Year : 2024

OPTIMAL DIVIDEND STRATEGIES AND REINSURANCE FROM AN OPTIMAL TRANSPORT PERSPECTIVE

Garcia Flores Brandon Israel

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

OPTIMAL DIVIDEND STRATEGIES AND REINSURANCE FROM AN OPTIMAL TRANSPORT PERSPECTIVE

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Doctorat en Sciences actuarielles

par

Brandon Israel GARCIA FLORES

Directeur de thèse
Prof. Hansjoerg Albrecher

Jury

Prof. Boris Nikolov, président
Prof. Peter Hieber, expert interne
Prof. Beatrice Acciaio, experte externe
Prof. Hanspeter Schmidli, expert externe

LAUSANNE
2024

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

OPTIMAL DIVIDEND STRATEGIES AND REINSURANCE FROM AN OPTIMAL TRANSPORT PERSPECTIVE

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Doctorat en Sciences actuarielles

par

Brandon Israel GARCIA FLORES

Directeur de thèse
Prof. Hansjoerg Albrecher

Jury

Prof. Boris Nikolov, président
Prof. Peter Hieber, expert interne
Prof. Beatrice Acciaio, experte externe
Prof. Hanspeter Schmidli, expert externe

LAUSANNE
2024

UNIL | Université de Lausanne
Faculté des hautes études commerciales
Internef - 261
CH-1015 Lausanne

IMPRIMATUR

La Faculté des hautes études commerciales de l’Université de Lausanne autorise l’impression de la thèse de doctorat rédigée par

Brandon Israel GARCIA FLORES

intitulée

Optimal Dividend Strategies and Reinsurance From an Optimal Transport Perspective

sans se prononcer sur les opinions exprimées dans cette thèse.

Lausanne, le 29.01.2024

Professeure Marianne Schmid Mast, Doyenne

Thesis committee

Prof. **Hansjörg Albrecher**
Thesis supervisor,
Department of Actuarial Science, University of Lausanne, Switzerland.

Prof. **Boris Nikolov**
President of the jury,
Department of Finance, University of Lausanne, Switzerland.

Prof. **Peter Hieber**
Internal expert,
Department of Actuarial Science, University of Lausanne, Switzerland.

Prof. **Beatrice Acciaio**
External expert,
Department of Mathematics, ETH Zürich, Switzerland.

Prof. **Hanspeter Schmidli**
External expert,
Department of Mathematics, University of Cologne, Germany.

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Brandon Israel GARCIA FLORES

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: 27.11.2023

Prof. Hansjörg 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

Brandon Israel GARCIA FLORES

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: 27.11.2023

Prof. Peter HIEBER
Internal expert

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Brandon Israel GARCIA FLORES

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: 27.11.23

Prof. Beatrice ACCIAIO
External expert

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Brandon Israel GARCIA FLORES

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: 27.11.23

Prof. Hanspeter SCHMIDLI
External expert

Acknowledgements

I would like to start by expressing my deepest gratitude to my advisor, Prof. Hansjörg Albrecher, for his unwavering support and guidance throughout the entire journey of my PhD. His mentorship has not only helped me shape the direction of my research, but of my person as a whole.

I am profoundly thankful to the members of my thesis committee, Prof. Peter Hieber, Prof. Beatrice Acciaio and Prof. Hanspeter Schmidli for their insightful comments and suggestions, which helped me improve the quality of the thesis.

I am also thankful to Prof. Christian Hipp and Prof. Acciaio for their contributions and sharing of knowledge, which were instrumental for the writing of Chapters 4 and 5. A further third expression of my gratitude goes to Prof. Acciaio for her availability and patience when discussing topics on optimal transport, a topic of which I am very fond now, and for her continuous support in the final stage of my PhD.

I am grateful to all the professors, colleagues and ex-colleagues at DSA and UNIL for all the comments, discussions, coffee breaks and games of durak, which in a way or another helped me navigate throughout my studies. Special thanks go to Alaric, Aleksandr, Andrey, Dina, Francesco, José Carlos, José Miguel, Laura, Peter and Sascha.

At last but not at least, I would like to thank my family and friends, for their continuous support, in person or in distance. Thank you to Isaías, Chuck, Carla, Edith, Guillem, Jonas, Martin, Jorge, Oscar and Laura. Special thanks go to my grandparents, mother and siblings, whose love and cheerfulness could be felt thousands of kilometers away. A warm thank-you also goes to Anette, Jesper, Rasmus and Marie for their kind encouragement and for making me felt welcome in their family. Finally, I would like to thank Cecilie for her patience, understanding, support, trust and endless love. This work would not have been possible without her.

Abstract

This thesis presents an in-depth exploration of three topics in non-life insurance: the Greenwood statistic, optimal dividend-payment strategies, and optimal reinsurance contracts.

Chapter 1 introduces some basic concepts and core ideas, establishing a context for the work explored in each of the subsequent chapters. Chapter 2 includes a study on generalizations of the Greenwood statistic, focusing on its asymptotic behaviour under heavy-tailed distributions. Our results generalize those found in, e.g., Albrecher and Teugels [**8**] and Lepage et al. [**85**]. In Chapter 3 the thesis revisits the calculation of optimal dividend band strategies for an insurance portfolio by employing two numerical optimization techniques tailored to the problem, which allow to more efficiently re-derive some previously known results as well as new ones, including an optimal 4-band strategy. In a more stringent scenario, Chapter 4 builds upon these techniques, focusing now on balancing the financial benefits of dividend payments with the long-term solvency of an insurance company measured through the infinite-time-horizon probability of ruin. This is done by considering a particular kind of strategy inspired by band strategies, and the problem is approached for a general spectrally-negative Lévy process. One of our key findings is that these strategies perform outstandingly well, giving sometimes a performance comparable to the one of the unconstrained problem, while in addition respecting restrictive ruin probability constraints. Finally, in Chapter 5, we take the perspective of optimal transport to reconsider the problem of optimal reinsurance. This approach allows to provide alternative proofs of classical optimal reinsurance problems, as well as deriving new solutions that were not achievable before. The chapter concludes by identifying situations under which additional external randomness (like an independent lottery) can increase the efficiency of a reinsurance contract for all involved parties.

Contents

CHAPTER 1

Introduction

Within the scope of financial security, non-life insurance stands as a necessary tool for safeguarding individuals and enterprises against a spectrum of risks that could potentially cause economic instability. This thesis aims at expanding the mathematical toolbox of classical actuarial practice and widening the understanding of a selection of topics in non-life insurance, namely, those related to the Greenwood statistic, optimal dividend-payment strategies, and optimal reinsurance contracts.

The Greenwood statistic, traditionally used in demography and epidemiology for assessing the clustering of events, has found applications in the domain of finance, via the Sharpe ratio, and survival analysis, due to its relevance quantifying the variability of interval estimates. Within insurance, this statistic is of particular relevance in the case when the claims are heavy-tailed, and traditional statistical measures, such as the expectation and variance, might fail to even be defined. Due to its natural connection to the estimation of the coefficient of variation, the Greenwood statistic offers a different perspective on risk assessment and pricing strategies. Our research studies the asymptotic properties of several variants and extensions of this statistic, in particular, of its moments, providing further insights into the limiting distribution of such a measure and revealing some appealing mathematical structures.

Moving from the assessment of risk to strategies for surplus distribution, we approach the area of optimal dividend payment strategies. The study of optimal dividend strategies, as a part of actuarial literature, finds its roots in the seminal work of Bruno de Finetti, who introduced the idea as a way of addressing the deficiencies of the simple models of the time. The typical assessment criterion of the probability of ruin had the unrealistic implication that the capital of an insurance company could grow without bounds, assuming that it does not get ruined first. De Finetti suggested instead to assess the value of an insurance portfolio economically by the present value of aggregate future dividends that the company could pay to its investors, and the question that ensued was establishing which was the best way of doing so. Attempts at answering this question have resulted in fructiferous developments in insurance literature, and while theoretical results have been established, only few explicit examples have been obtained for even some of the simplest models. In this thesis we analyze the use of some numerical optimization techniques and obtain explicit results for the original question, as well as the one in which we also take into account the long-term safety of the insurance company as a side constraint in the optimization process.

Complementing the discussion of dividend policies, the study of optimal reinsurance contracts constitutes another classical problem within the actuarial literature. As part of the insurance industry, reinsurance is a fundamental tool for risk management, capital optimization, and it ensures the stability and sustainability of the insurance market. While the literature in optimal reinsurance is vast, and diverse considerations have been made to address the problem, an intuitive practical constraint when looking for an optimal contract is its deterministic nature, i.e., the requirement that the reinsured amount is fully identified once the claim size of the first line insurance company is known. Inspired by the area of optimal transport, in this thesis we challenge this assumption by allowing exogeneous random mechanisms to have an effect in the determinacy of optimal reinsurance contracts. We show how by making this relaxation, a variety of well-established results within the area can be re-derived, we derive several new optimality results and provide a methodology for the study of general constrained reinsurance problems.

The remainder of this chapter will be used to provide motivation and background relevant to the research areas of this thesis, as well as summarizing some of the key findings of the thesis. The style of the exposition will be rather informal (in mathematical terms) and due to the vast amount of work in these fields, the overview will by no means be exhaustive and will only touch upon several important aspects. Many further references will be given throughout later chapters of the thesis.

1.1. Extreme value analysis

As mentioned before, when studying the Greenwood statistic, we will solely focus on the case where the random variables involved in the analysis are heavy-tailed. A random variable X is heavy-tailed if $\mathbb{E}[e^{tX}] = \infty$ for every $t > 0$ (see e.g. [**99**]). While this is the most common use of the term, our interpretation of this property will be slightly more restricted, namely that the underlying distribution has a regularly varying tail (or equivalently, it is a Pareto-type distribution). We say that a distribution function F has a regularly varying tail if there exists a $\gamma > 0$ such that for any $\omega > 0$, we have

$$\lim_{x \to \infty} \frac{1 - F(\omega x)}{1 - F(x)} = \omega^{-1/\gamma}.$$

The motivation for restricting the analysis in the thesis to Pareto-type distributions comes from extreme value theory (EVT), which stands as the primary theoretical and statistical tool used for dealing with heavy-tailed random variables in risk management (see [**93**]). Following EVT, this is the *only* kind of tail-behaviour that we need to consider when dealing with the asymptotic behaviour of block maxima of heavy-tailed distributions. In this section we will justify this statement and summarize some of the facts and propositions from the area that will be implicitly used in the sequel. The exposition here is mostly based on [**27**] (but see also [**32**] and [**59**] for further references on regular variation and EVT within the context of insurance and finance).

The main concern of EVT is the study and analysis of extreme values. Given an i.i.d. sample of continuous random variables $X_1, \ldots, X_n$, the traditional way of studying extreme values is by assessing the properties of the sample maximum

$$X_{n,n} = \max(X_1, \ldots, X_n).$$

Arguing from a statistical point of view, one would like to handle this analysis without relying on the common distribution of the X_i's, which in applications is often unknown. A natural alternative is to assume that the sample is large enough so that one can use asymptotic properties to make deductions. Similar to the situation in the central limit theorem, one is then faced with the question of finding sequences of real numbers $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ with $b_n > 0$ such that

$$\mathbb{P}\left[\frac{X_{n,n} - a_n}{b_n} \leq x\right] \to G(x)$$

as $n \to \infty$ for every x for which G is continuous. The question now is to identify all the possible (non-degenerate) distributions G which are allowed to appear on the right-hand side of the previous equation, as well as characterizing the distributions of the X_i's for which such G's can be achieved. The answer to the first question is given by the Fisher-Tippett theorem.

THEOREM 1.1.1 (cf. [59, Theorem 3.2.3]). *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. If there exists some non-generate 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 there exists a $\gamma \in \mathbb{R}$ such that G has to be the form*

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

We interpret the case $\gamma = 0$ as $G_0(x) = \exp(e^{-x})$. The quantity γ is called the extreme value index (EVI) and gives an indication on the support of the distribution and the behaviour of the tails of the distribution. In particular, for $\gamma > 0$, the support of the distribution contains the positive half-line and is bounded below by $-1/\gamma$. As this is the main case of interest in the context of heavy tails, we will assume from now on that $\gamma > 0$.
Having answered the first question, we now turn towards the classification problem. Letting F denote the common distribution function of the X_i's, it turns out that the existence of sequences $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ with $b_n > 0$ making the sequence

$$\frac{X_{n,n} - a_n}{b_n}$$

convergent is intimately related to the existence of a positive function a such that the limit

$$\lim_{x\to\infty} \frac{U(xu) - U(x)}{a(x)} \tag{1}$$

exists for every $u > 0$, where U is the tail quantile function of F,

$$U(y) = \inf\{x \mid 1 - F(x) \leq y^{-1}\}.$$

Denoting the limit in (1) by $h(u)$, we have

Proposition 1.1.2 (c.f. [27, Proposition 2.2]). *When G is non-degenerate, the only possible values for h are given by*

$$h_\gamma(u) = \frac{u^\gamma - 1}{\gamma},$$

where $\gamma \in \mathbb{R}$ is the EVI.

Recalling that $\gamma > 0$, it turns out that

$$\lim_{x\to\infty} \frac{U(xu) - U(x)}{a(x)} = \frac{u^\gamma - 1}{\gamma}$$

if and only if U has the form $U(x) = x^\gamma \ell_U(x)$ for some slowly varying function ℓ_U, i.e., a function satisfying

$$\lim_{x\to\infty} \frac{\ell(xu)}{\ell(x)} = 1, \quad u > 0.$$

This property is summarized by saying that U is a regularly varying function. In this case, we can take $a = \gamma U$.
Notice that all these properties are expressed in terms of U and while this function is determined by F, one would like to rephrase the properties in terms of F. However, in this scenario, there is a very simple relationship between the behaviour at infinity of F and U:

$$1 - F(x) = x^{-1/\gamma}\ell_F(x) \text{ if and only if } U(x) = x^\gamma \ell_U(x),$$

where ℓ_F and ℓ_U are slowly varying functions (and ℓ_F is the so called *de Bruijn conjugate* of ℓ_U). In summary, we therefore see that the sequence of (normalized) maxima $X_{1,1}, X_{2,2}, \ldots$ arising from a heavy-tailed distribution F converges (in distribution) to a non-degenerate limit if and only if F is of Pareto-type. Hence, when focusing on the asymptotic behaviour of

block maxima for heavy-tailed distributions, it is enough to only consider Pareto-type distributions.
Using heavy-tailed distributions, Chapter 2 derives results for the asymptotic expectation of the generalized Greenwood statistic. Given a sequence of random variables $X_1, X_2, \ldots$, $\nu, \theta, \eta > 0$ and integers $0 \le r < s$, this statistic is defined as

$$T_{n,s,r}(\nu,\theta,\eta) := \frac{\left(X_{n-r,n}^{\nu} + \cdots + X_{n-s,n}^{\nu}\right)^{\theta}}{(X_{n-r,n} + \cdots + X_{n-s,n})^{\eta}}. \tag{2}$$

The reason for this definition lies in the works of Albrecher and Teugels [**8**] and Lepage et al. [**85**]. In [**8**], it was shown that, when the EVI is larger than 1, $\mathbb{E}[T_{n,n-1,0}(2,k,2k)]$ converges to a polynomial of degree k in $\alpha := 1/\gamma$. The simplicity of this structure motivates us to inquire about the behaviour at infinity of the expectation when we replace the 2 in the exponent by an arbitrary positive number. Similarly, in [**85**] it was proved that for $0 < \alpha < 1$,

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{\max(X_1, \ldots, X_n)}\right] = \frac{\gamma}{\gamma - 1},$$

which motivates to study how many terms in the sum are relevant for the asymptotic behaviour. Notice that for these two results one requires $0 < \alpha < 1$, which corresponds to the case of infinite mean of the X_i's, although in our study we allow α to take any positive value. Notably, we find that, when $\nu > \alpha$, $\theta \in \mathbb{N}$ and $0 < \alpha < 1$, $\lim \mathbb{E}[T_{n,n-1,0}(\nu,\theta,\nu\theta)]$ has a particularly nice structure which reduces to a polynomial when ν is an integer, generalizing the results from [**8**]. For all the other possible combinations ($\nu < \alpha$, $\nu = \alpha$, $\alpha = 1$, etc.) the expectation converges to either zero or infinity and we provide the asymptotic rate of this limit behaviour. Concerning different terms involved in the sum (i.e., $r \neq 0$ or $s \neq n-1$), the previous results show that the case $0 < \alpha < 1$ is the only one of relevance and we obtain an explicit expression for $\lim \mathbb{E}[T_{n,s,r}(\nu,\theta,\nu\theta)]$ in terms of incomplete gamma functions.

