Then, the safety of an insurance portfolio is measured by whether at some point in time the insurer may not be able to pay the claims occurred in the portfolio with the existing surplus level.

The traditional model for $C(t)$ is based on the early work of Filip Lundberg, who introduced the compound Poisson risk process in his famous doctoral thesis in 1903. Later, Harald Cramér mentioned the importance of Lundberg's ideas and incorporated them into the theory of stochastic processes. The resulting model is hence often referred to as Cramér-Lundberg model and it is one of the classical building blocks of Risk Theory.

The study of ruin probabilities has attracted enormous attention in the literature on applied probability theory, see for example Bühlmann [48], Seal [167, 168], Gerber [102], Grandell [108], Daykin et al. [61], De Vylder [68], Dickson [88], Rolski et al. [161] and Asmussen and Albrecher [22] for surveys. In the following section we cite some results that are standard in the ruin theory literature, and that are relevant for subsequent chapters of this thesis.

In the Cramér-Lundberg model it is assumed that the premium income is continuous and that in any time interval it is proportional to the length of this interval. The aggregate claims follow a compound Poisson process $S(t)$ with Poisson parameter λ . The surplus process of the insurance portfolio is then given by

$$C(t) = u + ct - S(t), \quad t \geq 0,$$

with initial surplus $C(0) = u$. Here c is the premium income per unit of time, and the claim sizes modeled by positive i.i.d. random variables with cumulative distribution function $F_X(x)$ and finite mean $\mu > 0$.

In reality the mean number of claims in a time interval will not be constant throughout time. In fact, it might be periodic over time or more generally vary according to a function $\lambda(t)$. Moreover, both the number of claims and premium income will also depend on the volume of the portfolio at time t , say $v(t)$. In other words, the claim number process is an inhomogeneous Poisson process with mean value function $\mu(t) = \int_0^t \lambda(s)v(s)ds$ with inverse function $\mu^{-1}(t)$. Then, from Section 1.1.2 we know that $\tilde{N}(\mu^{-1}(t))$ is a standard homogeneous Poisson process. Also, let the premium income vary with time such that the premium income in the interval $(0, t]$ is $c\mu(t)$ for some constant c . Then, consider the process $\tilde{C}(t) = C(\mu^{-1}(t))$

$$\tilde{C}(t) = u + c\mu(\mu^{-1}(t)) - \sum_{i=1}^{\tilde{N}(\mu^{-1}(t))} X_i = u + ct - \sum_{i=1}^{\tilde{N}_1(t)} X_i,$$

which is also a Cramér-Lundberg process. Therefore, we should not consider time to be real time but operational time.

It is important for an insurance company that $C(t)$ stays above a certain level given by internal or legal requirements. By adjusting the initial surplus it is no loss of generality to consider the level 0, and ruin is defined as the event that $C(t)$ drops below zero for the first time, i.e., when the total of incurred claims are larger than the initial surplus u plus the earned premiums. Define the ruin time

$$\tau(u) = \inf\{t > 0 : C(t) < 0\}.$$

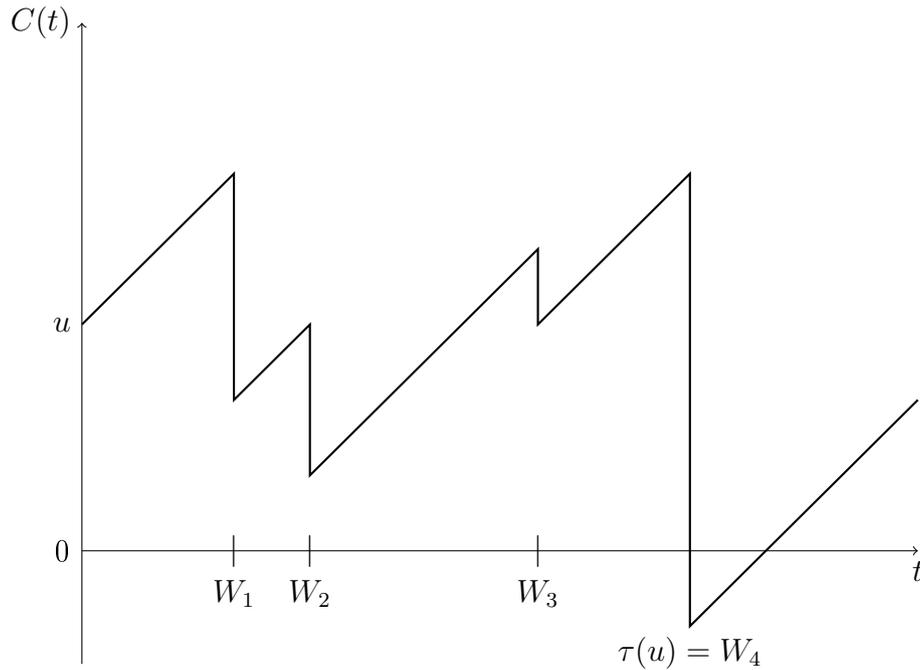


Figure 1.1: Sample path of the risk process $C(t)$ with initial surplus u . Ruin occurs at $\tau(u) = W_4$.

The probability of (ultimate) ruin is then defined as

$$\psi(u) = \mathbb{P}\left(\inf_{t>0} C(t) < 0\right) = \mathbb{P}(\tau(u) < \infty), \quad (1.6)$$

and $\phi(u) = 1 - \psi(u)$ is the survival probability. Note that $\tau(u)$ is a defective random variable. It is often more convenient to work with the loss process $R(t) = S(t) - ct$, for $t \geq 0$ and its maximum $L = \sup_{t \geq 0} R(t)$. The ruin probability can alternatively be written as $\psi(u) = \mathbb{P}(L > u)$. Denote the claim times by W_1, W_2, \dots with the convention that $W_0 = 0$. Figure 1.1 shows a sample path of the Cramér-Lundberg surplus process that is ruined at the fourth claim arrival.

In order to avoid $\psi(u) = 1$ for all u , one requires that the surplus process satisfies the net profit condition $c - \lambda\mu > 0$ (see, for example Asmussen and Albrecher [22, Ch. 4.1, Corollary 1.4]). In other words, the premium received per time unit need to exceed the expected claim payments per time unit. A further related basic quantity is the safety loading

$$\theta = \frac{c - \lambda\mu}{\lambda\mu} > 0$$

defined as the relative amount by which the premium income c exceeds the average amount of claim per unit of time.

1.3.1 An integro-differential equation

By taking advantage of the Markovian structure of the surplus process one can obtain a renewal equation for the survival probability. Conditioning on the time of

the first claim arrival, one obtains, noting that $\phi(u) = 0$ for $u < 0$,

$$\phi(u) = e^{-\lambda h} \phi(u + ch) + \int_0^h \lambda e^{-\lambda t} \int_0^{u+ct} \phi(u + ct - x) dF_X(x) dt.$$

Intuitively, the first term on the right-side of the equation describes the case where no claim occurs on $(0, h]$: due to the Markov property of the process, a new Cramér-Lundberg process then starts at the higher surplus $u + ch$. The second term describes the situation where a claim indeed occurs at some point $t \in (0, h]$, so that one is either ruined if the claim amount exceeds the available surplus $u + ct$ at time t or one restarts the evaluation at a new surplus $u + ct - x$.

It is possible to show that $\phi(u)$ is continuous and differentiable everywhere on $[0, \infty)$ with the exception of the countable set where $F_X(x)$ is not continuous (see Rolski et al. [161, Ch. 5.3]). Using these results, at the continuity points of $F_X(x)$ one can therefore write

$$\phi'(u) = \frac{\lambda}{c} \phi(u) - \frac{\lambda}{c} \int_0^u \phi(u - x) dF_X(x), \quad (1.7)$$

along with the boundary condition $\lim_{u \rightarrow \infty} \phi(u) = 1$. The latter is due to the positive drift of the process imposed by the net profit condition. Integrating both sides over the interval $(0, u]$ and doing the corresponding algebraic manipulations one obtains an integral equation for the ruin probability (see [161, Theorem 5.3.2]). Furthermore, given the boundary condition it follows that

$$\psi(0) = \frac{\lambda \mu}{c} = \frac{1}{1 + \theta}.$$

This result is quite remarkable: for zero initial surplus the ruin probability depends on the claim size distribution only through its mean and is independent of the specific form of the claim size distribution $F_X(x)$. Furthermore, one sees that it only depends on the safety loading θ .

1.3.2 The Laplace transform and the Pollaczek-Khintchine formula

Consider the Laplace transform of the survival probability defined as

$$\widehat{\phi}(s) = \int_0^\infty e^{-su} \phi(u) du, \quad \text{for } s > 0.$$

Multiplying (1.7) by e^{-su} and then integrating, one obtains

$$c \int_0^\infty e^{-su} \phi'(u) du - \lambda \int_0^\infty e^{-su} \phi(u) du + \lambda \int_0^\infty \int_0^u e^{-su} \phi(u - x) dF_X(x) du = 0,$$

for $s > 0$. By rearranging the third term of the integral and noting that the convolution of two functions is the product of their Laplace transforms, it follows that

$$\int_0^\infty \int_0^u e^{-su} f_X(x) \phi(u - x) dx du = \widehat{\phi}(s) \cdot M_X(-s),$$

where $M_X(-s)$ is the Laplace transform of the claim size distribution.

Thus one obtains the equation

$$c(s\widehat{\phi}(s) - \phi(0)) - \lambda\widehat{\phi}(s) + \lambda\widehat{\phi}(s)M_X(-s) = 0,$$

which has solution

$$\widehat{\phi}(s) = \frac{c\phi(0)}{cs - \lambda(1 - M_X(-s))} = \frac{c - \lambda\mu}{cs - \lambda(1 - M_X(-s))} = \frac{1}{s} \frac{\frac{\theta}{1+\theta}}{1 - \frac{1}{1+\theta} \left(\frac{1 - M_X(-s)}{s\mu} \right)}. \quad (1.8)$$

The Laplace transform of the ruin probability can be easily expressed as

$$\widehat{\psi}(s) = \int_0^\infty \psi(u)e^{-su} du = \frac{1}{s} - \widehat{\phi}(s).$$

From (1.8) one easily obtains a representation formula for $\psi(u)$ (the celebrated Pollaczek-Khinchine formula, cf. [161, Theorem 5.3.4])

$$\psi(u) = \mathbb{P}(L > u) = \frac{\theta}{1 + \theta} \sum_{k=0}^{\infty} \left(\frac{1}{1 + \theta} \right)^k \overline{F}_{X_I}^{*k}(u), \quad (1.9)$$

where $\overline{F}_{X_I}^{*n}$ denotes the n -th fold convolution of the survival distribution of the integrated tail distribution of X given by $F_{X_I}(x) = \mu^{-1} \int_0^x (1 - F_X(x)) dx$, $x \geq 0$, and Laplace transform $M_{X_I}(-s) = \frac{1 - M_X(-s)}{s\mu}$.

Thus, expression (1.9) represents L as a geometric compound random variable (cf. [22, Ch. III.5]) and may be written as $L = \sum_{k=1}^M L_i$, with M being the number of record highs and L_i ($i = 1, 2, \dots$) the respective overshoots. Moreover, M has a geometric distribution with parameter $\psi(0) = 1/(1 + \theta)$ and $F_{L_i}(x) = F_{X_I}(x)$ (cf. [22, Ch. IV.3a, Corollary 3.1]). Figure 1.2 illustrates the latter on the claim surplus process.

1.3.3 Extensions of the classical risk model

Over the years many extensions of the Cramér-Lundberg model have been proposed, and we briefly comment on some of those (we refer the enthusiastic reader to Asmussen and Albrecher [22] for a detailed overview).

In order to generalize the Cramér-Lundberg model, modern risk theory now often uses more general Lévy risk processes. One motivation is to directly model the aggregate claims by infinitely divisible distributions rather than compound the single claims as in the Cramér-Lundberg model. Early references of this approach are Dufresne et al. [91] where the compound Poisson process is replaced by a Gamma process, Dufresne and Gerber [90] who added a diffusion component or Morales and Schoutens [148]. This direction of research has seen a tremendous growth in recent years, see Asmussen and Albrecher [22, Ch. 11] and Kyprianou [133, 134] for details.

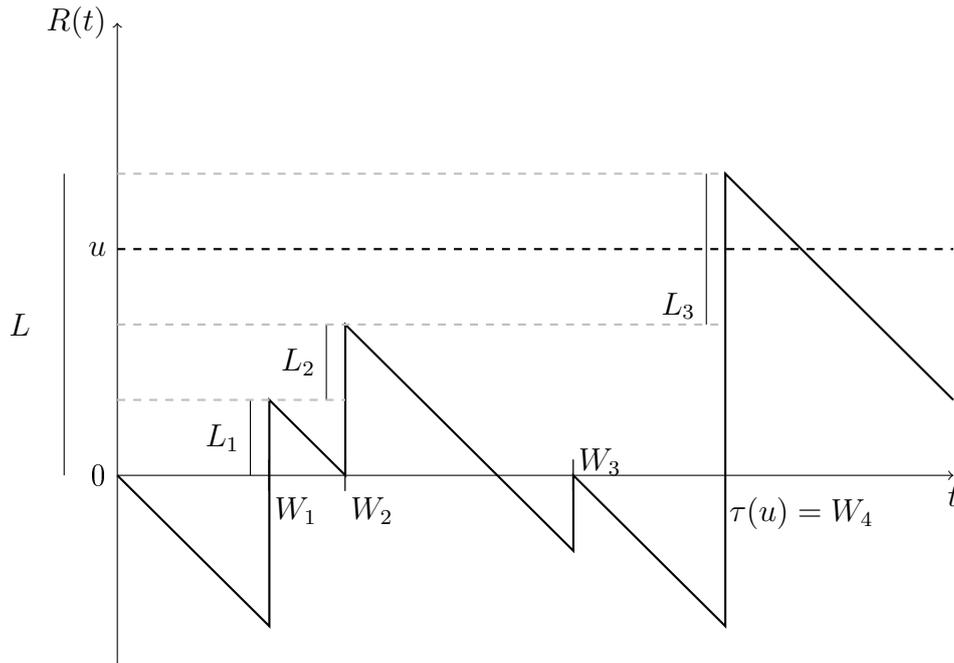


Figure 1.2: Sample path of the risk process $R(t)$ with initial surplus u . Ruin occurs at $\tau(u) = W_4$. Also, the quantities L, L_1, L_2 and L_3 are depicted.

Since de Finetti's work [62], it has also been of interest to consider decision rules under which the insurance company pays dividends to its shareholders and more precisely to specify the optimal strategy to pay dividends over the life time of the insurance portfolio, see Gerber [101], Avanzi [25], Albrecher and Thonhauser [16] or Azcue and Muler [29] for overviews. Other extensions of the classical model include stochastic investment returns (cf. Delbaen and Haezendonck [73]), taxes (cf. Albrecher and Hipp [12]) or reinsurance (see e.g. Schmidli [163]). Similar to dividend strategies, adding such features introduces decision theoretical aspects, such as how much to dynamically invest in risky assets in a surplus-dependent way or what is the optimal reinsurance cover to purchase.

As mentioned in Section 1.1, the homogeneous Poisson model is often not a good description of reality and more realistic models are needed to model the claim arrival process. In that regard, the renewal risk model, also known as the Sparre-Anderson model (cf. [172]), replaces the exponential distribution of the time in between arrivals by a more general distribution. This extension introduces a certain degree of contagion into the model, as the lack-of-memory property of the claim arrival in the Cramér-Lundberg model is lost. More generally, as discussed in the beginning, one can introduce inhomogeneity into the process by means of an intensity function $\lambda(t)$ describing diverse fluctuations over time in order to account for effects such as seasonality, insurance cycles, etc. A natural way to achieve this is by means of an inhomogeneous Poisson model or, more generally, a Cox process. For example, Björk and Grandell [39] assume a risk process with a Cox process for counting claim arrivals. Using a martingale approach they gave extensions of the classical Lundberg inequality. Nevertheless, Cox processes are too general and therefore one needs more

concrete assumptions. For example, Albrecher and Asmussen [6] considered a claim arrival process that is a superposition of a homogeneous Poisson process and a Cox process with a Poisson shot noise intensity process. As described in Section 1.1.4, Markov-modulated Poisson processes are another appealing way to allow for inhomogeneity. Asmussen [21] considered risk processes with the property that the rate λ of the Poisson arrival process and claim size distribution F are not fixed in time but depend on the state of an underlying Markov jump process $\{J(t); t \geq 0\}$ with a finite state space such that $\lambda = \lambda_i$ and $F = F_i$ when $J(t) = i$. Thus, $J(t)$ describes the environmental conditions for the risk process, see Asmussen and Albrecher [22, Ch. VII] for a comprehensive treatment of risk theory in a Markovian environment.

When no closed-form solutions are available for the ruin probability, several numerical methods are available, for example Laplace transform inversion, matrix-analytic methods or numerical solutions of differential and integral equations (cf. [22, Ch. I.4]). Another classical approach is to use approximations for the ruin probability for large values of the initial surplus. The Cramér-Lundberg approximation $\psi(u) \sim C \cdot e^{-Ru}$, $u \rightarrow \infty$ is one of the most celebrated results of risk theory, where $C = (c - \lambda\mu)/(\lambda M'_X(R) - c)$ and $R > 0$ is the solution of the Lundberg equation $\lambda(M_X(R) - 1) - cR = 0$ provided that $M_X(R)$ exists. This approximation is renowned for being quite accurate for intermediate values of u already and not only for large initial surplus. It holds for exponentially bounded claim distributions and has generalizations to some more general surplus processes. Various approximations exist for the ruin probability in the compound Poisson model (cf. Asmussen and Albrecher [22, Ch. IV.7]). In particular, if information about the claim arrivals intensity, claim size moments, and premium rate is available, one idea introduced in De Vylder [65] is the following. Given the explicit ruin probability formula for exponential claims, one can approximate $C(t)$ by a surplus process $\tilde{C}(t)$ with exponentially distributed claims by matching the observed available information and the theoretical moments under this assumption. That approximation has proved to be quite accurate for light-tail claim distributions. Extensions of the latter to approximations with more general claim distributions can be found, for example, in Badescu and Standford [32]. Moreover, Gerber et al. [103] study an extension of this approximation to study the effects of external mechanisms such as dividend payments. If on the other hand $F_{X_I}(x)$ is sub-exponential, one has the so-called large claim approximation $\psi(u) \sim \frac{\lambda\mu}{c-\lambda\mu}(1 - F_{X_I}(u))$. Some of the early references on such approximations are Teugels and Veraverbeke [177], von Bahr [179] and Embrechts and Veraverbeke [95]. Stronger results under additional assumptions on the tail of the distribution can for instance be found in Ramsay [156] and Albrecher and Kortschak [15] for Pareto claim sizes, and Burr claim sizes are studied in Kortschak and Albrecher [132]. In Albrecher et al. [13] higher-order expansions for ruin probabilities in renewal risk models with sub-exponential claim sizes are obtained. Moreover, Kalashnikov [128] provides upper bounds for the survival probability in the presence of heavy tails.

If only partial information about the surplus process is known (for example, as mentioned above, information about the claim arrival intensity or moments of the individual claim sizes), besides approximations one often tries to find bounds for

the ruin probability. In Chapter 4, we explore this situation and provide upper and lower bounds for the ruin probability when only information about the first moments of the claim size distribution is available.

1.4 Stochastic orders

A particular task of actuarial practice is to compare the attractiveness of different portfolios of risks, which often boils down to ordering of risks. One first criterion could be to compare the resulting ruin probability (or some other risk measure). Often such assignment has to be carried out with only partial knowledge about the underlying random variables. For instance, information about the mean and the variance, or maybe higher order moments, of the claim size may be available, but not a full description of the risks under examination. Hence, there is a need for stochastic ordering relations, which can lead to bounds on the risk measures under consideration. As argued in Brockett [47], if confronted with several portfolio of risks, given an initial surplus and safety loading, the actuary might opt for the one with the smallest ruin probability. However, if encountered with incomplete available information, the decision needs to be made based on lower and upper bounds for the ruin probability.

The derivation of stochastic bounds is an active area of research and has found applications in various fields. Stochastic order relations constitute an important tool in the analysis of actuarial problems. For instance, they can be used to compare complex models possibly leading to more conservative decisions. In an actuarial context, stochastic orderings were first described in the seminal papers by Borch [41], Bühlmann et al. [49] and Goovaerts et al. [105]. Classic additional references are Shaked and Shantikumar [170, 171] and Müller and Stoyan [149]; see also Kaas et al. [126] and Denuit et al. [77] for an actuarial context.

Given the natural connection between stochastic ordering and utility theory, in the actuarial literature most of the order relations have an integral form. Namely, given two random variables X and Y and a class of measurable functions \mathcal{G} , we say that X precedes Y in the $\leq_{\mathcal{G}}$ -sense ($X \leq_{\mathcal{G}} Y$) if and only if $\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y))$ for all the functions $\phi \in \mathcal{G}$. The function ϕ can be thought of as the preferences of an economic agent (its utility function). In this thesis we do not pursue interpretations of the latter kind, however we do make use of certain types of stochastic order relationships of the integral form and in this section we briefly comment on three of them that will be relevant in Chapter 4.

We say that a random variable X precedes a random variable Y in stop-loss order if the X has less weight in the tail than Y , in the sense that the expected value above a certain deductible d (stop-loss premium) is larger for Y than for X , and this is true for each possible d . A formal definition of the latter is as follows.

Definition 1.4.1. *A random variable X precedes a random variable Y in stop-loss order ($X \leq_{sl} Y$), if their stop-loss premiums are ordered uniformly:*

$$\mathbb{E}((X - d)_+) \leq \mathbb{E}((Y - d)_+),$$

for all deductibles (or retentions) $d \geq 0$, where $z_+ = \max(z, 0)$.

Stop-loss order has great appeal to actuaries given that it reflects the common risk preferences of all risk averse decision makers, i.e. economic agents with concave increasing utility functions. Furthermore, stop-loss order is preserved under many important mathematical operations relevant for actuarial purposes. For example, it is preserved under convolution, compounding, and mixing (cf. [77, Ch. 3.2.2]).

Denuit [75] studied the class of functions $\phi(x) = 1 - e^{-sx}$ where s is a positive number. This class of functions give rise to the Laplace transform order, which can be defined as follows:

Definition 1.4.2. *Given two non-negative random variables X and Y , X precedes Y in Laplace transform order ($X \leq_{Lt} Y$) if*

$$\mathbb{E}(\exp(-sX)) \leq \mathbb{E}(\exp(-sY)),$$

holds for all $s \geq 0$.

Moreover, as argued in [75], $X \leq_{Lt} Y$ can be interpreted as a preferred income X over an income Y by all decision makers with an exponential utility function. An extensive treatment of the Laplace transform order can be found in Shaked and Shantikumar [171, Ch. 5].

As mentioned before, in many situations the actuary only has access to certain moments of the distribution, hence a great amount of research has been devoted to the derivation of upper and lower bounds on quantities in the integral form $\mathbb{E}(\nu(X))$, for some function ν , and X belonging to a certain class of functions satisfying some moment constraints. Early examples of this are De Vylder [66, 67], De Vylder and Goovaerts [71], Kaas and Goovaerts [125] and Hürlimann [116] where bounds for functions of random variables under integral constraints were analyzed.

In that regard, Denuit et al. [85, 76] introduced the concept of m -convex orders, which is formally defined as follows.

Definition 1.4.3. *Let \mathcal{S} be a subinterval of the real line (in this thesis closed intervals). For a positive integer m , consider the class \mathcal{M}_{m-cx} of all functions $\phi : \mathcal{S} \rightarrow \mathbb{R}$ whose m -th derivative $\phi^{(m)}$ exists and satisfies $\phi^{(m)}(x) \geq 0$, for all $x \in \mathcal{S}$, or which are limits of sequences of functions whose m -th derivative is continuous and non-negative on \mathcal{S} .*

Let X and Y be two random variables that take on values in \mathcal{S} . Then, X is said to be smaller than Y in the m -convex order ($X \leq_{m-cx} Y$) if $\mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y))$ for all functions $\phi \in \mathcal{M}_{m-cx}$ provided the expectations exist.

For a comprehensive treatment of the theory of m -convex orders and effective methods for deriving the support points and associated probabilities of the stochastic extrema, we refer to the reader to Denuit et al. [86], Shaked and Shantikumar [171, Sec. 3.1] and the references therein. The reader interested in how to approximate theoretical premium calculation principles using m -convex extrema is referred to

Denuit [74]. Denuit et al. [84] applies decision theory using the concept of m -convex pain functions. Here pain functions are defined as $v(x) = -u(x)$ for $x \in \mathbb{R}$, with $u(\cdot)$ being a non-decreasing utility function for a decision-maker. Then, $v(x)$ might be viewed as the pain associated with a debt of amount x monetary units. For a more recent application to assess the difference in payment between an agent incurred loss and the compensation from a structured risk transfer mechanism such as index-based insurance or reinsurance contracts see Lefèvre et al. [138], where an Enterprise Risk Management framework is defined to assess these hedging imperfections. The concept of m -convex orderings for comparing arithmetic risks was developed in Denuit and Lefèvre [82] and Denuit et al. [83]. In these contributions, given a non-degenerate moment space with fixed m moments, explicit formulas for the discrete m -convex extremal distribution were derived for $m = 1, 2, 3$. Furthermore, Courtois et al. [56] extended the previous references by proposing a method for deriving expressions for general non-negative integer m .

1.5 Main contributions of this thesis

This thesis contains results in the area of modeling and assessing risk of insurance portfolios, which are structured in three chapters. The results of Chapter 2 have already been published in the *North American Actuarial Journal*, the content of Chapter 3 is accepted for publication in *ASTIN Bulletin* and Chapter 4 is submitted for publication.

Motivated by Selch and Scherer [169], in Chapter 2 we propose a non-stationary extension of the construction principle of Lévy-subordinated Poisson processes for claim frequency modeling, introduced in the before-mentioned reference, by enabling trends and seasonal behavior in the underlying subordinator. Furthermore, we argue that this will allow to keep all the advantages of the Selch-Scherer model, but in addition account for non-stationary behavior, leading to a flexible, yet parsimonious continuous-time model for claim counts that accounts for both clustering and non-stationarities. This is particularly relevant in insurance practice where claims often occur in clusters and their arrivals may depend on various external and time-dependent factors. Also, by means of a case study we show how the incorporation of such non-stationarities can accommodate characteristics of real data such as changes in policy volumes over time.

Chapter 3 deals with heavy-tail modeling. In particular, it provides a model for claim sizes where the Pareto behavior only sets in after a certain threshold and such a behavior does not extend indefinitely due to truncation or tempering effects in the data. In that regard, it generalizes earlier results of Meerschaert et al. [145] and Raschke [157] by considering the tempering of a Pareto-type distribution with a general Weibull distribution in a peaks-over-threshold approach. This provides a relevant model for insurance applications where often the Hill plot shows a sharp increase when considering a small number of order statistics only, as the data become less heavy-tailed further out in the tails. For example, claim payments are influenced by claim management and claims may for instance be subject to a higher level of

inspection at highest damage levels leading to weaker tails than apparent from modal claims.

Finally, Chapter 4 deals with ruin theory under incomplete information, namely in the circumstances where only limited information about the claim size distribution is available. This is in fact a classical problem posed many years ago by the Swiss Actuary Hans Schmitter. To revisit this classical problem from a fresh perspective, we consider random initial surplus, leading to considerably more amenable expressions for the ruin probability in the Cramér-Lundberg model. We show that the computational simplification that is obtained by randomizing the initial surplus allows to connect the problem to m -convex ordering to obtain analytical bounds for the randomized ruin probability. In addition, we show that the solution to the classical problem with deterministic initial surplus level can conveniently be approximated via Erlang(k)-distributed initial surplus for a large k .

Chapter 2

Fitting non-stationary Cox processes: An application to fire insurance data

This chapter is based on the following article:

H. Albrecher, J.C. Araujo-Acuna, and J. Beirlant. Fitting nonstationary Cox processes: An application to fire insurance data. *North American Actuarial Journal*. *To appear*, 2020.

Abstract. In insurance practice, claims often occur in clusters and their arrivals may depend on various external and time-dependent factors. In this article, we propose a statistical approach for modeling claim arrivals by considering clustered arrivals and non-stationarity simultaneously. To this end, we extend the Cox process methodology with Lévy subordinators presented in Selch and Scherer [169] relaxing the stationarity of increments assumption. A particular special case of the proposed approach is a dynamic and flexible model of negative binomially distributed claim numbers with trends and seasonal variations of the parameters. For illustration purposes, we fit the model to a fire insurance portfolio and show that it allows the modeling of cluster occurrences in a seasonal pattern while preserving overdispersion, which is frequently observed in claim count data. We illustrate its use in forecasting and Value-at-Risk and Expected Shortfall computations of the aggregate insurance risk. Finally, we provide a multivariate extension of the model, where simultaneous cluster arrivals in different components are generated by a non-stationary common subordinator.

2.1 Introduction

Modelling the occurrence of insurance claims has a long history in actuarial science (see e.g. Denuit et al. [87]). In practice one often has annual claim counts available and, correspondingly, the focus has traditionally been on fitting discrete positive random variables to these data, with Poisson and negative binomial models (reflecting over-dispersion) being particularly popular, even today. If more refined information on the time of claim occurrences is available, then one can naturally use the fine structure of the data to test and challenge such simple model assumptions, and in a number of cases it turns out that counting processes can considerably improve the understanding of the underlying mechanisms of claim occurrence and their consequences for insurance risk management.

If one has access to an even more comprehensive record of the claim history and individual characteristics of the policyholders, then the claim counts of insureds can also be analyzed using count regression techniques. Starting with the contribution of Jorgenson [121] and the development of Generalized Linear Models (GLMs) (cf. [151]), the development of models for count data has advanced at great pace. In this context, one usually starts with Poisson, or negative binomial, regression as a building block and aims to identify risk factors to predict the expected occurrence of insurance claims based on the policyholders' risk characteristics. Standard references for this type of models are Gourieroux et al. [106], Hausman et al. [113], Machado and Santos Silva [142] and comprehensive treatments are given by Winkelmann [183], Frees [97], Cameron and Trivedi [51] and most recently Denuit et al. [78]. In recent decades and with advances in technological developments, insurance companies have been collecting policyholders' information for several years, which motivated the introduction of longitudinal data models in actuarial science in order to exploit this panel data structure. For example, by including random or subject-specific effects in the structure for the mean, GLMs can be extended to Generalized Linear Mixed Models (GLMMs), and these random effects can model the correlation structure between observations on the same subject and take account of heterogeneity among subjects; cf. Frees et al. [98, 99] and Antonio and Beirlant [18] who discuss GLMMs and their applications in ratemaking. Moreover, if the actuary has strong reasons to believe that nonlinear effects of continuous covariates are present, then one can explore Generalized Additive Models (GAMs), in which continuous covariates are included into the model in a semiparametric additive predictor; see, for example, Denuit and Lang [81] and Antonio and Valdez [19] for a comprehensive survey of statistical tools for risk classification used in insurance. Furthermore, Boucher et al. [42] provide an extensive selection of panel data models based on Poisson and negative binomial distributions, which allow for time dependences between observations.

This article, however, deals with a general modeling situation without such detailed additional information but with the goal to provide a flexible stochastic process framework for the description of the counting procedure over time in a portfolio. Though the Poisson distribution for annual claim counts is naturally embedded in Poisson processes, the analogue for negative binomial claim counts is somewhat trickier. Indeed, it is well-known that randomizing the rate of a Poisson random

variable according to a gamma distribution leads to a negative binomial random variable. However, rather than the often suggested mixed Poisson process, it is a Poisson process subordinated by a gamma process that embeds the negative binomial claim count model into a corresponding model in a natural way. The concept of subordination was only introduced recently in a concise way into the actuarial literature by Selch and Scherer [169]; see also Albrecher et al. [7, Ch.V] for a detailed discussion. A particular advantage of the subordination construction is that, in contrast to the ordinary Poisson process, it allows for clusters of claim arrivals, which are often observed in the data. Another advantage is that the subordination construction leads to intuitive multivariate models, where a vector of initially independent Poisson processes becomes dependent by means of a common time transformation of each component; see e.g. Selch and Scherer [169]. However, the Lévy subordinator construction proposed in [169] cannot capture possible non-stationarity in the data, which can be an important feature in datasets.

In this article we propose a non-stationary extension of the construction principle of Lévy-subordinated Poisson processes for claim frequency modeling introduced in Selch and Scherer [169], by enabling trends and seasonal behavior in the underlying subordinator. This will preserve all of the advantages of the Selch-Scherer model, but also account for non-stationary behavior, leading to a flexible yet parsimonious continuous-time model for claim counts that accounts for both clustering and non-stationarity. In Section 2.2 we start by motivating the approach for a concrete dataset of Dutch fire insurance claims. Section 2.3 introduces the model in detail and discusses simulation aspects that are used later in the implementation. Section 2.4 presents a maximum likelihood procedure for the estimation of the model parameters from data. In Section 2.5, we then apply the model to the Dutch fire insurance dataset together with an analysis of the impact of the new model on risk measures computations of the aggregate insurance risk. Section 2.6 discusses a multivariate extension and illustrates its performance on the well-known multivariate Danish fire insurance dataset. Section 2.7 concludes.

2.2 A motivating example

Let us revisit a dataset used in Albrecher et al. [7, Sec.1.3.2] on Dutch fire insurance claims over the period 2000-2014, containing the exact date of each fire incident and the building type. We will focus here on the sub-portfolio of commercial buildings. An interesting feature of the latter are the daily cluster sizes for the aggregate process, ranging from zero to four. Figure 2.1 depicts the aggregate process and Figure 2.2 shows the cluster arrivals. The semi-annual moving averages (given in red) already suggest that non-stationarity may be present in the data.

The homogeneous Poisson process commonly acts as a starting point of any statistical analysis of a claim count process when the claim arrival data are sufficiently specific, see Mikosch [146], Rolski et al. [161], Grandell [109], or Albrecher et al. [7, Ch.V] for more details. Figure 2.1 depicts asymptotic 95% confidence intervals for a homogeneous Poisson assumption, suggesting a rejection of the latter. One can

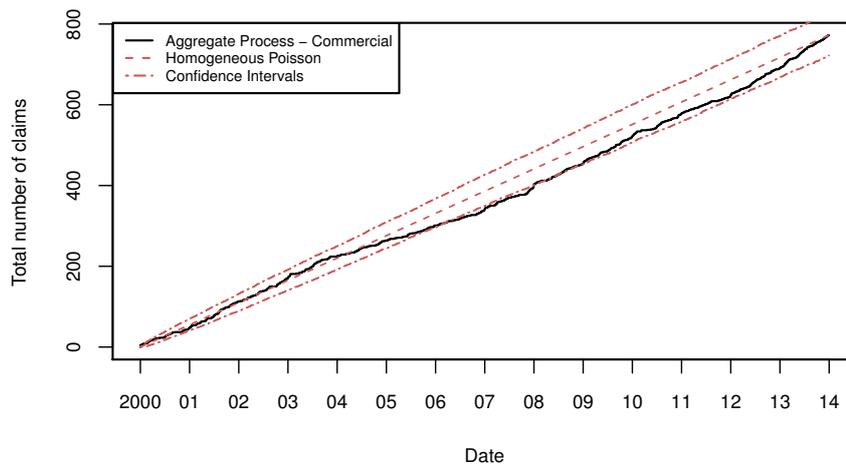


Figure 2.1: Dutch fire insurance data: aggregate claim number as a function of time together with 95% confidence intervals under a homogeneous Poisson assumption.

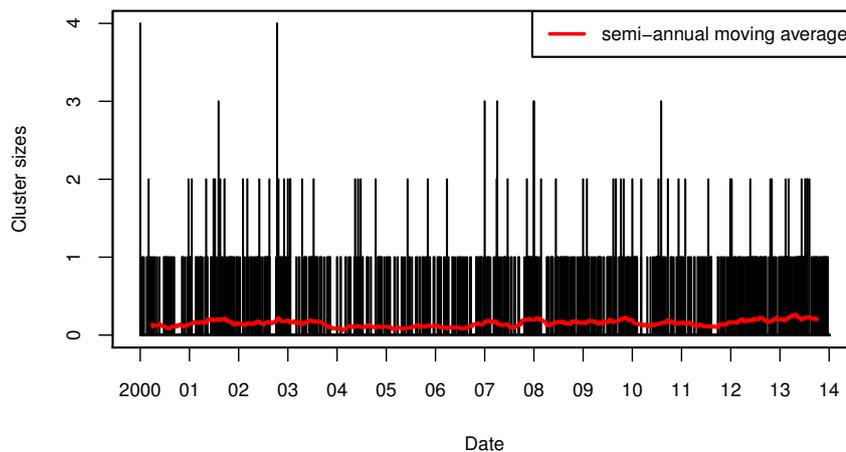


Figure 2.2: Dutch fire insurance data: daily cluster sizes and semi-annual moving average (solid line) of claim numbers against time.

perform a suitably adapted Kolmogorov-Smirnov test for uniformity for the claim arrivals over the considered time interval (see [7, p.173]), which indeed rejects a homogeneous Poisson process, with a P-value of 0.00093. See also Figure 2.3, which depicts the pairs (j, T_j) , $j = 1, \dots, n$, where T_j denotes the time of the claim arrival of the j th claim, and the resulting plot deviates considerably from linearity that would be required to fulfill the uniform distribution assumption of the arrival times under a homogeneous Poisson process.

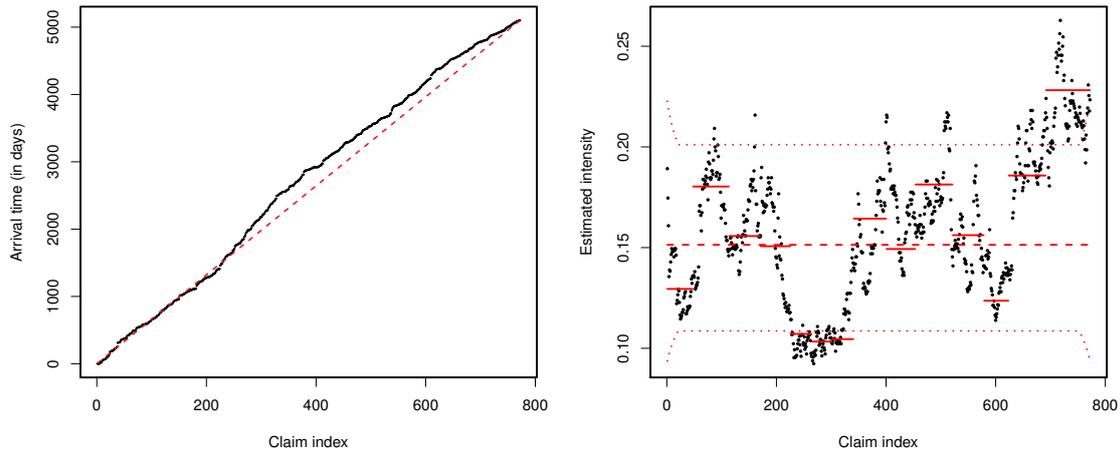


Figure 2.3: Dutch fire insurance data: time points (j, T_j) , $j = 1, \dots, n$ (left); moving average of the estimated intensity function ($h = 20$) (dots) together with the reciprocal of the annual expected inter-arrival times (piecewise constant solid line), the intensity estimate of a homogeneous Poisson process (dashed line) together with 95% confidence intervals (dotted lines) (right).

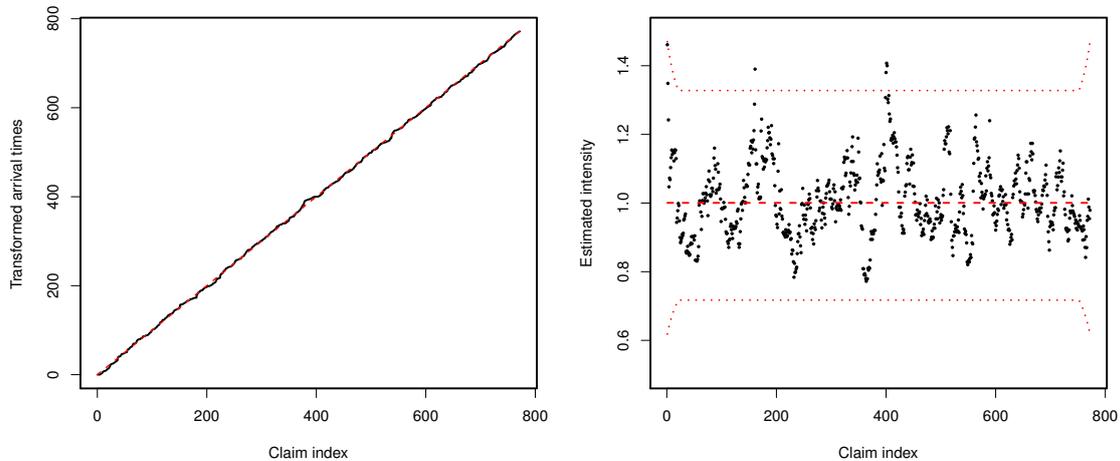


Figure 2.4: Dutch fire insurance data: transformed time points $(j, \mu(T_j))$, $j = 1, \dots, n$ (left); moving average of the estimated intensity function ($h = 50$) after time transformation together with 95% confidence intervals (dotted lines) (right).

In order to describe how the intensity λ varies over time, one can consider, for example, a moving average of the waiting times $W_j = T_j - T_{j-1}$, plotting

$$1/\hat{\lambda}_i = \frac{1}{\min(n, i+h) - \max(1, i-h) + 1} \sum_{j=\max(1, i-h)}^{\min(n, i+h)} W_j \quad (2.1)$$

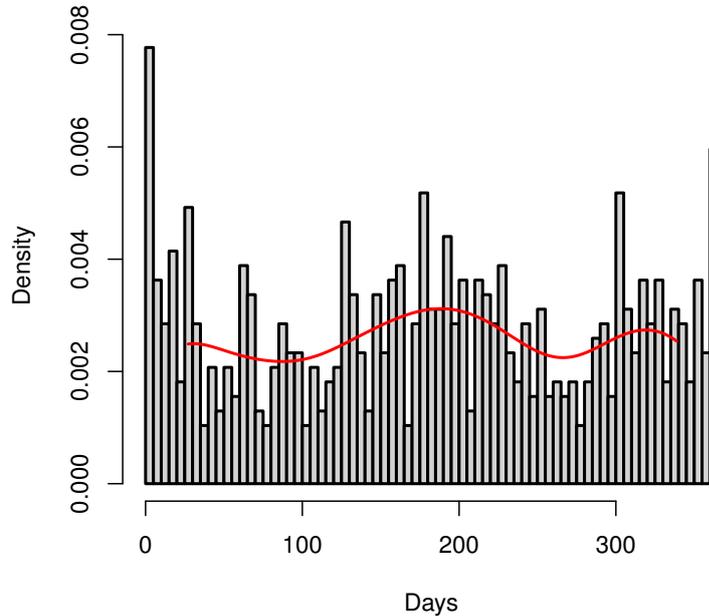


Figure 2.5: Dutch fire insurance data: histogram of the claim arrival days within the year of occurrence.

against i ($i = 1, \dots, n$) for some bandwidth $2h$ (cf. Mikosch [146]). Though for a homogeneous Poisson process no trends should be visible in such a plot, the right-hand-side in Figure 2.3 clearly indicates time variation (note that this plot slightly differs from Figure 5.6 (top left and top right) in [7, Sec.5], because we only consider the portfolio of commercial buildings).

A natural next step may then be to suspect an inhomogeneous Poisson model for the arrival of claims; see, for instance, Lu and Garrido [141], Mikosch [146, Sec.2.1.7] or Albrecher et al. [7, Sec.5.5.2]. We estimate the respective mean value function $\mu(t) = \int_0^t \lambda(s) ds$ using (2.1). If one then transforms the observation times T_j to $\mu(T_j)$ and renormalizes, under the assumption of an inhomogeneous Poisson process the new claim arrival process should again follow a homogeneous Poisson process. Figure 2.3 (right) gives the estimated yearly average intensity value (piecewise constant solid line). Time transforming according to the latter, the mean value function μ is piecewise linear with different slopes for each year. Figure 2.4 (left) shows that the resulting time-changed arrival times agree nicely with a linear pattern as required for a homogeneous Poisson process. However, the Shapiro-Wilk test for exponentiality for the time-changed waiting times rejects a Poisson model with a P-value of 0.007. Figure 2.4 (right) illustrates that the moving averages of the time-changed process still contains some significant structure. In addition, it is clear that the multiple claim arrivals per day observed in Figure 2.2 could not be reproduced by an inhomogeneous Poisson process with reasonable intensity patterns (see [7, Sec. 5.5.2] for a further discussion of this point). Finally, the histogram of claim arrivals

as a function of the day within the year of occurrence in Figure 2.5 reveals a seasonal pattern in the process. There is more activity in the middle of the year and in the period December-January, which may be attributed to the dry summer season and increased firework activity during the holiday season at the end of the year.

Motivated by this example, we hence look for a general, yet parsimonious possibility to incorporate both clustering and non-stationarity in the claim count model, which we will pursue in the sequel.

2.3 The non-stationary Cox model

2.3.1 Poisson processes directed by a Lévy subordinator

Let $\tilde{N}_\lambda(t)$ denote a homogeneous Poisson process with intensity parameter λ and let $N_L(t) := \tilde{N}_\lambda(L(t))$ be a doubly stochastic Poisson (*Cox process*) directed by a Lévy subordinator L . The Lévy subordinator L acts as a random operational time, also referred to as a *stochastic clock*. $\tilde{N}_\lambda(L(t))$ is then again a Lévy process with independent and stationary increments; cf. Cox [57] and Ammeter [17] for early references and for applications of shot noise Cox processes into insurance claim count modeling (Liu [140] and Avanzi et al. [28, 27]). The random time jumps produced by the subordinator L leads to simultaneous claim arrivals in physical time. Though subordinators are a classical modeling tool in finance (see, for instance, Schoutens [166] and Kyprianou [133]), their application for the modeling of insurance claim counts has only been suggested recently by Selch & Scherer [169].

Assume that the process N_L is observed at discrete points in time, $0 := t_0 < t_1 < \dots < t_n := T$ ($n \in \mathbb{N}$) with T being a finite time horizon. Further let $\Delta k_j := k_j - k_{j-1}$ and $\Delta L(t_i)$ the increment of the process L in the interval $(t_{i-1}, t_i]$. The joint probability mass function (pmf) of N_L at discrete time points is then simply given by

$$\begin{aligned} \mathbb{P}(N_L(t_1) = k_1, \dots, N_L(t_n) = k_n) &= \prod_{i=1}^n \mathbb{E} \left(\frac{e^{-\lambda(\Delta L(t_i))}}{\Delta k_i!} (\lambda(\Delta L(t_i)))^{\Delta k_i} \right) \\ &:= \prod_{i=1}^n \frac{(-\lambda)^{\Delta k_i}}{\Delta k_i!} \varphi_{\Delta L(t_i)}^{(\Delta k_i)}(\lambda), \end{aligned} \quad (2.2)$$

where $\varphi_{L(t)}^{(k)}(\lambda)$ is the k -th derivative of the Laplace transform of $L(t)$ (cf. [169]). If the times are equidistant with $\Delta t_i = h$, then by the property of stationary increments, the above expression can be further simplified to

$$\mathbb{P}(N_L(t_1) = k_1, \dots, N_L(t_n) = k_n) = \prod_{i=1}^n \frac{(-\lambda)^{\Delta k_i}}{\Delta k_i!} \varphi_{L(h)}^{(\Delta k_i)}(\lambda).$$

In the particular case of a gamma subordinator with density

$$f_{L(t)}(x) = \frac{\eta^{\alpha t}}{\Gamma(\alpha t)} x^{\alpha t - 1} e^{-\eta x}, \quad x > 0,$$

for given parameters $\alpha, \eta > 0$, one simply gets

$$\varphi_{L(h)}(u) = \left(1 + \frac{u}{\eta}\right)^{-\alpha h}.$$

Figure 2.6 shows one sample path of a gamma subordinator with parameters $\alpha = \lambda = 15$ and $T = 1$.

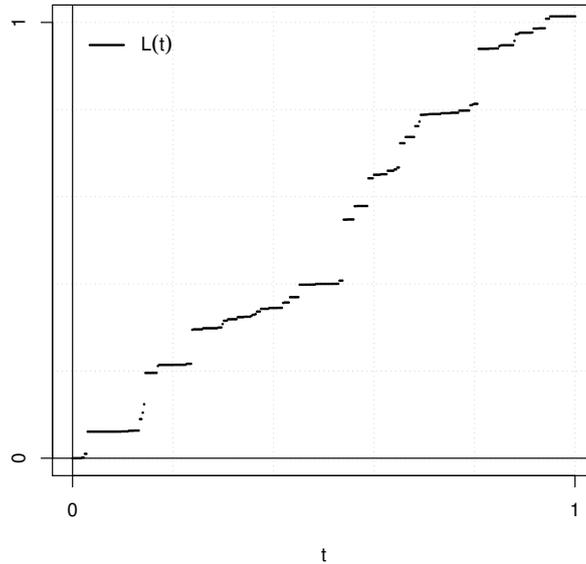


Figure 2.6: Sample path of a gamma subordinator with parameters $\alpha = \lambda = 15$ and $T = 1$.

In this case, the increments of $N_L(t)$ are independent and negative binomially distributed, which is a popular assumption for claim counts in insurance practice, with two parameters α, η and resulting overdispersion (cf. Albrecher et al. [7, Sec.5.2.4] for a more detailed discussion). As mentioned in Selch & Scherer [169, Sec. 2.1], some subordinator models lead to well known families of mixed Poisson distributions; in particular, they also exemplified the inverse Gaussian process as an alternative to model the random operational time. In the sequel, however, we focus on the gamma process and its extensions, given the prominent link with the negative binomial distribution and the wide acceptance of the latter in insurance practice.

2.3.2 Poisson processes directed by an additive process

In order to allow for non-stationarity but still maintaining the analytical tractability of the model, we now replace the Lévy subordinator L by an additive process M (see, for instance, Sato [162] and Cont et al. [55]).

A càdlàg stochastic process $M = \{M(t); t \geq 0\}$ is called an *additive process*, if it satisfies the following properties:

- (i) $M(0) = 0$ almost surely;
- (ii) Independent increments: for any $t_0 < t_1 < \dots < t_n$, the increments $\Delta M(t_i) := M(t_i) - M(t_{i-1})$ are mutually independent for $i = 1, \dots, n$;
- (iii) Stochastic continuity: for all $t \geq 0$ and $\varepsilon > 0$, one has $\lim_{s \rightarrow t} \mathbb{P}(|M(t) - M(s)| > \varepsilon) = 0$.

Define now the doubly stochastic Poisson process $N(t) := \tilde{N}_\lambda(M(t))$ directed by an additive process M . The pmf of N at some fixed time point $t \geq 0$ is

$$\mathbb{P}(N(t) = k) = \frac{(-\lambda)^k}{k!} \varphi_{M(t)}^{(k)}(\lambda), \quad k \geq 0, \quad (2.3)$$

where $\varphi_{M(t)}^{(k)}(\lambda)$ denotes the k -th derivative of the Laplace transform of M . Analogous to the Lévy case, time is now directed by the process M ; we therefore require that the directing additive process has almost surely non-decreasing paths. In this way it is guaranteed that the process can only go forward in time. Note that a Lévy process directed by an additive process is itself an additive process (for a related approach in the construction of exchangeable exogenous shock models or in a mathematical finance context, together with some theoretical properties of additive subordination, see Mai et al. [143, Sec. 3] and Li et al. [139], respectively). Correspondingly, by the independent increments property we then have that the joint pmf of N is given by

$$\mathbb{P}(N(t_1) = k_1, \dots, N(t_n) = k_n) = \prod_{i=1}^n \frac{(-\lambda)^{\Delta k_i}}{\Delta k_i!} \varphi_{\Delta M(t_i)}^{(\Delta k_i)}(\lambda), \quad (2.4)$$

for any $n \in \mathbb{N}$, $0 := t_0 \leq t_1 \leq \dots \leq t_n$, and $0 := k_0 \leq k_1 \leq \dots \leq k_n$ for $k_j \in \mathbb{N}$, $j = 1, \dots, n$.

We will now focus on the non-stationary gamma process with time-dependent parameter $\alpha(t) > 0$ and constant parameter $\eta > 0$. M then has the density

$$f_{M(t)}(x) = \frac{\eta \int_0^t \alpha(s) ds}{\Gamma(\int_0^t \alpha(s) ds)} x^{\int_0^t \alpha(s) ds - 1} e^{-\eta x}, \quad x > 0,$$

with Laplace transform

$$\varphi_{M(t)}(u) = \left(1 + \frac{u}{\eta}\right)^{-\int_0^t \alpha(s) ds}.$$

Similar to the gamma process, the time-dependent parameter $\alpha(t)$ controls the intensity of the jumps (or, essentially, the time scaling of the process), and the η parameter fixes the decay rate of big jumps; cf. Cont and Tankov [55, Sec. 4.2.2]. The resulting marginal distribution of $N(t)$ is then a negative binomial distribution with parameters $\left(\int_0^t \alpha(s) ds, \frac{\lambda}{\eta + \lambda}\right)$. The derivatives of the Laplace transform $\varphi_{M(t)}^{(k)}(\lambda)$ are given by

$$\varphi_{M(t)}^{(k)}(\lambda) = \left\{ \prod_{j=1}^k \left(\int_0^t \alpha(s) ds + j - 1 \right) \right\} \left(\frac{1}{\eta + \lambda} \right)^k \left(\frac{\eta}{\eta + \lambda} \right)^{\int_0^t \alpha(s) ds}$$

and thus

$$\mathbb{P}(N(t) = k) = \frac{(-\lambda)^k}{k!} \varphi_{M(t)}^{(k)}(\lambda) = \binom{\int_0^t \alpha(s) ds + k - 1}{k} \left(\frac{\lambda}{\eta + \lambda} \right)^k \left(\frac{\eta}{\eta + \lambda} \right)^{\int_0^t \alpha(s) ds} \quad (2.5)$$

which is indeed the negative binomial density.

Since in the sequel we will need to simulate from this process $N(t)$, we briefly describe an efficient way to do that. For a general reference on the Monte Carlo simulation of Cox processes, see for instance Korn et al. [131]. It is not difficult to adapt the general methodology to the case of an additive process: Let $\Delta_i M := M(t_i) - M(t_{i-1})$ denote the distribution of the jump size, where $t_i := iT/n$, $i = 0, \dots, n$ and $\Delta t_i = \frac{1}{n}$.

Simulation of $M(t)$

- 1: Set $M(0) := 0$.
 - 2: **for** $i = 0$ to n **do**
 - 3: Simulate a random number $\Delta_i M$ from the distribution of $M(t_i) - M(t_{i-1})$, independent of previous increments.
 - 4: Define $M(t) := M(t_{i-1}) + \Delta_i M$.
 - 5: Set $M(t) := M(t_{i-1}) \forall t \in (t_{i-1}, t_i)$.
-

Given the sample path of $\{\hat{M}(t)\}_{t \in [0, T]}$ of $M(t)$ generated above, a sample path $\{\hat{N}(t)\}_{t \in [0, T]}$ of $N(t)$ can now be simulated using the following algorithm, cf. Selch & Scherer [169, Sec. 1.2].

Simulation of $N(t)$

- 1: $i \leftarrow 0; S \leftarrow 0$
- 2: **while** $S \leq \hat{M}(T)$ **do**
- 3: Generate a random number $W_i \sim \text{Exp}(\lambda)$ independent of all previous such random numbers.
- 4: $i \leftarrow i + 1; S \leftarrow S + W_i$
- 5: Compute claim arrival times according to

$$T_i := \inf\{t > 0 : W_1 + \dots + W_i \leq \hat{M}(t)\}, \quad i \in \mathbb{N}$$

- 6: Determine the sample path $\{\hat{N}(t)\}_{t \in [0, T]}$ from the sample claim arrival times using

$$\hat{N}(t) = \sum_{i=1}^{\infty} \mathbf{1}_{\{W_1 + \dots + W_i \leq \hat{M}(t)\}}, \quad t \geq 0$$

2.4 Parameter estimation

Let us now consider the estimation of the parameters λ of the underlying Poisson process \tilde{N}_λ and the parameter vector Θ of the additive subordinator M from historical claim occurrence times using maximum likelihood on the basis of (2.4). Assume that $N(t)$ is observed at finite time instances $0 := t_0 < t_1 < \dots < t_n := T$, where T is a finite time horizon and $n \in \mathbb{N}$. Let \hat{k}_i denote the observations of the process $N(t)$ at time t_i for $i = 1, \dots, n$. Clearly, the increments $\Delta \hat{k}_i = \hat{k}_i - \hat{k}_{i-1}$, $i = 1, \dots, n$ are independent.

Based on (2.3) and (2.4), the log-likelihood function can be written as

$$\begin{aligned} \ell(\cdot; \Theta) &= \log \left\{ \prod_{i=1}^n \mathbb{P}(\Delta N(t_i) = \Delta \hat{k}_i) \right\} = \sum_{i=1}^n \log \left\{ \frac{(-\lambda)^{\Delta \hat{k}_i}}{\Delta \hat{k}_i!} \varphi_{\Delta M(t_i)}(\lambda; \Theta) \right\} \quad (2.6) \\ &= \sum_{i=1}^n \Delta \hat{k}_i \log(\lambda) - \sum_{i=1}^n \log(\Delta \hat{k}_i!) + \sum_{i=1}^n \log \left\{ (-1)^{\Delta \hat{k}_i} \varphi_{\Delta M(t_i)}(\lambda; \Theta) \right\}. \end{aligned}$$

One observes that (2.6) depends on the derivatives of order $\Delta \hat{k}_i$ of the Laplace transform of $M(t)$. Whenever the latter are available, one can take the derivative of ℓ with respect to each parameter to obtain the corresponding estimators. The parameters of $N(t)$ include λ of the Poisson process and the parameters $(\alpha(s), \eta)$ of the gamma subordinator M . Note that the final number of parameters depends on the specification of the intensity function $\alpha(s)$, and the particular shape will depend on the situation at hand.

For the fire insurance application of this paper, in view of Figure 2.5 we will consider a trend function $g(t) : [0, \infty) \rightarrow [0, \infty)$ as well as seasonal terms around the mean; that is,

$$\alpha(t) = g(t) + \sum_{k=1}^K \gamma_k \cos(2\pi\omega_k t + \phi_k) \quad (2.7)$$

where $\gamma_k > 0$, $\omega_k > 0$ and $0 < \phi_k < 2\pi$.

Naturally, one should be cautious for over-parametrization (e.g., with the above specification, our model has $2K + 3$ parameters). For the concrete estimation procedure, the constraints $\lambda > 0$, $\eta > 0$, and $\int_{t_{i-1}}^{t_i} \alpha(s) ds > 0$ for all $i = 1, \dots, n$ have to be taken into account.

Remark 1. For details on the moments of the process $N(t)$ see Selch and Scherer [169, Sec. 2.1, Theorem 2.3]. In particular, they show that the n^{th} moment of the process is given by

$$\mathbb{E}(N^n(t)) = \sum_{k=0}^n S(n, k) \mathbb{E}(M(t)^k) \cdot \lambda^k.$$

For the n^{th} central moment of the process, one then gets

$$\mathbb{E}\left(\{N(t) - \mathbb{E}(N(t))\}^n\right) = \sum_{k=0}^n \binom{n}{k} (-1)^k \left\{ \sum_{j=0}^{n-k} S(n-k, j) \mathbb{E}(M(t)^j) \lambda^j \right\} (\mathbb{E}(N(t)))^k,$$

where $S(n, k)$ are the Stirling numbers of the second kind (here the existence of higher moments of $M(t)$ is required).

When considering a gamma subordinator for $M(t)$, we get

$$\begin{aligned} \mathbb{E}(N(t)) &= \lambda \frac{\int_0^t \alpha(s) ds}{\eta}, \\ \text{Var}(N(t)) &= \lambda \frac{\int_0^t \alpha(s) ds}{\eta} + \lambda^2 \frac{\int_0^t \alpha(s) ds}{\eta^2}, \\ \mathbb{E}(\{N(t) - \mathbb{E}(N(t))\}^3) &= \lambda \frac{\int_0^t \alpha(s) ds}{\eta} + 3\lambda^2 \frac{\int_0^t \alpha(s) ds}{\eta^2} + 2\lambda^3 \frac{\int_0^t \alpha(s) ds}{\eta^3}. \end{aligned}$$

Remark 2. It is natural to use time normalization in the fitting procedure; that is, to choose η as the solution of the equality

$$\mathbb{E}(N(T)) = \lambda T.$$

We hence get $\eta = \frac{1}{T} \int_0^T \alpha(s) ds$ and the number of parameters is in fact reduced to $2K + 2$.

Remark 3. In practice, claim count data often exhibit overdispersion. As in the gamma-subordinated case of [169], we also get – for any choice of parameters – overdispersion for the non-stationary gamma-subordinated model:

$$\text{Disp}(N(t)) := \frac{\text{Var}(N(t))}{\mathbb{E}(N(t))} = \frac{\left(1 + \frac{\lambda}{\eta}\right) \left(\frac{\lambda}{\eta} \int_0^t \alpha(s) ds\right)}{\left(\frac{\lambda}{\eta} \int_0^t \alpha(s) ds\right)} = 1 + \frac{\lambda}{\eta} > 1. \quad (2.8)$$

Remark 4. Selch and Scherer [169, Sec. 4.2.2] suggest that in case of time-inhomogeneous patterns one could first (deterministically) time-change the process to remove the inhomogeneity and then proceed with the Lévy subordination for the transformed data. However, this approach would hinge on the specification of the deterministic inhomogeneity pattern and the following Cox process fit would be an independent second step. In fact, Avanzi et al. [26] followed a similar approach where first an exposure/volume component is calibrated according to the modelers domain knowledge and data analysis, which represents the known information. Afterwards, the values obtained from the exposure component serve as inputs to fit a Markov-modulated Poisson process, which will, intuitively, capture any residual temporal effect. The non-stationary Cox process approach suggested in this article offers a flexible and direct way to incorporate time-inhomogeneity, and after agreeing on the general shape of the parametric intensity function, all components in the model are estimated at the same time.

	λ	β_0	β_1	γ_1	ϕ_1	γ_2	ϕ_2	η	Disp	T
SS_1	3.00	4.0000		2.0000	1.5000			4.0000	1.7500	5
SS_2	2.00	5.0000		4.0000	2.0000	3.00	1.00	5.0000	1.4000	15
SS_3	3.00	4.0000	0.0010	2.0000	1.0000			5.0950	1.5888	3
SS_4	3.00	4.0000	0.0010	2.0000	1.0000			5.8250	1.5150	5
SS_5	0.17	0.9000	0.0005	0.7400	0.5000			2.7250	1.0697	10

Table 2.1: Simulation studies summary SS_1 to SS_5 . Disp = $1 + \lambda/\eta$.

2.4.1 Simulation study

Here we assess the performance of the maximum likelihood estimation method introduced above by means of a simulation study. Therefore, for different parameter specifications of the function $\alpha(t) = \beta_0 + \beta_1 s + \gamma_1 \cos(2\pi s/182.5 + \phi_1) + \gamma_2 \cos(2\pi s/365 + \phi_2)$ and time horizons T , we generate $m = 500$ sample paths of the claim counting process $N(t)$ using the algorithm described in Section 2.3. For each sample path, the model estimates were obtained using the MLE function described above using a gamma subordinator. The sample paths of the seasonal process are discretized with 365 steps per time unit, corresponding to daily observations of the process over the year, and the intensity functions are set such that the integrated intensity is strictly positive. Table 2.1 summarizes the parameter settings for the simulation study.

Larger estimation errors come from the simulated model with the lowest dispersion (model SS_5). This can be attributed to the fact that when simulating from models with dispersion coefficient close to one, a reasonable number of paths will show dispersion close to the unit value and therefore the MLE procedure has difficulties to fit an overdispersed model (see Remark 2), which results in several outliers. However, such drawback is not specific to the proposed non-stationary model, but arises under any Lévy setup as well. Consequently, in practice one needs to be cautious in the fitting procedure whenever the estimated dispersion is close to 1, because a possibly large value of η can be a sign of model rejection. One should validate the model using graphical methods as described in the case studies in the next sections.

Given the presence of outliers in case of SS_5 , the simulation results for the model parameters are summarized using the robust measures

$$\begin{aligned} \tilde{\lambda} &:= \widehat{\text{median}}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_m), & \widehat{\text{mad}}(\hat{\lambda}) &:= \frac{\widehat{\text{median}}(|\hat{\lambda}_1 - \tilde{\lambda}|, |\hat{\lambda}_2 - \tilde{\lambda}|, \dots, |\hat{\lambda}_m - \tilde{\lambda}|)}{\Phi^{-1}(3/4)} \\ \widehat{\text{rbias}}(\hat{\lambda}) &:= \frac{\tilde{\lambda} - \lambda}{\lambda}, & \widehat{\text{rmse}}^{\text{rob}}(\hat{\lambda}) &:= \left((\tilde{\lambda} - \lambda)^2 + \widehat{\text{mad}}(\hat{\lambda})^2 \right)^{1/2}, \end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative normal distribution function, $\widehat{\text{mad}}$ stands for median absolute deviation, $\widehat{\text{rbias}}$ for robust bias, and $\widehat{\text{rmse}}^{\text{rob}}$ for robust root mean squared error. For a detailed discussion on robust statistics, see Huber and Ronchetti

[115]. In case of SS_1 to SS_4 the results turned out to be close to the corresponding classical non-robust bias and RMSE measures.

The estimation results for the selected functions are summarized in Tables 2.2 to 2.6. It was found that the intensity parameter λ was accurately estimated, as can be seen from the given estimators: in all the cases the $\widehat{\text{mad}}$ values are small and the relative bias stays below 0.15%. However, one notices that the parameters are typically slightly underestimated. In models 1 to 4, the relative bias of the MLE estimates is predominantly smaller than 2%; therefore, the medians of the estimators are close to the true parameter values. The largest discrepancies occur for the sinusoidal phase parameters ϕ when a linear trend and a short/medium time horizon are considered. Both $\widehat{\text{mad}}$ and $\widehat{\text{rmse}}^{\text{rob}}$ show a consistent behavior across all simulated models.

	$\lambda^{SS_1} = 3.0$	$\beta_0^{SS_1} = 4.0$	$\gamma_1^{SS_1} = 2.0$	$\phi_1^{SS_1} = 1.5$
$\widehat{\text{median}}$	2.99562	4.00111	2.005741	1.47057
$\widehat{\text{mad}}$	0.053617	0.301535	0.182962	0.047360
$\widehat{\text{rbias}} (\%)$	-0.146119	0.02768	0.287013	-1.961781
$\widehat{\text{rmse}}^{\text{rob}}$	0.05380	0.30154	0.18305	0.05576

Table 2.2: Estimation results for SS_1

	$\lambda^{SS_2} = 2.0$	$\beta_0^{SS_2} = 5.0$	$\gamma_1^{SS_2} = 4.0$	$\phi_1^{SS_2} = 2.0$	$\gamma_2^{SS_2} = 3.0$	$\phi_2^{SS_2} = 1.0$
$\widehat{\text{median}}$	2.00283	5.04887	4.04427	1.96437	3.01221	0.98313
$\widehat{\text{mad}}$	0.02126	0.37337	0.30968	0.01806	0.2436	0.01637
$\widehat{\text{rbias}} (\%)$	0.14155	0.97746	1.10673	-1.78141	0.40714	-1.68657
$\widehat{\text{rmse}}^{\text{rob}}$	0.02145	0.37656	0.31283	0.03995	0.24391	0.02351

Table 2.3: Estimation results for SS_2

	$\lambda^{SS_3} = 3.0$	$\beta_0^{SS_3} = 4.0$	$\beta_1^{SS_3} = 0.001$	$\gamma_1^{SS_3} = 2.0$	$\phi_1^{SS_3} = 1.0$
$\widehat{\text{median}}$	2.99772	4.03477	0.00102	2.0072	0.96833
$\widehat{\text{mad}}$	0.06499	0.56382	0.00022	0.31371	0.07644
$\widehat{\text{rbias}} (\%)$	-0.0761	0.86913	1.69191	0.35986	-3.16739
$\widehat{\text{rmse}}^{\text{rob}}$	0.06503	0.56489	0.00022	0.31379	0.08274

Table 2.4: Estimation results for SS_3

	$\lambda^{SS_4} = 3.0$	$\beta_0^{SS_4} = 4.0$	$\beta_1^{SS_4} = 0.001$	$\gamma_1^{SS_4} = 2.0$	$\phi_1^{SS_4} = 1.0$
$\widehat{\text{median}}$	2.99671	4.04403	0.00101	2.02431	0.96806
$\widehat{\text{mad}}$	0.04793	0.48773	0.00014	0.26247	0.06522
$\widehat{\text{rbias}} (\%)$	-0.10959	1.10078	1.26307	1.21568	-3.19353
$\widehat{\text{rmse}}^{\text{rob}}$	0.04804	0.48971	0.00014	0.2636	0.07262

Table 2.5: Estimation results for SS_4

	$\lambda^{SS_5} = 0.17$	$\beta_0^{SS_5} = 0.90$	$\beta_1^{SS_5} = 0.0005$	$\gamma_1^{SS_5} = 0.74$	$\phi_1^{SS_5} = 0.50$
$\widehat{\text{median}}$	0.18959	0.92758	0.00053	0.79266	0.4544
$\widehat{\text{mad}}$	0.00731	0.39003	0.00022	0.34407	0.18411
$\widehat{\text{rbias}} (\%)$	-0.21629	3.06453	5.94046	7.11613	-9.11928
$\widehat{\text{rmse}}^{\text{rob}}$	0.00732	0.39100	0.00022	0.34807	0.18967

Table 2.6: Estimation results for SS_5

2.5 Case study: Dutch fire insurance data

We now return to the Dutch fire insurance dataset described in Section 2.2. We have seen there that both a homogeneous and inhomogeneous Poisson models did not lead to satisfactory results. We therefore proceed first to the gamma-subordinated Poisson process for a benchmark, and then will illustrate how the non-stationarities enabled by the model we proposed in this paper will lead to substantial improvements of model performance.

2.5.1 Model (M_0)

Let us first study as a reference model the stationary gamma subordinator model. This choice among other Lévy subordinators is motivated by the resulting marginal negative binomial distributions, which are very popular in market practice (for other possible choices of Lévy subordinators cf. [169]). In this case the model parameters are λ , α and η . Time normalization immediately entails $\alpha = \eta$, and solving the ML equations one obtains the parameters with log-likelihood ℓ , Akaike information criterion (AIC), and Bayesian information criterion (BIC) as given in Table 2.7.

One can construct confidence bands for each point in time of the aggregate process using Monte Carlo simulations of the process with the fitted parametric model. Simulating 1000 paths from the gamma subordinator model over the considered 14-year period and taking the 25-th and 975-th largest observations at each time t leads

$\hat{\lambda}$	$\hat{\alpha} = \hat{\eta}$	ℓ	AIC	BIC
0.151076	2.317070	-2'282.02	4'568.03	4'581.12

Table 2.7: Model M_0

to the 95% confidence band depicted in Figure 2.7. The observed process roughly fits within these bands. One observes, however, a systematic deviation around its mean. Moreover, the histogram of arrival days per year of the observed process is not reflected in simulated paths of the fitted process (cf. Figure 2.8), and one sees a need for incorporating seasonal effects.

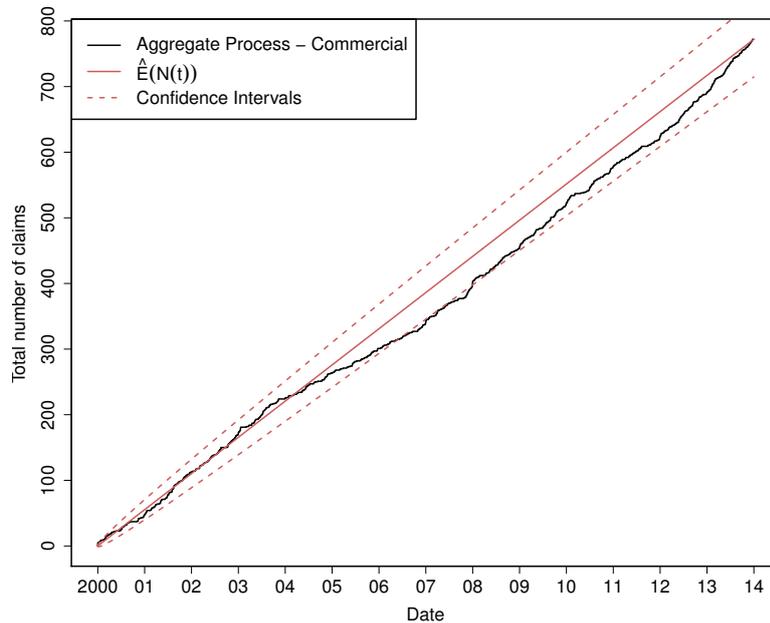


Figure 2.7: Dutch fire insurance data: observed process and simulated confidence band based on model (M_0) with constant α .

2.5.2 Non-stationary models

In order to narrow down the non-stationarity of this dataset that we would like to incorporate in the modeling, consider a moving average of the intensity function with a window of 182 days, given in Figure 2.9, together with the over-all mean of the process. The picture suggests a sudden drop in the intensity in 2003-2004, and a positive trend after that. This may indicate the presence of a change-point. In order to identify a possible candidate for the latter, we fit a regression model to the waiting times W_j (that is, the times between subsequent claim arrivals) of the form

$$W_j = (A_{01} + B_{11} \cdot j) \cdot 1_{j < x_0} + (A_{02} + B_{12} \cdot j) \cdot 1_{j \geq x_0}, \quad (2.9)$$

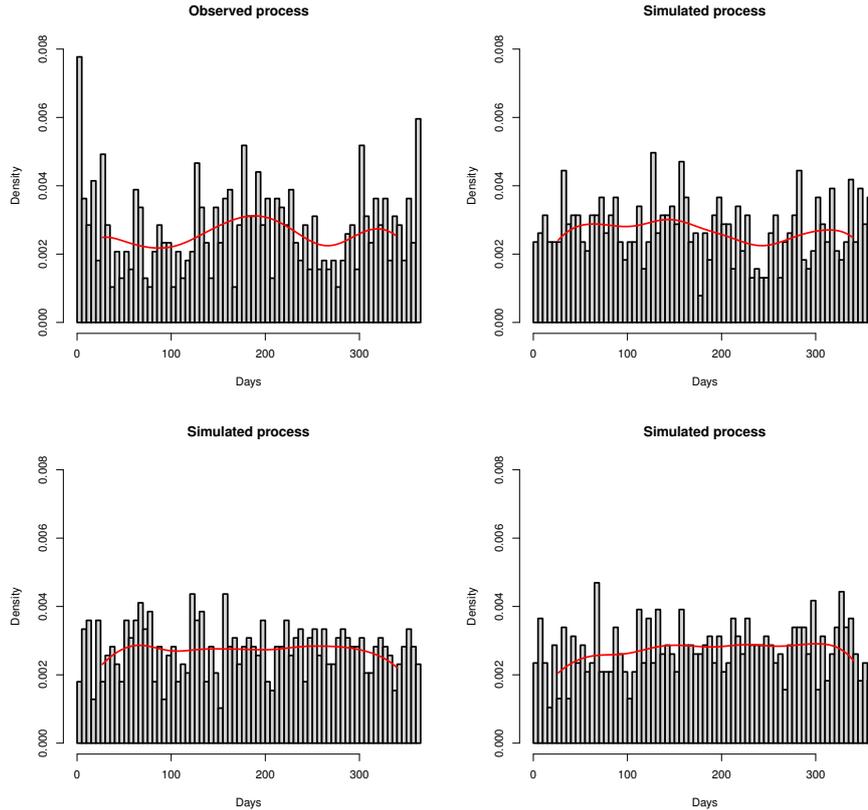


Figure 2.8: Dutch fire insurance data: histogram of the claim arrival days within the year of occurrence for the observed process (top left) and three simulated claim number processes following the fitted homogeneous gamma subordinator model (M_0) (top right and bottom).

where x_0 is the possible change-point. We then fit a model for all possible choices of x_0 and identify the one that maximizes the adjusted R-squared (see, e.g., [9] for a recent more general approach towards testing trends against change-points, and [93] for a related quickest detection formulation of possible change-points). The resulting estimated point x_0 is claim number 217, which corresponds to September 2003, in line with the visual impression from Figure 2.9.¹

The regression approach also indicates that B_{11} is not significant, so that eventually we may not include a linear trend in the first time interval. Figure 2.10 plots the resulting regression fit. In addition to trends and change-points, Figure 2.5 suggests seasonal time patterns. We allow these seasonalities to be different before and after the change-point. In order to narrow down the frequencies that we would like to consider in the model, we perform an exploratory analysis using a mean-corrected periodogram; see, e.g., Vere-Jones [178, Sec.3] for details. Figure 2.11 shows that

¹In fact, reapproaching the data provider on this matter, they confirmed that indeed due to bad fire insurance results, the Dutch insurance companies collectively decided to reunderwrite and clean up their fire portfolios at that point in time, with stricter acceptance criteria, tariff corrections and intensified attention for fire prevention, providing a causal reason for the change-point in the dynamics.

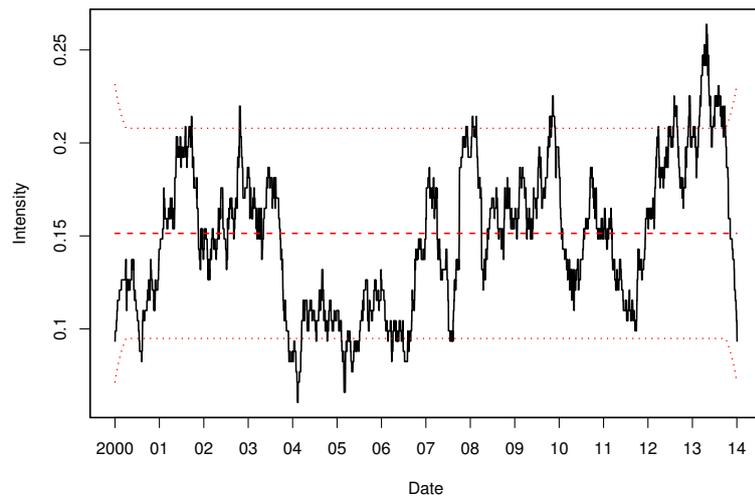


Figure 2.9: Dutch fire insurance data: moving average estimate of the intensity function using a semi-annual bandwidth, together with the intensity estimate of a homogeneous Poisson process (dashed line) together with 95% confidence intervals (dotted lines).

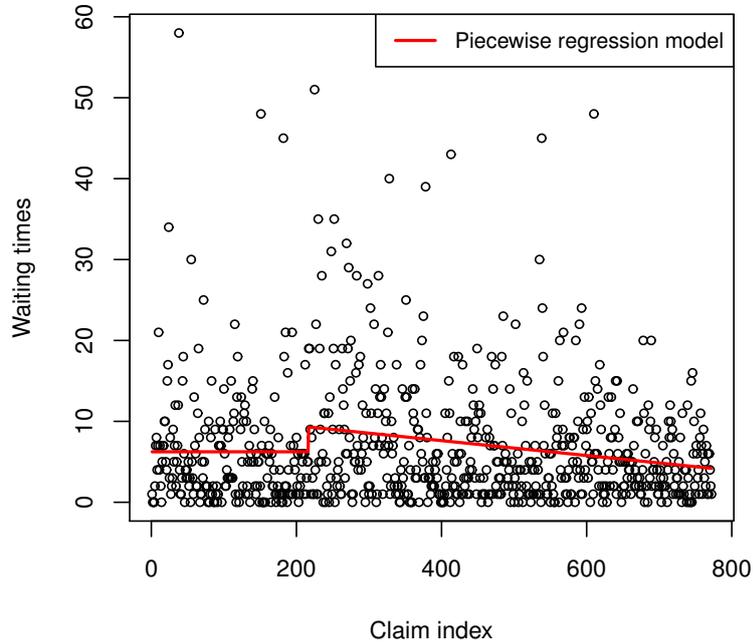


Figure 2.10: Dutch fire insurance data: change-point identification using (2.9).

the semi-annual frequency dominates the periodogram by a large margin and it is therefore reasonable to restrict the considered seasonality to such a semi-annual frequency. We hence decided to consider the following parametric family of intensity functions:

- Model ($M_{1,1}$): seasonal effect without trend:

$$\alpha_1(t) = \beta_{01} + \gamma_{11} \cos(2\pi(1/182.5)t + \phi_{11}),$$

- Model ($M_{1,2}$): seasonal effect with time trend:

$$\alpha_2(t) = \beta_{02} + \beta_{12}t + \gamma_{12} \cos(2\pi(1/182.5)t + \phi_{12}),$$

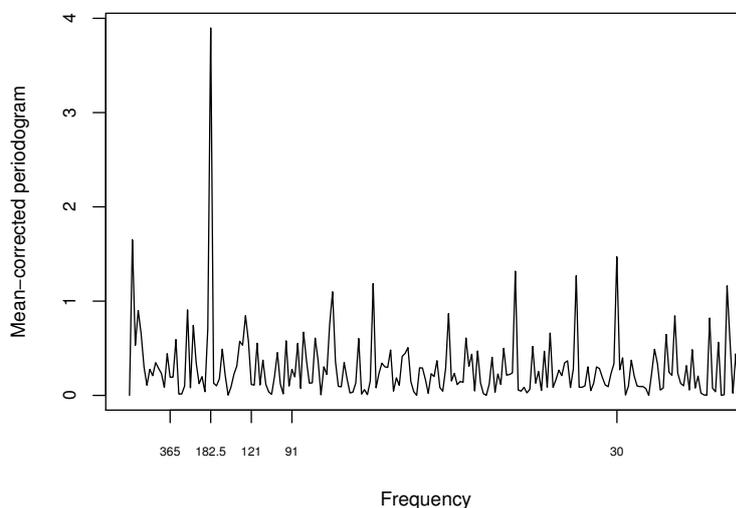


Figure 2.11: Dutch fire insurance data: mean-corrected periodogram.

Figure 2.12 shows the resulting model fit together with confidence bands separately before (left) and after (right) the change-point. Clearly, Model ($M_{1,1}$) seems to be a satisfactory fit before the change-point, but not thereafter (in other words, the linear trend is needed after the change-point). Conversely, Figure 2.13 illustrates the model fit with trend ($M_{1,2}$) separately for both time periods, and for the second time period including a trend leads to a much better fit than without. For the first time period, visually the fit also looks quite good in this case, but the AIC criterion would still suggest leaving out the trend for this case. We hence choose the following combination as the final model:

$$\text{Model } (M_2) : \alpha_{M_2}(t) = \alpha_1(t) \cdot 1_{t < x_0} + \alpha_2(s) \cdot 1_{t \geq x_0}. \quad (2.10)$$

Table 2.8 gives the final parameters estimated by maximum likelihood. One sees that this model improves the AIC value by 46 points compared to the benchmark model

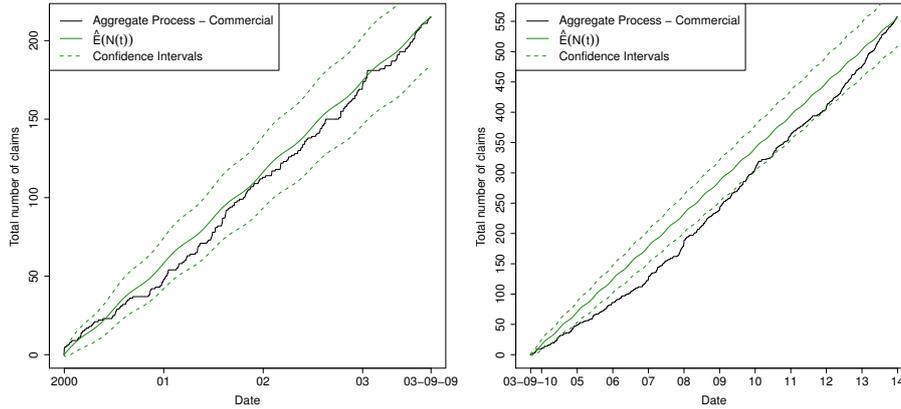


Figure 2.12: Dutch fire insurance data: observed process and simulated confidence band based on model $(M_{1,1})$. Left: time period 2000–09/2003 (AIC=1252.25, BIC=1273.07); Right: time period 09/2003–2013 (AIC=3292.17, BIC=3317.10)

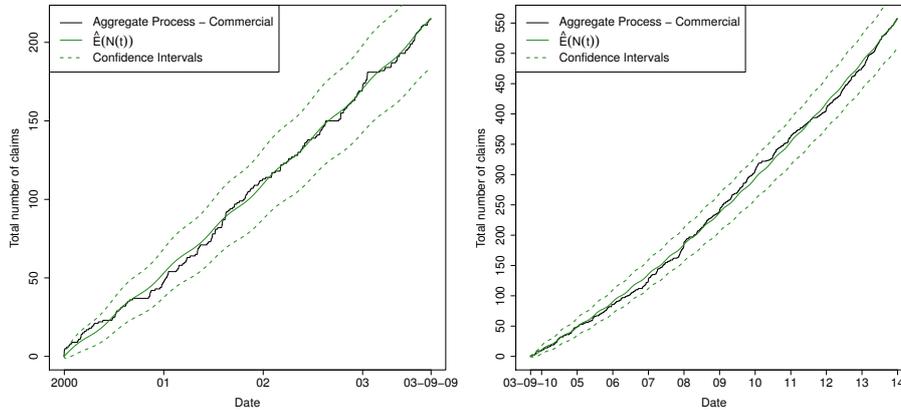


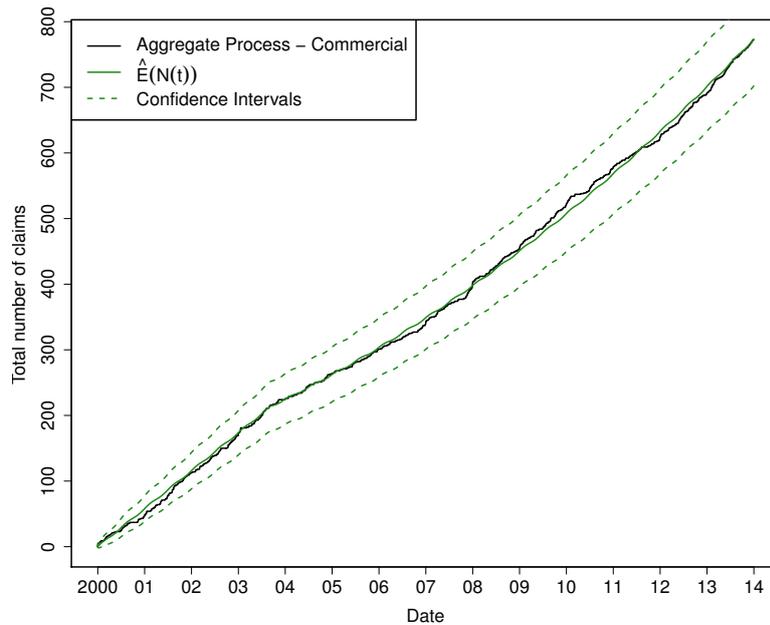
Figure 2.13: Dutch fire insurance data: observed process and simulated confidence band based on model $(M_{1,2})$. Left: time period 2000–09/2003 (AIC=1253.28, BIC=1279.31); Right: time period 09/2003–2013 (AIC=3269.75, BIC=3300.92).