1.2. Risk models for the surplus of an insurance company

As mentioned before, the idea of introducing dividend-payment strategies by De Finetti [**54**] has its roots in the deficiencies of the classical models used for modelling the surplus of an insurance company. Hence, before addressing the theory behind optimal dividend strategies, we present the basic models used in the area. In most of the models, the surplus process of a company is considered to change continuously with time. The traditional setting is based on the early works of Filip Lundberg [**92, 91**] and Harald Cramér [**52, 51**] and is henceforth often referred to as the Cramér-Lundberg model. This model is composed of four elements: the initial surplus u, a premium rate p, representing the amount of premium collected by the insurer per unit of time, a sequence of i.i.d. non-negative random variables $(Y_k)_{k\ge0}$ representing the size of the claims and a Poisson process $(N_t)_{t\ge0}$ of rate λ representing the arrival of the claims. With these definitions, the surplus at time t is given by $u + C_t$, where

$$C_t = pt - \sum_{k=1}^{N_t} Y_k, \quad t \ge 0.$$

It is typically assumed that $\mu := \mathbb{E}[Y_k] < \infty$, as otherwise the risks are not insurable. Further, it is expected that, by engaging in business, the insurer makes a profit. Since in this model the only source of income for the insurer is through the premium p, it is often assumed that p satisfies the *safety loading condition*, i.e., that it satisfies the inequality

$$p > \mu\lambda.$$

This simply means that, on average, the insurer will collect more premium than the claims being paid to the policyholders.
While the Cramér-Lundberg model has been the fundamental block of classical risk theory, several extensions have been introduced to encompass a wider set of processes and situations (see [**17, 99**] for surveys in this regard). Observe that, from a mathematical point of view, two of the fundamental properties of the Cramér-Lundberg model are the independence and the stationarity of the increments. The first property refers to the fact that for any two times $0 \leq s \leq t$, $C_t - C_s$ is independent of $\sigma((C_r)_{r\leq s})$, the σ-algebra generated by the process up to time s. The second property means that, for any $s, t \geq 0$, $C_{t+s} - C_s$ has the same distribution as C_t. Considering the right-continuity of the paths, these properties naturally lead to replace the Cramér-Lundberg model by the more general class of processes known as Lévy processes. Briefly speaking, a Lévy process is a process satisfying these three properties, i.e., a *càdlàg* stochastic process starting at zero with stationary and independent increments. The advantage of considering this kind of processes is that it includes several other processes commonly used in modelling, notably the case of the Brownian motion with drift (a diffusion). See e.g. [**57, 58**] or [**96**] for applications of these processes in risk theory or [**83**] and [**17**] for surveys. Two of the fundamental properties of Lévy processes are given by the following:

Theorem 1.2.1 (cf. [**83**, Theorem 1.3]). *A càdlàg stochastic process with independent increments $(C_t)_{t\geq 0}$ is a Lévy process if and only if there exist $a, \sigma \in \mathbb{R}$ and a measure Π concentrated on $\mathbb{R} \setminus \{0\}$ satisfying*

$$\int_{\mathbb{R}} (1 \wedge x^2)\, \Pi(dx) < \infty$$

and such that $\mathbb{E}[e^{i\theta C_t}] = e^{-t\Psi(\theta)}$, where

$$\Psi(\theta) = ia\theta + \frac{\sigma^2\theta^2}{2} + \int_{\mathbb{R}} (1 - e^{i\theta x} + i\theta x 1_{\{|x|<1\}}(x))\, \Pi(dx), \quad \theta \in \mathbb{R}.$$

Moreover, the triple (a, σ^2, Π) is unique.

The measure Π in the previous theorem is called the Lévy (characteristic) measure of the process and the function Ψ its characteristic exponent.

Theorem 1.2.2 (Itô-Lévy decomposition, cf. [**83**, Theorem 2.1]). *Given any $a, \sigma \in \mathbb{R}$ and a measure Π concentrated on $\mathbb{R} \setminus \{0\}$ satisfying*

$$\int_{\mathbb{R}} (1 \wedge x^2)\, \Pi(dx) < \infty,$$

there exists a probability space and three independent Lévy processes $C^{(1)}, C^{(2)}$ and $C^{(3)}$ in that space such that:

- *The process $C = C^{(1)} + C^{(2)} + C^{(3)}$ is a Lévy process with triple (a, σ^2, Π),*
- *$C^{(1)}$ is a Brownian motion with drift, $C_t^{(1)} = -at + \sigma B_t$, $t \geq 0$, where $(B_t)_{t\geq 0}$ is a standard Brownian motion,*
- *$C^{(2)}$ is a compound Poisson process $C_t^{(2)} = \sum_{k=1}^{N_t} Y_k, , t \geq 0$, where $(N_t)_{t\geq 0}$ is a Poisson process with rate $\Pi(\mathbb{R} \setminus (-1, 1))$ and $(Y_k)_{k\geq 0}$ is an i.i.d sequence of random variables with distribution $\Pi/\Pi(\mathbb{R} \setminus (-1, 1))$ and*
- *$C^{(3)}$ is a square integrable martingale with an almost surely countable number of jumps on each finite time interval, all of which are of magnitude less than 1, and with characteristic exponent*

$$\mathbb{E}[e^{i\theta C_t^{(3)}}] = \int_{0<|x|<1} (1 - e^{i\theta x} + i\theta x)\, \Pi(dx).$$

Thus, a Lévy process can be seen as an independent superposition of a Brownian motion, a compound Poisson process and a martingale which captures the information about the small jumps of the process. Within the insurance context, one does not deal with the whole class of Lévy processes, but instead with the smaller class of spectrally negative Lévy processes, which are processes with non-monotone paths for which $\Pi((0,\infty)) = 0$. As seen from the Itô-Lévy decomposition, the last property simply means that the process does not have any upward jumps, reinforcing the idea that the jumps can be interpreted as claims. Hence, the use of this class of processes preserves the idea that the surplus of the insurance company is the aggregate superposition of several independent claims, arriving sequentially through time, offset against a deterministic increasing process, corresponding to the accumulation of premiums, even when there are an almost surely infinite number of claims in any fixed time interval.
In the context of spectrally negative processes, the safety loading condition can be expressed as $\mathbb{E}[C_t] > 0$ for one (and hence, all) $t > 0$. In terms of the triple (a, σ^2, Π), the condition can be written as

$$-a + \int_{(-\infty,-1)} x \, \Pi(dx) > 0.$$

This inequality implies that both $\int_{(-\infty,-1)} x \, \Pi(dx) > -\infty$ and $a < 0$. The finiteness of the integral means that the overall (large) claims have finite expectation and, understanding $-a$ as the premium per time unit, the condition $\mathbb{E}[C_1] > 0$ then means that the premium is larger than the expected value of the large claims, in agreement with the Cramér-Lundberg scenario.
Having introduced the basic risk models used for the surplus of an insurance company, we now move towards pointing out an important fact: observe that for every $n \in \mathbb{N}$, we can write

$$C_n = C_1 + (C_2 - C_1) + \cdots (C_n - C_{n-1})$$

and due to the independence and stationarity of the increments, the sum on the right can be understood as a random walk, allowing the interpretation of Lévy processes as a continuous-time generalization of random walks. Considering that for the latter kind of processes, the Law of Large Numbers holds, we would expect something similar to hold for Lévy processes. This is indeed the case:

Theorem 1.2.3 (cf. [**83**, Theorem 7.2]). *If C is a Lévy process with $0 < \mathbb{E}[C_1] < \infty$, then*

$$\lim_{t\to\infty} \frac{C_t}{t} = \mathbb{E}[C_1] \quad \textit{a.s.}$$

In particular, $\lim_{t\to\infty} C_t = \infty$ *a.s.*

Returning to the setting of the surplus model, the previous theorem has some undesired implication: if we want to enforce the safety loading condition, which is natural in the context of the insurance business, then, in case the insurance company does not get ruined, it will eventually gather an arbitrarily large amount of surplus. The fact that these two extreme scenarios (either ruin or arbitrary richness) are the only two outcomes for the model, was what motivated de Finetti to modify the model by introducing the idea of assessing the value of an insurance portfolio by the present value of aggregate future dividends that the company could pay to its investors. In this new setting, the company pays dividends to shareholders in continuous time, and the objective is to identify the strategy that maximizes the expected sum of discounted dividend payments until the event of ruin. As pointed out in Chapter 3, when there are no further considerations (e.g. the probability of ruin), the innovative contributions of Gerber [**63**] and Azcue & Muler [**22**], as well as Avram et al. [**21**] showed that the optimal strategy is given by a *band strategy*. While this resolves the theoretical aspect of the

problem, the explicit identification of these strategies has been proven challenging, even for the simple case of the Cramér-Lundberg model (see, for instance, [**106**] or [**24**] for references into this specific problem with algorithms developed specifically for the subject). The purpose of Chapter III is therefore to identify the bands in these strategies, and for that matter we use two optimization techniques: an evolutionary strategy (see e.g. [**31**]) and a gradient-inspired method tailored to the problem. These methods allow us to more efficiently obtain some of the previously identified band strategies and to provide an example for which a four band strategy is optimal, thus providing further insight into the overall solution (see also [**28**] for further examples of explicit band strategies).
While these methods allow us to some extent to address the numerical complexities of the original question, a new problem arises: by following this kind of strategies, one disregards the long-term safety of the company, meaning that the infinite-time ruin probability typically becomes 1. Hence, by addressing one of the problems pointed out by de Finetti, one exacerbates the other. Further, ruin probability considerations have been a relevant topic in the area of stochastic control, as the problem exhibits the so-called time-inconsistency property, which prohibits the use of classical tools in the area (see [**10**], [**90**], [**86**], and [**67**] for contributions addressing directly or indirectly the ruin probability constraint). Chapter 4 examines a particular kind of strategy inspired by band strategies that addresses both dividends and a ruin constraint simultaneously, all under the full generality of spectrally-negative Lévy processes. Although this kind of strategies is given exogenously (and hence we cannot conclude any optimality property about it) one of our key findings is that the strategies perform outstandingly well, giving sometimes a performance comparable to the one of the unconstrained problem, while in addition respecting restrictive ruin probability constraints. This is in sharp contrast with some of the previously developed strategies in the literature, where the safety constraint typically decreased the dividend performance considerably (e.g. [**67**], [**109**], [**2**], [**111**]).

1.3. Optimal reinsurance

The topic of optimal reinsurance arises when an insurer has to make a choice among all the possible reinsurance options. This topic can be put into mathematical terms when we assume that the choice is solely based on optimizing an objective function based on some particular constraints. For example, the insurer might want to maximize expected utility after reinsurance and/or achieve a regulatory demand, both subject to a limit in the budget used for the premium. While there are several different ways of phrasing the problem, each corresponding to the particularities of the situation (see, e.g., [**1**] for a survey on the topic), in this thesis we will deal with the following setting: we assume that the first line insurer is interested in choosing the best reinsurance deal for its entire portfolio on a one-year basis contract. The portfolio is represented by a vector $X = (X_1, \dots, X_n)$ of risks, where each X_i represents, for example, the aggregate amounts of different lines of business for the insurer, though more general interpretations can be made. We assume that there is no uncertainty in the distribution of the portfolio and the insurer has full-knowledge of the distribution of X. In this context, a classical reinsurance contract is given by a collection of functions $f_i : \mathbb{R}_+ \to \mathbb{R}_+$, $i = 1, \dots, n$ such that $f_i(x) \leq x$ for every $x \geq 0$ and $i = 1, \dots, n$. These functions can be understood as rules that determine the split between the insurer and reinsurer by means of the identity

$$X_i = f_i(X_i) + (X_i - f_i(X_i)).$$

The requirement $0 \leq f_i(x) \leq x$ is thus enforced to ensure that both $f_i(X_i)$ and $(X_i - f_i(X_i))$ are non-negative, so $f_i(X_i)$ can then be understood as the amount retained by the insurer (or, equivalently, the reinsured amount, observing that the definition is symmetric in this regard).

Reinsurance contracts are compared by means of a *risk measure*, i.e., a function $\mathcal{P} : \mathscr{F} \to \overline{\mathbb{R}}$, where $\mathscr{F}$ is the Cartesian product of n function spaces (commonly the n-fold product of an L^p space for some $p \geq 1$). In a very general manner, the optimal reinsurance problem then can be phrased in the following way: given a subset $\mathcal{S} \subset \mathscr{F}$, find a reinsurance contract $(f_1^*, \ldots, f_n^*)$ such that

$$\mathcal{P}(f_1^*(X_1), \ldots, f_n^*(X_n)) = \inf_{(f_1(X_1), \ldots, f_n(X_n)) \in \mathcal{S}} \mathcal{P}(f_1(X_1), \ldots, f_n(X_n)). \tag{3}$$

The set $\mathcal{S}$ represents the constraints the insurer might have for the problem (e.g., the budget requirement from before). Assuming the risk measure is bounded from below, so that the infimum is finite, one immediately faces the following questions:

Do optimal reinsurance contracts exist?

And if they exist, can one identify them?

Are they unique?

One can easily see the difficulty of these questions by making a first approximation to the problem: in most optimization problems, one of the most basic assumptions is some sort of continuity/compactness requirement in order to ensure that the infimum in (3) is attained. In this scenario, one could then require $\mathcal{S}$ to be compact and $\mathcal{P}$ to be continuous (lower-semicontinuous is sufficient). While this does ensure the existence of a vector $(Y_1, \ldots, Y_n) \in \mathcal{S}$ minimizing $\mathcal{P}$, observe that we require the vector $(Y_1, \ldots, Y_n)$ to be reached from $(X_1, \ldots, X_n)$ by means of a reinsurance contract, i.e., we require $(Y_1, \ldots, Y_n)$ to be in the set

$$\mathcal{S} \cap \{(f_1(X_1), \ldots, f_n(X_n)) \in \mathscr{F} \mid 0 \leq f_i(x) \leq x, i = 1, \ldots, n\}. \tag{4}$$

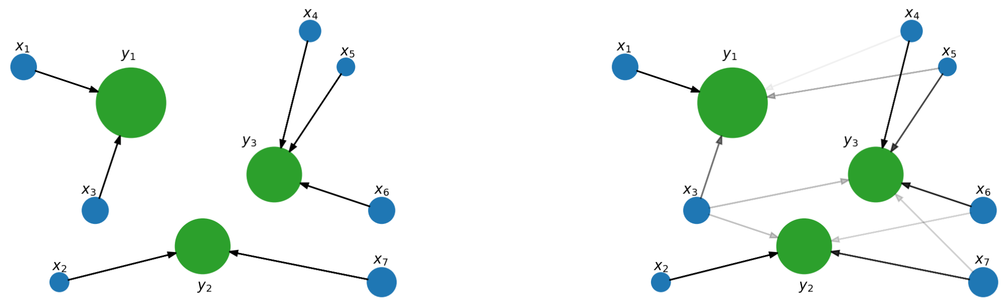
Now, while it is clear that the set on the right of the intersection in (4) is convex, it is not necessarily closed, so even adding the compactness assumption to $\mathcal{S}$ does not seem to help much.

The problems arising from the questions of existence and uniqueness have been the subject of several works in the literature and even for the case $n = 1$, one faces challenges. The usual way in which one deals with this is by either considering simple or no constraints and families of risk measures with particular helpful properties (e.g., convex risk measures, like in [**25, 26**] and [**48**]) thus ensuring existence and uniqueness at once, or by making the constraints so particular that the problem is either reduced in dimensionality or making the set in (4) compact (e.g., requiring that the only contracts under consideration are quota share or stop-loss/excess-of-loss contracts or convex combinations of them, see also [**110**] and [**42**]). By changing the perspective of the problem and trying to study this problem with a rather general setting, the purpose of the last chapter of this thesis will be to study the existence and identification of optimal reinsurance contracts. While the question of uniqueness is relevant in its own right, we will not directly address it, considering that the question of identification is challenging in itself.

The change of perspective mentioned in the previous paragraph is the following: from the point of view of the first line insurer, seeking the best optimal reinsurance contract can be thought of as finding the best way of "moving" some of its own risk to the reinsurer, where there is a "cost" of moving (or not) each part of it. The cost is indicated by $\mathcal{P}$ and while moving the risk, one would like to maintain the constraints specified by $\mathcal{S}$. Further, observe that due to the randomness in the outcome of X, the risk carried by X (in the future) is actually measured (in the present) in terms of its distribution rather than in the realization of X. Calling μ the distribution of portfolio, one could informally state the reinsurance problem as: "find the optimal way of moving the risk (mass) of μ to the reinsurer with cost $\mathcal{P}$ under constraints $\mathcal{S}$".

The idea of optimally moving (or transporting) mass with a predefined cost lies at the core of the area of optimal transport (OT). See [104], [11] or [113] for some standard references in the area. The rest of this section will be used to explain how, by borrowing one of the key ideas from OT, one can easily ensure the existence of optimal reinsurance contracts. The challenge becomes then to identify these contracts.

The key idea from OT is the passage from the Monge formulation of the optimal transportation problem to the Kantorovich formulation. The original problem considered by Monge in [95] is as follows: assume you have a certain amount of sand produced at some mines which needs to be transported to factories where it will be processed for further use. The overall production of sand from the mines matches exactly the sand required by the factories and the transportation is conducted according to a transportation plan, i.e., a plan specifying to which factory should the sand from a specific mine be sent in such a way that the sand requirements are satisfied for all factories and we require the mines to supply exactly one factory. Assuming there are n mines, m factories and a cost $c_{i,j}$ of sending sand from the mine in position x_i to the factory in position y_j, the question is then to find an optimal transportation plan that minimizes the cost. Figure 1A exemplifies the scenario of a transportation plan for the case where there are 7 mines and 3 factories.



(A) Monge formulation. (B) Kantorovich formulation.

FIGURE 1. Transportation plans under the Monge and Kantorovich formulation of OT.

While the problem is relatively easy to formulate, one can just as easily see that, except for a few exceptions, the problem might not even have a solution (consider the case of, say, 2 factories each producing 3 units of sand and 3 factories, each requiring 2 units). The problem arises due to the requirement of each mine supplying at most one factory. It is clear that, by dropping this requirement, one obtains more flexibility and hopefully the existence of optimal transportation plans. By relaxing this assumption one arrives at the Kantorovich formulation of OT, in which a transportation plan specifies instead the proportion of sand produced in x_i that should be sent to y_j. One can then, using standard linear programming tools, prove that optimal transportation plans exist. Figure 1B exemplifies a transportation map for the Kantorovich formulation for the same case as before, in which the darkness of the lines represents the proportion of sand carried over on that route.
Having explained the Monge-Kantorovich formulation of OT, one might wonder how this relates to the (re)insurance context. Notice that we can normalize the sand production and requirements of the mines and factories, thus assuming that the total production and consumption of sand equals 1 unit. By doing so, we are defining two probability measures ν_1

and ν_2 on the sets $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$ respectively. A transportation plan π therefore becomes a probability measure on $\{x_1, \dots, x_n\} \times \{y_1, \dots, y_m\}$ with marginals ν_1 and ν_2 respectively (a "coupling" of ν_1 and ν_2), and one moves mass from ν_1 to ν_2 according to π. Replacing "transportation plans" by "reinsurance contracts", and recalling that, with cost $\mathcal{P}$, we want to move mass from μ to the reinsurer, we can finally, in a slightly informal way, reformulate the reinsurance problem as follows:

Find a reinsurance contract $\eta \in \mathcal{S}$ that minimizes $\mathcal{P}$ and that couples μ with some (unknown) distribution ν.