(M_0) . We also see that the semi-annual seasonal effect is slightly more pronounced during the second period ($\hat{\gamma}_{11} < \hat{\gamma}_{12}$).

Figure 2.14 then shows that the final fitted model is situated in the center of the simulated confidence band based on (M_2) and Figure 2.15 shows cluster sizes of five simulated paths from the latter model. The histograms of the claim arrival days within the year of occurrence for some simulated processes from model (M_2) now correspond well with the seasonal pattern for the observed process; see Figure 2.16.

Remark. We compare the non-stationary Cox model fit with that of the inhomogeneous Poisson process with the same intensity function (2.10). Table 2.9 gives the final parameters estimated by maximum likelihood. Consistent with the non-stationary Cox model, we notice that the semi-annual seasonal effect is slightly more

$\hat{\lambda}$	$\hat{\beta}_{01}$	$\hat{\gamma}_{11}$	$\hat{\phi}_{11}$	$\hat{\beta}_{02}$	$\hat{\beta}_{12}$	$\hat{\gamma}_{12}$	$\hat{\phi}_{12}$
0.151375	2.681373	0.719005	0.049676	0.927540	0.000492	0.731791	0.322192
$\hat{\eta}$	ℓ		AIC		BIC		
2.561631	-2'252.85		4'521.69		4'574.02		

Table 2.8: Model (M_2)Figure 2.14: Dutch fire insurance data: simulated confidence intervals and observed aggregate data for Model (M_2).

pronounced during the second period ($\hat{\gamma}_{11} < \hat{\gamma}_{12}$).

Figure 2.17 shows that the final fitted model is situated in the center of the simulated confidence band. When visually comparing the fits, we notice that there is no apparent difference when compared to Figure 2.14. However, when comparing the cluster sizes, the inhomogeneous Poisson process underestimates the number of clusters of size 4. In order to investigate this further, Table 2.10 compares the cluster structure of the observed process with 500 paths coming from M_2 and the inhomogeneous Poisson fit. Both models show similar cluster behavior, but the non-stationary Cox fit performs better in reproducing large clusters, as expected. Appendix A provides a model comparison analysis similar as in Avanzi et al. [26, Sec. 5.6].

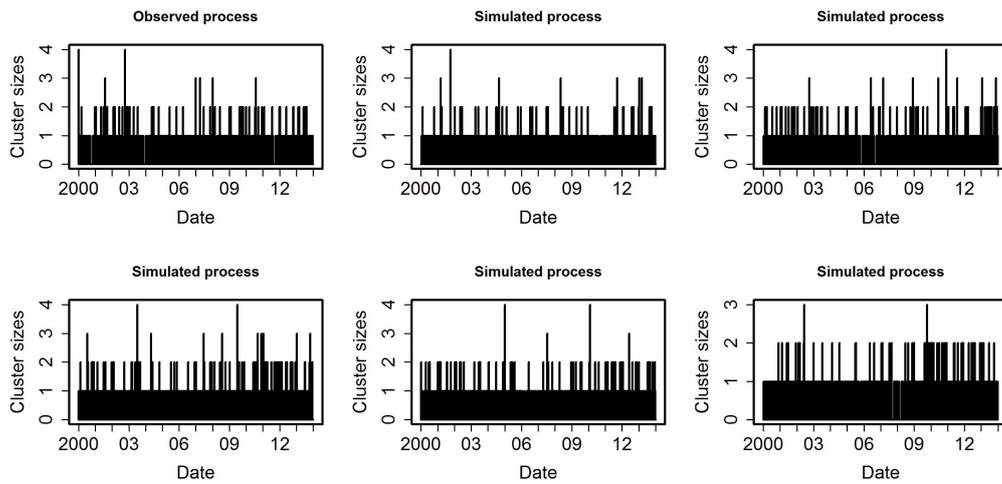


Figure 2.15: Dutch fire insurance data: Observed clusters (top left) versus clusters of five different simulated paths from M_2 .

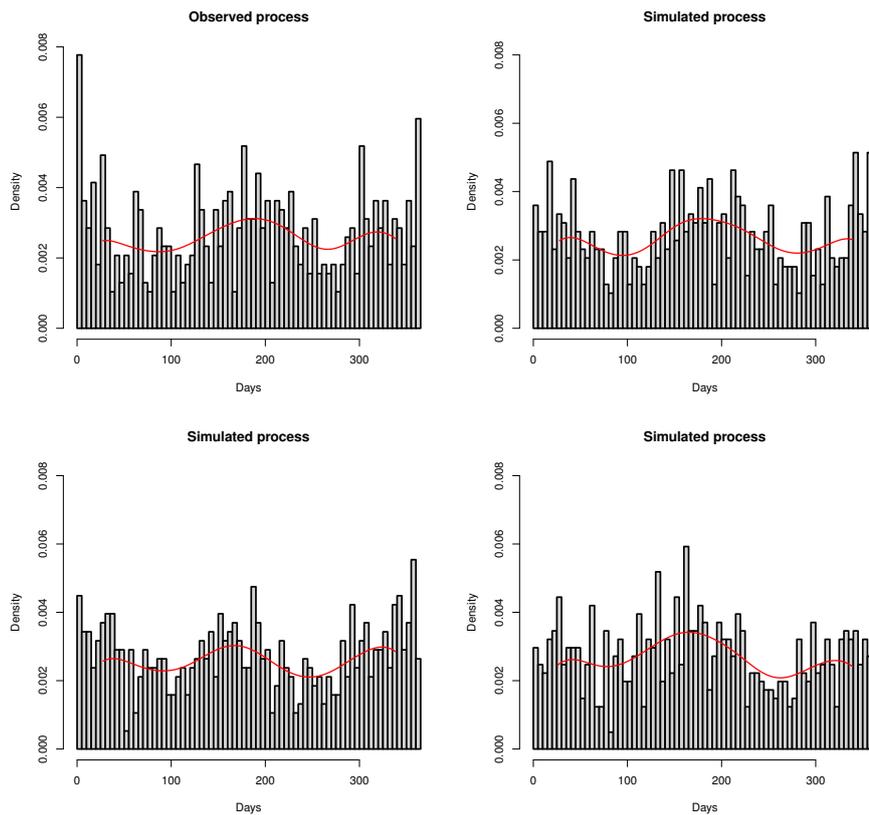
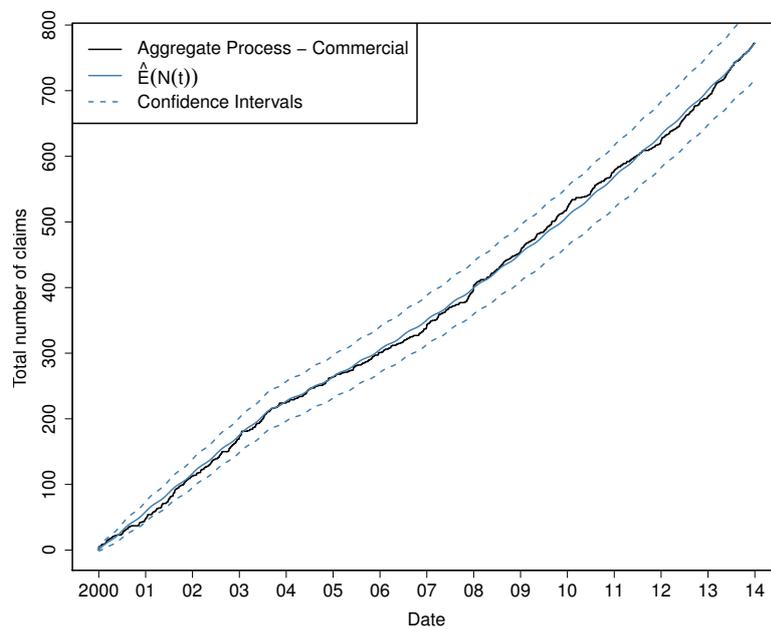


Figure 2.16: Dutch fire insurance data: histogram of the claim arrival days within the years of occurrence for the observed process (top left) and three simulated claim number processes following (M_2) (top right and bottom).

$\hat{\beta}_{01}$	$\hat{\gamma}_{11}$	$\hat{\phi}_{11}$	$\hat{\beta}_{02}$	$\hat{\beta}_{12}$	$\hat{\gamma}_{12}$	$\hat{\phi}_{12}$
0.159673	0.041847	0.037468	0.055816	0.000029	0.043245	0.320861
ℓ		AIC		BIC		
-2'257.06		4'528.12		4'573.89		

Table 2.9: Estimation results for the inhomogeneous Poisson model

Figure 2.17: Dutch fire insurance data: simulated confidence intervals and observed aggregate data for Model (M_{inh}).

Number of claims (%)	2	3	4	5	≥ 6
Observed	1.0959	0.1174	0.0391	0.0000	0.0000
Simulated inh. Poisson	1.0297	0.0592	0.0028	0.0001	0.0000
Simulated M_2	1.2680	0.1137	0.0097	0.0009	0.0002

Table 2.10: Dutch fire insurance data: Percentage of number of claims per day in the observed process versus cluster sizes from 500 paths simulated from (M_2) and the inhomogeneous Poisson process.

2.5.3 Backtesting

In order to test the efficiency of the given modeling approach in prediction of future observations, we fit the model based on the data from the period January 2007 to June 2013 only. We shall consider a non-stationary gamma model $(M_{1,2})$, where the intensity parameter has a linear component and one seasonal component. The parameter estimates are given in Table 2.11.

$\hat{\lambda}$	$\hat{\beta}_{0,2}$	$\hat{\beta}_{1,2}$	$\hat{\gamma}_{1,2}$	$\hat{\phi}_{1,2}$
0.166948	4.108265	0.000431	1.551452	0.009816
	ℓ	AIC	BIC	
	-1'117.919	2'247.839	2'276.696	

Table 2.11: Model $(M_{1,2})$ applied to 01/2007 to 06/2013

The figures reveal that the trend (with coefficient $\hat{\beta}_{1,2}$) is negligible in this case, whereas the semi-annual frequency given by $\hat{\gamma}_{1,2}$ is important (see also Figure 2.18(left)). Moreover, the AIC value assessing the fit of this model shows a clear improvement over the AIC and a modest one over the BIC values of (M_0) , which are 2'266.09 and 2'277.64, correspondingly, for that same period.

Predicting the observations for the last six months of the year 2013 based on Monte Carlo simulation of the fitted model from Table 2.11 gives the confidence bands in Figure 2.18 (right). The observed process during these last six months stays nicely within this confidence band, and the predicted observations slightly underestimate the observed counts.

2.5.4 Risk assessment based on Model (M_2)

We now look into the impact of the different model assumptions on risk measures of the aggregate claim amount. Let us assume that the aggregate claim amount is given by the collective model $S(t) = \sum_{i=1}^{N(t)} X_i$, where the individual claim sizes X_i are independent and identically distributed with cumulative distribution function F_X (and independent of $N(t)$). Under the non-stationary Cox process assumption for $N(t)$ of this paper, one then has

$$F_{S(t)}(x) := \mathbb{P}(S(t) \leq x) = \sum_{n=0}^{\infty} \left\{ \frac{(-\lambda)^n}{n!} \varphi_{M(t)}^{(n)}(\lambda) \right\} F_X^{*n}(x), \quad x \geq 0. \quad (2.11)$$

There are several techniques available to determine the distribution function of the total claim size, for comprehensive recent surveys see, for example, Klugman et al. [130, Ch. 9] and [129, Ch. 6] or Albrecher et al. [7, Ch.6]. Here we use the

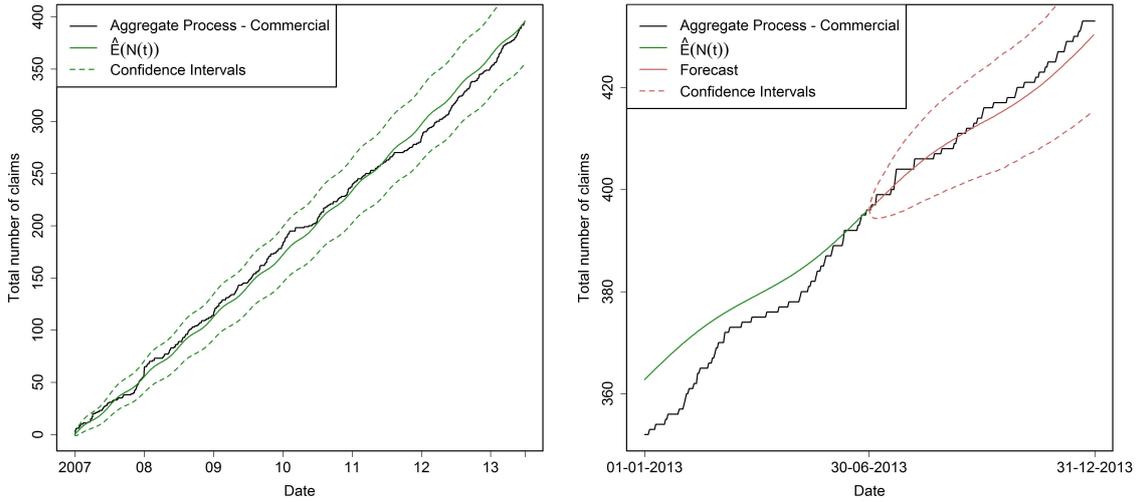


Figure 2.18: Dutch fire insurance data: simulated confidence intervals for model $(M_{1,2})$ and observed data for 2007–06/2013 (left) and 6-month horizon Monte Carlo forecast (right).

fast Fourier transform (FFT) method, which discretizes the claim sizes and then numerically inverts in an efficient way the characteristic function of $S(t)$

$$\mathbb{E}(e^{izS(t)}) = Q_{N(t)}(\mathbb{E}(e^{izX})),$$

exploiting complex analysis techniques (see, e.g., [7, Sec.6.4] for details). Here, $Q_{N(t)}$ denotes the probability generating function of $N(t)$, which in our context, for each t , is the one of a negative binomial distribution with the form

$$Q_{N(t)}(z) = \left(\frac{\eta}{\eta - \lambda(z - 1)} \right)^{\int_0^t \alpha(s) ds}, \quad \text{for } |z| < 1 + \frac{\lambda}{\eta}.$$

In the present application, the discretization of the claim size X is based on the approximative probability function $f_X(x_j) = P(X = x_j)$, $j = 0, \dots, M - 1$, where $M = 2^k$ for some integer k . For the individual claim size X , we choose here the distribution identified for this Dutch fire dataset in [7, Sec.4.3.2]. The discretization is performed using the rounding method such that the true cdf passes through the midpoints of the intervals $[x - h/2, x + h/2)$ for $x = x_0 + h, \dots, x_{M-1} - h$, $f_X(x_0) = F_X(x_0 + h/2)$ and step size $h > 0$. The procedure then finally yields a vector of dimension M with entries $f_{S(t)}(x_j) = P(S = x_j)$, $j = 0, \dots, M - 1$.

The Value-at-Risk (VaR) and the conditional tail expectation (CTE) at significance level p of the aggregate claim amount are defined by

$$\text{VaR}_p(S(t)) = \inf\{x | F_{S(t)}(x) \geq p\},$$

and

$$\text{CTE}_p(S(t)) = \mathbb{E}\left(S(t) \mid S(t) > \text{VaR}_p(S(t))\right) = \frac{1}{1 - p} \int_{\text{VaR}_p(S(t))}^{\infty} x f_{S(t)}(x) dx. \quad (2.12)$$

Once the vector with entries $f_{S(t)}(x_j)$, $j = 0, \dots, M - 1$, is available, one can numerically estimate $\text{VaR}_p(S(t))$ by finding the p -quantile of the estimated cdf. Subsequently, the CTE is then estimated by approximating the integral in (2.12) numerically.

We estimated the 365-days VaR and CTE at the significance levels of 95%, 99%, and 99.5% for both the fitted models (M_0) and (M_2) for the beginning of each year over the period 2000-2013. Table 2.12 and Table 2.13 present the VaR and CTE values computed for some selected years of the observed period.

$S(365)$	VaR_{95}	VaR_{99}	$\text{VaR}_{99.5}$
M_0	381'217'800.00	517'477'900.00	600'596'000.00
$M_{2,2000}$	396'273'700.00	535'193'600.00	619'909'700.00
$M_{2,2003}$	360'060'300.00	492'492'900.00	573'333'700.00
$M_{2,2004}$	279'057'100.00	395'494'400.00	466'719'300.00
$M_{2,2006}$	324'337'100.00	449'996'300.00	526'775'400.00
$M_{2,2009}$	390'454'200.00	528'355'100.00	612'463'400.00
$M_{2,2010}$	412'114'100.00	553'765'200.00	640'108'300.00
$M_{2,2013}$	476'192'500.00	628'319'000.00	720'876'400.00

Table 2.12: VaR estimates for the aggregate claim amount using models (M_0) and (M_2).

$S(365)$	CTE_{95}	CTE_{99}	$\text{CTE}_{99.5}$
M_0	478'982'852.00	684'666'467.00	817'475'670.00
$M_{2,2000}$	495'877'084.67	705'264'162.14	840'324'391.54
$M_{2,2003}$	455'175'006.37	655'533'949.45	785'098'440.98
$M_{2,2004}$	362'988'767.88	540'593'854.37	656'063'739.22
$M_{2,2006}$	414'735'373.87	605'541'563.58	729'228'330.64
$M_{2,2009}$	489'354'409.68	697'331'513.26	831'537'372.76
$M_{2,2010}$	513'598'526.50	726'749'411.99	864'082'258.09
$M_{2,2013}$	584'839'923.07	812'210'387.31	958'019'659.16

Table 2.13: CTE estimates for the aggregate claim amount using models (M_0) and (M_2).

Figure 2.19 plots the 99.5% VaR and 99% CTE values for the entire 14-years period. One sees that moving from the stationary to the non-stationary model has substantial consequences on the respective risk measures (and correspondingly capital requirements). In particular, the simpler model (M_0) would substantially misspecify the respective values.

In order to study the effect of seasonality within a given year we compute the VaR at time $t = 90, 100, \dots, 365$ days for the year 2009 at significance level 99.5% for both models (M_0) and (M_2). Figure 2.20 (left) shows a linear pattern for the VaR values under (M_0), whilst for (M_2) some variation is observed within the year. Figure 2.20

(right) depicts the difference of VaR-values between the two models. Note that the differences in the VaR values at the local minima at days $t = 120$ and $t = 310$ differ. One clearly sees the signature of both the trend and the semi-annual seasonality term in the intensity function.

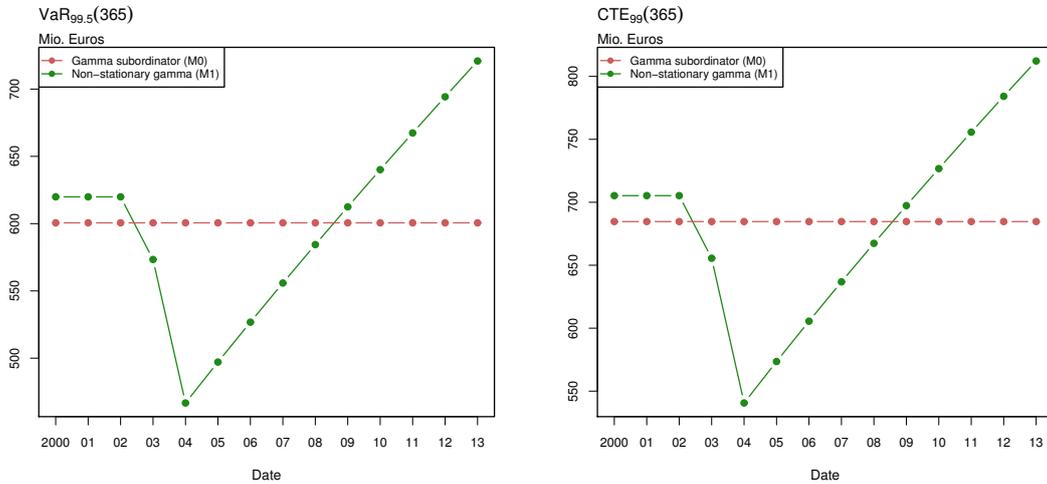


Figure 2.19: Dutch fire insurance data: $\text{VaR}_{99.5}$ and $\text{CTE}_{0.99}$ for models (M_0) and (M_2)

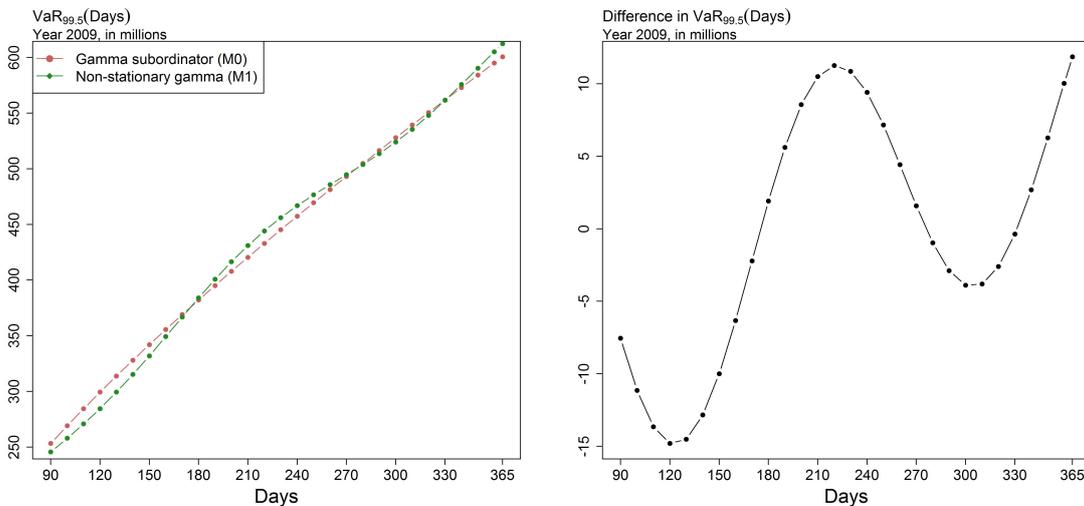


Figure 2.20: Dutch fire insurance data: $\text{VaR}_{99.5}^{(2009)}(t)$ for $t = 90, \dots, 365$ for models (M_0) and (M_2) (left) and absolute difference in mio. Euro (right).

2.6 A multivariate extension and case study

As outlined in the introduction, one advantage of subordinator Cox models is that they lead to natural multivariate models, where the dependence between different lines of business is introduced by sharing the same realization of the subordinator for otherwise independent Poisson processes (see Scherer and Selch [169] for details). In this section, we would like to illustrate this idea in our non-stationary setup.

Let $\{N_j(t); t \geq 0\}$ count the claims arriving in each line of business (or portfolio) $j = 1, \dots, d$ up to time $t \geq 0$. The multivariate Cox process $\mathbf{N} = (N_1, \dots, N_d)$ is defined through

$$\begin{aligned} \mathbf{N} &= \widetilde{\mathbf{N}}_\lambda(M(t)) \\ &= \left\{ (\widetilde{N}_{1,\lambda_1}(M(t)), \dots, \widetilde{N}_{d,\lambda_d}(M(t))); t \geq 0 \right\}. \end{aligned}$$

Note that we allow for d different marginal intensities λ_j , while the common $M(t)$ introduces dependence between the d lines of business. In particular, this construction can now lead to common jumps in several components at the same time. Expression (2.4) for the finite-dimensional distributions of \mathbf{N} with time points $0 := t_0 \leq t_1 \leq \dots \leq t_n$, and $0 := \mathbf{k}_0 \leq \mathbf{k}_1 \leq \dots \leq \mathbf{k}_n$ for with $\mathbf{k}_i = (k_i^{(1)}, \dots, k_i^{(d)})$, $k_i^{(j)} \in \mathbb{N}$, $i = 1, \dots, n$, now generalizes to

$$\mathbb{P}(\mathbf{N}(t_1) = \mathbf{k}_1, \dots, \mathbf{N}(t_n) = \mathbf{k}_n) = \prod_{i=1}^n \frac{(-\boldsymbol{\lambda})^{|\Delta \mathbf{k}_i|}}{\Delta \mathbf{k}_i!} \varphi_{\Delta M(t_i)}^{(|\Delta \mathbf{k}_i|)}(|\boldsymbol{\lambda}|), \quad (2.13)$$

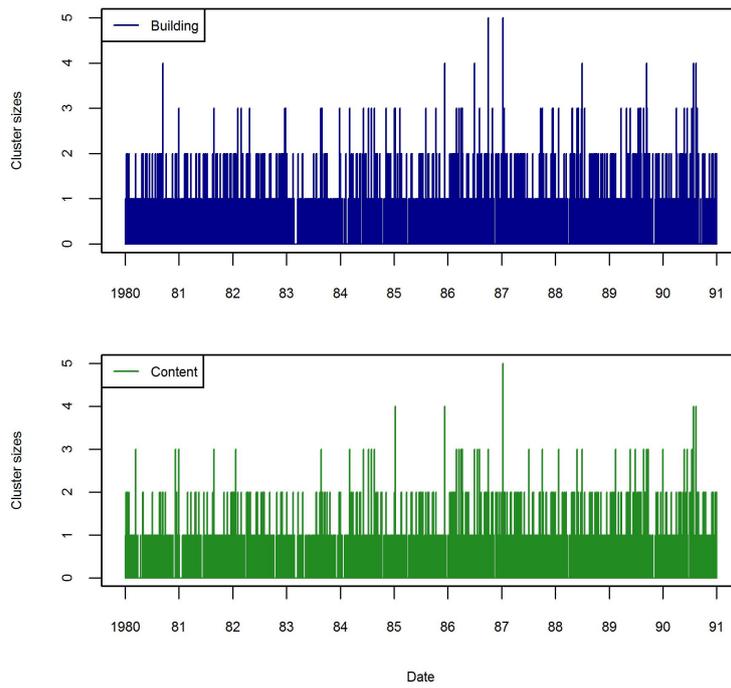
with $\Delta \mathbf{k}_i = (k_i^{(1)} - k_{i-1}^{(1)}, \dots, k_i^{(d)} - k_{i-1}^{(d)})$, $\Delta t_i = t_i - t_{i-1}$ for $i = 1, \dots, n$. Here we use the multi-index notation $|\mathbf{x}| = x_1 + \dots + x_k$, $\mathbf{k}! = k_1! \dots k_d!$ and $\mathbf{x}^{\mathbf{k}} = x_1^{k_1} \dots x_d^{k_d}$.

For the estimation we again assume that the multivariate process \mathbf{N} is observed up to a time horizon $T > 0$ on a discrete time grid $0 := t_0 \leq t_1 \leq \dots \leq t_n = T$.

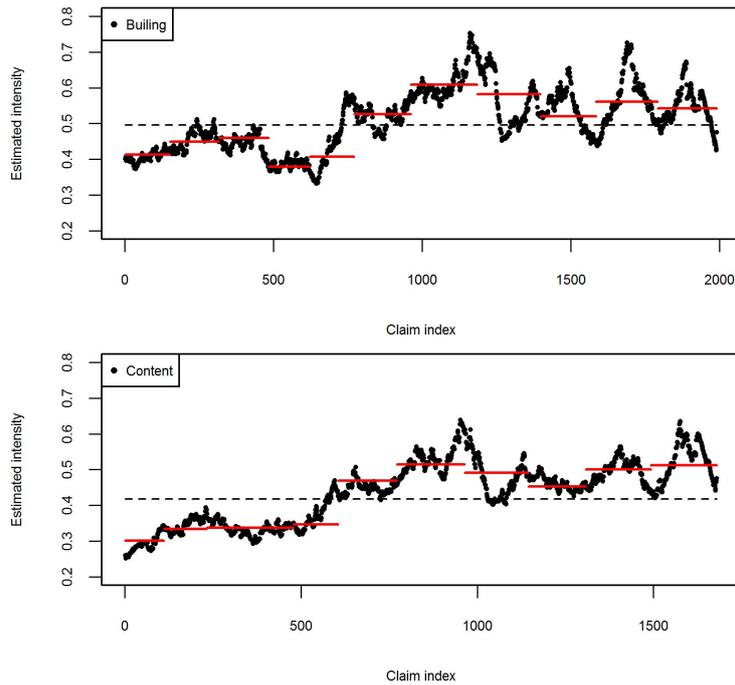
Similar to the univariate case, we denote by $\hat{\mathbf{k}}_i = (\hat{k}_i^{(1)}, \dots, \hat{k}_i^{(d)})$ the observations of the process \mathbf{N} at time t_i for $i = 1, \dots, n$. Since \mathbf{N} has independent increments, it follows that $\Delta \hat{\mathbf{k}}_i = \hat{\mathbf{k}}_i - \hat{\mathbf{k}}_{i-1}$, $i = 1, \dots, n$ are i.i.d. observations of $\Delta \mathbf{N}$. Using (2.13), the log-likelihood function is given by

$$\begin{aligned} \ell(\cdot; \Theta) &= \sum_{i=1}^n |\Delta \hat{\mathbf{k}}_i| \log(\boldsymbol{\lambda}) - \sum_{i=1}^n \log(\Delta \hat{\mathbf{k}}_i!) + \sum_{i=1}^n \log \left\{ (-1)^{|\Delta \hat{\mathbf{k}}_i|} \varphi_{\Delta M(t_i)}^{|\Delta \hat{\mathbf{k}}_i|}(|\boldsymbol{\lambda}|; \Theta) \right\} \\ &= \sum_{j=1}^d \hat{k}_j^{(n)} \log(\lambda_j) - \sum_{j=1}^d \sum_{i=1}^n \log(\hat{k}_i^{(j)}!) + \sum_{i=1}^n \log \left\{ (-1)^{|\Delta \hat{\mathbf{k}}_i|} \varphi_{\Delta M(t_i)}^{|\Delta \hat{\mathbf{k}}_i|}(|\boldsymbol{\lambda}|; \Theta) \right\}. \end{aligned} \quad (2.14)$$

Consequently, optimizing (2.14) with respect to the intensity parameter $\boldsymbol{\lambda}$ of the underlying Poisson process and the parameter vector Θ of the additive subordinator leads to the corresponding MLE estimators of the observed process. Note that the second term of the second equation in (2.14) is independent of the parameters and thus can be neglected in the optimization.



(a) Daily cluster sizes for the building (top) and content (bottom) categories.



(b) Moving average of the estimated intensity function ($h = 20$) (dots) together with the reciprocal of the annual expected inter-arrival times (piecewise constant solid line) and the intensity estimate of a homogeneous Poisson process (dashed line) for the building (top) and content (bottom) categories.

Figure 2.21: Danish fire: daily cluster sizes (top) and moving average of the estimated intensity function (bottom).

2.6.1 Case study: Danish fire insurance data.

We now illustrate the performance of the ML estimation of this multivariate process with the well-known Danish fire data set; see McNeil [144]. For the present purpose, we use the occurrence dates between 1980 and 1991 of the 2167 fire events for the two subportfolios building and content. Figure 2.21a shows the cluster arrivals for both building and content. As noted in Scherer and Selch [169, Sec 3.3], we observe simultaneous cluster arrivals; as for example, at the start of 1987. A multivariate Cox process with a common subordinator hence seems to be a reasonable model for this data set. Figure 2.21b depicts (2.1) for both lines. One clearly observes a similar trend for both components, which is an argument in favor of the model discussed here.

Similar to Section 2.5, we consider the multivariate gamma-subordinated Poisson process as a benchmark model. In this case the model parameters are λ_1 , λ_2 , α and η . From time normalization we have $\alpha = \eta$. Optimizing (2.14) one obtains the MLE parameters with log-likelihood ℓ , AIC and BIC as given in Table 2.14.

$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\alpha} = \hat{\eta}$	ℓ	AIC	BIC
0.495611	0.418039	0.671440	-6'566.72	13'139.44	13'160.41

Table 2.14: Danish fire: Gamma subordinator fit.

From Figure 2.22 we see that the observed processes systematically deviates from a linear increase of the Lévy subordinator. Therefore, one might consider a non-stationary gamma process with intensity function

$$\text{Model } (M_{\text{Dan}}) : \alpha_{M_{\text{Dan}}}(s) = \beta_0 + \beta_1 s.$$

Table 2.15 and Figure 2.23 summarize the fit. When comparing the AIC and BIC for both models, the non-stationary Cox model appears to be a better choice. The estimated mean now closely describes the observed data. Figure 2.24 compares the observed cluster arrivals with the one of three simulated paths from our model, and one visually concludes that the model produces larger clusters than the observed process. Table 2.17 compares observed and simulated simultaneous cluster sizes quantitatively, from which one concludes that medium range simultaneous cluster sizes are captured nicely with this model.

$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\eta}$
0.495658	0.418166	0.523972	8.06×10^{-5}	0.685756
ℓ		AIC		BIC
-6'542.33		13'096.66		13'138.61

Table 2.15: Model ($M_{\text{Dan}}^{\text{sgs}}$)

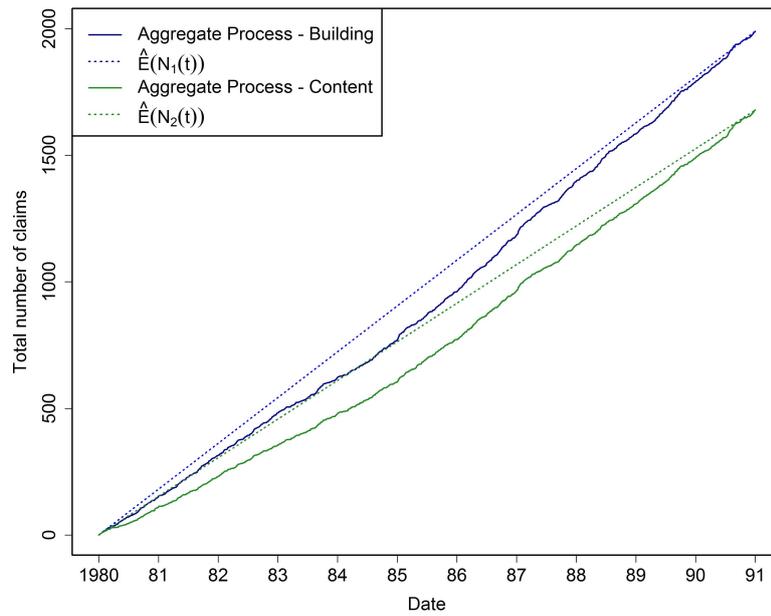


Figure 2.22: Danish fire: observed process and fitted gamma subordinator (dotted line) for each portfolio component.

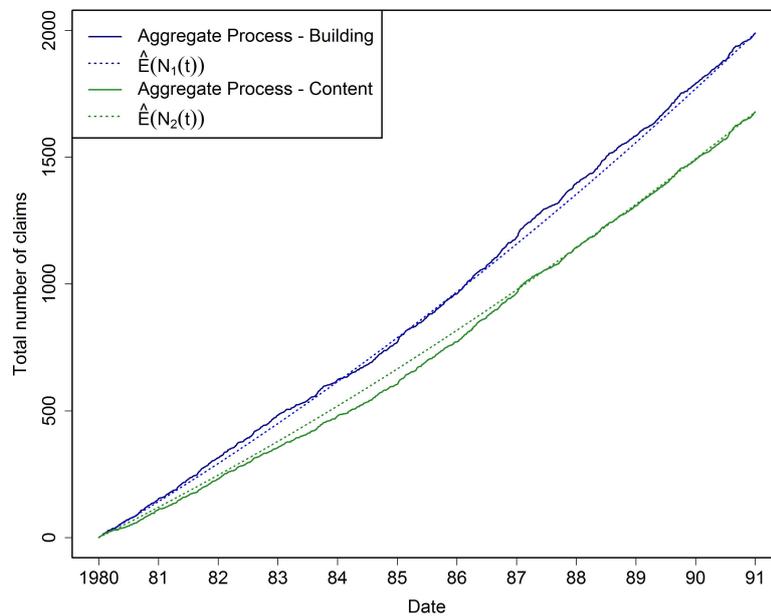


Figure 2.23: Danish fire: observed process and fitted non-stationary gamma subordinator (dotted line) for each portfolio component.

Remark. We also compare the fit of the non-stationary gamma Cox model with a

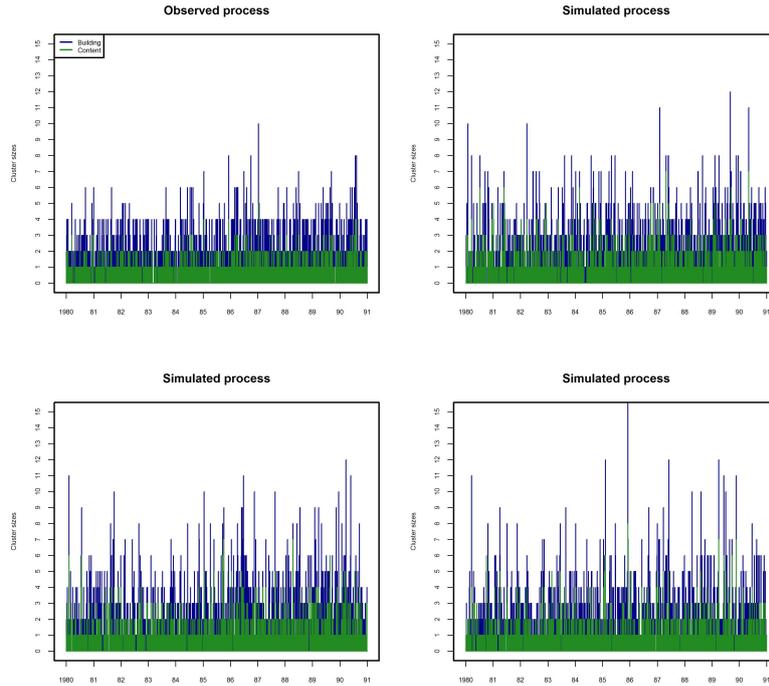


Figure 2.24: Danish fire: observed process (top left) and cluster sizes of four simulated claim number process following the non-stationary Cox model from Table 2.15.

bivariate inhomogeneous Poisson process given by

$$\mathbf{N}^{\text{inh}} = \left\{ (\tilde{N}_{1,\lambda_1}(\mu(t)), \tilde{N}_{2,\lambda_2}(\mu(t))); t \geq 0 \right\},$$

with a common intensity function $\mu(t) = \int_0^t (\beta_0 + \beta_1 s) ds$. The resulting ML estimators are summarized in Table 2.16. The plot analogous to Figure 23 results in the very same shape, so we refrain from showing it here. In terms of AIC and BIC criteria, this model also seems to be reasonable for the purpose. However, Figure 2.25 and Table 2.17 illustrate that the observed cluster behavior in the data is much better represented by the Cox model than by the inhomogeneous Poisson alternative.

$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$
0.492836	0.415844	0.757557	1.24×10^{-4}
ℓ	AIC		BIC
-5'907.02	11'822.03	11'850.00	

Table 2.16: Model ($M_{\text{Dan}}^{\text{inh}}$)

Cluster size	2	3	4	5	6	≥ 7
Observed	4.757161	0.647572	0.099626	0.024907	0.000000	0.000000
Simulated ($M_{\text{Dan}}^{\text{inh}}$)	0.462964	0.011905	0.000100	0.000000	0.000000	0.000000
Simulated (M_{Dan})	1.397908	0.345205	0.089813	0.024010	0.006326	0.002192

Table 2.17: Danish fire: Percentage of simultaneous clusters arrivals for the observed process and 500 simulated paths of the fitted multivariate inhomogeneous Poisson process ($M_{\text{Dan}}^{\text{inh}}$) and non-stationary Gamma subordinator (M_{Dan}).

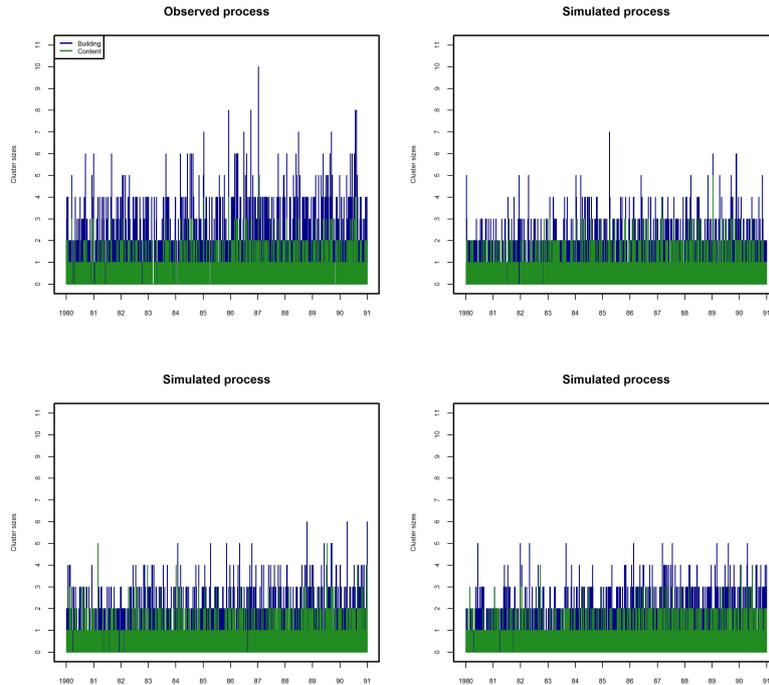


Figure 2.25: Danish fire: observed process (top left) and cluster sizes of four simulated claim number process of the fitted model ($M_{\text{Dan}}^{\text{inh}}$).

2.7 Conclusion

We showed how to extend over-dispersed Cox models that account for cluster arrivals in insurance claims modeling by seasonal patterns in a natural way. Concretely, when the seasonality is introduced in the subordinator directly, one can estimate all parameters conveniently at the same time. We exemplified the approach alongside a concrete fire insurance dataset and illustrated how one can in a pre-analysis step possibly formulate suitable parametric families for the time-dependent intensity functions that are then used in the estimation procedure. The resulting parsimonious non-stationary Cox process led to a substantial improvement in AIC and BIC values when compared to the stationary model and allows capturing seasonal characteristics that may be inherent in the data. Our approach to specify the parametric form

of the intensity function allowed to preserve tractability, but clearly modifications of that step may be required for other datasets. We also performed a rather promising back-testing procedure. Whereas the focus in this paper was on gamma subordinators leading to negative binomial marginal distributions for the claim counts, the procedure can be applied to any other subordinator choice as well. We also illustrate how such subordinator models can be used to formulate non-stationary multivariate claim number processes, where the dependence is introduced by a common subordinator, generating common cluster arrivals in its components as often observed in insurance applications.

2.8 Appendix A: Model comparison

	Hom. Poisson	M_0	M_{inh}	M_2
Σ residuals	0.000000	0.000000	-0.364845	-1.527825
Σ absolute residuals	1'333.10	1'333.10	1'318.07	1'318.78
Σ residuals squared	827.37	827.37	818.54	818.55
L-B test p -value (lag = 91)	0.250472	0.250472	0.680661	0.680675
L-B test p -value (lag = 182)	0.682880	0.682880	0.987985	0.987986
L-B test p -value (lag = 365)	0.135887	0.135887	0.635085	0.634794
Runs test p -value	0.000142	0.000142	0.001273	0.001275

Table 2.18: Summary statistics for model comparison.

This Appendix provides a comparison between the four models considered in Section 2.5. The comparison is done using the same tests as in Avanzi et al. [26, Sec. 5.6]². For each of the models, the predicted mean per day and daily residuals were calculated. Due to the time normalization assumption, the mean of the homogeneous Poisson model and the stationary gamma subordinator model (M_0) is the same. Moreover, the mean of the inhomogeneous Poisson model (M_{inh}) and the non-stationary Cox model (M_2) also coincide. Therefore, for each pair the sample residuals agree.