In Chapter 5, we formalize this statement by means of random reinsurance contracts. Unlike traditional contracts, these do not determine the reinsured amount solely based on the value of the original claim size. Instead, they introduce an additional, extrinsic source of randomness. Our investigations lead to new insights about optimal reinsurance. Notably, we find that under certain conditions – specifically, when $\mathcal{P}$ is lower semi-continuous and S is closed – optimal reinsurance contracts are guaranteed to exist, in sharp contrast to the classical formulation. One of our key discoveries is that, in many cases, the problem of finding the optimal reinsurance contract can be simplified to a finite-dimensional optimization problem with constraints and we develop a method for solving the resulting types of problems. The scenarios we consider are relevant to both practitioners and academics, highlighting the broad applicability of our method. In situations where reducing the problem to a finite dimension is not feasible, we demonstrate that it can often be framed as a constrained OT problem. This allows us to apply techniques from optimal transport to further understand and solve these reinsurance problems. We manage to use this new framework to provide alternative proofs of classical optimal reinsurance problems, provide extended and new solutions that were not achievable before and also identify situations under which additional external randomness (like an independent lottery) can increase the efficiency of a reinsurance contract for all involved parties.

With the established link between optimal transport and optimal reinsurance, we hope that our findings not only open the doors for future research but also challenge the traditional thinking in the field. The realization that, in some scenarios, truly random contracts are optimal is an invitation to reconsider longstanding approaches to address optimal reinsurance problems.

CHAPTER 2

Asymptotic analysis of generalized Greenwood statistics for very heavy tails

This chapter is based on the following article:

H. Albrecher, and B. Garcia Flores. Asymptotic analysis of generalized Greenwood statistics for very heavy tails. *Statistics & Probability Letters* 185 (2022): 109429.

Abstract. We consider some variants of the classical Greenwood statistic and analyze their asymptotic properties for regularly varying random variables with arbitrary index of variation. We also investigate the convergence rate of these asymptotics and study how many terms are asymptotically relevant for the resulting expressions. This naturally generalizes and unifies some earlier results in the literature.

2.1. Introduction

Consider a sequence $(X_i)_{i\geq 1}$ of independent and identically distributed non-negative random variables with cumulative distribution function (cdf) F. The Greenwood statistic T_n is defined as

$$T_n = \frac{X_1^2 + \cdots + X_n^2}{(X_1 + \cdots + X_n)^2}, \tag{5}$$

cf. [**68**]. This quantity appears naturally in various different contexts. For instance, when considering the spacings $D_i = \frac{X_i}{\sum_{i=1}^n X_i}$ with $\sum_{i=1}^n D_i = 1$, then $T_n = \sum_{i=1}^n D_i^2$ is a tool to test for clustering or heterogeneity. Indeed, $1/n \leq T_n \leq 1$, with the value of 1 occurring when all but one D_i are zero (extreme clustering), and $1/n$ occurring if all spacings $D_1 = \ldots = D_n$ are equal (homogeneity), see Arendarczyk et al. [**12**] for a recent contribution on the respective statistical testing procedures. Another application area stems from the relation $n\,T_n = 1 + (\widehat{\mathrm{CoV}(X)})^2$, where $\widehat{\mathrm{CoV}(X)}$ is the sample coefficient of variation of a dataset of n observations (see e.g. Albrecher et al. [**6**], Castillo et al. [**43**]). T_n also appears in financial applications via the Sharpe ratio [**107**], and in the self-normalized sum $T_n^{-1/2}$ and the Student-t statistic $\sqrt{(n-1)/(nT_n - 1)}$, see e.g. Chistyakov and Götze [**49**]. Finally, T_n appears when testing Taylor's law in the biological sciences (also called fluctuation scaling in the physical sciences), which is the empirical observation that often the sample variance is roughly proportional to some power of the sample mean in a set of samples (see e.g. Brown et al. [**40**], Brown and Cohen [**39**], De la Pena et al. [**55**] and Cohen, Davis and Samorodnitsky [**50**]).

Most of the above references deal with an underlying cdf F that is heavy-tailed, and that case will also be the focus of the present paper. In fact, we will consider the case of a regularly varying tail

$$1 - F(x) \sim \ell(x)\, x^{-\alpha}, \quad x \to \infty, \tag{6}$$

with $\alpha > 0$ and $\ell(x)$ denoting any slowly varying function (i.e. $\ell(tx)/\ell(x) \to 1$ as $x \to \infty$ for all $t > 0$). Note that the symbol $\sim$ refers to the ratio of both sides converging to 1 in the limit. In extreme value analysis, F is then also referred to have a Pareto-type tail (see e.g. Beirlant et al. [**27**]). While in our study we allow α to take any positive value, our particular interest is in the case $0 < \alpha < 1$. This is the situation of very heavy tails, for which not only the variance, but even the mean of the marginal random variables X_i does not exist. In that case, beyond the relevance for applications, the study of T_n is interesting in its own right from a mathematical point of view, as the fluctuations in the numerator and the ones in the denominator of (5) tame each other in an intriguing way. In Albrecher and Teugels [**8**] it was shown that $\mathbb{E}[T_n^k]$ tends to a polynomial of degree k in α as $n \to \infty$, the coefficients of which can be explicitly determined. Subsequently, it was shown in Albrecher et al. [**9**] that these coefficients can be read off from the bivariate Taylor expansion of a simple continued fraction, and the coefficients turn out to be identical to the ones appearing in an enumeration problem in topology when counting the possible number of rooted maps on orientable surfaces, without regard to genus, with respect to edges and vertices.

Motivated by the simplicity of the above-mentioned polynomial, in this paper we will study the asymptotic behavior of moments of several variants and extensions of the Greenwood statistic (5), including arbitrary powers in (5) and its reciprocal values. Furthermore, it is well-known from extreme value theory that in the presence of heavy tails, the largest terms dominate the sum in a certain way. For instance, for $0 < \alpha < 1$ one has [**85**]

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{\max(X_1, \ldots, X_n)}\right] = \frac{1}{1-\alpha}.$$

It is hence of interest to see how many terms in the sums in (5) and its extensions are relevant for the established asymptotic behavior, and it turns out that a rather explicit corresponding result can be obtained.

The rest of the paper is organized as follows. Section 2.2 defines the quantities under consideration. In Section 2.3 asymptotic properties of the generalized Greenwood statistic are derived, and a result on the convergence rate to the limit is established for the case of a pure Pareto distribution. Section 2.4 finally addresses the question how the obtained limits change when only some largest order statistics are kept in the statistic. An explicit limit result is obtained and illustrated for the case of the classical Greenwood statistic.

2.2. Definitions and preliminaries

Let $X_1, X_2, \ldots$ be an independent and identically distributed (i.i.d.) sequence of non-negative random variables with regularly varying tail (6) and index $\alpha > 0$. For $\nu, \theta, \eta > 0$, define the generalized Greenwood statistic

$$T_n(\nu,\theta,\eta) = \frac{(X_1^\nu + \cdots + X_n^\nu)^\theta}{(X_1 + \cdots + X_n)^\eta} \tag{7}$$

where it is understood that $T_n = 0$ whenever the the denominator is zero. First of all, the special case

$$T_n(m,1,m) = \frac{X_1^m + \cdots + X_n^m}{(X_1 + \cdots + X_n)^m}, \quad m \in \mathbb{N} \tag{8}$$

is of particular interest, as it directly extends the Greenwood statistic (5) and allows to see to what extent properties of the case $m = 2$ carry over to higher powers. At the same time, the generality of (7) allows to also analyse the multiplicative inverse of T_n: If we define $Y_n = X_n^\nu$ (which entails $1 - F_Y(x) \sim \ell(x)x^{-\alpha/\nu}$), (7) reads

$$T_n(\nu, \theta, \eta) = \frac{(Y_1 + \cdots + Y_n)^\theta}{((Y_1)^{1/\nu} + \cdots + (Y_n)^{1/\nu})^\eta}.$$

For instance,

$$T_n(1/2, 2, 1) = \frac{(Y_1 + \cdots + Y_n)^2}{Y_1^2 + \cdots + Y_n^2}$$

then corresponds to the reciprocal value of (5), if we evaluate it for the new value $\alpha^* = \alpha/\nu$.

2.3. Asymptotic results for the extended Greenwood statistic

In this section we will derive an expression for $\lim_{n\to\infty} \mathbb{E}[T_n(\nu, \theta, \eta)]$ given in (7). In the same spirit as in [7], the behavior changes according to the value of α, and we will distinguish the respective cases. For $x \geq 0$, let μ_x be the quantity given by

$$\mu_x = \int_0^\infty t^x \, dF(t),$$

so that μ_x coincides with the usual moment of order x when x is an integer. Observe that μ_x is finite whenever $x < \alpha$. Moreover, let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and define for positive integers m and p

$$\begin{aligned} S'_m(p) &= \left\{ \boldsymbol{\tau} \in (\mathbb{N}_0)^p \mid \sum_{i=1}^p \tau_i = m \right\}, \\ S_m(p) &= \{ \boldsymbol{\tau} \in S'_m(p) \mid \tau_i > 0, \ i = 1, \ldots, p \}. \end{aligned}$$

Clearly, we require $p \leq m$ for $S_m(p)$ to be non-empty. For $\boldsymbol{\tau} \in S_m(p)$, let $G_{\boldsymbol{\tau}}$, $E_{\boldsymbol{\tau}}$ and $F_{\boldsymbol{\tau}}$ be the sets given by

$$\begin{aligned} E_{\boldsymbol{\tau}} &= \{ i \in \{1, \ldots, p\} \mid \tau_i \nu > \alpha \} \\ F_{\boldsymbol{\tau}} &= \{ i \in \{1, \ldots, p\} \mid \tau_i \nu \leq \alpha, \mu_{\tau_i \nu} < \infty \} \end{aligned}$$

and $G_{\boldsymbol{\tau}} = \{1, \ldots, p\} \setminus (E_{\boldsymbol{\tau}} \cup F_{\boldsymbol{\tau}})$. Observe that $G_{\boldsymbol{\tau}}$ is non-empty only when $\mu_\alpha = \infty$. Let $e_{\boldsymbol{\tau}}$ and $f_{\boldsymbol{\tau}}$ denote the number of elements of $E_{\boldsymbol{\tau}}$ and $F_{\boldsymbol{\tau}}$ respectively, and $\sigma_{\boldsymbol{\tau}}$ and $\varsigma_{\boldsymbol{\tau}}$ the sum of those τ_i's for which $i \in E_{\boldsymbol{\tau}}$ or $i \in F_{\boldsymbol{\tau}}$ respectively. Finally, for an arbitrary real number β and $\boldsymbol{\tau} \in S'_m(p)$, we use the short-hand notation

$$\begin{bmatrix} \beta \\ \boldsymbol{\tau} \end{bmatrix} = \frac{\beta(\beta - 1) \cdots (\beta - m + 1)}{\tau_1! \cdots \tau_p!}.$$

For the discussion that follows, assume $0 < \alpha < 1$ and $\alpha < \nu$. Let $E_1, E_2, \ldots$ be an i.i.d. sequence of exponential random variables and let $\Gamma_n = \sum_{i=1}^n E_i$. Using Lemma 1 of LePage et al. [85] (see also [7, Rem.3]) and the Continuous Mapping Theorem, one sees that

$$T_n(\nu, \theta, \eta) \xrightarrow{d} \frac{\left(\sum_{k=1}^\infty \Gamma_k^{-\nu/\alpha}\right)^\theta}{\left(\sum_{k=1}^\infty \Gamma_k^{-1/\alpha}\right)^\eta}.$$

as $n \to \infty$. Factoring the denominator in the above limit as $(\cdot)^{\eta - \theta\nu}(\cdot)^{\theta\nu}$, one sees, together with Slutsky's Theorem, that the only limit different from 0 or ∞ is obtained for the case

$\eta = \theta\nu$ which is therefore the only case of interest. If furthermore $\nu \geq 1$, then the limit is bounded by 1, and we obtain by dominated convergence that

$$\lim_{n\to\infty} \mathbb{E}[T_n(\nu,\theta,\eta)] = \mathbb{E}\left[\frac{\left(\sum_{k=1}^{\infty}\Gamma_k^{-\nu/\alpha}\right)^{\theta}}{\left(\sum_{k=1}^{\infty}\Gamma_k^{-1/\alpha}\right)^{\eta}}\right]. \tag{9}$$

While one might hope at this point that results along the lines of [**6**, Theorem 2.1] might lead to a path towards evaluating the right-hand side of (9) more explicitly, this does not seem to be practically feasible, so that we follow another approach in what follows (which also allows to consider the case $\alpha < \nu < 1$).

For technical reasons, we restrict our analysis to the case of integer values for θ, which allows us to expand the power in the numerator of (7) into a finite sum of products of the form $X_1^{\tau_1\nu}\cdots X_n^{\tau_n\nu}$. In any case, this is also the most relevant situation in applications.

We begin with a lemma that will be useful in subsequent sections as well.

Lemma 2.3.1. *For $\theta \in \mathbb{N}$ and $n > \theta$, we have*

$$\mathbb{E}[T_n(\nu,\theta,\eta)] = \frac{1}{\Gamma(\eta)}\sum_{p=1}^{\theta}\sum_{\boldsymbol{\tau}\in S_\theta(p)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\binom{n}{p}\int_0^\infty w^{\eta-1}\left(\int_0^\infty e^{-wt}dF(t)\right)^{n-p}\prod_{i=1}^{p}\int_0^\infty x^{\tau_i\nu}e^{-wx}dF(x)dw. \tag{10}$$

Proof. In analogy to the proof for $\nu = 2$ derived in [**8**], we start with the identity

$$\frac{1}{\beta^\eta} = \frac{1}{\Gamma(\eta)}\int_0^\infty w^{\eta-1}e^{-\beta w}dw \tag{11}$$

valid for any $\beta > 0$. Letting $A = \{X_1 + \cdots + X_n > 0\}$, (11) and Tonellis's theorem imply that

$$\mathbb{E}[T_n(\nu,\theta,\eta)1_A] = \frac{1}{\Gamma(\eta)}\int_0^\infty w^{\eta-1}\mathbb{E}\left[\left(\sum_{i=1}^{n}X_i^\nu\right)^{\theta} e^{-w\sum_1^n X_i}1_A\right]dw. \tag{12}$$

where 1_A is the indicator function of A. Now, due to positivity, $A = \{X_1^\nu + \cdots + X_n^\nu > 0\} = \{T_n > 0\}$, so equality in (12) trivially holds when replacing A by its complement (since both sides are zero). Adding up we arrive at

$$\mathbb{E}[T_n(\nu,\theta,\eta)] = \frac{1}{\Gamma(\eta)}\int_0^\infty w^{\eta-1}\mathbb{E}\left[\left(\sum_{i=1}^{n}X_i^\nu\right)^{\theta} e^{-w\sum_1^n X_i}\right]dw. \tag{13}$$

Since according to the notation introduced above

$$\left(\sum_{i=1}^{n}X_i^\nu\right)^{\theta} = \sum_{\boldsymbol{\tau}\in S'_\theta(n)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}X_1^{\tau_1\nu}\cdots X_n^{\tau_n\nu}, \tag{14}$$

we deduce

$$\mathbb{E}[T_n(\nu,\theta,\eta)] = \frac{1}{\Gamma(\eta)}\int_0^\infty w^{\eta-1}\sum_{\boldsymbol{\tau}\in S'_\theta(n)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\prod_{i=1}^{n}\mathbb{E}\left[X_i^{\tau_i\nu}e^{-wX_i}\right]dw. \tag{15}$$

When $n > \theta$, some of the τ_i must be zero. Considering only ordered tuples with non-zero elements, we see that for a fixed $p \leq m$, there will be $\binom{n}{p}$ n-tuples with p non-zero elements.

Hence
(16)
$$\sum_{\boldsymbol{\tau}\in S'_\theta(n)} \begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix} \prod_{i=1}^{n} \mathbb{E}\left[X_i^{\tau_i\nu} e^{-wX_i}\right] = \sum_{p=1}^{\theta} \sum_{\boldsymbol{\tau}\in S_\theta(p)} \begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix} \binom{n}{p} \left(\int_0^\infty e^{-wt} dF(t)\right)^{n-p} \prod_{i=1}^{p} \int_0^\infty x^{\tau_i\nu} e^{-wx} dF(x).$$

Plugging (16) into (15) concludes the proof. □

Let U denote the tail quantile function of F, i.e.,

$$U(s) = F^{-1}(1 - s^{-1}),$$

where F^{-1} is the (generalized) left-continuous inverse function of F. Similarly, let $\widetilde{U}$ be the tail quantile function associated with the distribution $\widetilde{F}$ given by

$$\widetilde{F}(x) = 1 - \alpha x^{-\alpha} \int_0^x t^{\alpha-1}(1 - F(t))dt.$$

Define $\lambda := \alpha^{-1}$. The following sequence of results generalizes Theorems 3.1-3.4 of [**8**].