The summary statistics of the residuals are given in Table 2.18. Moreover, the latter also gives the p -value for the Ljung-Box (L-B) test for autocorrelation for quarterly, semi-annual, and annual lags. Also, the p -value for the Wolf-Waldowitz runs test for randomness/persistence is given. As previously mentioned, the first two and last two columns coincide given that the residuals for the stationary models are the same and likewise for the non-stationary ones. Based on this analysis, we see that the inhomogeneous Poisson and non-stationary gamma model produced a better fit to the data, however the improvement does not seem to be significant.

Figure 2.26 depicts the estimated daily mean for the gamma subordinator (first row, left) and the corresponding sample residuals (first row, right). Also, the daily mean for the non-stationary gamma model (third row, left) and its residuals (third row, right) are included. Moreover, the second and fourth row depict the sample autocorrelation (left) and partial autocorrelation (right) for the residuals of both models. Furthermore, no serial correlation is revealed for any of the models.

²The code for this analysis was based on Avanzi and co-authors R code. The latter can be found at: <https://github.com/agi-lab/reserving-MMNPP>

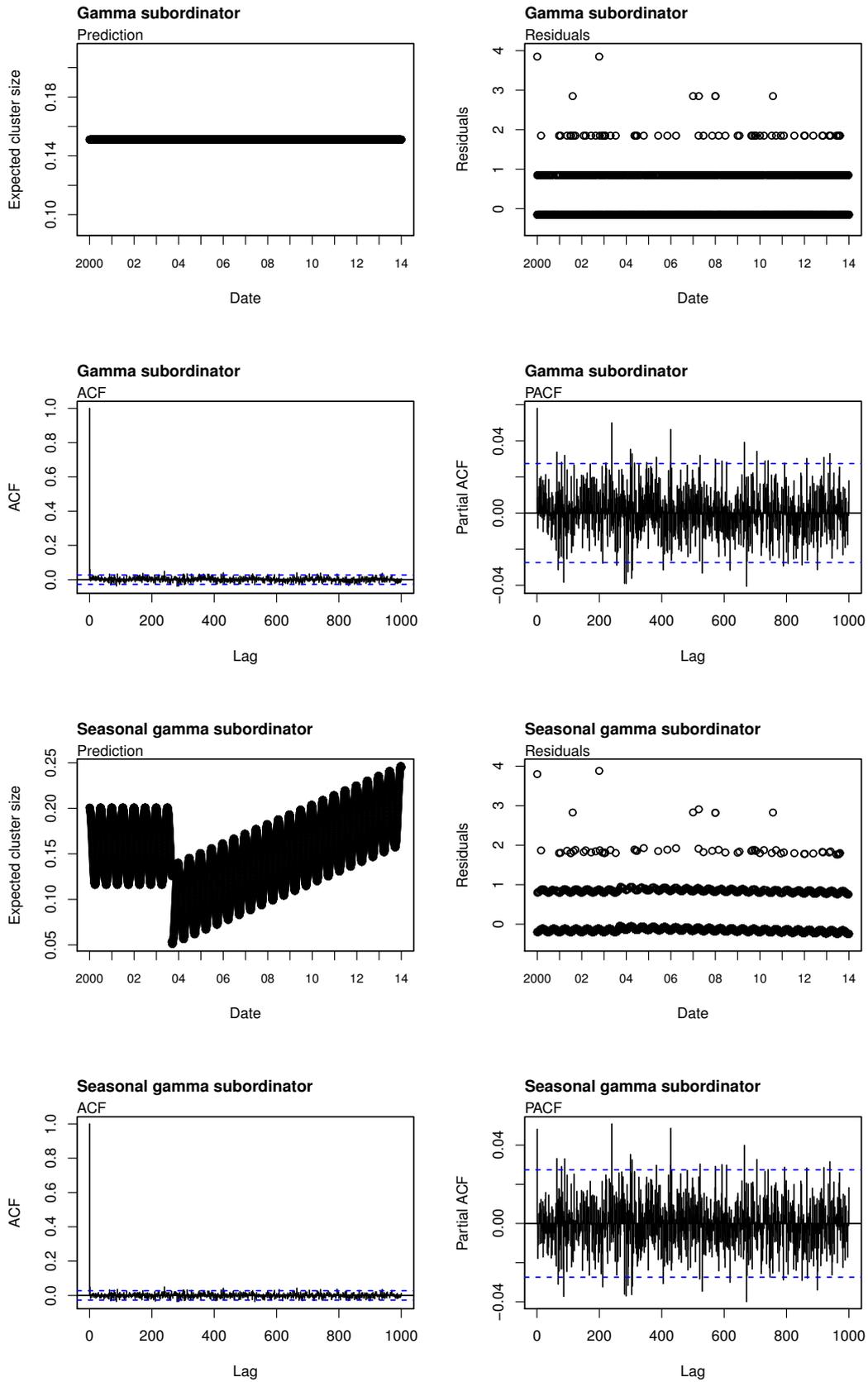


Figure 2.26: Dutch fire insurance data: Predictions and residuals for model M_0 (top 2 rows) and model M_2 (bottom 2 rows).

Chapter 3

Tempered Pareto-type modeling using Weibull distributions

This chapter is based on the following article:

H. Albrecher, J.C. Araujo-Acuna, and J. Beirlant. Tempered Pareto-type modeling using Weibull distributions. *ASTIN Bulletin*. To appear, 2020.

Abstract. In various applications of heavy-tail modeling, the assumed Pareto behavior is tempered ultimately in the range of the largest data. In insurance applications, claim payments are influenced by claim management and claims may, for instance, be subject to a higher level of inspection at highest damage levels leading to weaker tails than apparent from modal claims. Generalizing earlier results of Meerschaert et al. [145] and Raschke [157], in this paper we consider tempering of a Pareto-type distribution with a general Weibull distribution in a peaks-over-threshold approach. This requires to modulate the tempering parameters as a function of the chosen threshold. Modelling such a tempering effect is important in order to avoid overestimation of risk measures such as the Value-at-Risk (VaR) at high quantiles. We use a pseudo maximum likelihood approach to estimate the model parameters, and consider the estimation of extreme quantiles. We derive basic asymptotic results for the estimators, give illustrations with simulation experiments and apply the developed techniques to fire and liability insurance data, providing insight into the relevance of the tempering component in heavy-tail modeling.

3.1 Introduction

Probability distributions with power-law tails are extensively used in various fields of applications including insurance, finance, information technology, mining of precious stones and language studies (see e.g. [150] for a recent overview). In extreme value methodology such applications are appropriately modeled using the concept of Pareto-type models such that a variable X of interest satisfies

$$\mathbb{P}(X > x) = x^{-\alpha}\ell(x), \quad (3.1)$$

with $\alpha > 0$ and some slowly varying function ℓ satisfying

$$\frac{\ell(tx)}{\ell(t)} \rightarrow 1 \text{ as } t \rightarrow \infty \text{ for every } x > 0. \quad (3.2)$$

In addition to the (pure) Pareto distribution, further examples from this model are the Burr, Fréchet, t and log-gamma distribution (see Beirlant et al. [36, Ch. 2] for an overview). Often the power-law behavior does not extend indefinitely due to some truncation or tapering effects. In Beirlant et al. [33], estimation of truncated tails was developed in a peaks-over-threshold (POT) approach for Pareto-type tails, and other max-domains of attraction were dealt with in Beirlant et al. [34]. Inspired by applications in geophysics and finance, Meerschaert et al. [145] discussed parameter estimation under exponential tempering of a simple Pareto law with survival function

$$\mathbb{P}(X > x) = cx^{-\alpha}e^{-\beta x}, \quad (3.3)$$

where $\alpha, \beta > 0$ and $c > 0$ is a scale parameter. In the context of insurance data, Raschke [157] recently discussed the use of the more general Weibull tempering of a simple power law with survival function

$$\mathbb{P}(X > x) = cx^{-\alpha}e^{-(\beta x)^\tau}, \quad (3.4)$$

with $c, \alpha, \beta, \tau > 0$.

However, typically the power-law behavior only sets in from some threshold t on, rather than from the lowest measurements as assumed when using the simple Pareto model. The Pareto-type model (3.1) allows for flexible modeling of this behavior. In this paper we therefore want to study *Weibull tempered Pareto-type distributions* with survival function

$$\mathbb{P}(X > x) = x^{-\alpha}\ell(x)e^{-(\beta x)^\tau}, \quad (3.5)$$

with ℓ a slowly varying function, $\alpha = 1/\gamma > 0$ controlling the power-law tail with extreme value index γ , and β, τ governing the Weibull tempering. Note that for $\tau = 1$ one recovers the exponential tempering. Moreover, for $\tau > 1$ the Weibull distribution has lighter tails than exponential and therefore one expects the tempering to be quite strong, or closer to truncation. On the other hand, when $\tau < 1$ the Weibull distribution has heavier tails than exponential and consequently the

tempering is expected to be weaker.

We illustrate the need for such Weibull tempering of a Pareto-type tail with the Norwegian fire insurance data set discussed in Beirlant et al. [36], which contains the year of occurrence of the claim and the claim value (in thousand Kroner) from 1972 until 1992, see also Brazauskas and Kleefeld [44, Sec. 2] for a detailed description of the data. In Figure 3.1 these data are plotted by year of occurrence, next to a log-log plot (Pareto QQ-plot)

$$\left(-\log\left(1 - \frac{j}{n+1}\right), \log X_{j,n}\right), \quad j = 1, \dots, n,$$

where $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ denote the ordered data from a sample of size n . Strict Pareto behavior corresponds to an overall linear log-log plot, but linearity only arises approximately at the top 5000 observations. Note also the bending at the largest observations in the upper right corner in the log-log plot. This tapering near the highest observations often occurs with insurance claim data and typically is due to a stricter claim management policy for the larger claims or due to a policy coverage limit or probable maximum loss. Nevertheless, under the assumption that such upper bounds exist, truncated Pareto-type models might be more suitable. This tapering is also visible when plotting the pseudo maximum likelihood estimator $\hat{\alpha}_k^H = 1/H_{k,n}$ of α under (3.1) (cf. bottom plot in Figure 3.1), where $H_{k,n}$ denotes the Hill estimator [114]

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^k \log \frac{X_{n-j+1,n}}{X_{n-k,n}}. \quad (3.6)$$

The latter can be considered as an estimator of the slope in the log-log plot when restricting to the top $k+1$ observations. In that sense, the statistics $H_{k,n}$ can be considered as derivatives of the Pareto QQ-plot at the top k observations. Here, the values $\hat{\alpha}_k$ exhibit a stable area for $1000 \leq k \leq 5000$ which expresses power-law behavior beyond $X_{n-1000,n}$, and make a sharp increase at the smallest k values due to tapering. This sharp increase behavior is referred to as ‘‘Hill horror plot’’ in the literature (cf. Resnick [159]) and it arises from the difficulty that the Hill estimator cannot correctly adjust non-Pareto-like tails. The latter often occur for insurance data, since – due to truncation and tempering effects – the data become less heavy-tailed further further out in the tails. Following the QQ- and derivative plot methodology from Chapter 4 in Albrecher et al. [7], one can construct a Weibull QQ-plot $(\log(-\log(1 - \frac{j}{n+1})), \log X_{j,n})$, $j = 1, \dots, n$, and its derivative plot in order to verify the Weibull nature of the tempering as proposed in (3.5). A Weibull tail is observed when a linear behavior is apparent in that QQ-plot at some top portion of the data, which can then be confirmed by a constant derivative plot in that region. For the present case, Figure 3.2 shows that the derivative plot becomes constant on average when $\log X > 11$, corresponding to a linear Weibull pattern in the QQ-plot at the top observations with vertical coordinate larger than 11.

As a second example, a tapering effect is also observed in the Secura Belgian Re data set from Beirlant et al. [36]. We refer the reader to Beirlant et al. [35, Sec. 1.3.3

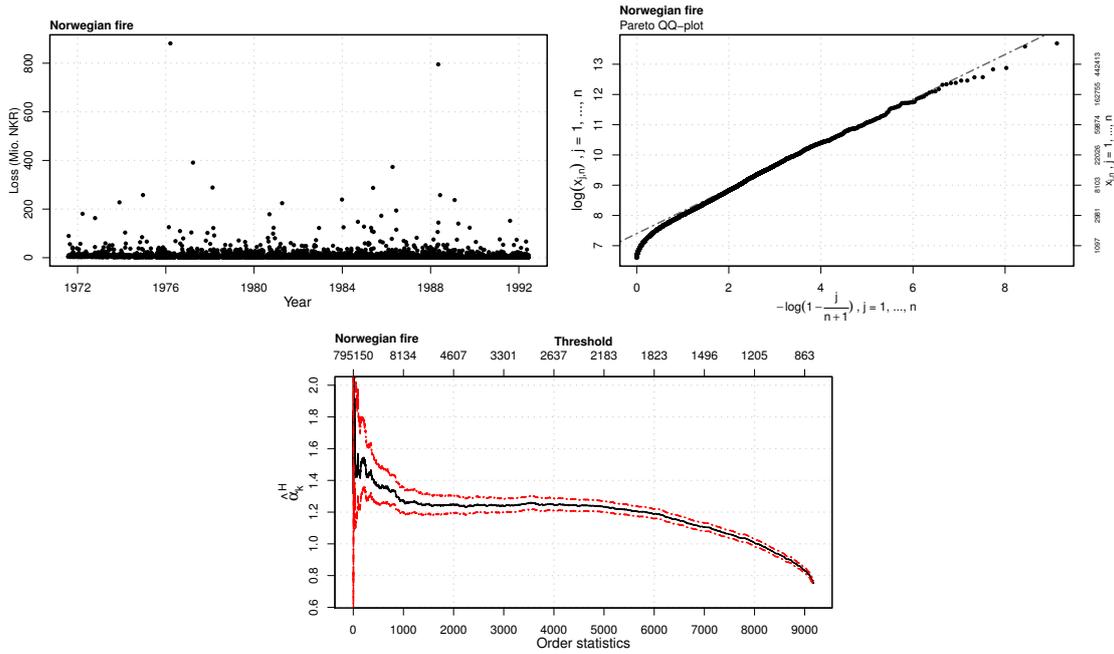


Figure 3.1: Norwegian fire claim data: claim sizes as a function of occurrence time (top left), log-log plot (top right) and $\hat{\alpha}_k^H$ estimates with 95% confidence interval (bottom).

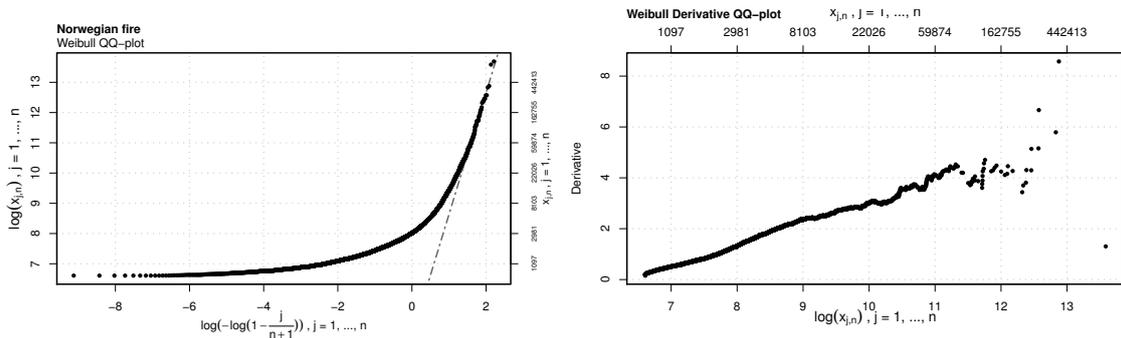


Figure 3.2: Norwegian Fire claim data: Weibull QQ-plot (left) and Weibull derivative plot (right).

& Sec. 6.2] for further details about the data set. The Pareto QQ-plot in Figure 3.3 shows a linear pattern from $\log X > 15$, but bending is visible near the top 10 observations, leading to higher values of $\hat{\alpha}_k^H$ at $k \leq 10$. The Weibull derivative plot shows an ultimately decreasing behavior at the largest 10 observations. This then could lead to truncated Pareto modeling rather than Weibull tempering of a Pareto-type tail, as discussed in detail in Beirlant et al. [33].

In this paper, we complement the graphical and exploratory analysis of Weibull tempering of Pareto-type tails as illustrated above with a mathematical analysis of model (3.5). This can be considered as an alternative to the truncated Pareto-type distributions X discussed in [33] which were defined by $X =_d Y|Y < T$ for some

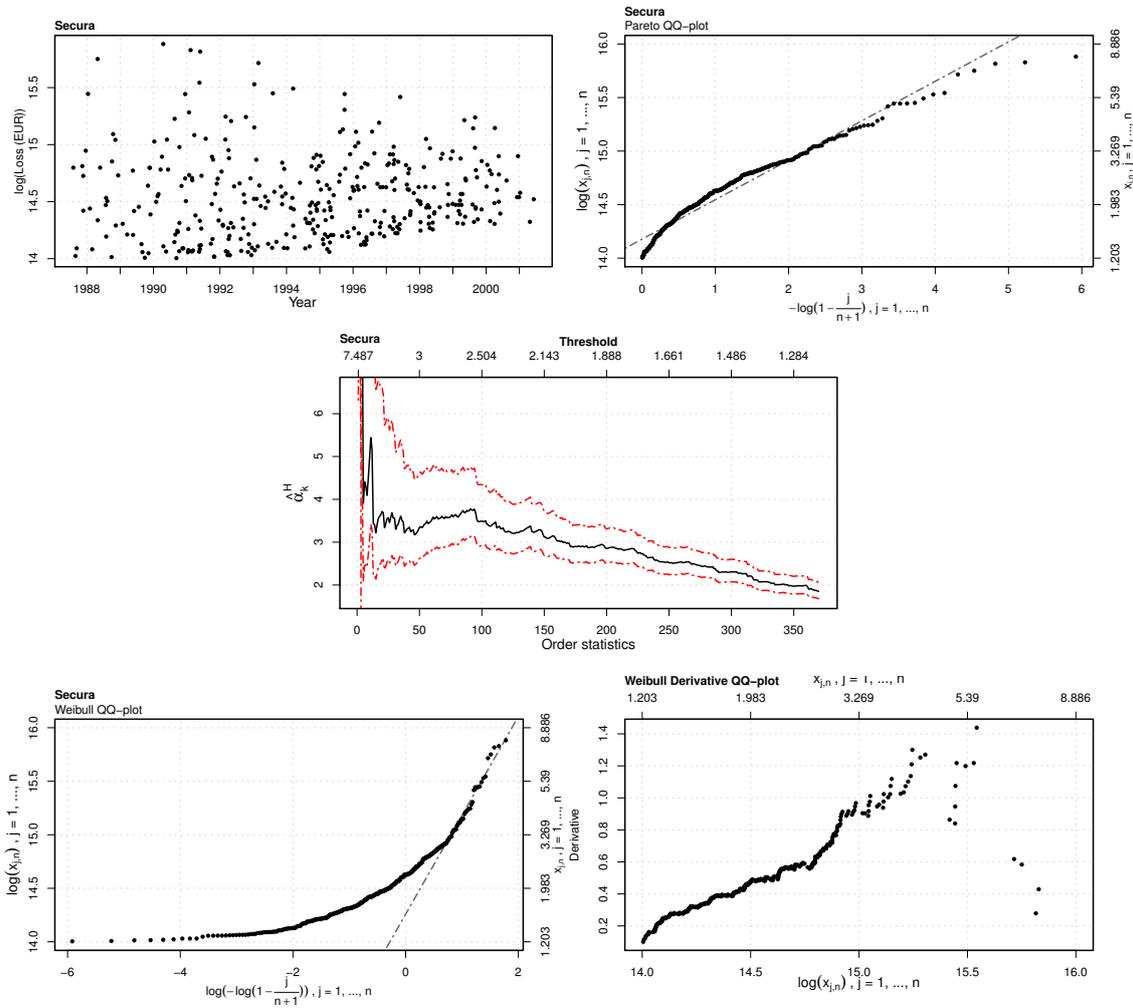


Figure 3.3: Secura Belgian Re claim data: $\log(\text{Claim sizes})$ as a function of the year of occurrence (top left), log-log plot (top right), $\hat{\alpha}_k^H$ estimates with 95% confidence interval (middle), Weibull QQ-plot (bottom left) and Weibull derivative plot (bottom right).

high value of T and Y satisfying Pareto-type behaviour (3.1). Truncation also leads to tapering and appears, for instance, in modeling of earthquake energy levels on the basis of the Gutenberg-Richter law. From the viewpoint of truncation, model (3.5) corresponds to $X = \min(Y, W)$ with Y and W independent, Y being Pareto-type distributed and W Weibull distributed with $\mathbb{P}(W > x) = e^{-(\beta x)^\tau}$. Such a model is intended to describe situations where a gradual transit from a power-law decay to an exponentially fast decay is observed as one goes further into the tail. In view of the general nature of the Pareto-type models (3.1), this approach will not be able to capture the characteristics over the whole range of the distribution but focuses rather on the largest observations above some threshold $X_{n-k, n}$. However, if appropriate such tempered tail fits could be spliced with different methods to describe the data below the chosen $X_{n-k, n}$, as it was done before to obtain composed models with a Pareto or generalized Pareto tail fit, see, for instance, Reynkens et al. [160] for mixed Erlang compositions with Pareto tails, Brazauskas and Kleefeld [44] for log-normal and Weibull models spliced with Pareto tail fits, and Raschke [157] for Pareto-Pareto or cascade Pareto modeling. Albrecher et al. [8] considered a parsimonious and versatile family of distributions for the modeling of heavy-tailed risks using the class of matrix Mittag-Leffler distributions.

In Section 3.2, we position the tempered Pareto-Weibull model in a POT approach allowing $\beta \rightarrow 0$ as the threshold $t \rightarrow \infty$ and study pseudo-maximum likelihood estimation providing basic asymptotic theory. We also discuss estimation of extreme return levels and return periods. Proofs of mathematical results are deferred to the Appendix. In Section 3.3 we provide simulation results, and in Section 3.4 we complete the analysis of the Norwegian fire and the Belgian liability insurance data sets based on the obtained results. Section 3.5 concludes.

3.2 Tempered Pareto-type modeling and estimation

Let $X = \min(Y, W)$ with Y and W independent, where Y is Pareto-type distributed following (3.1) and

$$\mathbb{P}(W > x) = e^{-(\beta x)^\tau} \text{ for } x > 0.$$

The survival function of X is then given by

$$\mathbb{P}(X > x) := \bar{F}(x) = x^{-\alpha} \ell(x) e^{-(\beta x)^\tau}.$$

For the POT distribution of $\frac{X}{t} | X > t$ for some threshold $t > 0$, we obtain for $x > 1$

$$\begin{aligned} \bar{F}_t(x) &:= \mathbb{P}\left(\frac{X}{t} > x | X > t\right) \\ &= \frac{\mathbb{P}(X > tx)}{\mathbb{P}(X > t)} \\ &= \frac{(tx)^{-\alpha} \ell(tx) e^{-(\beta tx)^\tau}}{t^{-\alpha} \ell(t) e^{-(\beta t)^\tau}} \\ &= x^{-\alpha} \frac{\ell(tx)}{\ell(t)} e^{-(\beta t)^\tau (x^\tau - 1)}. \end{aligned}$$

By definition $\ell(xt)/\ell(t) \approx 1$ for large enough thresholds t . We then assume that at some large values of t , the parameter β is inversely proportional to t , so that a simple Pareto-Weibull model (3.4) provides an appropriate fit to the POTs X/t ($X > t$), at least better than the simple Pareto fit with distribution function $1 - x^{-\alpha}$ as used in classical extreme value methodology for Pareto-type tails. In order to formalize the above, one takes the limit for $t \rightarrow \infty$ which necessarily requires $\beta = \beta_t \downarrow 0$ as $t \uparrow \infty$. The model considered in this paper is then formally given by

(\mathcal{M}) The POT distribution \overline{F}_t satisfies

$$\overline{F}_t(x) \rightarrow \overline{F}_{\alpha, \beta_\infty, \tau}(x) := x^{-\alpha} e^{-\beta_\infty^\tau (x^\tau - 1)}, \text{ as } t \rightarrow \infty \text{ for every } x > 1,$$

where

- a) (**rough tempering**) $\beta = \beta_t$ satisfies $\beta_t t \rightarrow \beta_\infty > 0$, with $\beta_\infty < \infty$, corresponding to the situation where the deviation from the Pareto behavior due to Weibull tempering will be visible in the data from t on and the approximation of the POT distribution using the limit distribution $\overline{F}_{\alpha, \beta_\infty, \tau}$ appears more appropriate than using $\overline{F}_{\alpha, 0, \tau} = x^{-\alpha}$, the simple Pareto distribution;
- b) (**light tempering**) $\beta = \beta_t$ satisfies $\beta_t t \rightarrow 0$, corresponding to

$$\overline{F}_t(x) \rightarrow x^{-\alpha}, \quad x > 1,$$

in which case the tempering is hardly or not visible in the data above t . It will then be practically impossible to discriminate light tempering from no tempering.

In other words, under light tempering the tail index α can be estimated using the traditional methods for Pareto-type tails, see for example Beirlant et al. [36, Ch. 2] or Beirlant et al. [35, Ch. 4] for a thorough description of classical methods available or, for example, Bladt et al. [40] for a recent contribution using a trimmed estimator. In this contribution we only discuss the situation where there is a clear deviation from the pure Pareto behavior, i.e. rough tempering. Nevertheless, as it will become apparent from the simulation studies in Section 3.3, under light-tempering the methods proposed in this contribution provide similar results as the traditional methods for Pareto-type tails.

3.2.1 Statistical estimation

Given a particular threshold t , the quasi-likelihood procedure consists of fitting the limit distribution in (\mathcal{M}) to the POT data

$$\frac{X_j}{t} \text{ when } X_j > t, \quad j = 1, \dots, n.$$

We also use the notation $\lambda = \beta_\infty^\tau$, so that the limit distribution in (\mathcal{M}) is given by

$$\overline{F}_{\alpha, \lambda, \tau}(x) = x^{-\alpha} e^{-\lambda(x^\tau - 1)}, \quad x > 1.$$

The log-likelihood is then given by

$$\begin{aligned} \log L(\alpha, \lambda, \tau) = & -(1 + \alpha) \sum_{j=1}^n \log \left(\frac{X_j}{t} \right) 1_{(X_j > t)} - \lambda \sum_{j=1}^n \left(\left(\frac{X_j}{t} \right)^\tau - 1 \right) 1_{(X_j > t)} \\ & + \sum_{j=1}^n \log \left(\alpha + \lambda \tau \left(\frac{X_j}{t} \right)^\tau \right) 1_{(X_j > t)}. \end{aligned} \quad (3.7)$$

In extreme value methodology the choice of a threshold t is an important matter. A common practice is to select the $(k+1)$ -largest observation $x_{n-k,n}$ for some $k \in \{4, \dots, n-1\}$ as the threshold t , and to plot the resulting estimates as a function of the inverse rank k . Many authors then suggest to find k in a stable portion of these plots, if available. Data-driven choices of k are sometimes available minimizing the asymptotic mean-squared error based on asymptotic results that describe the bias and variance for intermediate k sequences. While an asymptotic result is presented below in Theorem 2.1, we here present an approach focusing on the goodness-of-fit of the tempering model to the POT data above the different thresholds $x_{n-k,n}$, using a QQ-plot approach. Then, for a given value of τ , one finds the least-squares line that minimizes

$$\left(-\log \left(1 - \hat{F}_k(V_{j,k}) \right), \alpha \log V_{j,k} + \tau \beta_\infty^\tau h_\tau(V_{j,k}) \right), \quad j = 1, \dots, k, \quad (3.8)$$

with $h_\tau(x) = (x^\tau - 1)/\tau$, the POT data $V_{j,k} = X_{n-j+1,n}/X_{n-k,n}$, $j = 1, \dots, k$, and \hat{F}_k denoting the empirical distribution function based on those POTs. Therefore, since $\hat{F}_k(V_{j,k}) = \frac{j}{k+1}$, one is led to minimize

$$WLS(V_{j,k}; \alpha_k, \delta_k, \tau_k) := \sum_{j=1}^k w_{j,k} \left(\frac{1}{\alpha} \log \frac{k+1}{k-j+1} - \log V_{j,k} - \delta h_\tau(V_{j,k}) \right)^2, \quad (3.9)$$

with respect to α and $\delta = \tau \beta_\infty^\tau$, where $\{w_{j,k}, j = 1, \dots, k\}$ are appropriate weights. In particular, if $w_{j,k} = 1/\log \left(\frac{k+1}{k-j+1} \right)$ when $\delta \downarrow 0$, that is without tempering, we recover the classical Hill estimator $H_{k,n}$.

Optimization using (3.9) also leads to an adaptive selection method for choosing k which gives appropriate estimates for $(\alpha, \tau, \beta_\infty)$, choosing the k for which the weighted least-squares (WLS) value is minimal:

$$\hat{k} = \arg \min_k SS_k \quad (3.10)$$

with

$$SS_k = \sum_{j=1}^k \frac{1}{\log \left(\frac{k+1}{k-j+1} \right)} \left(\frac{1}{\hat{\alpha}_k^W} \log \left(\frac{k+1}{k-j+1} \right) - \log V_{j,k} - \hat{\delta}_k^W h_{\hat{\tau}_k^W}(V_{j,k}) \right)^2. \quad (3.11)$$

Since for $\tau \rightarrow 0$ the parameters α and τ become non-identifiable, numerical issues will arise during the statistical estimation procedure when directly optimizing

Algorithm 1 Estimation of $(\hat{\alpha}_k^W, \hat{\beta}_{\infty,k}^W, \hat{\tau}_k^W)$ and $(\hat{\alpha}_k^M, \hat{\beta}_{\infty,k}^M, \hat{\tau}_k^M)$

- 1: **set** $\tilde{\tau}_1 < \tilde{\tau}_2 < \dots < \tilde{\tau}_m, m \in \mathbb{N}$
 - 2: **for** $k = 1, 2, \dots, n - 1$ **do**
 - 3: **for** $i = 1, 2, \dots, m$ **do**
 - 4: **Optimization step.** Set

$$\left(\hat{\alpha}_{k,\tilde{\tau}_i}, \hat{\delta}_{k,\tilde{\tau}_i} \right) := \arg \min_{(\alpha > 0, \delta > 0)} WLS(V_{j,k}; \alpha, \delta, \tilde{\tau}_i)$$
 - 5: $\widehat{WLS}_{k,\tilde{\tau}_i} \leftarrow WLS(V_{j,k}; \hat{\alpha}_{k,\tilde{\tau}_i}, \hat{\delta}_{k,\tilde{\tau}_i}, \tilde{\tau}_i)$
 - 6: $\hat{\lambda}_{k,\tilde{\tau}_i} \leftarrow \hat{\delta}_{k,\tilde{\tau}_i} / \tilde{\tau}_i$
 - 7: $\widehat{\log L}_{k,\tilde{\tau}_i} \leftarrow \log L(V_{j,k}; \hat{\alpha}_{k,\tilde{\tau}_i}, \hat{\lambda}_{k,\tilde{\tau}_i}, \tilde{\tau}_i)$
 - 8: **Set**

$$\left(\hat{\alpha}_k^W, \hat{\delta}_k^W, \hat{\tau}_k^W \right) := \arg \min_{(\hat{\alpha}_k, \hat{\delta}_k, \hat{\tau}_k)} \left\{ \widehat{WLS}_{k,\tilde{\tau}_i}; i = 1, \dots, m \right\}$$
 - 9: $\hat{\beta}_{\infty,k}^W \leftarrow \left(\hat{\delta}_k^W / \hat{\tau}_k^W \right)^{1/\hat{\tau}_k^W}$
 - 10: **Set**

$$\left(\hat{\alpha}_k^M, \hat{\lambda}_k^M, \hat{\tau}_k^M \right) := \arg \max_{(\hat{\alpha}_k, \hat{\lambda}_k, \hat{\tau}_k)} \left\{ \widehat{\log L}_{k,\tilde{\tau}_i}; i = 1, \dots, m \right\}$$
 - 11: $\hat{\beta}_{\infty,k}^M \leftarrow \left(\hat{\lambda}_k^M \right)^{1/\hat{\tau}_k^M}$
 - 12: **return** $(\hat{\alpha}_k^W, \hat{\beta}_{\infty,k}^W, \hat{\tau}_k^W)$ and $(\hat{\alpha}_k^M, \hat{\beta}_{\infty,k}^M, \hat{\tau}_k^M)$, for $k = 1, 2, \dots, n - 1$.
-

the likelihood, or when minimizing (3.9). However, fixing a value of τ during the calibration procedure reduces numerical instabilities. The optimization procedure Algorithm 1 which is used in the simulations and cases, leads to WLS estimates $(\hat{\alpha}_k^W, \hat{\beta}_{\infty,k}^W, \hat{\tau}_k^W)$ and maximum likelihood estimates $(\hat{\alpha}_k^M, \hat{\beta}_{\infty,k}^M, \hat{\tau}_k^M)$, starting from a grid of m initial τ values $\tilde{\tau}_1 < \tilde{\tau}_2 < \dots < \tilde{\tau}_m, m \in \mathbb{N}$.

3.2.2 Return periods and extreme quantiles

In order to estimate return periods of the type $1/\mathbb{P}(X > z)$ for some large outcome level z , we use the approximation

$$\frac{\mathbb{P}(X > tx)}{\mathbb{P}(X > t)} \approx x^{-\alpha} e^{-\lambda \tau h_{\tau}(x)}$$

with t large, so that setting $tx = z$ and $t = x_{n-k,n}$ for some k , we obtain the estimators for $\mathbb{P}(X > z)$

$$\hat{P}_{z,k}^W = \frac{k+1}{n+1} \left(\frac{z}{x_{n-k,n}} \right)^{-\hat{\alpha}_k^W} \exp \left(-\hat{\lambda}_k^W \hat{\tau}_k^W h_{\hat{\tau}_k^W}(z/x_{n-k,n}) \right) \quad (3.12)$$

and similarly $\hat{P}_{z,k}^M$, where $\mathbb{P}(X > t) = \mathbb{P}(X > x_{n-k,n})$ is estimated using the empirical proportion $(k+1)/(n+1)$.

The value z solving the equation

$$\frac{k+1}{n+1} \left(\frac{z}{x_{n-k,n}} \right)^{-\hat{\alpha}_k^W} \exp \left(-\hat{\lambda}_k^W \hat{\tau}_k^W h_{\hat{\tau}_k^W}(z/x_{n-k,n}) \right) = p, \quad (3.13)$$

for a given value $p \leq \frac{1}{n}$, then yields an estimator $\hat{Q}_{p,k}^W$ for the extreme quantile or return level $Q(1-p)$, and hence for Value-at-Risk (VaR_p) at extreme quantile levels $1-p$. Similarly, one obtains the estimator $\hat{Q}_{p,k}^M$.

3.2.3 Asymptotic distribution of the maximum likelihood estimators

We end this section stating the asymptotic distribution of the maximum likelihood estimators $\hat{\alpha}_t, \hat{\lambda}_t, \hat{\tau}_t$. The likelihood equations in (α, λ, τ) are given by

$$\begin{aligned} \sum_{j=1}^n \left\{ \alpha + \lambda \tau \left(\frac{X_j}{t} \right)^\tau \right\}^{-1} 1_{(X_j > t)} &= \sum_{j=1}^n \log \left(\frac{X_j}{t} \right) 1_{(X_j > t)}, \\ \sum_{j=1}^n \frac{\left(\frac{X_j}{t} \right)^\tau}{\alpha + \lambda \tau \left(\frac{X_j}{t} \right)^\tau} 1_{(X_j > t)} &= \sum_{j=1}^n h_\tau \left(\frac{X_j}{t} \right) 1_{(X_j > t)}, \\ \sum_{j=1}^n \frac{\left(\frac{X_j}{t} \right)^\tau \log \left(\frac{X_j}{t} \right)}{\alpha + \lambda \tau \left(\frac{X_j}{t} \right)^\tau} 1_{(X_j > t)} &= \sum_{j=1}^n \left(\frac{X_j}{t} \right)^\tau \log \left(\frac{X_j}{t} \right) 1_{(X_j > t)}. \end{aligned}$$

We further assume classical second-order slow variation

$$\frac{\ell(ty)}{\ell(t)} = 1 + Dt^\rho h_\rho(y), \quad \text{with } D \in \mathbb{R}, \rho < 0, \quad (3.14)$$

and set $\hat{\boldsymbol{\theta}}_t = (\hat{\alpha}_t, \hat{\lambda}_t, \hat{\tau}_t)^t$ and $\boldsymbol{\theta} = (\alpha, \lambda, \tau)^t$.

Theorem 3.2.1. *Under $\bar{F}(x) = x^{-\alpha} \ell(x) e^{-\beta x^\tau}$ satisfying (\mathcal{M}) with $\beta_\infty > 0$ and ℓ satisfying (3.14), we have as $n, t \rightarrow \infty$ such that $n\bar{F}(t) \rightarrow \infty$ and $\sqrt{n\bar{F}(t)} t^\rho \rightarrow \nu > 0$ that*

$$\sqrt{n\bar{F}(t)} \left(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta} \right) \rightarrow_d \mathcal{N}_3 \left(D\nu \mathbf{I}^{-1} \mathbf{b}, \mathbf{I}^{-1} \right)$$

with $\mathbf{I} \in \mathbb{R}^{3 \times 3}$ symmetric and $\mathbf{b} \in \mathbb{R}^{3 \times 1}$ and

$$\begin{aligned}
I_{1,1} &= \int_1^\infty \frac{u^{-\alpha-1} e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau} du, \\
I_{2,2} &= \tau^2 \int_1^\infty \frac{u^{2\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau} du, \\
I_{3,3} &= \lambda \int_1^\infty \left\{ \log^2(u) - \frac{2 \log u}{\alpha + \lambda\tau u^\tau} \right. \\
&\quad \left. + \frac{\lambda u^\tau (1 + 2\tau \log u) - \alpha\tau (\log u)^2}{(\alpha + \lambda\tau u^\tau)^2} \right\} u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)} (\alpha + \lambda\tau u^\tau) du, \\
I_{1,2} &= \tau \int_1^\infty \frac{u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau} du, \\
I_{1,3} &= \lambda \int_1^\infty (1 + \tau \log u) \frac{u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)}}{\alpha + \lambda\tau u^\tau} du, \\
I_{2,3} &= \int_1^\infty \left\{ \log u - \frac{\alpha(1 + \tau \log u)}{(\alpha + \lambda\tau u^\tau)^2} \right\} u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)} (\alpha + \lambda\tau u^\tau) du, \\
b_1 &= \int_1^\infty \left(\frac{1}{\alpha + \lambda\tau u^\tau} - \log u \right) u^{-\alpha-1} e^{-\lambda\tau h_\tau(u)} [h_\rho(u)(\alpha + \lambda\tau u^\tau) - u^\rho] du, \\
b_2 &= \int_1^\infty \left(\frac{\tau u^\tau}{\alpha + \lambda\tau u^\tau} - \tau h_\tau(u) \right) u^{-\alpha-1} e^{-\lambda\tau h_\tau(u)} [h_\rho(u)(\alpha + \lambda\tau u^\tau) - u^\rho] du, \\
b_3 &= \lambda \int_1^\infty \left(\frac{1 + \tau \log u}{\alpha + \lambda\tau u^\tau} - \log u \right) u^{\tau-\alpha-1} e^{-\lambda\tau h_\tau(u)} [h_\rho(u)(\alpha + \lambda\tau u^\tau) - u^\rho] du.
\end{aligned}$$

The derivation of this result is postponed to the Appendix A.

3.3 Simulation results

The finite sample behavior of the estimators $(\hat{\alpha}_k^W, \hat{\tau}_k^W)$ and $(\hat{\alpha}_k^M, \hat{\tau}_k^M)$ and the resulting tail probabilities $\hat{P}_{z,k}^W, \hat{P}_{z,k}^M$ and extreme quantiles $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ resulting from Algorithm 1, (3.12) and (3.13), respectively, have been studied through an extensive Monte Carlo simulation procedure. For each setting, 500 runs with sample size $n = 500$ were performed. The mean and root mean squared error (RMSE) of the estimators are presented for the following models:

(a) Burr-Weibull(α, ξ, τ, β) model with Burr survival function is given by

$$1 - F_Y(y) = (1 + y^{-\xi\alpha})^{1/\xi}, \quad y > 0, \quad \alpha > 0, \quad \xi < 0.$$

Here (3.14) is satisfied with $\rho = \xi\alpha$. We used $(\alpha, \xi, \tau, \beta) = (2, -1, 1.50, 0.50)$ and $(2, -1, 0.50, 0.20)$.

- (b) Fréchet-Weibull(α, τ, β) model with the Fréchet survival function

$$1 - F_Y(y) = \exp(-y^{-\alpha}), \quad y > 0, \quad \alpha > 0.$$

Here (3.14) is satisfied with $\rho = -\alpha$. We used $(\alpha, \tau, \beta) = (2, 2, 0.50)$ and $(2, 0.50, 0.20)$.

- (c) Pareto-Weibull(α, τ, β) model using the Pareto distribution

$$1 - F_Y(y) = y^{-\alpha}, \quad y > 1, \quad \alpha > 0.$$

Here $\ell(x) = 1$. We used $(\alpha, \tau, \beta) = (1, 2, 0.20)$.

- (d) In order to study the behavior of the estimators under Weibull tempering of a heavy-tailed distribution outside the Pareto-type family we simulated from a tempered log-normal distribution with parameters $\mu = 0$ and $\sigma = 10$.

The plots of these simulation studies are collected in Appendix B.

In the plots concerning the estimation of α , we also plot the results for the Hill estimator $H_{k,n}$, while in case of the tail quantile estimates $\hat{Q}_{p,k}^W$ and $\hat{Q}_{p,k}^M$ we also provide the results for the Weissman [181] estimator $\hat{Q}_{p,k}^H = X_{n-k,n} \left(\frac{k}{np}\right)^{1/\hat{\alpha}_k^H}$. Finally, we also present the boxplots of the estimates when using the adaptive choice \hat{k} given in (3.10) for k . The characteristics for the tail probability estimators $\hat{P}_{z,k}^W, \hat{P}_{z,k}^M$ are quite comparable to those of the extreme quantiles, and are omitted here.

From the mean and RMSE plots one notes that the results for the MLE results $\hat{\alpha}^M, \hat{\tau}^M$ and \hat{Q}_p^M improve upon the weighted least squares-based results. Namely, in most of the analyzed cases the MLE results show a smaller bias and RMSE when compared to the weighted least squares based results. The results with the adaptive choice \hat{k} of k are promising, and again best for the MLE results. In case $\tau > 1$ (see Figures 3.4, 3.5, 3.8, 3.9, 3.12, 3.13, 3.14 and 3.15) when the tempering is quite strong, the results for the proposed methods improve upon the classical estimators $H_{k,n}$ and $\hat{Q}_{p,k}^H$, which ignore the tempering in the simulated samples. Note that in these cases the VaR estimates based on the MLE parameters taken at the adaptive value \hat{k} show a rather small bias, even in case of the log-normal model which is situated outside our Pareto-type model assumption.

In case $\tau < 1$ (see Figures 3.6, 3.7, 3.10 and 3.11), hence under weaker tempering, the bias and RMSE results are comparable with the classical estimators. The VaR estimates at \hat{k} tend to overestimate the correct value. As will become clear from the case studies in the next section, the Pareto and tempered Pareto fits can lead to quite different extreme tail fits *per sample*. Namely, even though the performance of the methods presented here and the classical methods are similar under weaker tempering, ignoring the tempering leads to overfitting tail probabilities and quantiles.

We conclude that the use of classical estimators ignoring the tempering effect leads to overestimation of the risk measures, while the proposed method provides reasonable

VaR estimates especially for larger values of $\tau > 1$. In case of smaller tempering with a heavier Weibull tail, improvements can be made concerning the adaptive choice of k . Another possibility is to search for bias reduced estimators as available in the non-tempering literature (see, for instance, Chapters 3 and 4 in [35]).

3.4 Insurance cases

We now apply the presented methods to the Norwegian and the Secura Re Belgian data sets introduced in Section 1. In addition, we contrast the tail index estimates $\hat{\alpha}_{k,n}^W$ and $\hat{\alpha}_{k,n}^M$ with the values obtained for the truncated Pareto-type model proposed in Beirlant et al. [33], where $\hat{\alpha}_{k,n}^T$ is obtained as the solution to

$$H_{k,n} = \frac{1}{\alpha_{k,n}^T} + \frac{R_{k,n}^{\alpha_{k,n}^T} \log(R_{k,n})}{1 - R_{k,n}^{\alpha_{k,n}^T}},$$

with $R_{k,n} = X_{n-k,n}/X_{n,n}$. The latter estimator was first proposed in Aban et al. [1] as the conditional MLE based on the $k + 1$ ($0 \leq k < n$) largest order statistics representing only the portion of the tail where the truncated Pareto approximation holds, see also [7, Sec 4.2.3].

We then also measure the goodness-of-fit using QQ-plot (3.8) and the analogous expression for the truncated model. The results are discussed here and the figures are presented in Appendix C.

For the Norwegian fire insurance data set, we find $\hat{k} = 4920$ from the plot of SS_k from (3.11) in Figure 3.16, where also the different parameter estimates as a function of k can be found. The log-log plot based on (3.8) at $k = 4920$ shows a good tail fit for the tempered Pareto model, in contrast to the simple Pareto fit which will overfit tail probabilities and quantiles. This can be seen from Figure 3.17 where for larger k , the classical Weissman estimates $\hat{Q}_{1/(cn),k}^H$ ($c = 1, 2$) lead to much larger estimates than those based on the proposed tempering modeling. Only when k is really small, that is, when restricting to the data situated in the bottom curved area of the log-log plot, the classical linear Pareto fit is able to provide a reasonable representation of the most extreme data. Finally, note from the log-log plot in Figure 3.16 that the truncated Pareto fit follows the linear Pareto fit except for the two final extreme points after which a sharp deviation is observed up to an estimated finite truncation point T estimated at $\hat{T}_{\hat{k}} = 1, 211, 106$, when using the estimation method proposed in [33, Sec. 3, Eq. 19].

In order to illustrate the possibility of extending the proposed method in a time-dependent regression context, we fitted the approach to 3-years sliding time windows. The size of the windows was selected to have at least 300 observations at each point in time. Figure 3.18 shows the estimated VaR at 99.5% (top) and 99.9% (bottom) using the tempered Pareto approach with \hat{k} selected using the proposed adaptive procedure, next to simple Pareto and truncated Pareto modeling. We also compare with the observed quantiles obtained using the standard R function, which estimates

the quantiles as weighted averages of consecutive order statistics. The VaR values based on the tempered Pareto model are situated between the observed and the Pareto and truncated Pareto fits, from which one can conclude that the tempered tail behavior observed for the complete data set in the bottom frame in Figure 3.16 is also present conditional on a time window, leading to overestimation when using classical methods that ignore the proposed tempering. It is also worth noticing that the VaR at 99.5% values exhibits an overall decreasing trend with some stable behavior between 1979 and 1987. Figures 3.19 and 3.20 show the respective VaR estimates for all values of k for some selected time windows.

In Figure 3.21, the respective results are given for the Secura Re Belgium data set. Here the best tempered Pareto fit is found at $\hat{k} = 147$, with the corresponding log-log plot given in the bottom figure. Here the tempered Pareto WLS fit closely follows the linear Pareto fit, while the MLE fit shows too much bending near the largest data. Both the Pareto and WLS tempered Pareto fit do miss the deviation at the top two data, which however is taken into account in the truncated Pareto analysis with $\hat{T}_{\hat{k}} = 8,967,620 = e^{16.009}$. While this deviation can be considered as statistically non-significant, it makes sense to consider the truncated Pareto fit here since Belgian car insurance contracts do show explicit upper limits. Another motivation for a truncated model is that the extreme quantile estimates $\hat{Q}_{1/(cn),\hat{k}}^M$ hardly change from $c = 1$ to $c = 2$, namely around the value e^{16} .

3.5 Conclusion

In this paper we addressed the fitting of Pareto-type distributions with a tempering component of Weibull type at large values. We extend earlier results for exponential tempering on strict Pareto tails, provide a Peaks over Threshold (POT) approach, develop estimation procedures and provide asymptotic properties of the proposed estimators. Finally, we present a simulation study and also apply the developed methods to actual insurance data, discussing challenges in the implementation and how to overcome them. The estimation of VaR values at extreme quantile levels shows improvements compared to more classical extreme value estimation methods that ignore the considered tempering effect. These improvements are more pronounced with growing tempering effect.

Further research concerning the generalization to a regression context and the use of tempered Pareto-Weibull models in composed or splicing models will be taken up in the future.

3.6 Appendix A: Proof of Theorem 3.2.1

Using Taylor expansions of the likelihood equations in $\hat{\boldsymbol{\theta}}_t$ around the correct value $\boldsymbol{\theta}$ leads to the following system of three equations, with $\boldsymbol{\theta} = (\tilde{\alpha}, \tilde{\lambda}, \tilde{\tau})$ situated in between $\hat{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}$:

$$\begin{aligned}
& \sqrt{n\bar{F}(t)}(\hat{\alpha}_t - \alpha) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \frac{1}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)} \\
& + \sqrt{n\bar{F}(t)}(\hat{\lambda}_t - \lambda) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \frac{\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)} \\
& + \sqrt{n\bar{F}(t)}(\hat{\tau}_t - \tau) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \frac{\tilde{\lambda}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\left(1 + \tilde{\tau}\log\frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)} \\
& = \sqrt{n\bar{F}(t)} \left(\frac{1}{n\bar{F}(t)} \sum_{j=1}^n \left\{ \frac{1}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \log\frac{X_j}{t} \right\} 1_{(X_j > t)} \right) \tag{3.15}
\end{aligned}$$

$$\begin{aligned}
& \sqrt{n\bar{F}(t)}(\hat{\alpha}_t - \alpha) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \frac{\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)} \\
& + \sqrt{n\bar{F}(t)}(\hat{\lambda}_t - \lambda) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \frac{\tilde{\tau}^2\left(\frac{X_j}{t}\right)^{2\tilde{\tau}}}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)} \\
& + \sqrt{n\bar{F}(t)}(\hat{\tau}_t - \tau) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \left(\frac{\tilde{\alpha}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\left(1 + \tilde{\tau}\log\frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} - \left(\frac{X_j}{t}\right)^{\tilde{\tau}}\log\frac{X_j}{t} \right) 1_{(X_j > t)} \\
& = \sqrt{n\bar{F}(t)} \left(\frac{1}{n\bar{F}(t)} \sum_{j=1}^n \left\{ \frac{\tau\left(\frac{X_j}{t}\right)^\tau}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \left(\frac{X_j}{t}\right)^\tau + 1 \right\} 1_{(X_j > t)} \right) \tag{3.16}
\end{aligned}$$

$$\begin{aligned}
& \sqrt{n\bar{F}(t)}(\hat{\alpha}_t - \alpha) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \frac{\tilde{\lambda}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\left(1 + \tilde{\tau}\log\frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} 1_{(X_j > t)} \\
& + \sqrt{n\bar{F}(t)}(\hat{\lambda}_t - \lambda) \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \left(\frac{\tilde{\alpha}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\left(1 + \tilde{\tau}\log\frac{X_j}{t}\right)}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} - \left(\frac{X_j}{t}\right)^{\tilde{\tau}}\log\frac{X_j}{t} \right) 1_{(X_j > t)} \\
& + \sqrt{n\bar{F}(t)}(\hat{\tau}_t - \tau) \frac{\tilde{\lambda}}{n\bar{F}(t)} \sum_{j=1}^n \left(\frac{\tilde{\lambda}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\left(1 + 2\tilde{\tau}\log\frac{X_j}{t}\right) + \tilde{\alpha}\tilde{\tau}\left(\log\frac{X_j}{t}\right)^2}{\left(\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}\right)^2} \right. \\
& \quad \left. - \frac{2\log\frac{X_j}{t}}{\tilde{\alpha} + \tilde{\lambda}\tilde{\tau}\left(\frac{X_j}{t}\right)^{\tilde{\tau}}} + \left(\log\frac{X_j}{t}\right)^2 \right) \left(\frac{X_j}{t}\right)^{\tilde{\tau}} 1_{(X_j > t)} \\
& = \sqrt{n\bar{F}(t)} \left(\frac{\lambda}{n\bar{F}(t)} \sum_{j=1}^n \left\{ \frac{\left(\frac{X_j}{t}\right)^\tau\left(1 + \tau\log\frac{X_j}{t}\right)}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \left(\frac{X_j}{t}\right)^\tau\log\frac{X_j}{t} \right\} 1_{(X_j > t)} \right) \tag{3.17}
\end{aligned}$$

The coefficients of $\sqrt{n\bar{F}(t)}(\hat{\alpha}_t - \alpha)$, $\sqrt{n\bar{F}(t)}(\hat{\lambda}_t - \lambda)$ and $\sqrt{n\bar{F}(t)}(\hat{\tau}_t - \tau)$ on the left hand sides of (3.15), (3.16) and (3.17) now converge in probability to the corresponding elements of \mathbf{I} . For instance for

$$\mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau) := \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \frac{1}{\left(\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau\right)^2} \mathbf{1}_{(X_j > t)}$$

we have

$$\begin{aligned} \mathbb{E}(\mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau)) &= - \int_t^\infty \frac{1}{\left(\alpha + \lambda\tau\left(\frac{x}{t}\right)^\tau\right)^2} d\frac{\bar{F}(x)}{\bar{F}(t)} \\ &= - \int_1^\infty \frac{1}{\left(\alpha + \lambda\tau u^\tau\right)^2} d\bar{F}_t(u) \\ &\rightarrow - \int_1^\infty \frac{1}{\left(\alpha + \lambda\tau u^\tau\right)^2} d\bar{F}_{\alpha,\lambda,\tau}(u) = I_{1,1}, \end{aligned}$$

as $t \rightarrow \infty$ using the consistency of ML estimators and assumption (\mathcal{M}) . The convergence of $\mathbf{I}_{1,1,n,t}(\tilde{\alpha}, \tilde{\lambda}, \tilde{\tau})$ to $\mathbf{I}_{1,1}$ then follows from

$$\text{Var}(\mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau)) = O\left((n\bar{F}(t))^{-1}\right) \text{ and } \mathbf{I}_{1,1,n,t}(\tilde{\alpha}, \tilde{\lambda}, \tilde{\tau}) - \mathbf{I}_{1,1,n,t}(\alpha, \lambda, \tau) = o_p(1)$$

as $n, t \rightarrow \infty$ using the consistency of the ML estimators.

Next the asymptotic normal distribution of the right hand sides of (3.15)-(3.17)

$$\begin{aligned} &\sqrt{n\bar{F}(t)} \left(\frac{1}{n\bar{F}(t)} \sum_{j=1}^n \left\{ \frac{1}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \log \frac{X_j}{t} \right\} \mathbf{1}_{(X_j > t)} \right. \\ &\quad , \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \left\{ \frac{\tau\left(\frac{X_j}{t}\right)^\tau}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \left(\frac{X_j}{t}\right)^\tau + 1 \right\} \mathbf{1}_{(X_j > t)} \\ &\quad \left. , \frac{\lambda}{n\bar{F}(t)} \sum_{j=1}^n \left\{ \frac{\left(\frac{X_j}{t}\right)^\tau (1 + \tau \log \frac{X_j}{t})}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \left(\frac{X_j}{t}\right)^\tau \log \frac{X_j}{t} \right\} \mathbf{1}_{(X_j > t)} \right) \quad (3.18) \end{aligned}$$

is derived.

Concerning the first component

$$\begin{aligned} \frac{1}{n\bar{F}(t)} \mathbb{E} \left(\sum_{j=1}^n \left\{ \frac{1}{\alpha + \lambda\tau\left(\frac{X_j}{t}\right)^\tau} - \log \frac{X_j}{t} \right\} \mathbf{1}_{(X_j > t)} \right) &= - \frac{1}{\bar{F}(t)} \int_t^\infty \left\{ \frac{1}{\alpha + \beta_\infty \frac{x}{t}} - \log \left(\frac{x}{t} \right) \right\} d\bar{F}(x) \\ &= - \int_1^\infty \left\{ \frac{1}{\alpha + \lambda\tau u^\tau} - \log u \right\} d\bar{F}_t(u), \end{aligned}$$

with $\bar{F}_t(u) = \mathbb{P}(X/t > u | X > t) = u^{-\alpha} (1 + Dt^\rho h_\rho(u)) e^{-\lambda(u^\tau - 1)}$ using the second order slow variation condition (3.14), so that

$$- \frac{d\bar{F}_t(u)}{du} = u^{-\alpha-1} e^{-\lambda(u^\tau - 1)} (\alpha + \lambda\tau u^\tau) + Dt^\rho u^{-\alpha-1} e^{-\lambda(u^\tau - 1)} \{h_\rho(u)[\alpha + \lambda\tau u^\tau] - u^\rho\}.$$

Using partial integration one easily checks that

$$\int_1^\infty \left\{ \frac{1}{\alpha + \lambda\tau u^\tau} - \log u \right\} u^{-\alpha-1} e^{-\lambda(u^\tau-1)} (\alpha + \lambda\tau u^\tau) du = 0,$$

so that the expected value of the first component is given by $Dt^\rho b_1$, leading to the asymptotic bias expression of $\hat{\alpha}_t$ as given in Theorem 2.1, and similar calculations lead to the bias of $\hat{\lambda}_t$ and $\hat{\tau}_t$.

So it remains to derive the asymptotic variances and covariances of the vector in (3.18). The variance of the first component is derived from

$$\begin{aligned} & \frac{1}{n\bar{F}(t)} \sum_{j=1}^n \mathbb{E} \left\{ \frac{1}{\alpha + \lambda\tau \left(\frac{X_j}{t}\right)^\tau} - \log \frac{X_j}{t} \right\}^2 1_{(X_j > t)} \\ &= \frac{1}{\bar{F}(t)} \mathbb{E} \left(\left\{ \frac{1}{\alpha + \lambda\tau \left(\frac{X}{t}\right)^\tau} - \log \frac{X}{t} \right\}^2 1_{(X > t)} \right) \\ &= - \int_1^\infty \left(\frac{1}{(\alpha + \lambda\tau u^\tau)^2} - \frac{2 \log u}{\alpha + \lambda\tau u^\tau} + (\log u)^2 \right) d\bar{F}_t(u) \\ &\rightarrow - \int_1^\infty \left(\frac{1}{(\alpha + \lambda\tau u^\tau)^2} - \frac{2 \log u}{\alpha + \lambda\tau u^\tau} + (\log u)^2 \right) du^{-\alpha} e^{-\lambda(u^\tau-1)}, \end{aligned}$$

as $n, t \rightarrow \infty$. Using partial integration one finds that $\int_1^\infty \left(\frac{2 \log u}{\alpha + \lambda\tau u^\tau} - (\log u)^2 \right) du^{-\alpha} e^{-\lambda(u^\tau-1)} = 0$, so that the asymptotic variance of the first component in (3.18) equals $\mathbf{I}_{1,1}$. In the same way one finds that the asymptotic variance covariance matrix of (3.18) equals \mathbf{I} .

Hence

$$(\mathbf{I} + o_p(1)) \sqrt{n\bar{F}(t)} (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}) = \mathcal{N}_3((D\nu)\mathbf{b}, \mathbf{I}) + o_p(1), \quad (3.19)$$

from which the result follows.

3.7 Appendix B: Simulation results

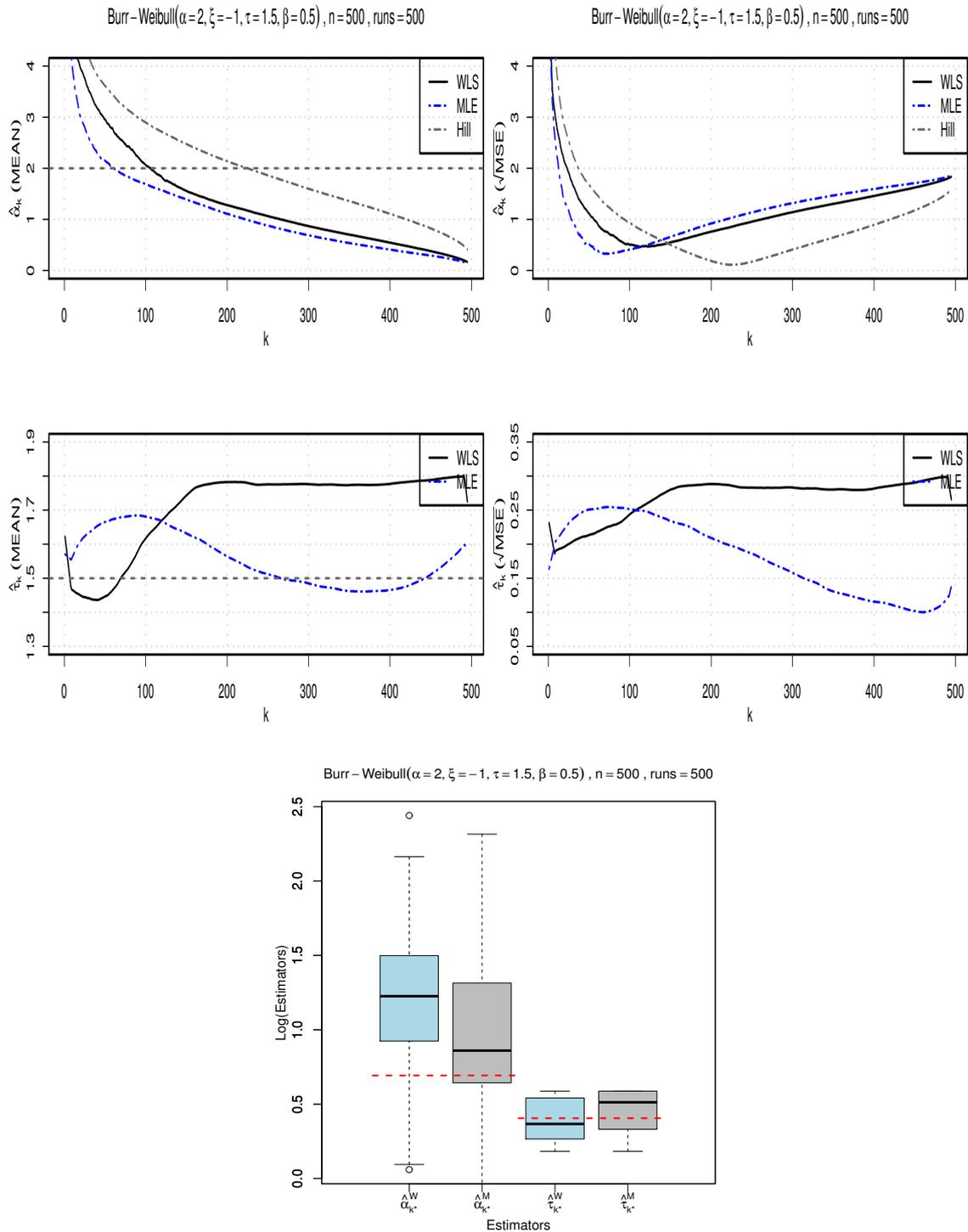


Figure 3.4: Burr-Weibull(2.0, -1.0, 1.5, 0.5). *Top*: Mean (left) and RMSE (right) of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle*: Mean (left) and RMSE (right) of $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ as a function of k ; *Bottom*: Boxplots of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

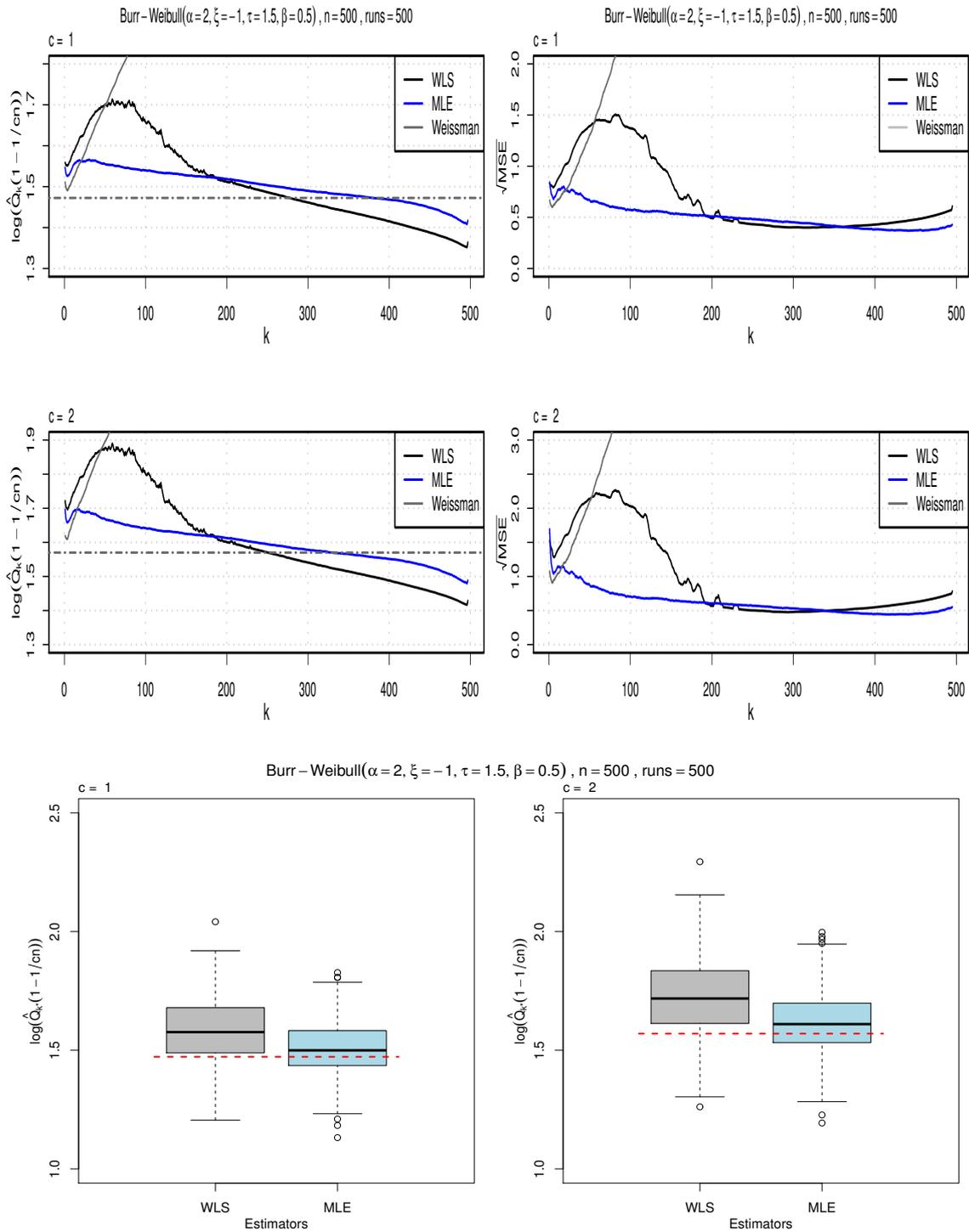


Figure 3.5: Burr-Weibull(2.0, -1.0, 1.5, 0.5): quantile estimates $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k . Bottom line: boxplots of $\hat{Q}_{p,\hat{k}}^W, \hat{Q}_{p,\hat{k}}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

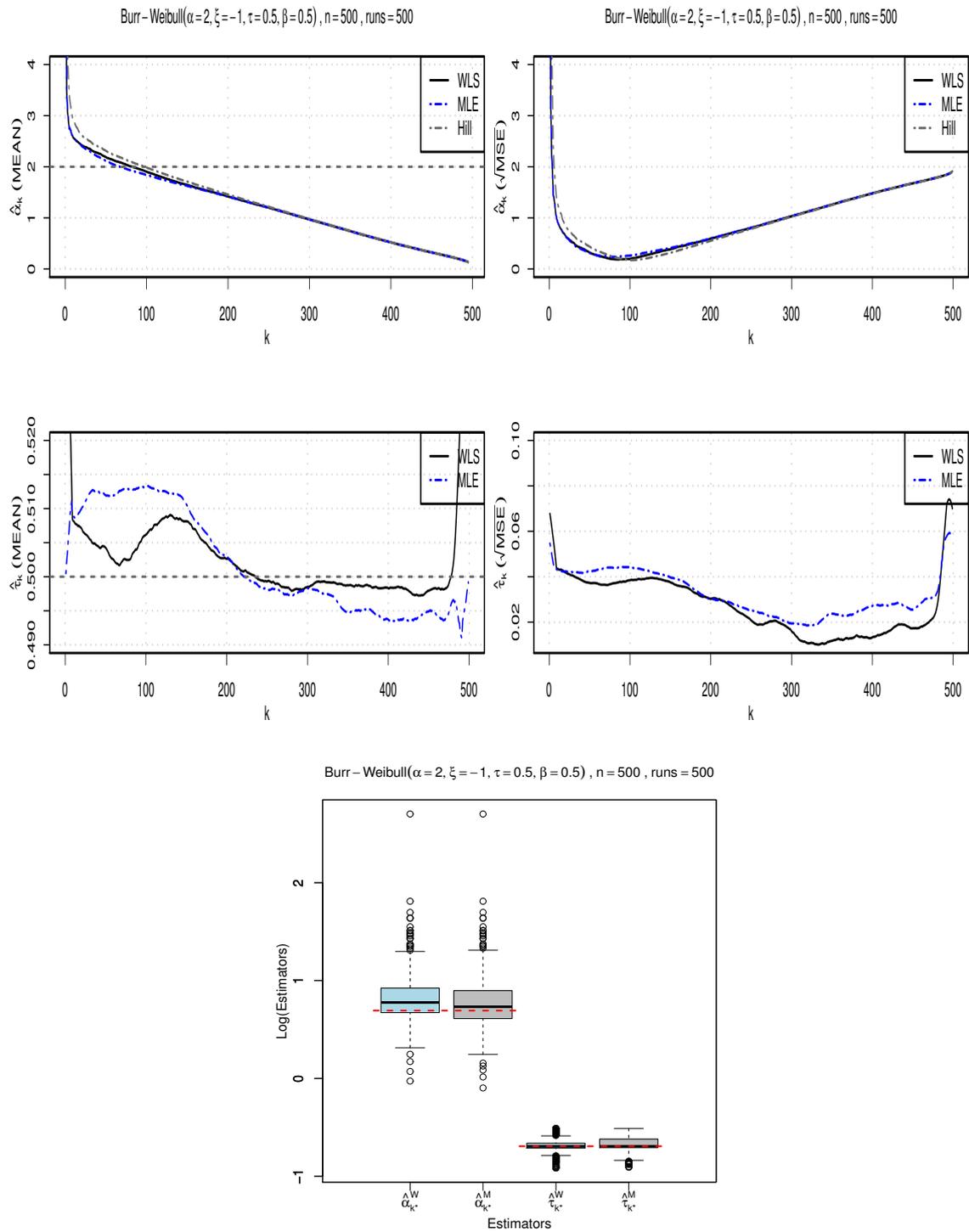


Figure 3.6: Burr-Weibull(2.0, -1.0, 0.5, 0.5). *Top:* Mean (left) and RMSE (right) of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle:* Mean (left) and RMSE (right) of $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ as a function of k ; *Bottom:* Boxplots of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

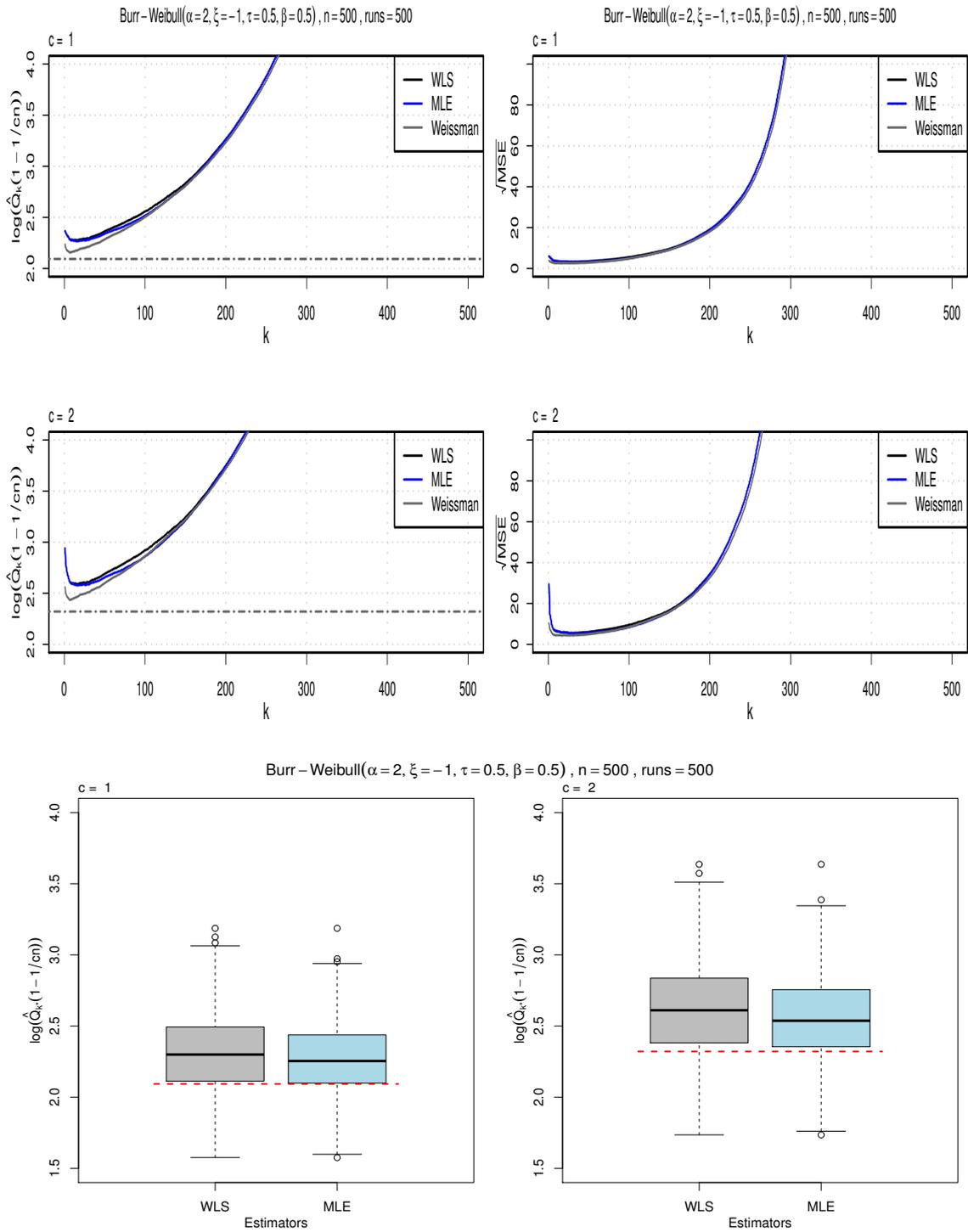


Figure 3.7: Burr-Weibull(2.0, -1.0, 0.5, 0.5): quantile estimates $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k . Bottom line: boxplots of $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

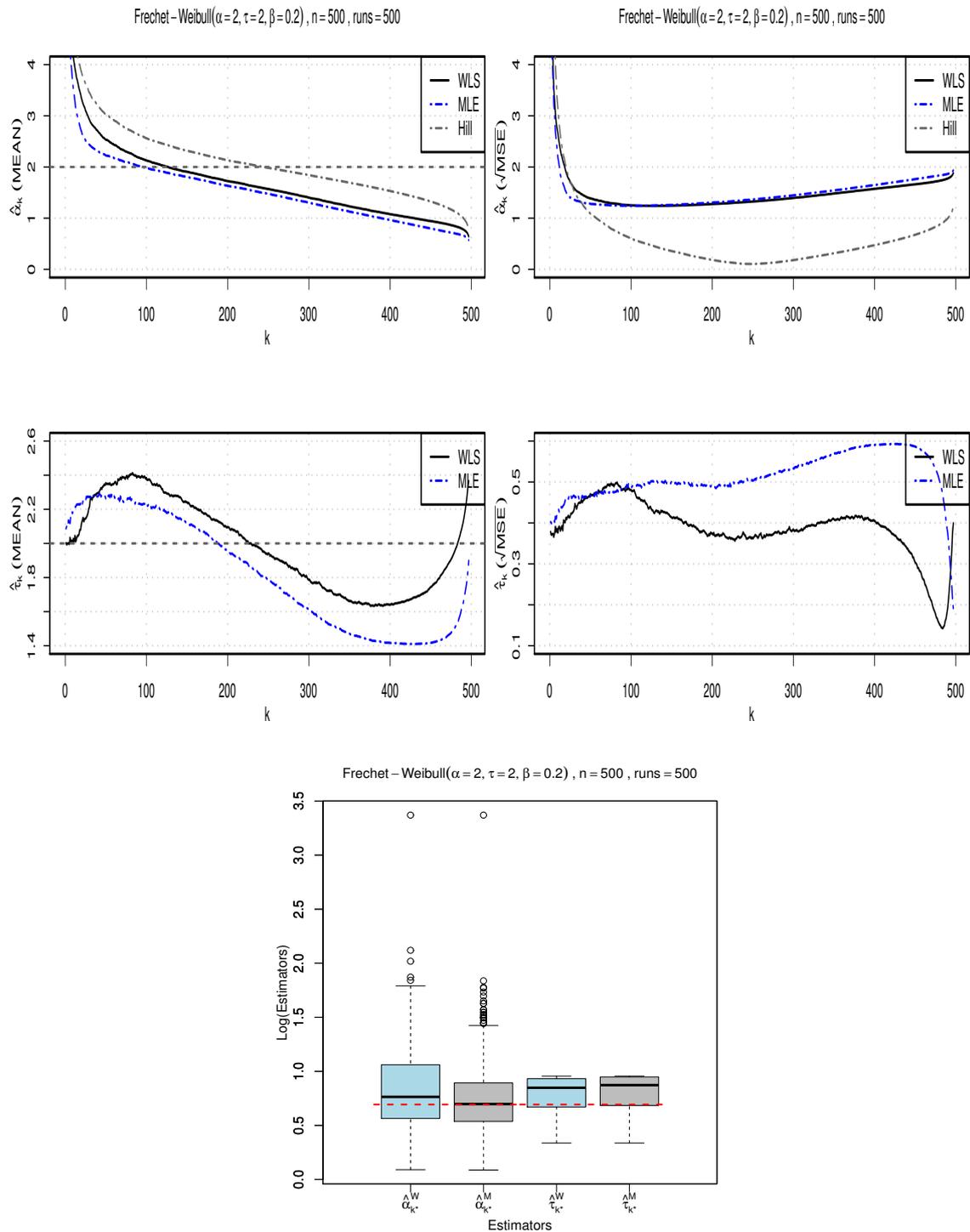


Figure 3.8: Fréchet-Weibull(2.0, 2.0, 0.2). *Top:* Mean (left) and RMSE (right) of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle:* Mean (left) and RMSE (right) of $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ as a function of k ; *Bottom:* Boxplots of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

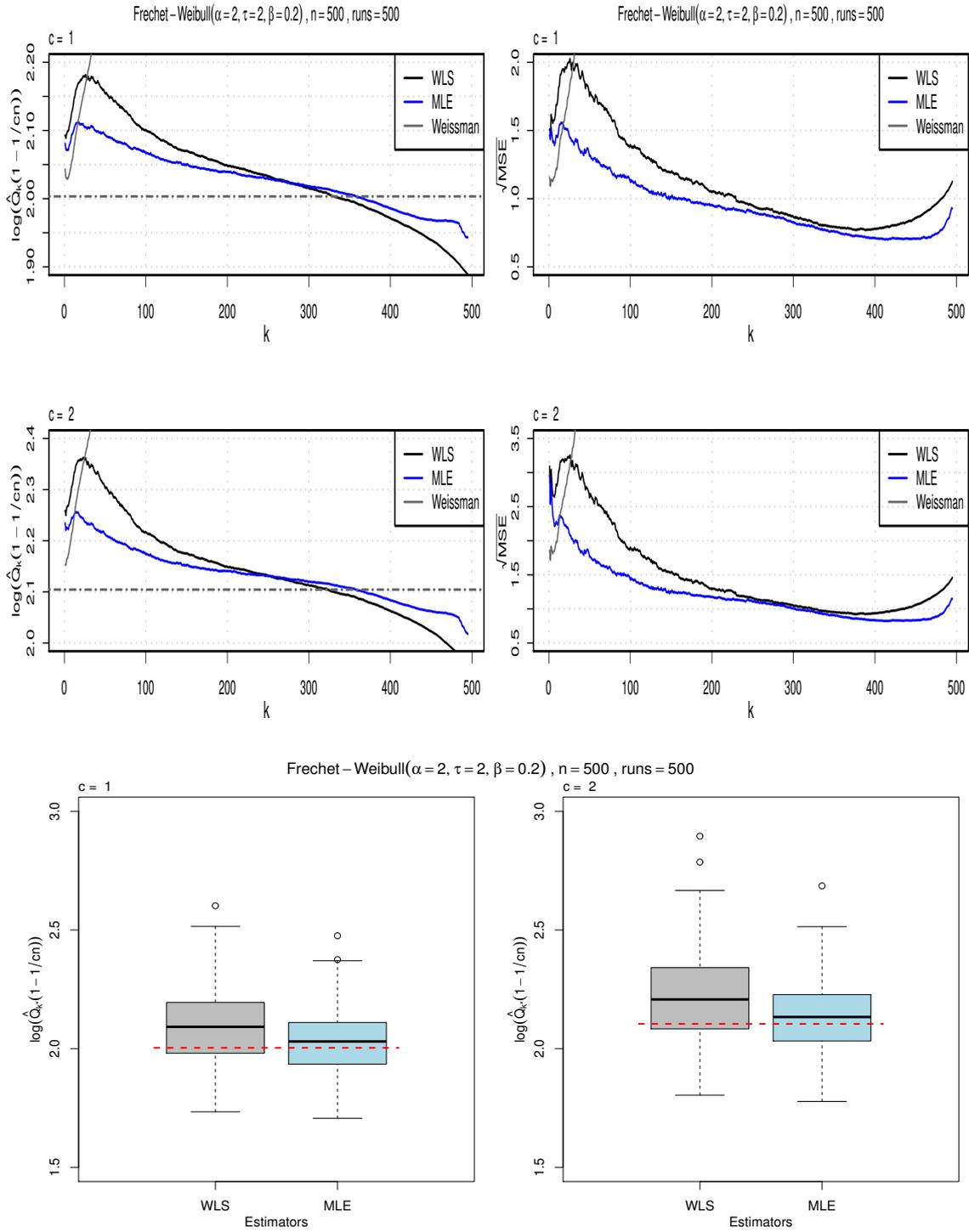


Figure 3.9: Frechet-Weibull(2.0, 2.0, 0.2): quantile estimates $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k . Bottom line: boxplots of $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

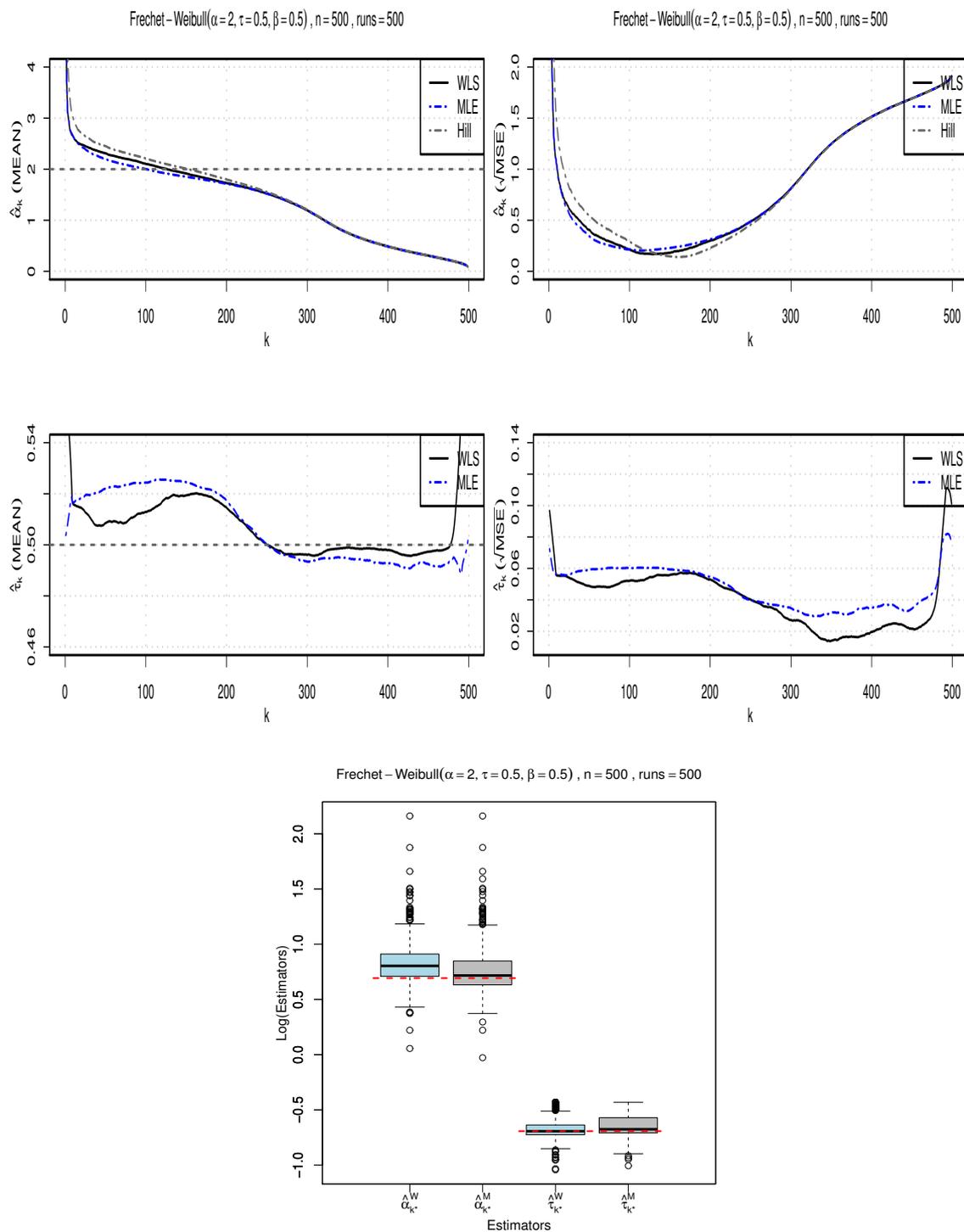


Figure 3.10: Fréchet-Weibull(2.0, 0.50, 0.5). *Top:* Mean (left) and RMSE (right) of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle:* Mean (left) and RMSE (right) of $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ as a function of k ; *Bottom:* Boxplots of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