Proposition 2.3.2. *If $\eta = \theta\nu$ with $\theta \in \mathbb{N}$, then, for $0 < \alpha < 1$, the asymptotic behavior of $\mathbb{E}[T_n(\nu, \theta, \eta)]$ is given by*
(17)
$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{1}{\Gamma(\eta)} \sum_{p=1}^{\theta} \sum_{\boldsymbol{\tau}\in S_\theta(p)} \begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix} \frac{\alpha^{e_{\boldsymbol{\tau}}-1}\Gamma\left((\eta-\sigma_{\boldsymbol{\tau}}\nu)\lambda + g_{\boldsymbol{\tau}}\right)}{p!\Gamma(1-\alpha)^{(\eta-\sigma_{\boldsymbol{\tau}}\nu)\lambda+g_{\boldsymbol{\tau}}}} \pi_1(\boldsymbol{\tau})\pi_2(\boldsymbol{\tau}) \frac{n^{f_{\boldsymbol{\tau}}} \log(n)^{p-e_{\boldsymbol{\tau}}-f_{\boldsymbol{\tau}}}}{U(n)^{\varsigma_{\boldsymbol{\tau}}\nu}},$$

where π_1 and π_2 are given by

$$\pi_1(\boldsymbol{\tau}) = \prod_{i\in E_{\boldsymbol{\tau}}} \Gamma(\tau_i\nu - \alpha) \quad \textit{and} \quad \pi_2(\boldsymbol{\tau}) = \prod_{i\in F_{\boldsymbol{\tau}}} \mu_{\tau_i\nu}.$$

In particular, for $\nu > \alpha$,

$$\lim_{n\to\infty} \mathbb{E}[T_n(\nu,\theta,\eta)] = \frac{1}{\Gamma(\eta)} \sum_{p=1}^{\theta} \sum_{\boldsymbol{\tau}\in S_\theta(p)} \begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix} \frac{\alpha^{p-1}}{p\Gamma(1-\alpha)^p} \prod_{i=1}^{p} \Gamma(\tau_i\nu - \alpha). \tag{18}$$

For $\nu < \alpha$ or $\nu = \alpha$ and $\mu_\alpha < \infty$ we have

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{\mu_\nu^\theta \Gamma(\eta\lambda)}{\alpha\Gamma(\eta)\Gamma(1-\alpha)^{\eta\lambda}} \frac{n^\theta}{U(n)^{\theta\nu}}, \tag{19}$$

and for $\nu = \alpha$ and $\mu_\alpha = \infty$ we have

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{\Gamma(\theta)}{\alpha\Gamma(\theta\alpha)\Gamma(1-\alpha)^\theta} \log(n)^\theta. \tag{20}$$

Proof. Consider the integrals appearing in (10). Integrals of this kind have been studied extensively in [7], so here we only sketch the methodology for working with them: With the change of variable $w = v/U(n)$, we obtain a factor $U(n)^{-\eta}$. But the latter can be written as

$$U(n)^{-\eta} = U(n)^{-\nu(\theta-m)} U(n)^{-\tau_1\nu} \cdots U(n)^{-\tau_p\nu} \tag{21}$$

and distributed accordingly between the factors in the inner integrals of (10). Using the changes of variables $t = U(n/s)$ and $x = U(n/h)$, we have

$$\begin{aligned}
\left(\int_0^\infty e^{-wt} dF(t)\right)^{n-p} &= \left(\frac{1}{n}\int_0^n \exp\left(-v\frac{U(n/s)}{U(n)}\right) ds\right)^{n-p} \\
&\sim \left(\frac{1}{n}\int_0^n \exp\left(-vs^{-\lambda}\right) ds\right)^{n-p} \\
&= \left(\exp\left(-vn^{-\lambda}\right) - \frac{v^\alpha \Gamma(1-\alpha, vn^{-\lambda})}{n}\right)^{n-p} \\
&\sim \exp\left(-v^\alpha \Gamma(1-\alpha)\right).
\end{aligned} \tag{22}$$

Here $\Gamma(s, x)$ denotes the upper incomplete gamma function. Similarly, for $i \in G_{\boldsymbol{\tau}}$, we have for the other integral in (10),

$$\begin{aligned}
\frac{1}{U(n)^{\tau_i\nu}}\int_0^\infty x^{\tau_i\nu} e^{-wx} dF(x) &= \frac{1}{n}\int_0^n \frac{U(n/h)^{\tau_i\nu}}{U(n)^{\tau_i\nu}} \exp\left(-v\frac{U(n/h)}{U(n)}\right) dh \\
&\sim \frac{1}{n}\int_0^n h^{-\lambda\tau_i\nu} \exp\left(-vh^{-\lambda}\right) dh \\
&= \frac{\alpha v^{\alpha-\tau_i\nu}}{n}\,\Gamma(\tau_i\nu - \alpha, vn^{-\lambda}) \\
&\sim \frac{\alpha v^{\alpha-\tau_i\nu}}{n}\Gamma(\tau_i\nu - \alpha),
\end{aligned} \tag{23}$$

for $i \in F_{\boldsymbol{\tau}}$ we have

$$\begin{aligned}
\frac{1}{U(n)^{\tau_i\nu}}\int_0^\infty x^{\tau_i\nu} e^{-wx} dF(x) &= \frac{1}{U(n)^{\tau_i\nu}}\int_0^\infty x^{\tau_i\nu} e^{-vx/U(n)} dF(x) \\
&\sim \frac{\mu_{\tau_i\nu}}{U(n)^{\tau_i\nu}},
\end{aligned} \tag{24}$$

and finally for $i \in G_{\boldsymbol{\tau}}$

$$\begin{aligned}
\frac{1}{U(n)^{\tau_i\nu}}\int_0^\infty x^{\tau_i\nu} e^{-wx} dF(x) &= \frac{1}{U(n)^{\alpha}}\int_0^\infty x^{\alpha} e^{-vx/U(n)} dF(x) \\
&\sim v^{-\alpha}(1 - \widetilde{F}(U(n)/v)) \\
&\sim 1 - \widetilde{F}(U(n)) \\
&\sim \frac{\log(n)}{n}.
\end{aligned} \tag{25}$$

It hence follows from (10), (22), (23), (24) and (25), together with the change of variable $vz^{-\lambda} = u$ that

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \int_0^\infty \sum_{p=1}^{\theta} \sum_{\boldsymbol{\tau}\in S_\theta(p)} \begin{bmatrix}\theta \\ \boldsymbol{\tau}\end{bmatrix} \frac{\alpha^{e_{\boldsymbol{\tau}}} v^{\eta+e_{\boldsymbol{\tau}}\alpha-\sigma_{\boldsymbol{\tau}}\nu-1}}{p!} e^{-v^\alpha\Gamma(1-\alpha)} \pi_1(\boldsymbol{\tau})\pi_2(\boldsymbol{\tau}) \frac{n^{f_{\boldsymbol{\tau}}}\log(n)^{p-e_{\boldsymbol{\tau}}-f_{\boldsymbol{\tau}}}}{U(n)^{\varsigma_{\boldsymbol{\tau}}\nu}} dv.$$

However, we see that the integral over v reduces to

$$\int_0^\infty v^{\eta+e_{\boldsymbol{\tau}}\alpha-\sigma_{\boldsymbol{\tau}}\nu-1} e^{-v^\alpha\Gamma(1-\alpha)} dv = \frac{\lambda\Gamma((\eta-\sigma_{\boldsymbol{\tau}}\nu)\lambda + e_{\boldsymbol{\tau}})}{\Gamma(1-\alpha)^{(\eta-\sigma_{\boldsymbol{\tau}}\nu)\lambda+e_{\boldsymbol{\tau}}}}$$

which proves (17).

If $\alpha < \nu$, then, for every $p \in \{1, \ldots, \theta\}$ and $\boldsymbol{\tau} \in S_\theta(p)$, we have $E_{\boldsymbol{\tau}} = \{1, \ldots, \theta\}$, $e_{\boldsymbol{\tau}} = p$ and $\sigma_{\boldsymbol{\tau}} = \theta$. Hence, we see that $\mathbb{E}[T_n(\nu, \theta, \eta)]$ has constant asymptotic behavior, proving (18).

Similarly, in the case $\nu \le \alpha$, we observe that the sequence $n^{f_{\boldsymbol{\tau}}} U(n)^{-\varsigma_{\boldsymbol{\tau}}\nu} \log(n)^{p-e_{\boldsymbol{\tau}}-f_{\boldsymbol{\tau}}}$ is regularly varying of order $f_{\boldsymbol{\tau}} - \varsigma_{\boldsymbol{\tau}}\nu\lambda$. Since $f_{\boldsymbol{\tau}} \le \varsigma_{\boldsymbol{\tau}}$, we see that $f_{\boldsymbol{\tau}} - \varsigma_{\boldsymbol{\tau}}\nu\lambda \le f_{\boldsymbol{\tau}}(1-\nu\lambda)$, so the order is maximized when there are as many ones as possible in $\boldsymbol{\tau}$. For $\nu < \alpha$ or $\nu = \alpha$ and $\mu_\alpha < \infty$, this happens precisely when $p = \theta$ (and $\boldsymbol{\tau}$ is the only element of $S_\theta(\theta)$). Hence, the dominating term in (17) is the last element in the sum over p, which proves (19). For $\nu = \alpha$ and $\mu_\alpha = \infty$, we observe that $F_{\boldsymbol{\tau}} = \varnothing$, so $f_{\boldsymbol{\tau}} = \varsigma_{\boldsymbol{\tau}} = 0$. Hence, the asymptotic behavior of (18) is determined by the terms $\log(n)^{p-e_{\boldsymbol{\tau}}}$. As the largest exponent is achieved when $p = \theta$, the dominating term is once again the last one in the sum over p, so we obtain (20). $\square$

Remark 2.3.3. Using the property $\Gamma(x+1) = x\Gamma(x)$ of the gamma function, it is evident that when ν is an integer, the limit (18) is always a polynomial in α with rational coefficients. $\square$

We consider now the case $\alpha = 1$. Since for $\nu = 1$, $\mathbb{E}[T_n(\nu, \theta, \theta\nu)] = 1$ for every n and θ, we exclude this case from the next proposition.

Proposition 2.3.4. *If $\eta = \theta\nu$ with $\theta \in \mathbb{N}$, $\alpha = 1$, $\nu \neq 1$, and $\mu_1 = \infty$, then the asymptotic behavior of $\mathbb{E}[T_n(\nu, \theta, \eta)]$ is given by*

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \sum_{p=1}^{\theta} \sum_{\boldsymbol{\tau}\in S_\theta(p)} \begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix} \frac{\Gamma\left(\eta - \sigma_{\boldsymbol{\tau}}\nu + e_{\boldsymbol{\tau}}\right)}{p!\Gamma(\eta)} \pi_1(\boldsymbol{\tau})\pi_2(\boldsymbol{\tau}) \frac{n^{f_{\boldsymbol{\tau}}} U(n)^{e_{\boldsymbol{\tau}}}}{\widetilde{U}(n)^{\varsigma_{\boldsymbol{\tau}}\nu + e_{\boldsymbol{\tau}}}}. \tag{26}$$

Hence, for $\nu > \alpha$,

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{1}{\theta\nu - 1}\frac{U(n)}{\widetilde{U}(n)}, \tag{27}$$

and for $\nu < \alpha$

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \mu_\nu^\theta \frac{n^\theta}{\widetilde{U}(n)^{\theta\nu}}. \tag{28}$$

Proof. Starting again with the integrals in (10), we make the change of variable $w = v/\widetilde{U}(n)$ and distribute the factor $\widetilde{U}(n)^{-\eta}$ thus obtained as in (21). Now, since $\mu_1 = \infty$, we have

$$1 - \int_0^\infty e^{t/x} dF(t) \sim 1 - \widetilde{F}(x).$$

Correspondingly, for the integral in t in (10) we obtain

$$\left(\int_0^\infty e^{-wt} dF(t)\right)^{n-p} \sim e^{-v}. \tag{29}$$

For the integral in x, by making the changes of variable $x = U(n/h)$ and $h = vr^{-1}U(n)/\widetilde{U}(n)$, we have, for $i \in E_{\boldsymbol{\tau}}$,

$$\frac{1}{\widetilde{U}(n)^{\tau_i\nu}} \int_0^\infty x^{\tau_i\nu} e^{-wx} dF(x) \sim \frac{v^{1-\tau_i\nu}}{n}\frac{U(n)}{\widetilde{U}(n)}\Gamma\left(\tau_i\nu - 1\right). \tag{30}$$

Similarly to (24) and (25), we have for $i \in F_{\boldsymbol{\tau}}$,

$$\frac{1}{\widetilde{U}(n)^{\tau_i\nu}} \int_0^\infty x^{\tau_i\nu} e^{-wx} dF(x) \sim \frac{\mu_{\tau_i\nu}}{\widetilde{U}(n)^{\tau_i\nu}}, \tag{31}$$

and for $i \in G_{\boldsymbol{\tau}}$

$$\frac{1}{\widetilde{U}(n)^{\tau_i\nu}}\int_0^\infty x^{\tau_i\nu}e^{-wx}dF(x) \sim \frac{1}{n}. \tag{32}$$

Plugging (29), (30), (31) and (32) into (10), we obtain

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \sum_{p=1}^{\theta}\sum_{\boldsymbol{\tau}\in S_\theta(p)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\frac{\Gamma(\eta+e_{\boldsymbol{\tau}}-\sigma_{\boldsymbol{\tau}}\nu)\pi_1(\boldsymbol{\tau})\pi_2(\boldsymbol{\tau})}{p!\Gamma(\eta)}\frac{n^{f_{\boldsymbol{\tau}}}U(n)^{e_{\boldsymbol{\tau}}}}{\widetilde{U}(n)^{\varsigma_{\boldsymbol{\tau}}\nu+e_{\boldsymbol{\tau}}}},$$

which proves (26).

If $\alpha < \nu$, (26) reduces to

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \sum_{p=1}^{\theta}\sum_{\boldsymbol{\tau}\in S_\theta(p)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\frac{1}{p\Gamma(\eta)}\frac{U(n)^p}{\widetilde{U}(n)^p}\prod_{i=1}^{p}\Gamma(\tau_i\nu-1).$$

However, since $U(n)/\widetilde{U}(n) \to 0$ as $n\to\infty$, the asymptotic behavior of $\mathbb{E}[T_n(\nu,\theta,\eta)]$ is determined by the (eventually) largest term of the sum, i.e., the one with $p=1$, so we obtain

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{\Gamma(\theta\nu-1)}{\Gamma(\eta)}\frac{U(n)}{\widetilde{U}(n)} = \frac{1}{\theta\nu-1}\frac{U(n)}{\widetilde{U}(n)}$$

as desired.

An argument similar to the one given for (19) then proves (28). □

Finally, we consider the case $\alpha \geq 1$:

Proposition 2.3.5. *If $\eta=\theta\nu$ with $\theta\in\mathbb{N}$, $\alpha>1$ or $\alpha=1$ and $\mu_1<\infty$, then the asymptotic behavior of $\mathbb{E}[T_n(\nu,\theta,\eta)]$ is given by*

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{1}{\Gamma(\eta)}\sum_{p=1}^{\theta}\sum_{\boldsymbol{\tau}\in S_\theta(p)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\frac{\alpha^{e_{\boldsymbol{\tau}}}\Gamma(\eta+\alpha e_{\boldsymbol{\tau}}-\sigma_{\boldsymbol{\tau}}\nu)}{p!\mu_1^{\eta+\alpha e_{\boldsymbol{\tau}}-\sigma_{\boldsymbol{\tau}}\nu}}\pi_1(\boldsymbol{\tau})\pi_2(\boldsymbol{\tau})\frac{U(n)^{\alpha e_{\boldsymbol{\tau}}}(1-\widetilde{F}(n))^{p-e_{\boldsymbol{\tau}}-f_{\boldsymbol{\tau}}}}{n^{(\alpha+1)e_{\boldsymbol{\tau}}+\varsigma_{\boldsymbol{\tau}}\nu-p}}. \tag{33}$$

Therefore, for $\nu>\alpha$ or $\nu=\alpha$, $\mu_\alpha=\infty$ and $\theta>1$,

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{\Gamma(\alpha+1)\Gamma(\theta\nu-\alpha)}{\mu_1^\alpha\Gamma(\eta)}\left(\frac{U(n)}{n}\right)^\alpha. \tag{34}$$

For $\nu=\alpha$, $\mu_\alpha=\infty$ and $\theta=1$ we have

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{1}{\mu_1^\eta}\frac{1-\widetilde{F}(n)}{n}, \tag{35}$$

and for $\nu<\alpha$ or $\nu=\alpha$ and $\mu_\alpha<\infty$,

$$\mathbb{E}[T_n(\nu,\theta,\eta)] \sim \frac{\mu_\nu^\theta}{\mu_1^\eta}n^{\theta(1-\nu)}. \tag{36}$$

Proof. Proceeding in the same fashion, we make the change of variable $w=v/n$ and distribute the factor $n^{-\eta}$ similarly to (21). Hence, for the integral in t we obtain

$$\left(\int_0^\infty e^{-wt}dF(t)\right)^{n-p} \sim e^{-\mu v}. \tag{37}$$

For the integral in x, by making the changes of variable $x = U(n/h)$ and $h = vr^{-\lambda}U(n)/n$, we have, for $i \in E_{\boldsymbol{\tau}}$

$$\frac{1}{n^{\tau_i \nu}} \int_0^\infty x^{\tau_i \nu} e^{-wx} dF(x) \sim \frac{\alpha v^{\alpha - \tau_i \nu} U(n)^\alpha}{n^{\alpha+1}} \Gamma(\tau_i \nu - \alpha). \tag{38}$$

The behavior for $i \in F_{\boldsymbol{\tau}}$ and $i \in G_{\boldsymbol{\tau}}$ is still similar to (24) and (25), so for $i \in F_{\boldsymbol{\tau}}$, we obtain

$$\frac{1}{n^{\tau_i \nu}} \int_0^\infty x^{\tau_i \nu} e^{-wx} dF(x) \sim \frac{\mu_{\tau_i \nu}}{n^{\tau_i \nu}}, \tag{39}$$

and for $i \in G_{\boldsymbol{\tau}}$

$$\frac{1}{n^{\tau_i \nu}} \int_0^\infty x^{\tau_i \nu} e^{-wx} dF(x) \sim 1 - \widetilde{F}(n). \tag{40}$$

Therefore,

$$\mathbb{E}[T_n(\nu, \theta, \eta)] \sim \frac{1}{\Gamma(\eta)} \sum_{p=1}^{\theta} \sum_{\boldsymbol{\tau} \in S_\theta(p)} \begin{bmatrix} \theta \\ \boldsymbol{\tau} \end{bmatrix} \frac{\alpha^{e_{\boldsymbol{\tau}}} \Gamma(\eta + \alpha e_{\boldsymbol{\tau}} - \sigma_{\boldsymbol{\tau}} \nu)}{p! \mu_1^{\eta + \alpha e_{\boldsymbol{\tau}} - \sigma_{\boldsymbol{\tau}} \nu}} \pi_1(\boldsymbol{\tau}) \pi_2(\boldsymbol{\tau}) \frac{U(n)^{\alpha e_{\boldsymbol{\tau}}} (1 - \widetilde{F}(n))^{p - e_{\boldsymbol{\tau}} - f_{\boldsymbol{\tau}}}}{n^{(\alpha+1)e_{\boldsymbol{\tau}} + \varsigma_{\boldsymbol{\tau}} \nu - p}}$$

as desired. Now, if $\nu > \alpha$, this expression reduces to

$$\mathbb{E}[T_n(\nu, \theta, \eta)] \sim \frac{1}{\Gamma(\eta)} \sum_{p=1}^{\theta} \frac{\alpha^p \Gamma(p\alpha)}{p! \mu^{p\alpha}} \left(\frac{U(n)}{n} \right)^{p\alpha} \sum_{\boldsymbol{\tau} \in S_\theta(p)} \begin{bmatrix} \theta \\ \boldsymbol{\tau} \end{bmatrix} \prod_{i=1}^{p} \Gamma(\tau_i \nu - \alpha).$$

Since $U(n)$ is regularly varying with index $1/\alpha < 1$, we have $U(n)/n \to 0$ as $n \to \infty$, so the asymptotic behavior is dominated by the term with $p = 1$, i.e.,

$$\mathbb{E}[T_n(\nu, \theta, \eta)] \sim \frac{\Gamma(\alpha + 1)\Gamma(\theta\nu - \alpha)}{\mu^\alpha \Gamma(\eta)} \left(\frac{U(n)}{n} \right)^\alpha$$

which proves (34).

For $\nu \leq \alpha$, we observe that the sequence

$$\frac{U(n)^{\alpha e_{\boldsymbol{\tau}}} (1 - \widetilde{F}(n))^{p - e_{\boldsymbol{\tau}} - f_{\boldsymbol{\tau}}}}{n^{(\alpha+1)e_{\boldsymbol{\tau}} + \varsigma \nu - p}}$$

is regularly varying with index $\alpha f_{\boldsymbol{\tau}} - \varsigma_{\boldsymbol{\tau}} \nu - p(\alpha - 1)$. Since $f_{\boldsymbol{\tau}} \leq p$ and $f_{\boldsymbol{\tau}} \leq \varsigma_{\boldsymbol{\tau}}$, the order is bounded by $f_{\boldsymbol{\tau}}(1 - \nu)$ and when $\nu < \alpha$ or $\nu = \alpha$ and $\mu_\alpha < \infty$, this bound is reached for $p = \theta$, obtaining (36). If otherwise, $\nu = \alpha$ and $\mu_\nu = \infty$, then $f_{\boldsymbol{\tau}} = \varsigma_{\boldsymbol{\tau}} = 0$, so the term involving n in (34) is given by

$$\frac{U(n)^{\alpha e_{\boldsymbol{\tau}}} (1 - \widetilde{F}(n))^{p - e_{\boldsymbol{\tau}}}}{n^{(\alpha+1)e_{\boldsymbol{\tau}} - p}}.$$

This term defines a regularly varying sequence of order $p(1 - \alpha)$ and since in this case $\alpha > 1$, the dominating term happens when $p = 1$, obtaining (34) or (35) depending on the value of θ. ☐

Remark 2.3.6. Observe that (36) provides a correction for the case $\theta + 1 > \alpha$ considered in Theorem 3.4 of [**8**]. ☐

We can now work out particular explicit cases of Formula (18) for $0 < \alpha < 1$. To begin with, for $\theta = 1$ (18) simplifies to

$$\lim_{n \to \infty} \mathbb{E} \left[\frac{X_1^\nu + \cdots + X_n^\nu}{(X_1 + \cdots + X_n)^\nu} \right] = \frac{\Gamma(\nu - \alpha)}{\Gamma(\nu)\Gamma(1 - \alpha)}. \tag{41}$$

If furthermore $\nu = m \in \mathbb{N}$ is an integer, one receives for any m the pleasant general formula

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\left[\frac{X_1^m + \cdots + X_n^m}{(X_1+\cdots+X_n)^m}\right] &= \lim_{n\to\infty} \mathbb{E}\left[T_n(m,1,m)\right] = \frac{1}{(m-1)!}\prod_{i=1}^{m-1}(i-\alpha) \\ &= (1-\alpha)\left(1-\frac{\alpha}{2}\right)\cdots\left(1-\frac{\alpha}{m-1}\right),\end{aligned} \tag{42}$$

extending the classical result $\lim_{n\to\infty} \mathbb{E}\left[T_n(2,1,2)\right] = 1-\alpha$ in a natural way. Figure 1 depicts (42) as a function of α ($0 < \alpha < 1$) for $m = 2, 3, 4, 5$ (solid lines). For $m = 4$, it is also of interest to compare the resulting expression to

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{(X_1^2+\cdots+X_n^2)^2}{(X_1+\cdots+X_n)^4}\right] = \lim_{n\to\infty} \mathbb{E}\left[T_n(2,2,4)\right] = (1-\alpha)(1-2\,\alpha/3),$$

which was already obtained in **[8]** and is depicted as the dashed line in Figure 1. Indeed, the latter contains all the additional mixed positive terms in the numerator, and we have $(1-\alpha)(1-2\,\alpha/3) > (1-\alpha)(1-\alpha/2)(1-\alpha/3)$ for all $0 < \alpha < 1$.

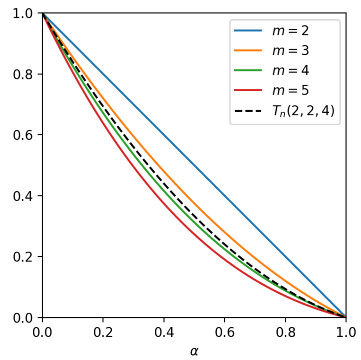


FIGURE 1. $\lim_{n\to\infty} \mathbb{E}\left[T_n(m,1,m)\right]$ as found in (42) for $m = 2, 3, 4, 5$ (solid lines) and $\lim_{n\to\infty} \mathbb{E}\left[T_n(2,2,4)\right]$ (dashed line).

For the limiting behavior of the second and third moment of (8) we for instance get the explicit formulas

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}\left[T_n^2(m,1,m)\right] &= \lim_{n\to\infty} \mathbb{E}\left[T_n(m,2,2m)\right] \\ &= \frac{1}{(2m-1)!}\prod_{i=1}^{m-1}(i-\alpha)\left(\prod_{i=1}^{m}(i+m-1-\alpha) + \alpha\prod_{i=1}^{m-1}(i-\alpha)\right) \\ \lim_{n\to\infty} \mathbb{E}\left[T_n^3(m,1,m)\right] &= \lim_{n\to\infty} \mathbb{E}\left[T_n(m,3,3m)\right] \\ &= \frac{1}{(3m-1)!}\prod_{i=1}^{m-1}(i-\alpha)\left(\prod_{i=1}^{2m}(i+m-1-\alpha)\right. \\ &\qquad \left. +3\alpha\prod_{i=1}^{2m-1}(i-\alpha) + 2\alpha^2\prod_{i=1}^{m-1}(i-\alpha)^2\right).\end{aligned}$$

While we know from Remark 2.3.3 that the limit of the kth moment of $T_n(m,1,m)$ is a polynomial in α, in analogy to Theorem 3.1 in **[8]** one might furthermore conjecture that its degree is $k(m-1)$ for each $m \in \mathbb{N}$. Indeed,

$$\prod_{i=1}^{m}(i+m-1-\alpha) = q_1(\alpha) + (-1)^{m-1}\sum_{i=1}^{m}(i+m-1)\alpha^{m-1} + (-1)^m\alpha^m$$

and

$$\alpha\prod_{i=1}^{m-1}(i-\alpha) = q_2(\alpha) + (-1)^{m-2}\sum_{i=1}^{m-1} i\alpha^{m-1} + (-1)^{m-1}\alpha^m$$

where q_1 and q_2 are polynomials of degree $m-2$. Therefore, we see that $\prod_{i=1}^{m}(i+m-1-\alpha)+\alpha\prod_{i=1}^{m-1}(i-\alpha)$ is a polynomial of degree $m-1$ with leading coefficient $(-1)^{m-1}m^2$, so we can conclude that $\lim_{n\to\infty}\mathbb{E}\left[T_n(m,2,2m)\right]$ is a polynomial of degree $2(m-1)$. Similar computations also establish that $\lim_{n\to\infty}\mathbb{E}\left[T_n(\alpha,m,3,3m)\right]$ is polynomial of degree $3(m-1)$. However, to prove that conjecture for arbitrary k seems to be difficult.

Choosing $\alpha^* = \alpha/m$, $\nu = 1/m$, $\theta = k/m$ and $\eta = k$, we deduce from (18) for the reciprocal moments

(43)
$$\lim_{n\to\infty}\mathbb{E}\left[\frac{(X_1+\cdots+X_n)^{mk}}{(X_1^m+\cdots+X_n^m)^k}\right] = \frac{1}{\Gamma(k)}\sum_{p=1}^{mk}\sum_{\boldsymbol{\tau}\in S_{mk}(p)}\begin{bmatrix} mk \\ \boldsymbol{\tau}\end{bmatrix}\frac{\alpha^{p-1}}{m^{p-1}p\Gamma(1-\alpha/m)^p}\prod_{i=1}^{p}\Gamma\left(\frac{\tau_i-\alpha}{m}\right).$$

For $m=2$ and $k=1$, this leads to

$$\lim_{n\to\infty}\mathbb{E}\left[\frac{(X_1+\cdots+X_n)^2}{X_1^2+\cdots+X_n^2}\right] = 1+\frac{\alpha}{2}\left(\frac{\Gamma(\frac{1}{2}-\alpha/2)}{\Gamma(1-\alpha/2)}\right)^2 = 1+\frac{\alpha B(1/2,(1-\alpha)/2)^2}{2\pi},$$

where $B(x,y)=\Gamma(x)\Gamma(y)/\Gamma(x+y)$ denotes the Beta function. Since $B(1/2,(1-\alpha)/2)^2$ is not expressible in terms of elementary functions, we can in general not expect (43) to be expressible in a simpler way.

It is of interest to assess the speed of convergence in Proposition 2.3.2, in particular when using Greenwood statistics for statistical estimation procedures (cf. Castillo et al. [**43**]). In fact, for the pure Pareto case (i.e. with the slowly varying function $\ell(x)\equiv 1$), a quantitative result can be obtained:

Proposition 2.3.7. *If*

$$F(x) = 1-x^{-\alpha},\ x>1, \tag{44}$$

and $0<\alpha<1<\nu$, *then*

$$\left|\mathbb{E}\left[T_n(\nu,1,\nu)\right]-\frac{\Gamma(\nu-\alpha)}{\Gamma(\nu)\Gamma(1-\alpha)}\right| \le \frac{C}{n^{1-\alpha}} + D\,\frac{\Gamma(\nu-1,n)}{n^{\nu-1}}$$

for all n *and some constants* C *and* D.

Proof. Observe that under (44) we have $U(s)=s^\lambda$, which implies that equality holds in the second line of equations (22), (23) and (25). Hence, after the change of variable $v=n^\lambda u$, (15) becomes

$$\mathbb{E}[T_n(\nu,\theta,\nu)] = \frac{1}{\Gamma(\nu\theta)}\sum_{p=1}^{\theta}\sum_{\boldsymbol{\tau}\in S_\theta(p)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\binom{n}{p}\alpha^p\cdot \tag{45}$$
$$\cdot\int_0^\infty u^{p\alpha-1}\left(\exp(-u)-u^\alpha\Gamma(1-\alpha,u)\right)^{n-p}\prod_{i=1}^{p}\Gamma(\tau_i\nu-\alpha,u)du.$$

As $\theta = 1$, this reduces to

$$\mathbb{E}[T_n(\nu, 1, \nu)] = \frac{n\alpha}{\Gamma(\nu)} \int_0^\infty u^{\alpha-1} \left(\exp(-u) - u^\alpha \Gamma(1-\alpha, u)\right)^{n-1} \Gamma(\nu - \alpha, u) du.$$

By observing that the derivative of the mapping $u \mapsto \exp(-u) - u^\alpha \Gamma(1-\alpha, u)$ is given by $u \mapsto -\alpha u^{\alpha-1}\Gamma(1-\alpha, u)$, we can integrate by parts to obtain

$$\mathbb{E}[T_n] = \frac{\Gamma(\nu-\alpha)}{\Gamma(\nu)\Gamma(1-\alpha)} + \frac{1}{\Gamma(\nu)} \int_0^\infty \left(\exp(-u) - u^\alpha \Gamma(1-\alpha, u)\right)^n H'(u) du, \tag{46}$$

where $H(u) = \Gamma(\nu - \alpha, u)/\Gamma(1-\alpha, u)$. In view of (41) we can therefore focus on the approximation error

$$\xi_n := \frac{1}{\Gamma(\nu)} \int_0^\infty \big(\exp(-u) - u^\alpha \Gamma(1-\alpha, u)\big)^n H'(u) du$$

for any finite n. While this expression cannot be evaluated exactly, we can obtain bounds for the convergence rate of ξ_n to zero.

We need to derive some facts about H and its derivative. Observe that H' is given by

$$H'(u) = \frac{u^{-\alpha}e^{-u}\Gamma(\nu-\alpha, u) - u^{\nu-\alpha-1}e^{-u}\Gamma(1-\alpha, u)}{\Gamma(1-\alpha, u)^2} = \frac{u^{-\alpha}e^{-u}}{\Gamma(1-\alpha, u)} \left(\frac{\Gamma(\nu-\alpha, u)}{\Gamma(1-\alpha, u)} - u^{\nu-1}\right).$$

By means of, for example, de l'Hôpital's rule we can show that

$$\lim_{u\to\infty} \frac{u^{-\alpha}e^{-u}}{\Gamma(1-\alpha, u)} = 1, \quad \lim_{u\to\infty} \frac{\Gamma(\nu-\alpha, u)}{u^{\nu-1}\Gamma(1-\alpha, u)} = 1 \quad \text{and} \quad \lim_{u\to\infty} u\left(\frac{\Gamma(\nu-\alpha, u)}{u^{\nu-1}\Gamma(1-\alpha, u)} - 1\right) = \nu - 1. \tag{47}$$

The first two limits, together with the fact that $\nu \geq 1$, imply that the behavior of H close to zero is dominated by the factor $u^{-\alpha}$, whereas the behavior at infinity is dominated by $H(u) - u^{\nu-1}$. Moreover, using once again the fact that $\nu \geq 1$, we notice that

$$u^{\nu-1} \int_u^\infty t^{-\alpha} e^{-t} dt \leq \int_u^\infty t^{\nu-\alpha-1} e^{-t}\, dt,$$

which implies $H(u) \geq u^{\nu-1}$ for every $u \geq 0$. Thus, H' is positive and H is increasing. Splitting the integral defining ξ_n, we get

$$\mid \xi_n \mid \leq C_1 \int_0^1 u^{-\alpha} \exp(-nu) du + C_2 \int_1^\infty \exp(-nu)(H(u) - u^{\nu-1}) du \tag{48}$$

where

$$C_1 = \frac{\Gamma(\nu-\alpha, 1)}{\Gamma(\nu)\Gamma(1-\alpha, 1)^2} \quad \text{and} \quad C_2 = \frac{e^{-1}}{\Gamma(\nu)\Gamma(1-\alpha, 1)}.$$

Now, the change of variable $u = nv$ shows that

$$\int_0^1 u^{-\alpha} \exp(-nu) du \leq \frac{\Gamma(1-\alpha)}{n^{1-\alpha}}.$$

Finally, the third limit in (47) shows that there exists a constant C_3 such that $H(u) - u^{\nu-1} \leq C_3 u^{\nu-2}$ for every $u \geq 1$. Hence,

$$\int_1^\infty \exp(-nu)(H(u) - u^{\nu-1})\, du \leq C_3 \int_1^\infty u^{\nu-2} \exp(-nu)\, du = \frac{C_3\Gamma(\nu-1, n)}{n^{\nu-1}}$$

which finishes the proof by taking $C = C_1$ and $D = C_2C_3$. □

Remark 2.3.8. For arbitrary $\theta > 1$, a similar procedure can lead to an explicit expression for the approximation error ξ_n as well. In that case, however, one needs to integrate by parts p times the integral in (45) in order to get rid of the factorial coming from $\binom{n}{p}$ and in order to make the factor $\Gamma(1-\alpha)^p$ appear in the denominator. Considering the product inside the integral, this procedure makes the overall integrand quite hard to handle. □

2.4. How many terms are asymptotically relevant?

In view of the subexponentiality of Pareto-type distributions, and in the spirit of [7], it is natural to inquire to what extent the largest terms in the sums in the numerator and denominator in (7) dominate. Concretely, do we get to a similar limiting behavior already when only taking some largest terms in both?

To address this question, consider the order statistics $X_{1,n} \leq \ldots \leq X_{n,n}$ of the sample $X_1, \ldots, X_n$ and, for a non-decreasing function $r : \mathbb{N} \to \mathbb{N}$ with $r(n) \leq n$, define

$$T_{n,n-r(n)} = \frac{\left(X_{n,n}^{\nu} + \cdots + X_{r(n),n}^{\nu}\right)^{\theta}}{(X_{n,n} + \cdots + X_{r(n),n})^{\eta}}. \tag{49}$$

For the following, we will consider only the case $0 < \alpha < 1$ and $\nu > \alpha$, which is the only case for which the asymptotic behavior of $\mathbb{E}[T_n(\nu, \theta, \theta\nu)]$ is not zero or infinity.

For $q > 0$, let ℓ^q denote the space of q-summable sequences, regarded as a subspace of $\mathbb{R}^{\mathbb{N}}$ with the product topology. Define

$$C = \{(z_1, z_2, \ldots) \in \mathbb{R}^{\mathbb{N}} \mid z_1 > 0, z_1 \geq z_2 \geq \cdots \geq 0\},$$
$$D = C \cap \bigcup_{q>\alpha} \ell^q.$$

We have the following.

Proposition 2.4.1. *Let $E_1, E_2, \ldots$ be an i.i.d. sequence of exponential random variables and let $\Gamma_n = \sum_{i=1}^n E_i$. If $\lim_{n\to\infty} n - r(n) = s$, with $s \in \mathbb{N} \cup \{\infty\}$, then the sequence*

$$Z_n = (X_{n,n}, \ldots, X_{r(n),n}, 0, 0, \ldots)/U(n)$$

converges in distribution to

$$Z = (\Gamma_1^{-\lambda}, \Gamma_2^{-\lambda}, \ldots, \Gamma_{s+1}^{-\lambda}, 0, \ldots)$$

regarded as random elements in D. Here $\lambda = \alpha^{-1}$ and there are no zeros at the end of the vector defining Z if s is infinite.