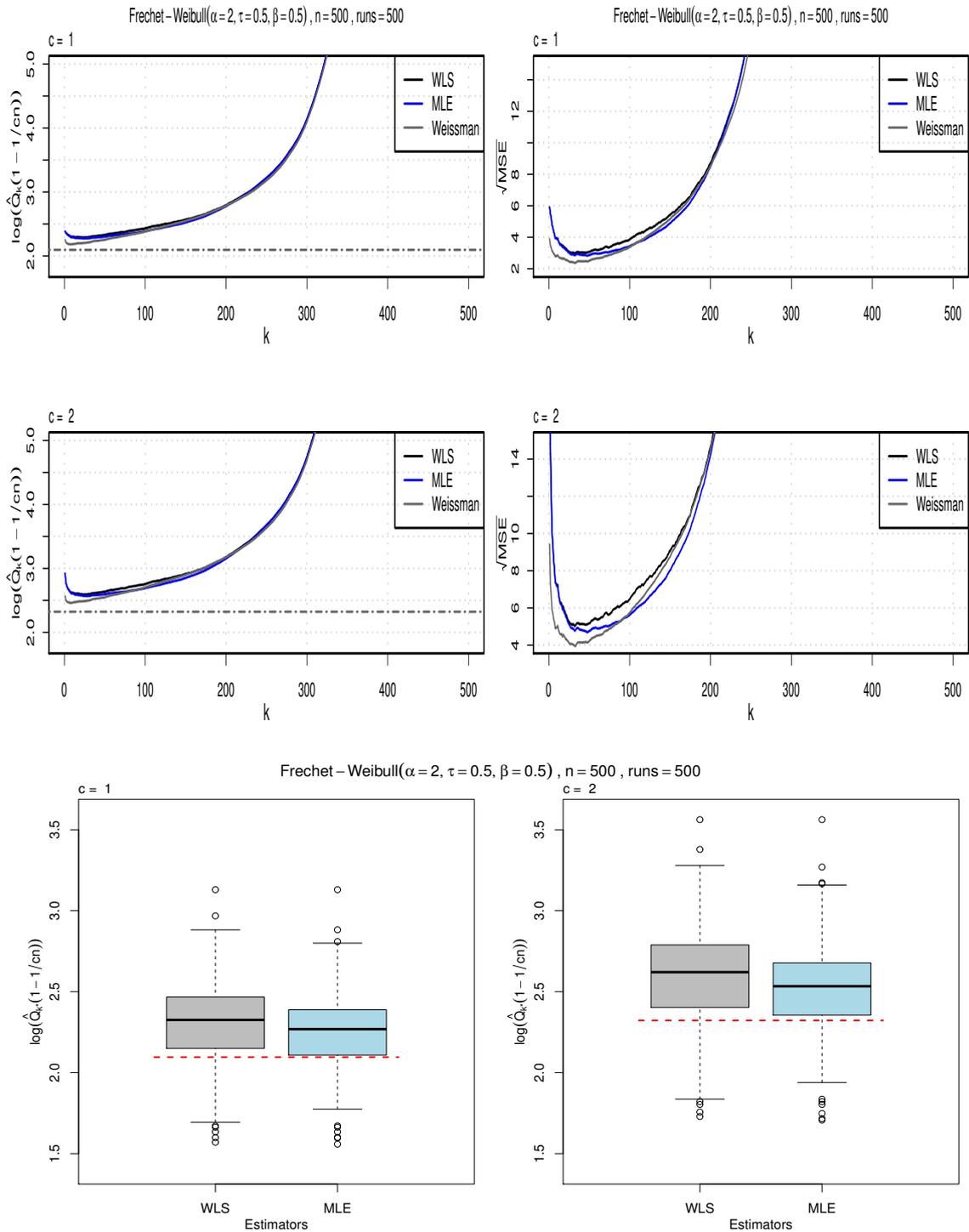


Figure 3.11: Fréchet-Weibull(2.0, 0.5, 0.5): quantile estimates $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k . Bottom line: boxplots of $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

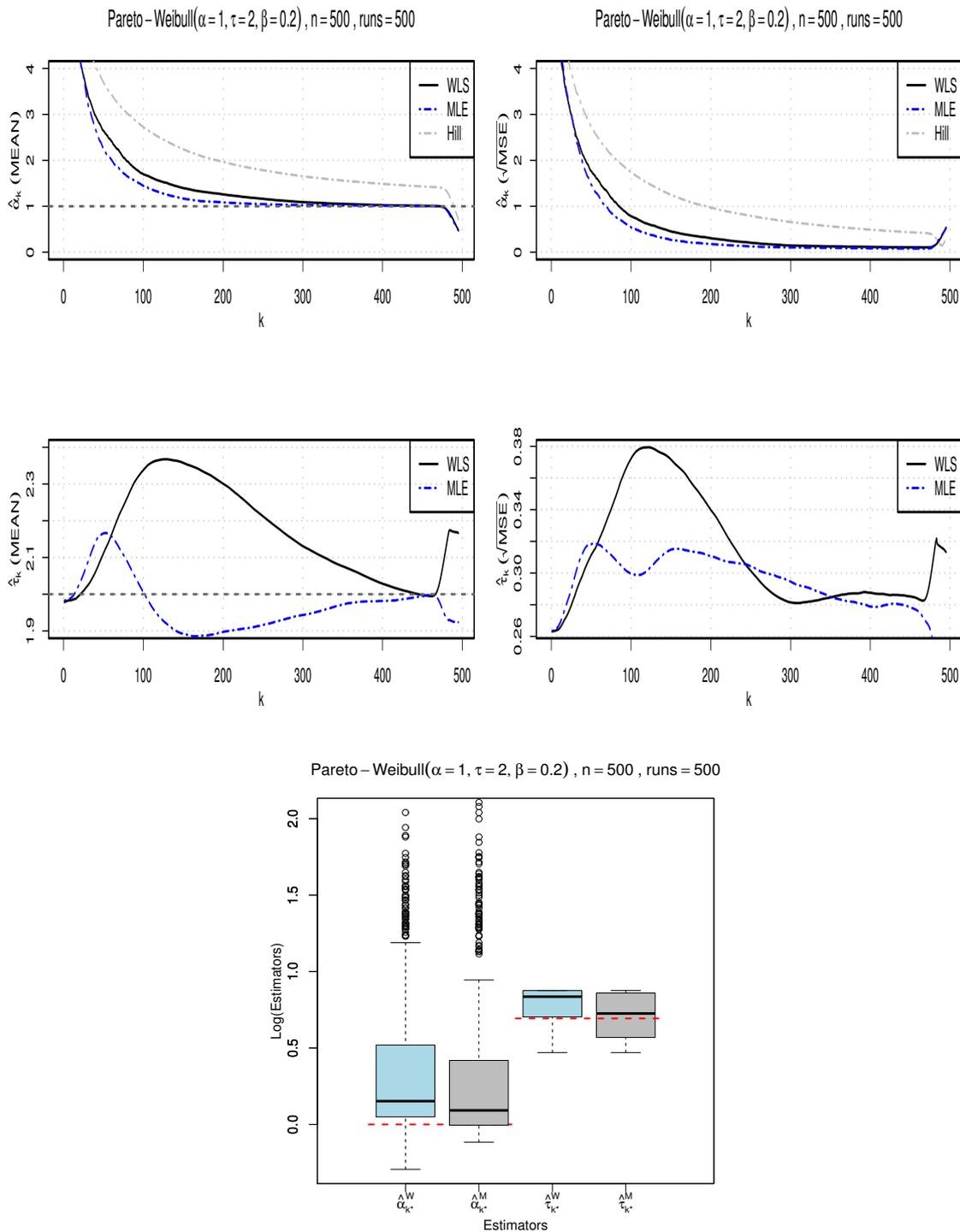


Figure 3.12: Pareto-Weibull(1.0, 2.0, 0.2). *Top*: Mean (left) and RMSE (right) of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ as a function of k ; *Middle*: Mean (left) and RMSE (right) of $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ as a function of k ; *Bottom*: Boxplots of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

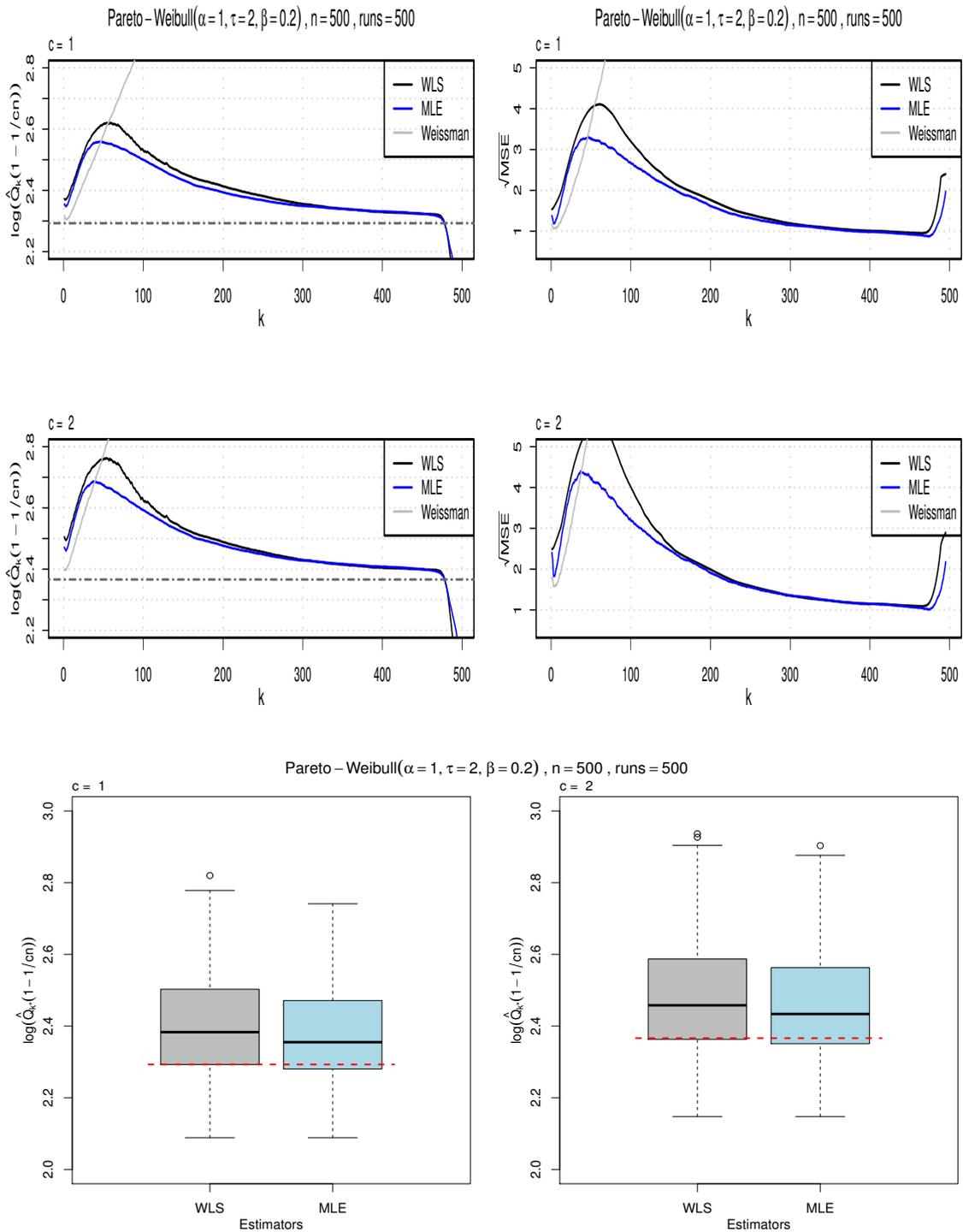


Figure 3.13: Pareto-Weibull(1.0, 2.0, 0.2): quantile estimates $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 1$ (top) and $c = 2$ (middle). Means (left) and RMSE (right) as a function of k . Bottom line: boxplots of $\hat{Q}_{p,k}^W, \hat{Q}_{p,k}^M$ with $c = 1$ (left) and $c = 2$ (right). Horizontal dashed lines indicate the real parameters.

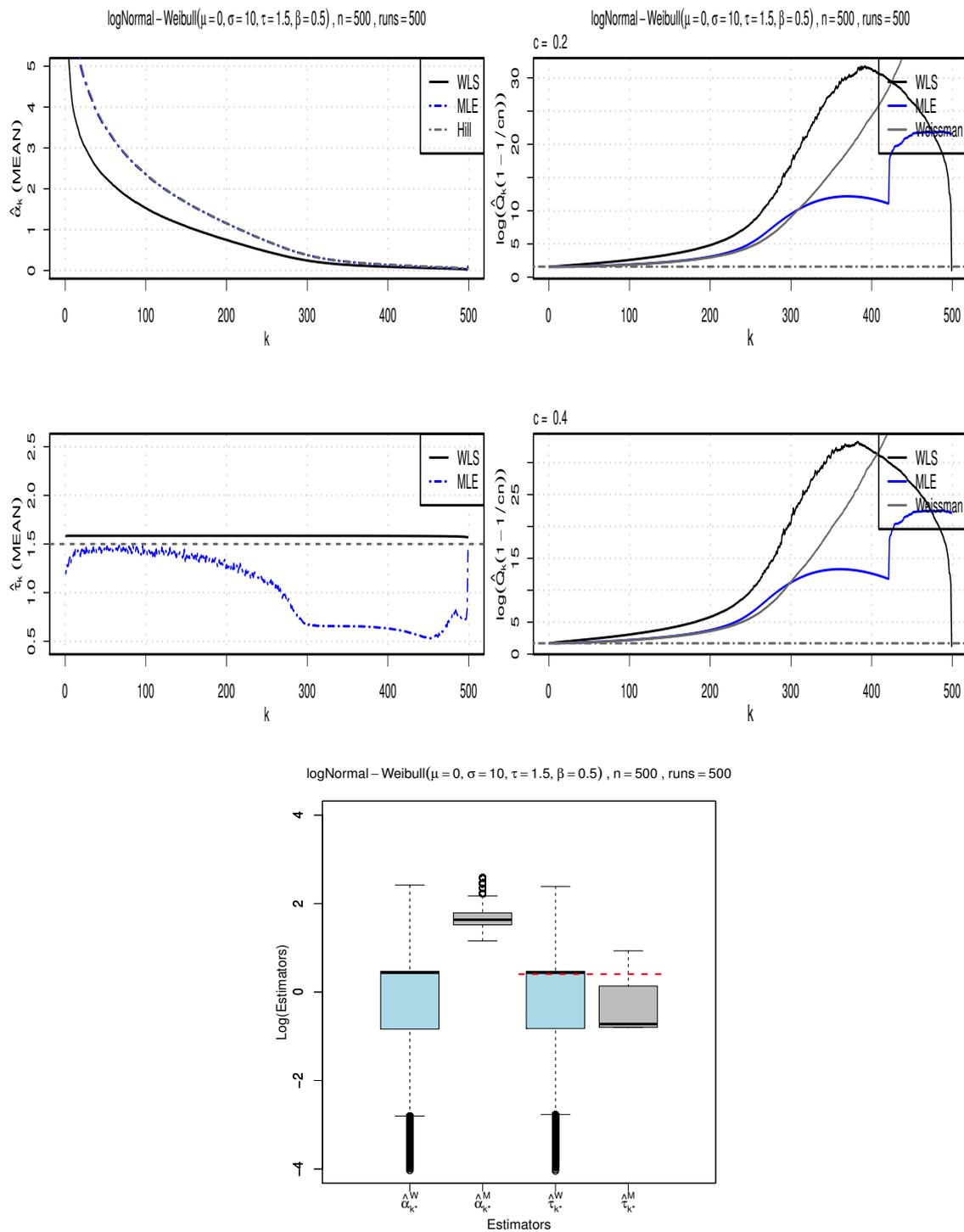


Figure 3.14: log-normal-Weibull(0.0, 100, 1.5, 0.5). *Top left:* $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$ and $H_{k,n}$ mean estimates as a function of k . *Middle left:* $\hat{\tau}_k^W$, $\hat{\tau}_k^M$ mean estimates as a function of k . *Right:* quantile estimates $\hat{Q}_{p,k}^W$, $\hat{Q}_{p,k}^M$ with $p = \frac{1}{cn}$ with $c = 0.2$ (top) and $c = 0.4$ (middle). *Bottom:* boxplots of $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $\hat{\tau}_k^W$ and $\hat{\tau}_k^M$ (log-scale). Horizontal dashed lines indicate the real parameters.

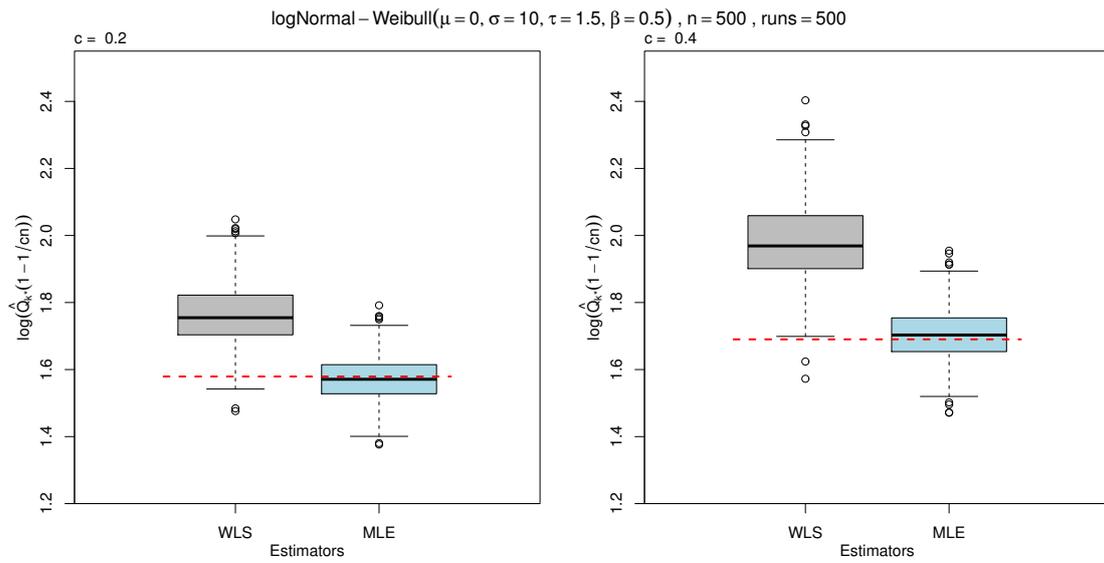


Figure 3.15: log-normal-Weibull(0.0, 100, 1.5, 0.5): Boxplots of $\hat{Q}_{p,\hat{k}}^W, \hat{Q}_{p,\hat{k}}^M$ with $c = 0.2$ (left) and $c = 0.4$ (right). Horizontal dashed lines indicate the real values.

3.8 Appendix C: Insurance cases

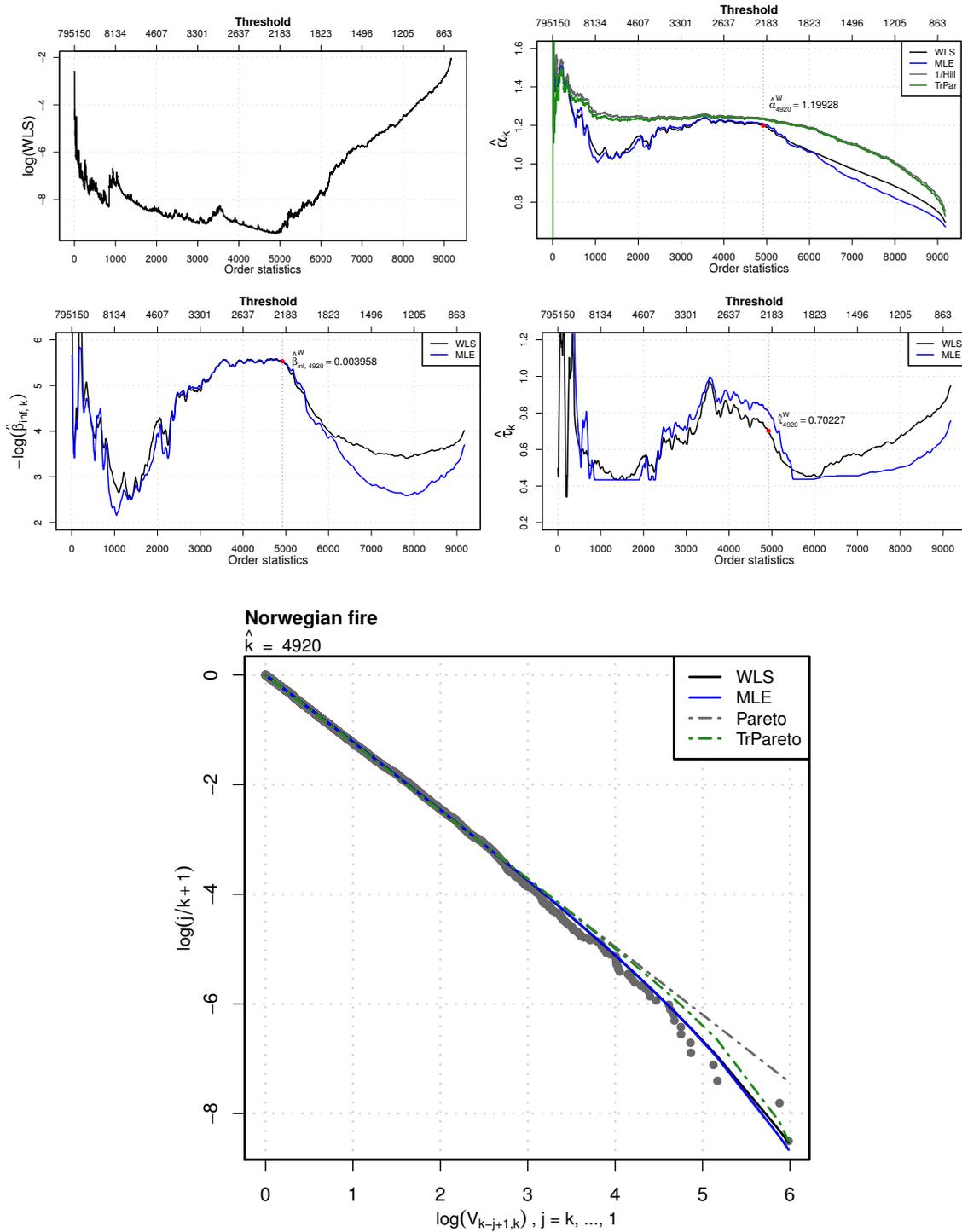


Figure 3.16: *Norwegian fire insurance data*: Top left: SS_k from (3.11); Top right: $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $H_{k,n}$ and $\hat{\alpha}_k^T$; Middle left: $-\log \hat{\beta}_{\infty,k}^W$, $-\log \hat{\beta}_{\infty,k}^M$; Middle right: $\hat{\tau}_k^W$, $\hat{\tau}_k^M$; Bottom: log-log plot with fit obtained from (3.8) with $k = \hat{k} = 4915$ using MLE and WLS estimates, next to Pareto and truncated Pareto fit.

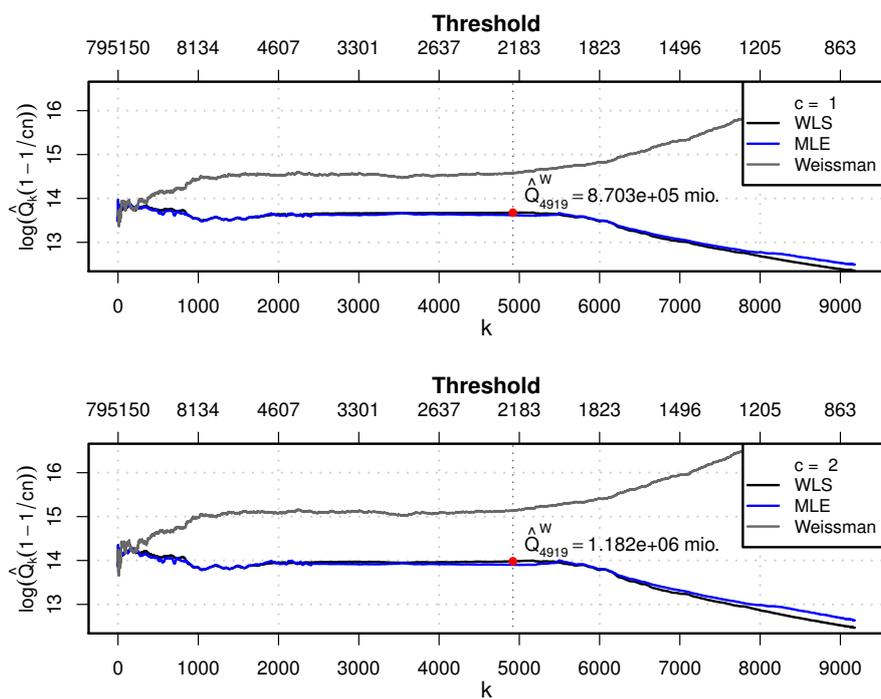


Figure 3.17: Norwegian fire insurance data: $\hat{Q}_{p,k}^W$, $\hat{Q}_{p,k}^M$ and $\hat{Q}_{p,k}^H$ quantile estimates with $p = 1/n$ (top) and $p = 1/(2n)$ (bottom).

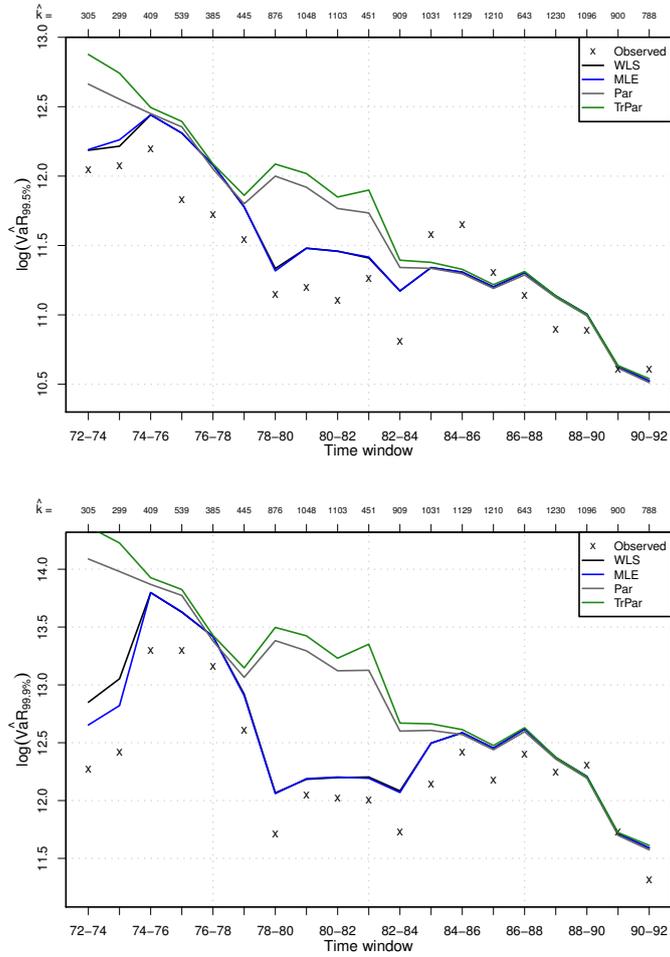


Figure 3.18: *Norwegian fire insurance data*: $\log\text{-VaR}(99.5\%)$ (top) and $\log\text{-VaR}(99.5\%)$ (bottom) at \hat{k} for tempered model (black and blue lines), Pareto (grey), truncated Pareto (green) and observed values (x). For each time window, \hat{k} is displayed at the top margin.

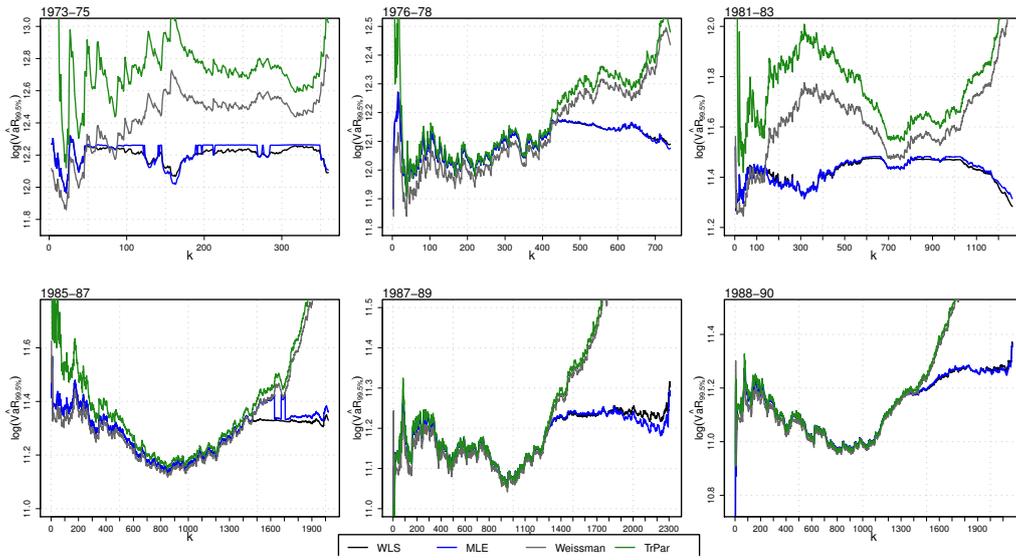


Figure 3.19: Norwegian fire insurance data: $\log VaR(99.5\%)$ for tempered model (black and blue lines), Pareto (grey) and truncated Pareto (green) for selected time windows.

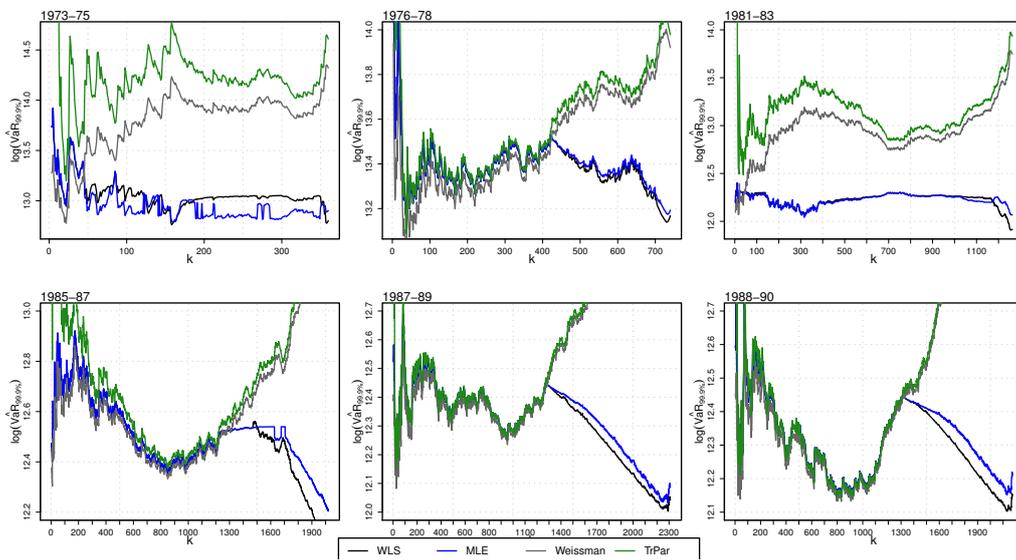


Figure 3.20: Norwegian fire insurance data: $\log VaR(99.9\%)$ for tempered model (black and blue lines), Pareto (grey), truncated Pareto (green) for selected time windows.

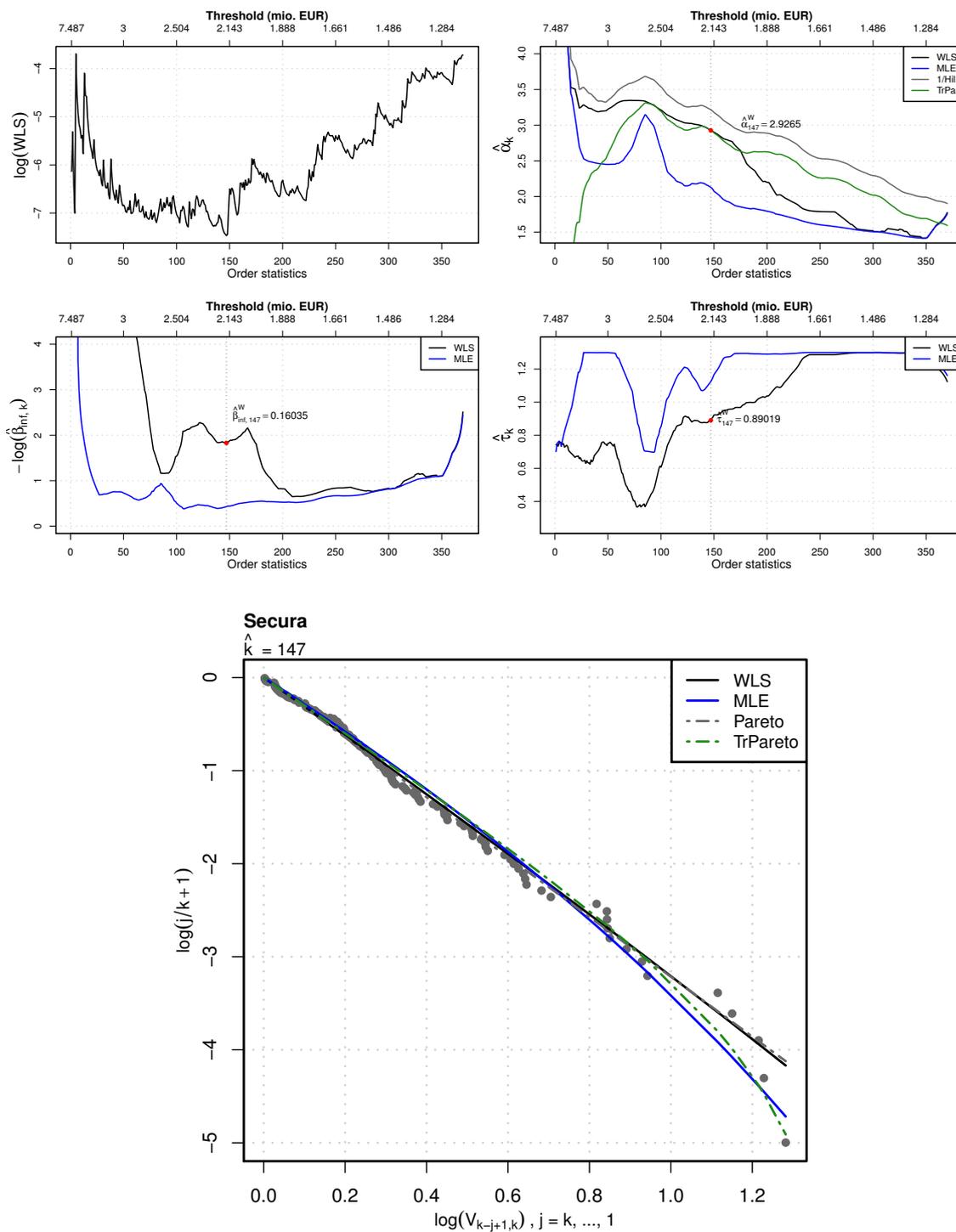


Figure 3.21: *Secura* data set: Top left: SS_k from (3.11); Top right: $\hat{\alpha}_k^W$, $\hat{\alpha}_k^M$, $H_{k,n}$ and $\hat{\alpha}_k^T$; Middle left: $-\log \hat{\beta}_{\infty,k}^W$, $-\log \hat{\beta}_{\infty,k}^M$; Middle right: $\hat{\tau}_k^W$, $\hat{\tau}_k^M$; Bottom: log-log plot with fit obtained from (3.8) with $k = \hat{k} = 147$ using MLE and WLS estimates, next to Pareto and truncated Pareto fit.

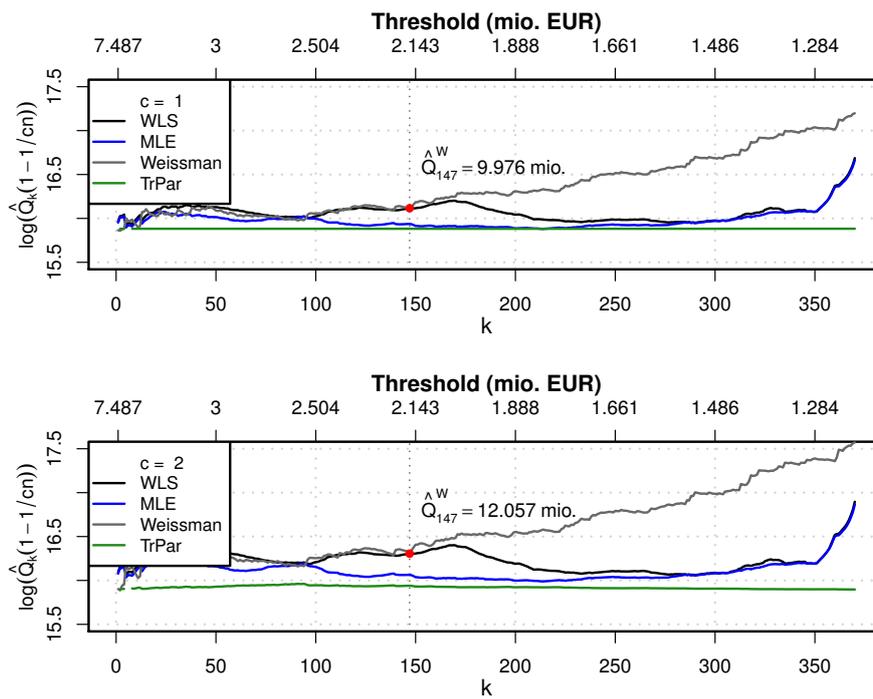


Figure 3.22: *Secura data set*: $\hat{Q}_{p,k}^W$, $\hat{Q}_{p,k}^M$ and $\hat{Q}_{p,k}^H$ quantile estimates with $p = 1/n$ (top) and $p = 1/(2n)$ (bottom).

Chapter 4

On the randomized Schmitter problem

This chapter is based on the following article:

H. Albrecher and J.C. Araujo-Acuna. On the randomized Schmitter problem. *Submitted*, 2020.

Abstract. We revisit the classical Schmitter problem in ruin theory and consider it for randomly chosen initial surplus level U . We show that the computational simplification that is obtained for exponentially distributed U allows to connect the problem to m -convex ordering, from which simple and sharp analytical bounds for the ruin probability are obtained, both for the original (but randomized) problem and for extensions involving higher moments. In addition, we show that the solution to the classical problem with deterministic initial surplus level can conveniently be approximated via Erlang(k)-distributed U for sufficiently large k , utilizing the computational advantages of the advocated randomization approach.

4.1 Introduction

At the ASTIN Colloquium 1990 in Montreux, the Swiss actuary Hans Schmitter presented an algorithm for the exact evaluation of the ruin probability $\psi(u)$ of a Cramér-Lundberg surplus process for an insurance portfolio with initial surplus u , for the case when the claim amount distribution is discrete on a finite range (Schmitter [165]). Also inspired by Bowers [43], he then posed the following question: If the individual claims are known to have mean μ and variance σ^2 , which claim size distributions minimize or maximize the ruin probability for a given u , respectively?

I.e.,

$$\begin{aligned} & \min/\max \quad \psi(u) \\ & \text{subject to} \quad X \text{ is a non-negative random variable,} \\ & \quad \quad \quad \text{with } \mathbb{E}(X) = \mu \text{ and } \text{Var}(X) = \sigma^2, \end{aligned}$$

where X is the random variable describing the individual claim sizes. This problem was then further discussed by Brockett et al. [47] and taken up in Kaas [122], where it was also extended to the related problem of finding extremal values of stop-loss premiums for compound Poisson distributions with similar moment restrictions. Much later, De Vylder et al. [70, 72] provided a numerical solution to the Schmitter problem based on a renewal equation that approximates the classical ruin model using a discrete time grid and partially solved the original problem in [69].

While on the basis of these contributions the problem can be considered as quite well understood, it was never solved in full generality. Correspondingly, despite the considerable time that has passed since then and the gradual shift of criteria for solvency considerations in insurance practice in the meantime, we would like to add an additional layer of complexity and understanding of the Schmitter problem in this paper by taking the perspective of a randomized initial surplus level.

Randomization as a principle has proven to be a very useful tool in risk theory leading to simpler expressions (see e.g. Albrecher et al. [10], Ivanovs [117]) or even unexpected identities (Albrecher and Ivanovs [14]), but particularly also to considerable computational advantages (cf. Carr [52], Avram et al. [23], Albrecher and Goffard [11]). The idea for the latter computational approach is to replace a deterministic quantity by a random variable with matching expected value, often with the advantage of smoothing the corresponding computational problem, leading to simpler and amenable expressions. In a final step, if possible the variance of that random variable is reduced considerably such that the resulting value can be an excellent approximation of the original computationally complex problem (“Erlangization”).

From a practical viewpoint, the classical risk model can be interpreted in two different ways: either representing the surplus process, or from the accounting perspective of assets and liabilities of the insurance company. According to the latter perspective, an insurance company has to regularly evaluate the value of its assets and its liabilities. Then, the value of the assets minus the value of the liabilities gives the available capital, or initial surplus. However, at the moment of valuation the exact value of the total assets might not be known with certainty and hence it might be thought of in terms of a random initial surplus.

In our setting, we replace the deterministic initial surplus level u by an exponentially distributed random variable U with mean u . The expected value of the resulting ruin probability can then be expressed in terms of the (simpler) Laplace transform of the classical ruin probability under the Cramér-Lundberg model. At this level, analytical lower and upper bounds for the ruin probability in the Schmitter problem can then be established utilizing the strong results in the theory of m -convex orders obtained by Denuit et al. [76, 85] (see also [138] for a recent application of this ordering concept). In addition, the generality of the latter results in fact allows

to give sharp upper and lower bounds for the ruin probability when more than two moments of the underlying claim size distribution are specified, which can be seen as an extension of the Schmitter problem that naturally narrows the gap between the upper and lower bound. For a comprehensive survey of stochastic orderings we refer to the monographs by Kaas et al. [126], Shaked and Shanthikumar [170, 171] and Müller and Stoyan [149]. More recent treatments in a specifically actuarial context include Kaas et al. [124, Ch.7] and Asmussen and Steffensen [24, Ch.8].

Eventually, we are also interested in using these explicit expressions of the randomized model to approximate the classical situation of deterministic initial surplus level u . Developing the results further towards Erlang(k) distributed initial surplus, for increasing k (maintaining the expected value at u) this provides increasingly accurate approximations for the classical deterministic case, expressed through the explicit formulas of the randomized model.

The remaining paper is structured as follows. First, Section 4.2 recapitulates the model setting and summarizes relevant results from the existing literature. In Section 4.3 we then analyze the problem for an exponentially distributed initial surplus level U . We obtain an expression for the corresponding (expected) ruin probability in terms of the Laplace transform of the classical ruin probability in the Cramér-Lundberg model, and provide sharp lower and upper bounds for it when the claim size is bounded. We also provide corresponding bounds in the case of more than two pre-specified moments of the claim size distribution. Moreover, we illustrate the resulting interval for particular numerical parameters and place various concrete (truncated) claim size distributions within these bounds. In Section 4.4, we expand the randomization idea towards Erlang(k)-distributed initial surplus, and in the spirit of Asmussen et al. [23] we approximate the ruin probability with deterministic surplus via Erlangization and Richardson extrapolation. We give numerical illustrations which show that the known and somewhat curious kinks in the graphs of the known optimal solutions of the classical Schmitter problem can be smoothly approximated with this randomization approach. In some cases, a small value of k is already sufficient for a good approximation, in others the value of k has to be quite considerable. Section 4.5 concludes.