Proof. This result is an easy extension of Lemma 1 in [85], also being based on the convergence of the finite-dimensional distributions. The main difference apart from the change in the number of terms defining the vectors is the definition of the state space of the Z_n and Z here. Observe first that since

$$\bigcup_{q>\alpha} \ell^q = \bigcup_{\substack{q>\alpha,\\ q\in\mathbb{Q}}} \ell^q$$

and ℓ^q is Borel in the product topology for each $q > 0$, then D is also a Borel subset of $\mathbb{R}^{\mathbb{N}}$. As $Z_n \in D$ and $Z \in D$ a.e., we can restrict the state space from C to D without affecting the convergence in distribution □

The relevance of this proposition is the following: Extending the respective remarks of Section 2.2, if $\eta = \theta\nu$, the function $f : D \to \mathbb{R}$ given by

$$f(z_1, z_2, \dots) = \frac{\left(\sum_{k=1}^{\infty} z_k^{\nu}\right)^{\theta}}{\left(\sum_{k=1}^{\infty} z_k\right)^{\eta}}$$

is continuous, so by the Continuous Mapping Theorem we have

$$T_n = f(Z_n) \xrightarrow{d} f(Z) = \frac{\left(\sum_{k=1}^{s+1} \Gamma_k^{-\lambda\nu}\right)^{\theta}}{\left(\sum_{k=1}^{s+1} \Gamma_k^{-\lambda}\right)^{\eta}}.$$

Moreover, for $\nu \geq 1$, f is bounded by 1, so

$$\lim_{n\to\infty} \mathbb{E}[T_n] = \mathbb{E}\left[\frac{\left(\sum_{k=1}^{s+1} \Gamma_k^{-\lambda\nu}\right)^{\theta}}{\left(\sum_{k=1}^{s+1} \Gamma_k^{-\lambda}\right)^{\eta}}\right].$$

If $\alpha < \nu < 1$, we can define a sequence of functions $f_m : D \to \mathbb{R}$ given by

$$f_m(z_1, z_2, \dots) = \frac{\left(\sum_{k=1}^{m} z_k^{\nu}\right)^{\theta}}{\left(\sum_{k=1}^{m} z_k\right)^{\eta}}.$$

We notice that f_m is bounded by $m^{\theta(1-\nu)}$ and $f_m \to f$ pointwise in D. Since in this case $f_m(Z) \leq f_{m+1}(Z)$, monotone convergence applies and at least we can conclude

$$\lim_{m\to\infty}\lim_{n\to\infty} \mathbb{E}[f_m(Z_n)] = \lim_{m\to\infty} \mathbb{E}\left[\frac{\left(\sum_{k=1}^{\min(m,s+1)} \Gamma_k^{-\lambda\nu}\right)^{\theta}}{\left(\sum_{k=1}^{\min(m,s+1)} \Gamma_k^{-\lambda}\right)^{\eta}}\right] = \mathbb{E}\left[\frac{\left(\sum_{k=1}^{s+1} \Gamma_k^{-\lambda\nu}\right)^{\theta}}{\left(\sum_{k=1}^{s+1} \Gamma_k^{-\lambda}\right)^{\eta}}\right].$$

As in Section 2.2, we conclude that only the case $\eta = \theta\nu$ is of interest, as otherwise Slutsky's Theorem together with the fact that $U(n) \to \infty$ implies that T_n converges in distribution to zero or infinity. We also observe that only the case of finite s will lead to different results than $r(n) = 1$ (i.e. when keeping all the terms). Aiming for a little bit more generality, we will therefore focus on the quantity

$$T_{n,s,r}(\nu, \theta, \eta) := \frac{\left(X_{n-r,n}^{\nu} + \cdots + X_{n-s,n}^{\nu}\right)^{\theta}}{\left(X_{n-r,n} + \cdots + X_{n-s,n}\right)^{\eta}}.$$

with $0 \leq r < s$.

Proposition 2.4.2. *The limit of the expectations* $\mathbb{E}[T_{n,s,r}]$ *is given by*

$$\lim_{n\to\infty} \mathbb{E}[T_{n,s,r}(\nu,\theta,\eta)] = \frac{s!}{\Gamma(\eta) r! (s-r-1)!} \sum_{m=0}^{\theta} \int_0^{\infty} u^{\nu(\theta-m)-\alpha-1} e^{-u} \Theta_{m,s,r}(u) du, \tag{50}$$

where

$$\Theta_{m,s,r}(u) = \alpha^{s-r} u^{(s+1)\alpha} \sum_{p=0}^{m} \binom{\theta}{p} \sum_{\boldsymbol{\tau} \in S'_{m-p}(s-r-1)} \begin{bmatrix} \theta - p \\ \boldsymbol{\tau} \end{bmatrix} \int_u^{\infty} v^{p\nu-(r+1)\alpha-1} e^{-v} \prod_{i=1}^{s-r-1} \int_u^{v} z^{\tau_i\nu-\alpha-1} e^{-z} dz dv.$$

Proof. We only need to adapt the proof of Proposition 2.3.2. Using the same identity as in (13), we write the corresponding expectation as

(51)
$$\mathbb{E}\left[\left(X_{n-r,n}^{\nu}+\cdots+X_{n-s,n}^{\nu}\right)^{\theta}e^{-w\sum_{n-s}^{n-r}X_{i,n}}\right]$$
$$=\sum_{m=0}^{\theta}\sum_{p=0}^{m}\binom{\theta}{p}\sum_{\boldsymbol{\tau}\in S'_{m-p}(s-r-1)}\begin{bmatrix}\theta-p\\ \boldsymbol{\tau}\end{bmatrix}\mathbb{E}\left[X_{n-s,n}^{\nu(\theta-m)}X_{n-r,n}^{p\nu}X_{n-s+1,n}^{\tau_1\nu}\cdots X_{n-r-1,n}^{\tau_{s-r-1}\nu}e^{-w\sum_{n-s}^{n-r}X_{i,n}}\right],$$

Conditioning first on $X_{n-s,n}$ and then on $X_{n-r,n}$, we can rewrite the innermost sum as

(52)
$$\sum_{\boldsymbol{\tau}\in S'_{m-p}(s-r-1)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\mathbb{E}\left[X_{n-s,n}^{\nu(\theta-m)}X_{n-r,n}^{p\nu}X_{n-s+1,n}^{\tau_1\nu}\cdots X_{n-r-1,n}^{\tau_{s-r-1}\nu}e^{-w\sum_{n-s}^{n-r}X_{i,n}}\right]=$$
$$\begin{bmatrix}n\\ \sigma\end{bmatrix}\sum_{\boldsymbol{\tau}\in S'_{m-p}(s-r-1)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\int_0^{\infty}y^{\nu(\theta-m)}e^{-wy}F^{n-s-1}(y)I(y)dF(y),$$

where $\sigma=(1,1,n-s-1,r,s-r-1)$ and
$$I(y)=\int_y^{\infty}x^{p\nu}e^{-wx}\bar{F}^r(x)\prod_{i=1}^{s-r-1}\left(\int_y^x z^{\tau_i\nu}e^{-wz}dF(z)\right)dF(x).$$
Substitutions similar to those used in equations (22) or (23) then lead to (50). □

Writing $T_{n,s}$ instead of $T_{n,s,0}$, we can easily derive a formula for the case where only the largest $s+1$ statistics are considered in the quotient.

Corollary 2.4.3. *The limit of the expectations* $\mathbb{E}[T_{n,s}]$ *is given by*

$$\lim_{n\to\infty}\mathbb{E}[T_{n,s}(\nu,\theta,\eta)]=\frac{1}{\Gamma(\eta)}\sum_{m=0}^{\theta}\int_0^{\infty}u^{\nu(\theta-m)-\alpha-1}e^{-u}\Theta_{m,s}(u)du,\tag{53}$$

where
$$\Theta_{m,s}(u)=\alpha^s u^{(s+1)\alpha}\sum_{\boldsymbol{\tau}\in S'_m(s)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\prod_{i=1}^{s}\int_u^{\infty}v^{\tau_i\nu-\alpha-1}e^{-v}dv.$$

Proof. This follows immediately from Proposition 2.4.2 after setting $r=0$ and observing that
$$\sum_{p=0}^{m}\binom{\theta}{p}\sum_{\boldsymbol{\tau}\in S'_{m-p}(s-1)}\begin{bmatrix}\theta-p\\ \boldsymbol{\tau}\end{bmatrix}\int_u^{\infty}v^{p\nu-\alpha-1}e^{-v}\prod_{i=1}^{s-1}\int_u^{v}z^{\tau_i\nu-\alpha-1}e^{-z}dzdv=$$
$$\frac{1}{s}\sum_{\boldsymbol{\tau}\in S'_m(s)}\begin{bmatrix}\theta\\ \boldsymbol{\tau}\end{bmatrix}\prod_{i=1}^{s}\int_u^{\infty}v^{\tau_i\nu-\alpha-1}e^{-v}dv.$$
□

In the definition of $\Theta_{m,s,r}$ and $\Theta_{m,s}$ we allow the $\boldsymbol{\tau}$'s to vary over sets of the form $S'_m(s)$. Switching over to sums over $S_m(p)$, we can write
$$\Theta_{0,s,r}(u)=\alpha^{s-r}u^{(s+1)\alpha}\int_u^{\infty}v^{-(r+1)\alpha-1}e^{-v}\left(\int_u^{v}z^{-\alpha-1}e^{-z}dz\right)^{s-r-1}dv$$

and for $m \geq 1$,

$$\Theta_{m,s,r}(u) = \alpha^s u^{(s+1)\alpha} \sum_{p=0}^{m} \binom{\theta}{p} \sum_{q=1}^{\min(\rho,m-p)} \binom{\rho}{q} \left(\int_u^\infty v^{-\alpha-1} e^{-v} dv \right)^{\rho-q} \sum_{\boldsymbol{\tau} \in S_q(m-p)} \begin{bmatrix} \theta \\ \boldsymbol{\tau} \end{bmatrix} \Gamma(\tau_i \nu - \alpha, u, v).$$

with $\rho = s - r - 1$ and $\Gamma(x,u,v) = \int_u^v t^{x-1} e^{-t} dt$. Similarly,

$$\Theta_{0,s}(u) = \alpha^s u^{(s+1)\alpha} \left(\int_u^\infty v^{-\alpha-1} e^{-v} dv \right)^s$$

and for $m \geq 1$,

$$\Theta_{m,s}(u) = \alpha^s u^{(s+1)\alpha} \sum_{p=1}^{\min(s,m)} \binom{s}{p} \left(\int_u^\infty v^{-\alpha-1} e^{-v} dv \right)^{s-p} \sum_{\boldsymbol{\tau} \in S_p(m)} \begin{bmatrix} \theta \\ \boldsymbol{\tau} \end{bmatrix} \Gamma(\tau_i \nu - \alpha, u).$$

Remark 2.4.4. The discussion following Proposition 2.4.1 also shows that indeed

$$\lim_{s\to\infty} \lim_{n\to\infty} \mathbb{E}[T_{n,s}] = \lim_{n\to\infty} \mathbb{E}[T_n]$$

monotonically. Given the form of the first double limit, one could wonder whether some sort of limiting function Θ_m exists for the $\Theta_{m,s}$ that could work as sort of kernels for $\lim_{n\to\infty} \mathbb{E}[T_n]$. However, observe that for every $0 \leq p < s$ and $u > 0$ we have

$$\alpha^{s-p} u^{(s-p)\alpha} \left(\int_u^\infty v^{-\alpha-1} e^{-v} dv \right)^{s-p} = \left(e^{-u} - u^\alpha \int_u^\infty v^{-\alpha} e^{-v} dv \right)^{s-p}.$$

Since the expression inside the parenthesis on the right hand side of the equality is clearly positive and smaller than 1, we deduce that $\lim_{s\to\infty} \Theta_{s,m} = 0$ pointwise and $\Theta_{s,m}$ cannot converge to $\Theta_m \neq 0$ pointwise or in L^q for any q. $\square$

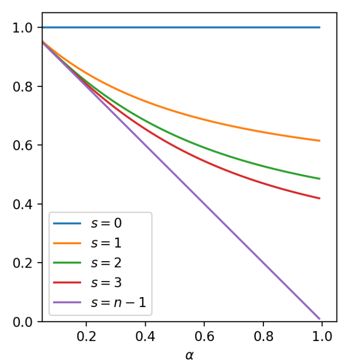


FIGURE 2. $\lim_{n\to\infty} \mathbb{E}[T_{n,s}(2,1,2)]$ for $s = 0, 1, 2, 3, n-1$.

As an example, let us consider in more detail the specific case of squares:

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{X_{n,n}^2 + \cdots + X_{n-s,n}^2}{(X_{n,n} + \cdots + X_{n-s,n})^2}\right] = \lim_{n\to\infty} \mathbb{E}[T_{n,s}(2,1,2)]$$

$$= \alpha^{s+1} \int_0^\infty u^{s\alpha-1} e^{-u} \left[u^2 \left(\int_u^\infty v^{-\alpha-1} e^{-v} dv\right)^s + s \left(\int_u^\infty v^{-\alpha} e^{-v} dv\right) \left(\int_u^\infty v^{-\alpha-1} e^{-v} dv\right)^{s-1}\right] du$$

$$= (s+1)\alpha^{s+1} \int_0^\infty u^{(s+1)\alpha-1} \left(\int_u^\infty v^{1-\alpha} e^{-v} dv\right) \left(\int_u^\infty v^{-\alpha-1} e^{-v} dv\right)^s du$$

$$= (s+1)\alpha^{s+1} \int_0^\infty u^{(s+1)\alpha-1} \Gamma(2-\alpha, u) \big(\Gamma(-\alpha, u)\big)^s du,$$

which quantifies what extent of the limit $1-\alpha$ is already formed by keeping only the $s+1$ largest terms in the sums of the numerator and denominator of the classical Greenwood statistic. Figure 2 depicts this limit as a function of α for keeping the one, two, three, four largest terms only, and compares it with the case of keeping all terms ($s = n-1$). One observes that for very small α these few largest terms already capture the overall behavior quite accurately, which is well in line with the intuition that in that situation the largest terms strongly dominate the sum. In contrast, for values of α closer to 1, significantly more largest order statistics are needed to closely approximate the asymptotic behavior of the Greenwood statistic.

CHAPTER 3

Optimal dividend bands revisited: a gradient based method and evolutionary algorithms

This chapter is based on the following article:

H. Albrecher, and B. Garcia Flores. Optimal dividend bands revisited: a gradient-based method and evolutionary algorithms. *Scandinavian Actuarial Journal* 2023.8 (2023): 788-810.

Abstract.We reconsider the study of optimal dividend strategies in the Cramér-Lundberg risk model. It is well-known that the solution of the classical dividend problem is in general a band strategy. However, the numerical techniques for the identification of the optimal bands available in the literature are very hard to implement and explicit numerical results are known for very few cases only. In this paper we put a gradient-based method into place which allows to determine optimal bands in more general situations. In addition, we adapt an evolutionary algorithm to this dividend problem, which is not as fast, but applicable in considerable generality, and can serve for providing a competitive benchmark. We illustrate the proposed methods in concrete examples, reproducing earlier results in the literature as well as establishing new ones for claim size distributions that could not be studied before.

3.1. Introduction

Consider the optimal dividend problem for an insurance company whose surplus process evolves according to the Cramér-Lundberg model (see e.g. [**17**]). The company pays dividends to shareholders in continuous time according to some admissible strategy π, and the objective is to identify the strategy that maximizes the expected sum of discounted dividend payments until the event of ruin. If there are no constraints on the size of the payments, Gerber [**63**] showed that such a strategy always exists and is given by a *band strategy* that partitions the interval $[0,\infty)$ into three sets $\mathscr{B}_0, \mathscr{B}_c$ and $\mathscr{B}_\infty$: whenever the current surplus level is in $\mathscr{B}_0$, no dividends are paid, when the current surplus level is in $\mathscr{B}_c$, all incoming premium is paid as dividends so that the same surplus level is maintained until the next claim arrives, and when the current surplus level is in $\mathscr{B}_\infty$, a lump sum payment to the first surplus level outside of $\mathscr{B}_\infty$ is carried out. This leads to a cascading strategy towards ruin (see e.g. the sample path illustration in Figure 1 later). Since Gerber's result, the optimal dividend problem and variants have been studied under many different and more general model assumptions and under various constraints (see for instance Avanzi [**18**] and Albrecher & Thonhauser [**10**] for surveys).

For many claim size distributions of practical relevance, the above band strategy turns out to collapse to a *barrier strategy* (i.e., $|\mathscr{B}_c| = 1$), see e.g. Gerber & Shiu [**66**], and Avram et al. [**20**] and Loeffen [**89**] for sufficient conditions on the Lévy measure under which a barrier strategy is optimal in the general framework of spectrally negative Lévy processes. In such a case the determination of the respective optimal barrier is rather straight-forward, if the scale function corresponding to the underlying surplus process is explicitly available (see e.g. Hubalek & Kyprianou [77]). In contrast, already for the classical Cramér-Lundberg model it turns out to be surprisingly challenging to (even numerically) identify the sets $\mathscr{B}_0, \mathscr{B}_c, \mathscr{B}_\infty$ for more involved claim size distributions.

The value function of the optimal dividend problem has been identified as a viscosity solution of the corresponding Hamilton-Jacobi-Bellman equation by Azcue & Muler [22], which led to an iterative procedure for the numerical determination of the optimal bands (cf. [22, 23] as well as Schmidli [106] and Berdel [28]). However, when following the respective algorithm, one faces two main difficulties:

(i) There is no complete understanding of the internal mechanism which furnishes the sets $\mathscr{B}_0, \mathscr{B}_c$ and $\mathscr{B}_\infty$ given an arbitrary set of parameters of the Cramér-Lundberg model, so that, as of now, concrete numerical solutions are known only for very few concrete and simple claim distributions.
(ii) The numerical approaches suggested in the literature require the solution of several differential equations, which can be very expensive in terms of computation time.

While the numerical algorithms suggested so far were a by-product of the meticulous study of the existence and uniqueness of the solution of the optimal dividend problem and its characterization, in this paper we would like to take a different route. Relying on the fact that a band strategy is optimal and assuming that there are only finitely many such bands, we are interested to see if there are numerical alternatives to determine these optimal bands more efficiently and/or generally. In particular, we propose two respective numerical algorithms that differ from previous approaches.