4.2 Preliminaries and previous results

Consider the classical Cramér-Lundberg model with surplus process

$$C(t) = u + ct - S(t),$$

at time $t \geq 0$, where u is the initial surplus level. Here, $S(t) = X_1 + \dots + X_{N(t)}$ denotes the aggregate claims up to time t , where the number of claims $\{N(t); t \geq 0\}$ up to time t refers to a homogeneous Poisson process with rate $\lambda > 0$ and the claim sizes $X_i, i = 1, 2, \dots$, are independent and identically distributed random variables with distribution function F_X and expected value $\mathbb{E}(X_1) = \mu$, independent of $\{N(t); t \geq 0\}$. The premium income per unit of time is $c = (1 + \theta)\lambda\mu$, where $\theta > 0$

is the safety loading. Define the associated aggregate loss process as $R(t) = S(t) - ct$, for $t \geq 0$. The probability $\psi(u)$ of ultimate ruin is the probability that the surplus process $C(t)$ ever drops below zero,

$$\psi(u) = \mathbb{P} \left(\inf_{t \geq 0} C(t) < 0 \right) = \mathbb{P} \left(\sup_{t \geq 0} R(t) > u \right).$$

The maximal aggregate loss $L = \sup_{t \geq 0} R(t)$ can be decomposed as the sum of ladder heights, i.e. as the sum of the amounts by which record lows (here denoted by L_1, L_2, \dots) in the insurer's surplus $C(t)$ appear. Furthermore, the distribution of the L_i ($i = 1, 2, \dots$) is given by the integrated tail distribution $F_{L_i}(x) = \mu^{-1} \int_0^x (1 - F_X(z)) dz$, $x > 0$. It is well known that $\psi(u)$ is given explicitly by the Pollaczek-Khinchine formula

$$\psi(u) = \frac{\theta}{1 + \theta} \sum_{k=0}^{\infty} \left(\frac{1}{1 + \theta} \right)^k (1 - F_{L_i}^{*k}(u)), \quad (4.1)$$

where $F_{L_i}^{*k}$ denotes the k -fold convolution of the ladder height distribution (see e.g. [22, Th.IV.2.1]). The latter expression shows that L is a compound geometric random variable and may be written as $L = \sum_{k=1}^M L_i$, with M being the number of ladder heights. It is easy to see that M has a geometric distribution with parameter $\psi(0) = 1/(1 + \theta)$ (see, for example [22, Cor.IV.3.1]). The Laplace transform of (4.1) is well-known to be

$$\widehat{\psi}(s) = \int_0^{\infty} e^{-su} \psi(u) du = \frac{1}{s} - \frac{c - \lambda\mu}{cs - \lambda(1 - M_X(-s))}, \quad (4.2)$$

where $M_X(-s) = \int_0^{\infty} e^{-sx} dF_X(x)$ is the Laplace-Stieltjes transform of X (cf. [161, Th.5.3.3] or [22, Cor.IV.3.4]).

For the case when the claim amount distribution has discrete support $\{x_1, x_2, \dots, x_m\}$ (with probabilities p_1, p_2, \dots, p_m), Schmitter [165] gave an explicit expression to compute $\psi(u)$ in the form

$$\psi(u) = 1 - \frac{\theta}{1 + \theta} \sum_{l_1, \dots, l_m} (-z_m)^{l_1 + \dots + l_m} e^{z_m} \prod_{j=1}^m \frac{p_j^{l_j}}{l_j!},$$

where $z_m = (u - l_1 \cdot x_1 - \dots - l_m \cdot x_m)_+ / \mu \cdot (1 + \theta)$ and $z_+ = \max(z, 0)$.

In the context of the Schmitter problem, 2-point distributions for the claim size play a special role. If X assumes the values x_1 with probability p and $x_2 > x_1$ with probability $1 - p$, then for fixed mean $\mu > 0$ and variance $\sigma^2 > 0$ we simply have

$$\mu = x_1 \cdot p + x_2 \cdot (1 - p) \quad \text{and} \quad \sigma^2 + \mu^2 = x_1^2 \cdot p + x_2^2 \cdot (1 - p),$$

or correspondingly

$$x_1 = \mu - \frac{\sigma^2}{x_2 - \mu}, \quad x_2 = \mu + \frac{\sigma^2}{\mu - x_1} \quad \text{and} \quad p = \frac{\sigma^2}{\sigma^2 + (\mu - x_1)^2}, \quad (4.3)$$

Notice that x_2 is increasing in x_1 . Moreover, one has the relationships

$$\frac{\sigma^2}{\mu^2 + \sigma^2} \leq p < 1, \quad 0 \leq x_1 < \mu, \quad \text{and} \quad \frac{\mu^2 + \sigma^2}{\mu} \leq x_2$$

(see e.g. Kaas et al. [126, Ch 10.2]). If we additionally assume that $X \in [0, b]$, naturally $x_2 \leq b$, and we have $0 \leq \mu \leq b$ and $0 \leq \sigma^2 \leq \mu(b - \mu)$. The following two extremal cases will be particularly relevant later. Namely, $X = \{0, 0^* := (\mu^2 + \sigma^2)/\mu\}$ and so $p = \sigma^2/(\sigma^2 + \mu^2)$ and $X = \{b^* := \mu + \sigma^2/(\mu - b), b\}$. In here, x^* denotes the function that assigns to x the unique real number such that the random variable $X = \{x, x^*\}$ has mean μ and variance σ^2 . Note that if b is not bounded, then as $x_1 \uparrow \mu$, $p \uparrow 1$ and $x_2 \rightarrow \infty$: while the probability mass at x_2 becomes arbitrarily small, it significantly contributes to the variance.

For any non-negative loss variable X , the stop-loss premium π_X is defined by

$$\pi_X(d) = \mathbb{E}((X - d)_+) = \int_d^\infty (1 - F_X(z))dz, \quad \text{for } d \geq 0.$$

Note that there is a one-to-one relation between the integrated tail distribution of X and its stop-loss premium, namely $F_{L_i}(z) = 1 - \pi_X(z)/\mu$. One important concept in the theory of risk ordering is the stop-loss order. Concretely, a random variable X is said to be less risky than another random variable Y in stop-loss order ($X \leq_{sl} Y$) if $\pi_X(d) \leq \pi_Y(d)$ for all retentions $d \geq 0$. The problem of finding bounds for stop-loss premiums is a classical topic in actuarial science, see for example Bühlmann et al. [49], Kaas and Goovaerts [123] and Steenackers and Goovaerts [174]. For a study of the relation between stop-loss premiums and their associated ruin probabilities as well as general upper bounds for both stop-loss premiums and ruin probabilities see Cai and Garrido [50] and the references therein.

A consequence of the above concept is that if for two Cramér-Lundberg risk processes with equal premium per unit of time and claim intensity parameter, but different claim sizes, say X and Y , with $X \leq_{sl} Y$ we have $\psi_X(u) \leq \psi_Y(u)$ for all $u \geq 0$ (see [126, Ch.8.2, Th.2.1]). Correspondingly, the Schmitter problem may be seen as being reduced to finding extremal distributions in the stop-loss order in the class of random variables in $[0, b]$ with mean μ and variance σ^2 . However, as pointed out in Brockett et al. [47] there are no extremal distributions in terms of stop-loss order in such a class.

Nevertheless one can construct stop-loss transforms in the corresponding range (bounded or not) with the given mean, but with minimal variance, larger than the given one. For two given moments, the latter is achieved by constructing a polynomial of degree 2 above the function $(X - d)_+$ which is tangent to this function in 2 points. The abscissas of these points will be the mass points. For a comprehensive description of this construction see Kaas et al. [126, Ch.10]. In the following we briefly state its main consequences.

For unbounded X with mean μ and variance σ^2 , the maximal stop-loss premium at fixed retention d is attained by a random variable Z with support $\{r, r^*\}$, where $r, r^* = d \mp \sqrt{(\mu - d)^2 + \sigma^2}$, and from (4.3) then $\mathbb{P}(X = r) = \sigma^2/(\sigma^2 + (\mu - r)^2)$.

Note that $\{r, r^*\}$ is the 2-point support that has d in the middle. If $X \in [0, b]$, this 2-point distribution still gives an upper bound, but it is no longer always sharp. Theorem 2.3 in [126, Ch.10] provides a sharp upper bound for stop-loss premiums for $X \in [0, b]$: For given retention d , the maximal stop-loss premium is attained by the distribution with the mass points

$$\begin{aligned} \{0, 0^*\} & \text{ if } 0 \leq d \leq \frac{1}{2}0^*, \\ \{r, r^*\} & \text{ if } \frac{1}{2}0^* \leq d \leq \frac{1}{2}(b + b^*), \\ \{b^*, b\} & \text{ if } \frac{1}{2}(b + b^*) \leq d \leq b \end{aligned}$$

with the notation introduced before. However, these results do not provide an upper bound for the ruin probability in the Schmitter problem, because it is not the same extremal distribution across all values of d , but the latter would be needed to bequeath the dominance in terms of the stop loss premium from the integrated tail to all its convolutions in (4.1). However, Kaas [122] showed that if X has lower stop-loss premiums than Y on the interval $[0, u]$, then the same property holds for compound sums with N terms of these random variables respectively, and ruin probabilities with an initial surplus u are lower for X than for Y . That is, for values of u smaller than $\frac{1}{2}0^*$, the ruin probability is maximized by the 2-point claim random variable $X = \{0, 0^*\}$. Consequently, in terms of the upper bound the Schmitter problem is solved for small values of the initial surplus u .

De Vylder and Marceau [72] and De Vylder et al. [70] provided numerical solutions to the problem based on a renewal equation in a discretized risk model. By restricting to lattice distributions, they used the method of linear combinations (see also Kaas et al. [127, Sec.3]) to obtain optimal solutions to the problem. They noted that for $u \gg b$ the maximal ruin probability was given by the 2-point claim random variable $X = \{b^*, b\}$. In fact, De Vylder et al. [69] then proved that there exists a constant $c > 0$ such that for all $u \geq c$ the maximal ruin probability is given by that 2-point claim random variable. However, the concrete value of c as well as the optimal result for intermediate values of u seem to still not be settled up to this day.

The minimal stop-loss premium for risks X with mean μ is given by $(\mu - d)_+$ for all retentions $d \geq 0$, i.e. it is attained by the defective random variable Z concentrated at μ , implying $Z \leq_{sl} X$ and therefore $\psi_Z(u) \leq \psi_X(u)$ for all u . However, Z does not fulfill the variance constraint, so that this is not a valid solution to the Schmitter problem. It does provide a general lower bound for its solution though, and for unbounded X the variance constraint can then be satisfied by adding an ε ($\downarrow 0$) mass at infinity, see also [22, Cor. IV.8.4].

4.3 Exponentially distributed initial surplus

Let us now replace the deterministic initial surplus u by a random variable U that has an exponential distribution with parameter $s > 0$. The redefined surplus process then is

$$C^R(t) = U + ct - S(t), \quad t \geq 0,$$

where c and $S(t)$ are defined as in the classical ruin model. Using the convenient fact that this choice of U simply puts us in the framework of Laplace transforms, due to (4.2) the ruin probability $\psi_U(s) := \mathbb{P}(C^R(t) < 0 \text{ for some } t > 0)$ is then given by

$$\psi_U(s) := \mathbb{E}(\psi(U)) = \int_0^\infty \psi(u) s e^{-su} du = s \cdot \widehat{\psi}(s) = 1 - s \cdot \frac{c - \lambda \mu}{cs - \lambda(1 - M_X(-s))}. \quad (4.4)$$

Since the randomization of the initial surplus corresponds to a probability-weighted averaging over situations with deterministic surplus, it is clear that this step leads to a smoothing of the ruin probability shape. Figure 4.1 compares the ruin probabilities $\psi(u)$ for deterministic surplus $u = \{1.5, 4.5, 9.0\}$ and $\theta = 0.5$ (the parameters from Kaas [122, Fig.1]) with the corresponding randomized quantities of the same expected initial surplus $\mathbb{E}(U) = 1/s = u$ for 2-point distributions with given mean $\mu = 3$ and variance $\sigma^2 = 1$. One observes that the sensitivity w.r.t. the choice of the only free parameter x_1 is substantially different, and the somewhat curious shape change for increasing u from the classical deterministic case is indeed evened out.

Let us now look at the randomized and extended Schmitter problem

$$\begin{aligned} & \min/\max \quad \psi_U(s) \\ & \text{subject to} \quad \mathbb{E}(X^k) = \mu_k, \text{ for } k = 1, 2, \dots, m \end{aligned}$$

with possibly more than two fixed moments of the claims size distribution. Inspired by Kaas [122], using the maximal aggregate loss L and assuming that the moments of the claim size are finite, one can express the ruin probability in terms of the claim size moments, namely

$$\begin{aligned} \psi_U(s) &= \int_0^\infty \psi(u) s e^{-su} du = \sum_{k=0}^{\infty} (-1)^k \frac{s^{k+1}}{k!} \int_0^\infty u^k \psi(u) du \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{s^{k+1}}{k!} \int_0^\infty u^k \mathbb{P}(L > u) du = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{s^k}{k!} \mathbb{E}(L^k) \\ &= \sum_{k=1}^{\infty} (-1)^{k-1} \frac{s^k}{k!} \mathbb{E} \left(\mathbb{E} \left(\sum_{l_1+l_2+\dots+l_M=k} \binom{k}{l_1, l_2, \dots, l_M} \prod_{j=1}^M L_j^{l_j} \middle| M \right) \right). \end{aligned}$$

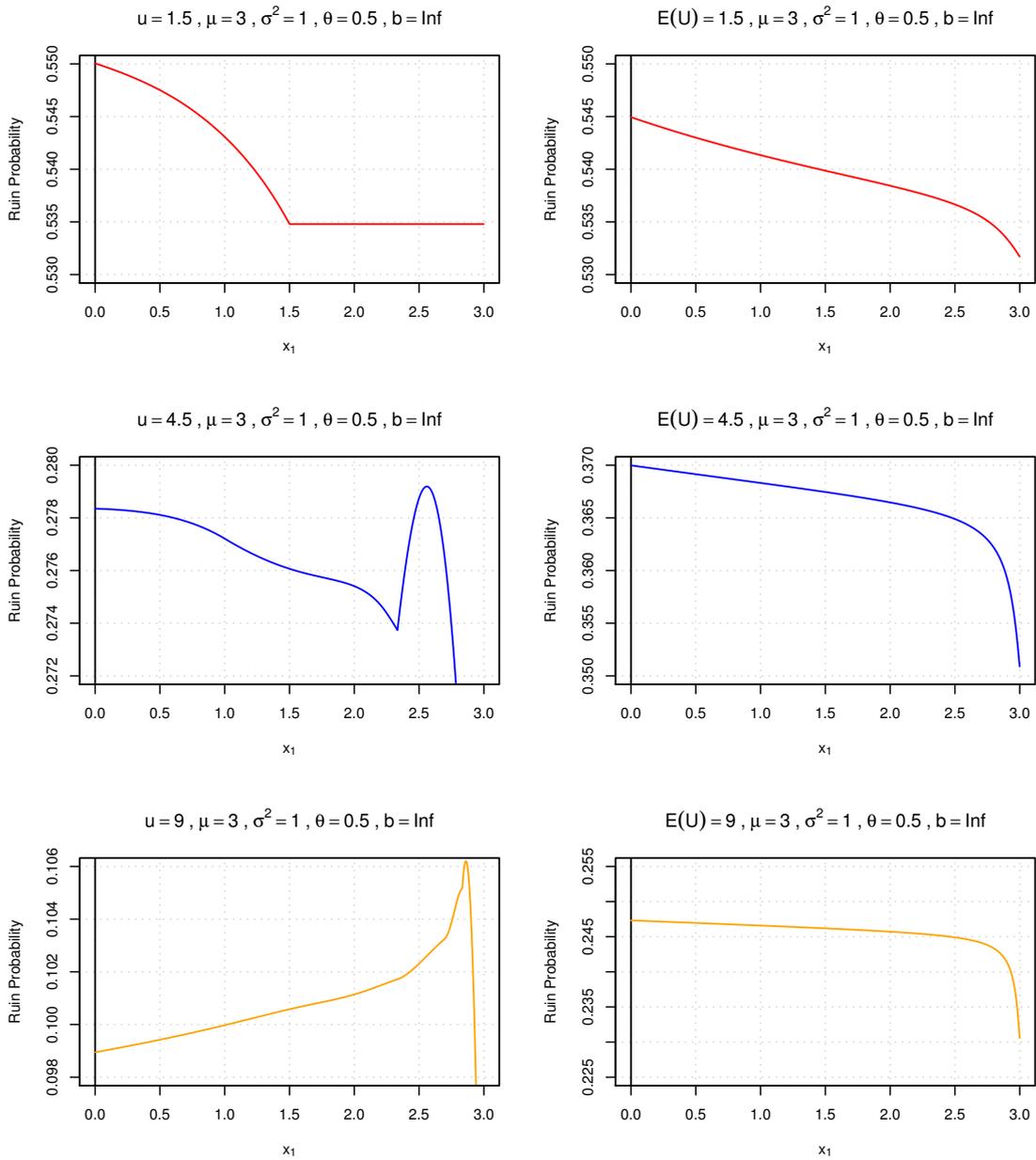


Figure 4.1: Ruin probabilities as a function of x_1 for $\mu = 3$, $\sigma^2 = 1$, $\theta = 0.5$ for the three deterministic surplus levels $u = 1.5, 4.5, 9$ (left column) and the randomized counterpart with the same expected initial surplus level (right column).

The first four terms of this series are given by

$$\begin{aligned}
E(L) &= \mathbb{E}(M)\mathbb{E}(L_1) = \frac{1}{2\theta\mu}\mathbb{E}(X^2) \\
E(L^2) &= \mathbb{E}(M)\mathbb{E}(L_1^2) + \mathbb{E}(M(M-1))\mathbb{E}^2(L_1) = \frac{1}{3\theta\mu}\mathbb{E}(X^3) + \frac{1}{2\theta^2\mu^2}\mathbb{E}(X^2) \\
E(L^3) &= \frac{1}{4\theta\mu}\mathbb{E}(X^4) + \frac{1}{\theta^2\mu^2}\mathbb{E}(X^3)\mathbb{E}(X^2) + \frac{3}{4\theta^3\mu^3}\mathbb{E}^3(X^2) \\
E(L^4) &= \frac{1}{5\theta\mu}\mathbb{E}(X^5) + \frac{1}{\theta^2\mu^2}\mathbb{E}(X^4)\mathbb{E}(X^2) + \frac{2}{3\theta^2\mu^2}\mathbb{E}^2(X^3) \\
&\quad + \frac{1}{6\theta^3\mu^3}\mathbb{E}(X^3)\mathbb{E}^2(X^2) + \frac{3}{2\theta^4\mu^4}\mathbb{E}^4(X^2)
\end{aligned}$$

Hence, if the first m moments of X are given, then one can approximate

$$\psi_U(s) \approx s\mathbb{E}(L) - \frac{s^2}{2}\mathbb{E}(L^2) + \cdots + (-1)^{m-1}\frac{s^m}{m!}\mathbb{E}(L^m) \quad (4.5)$$

and investigate the behavior with respect to the highest moment. For example, for $m = 2$

$$\psi_U(s) \approx s\frac{\sigma^2 + \mu^2}{2\theta\mu} - \frac{s^2}{6\theta\mu}\mathbb{E}(X^3) - s^2\frac{(\sigma^2 + \mu^2)^2}{4\theta^2\mu^2}.$$

Therefore, distributions with large third moment will make $\psi_U(s)$ small and vice versa. For 2-point distributions, a simple calculation shows that, $\frac{\partial}{\partial x_1}\mathbb{E}(X^3) = \sigma^2 + \left(\frac{\sigma^2}{\mu - x_1}\right)^2 > 0$ and $\frac{\partial^2}{\partial x_1^2}\mathbb{E}(X^3) = 2\frac{\sigma^2}{(\mu - x_1)^3} > 0$. Thus, for $x_1 \in [0, \mu)$, its third moment is increasing and convex, so the maximum will be at $x_1 = 0$ and the minimum at $x_1 \rightarrow \mu$. In fact, for deterministic surplus and 2-point distributions, Kaas [122] argued that as $\int_0^\infty \psi(u)du = \mathbb{E}(L)$ does not depend on x_1 and $\int_0^\infty u\psi(u)du = \mathbb{E}(L^2)$ increases linearly with the third moment of the claim distribution, so that for small u , the ruin probability will be large for $x_1 = 0$.

While these considerations are intuitive, from (4.4) it becomes clear that for the extremal values of the randomized ruin probability it suffices to minimize (maximize) the Laplace transform of the individual claim sizes, i.e. to find extremal random variables in the Laplace transform order. The Laplace transform order has been introduced by Rolski and Stoyan [176] to compare waiting times in queuing theory. In actuarial science, Denuit [75] studied both univariate and multivariate versions of the Laplace transform order and gave several actuarial applications. We can now give sharp bounds for the randomized Schmitter problem for two given moments.

Proposition 4.3.1. *Let X be a non-negative random variable with mean μ and variance σ^2 . Then*

$$1 - s \cdot \frac{c - \lambda\mu}{cs - \lambda(1 - e^{-s\mu})} \leq \psi_U(s) \leq 1 - s \cdot \frac{c - \lambda\mu}{cs - \lambda\left(1 - \frac{\sigma^2}{\sigma^2 + \mu^2} - \frac{\mu^2}{\sigma^2 + \mu^2}e^{-s(\mu + \sigma^2/\mu)}\right)}.$$

Proof. Note that $e^{-s\mu}$ is the Laplace transform of a random variable Z degenerate at μ . Moreover, $\frac{\sigma^2}{\sigma^2+\mu^2} + \frac{\mu^2}{\sigma^2+\mu^2}e^{-s(\mu+\sigma^2/\mu)}$ is the Laplace transform of a random variable X with mean μ , variance σ^2 and such that $\mathbb{P}(X = 0) = 1 - \mathbb{P}(X = (\mu^2 + \sigma^2)/\mu) = \frac{\sigma^2}{\sigma^2+\mu^2}$. Therefore, as maximizing the Laplace transform of the individual claim sizes minimizes $\psi_U(s)$ and vice versa, it suffices to show that $X \leq_{\text{Lt}} Y \leq_{\text{Lt}} Z$. The proof of the latter can be found in Shaked and Shanthikumar [171, Ch. 5, Theorem 5.A.21]. \square

It is worth noticing that the distribution maximizing the randomized ruin probability coincides with the 2-point distribution that maximizes the ruin probability under deterministic surplus for small values of u . This is rather intuitive, since $\psi_U(s)$ is a weighted average of $\psi(u)$ with a lot of weight on small values of u .

If more moments of the claim size X in $[0, b]$ are specified, then one can obtain tighter upper and lower bounds for the randomized ruin probability. In view of (4.4), this reduces to the derivation of bounds for the Laplace transform of X in the moment space $\mathcal{B}_S([0, b]; \mu_1, \mu_2, \dots, \mu_m)$ of all risks X with range $[0, b]$ such that $\mathbb{E}(X^k) = \mu_k$ for $k = 1, 2, \dots, m$. Fortunately, our context fits exactly into the framework of Denuit et al. [76, 85] who constructed lower and upper stochastic bounds for a given set of risks using m -convex stochastic orders. More precisely, consider the class \mathcal{M}_{m-cx} of all functions $\phi : [0, b] \rightarrow \mathbb{R}$ whose $(m+1)$ -th derivative $\phi^{(m+1)}(x)$ exists and is non-negative, for all $x \in [0, b]$, or which are limits of sequences of functions whose $(m+1)$ -th derivative is continuous and non-negative on $[0, b]$. Define the partial order relation \leq_{m-cx} among elements in \mathcal{B}_S as

$$X \leq_{m-cx} Y \text{ if and only if } \mathbb{E}(\phi(X)) \leq \mathbb{E}(\phi(Y)) \text{ for all functions } \phi \in \mathcal{M}_{m-cx}, \quad (4.6)$$

provided the expectations exists. It is then possible to determine two discrete risks $X_{\min}^{(m)}$ and $X_{\max}^{(m)}$, in $\mathcal{B}_S([0, b]; \mu_1, \mu_2, \dots, \mu_m)$ with probability masses depending on the moment set $(\mu_1, \mu_2, \dots, \mu_m)$ such that

$$X_{\min}^{(m)} \leq_{m-cx} X \leq_{m-cx} X_{\max}^{(m)} \text{ for all } X \in \mathcal{B}_S. \quad (4.7)$$

Explicit descriptions for the distribution functions of the extrema up to $m = 4$ are obtained in Denuit et al. [76, Table 1, Table 2]. Moreover, the latter reference also provided the extrema with respect to the order \leq_{m-cx} when not only the first $m-1$ moments and the support are given, but also when the density function of X is known to be unimodal with a known mode.¹

Let $X \in [0, b]$, $b > 0$, with moments $(\mu_1, \mu_2, \dots, \mu_m)$. Since $\phi(x) = (-1)^{m+1}e^{-sx}$ belongs to \mathcal{M}_{m-cx} for $s > 0$, we get from (4.6) and (4.7) that

$$\begin{aligned} M_{X_{\max}^{(m)}}(-s) &\leq M_X(-s) \leq M_{X_{\min}^{(m)}}(-s), \text{ for } m+1 \text{ odd} \\ M_{X_{\min}^{(m)}}(-s) &\leq M_X(-s) \leq M_{X_{\max}^{(m)}}(-s), \text{ for } m+1 \text{ even,} \end{aligned} \quad (4.8)$$

which can then be translated to bounds for $\psi_U(s)$. The bounds for the Laplace transform using m -convex risks were already described in Denuit et al. [86], extending

¹Note that in this paper we use the notation \leq_{m-cx} to denote m -convexity whenever the first m moments are available and not $(m-1)$ as it is standard in the m -convex risk literature.

earlier work of Eckberg [92], Whitt [182] and Lefèvre et al. [137]. In particular, [92] derived bounds for the Laplace transform up to the third moment using the theory of Chebychev systems and applied the bounds to problems in queuing and traffic theory. Moreover, the latter reference provides bounds for the case where no upper bound is known. We would also like to mention that, closely related to the theory of m -convex stochastic orders, using Markov-Krein theory and the theory of Chebychev systems, Brockett and Cox [45, 46] obtained similar upper and lower bounds for the expected value of a function of some random variable with given moments. Also, De Vylder [66, 67], De Vylder and Goovaerts [71], Kaas and Goovaerts [125] and Hürlimann [116] examined related bounding problems.

Using (4.8) we can give explicit bounds for the ruin probability with exponentially distributed initial surplus in terms of the given parameters. For reference, we restate here the respective bounds given in [76, Table 1, Table 2]) in terms of ruin probabilities when up to 4 moments of X are given:

Case $m = 1$. If μ_1 is given, then $X_{\min}^{(1)}$ is a random variable degenerate at μ_1 , and

$$X_{\max}^{(1)} = \begin{cases} 0 & \text{with } p = 1 - \frac{\mu_1}{b}, \\ b & \text{with } 1 - p = \frac{\mu_1}{b}. \end{cases}$$

Therefore,

$$\psi_U^{\min}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \lambda(1 - e^{-s\mu_1})},$$

$$\psi_U^{\max}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \frac{\lambda\mu_1}{b}(1 - e^{-sb})}.$$

Case $m = 2$. If μ_1 and μ_2 are given, then

$$X_{\min}^{(2)} = \begin{cases} 0 & \text{with } p = 1 - \frac{\mu_1^2}{\mu_2}, \\ \frac{\mu_2}{\mu_1} & \text{with } 1 - p = \frac{\mu_1^2}{\mu_2}, \end{cases} \quad X_{\max}^{(2)} = \begin{cases} \frac{b\mu_1 - \mu_2}{b - \mu_1} & \text{with } p = \frac{(b - \mu_1)^2}{(b - \mu_1)^2 + \mu_2 - \mu_1^2}, \\ b & \text{with } 1 - p = \frac{\mu_2 - \mu_1^2}{(b - \mu_1)^2 + \mu_2 - \mu_1^2}. \end{cases}$$

In this case, it can be seen that

$$\psi_U^{\min}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \lambda\left(1 - \frac{(b - \mu_1)^2}{\sigma^2 + (b - \mu_1)^2} e^{-s(\mu_1 - \sigma^2/(b - \mu_1))} - \frac{\sigma^2}{\sigma^2 + (b - \mu_1)^2} e^{-sb}\right)},$$

$$\psi_U^{\max}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \frac{\lambda\mu_1^2}{\sigma^2 + \mu_1^2}(1 - e^{-s(\mu_1 + \sigma^2/\mu_1)})}.$$

Note that for $b \rightarrow \infty$ the above expressions indeed converge to the bounds given in Proposition 4.3.1.

Case $m = 3$. If μ_1, μ_2 and μ_3 are given, then

$$X_{\min}^{(3)} = \begin{cases} r_+ = \frac{\mu_3 - \mu_1 \mu_2 + \sqrt{(\mu_3 - \mu_1 \mu_2)^2 - 4\sigma^2(\mu_1 \mu_3 - \mu_2^2)}}{2\sigma^2} & \text{with } p = \frac{\mu_1 - r_-}{r_+ - r_-}, \\ r_- = \frac{\mu_3 - \mu_1 \mu_2 - \sqrt{(\mu_3 - \mu_1 \mu_2)^2 - 4\sigma^2(\mu_1 \mu_3 - \mu_2^2)}}{2\sigma^2} & \text{with } 1 - p = 1 - \frac{\mu_1 - r_-}{r_+ - r_-}, \end{cases}$$

$$X_{\max}^{(3)} = \begin{cases} 0 & \text{with } p_3 = 1 - p_1 - p_2, \\ \frac{\mu_3 - b\mu_2}{\mu_2 - b\mu_1} & \text{with } p_1 = \frac{(\mu_2 - b\mu_1)^3}{(\mu_3 - b\mu_2)(\mu_3 - 2b\mu_2 + b^2\mu_1)}, \\ b & \text{with } p_2 = \frac{\mu_1 \mu_3 - \mu_2^2}{b(\mu_3 - 2b\mu_2 + b^2\mu_1)}. \end{cases}$$

Then, the bounds for the ruin probability are given by

$$\psi_U^{\min}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \lambda \left(1 - \left(1 - \frac{\mu_1 - r_-}{r_+ - r_-} \right) e^{-sr_-} - \left(\frac{\mu_1 - r_-}{r_+ - r_-} \right) e^{-sr_+} \right)},$$

$$\psi_U^{\max}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \lambda \left(p_1 \left(1 - e^{-s \frac{\mu_3 - b\mu_2}{\mu_2 - b\mu_1}} \right) + p_2 (1 - e^{-sb}) \right)}.$$

Case $m = 4$. If μ_1 and up to μ_4 are given, then

$$X_{\min}^{(4)} = \begin{cases} 0 & \text{with } 1 - p_+ - p_-, \\ t_+ = \frac{\mu_1 \mu_4 - \mu_2 \mu_3 + \sqrt{(\mu_1 \mu_4 - \mu_2 \mu_3)^2 - 4(\mu_1 \mu_3 - \mu_2^2)(\mu_2 \mu_4 - \mu_3^2)}}{2(\mu_1 \mu_3 - \mu_2^2)} & \text{with } p_+ = \frac{\mu_2 - t_- \mu_1}{t_+ (t_+ - t_-)}, \\ t_- = \frac{\mu_1 \mu_4 - \mu_2 \mu_3 - \sqrt{(\mu_1 \mu_4 - \mu_2 \mu_3)^2 - 4(\mu_1 \mu_3 - \mu_2^2)(\mu_2 \mu_4 - \mu_3^2)}}{2(\mu_1 \mu_3 - \mu_2^2)} & \text{with } p_- = \frac{\mu_2 - t_+ \mu_1}{t_- (t_- - t_+)}. \end{cases}$$

$$X_{\max}^{(4)} = \begin{cases} z_+ = \frac{(\mu_1 - b)(\mu_4 - b\mu_3) - (\mu_2 - b\mu_1)(\mu_3 - b\mu_2) + \sqrt{\rho}}{2((\mu_1 - b)(\mu_3 - b\mu_2) - (\mu_2 - b\mu_1)^2)} & \text{with } q_+ = \frac{\mu_2 - (b + z_-)\mu_1 + bz_-}{(z_+ - z_-)(z_+ - b)}, \\ z_- = \frac{(\mu_1 - b)(\mu_4 - b\mu_3) - (\mu_2 - b\mu_1)(\mu_3 - b\mu_2) - \sqrt{\rho}}{2((\mu_1 - b)(\mu_3 - b\mu_2) - (\mu_2 - b\mu_1)^2)} & \text{with } q_- = \frac{\mu_2 - (b + z_+)\mu_1 + bz_+}{(z_- - z_+)(z_- - b)}, \\ b & \text{with } 1 - q_+ - q_-. \end{cases}$$

Here,

$$\rho := ((\mu_1 - b)(\mu_4 - b\mu_3) - (\mu_2 - b\mu_1)(\mu_3 - b\mu_2))^2 - 4((\mu_1 - b)(\mu_3 - b\mu_2) - (\mu_2 - b\mu_1)^2)((\mu_2 - b\mu_1)(\mu_4 - b\mu_3) - (\mu_3 - b\mu_2)^2)$$

As can easily be verified,

$$\psi_U^{\min}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \lambda(1 - q_+ e^{-sz_+} - q_- e^{-sz_-} - (1 - q_+ - q_-)e^{-sb})},$$

$$\psi_U^{\max}(s) = 1 - s \cdot \frac{c - \lambda\mu_1}{cs - \lambda(p_+(1 - e^{-st_+}) + p_-(1 - e^{-st_-}))}.$$

4.3.1 Numerical illustrations.

De Vylder [68, Sec.II, Ch.3] gives conditions for the class of all vectors $(\mu_1, \mu_2, \dots, \mu_m) \in \mathbb{R}^m$ such that $\mathcal{B}_S([0, b]; \mu_1, \mu_2, \dots, \mu_m)$ is not empty. In Denuit et al. [76, Sec.4.1], this class of all possible moment sequences is denoted by $\mathcal{D}_m([0, b])$. Moreover, they

provided expressions for the topological interior, $\mathcal{D}_m \circ ([0, b])$, of $\mathcal{D}_m([0, b])$ up to $m = 4$. For completeness we cite the three cases relevant for our applications here, namely:

$$\mathcal{D}_1 \circ ([0, b]) = \{\mu_1 \in \mathbb{R} | 0 < \mu_1 < b\},$$

$$\mathcal{D}_2 \circ ([0, b]) = \{(\mu_1, \mu_2) \in \mathbb{R}^2 | \mu_1 \in \mathcal{D}_1 \circ ([0, b]) \text{ and } \mu_1^2 < \mu_2 < \mu_1 b\},$$

$$\mathcal{D}_3 \circ ([0, b]) = \{(\mu_1, \mu_2, \mu_3) \in \mathbb{R}^3 | (\mu_1, \mu_2) \in \mathcal{D}_2 \circ ([0, b]) \text{ and}$$

$$\frac{\sigma^2}{\mu_1}(\sigma^2 - \mu_1^2) - 2\mu_1^3 + 3\mu_1\mu_2 < \mu_3 < (b - \mu_1)\sigma^2 - \frac{\sigma^4}{b - \mu_1} - 2\mu_1^3 + 3\mu_1\mu_2\}.$$

Figure 4.2 depicts the sharp bounds for the ruin probability with $b = 100$, $\theta = 0.5$ and $s = 2/5$, i.e. $\mathbb{E}(U) = 2.5$. The upper left figure shows the bounds for $\mu_1 = 3.95$ as a function of μ_2 satisfying $(\mu_1, \mu_2) \in \mathcal{D}_2 \circ ([0, b])$. For this case, we also know the upper bound solution of the Schmitter problem with deterministic surplus and we can compare the two. It turns out that the upper bounds of the randomized and the deterministic case are remarkably close. The upper right figure shows the sharp ruin probability bounds for three given moments as a function of μ_3 satisfying $(3.95, 48.62, \mu_3) \in \mathcal{D}_3 \circ ([0, b])$. As remarked in the previous section, once sees that the ruin probability decreases with increasing third moment. As expected, the bounds are tighter as the knowledge of the second moment is incorporated. Finally, the graph at the bottom depicts the bounds of the generalized randomized Schmitter problem for given four moments of X as a function of μ_4 , leading to yet tighter bounds. It is worth noticing, that in this numerical illustration the values of the first three moments were selected in such a way that one finds a feasible set of distribution parameters for all of the distributions in the following numerical illustration.

In order to illustrate the performance of the bounds and how they improve with the addition of moment information, we explicitly calculate $\psi_U(s)$ for some chosen claim size distribution in each case, suitably truncated so that it fits into the model setup:

- **Case $m = 1$.** Truncated Exponential(λ) model with distribution function given by

$$F_X(x) = \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda b}}, \quad 0 < x \leq b, \quad \lambda > 0,$$

and Laplace transform

$$M_X(-s) = \frac{\lambda}{\lambda + s} \frac{1 - e^{-(\lambda+s)b}}{1 - e^{-\lambda b}}.$$

- **Case $m = 2$.**
 - Truncated Gamma(α, β) model with distribution function

$$F_X(x) = \frac{\gamma(\alpha, \beta x)}{\gamma(\alpha, \beta b)}, \quad 0 < x \leq b, \quad \alpha, \beta > 0,$$

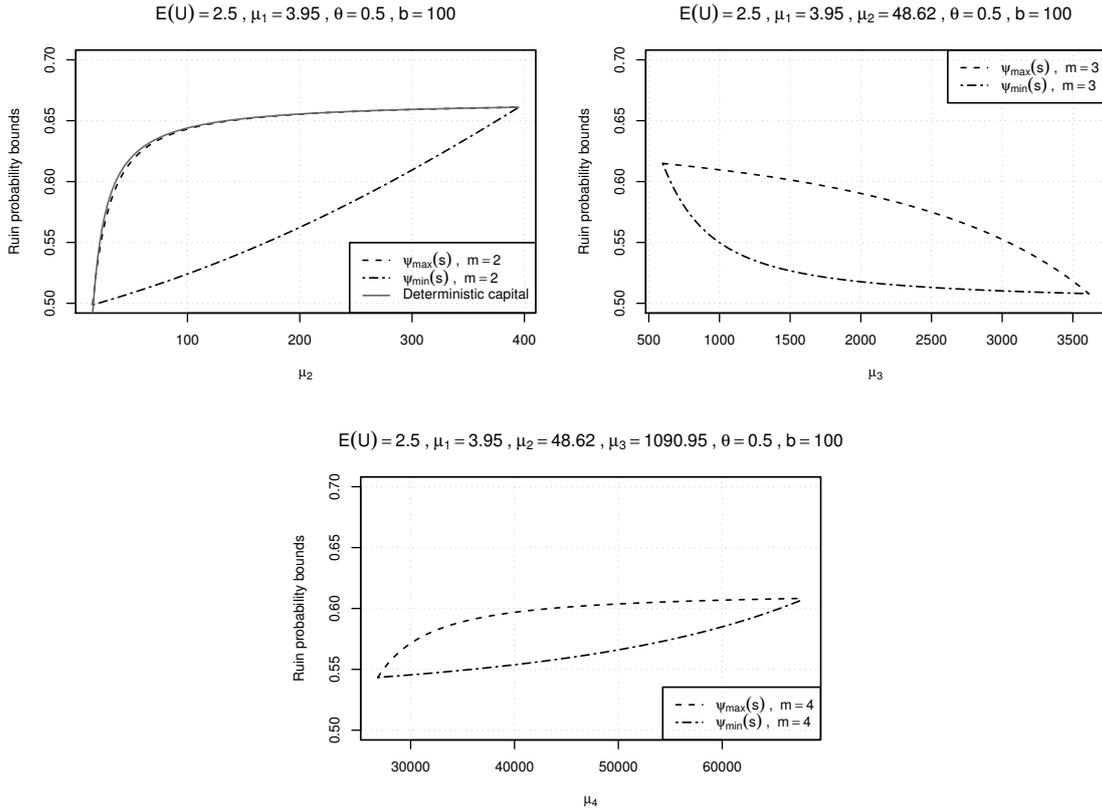


Figure 4.2: Sharp bounds for the randomized ruin probability $\psi_U(2/5)$, considered as a function of μ_2 , μ_3 and μ_4 respectively.

and Laplace transform

$$M_X(-s) = \left(\frac{\beta}{\beta + s} \right)^\alpha \frac{\gamma(\alpha, (\beta + s)b)}{\gamma(\alpha, \beta b)},$$

where $\gamma(\alpha, x) = \int_0^x z^{\alpha-1} e^{-z} dz$ is the lower incomplete gamma function.

– Truncated US-Pareto(α, η) (Lomax) model with distribution function

$$F_X(x) = \frac{1 - \left(\frac{\eta}{\eta+x} \right)^\alpha}{1 - \left(\frac{\eta}{\eta+b} \right)^\alpha}, \quad 0 \leq x \leq b, \quad \alpha, \eta > 0,$$

and Laplace transform

$$M_X(-s) = \frac{\alpha(\eta s)^\alpha e^{\eta s}}{1 - \left(\frac{\eta}{\eta+b} \right)^\alpha} (\Gamma(-\alpha, \eta s) - \Gamma(-\alpha, (\eta + b)s)),$$

where $\Gamma(\alpha, x) = \int_x^\infty z^{\alpha-1} e^{-z} dz$ is the upper incomplete gamma function.

- **Case** $m = 3$. Truncated generalized Gamma(α, β, τ) model with density and distribution function given by

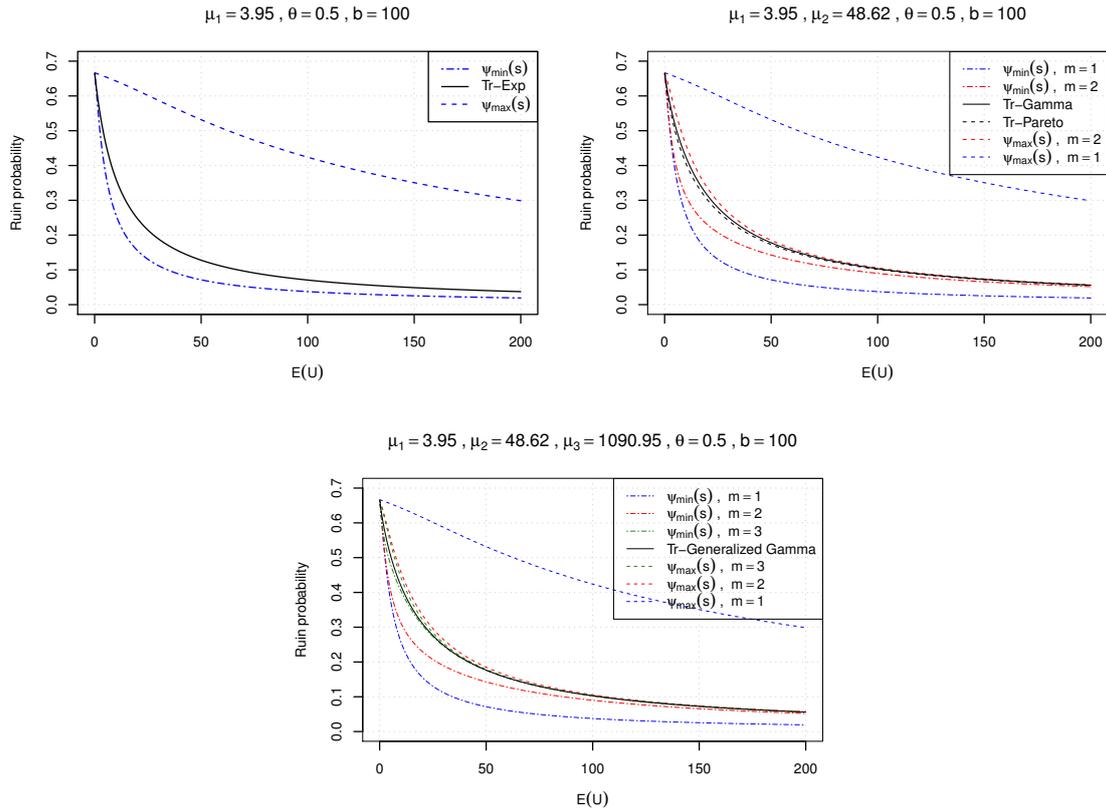
$$f_X(x) = \frac{\tau x^{\alpha\tau-1} \beta^{-\alpha\tau} e^{-(x/\beta)^\tau}}{\gamma(\alpha, (b/\beta)^\tau)}, \quad F_X(x) = \frac{\gamma(\alpha, (x/\beta)^\tau)}{\gamma(\alpha, (b/\beta)^\tau)}, \quad 0 < x \leq b, \quad \alpha, \beta\tau > 0,$$

and Laplace transform

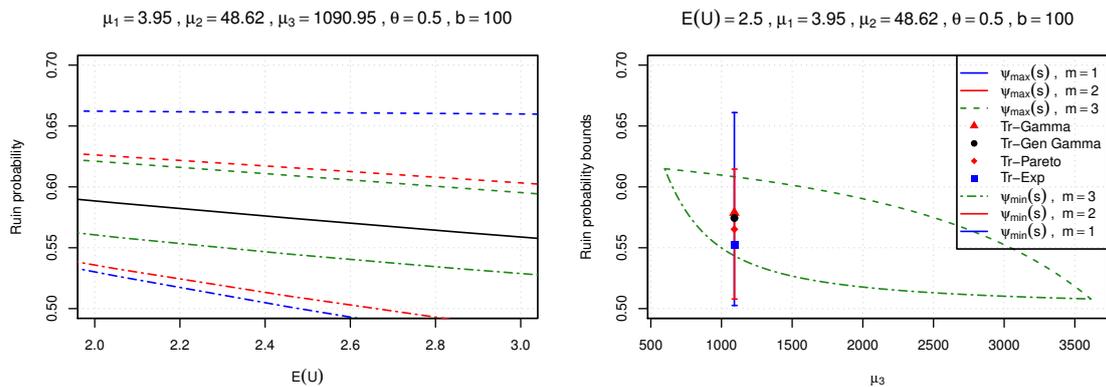
$$M_X(-s) = \sum_{k=0}^{\infty} \frac{(-\beta s)^k}{k!} \frac{\gamma(\alpha + k/\tau, (b/\beta)^\tau)}{\gamma(\alpha, (b/\beta)^\tau)}.$$

In each case, the distribution parameters were determined using the method of moments for the given moment values set in the example above. For further details on claim size distributions and truncation see, for example, [7, Sec.3.3 & Ch.4].

The results are given in Figure 4.3d where the exact ruin probabilities obtained using (4.4) together with the general bounds are plotted as a function of the expected initial surplus $\mathbb{E}(U) = 1/s$ for the same set of parameters as above. In particular, b was selected in such a way that no strong truncation effect is present in the distributions. One sees that, for fixed μ_1 only, the truncated exponential case is nicely between the sharp bounds. However, these bounds are very wide. When information about the second moment of X is included, the tightness of the bounds improves significantly. From (4.5), one would expect that to be the case only for small values of s where information about the first two moments provides a good approximation for the ruin probability. However, we can see that even for large values of s the improvement is considerable. The tightness of the interval for possible ruin probabilities becomes even more remarkable when the first three moments are fixed. This illustrates that in the context of ruin probabilities, the knowledge of the first few moments of the claim size distribution already provides a very accurate approximation. In a broader statistical context, for an account on reconstructions of arbitrary distributions from given moments, see e.g. Mnatsakanov [147]. Finally, for recent progress on the general probability level concerning criteria of moment-determinacy of distributions, see Yarovaya et al. [184].



(d) Ruin probabilities $\psi_U(s)$ as a function of the expected initial surplus. *Top left:* Bounds for $\psi_U(s)$ together with the case of truncated exponential claims for given μ_1 . *Top right:* Bounds for $\psi_U(s)$ together with the case of truncated gamma and truncated US-Pareto claims for given μ_1 and μ_2 . *Bottom:* Bounds for $\psi_U(s)$ together with truncated generalized gamma claims when μ_1, μ_2 and μ_3 are given.



(g) *Left:* Magnification around $\mathbb{E}(U) = 2.5$ of the ruin probabilities $\psi_U(s)$ and the upper/lower bounds in Figure 4.3d (bottom). *Right:* Bounds for $\psi_U(2/5)$, considered as a function of μ_3 (Figure 4.2 (top right)), with a vertical bar at $\mu_3 = 1090.95$ summarizing the upper/lower bounds and ruin probabilities in Figure 4.3d at $\mathbb{E}(U) = 2.5$.

Figure 4.3: Numerical illustration for bounds of ruin the ruin probability under exponentially distributed initial capital.

4.4 Erlang distributed initial surplus

A natural extension of exponentially distributed random surplus is now to consider Erlang distributed initial surplus. Concretely, consider U to be an Erlang(k, s) random variable E_k with density

$$f_{E_k}(x) = \frac{1}{(k-1)!} s^k x^{k-1} e^{-sx} \quad \text{for } k \geq 1, s > 0, x > 0,$$

We then get

$$\psi_E(k, s) := \mathbb{E}(\psi(E_k)) = \int_0^\infty \psi(u) \frac{s^k}{(k-1)!} u^{k-1} e^{-su} du = 1 + \frac{(-s)^k}{(k-1)!} \frac{\partial^{(k-1)}}{\partial s^{(k-1)}} \hat{\phi}(s).$$

Here $\hat{\phi}(s) = 1/s - \hat{\psi}(s)$ denotes the Laplace transform of the survival probability of the classical Cramér-Lundberg risk process and we observe that its derivatives w.r.t. the Laplace argument lead to an explicit expression for the case of random Erlang-distributed initial surplus. We focus here on the classical Schmitter setting with fixed mean and variance of the claim size distribution. In Figure 4.4 we depict the ruin probabilities for Erlang(k, s) distributed initial surplus for 2-point distributions as a function of x_1 for a given mean and variance, for two expected surplus levels.

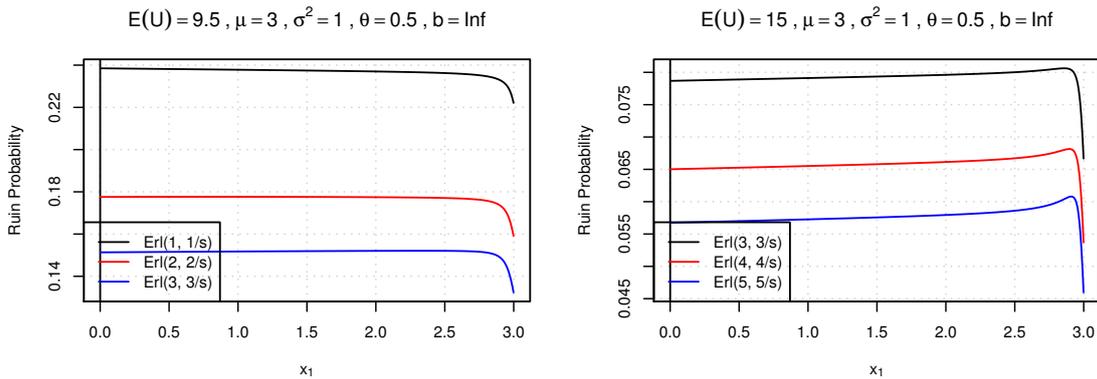


Figure 4.4: Ruin probabilities for Erlang($k, k/u$) distributed surplus as a function of x_1 for $\mu = 3$, $\sigma^2 = 1$, $\theta = 0.5$ and $\mathbb{E}(U)$ as specified.

In contrast to the exponential case ($k = 1$), there is unfortunately no direct relation between the optimization problem and the minimization (maximization) of the Laplace transform of the ruin probability. What we obtain is in fact an expression in terms of its $(k-1)$ -th derivative (with $\partial^{(0)}/\partial s^{(0)} \hat{\phi}(s) = \hat{\phi}(s)$). For example, for $k = 2$ we get

$$\begin{aligned} \psi_E(2, s) &= 1 + s^2 \frac{\partial}{\partial s} \hat{\phi}_u(s) = 1 + s^2 \frac{\partial}{\partial s} \frac{c - \lambda\mu}{cs - \lambda(1 - M_X(-s))} \\ &= 1 - s^2 \frac{c + \lambda \frac{\partial}{\partial s} M_X(-s)}{c - \lambda\mu} \left(\frac{c - \lambda\mu}{cs - \lambda(1 - M_X(-s))} \right)^2, \quad s > 0. \end{aligned}$$

Thus, for a fixed parameter s , in order to maximize our ruin probability, we need to minimize an expression that depends on both the Laplace transform of X and its first derivative.

Since the variance of a Erlang distribution goes to 0 as $k \rightarrow \infty$, one particular motivation to consider Erlang distributed initial surplus is as a tool to approximate the case of deterministic initial surplus, as in fact one has $\psi_E(k, s) \rightarrow \psi(u)$ as $k \rightarrow \infty$. The approximation $\psi(u) \approx \psi_E(k, s)$, or *Erlang smoothing*, was considered in Asmussen et al. [23] as a numerical scheme to approximate the finite horizon ruin probability by replacing the deterministic time horizon T by an “standardized” $\text{Erlang}(k, k/T)$ random variable, which for $k \rightarrow \infty$ becomes exact (see [22, Ch.IX.8] for a more general discussion, as well as Stanford et al. [173], Carr [52] and Kyprianou and Pistorius [135] for applications of this approach to other fields). Concerning the convergence rate with increasing k , for our context of random initial surplus one can adapt Theorem 6 of Asmussen et al. [23] in a straight-forward way to obtain the following result:

Proposition 4.4.1. *Let $u > 0$ be the expected initial surplus and let E_k denote the Erlang distribution with shape parameter k and mean u . Then $\psi_E(k, s) \rightarrow \psi(u)$ as $k \rightarrow \infty$. More precisely, for some constant C*

$$\psi_E(k, k/u) = \psi(u) + \frac{C}{k} + O(k^{-2}). \quad (4.9)$$

As already suggested in [23], a further improvement of accuracy for fixed k can be obtained by Richardson extrapolation. This is a general method (see e.g. Press et al. [155] for details) for computing an abstract quantity y (it could be an integral, a derivative, etc.) accurately using a sequence $y_k \rightarrow y$ for which the convergence rate is known,

$$y_k = y + \frac{c_1}{k} + \frac{c_2}{k^{1+\epsilon}} + \dots,$$

where c_1 is typically unknown but can be eliminated. In fact, setting $\tilde{y}_k = (k+1)y_{k+1} - ky_k$, we get that $\tilde{y}_k \rightarrow y$ and one obtains an improved approximation of convergence rate $O(k^{-1-\epsilon})$.

Translated into our context, we then get

$$\psi(u) \approx (k+1)\psi_E(k+1, (k+1)/u) - k\psi_E(k, k/u), \quad (4.10)$$

with an error rate of order $1/k^2$.

For an illustration of the method, consider the same example as in Figure 4.1, namely the set of 2-point distributions with mean $\mu = 3$, variance $\sigma^2 = 1$, and safety loading $\theta = 0.5$. Figure 4.5 shows the results of the approximation. One observes that the approximation of the deterministic case via the randomized initial surplus is quite satisfactory already for $k = 11$. The numerical approximation works well even for intermediate values of the initial surplus for which the ruin probability (and its kink) is difficult to approximate. In order to also reproduce the particular shape of that curve, higher values of k are however needed. It is worth to note the tremendous

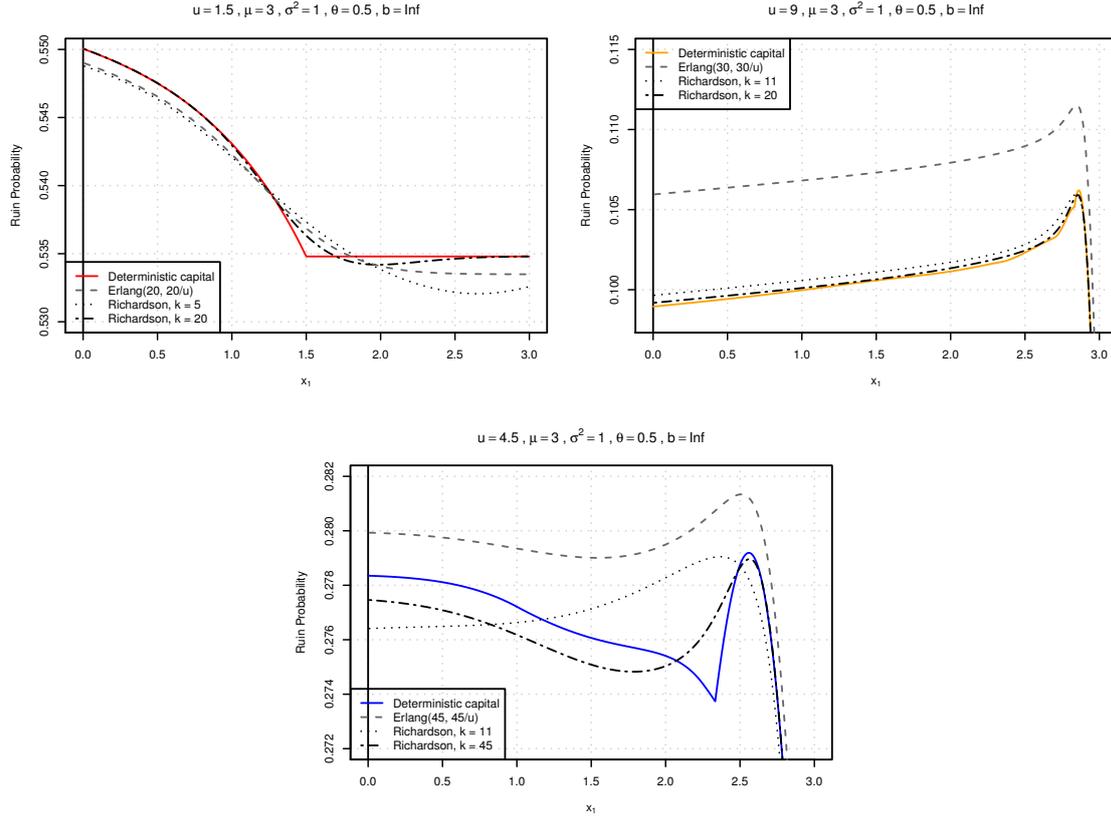


Figure 4.5: Ruin probabilities as a function of x_1 for $\mu = 3$, $\sigma^2 = 1$, $\theta = 0.5$ and three levels of initial surplus u (expected surplus $\mathbb{E}(U) = 1/s$, respectively).

improvement when employing Richardson extrapolation for larger values of u (cf. the graph for $u = 9$).

Remark. Analogous to the exponential initial surplus case, one can obtain an expression for the ruin probability in terms of moments of L . Concretely,

$$\begin{aligned}
 \psi_E(k, s) &= \int_0^\infty \psi(u) \frac{s^k}{(k-1)!} u^{k-1} e^{-su} du = \int_0^\infty \psi(u) \frac{s^k}{(k-1)!} \left((-1)^{k-1} \frac{\partial^{k-1}}{\partial s^{k-1}} e^{-su} \right) du \\
 &= \frac{s(-s)^{k-1}}{(k-1)!} \frac{\partial^{k-1}}{\partial s^{k-1}} \int_0^\infty \psi(u) e^{-su} du \\
 &= \frac{s(-s)^{k-1}}{(k-1)!} \frac{\partial^{k-1}}{\partial s^{k-1}} \sum_{j=0}^\infty \frac{(-s)^j}{j!} \int_0^\infty u^j \mathbb{P}(L > u) du \\
 &= \frac{s(-s)^{k-1}}{(k-1)!} \frac{\partial^{k-1}}{\partial s^{k-1}} \sum_{j=0}^\infty \frac{(-s)^j}{(j+1)!} \mathbb{E}(L^{j+1}) \\
 &= \sum_{j \geq k-1} (-1)^{j+k-1} \frac{s^{j+1}}{(j+1)!} \binom{j}{k-1} \mathbb{E}(L^{j+1}).
 \end{aligned}$$

Therefore, one might try to understand the behavior of the $\psi_E(k, s)$ by analyzing

the first, say, two terms of the previous series to obtain the approximation

$$\psi_E(k, s) \approx \frac{s^k}{k!} \mathbb{E}(L^k) - \frac{s^{k+1}}{(k+1)!} k \mathbb{E}(L^{k+1}).$$

In the exponential case $k = 1$, as the second moment of l involves the first three moments of X , this means that for given μ_1, μ_2 one could infer about the behavior of $\psi_U(s)$ by simply analyzing the first non-given moment, i.e. μ_3 . We see that the same line of reasoning applied to the Erlang(k) case needs the $(k+2)$ -th moment of X , already for the above first two terms. This is unfortunate, as the deterministic $\psi(u)$ will only be obtained for $k \rightarrow \infty$, and we see that even in this simple approximation higher-order moments of X already play a crucial role. This is in particular the case for moderate values of $\mathbb{E}(U)$, and in those cases we have indeed seen in the graphs above that a good approximation of the deterministic case needed large values of k .

4.5 Conclusion

In this paper we showed how randomization can be used to provide a solution to the Schmitter problem in ruin theory and its extension to higher moments. Linking this problem with established results in the theory of m -convex stochastic orders, we provided sharp bounds for the ruin probability under the assumption of an exponential initial surplus. For the more general case of Erlang distributed initial surplus, such analytical sharp bounds are not within reach. However, we showed how the deterministic classical case can be approximated by the simple expressions of the randomized case using Erlangization.

Bibliography

- [1] I. B. Aban, M. M. Meerschaert, and A. K Panorska. Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association*, (473):270–277, 2006.
- [2] H. Albrecher. Operational time. In *Encyclopedia of Actuarial Science*, volume 3, pages 1207–1208. Wiley, Chichester, 2004.
- [3] H. Albrecher and J.C. Araujo-Acuna. On the randomized Schmitter problem. *Submitted*, 2020.
- [4] H. Albrecher, J.C. Araujo-Acuna, and J. Beirlant. Fitting nonstationary Cox processes: An application to fire insurance data. *North American Actuarial Journal*. *To appear*, 2020.
- [5] H. Albrecher, J.C. Araujo-Acuna, and J. Beirlant. Tempered Pareto-type modeling using Weibull distributions. *ASTIN Bulletin*. *To appear*, 2020.
- [6] H. Albrecher and S. Asmussen. Ruin probabilities and aggregate claims distributions for shot noise Cox processes. *Scandinavian Actuarial Journal*, 2006(2):86–110, 2006.
- [7] H. Albrecher, J. Beirlant, and J. Teugels. *Reinsurance: Actuarial and Statistical Aspects*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2017.
- [8] H. Albrecher, M. Bladt, and M. Bladt. Matrix Mittag-Leffler distributions and modeling heavy-tailed risks. *Extremes*, 23(3), 2020.
- [9] H. Albrecher, M. Bladt, D. Kortschak, F. Prettenthaler, and T. Swierczynski. Flood Occurrence Change-Point Analysis in the Paleoflood Record from Lake Mondsee (NE Alps). *Global and Planetary Change*, 178:65 – 76, 2019.
- [10] H. Albrecher, E. C. K. Cheung, and S. Thonhauser. Randomized observation periods for the compound Poisson risk model: the discounted penalty function. *Scandinavian Actuarial Journal*, 2013(6):424–452, 2013.
- [11] H Albrecher and P.-O. Goffard. On the profitability of selfish blockchain mining under consideration of ruin. *arXiv Preprint*, 2020. arXiv:2010.12577.

- [12] H. Albrecher and C. Hipp. Lundberg's risk process with tax. *Blätter der DGVMF*, 28(1):13–28, 2007.
- [13] H. Albrecher, C. Hipp, and D. Kortschak. Higher-order expansions for compound distributions and ruin probabilities with subexponential claims. *Scandinavian Actuarial Journal*, 2010(2):105–135, 2010.
- [14] H. Albrecher and J. Ivanovs. Strikingly simple identities relating exit problems for Lévy processes under continuous and Poisson observations. *Stochastic Processes and their Applications*, 127(2):643 – 656, 2017.
- [15] H. Albrecher and D. Kortschak. On ruin probability and aggregate claim representations for Pareto claim size distributions. *Insurance: Mathematics and Economics*, 45(3):362–373, 2009.
- [16] H. Albrecher and S. Thonhauser. Optimality results for dividend problems in insurance. *Rev. Real Acad. Cienc. Exactas, Fis. y Nat.*, 103(2):295–320, 2009.
- [17] H. Ammeter. A Generalization of the Collective Theory of Risk in Regard to Fluctuating Basic-probabilities. *Scandinavian Actuarial Journal*, 31:171–198, 1948.
- [18] K. Antonio and J. Beirlant. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, 40(1):58–76, 2007.
- [19] K. Antonio and E. A. Valdez. Statistical concepts of a priori and a posteriori risk classification in insurance. *Advances in Statistical Analysis*, 96(2):187–224, 2012.
- [20] D. Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2009.
- [21] S. Asmussen. Risk theory in a Markovian environment. *Scandinavian Actuarial Journal*, 1989(2):69–100, 1989.
- [22] S. Asmussen and H. Albrecher. *Ruin Probabilities*. Advanced series on statistical science & applied probability, 14. World scientific, 2nd edition, 2009.
- [23] S. Asmussen, F. Avram, and M. Usabel. Erlangian approximations for finite-horizon ruin probabilities. *ASTIN Bulletin*, 32(2):267–281, 2002.
- [24] S. Asmussen and M. Steffensen. *Risk and Insurance: A Graduate Text*. Springer, 2020.
- [25] B. Avanzi. Strategies for dividend distribution: A review. *North American Actuarial Journal*, 13(2):217–251, 2009.
- [26] B. Avanzi, G. Taylor, B. Wong, and A. Xian. Modelling and understanding count processes through a Markov-modulated non-homogeneous Poisson process framework. *European Journal of Operational Research*, 290(1):177 – 195, 2021.

- [27] B. Avanzi, G. Taylor, B. Wong, and X. Yang. On the modelling of multivariate counts with Cox processes and dependent shot noise intensities. *Insurance: Mathematics and Economics*, 2021.
- [28] B. Avanzi, B. Wong, and X. Yang. A micro-level claim count model with overdispersion and reporting delays. *Insurance: Mathematics and Economics*, 71:1 – 14, 2016.
- [29] P. Azcue and N. Muler. *Stochastic optimization in insurance: a dynamic programming approach*. Springer, 2014.
- [30] A. Badescu, S. Lin, and D. Tang. A marked Cox model for the number of IBNR claims: Estimation and application. *ASTIN Bulletin*, pages 1–31, 2016.
- [31] A. Badescu, S. Lin, and D. Tang. A marked Cox model for the number of IBNR claims: Theory. *Insurance: Mathematics and Economics*, 69:29–37, 2016.
- [32] A. Badescu and D. Stanford. A generalization of the De Vylder approximation for the probability of ruin. *Economic Computation and Economic Cybernetics Studies and Research*, 40(3–4):245–265, 2006.
- [33] J. Beirlant, M. I. Fraga Alves, and I. Gomes. Tail fitting for truncated and non-truncated Pareto-type distributions. *Extremes*, (3):429–462, 2016.
- [34] J. Beirlant, M. I. Fraga Alves, and T. Reynkens. Fitting tails affected by truncation. *Electronic Journal of Statistics*, (1):2026–2065, 2017.
- [35] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2006.
- [36] J. Beirlant, J. Teugels, and P. Vynckier. *Practical Analysis of Extreme Values*. Leuven University Press Leuven, 1996.
- [37] J. Bertoin. *Lévy processes*, volume 121. Cambridge university press, 1996.
- [38] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [39] T. Björk and J. Grandell. Exponential inequalities for ruin probabilities in the Cox case. *Scandinavian Actuarial Journal*, 1988(1-3):77–111, 1988.
- [40] M. Bladt, H. Albrecher, and J. Beirlant. Threshold selection and trimming in extremes. *Extremes*, 23(4):629–665, 2020.
- [41] K. Borch. The utility concept applied to the theory of insurance. *ASTIN Bulletin*, 1(5):245–255, 1961.
- [42] J.P. Boucher, M. Denuit, and M. Guillén. Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. *Variance*, 2(1):135–162, 2008.

- [43] N. L. Bowers. An upper bound on the stop-loss net premium–actuarial note. *Transactions of the Society of Actuaries*, 21, 1969.
- [44] V. Brazauskas and A. Kleefeld. Modeling severity and measuring tail risk of norwegian fire claims. *North American Actuarial Journal*, (1):1–16, 2016.
- [45] P. L. Brockett and S. H. Cox. Optimal ruin calculations using partial stochastic information. *Transactions of the Society of Actuaries*, 1984.
- [46] P. L. Brockett and S. H. Cox. Insurance calculations using incomplete information. *Scandinavian Actuarial Journal*, 2:94–108, 1985.
- [47] P. L. Brockett, M. J. Goovaerts, and G. C. Taylor. The Schmitter problem. *ASTIN Bulletin*, 21(1):129–132, 1991.
- [48] H. Bühlmann. *Mathematical Methods in Risk Theory*, volume 172. Springer-Verlag, 1970.
- [49] H. Bühlmann, B. Gagliardi, H. U. Gerber, and E. Straub. Some inequalities for stop-loss premiums. *ASTIN Bulletin*, 9(1-2):75–83, 1977.
- [50] J. Cai and J. Garrido. Aging properties and bounds for ruin probabilities and stop-loss premiums. *Insurance: Mathematics and Economics*, 23(1):33 – 43, 1998.
- [51] A. C. Cameron and P. K. Trivedi. *Regression analysis of count data*. Cambridge university press, 1998.
- [52] P. Carr. Randomization and the American put. *The Review of Financial Studies*, 11(3):597–626, 1998.
- [53] V. Chavez-Demoulin, A. Davison, and A. McNeil. Estimating Value-at-Risk: A point process approach. *Quantitative Finance*, 5(2):227–234, 2005.
- [54] S. Coles. *An introduction to statistical modeling of extreme values*. Springer, 2001.
- [55] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. CRC press, 2003.
- [56] C. Courtois, M. Denuit, and S. Van Bellegem. Discrete s-convex extremal distributions: Theory and applications. *Applied Mathematics Letters*, 19(12):1367 – 1377, 2006.
- [57] D. Cox. Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society*, 17:129–157, 1955.
- [58] D. Cox and V. Isham. *Point Processes*, volume 12. Chapman & Hall/CRC, 1980.

- [59] A. Dassios and J. Jang. Pricing of catastrophe reinsurance and derivatives using the Cox process with shot noise intensity. *Finance and Stochastics*, 7(1):73–95, 2003.
- [60] A. Dassios and J. Jang. Kalman-Bucy filtering for linear systems driven by the Cox process with shot noise intensity and its application to the pricing of reinsurance contracts. *Journal of applied probability*, 42(1):93–107, 2005.
- [61] C. D. Daykin, T. Pentikäinen, and M. Pesonen. *Practical Risk Theory for Actuaries*. CRC Press, 1993.
- [62] B. De Finetti. Su un'impostazione alternativa della teoria collettiva del rischio. In *Transactions of the XVth international congress of Actuaries*, volume 2, pages 433–443, 1957.
- [63] L. de Haan. *On regular variation and its application to the weak convergence of sample extremes*. Mathematical Centre Tract 32, Amsterdam, 1970.
- [64] L. de Haan and A. Ferreira. *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.
- [65] E. De Vylder. A practical solution to the problem of ultimate ruin probability. *Scandinavian Actuarial Journal*, 1978(2):114–119, 1978.
- [66] E. De Vylder. Best upper bounds for integrals with respect to measures allowed to vary under conical and integral constraints. *Insurance: Mathematics and Economics*, 1(2):109–130, 1982.
- [67] E. De Vylder. Maximization, under equality constraints, of a functional of a probability distribution. *Insurance: Mathematics and Economics*, 2(1):1–16, 1983.
- [68] E. De Vylder. *Advanced risk theory: A self-contained introduction*. Editions de l'Université de Bruxelles, 1996.
- [69] E. De Vylder, M. Goovaerts, and E. Marceau. The bi-atomic uniform minimal solution of Schmitter's problem. *Insurance: Mathematics & Economics*, 20(1):59–78, 1997.
- [70] E. De Vylder, M. Goovaerts, and E. Marceau. The solution of Schmitter's simple problem: Numerical illustration. *Insurance: Mathematics & Economics*, 20(1):43–58, 1997.
- [71] E. De Vylder and M. J. Goovaerts. Analytical best upper bounds on stop-loss premiums. *Insurance: Mathematics and Economics*, 1(3):163–175, 1982.
- [72] E. De Vylder and E. Marceau. The numerical solution of Schmitter problems: Theory. *Insurance: Mathematics & Economics*, 19(1):1–18, 1996.
- [73] F. Delbaen and J. Haezendonck. Classical risk theory in an economic environment. *Insurance: Mathematics and Economics*, 6(2):85–116, 1987.

- [74] M. Denuit. The exponential premium calculation principle revisited. *ASTIN Bulletin*, 29(2):215–226, 1999.
- [75] M. Denuit. Laplace transform ordering of actuarial quantities. *Insurance: Mathematics and Economics*, 29(1):83–102, 2001.
- [76] M. Denuit, E. De Vylder, and C. Lefèvre. Extremal generators and extremal distributions for the continuous s-convex stochastic orderings. *Insurance: Mathematics and Economics*, 24(3):201–217, 1999.
- [77] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. *Actuarial theory for dependent risks: Measures, orders and models*. John Wiley & Sons, 2006.
- [78] M. Denuit, D. Hainaut, and J. Trufin. *Effective statistical learning methods for actuaries I: GLMs and extensions*. Springer Actuarial. Springer, 2019.
- [79] M. Denuit, D. Hainaut, and J. Trufin. *Effective statistical learning methods for actuaries III: Neural Networks and Extensions*. Springer Actuarial. Springer, 2019.
- [80] M. Denuit, D. Hainaut, and J. Trufin. *Effective statistical learning methods for actuaries II: Tree-Based Methods and Extensions*. Springer Actuarial. Springer, 2020.
- [81] M. Denuit and S. Lang. Non-life rate-making with Bayesian GAMs. *Insurance: Mathematics and Economics*, 35(3):627–647, 2004.
- [82] M. Denuit and C. Lefèvre. Some new class of stochastic order relations among arithmetic random variables, with applications in actuarial science. *Insurance: Mathematics and Economics*, 20:197–213, 1997.
- [83] M. Denuit, C. Lefèvre, and M. Mesfioui. On s-convex stochastic extrema for arithmetic risks. *Insurance: Mathematics and Economics*, 25(2):143–155, 1999.
- [84] M. Denuit, C. Lefèvre, and M. Scarsini. On s-convexity and risk aversion. *Theory and Decision*, 50(3):239–248, 2001.
- [85] M. Denuit, C. Lefèvre, and M. Shaked. The s-convex orders among real random variables, with applications. *Mathematical Inequalities and Their Applications*, 1(4):585–613, 1998.
- [86] M. Denuit, C. Lefèvre, and M. Shaked. On s-convex approximations. *Advances in Applied Probability*, 32(4):994–1010, 2000.
- [87] M. Denuit, X. Maréchal, S. Pitrebois, and J. Walhin. *Actuarial Modelling of Claim Counts*. John Wiley & Sons, 2007.
- [88] D. C. M. Dickson. *Insurance Risk and Ruin*. Cambridge University Press, 1999.

- [89] J. Dubourdieu. Remarques relatives à la théorie mathématique de l'assurance-accidents. *Bull. Trim. Inst. Actu. Français*, 44:79–146, 1938.
- [90] F. Dufresne and H. U. Gerber. Risk theory for the compound Poisson process that is perturbed by diffusion. *Insurance: Mathematics and Economics*, 10(1):51–59, 1991.
- [91] F. Dufresne, H. U. Gerber, and E. S.W. Shiu. Risk theory with the gamma process. *ASTIN Bulletin: The Journal of the IAA*, 21(2):177–192, 1991.
- [92] A. E. Eckberg. Sharp bounds on laplace-stieltjes transforms, with applications to various queueing problems. *Mathematics of Operations Research*, 2(2):135–142, 1977.
- [93] N. El Karoui, S. Loisel, and Y. Salhi. Minimax optimality in robust detection of a disorder time in doubly-stochastic Poisson processes. *The Annals of Applied Probability*, 27:2515–2538, 2017.
- [94] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.
- [95] P. Embrechts and N. Veraverbeke. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics*, 1(1):55–72, 1982.
- [96] R. Fisher and L. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. 24(2):180–190, 1928.
- [97] E. W. Frees. *Regression modeling with actuarial and financial applications*. International Series on Actuarial Science. Cambridge University Press, Cambridge, 2010.
- [98] E. W. Frees, V. R. Young, and Y. Luo. A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24(3):229–247, 1999.
- [99] E. W. Frees, V. R. Young, and Y. Luo. Case studies using panel data models. *North American Actuarial Journal*, 5(4):24–42, 2001.
- [100] T. C. Fung, A. L. Badescu, and S. Lin. Multivariate Cox hidden markov models with an application to operational risk. *Scandinavian Actuarial Journal*, 2019(8):686–710, 2019.
- [101] H. U. Gerber. *Entscheidungskriterien für den zusammengesetzten Poisson-Prozess*. PhD thesis, ETH Zurich, 1969.
- [102] H. U. Gerber. *An Introduction to Mathematical Risk Theory*. S.S. Huebner Foundation Monographs, University of Pennsylvania, 1979.

- [103] H. U. Gerber, E. S. W. Shiu, and N. Smith. Methods for estimating the optimal dividend barrier and the probability of ruin. *Insurance: Mathematics and Economics*, 42(1):243 – 254, 2008.
- [104] B. Gnedenko. Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.
- [105] M. Goovaerts, E. De Vylder, and J. Haezendonck. Ordering of risks: a review. *Insurance: Mathematics and Economics*, 1(2):131–161, 1982.
- [106] C. Gourieroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3):681–700, 1984.
- [107] J. Grandell. *Doubly stochastic Poisson processes*, volume 529. Springer, 1976.
- [108] J. Grandell. *Aspects of Risk Theory*. Springer-Verlag, 1991.
- [109] J. Grandell. *Mixed Poisson Processes*. CRC Press, 1997.
- [110] A. Guillou, S. Loisel, and G. Stupfler. Estimation of the parameters of a Markov-modulated loss process in insurance. *Insurance: Mathematics and Economics*, 53(2):388–404, 2013.
- [111] A. Guillou, S. Loisel, and G. Stupfler. Estimating the parameters of a seasonal Markov-modulated Poisson process. *Statistical Methodology*, 26:103–123, 2015.
- [112] P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(1):37–42, 1982.
- [113] J. A. Hausman, B. H. Hall, and Z. Griliches. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52(4):909–938, 1984.
- [114] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, (5):1163–1174, 1975.
- [115] P. J. Huber and E. M. Ronchetti. *Robust statistics*. John Wiley & Sons, Inc., 2009.
- [116] W. Hürlimann. Extremal moment methods and stochastic orders. *Application in Actuarial Science. Monograph manuscript*, 1998.
- [117] J. Ivanovs. A note on killing with applications in risk theory. *Insurance: Mathematics and Economics*, 52(1):29–34, 2013.
- [118] J. Jang and R. Oh. A review on Poisson, Cox, Hawkes, shot-noise Poisson and dynamic contagion process and their compound processes. *Annals of Actuarial Science*, pages 1–22, 2020.
- [119] A. F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171, 1955.

- [120] N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate discrete distributions*, volume 444. 2005.
- [121] D. W. Jorgenson. Multiple regression analysis of a Poisson process. *Journal of the American Statistical Association*, 56:235–245, 1961.
- [122] R. Kaas. The Schmitter problem and a related problem: A partial solution. *ASTIN Bulletin*, 21(1):133–146, 1991.
- [123] R. Kaas and M. J. Goovaerts. Best bounds for positive distributions with fixed moments. *Insurance: Mathematics and Economics*, 5(1):87 – 92, 1986.
- [124] R. Kaas, M. J. Goovaerts, J. Dhaene, and M. Denuit. *Modern actuarial risk theory: using R*. Springer Science & Business Media, 2008.
- [125] R. Kaas and M.J. Goovaerts. Bounds on stop-loss premiums for compound distributions. *ASTIN Bulletin*, 16(1):13–17, 1986.
- [126] R. Kaas, A. E. Van Heerwaarden, and M. Goovaerts. *Ordering of actuarial risks*. CAIRE. Brussels, 1994.
- [127] R. Kaas, M. Vanneste, and M. J. Goovaerts. Maximizing compound Poisson stop-loss premiums numerically with given mean and variance. *ASTIN Bulletin*, 22(2):225–233, 1992.
- [128] V. Kalashnikov. Bounds for ruin probabilities in the presence of large claims and their comparison. *North American Actuarial Journal*, 3(2):116–128, 1999.
- [129] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss models. Further topics*. John Wiley & Sons, 2013.
- [130] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss models: From data to decisions*. John Wiley & Sons, 5th edition edition, 2019.
- [131] R. Korn, E. Korn, and G. Kroisandt. *Monte Carlo Methods and Models in Finance and Insurance*. CRC Press, 2010.
- [132] D. Kortschak and H. Albrecher. An asymptotic expansion for the tail of compound sums of Burr distributed random variables. *Statistics & probability letters*, 80(7-8):612–620, 2010.
- [133] A. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer Science & Business Media, 2006.
- [134] A. Kyprianou. *Gerber–Shiu Risk Theory*. Springer Science & Business Media, 2013.
- [135] A. E. Kyprianou and M. R. Pistorius. Perpetual options and Canadization through fluctuation theory. *The Annals of Applied Probability*, 13(3):1077–1098, 2003.

- [136] D. Lando. On Cox processes and credit risky securities. *Review of Derivatives research*, 2(2-3):99–120, 1998.
- [137] C. Lefèvre, M. Hallin, and P. Narayan. On fractional linear bounds for probability generating functions. *Journal of Applied Probability*, 23:904–913, 1986.
- [138] C. Lefèvre, S. Loisel, and P. Montesinos. Bounding basis risk using s-convex orders on Beta-unimodal distributions. *Preprint. HAL Id: hal-02611208*, 2020.
- [139] J. Li, L. Li, and R. Mendoza-Arriaga. Additive Subordination and its Applications in Finance. *Finance and Stochastics*, 20(3):589–634, 2016.
- [140] C.-Y. Liu. Claim count modelling with shot noise Cox processes. 2012.
- [141] Y. Lu and J. Garrido. Doubly Periodic Non-Homogeneous Poisson Models for Hurricane Data. *Statistical Methodology*, 2(1):17–35, 2005.
- [142] J. A. F. Machado and J. M. C. Santos Silva. Quantiles for counts. *Journal of the American Statistical Association*, 100(472):1226–1237, 2005.
- [143] J. F. Mai, S. Schenk, and M. Scherer. Exchangeable exogenous shock models. *Bernoulli*, 22(2):1278–1299, 2016.
- [144] A. J. McNeil. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27(1):117–137, 1997.
- [145] M. M. Meerschaert, P. Roy, and Q. Shao. Parameter estimation for exponentially tempered power law distributions. *Communications in Statistics. Theory and Methods*, (10):1839–1856, 2012.
- [146] T. Mikosch. *Non-life Insurance Mathematics. An Introduction with the Poisson Process*. Springer-Verlag, Berlin, 2009.
- [147] R.M. Mnatsakanov. Hausdorff moment problem: reconstruction of distributions. *Statistics & Probability Letters*, 78(12):1612–1618, 2008.
- [148] M. Morales and W. Schoutens. A risk model driven by Lévy processes. *Applied Stochastic Models in Business and Industry*, 19(2):147–167, 2003.
- [149] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks*. Wiley Series in Probability and Statistics, 2002.
- [150] J. Nair, A. Wierman, and B. Zwart. *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Preprint, California Institute of Technology, 2020.
- [151] J. A. Nelder and R. W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384, 1972.
- [152] E. Ohlsson and B. Johansson. *Non-Life Insurance Pricing with Generalized Linear Models*. Springer, 2010.

- [153] H. Panjer. Recursive evaluation of a family of compound distributions. *ASTIN Bulletin*, 12(1):22–26, 1981.
- [154] P. Parodi. *Pricing in General Insurance*. CRC Press, 2014.
- [155] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes: The art of scientific computing, Third Edition*. Cambridge University Press, 2007.
- [156] C. M. Ramsay. A solution to the ruin problem for Pareto distributions. *Insurance: Mathematics and Economics*, 33(1):109–116, 2003.
- [157] M. Raschke. Alternative modelling and inference methods for claim size distributions. *Annals of Actuarial Science*, pages 1–19, 2020.
- [158] R.-D. Reiss and M. Thomas. *Statistical analysis of extreme values*. Birkhäuser, Basel, 2007.
- [159] S. I. Resnick. Heavy tail modeling and teletraffic data. *The Annals of Statistics*, 25(5):1805–1869, 1997.
- [160] T. Reynkens, R. Verbelen, J. Beirlant, and K. Antonio. Modelling censored losses using splicing: a global fit strategy with mixed Erlang and extreme value distributions. *Insurance: Mathematics and Economics*, 77:65–77, 2017.
- [161] T. Rolski, H. Schmidli, V. Schmidt, and J. L. Teugels. *Stochastic Processes for Insurance and Finance*. John Wiley & Sons, 1999.
- [162] K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999.
- [163] H. Schmidli. *Stochastic Control in Insurance*. Springer Science & Business Media, 2007.
- [164] T. Schmidt. Catastrophe insurance modeled by shot-noise processes. *Risks*, 2(1):3–24, 2014.
- [165] H. Schmitter. The ruin probability of a discrete claims distribution with a finite number of steps. In *XXII ASTIN Colloquium, Montreux, Switzerland*, 1990.
- [166] W. Schoutens. *Lévy Processes in Finance*. Wiley, 2003.
- [167] H. L. Seal. *The Stochastic Theory of a Risk Business*. Wiley, 1969.
- [168] H. L. Seal. *Survival Probabilities*. Wiley, 1978.
- [169] D. Selch and M. Scherer. *A Multivariate Claim Count Model for Applications in Insurance*. Springer International Publishing, 2018.
- [170] M. Shaked and J. G. Shanthikumar. *Stochastic orders and their applications*. Academic Press, New York, 1994.

- [171] M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007.
- [172] E. Sparre Andersen. On the collective theory of risk in case of contagion between claims. *Bulletin of the Institute of Mathematics and its Applications*, 12(2):275–279, 1957.
- [173] D. A. Stanford, F. Avram, A. L. Badescu, L. Breuer, A. Soares Da Silva, and G. Latouche. Phase-type approximations to finite-time ruin probabilities in the Sparre-Andersen and stationary renewal risk models. *ASTIN Bulletin*, 35(1):131–144, 2005.
- [174] A. Steenackers and M. J. Goovaerts. Bounds on stop-loss premiums and ruin probabilities. *Insurance: Mathematics and Economics*, 10(2):153–159, 1991.
- [175] B. Sundt and W. Jewell. Further results on recursive evaluation of compound distributions. *ASTIN Bulletin*, 12(1):27–39, 1981.
- [176] Rolski T. and Stoyan D. On the comparison of waiting times in GI/G/1 queues. *Operations Research*, 24(1):197–200, 1976.
- [177] J. L. Teugels and N. Veraverbeke. Cramér-type estimates for the probability of ruin. *C.O.R.E. Discussion Paper No 7316*, 1973.
- [178] D. Vere-Jones. On the Estimation of Frequency in Point-Process Data. *Journal of Applied Probability*, (Special Vol. 19A):383–394, 1982.
- [179] B. von Bahr. Asymptotic ruin probabilities when exponential moments do not exist. *Scandinavian Actuarial Journal*, 1975(1):6–10, 1975.
- [180] R. von Mises. La distribution de la plus grande de n valeurs. *Rev. math. Union interbalcanique*, 1:141–160, 1936.
- [181] I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, (364):812–815, 1978.
- [182] W. Whitt. Untold horrors of the waiting room: What the equilibrium distribution will never tell about the queue-length process. *Management Science*, 29(4):395–408, 1983.
- [183] R. Winkelmann. *Econometric analysis of count data*. Springer Science & Business Media, 2008.
- [184] E. Yarovaya, J. M. Stoyanov, and K. K. Kostyashin. On conditions for a probability distribution to be uniquely determined by its moments. *Theory of Probability & Its Applications*, 64(4):579–594, 2020.