The first one exploits the 'cascading' nature of band strategies and establishes a method based on gradients, when the value function is considered as a function of all band levels rather than the initial surplus level. This will lead to a rather fast numerical routine that can also be tailored to work for cases in which the scale function is not explicitly available, but its Laplace transform is (which is the case in general, since that Laplace transform is defined as the reciprocal of the (shifted) Laplace exponent of the underlying spectrally negative Lévy process). For every fixed number of bands, one then obtains an iterative procedure and can finally compare whether increasing the number of bands still improves the solution.

The second method also uses the explicit iterative formula for the expected discounted dividend payments for given band levels and uses it as the objective function in a general *evolutionary algorithm*. Evolutionary strategies (ES) have been applied in the past with some success in reinsurance problems where the evaluation of the function to optimize is only possible through numerical procedures due to the non-existence of explicit algebraic expressions (see, for example, Salcedo-Sanz et al. [103] and Roman et al. [100]). In our context, we propose the use of an evolutionary self-adaptive strategy based on the algorithm originally proposed in the survey paper by Beyer & Schwefel [31] which does not use the derivatives of involved functions and which can be easily implemented in common programming languages. This genetic algorithm is rather flexible and works particularly well in high-dimensional problems. We will hence adapt such an algorithm to the needs of our dividend problem, and indeed get to the same solutions as the other methods do. While in reasonably low dimension (like the dividend problem typically is, as there are only a few bands to consider) the computation time using this algorithm is not favorable when compared to the gradient-based approach, it is applicable in very general setups and can nicely serve as a benchmark for numerical procedures. Furthermore, it can also be useful in other application areas in risk theory, and since the idea and implementation of evolutionary algorithms may not be so commonly familiar in the risk theory community, we present the underlying principle and implementation variants in some detail in a separate section.

The rest of the paper is organized as follows. Section 3.2 contains some definitions and preliminaries on the model assumptions and the nature of the dividend band strategies. In Section 3.3 we summarize the previously available numerical procedures for the determination of optimal dividend bands. Section 3.4 then provides the expressions for the expected discounted payments that will be used in the numerical algorithms, in particular with respect to the 'cascading' view. Section 3.5 develops the gradient-based algorithm and discusses issues of its implementation. In Section 3.6 we give a general account of the idea behind evolutionary algorithms, and the necessary adaptations to the optimal dividend problem are discussed in Section 3.7. Finally, in Section 3.8 we provide numerical illustrations. We first re-derive the known optimal bands for the well-known example of an Erlang(2) claim size distribution derived in Azcue & Muler [**22**] as well as the mixture of Erlang claim size distribution established in Berdel [**28**]. We then establish a new instance of a mixure of Erlang distributions for which a 4-band strategy is optimal. Subsequently we use our algorithms to derive the optimal barrier level in a risk model with Pareto claim sizes (where a barrier is known to be optimal due to Loeffen [**89**]). We also implement an example with a mixture of Erlang and Pareto claims, which could not be handled with previously existing techniques and for which also two bands turn out to be optimal. Section 3.9 concludes.

3.2. Definitions and preliminaries

Consider the surplus process of an insurance portfolio in a Cramér-Lundberg model

$$C_t = u + pt - \sum_{k=1}^{N_t} Y_k, \quad t \geq 0,$$

with $(N_t)_{t\geq 0}$ a homogeneous Poisson process with rate λ representing the arrival of claims, $(Y_k)_{k\geq 0}$ a sequence of absolutely continuous i.i.d. claim size random variables with density f_Y and finite mean μ, and the premium rate p satisfying the positive safety loading condition $p = (1+\eta)\lambda\mu$ for some $\eta > 0$. Let $(\mathscr{F}_t)_{t\geq 0}$ be the usual augmentation of the filtration generated by $(C_t)_{t\geq 0}$.

The dividend strategy π is represented by the process $(U_t)_{\geq 0}$, where U_t are the dividends paid up to time t. A dividend strategy is called *admissible* if the associated process $(U_t)_{t\geq 0}$ is adapted to $(\mathscr{F}_t)_{t\geq 0}$, non-decreasing and càglàd. Denote by Π the set of all admissible strategies.

For an admissible strategy $\pi \in \Pi$ we denote by $X_t = C_t - U_t$ the surplus after dividend payments. Let

$$\tau_D = \inf\{t \geq 0 \mid X_t < 0\}$$

be the time until ruin, then

$$V_\pi(u) = \mathbb{E}\left[\int_0^{\tau_D} e^{-\delta t} dU_t \Big| X_0 = u\right]$$

is the expected value of the aggregated discounted dividends paid until ruin, where $\delta > 0$ is the force of interest. The objective is then to identify the strategy that maximizes V over all admissible strategies, that is, to find a strategy π^* such that

$$V_{\pi^*} = \sup_{\pi\in\Pi} V_\pi(x). \tag{54}$$

As pointed out in the introduction, the class of *band strategies* is known to be optimal in this case (cf. [**63**]). A band strategy is defined by a partition of the positive half-line $\mathbb{R} = \mathscr{B}_0 \cup \mathscr{B}_c \cup \mathscr{B}_\infty$ with the following properties:

- If $x \in \mathscr{B}_0$, there exists $\varepsilon > 0$ such that $[x, x+\varepsilon) \subset \mathscr{B}_0$.
- $\mathscr{B}_c$ is compact.
- $\mathscr{B}_\infty$ is open in $[0, \infty)$.
- If $x \notin \mathscr{B}_\infty$ and there is a sequence $(x_n) \subset \mathscr{B}_\infty$ such that $x_n \to x$, then $x \in \mathscr{B}_c$.
- $(\sup \mathscr{B}_c, \infty) \subset \mathscr{B}_\infty$.

$\mathscr{B}_0$ corresponds to all surplus levels for which $dU_t = 0$ (no dividends are being paid), $\mathscr{B}_c$ is the set of surplus levels for which $dU_t = p\,dt$ (all incoming premium is paid as dividends) and $\mathscr{B}_\infty$ is the set of surplus levels at which $U_{t+} - U_t = X_t - \sup\{b \in \mathscr{B}_c \mid X_t > b\}$ is applied (the smallest possible lump sum is paid with which one leaves the set $\mathscr{B}_\infty$). As the focus of this paper is to provide alternatives to numerically identifying the optimal bands, we restrict ourselves to finitely many bands, i.e., for given levels $a_0 = 0 \le b_0 < a_1 < \cdots < b_{m-1}$, the band strategy is given by

$$\mathscr{B}_0 = [0, b_0) \cup \bigcup_{k=1}^{m-1} [a_k, b_k), \quad \mathscr{B}_c = \bigcup_{k=1}^{m-1} \{b_k\}, \quad \mathscr{B}_\infty = \bigcup_{k=0}^{m-2} (b_k, a_{k+1}) \cup (b_{m-1}, \infty).$$

We refer to this strategy as an *m-band strategy* (see Figure 1 for an illustration of a sample path with a 2-band strategy).

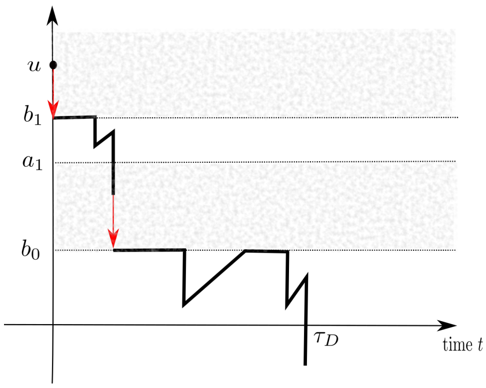


FIGURE 1. A sample path with a 2-band strategy

We conclude this section by making some remarks about notation: for a set $A \subset \mathbb{R}^n$, a function $f : A \to \mathbb{R}$, and a limit point $x \in A$, we denote by $f(x_1, \ldots, x_j-, \ldots, x_n)$ the limit $\lim_{y \to x} f(y)$ through points y for which $y_j < x_j$, given that this limit exists and similarly for $f(x_1, \ldots, x_j+, \ldots, x_n)$. The function f is then continuous at x if and only if $f(x_1, \ldots, x_j-, \ldots, x_n) = f(x_1, \ldots, x_j+, \ldots, x_n)$ for each $j = 1, \ldots, n$, provided all right-hand and left-hand limits exist. Finally, in order to avoid cumbersome notation, we denote the partial derivative in the i-th variable by $D_i f$.

3.3. The identification of bands in previous literature

As mentioned in the introduction, the explicit identification of optimal bands has proved challenging even in the classical Cramér-Lundberg risk model. In the following we briefly summarize the available approaches in the literature. The typical approach is to derive the Hamilton-Jacobi-Bellman (HJB) equation

$$\max\{1 - V'(x), \mathcal{L}(V)(x)\} = 0, \tag{55}$$

related to the stochastic control problem (54), where

$$\mathcal{L}(f)(x) = pf'(x) - (\lambda+\delta)f(x) + \lambda \int_0^x f(x-y) f_Y(y)\, dy, \quad x > 0 \tag{56}$$

is the infinitesimal generator of the discounted surplus process (see e.g. Azcue and Muler [**22, 23**]). Since V is typically not sufficiently regular to satisfy the needs of (55) as a classical solution, one needs to look for viscosity solutions, and it can be shown that $V(x)$ is the unique viscosity solution of (55) satisfying a growth condition and a particular initial condition [**22**]. The numerical approach to find this solution is then an iterative procedure. In particular, Schmidli [**106**] proposed an algorithm for finding the levels of the optimal bands. This algorithm was formalized by Berdel [**28**], who considered the problem for the general case of phase-type claim distributions (cf. Algorithm 1). Here, W_δ is the scale function (see

Input : Scale function W_δ and infinitesimal generator $\mathcal{L}$.
Output: Levels $B^* = (b_0^*, a_1^*, \dots, b_{M-1}^*)$ of the best band strategy.

1 **begin**
2 $\quad m := 0;$
3 $\quad b_0 := \sup\left\{ x \geq 0 \mid W_\delta'(x) = \inf_{y \geq 0} W_\delta'(y) \right\};$
4 $\quad V_0(x) := \begin{cases} W_\delta(x)/W_\delta'(b_0) & x \leq b_0 \\ V_0(b_0) + x - b_0 & x > b_0 \end{cases};$
5 $\quad$ **while** $\mathcal{L}(V_m)(x) > 0$ *for some* $x > b_m$ **do**
6 $\quad\quad \mathscr{G}_m := \{ f_{m+1}^a, a > b_m \mid f_{m+1}^a(x) = V_m(x),\ x \leq a$ and $\mathcal{L}(f)_{m+1}^a(x) = 0,\ x > a \};$
7 $\quad\quad a_{m+1} := \inf\left\{ a > b_m \mid \inf_{x > a} f_{m+1}^{a\prime}(x) = 1 \right\};$
8 $\quad\quad b_{m+1} := \sup\left\{ x > a_{m+1} \mid f_{m+1}^{a_{m+1}\prime}(x) = 1 \right\};$
9 $\quad\quad V_{m+1}(x) := \begin{cases} f_{m+1}^{a_{m+1}}(x) & x \leq b_{m+1} \\ f_{m+1}^{a_{m+1}}(b_{m+1}) + x - b_{m+1} & x > b_{m+1} \end{cases};$
10 $\quad\quad m := m + 1$
11 $\quad$ **end**
12 **end**

ALGORITHM 1: Schmidli's algorithm

e.g. [**17**, Ch.XI]). When the optimal strategy is in fact a finite band strategy, Algorithm 1 is guaranteed to converge. However, depending on the particular distribution of the claims, the procedure can be computationally complex, as can be seen from Lines 6 to 8. In each iteration of the algorithm, a family of functions $\mathscr{G}_m$ parametrized by the interval (b_m, ∞) is defined in such a way that for each $a > b_m$, the function f_{m+1}^a solves (55) on $x > a$ with boundary condition $f_{m+1}^a(a) = V_m(a)$. Apart from some cases where this can be done explicitly, (55) has in general to be solved numerically. While this might not represent a problem for a couple of values of $a > b_m$, the difficulty arises when we consider Lines 7 and 8, since they presuppose a full knowledge of the solutions in the entire interval (b_m, ∞) in order to compute the necessary extrema. As stated in [**28**], one can define $\bar{a}_{m+1} = \inf\{x > b_m \mid \mathcal{L}(V_m)(x) > 0\}$ and restrict the family to the interval $(b_m, \bar{a}_{m+1})$. However, this introduces another extremum to be computed and one still has to consider the trade-off that arises at each step of the procedure when choosing a grid fine enough to discretize this new interval.

An alternative iterative algorithm for finding the optimal bands is proposed in Avram et al. [**21**], using stochastic sub- and super-solutions of (55) (their approach is formulated for general spectrally negative Lévy processes and the inclusion of fixed transaction costs with every dividend payment). Similarly to Algorithm 1, optimal levels are found in a sort of "upwards" approach finding higher band levels at each step of the procedure. However, instead of solving integro-differential equations, each step consists of solving a stochastic control problem expressed through Gerber-Shiu functions. The advantage of this is that the problem is then reduced to finding the extrema of a low-dimensional function at each iteration, and there is no need for a full set of solutions of (55) as seen in, for example, Lines 7 and 8 of Algorithm 1. The Algorithm 2 presented later in this paper sets out from a top-down approach, and then also leads to a bottom-up procedure that is formulated via discounted deficit densities explicitly (rather than general Gerber-Shiu functions), so that it eventually can be interpreted as a particular customization and implementation of the algorithm by Avram et al., see the details below.

3.4. Properties of the value function

In this section we recollect some properties of V_π, which will form the basis for the implementation of the numerical algorithms presented later.

For a fixed set of levels $a_0 = 0 \leq b_0 < a_1 < \cdots < b_{m-1}$ of an m-band strategy π, we observe the following: for any $0 \leq k \leq m-1$ and initial capital u in $[a_k, a_{k+1})$ (here $a_0 = 0$ and $a_m = \infty$), the amount of dividends paid in a realization of the process will be the same as in a process with a barrier strategy with initial capital $u - a_k$ and barrier $b_k - a_k$, up until a claim makes the original process go below a_k. Denoting by V_b the value function of a barrier strategy with barrier b, space-homogeneity and the Markov property imply that

$$V_\pi(u) = V_{b_k - a_k}(u - a_k) + \mathbb{E}\left[V_\pi(a_k - D(u - a_k, b_k - a_k))\right], \quad a_k \leq u < a_{k+1}, \tag{57}$$

where $D(u - a_k, b_k - a_k)$ denotes the deficit at ruin of a process with initial capital $u - a_k$, for which a barrier strategy with barrier $b_k - a_k$ is applied. In many cases, the density of the deficit at ruin can be computed by means of Gerber-Shiu functions (see [**66, 83**]) and the dividends-penalty identity (see [**87, 64**]). Assume henceforth that $D(u - a_k, b_k - a_k)$ has a density, which we denote by $f_D(\cdot, u - a_k, b_k - a_k)$. We observe that the variable inside the expectation in (57) is non-zero only when the deficit is at most a_k. We can therefore rewrite (57), for any $k = 0, \ldots, m-1$, as

$$V_\pi(u) = V_{b_k - a_k}(u - a_k) + \int_0^{a_k} V_\pi(a_k - y) f_D(y, u - a_k, b_k - a_k) dy, \quad a_k \leq u < a_{k+1}. \tag{58}$$

This set of equations provides the central formulas for computing the value of V_π given a fixed set of levels: For $0 \leq u < a_1$, the value of $V_\pi(u)$ is equal to $V_{b_0}(u)$, which is given in terms of the scale function of the process. We can then plug in these values in the integral in Equation (58) to obtain the value of $V_\pi(u)$ for $a_1 \leq u < a_2$ and repeat this procedure in an iterative way to obtain the value of $V_\pi(u)$ for every u. The problem of evaluating $V_\pi(u)$ is therefore reduced to the computation of the scale function W_δ and the density f_D. However, with knowledge of the scale function, the latter can be computed by means of the formula

$$f_D(y, u, b) = \lambda \int_0^\infty \left(\frac{W_\delta(u) W_\delta'(b - z)}{W_\delta'(b)} - W_\delta(u - z) \right) f_Y(y + z)\, dz. \tag{59}$$

see, e.g., [**83**, Ch.X]. The setting in this last reference is that of general Lévy processes. A more basic approach is to consider first the density $f_{D^0}(y, u)$ of the deficit at ruin with initial capital u in the absence of a dividend strategy. Let $\hat{f}_Y(s)$ denote the Laplace transform of the claim size density f_Y. Following e.g. [**17**, Ch.XII], we know that $f_{D^0}(y,u)$ can be obtained as the inverse Laplace transform of

$$\int_0^\infty e^{-su} f_{D^0}(y,u)\,du = \frac{\lambda(\hat{w}(y,\rho)-\hat{w}(y,s))}{ps-\delta-\lambda+\lambda\hat{f}_Y(s)}, \tag{60}$$

where $\hat{w}:[0,\infty)^2\to\mathbb{R}$ is given by

$$\hat{w}(y,s) = \int_0^\infty e^{-su} f_Y(y+u)\,du. \tag{61}$$

From the dividends-penalty identity [**64**] we then have

$$f_D(y,u,b) = f_{D^0}(y,u) - \frac{W_\delta(u)}{W_\delta'(b)} D_2 f_{D^0}(y,b). \tag{62}$$

Note that we require W_δ to be differentiable in order to be able to use these formulas. Nonetheless, if for some $\alpha<1$ and $C>0$, we have $f_Y(x)\le Cx^{-1-\alpha}$ for x in some neighbourhood of the origin, then $W_\delta\in C^{q+2}(0,\infty)$ whenever $f_Y\in C^q(0,\infty)$ (see [**82**]), where $C^q(0,\infty)$ refers to the set of q-times continuously differentiable functions on the real positive line. A formula similar to (59) shows that the statement is also valid whenever we replace W_δ by f_{D^0}.

Equation (58) also reveals further properties of V_π when we shift the focus from the initial capital u to the band limits: for $m\ge 1$, we can identify the set of m-band strategies with the set

$$\mathcal{B}_m = \{x\in\mathbb{R}^{2m-1} \mid 0\le x_1\le\cdots\le x_{2m-1}\},$$

and, for fixed $u>0$, we can consider the function $V^m:\mathcal{B}_m\to[0,\infty)$ given by $x\mapsto V_x(u)$. We have the following:

Proposition 3.4.1. *If $f_Y\in C^q(0,\infty)$, $q\ge 2$, the function V^m is continuously differentiable in the interior of the set*

$$\mathcal{C}_m = \mathcal{B}_m\cap\{x\in\mathbb{R}^{2m-1} \mid x_{2j-2}\ne u, j=1,\dots,m\}.$$

When $m=1$, we take $\{x\in\mathbb{R}^{2m-1} \mid x_{2m-2}\ne u\}$ to be equal to $\mathbb{R}$.

Proof. We proceed by induction on m. For $m=1$, $\mathcal{C}_1=[0,\infty)$ and $\mathcal{D}_1=[0,u)\cup(u,\infty)$. For $0\le u<b$, we have

$$V^1(b) = \frac{W_\delta(u)}{W_\delta'(b)},\quad V^{1\prime}(b) = -\frac{W_\delta(u)W_\delta''(b)}{W_\delta'(b)^2},$$

while for $0\le b<u$,

$$V^1(b) = u-b+V_b(b) = u-b+\frac{W_\delta(b)}{W_\delta'(b)},\quad V^{1\prime}(b) = -\frac{W_\delta(b)W_\delta''(b)}{W_\delta'(b)^2}.$$

Since $V^1(u-)=V^1(u+)=W_\delta(u)/W_\delta'(u)$ and $V^{1\prime}(u-)=V^{1\prime}(u+)=W_\delta(u)W_\delta''(u)/W_\delta'(u)^2$, the claim follows. Now, assume the claim is true for some $m\in\mathbb{N}$. We can write $\mathcal{C}_{m+1}$ as $\mathcal{C}_{m+1}=A\cup B\cup C$, where

$$\begin{aligned}
A &= \{x\in\mathcal{C}_{m+1} \mid u<x_{2m-2}\},\\
B &= \{x\in\mathcal{C}_{m+1} \mid x_{2m-2}<u\le x_{2m-1}\},\\
C &= \{x\in\mathcal{C}_{m+1} \mid x_{2m-1}<u\}.
\end{aligned}$$

We observe the following: on A, $V^{m+1}=V^m\circ\pi$, where $\pi:\mathbb{R}^{2m+1}\to\mathbb{R}^{2m-1}$ is the projection onto the first $2m-1$ coordinates. Since π maps the interior of A into the interior of $\mathcal{C}_m$, by the induction hypothesis, it follows that V^{m+1} is continuously differentiable in that set. Now, from (58) we have, for $b = (b_0, a_1, \dots, b_m)$ in B

$$V(b) = \frac{W_\delta(u - a_m)}{W'_\delta(b_m - a_m)} + \int_0^{a_m} V_y^m(b) f_D(a_m - y, b_m - a_m, b_m - a_m) dy,$$

while for b in C,

$$V(b) = u - b_m + \frac{W_\delta(b_m - a_m)}{W'_\delta(b_m - a_m)} + \int_0^{a_m} V_y^m(b) f_D(a_m - y, b_m - a_m, b_m - a_m) dy.$$

From (60) and (62) we see that under the assumptions of the proposition, f_D is twice continuously differentiable. Hence, using the induction hypothesis once again, it follows that V is continuously differentiable in the interiors of B and C. Moreover, since

$$\begin{aligned} V(b_0, a_1, \dots, u-) &= V(b_0, a_1, \dots, u+), \\ D_{2m} V(b_0, a_1, \dots, u-) &= D_{2m} V(b_0, a_1, \dots, u+), \end{aligned}$$

and

$$D_{2m+1} V(b_0, a_1, \dots, u-) = D_{2m+1} V(b_0, a_1, \dots, u+),$$

we conclude that V is continuously differentiable in the interior of $\mathcal{C}_{m+1}$, concluding the proof. □

Remark 3.4.2. By considering instead the set

$$\mathcal{D}_m = \mathcal{B}_m \cap \{x \in \mathbb{R}^{2m-1} \mid x_j \neq u, j = 1, \dots, 2m-1\},$$

we can, in a similar manner, conclude that if $f_Y, W_\delta, f_{D^0} \in C^q(0, \infty)$, $q \geq 2$, then V^m is $q-1$ times differentiable in the interior of $\mathcal{D}_m$, since in this case one does not have to consider the "pasting" at the points where $b_j = u$.

3.5. A gradient-based method

From Proposition 3.4.1 and its proof, we can compute the gradient of the value function when we fix the initial capital u and we look at it as a function of the levels. Assuming that W_δ and f_{D^0} are twice differentiable and setting

$$\mathcal{E}_m = \mathcal{B}_m \cap \{x \in \mathbb{R}^{2m-1} \mid x_{2m-2} < u\},$$

we have for $b \in \mathcal{E}_m$ and $u < b_{m-1}$,

(63)
$$D_{2m-1}V^m(b) = -\frac{W_\delta(u-a_{m-1})W_\delta''(b_{m-1}-a_{m-1})}{W_\delta'(b_{m-1}-a_{m-1})^2} + \int_0^{a_{m-1}} D_3 f_D(a_{m-1}-y, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(b)dy,$$

(64)
$$\begin{aligned} D_{2m-2}V^m(b) = &\frac{W_\delta(u-a_{m-1})W_\delta''(b_{m-1}-a_{m-1}) - W_\delta'(u-a_{m-1})W_\delta'(b_{m-1}-a_{m-1})}{W_\delta'(b_{m-1}-a_{m-1})^2} \\ &+ \int_0^{a_{m-1}} D_1 f_D(a_{m-1}-y, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(b)dy \\ &- \sum_{i=2}^{3} \int_0^{a_{m-1}} D_i f_D(a_{m-1}-y, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(b)dy \\ &+ f_D(0, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(a_{m-1}), \end{aligned}$$

while for $u > b_{m-1}$

(65)
$$D_{2m-1}V^m(b) = -\frac{W_\delta(b_{m-1}-a_{m-1})W_\delta''(b_{m-1}-a_{m-1})}{W_\delta'(b_{m-1}-a_{m-1})^2} + \int_0^{a_{m-1}} D_3 f_D(a_{m-1}-y, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(b)dy,$$

(66)
$$\begin{aligned} D_{2m-2}V^m(b) = &\frac{W_\delta(b_{m-1}-a_{m-1})W_\delta''(b_{m-1}-a_{m-1})}{W_\delta'(b_{m-1}-a_{m-1})^2} - 1 \\ &+ \int_0^{a_{m-1}} D_1 f_D(a_{m-1}-y, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(b)dy \\ &- \sum_{i=2}^{3} \int_0^{a_{m-1}} D_i f_D(a_{m-1}-y, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(b)dy \\ &+ f_D(0, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) V_y^{m-1}(a_{m-1}), \end{aligned}$$

and, in both cases, for $1 \leq i < 2m-2$,

(67)
$$D_i V^m(b) = \int_0^{a_{m-1}} f_D(a_{m-1}-y, b_{m-1}-a_{m-1}, b_{m-1}-a_{m-1}) D_i V_y^{m-1}(b)dy.$$

We can solve these equations in an iterative manner to find candidate levels for the optimal band strategy whenever it is finite: call $b_0^* < a_1^* < \ldots < b_{M-1}^*$ the levels of the optimal strategy and assume for the moment that $b_0^* > 0$. Since the first level has to occur at the largest global minimum of W_δ', we have $b_0^* = \sup\{x \geq 0 \mid W_\delta'(x) = \inf_{y \geq 0} W_\delta'(y)\}$. We observe that $W_\delta''(b_0^*) = 0$, so $D_1 V^1(b_0^*) = 0$, independently of u.

If the barrier strategy at b_0^* is not optimal and u is such that $b_1^* < u < a_2^*$, b_0^*, a_1^* and b_1^* solve (65) to (67) when $m = 2$, since the global maximum is attained at the 2-band strategy with levels b_0^*, a_1^* and b_1^*. Hence, since (67) is zero regardless of a_1 and b_1 when $b_0 = b_0^*$, we see that a_1^* and b_1^* are within the solutions to (65) and (66). Moreover, these equations can be solved without regards to u, and, if we optimally choose the solution (so that we end up obtaining a_1^* and b_1^*) we will see that b_0^*, a_1^* and b_1^* solve (63) to (67) regardless of the value of $u > a_1^*$.

Continuing in this fashion, if a two-band strategy is not optimal, for $m = 3$ and $i = 2, 3$, we have,

$$\begin{aligned} D_i V(b) &= \int_0^{a_2} f_D(a_2 - y, b_2 - a_2, b_2 - a_2) D_i V_y^2(b) dy \\ &= \int_0^{a_1^*} f_D(a_2 - y, b_2 - a_2, b_2 - a_2) D_i V_y^2(b) dy \\ &\quad + \int_{a_1^*}^{a_2} f_D(a_2 - y, b_2 - a_2, b_2 - a_2) D_i V_y^2(b) dy. \end{aligned}$$

On the interval $(0, a_1^*)$, $V^2(b) = V^1(b)$ as functions of the initial capital, so $D_i V^2(b) = 0$. By the remarks of the previous paragraph, we also have $V_y^2(b) = 0$ for all $y > a_1^*$, so we see that (67) is always zero regardless of the value of a_2 and b_2. Hence, we can proceed again by solving (65) and (66), choosing an optimal solution and test whether we proceed further with another band. At the $m + 1$-th step, the equations that we need to solve can be written in a simpler form as

(68)
$$0 = \frac{W_\delta(b_m - a_m) W_\delta''(b_m - a_m)}{W_\delta'(b_m - a_m)^2} - \int_0^{a_m} D_3 f_D(a_m - y, b_m - a_m, b_m - a_m) V_y^m(b^*) dy,$$

(69)
$$1 = f_D(0, b_m - a_m, b_m - a_m) V_{a_m}^m(b) + \int_0^{a_m} D_1 f_D(a_m - y, b_m - a_m, b_m - a_m) V_y^m(b^*) dy$$

where $b^* = (b_0^*, a_1^*, \dots, b_{m-1}^*)$.

If $b_0^* = 0$, we can discard the equation for $D_1 V^m$ and work instead on the interior of the set

$$\mathcal{E}_m' = \{x \in \mathbb{R}^{2m-2} \mid b_0^* \leq x_1 \leq \cdots \leq x_{2m-3} \leq \min(u, x_{2m-2})\}$$

by realizing that (63) to (67) for $2 \leq i < 2m - 2$ carry over verbatim to this situation. The same argument then shows that we can use the same procedure for obtaining the optimal levels. The procedure is described in Algorithm 2.

Algorithm 2 starts by obtaining the first level b_0^* of the optimal band strategy. If the barrier strategy at b_0^* is optimal, the algorithm finishes. Otherwise, the algorithm enters into its main loop. After updating the number of bands, the loop proceeds to an application of an abstract `solve` function in Line 7, which is used to solve simultaneously Equations (65) and (66) (or equivalently, (68) and (69)) when the first $2m-1$ variables of $D_{2m}V$ and $D_{2m+1}V$ are fixed. In Line 8, the best levels a_{m-1}^* and b_{m-1}^* are chosen by selecting the couple that produces the best value for the value function when the initial capital is set to b_{m-1}^*. Line 9 uses the function `defineV` for creating the value function of the strategy with the levels $(b_0^*, a_1^*, \dots, b_{m-1}^*)$ found so far, finalizing the loop.

Input : Scale function W_δ and density of deficit f_D.
Output: Levels $B^* = (b_0^*, a_1^*, \dots, b_{M-1}^*)$ of the best band strategy.

begin
  $m := 0;$
  $b_0 := \sup\left\{x \geq 0 \mid W_\delta'(x) = \inf_{y\geq 0} W_\delta'(y)\right\};$
  $V^0(x) := \begin{cases} W_\delta(x)/W_\delta'(b_0) & x \leq b_0 \\ V^0(b_0) + x - b_0 & x > b_0 \end{cases};$
  while $\mathcal{L}(V^m)_x > 0$ *for some* $x > b_m$ **do**
    $m := m + 1;$
    $B^* = \texttt{solve}(D_{2m}V(b_0^*, a_1^*, \dots, b_{m-1}^*, a_m, b_m) = 0,$
      $D_{2m+1}V(b_0^*, a_1^*, \dots, b_{m-1}^*, a_m, b_m) = 0);$
    $a_m^*, b_m^* := \texttt{select}(B^*);$
    $V^m := \texttt{defineV}(b_0^*, a_1^*, \dots, b_m^*);$
  end
end

Algorithm 2: Gradient-based algorithm for optimal dividends

As explained before, Algorithm 2 can be considered as a particular implementation of the algorithm proposed by Avram et al. [**21**] but obtained after trying to solve the gradient equations in a sort of ”backward” way. It is similar to Algorithm 1 and Algorithm 4 in the sense that it is not possible to determine beforehand whether a finite band strategy is optimal or not. The advantage is, however, that one avoids having to fully specify the solutions to the HJB equation as in Algorithm 1 while also avoiding the randomness involved in Algorithm 4.

Remark 3.5.1. Note that, in principle, one could derive explicit expressions that the optimal band levels should satisfy by means of equations (68) and (69). However, even in the simple case of the claims following an Erlang(2) distribution, one already arrives at rather complicated expressions which involve combinations and products of exponentials and polynomials, and the resulting levels cannot be given in terms of elementary functions. However, these equations can still be solved through numerical methods, which is the basis of the gradient-based technique that is going to be introduced in the sequel.

3.6. Evolutionary strategies

Evolutionary strategies belong to a class of nature-inspired optimization algorithms which intend to mimic biological evolution by means of procedures roughly categorized as *mutation*, *recombination* and *selection* procedures which incorporate tasks that resemble the way evolution is carried out in nature. Starting with a set of candidate solutions (called the *parental population*), one produces a new set of candidate solutions by means of recombination and, through mutation, randomly alters it to form a second set of candidate solutions (called the *offspring population*). One then uses selection to filter out the best candidate solutions from these two populations and iterates the process, replacing the previous parental population with the new population thus obtained. In general, recombination, mutation and selection tasks are problem-dependent and adjusted according to a diverse set of criteria. ES are classically referred by the way the offspring population is generated, and the notation for expressing it is the $(\mu_{\mathrm{ES}}/\rho_{ES}\overset{+}{,}\lambda_{ES})$-*notation* (the subscripts ES are used here to differentiate these symbols from the previously defined μ and λ in earlier sections). In this notation, μ_{ES} refers to the size of the parental population at the beginning and end of each iteration, ρ_{ES} refers to the amount of parents involved in the creation of one single offspring, randomly chosen without replacement, and λ_{ES} refers to the number of offsprings created in each iteration. The symbols "+" and "," refer to the way selection is carried out: the first one indicates that the μ_{ES} members of the new parental population are going to be extracted from a set obtained by merging the parental population together with the offspring population, while the symbol "," indicates that the parental population is discarded after the creation of the λ_{ES} offsprings (so that in a $(\mu_{\text{ES}}/\rho_{ES}, \lambda_{ES})$ strategy we necessarily require $\lambda_{\text{ES}} \geq \mu_{\text{ES}}$).

The pseudo-code for the algorithm from Beyer [**31**] is presented below in Algorithm 3, formulated in terms of a maximization problem. As suggested before, the basic objects handled by ES are populations, which in this strategy are modeled by tuples of the form $(x, s, f(x))$. In this representation, x is simply a candidate solution belonging to the search space X. The element s is used as a set of parameters aiding in the mutation procedure of the members of the population and leading the self-adaptive properties of the strategy. The last element is the value of the function to optimize at x, which needs to be stored in order to select elements in each iteration.

Input : Function f to maximize in unconstrained object space E.
Output: Solution to problem $x^* = \underset{x \in E}{\operatorname{argmax}} f(x)$

1 **begin**
2 $\quad g := 0;$
3 $\quad$ `initialize(`$\mathscr{P}^{(0)} := \{(x_{0,k}, s_{0,k}, f(x_{0,k})) \mid k = 1, \ldots, \mu_{\text{ES}}\}$`);`
4 $\quad$ **repeat**
5 $\quad\quad$ **for** $l := 1$ **to** λ_{ES} **do**
6 $\quad\quad\quad \mathscr{S}_l :=$ `sample(`$\mathscr{P}^{(g)}, \rho_{\text{ES}}$`);`
7 $\quad\quad\quad \tilde{s}_l :=$ `s_recombination(`$\mathscr{S}_l$`);`
8 $\quad\quad\quad \tilde{x}_l :=$ `x_recombination(`$\mathscr{S}_l$`);`
9 $\quad\quad\quad \tilde{s}'_l :=$ `s_mutation(`$\tilde{s}_l$`);`
10 $\quad\quad\quad \tilde{x}'_l :=$ `x_mutation(`$\tilde{x}_l$,$\tilde{s}'_l$`);`
11 $\quad\quad\quad \tilde{F}_l := F(\tilde{x}'_l);$
12 $\quad\quad$ **end**
13 $\quad\quad \mathscr{O}^{(g)} := \{(\tilde{x}'_l, \tilde{s}'_l, \tilde{F}_l) \mid l = 1, \ldots, \lambda_{\text{ES}}\};$
14 $\quad\quad \mathscr{P}^{(g+1)} :=$ `Selection(`$\mathscr{P}^{(g)}$,$\mathscr{O}^{(g)}$,μ_{ES}`);`
15 $\quad\quad g := g + 1$
16 $\quad$ **until** terminal_condition;
17 **end**

ALGORITHM 3: The basic ES-algorithm

After initialization of the algorithm, which is usually carried out randomly, the algorithm enters into the main loop of the strategy for generating subsequent populations. This loop can roughly be described as an alternation of creating offsprings out of the parental population and selecting the replacing parental population out of these offsprings.

The process for creating offsprings is carried out from Line 5 to 13. Lines 6 to 8 carry out the recombination procedure by first extracting a subsample from $\mathscr{P}^{(g)}$ of size ρ_{ES} without replacement. In real-valued spaces, a common recombination operator is the arithmetic mean, so, for example, $\texttt{s\_recombination}(\mathscr{S}_l) = \rho_{\text{ES}}^{-1} \sum_{\mathscr{S}_l^s} s_k^g$ where $\mathscr{S}_l^s$ is the set of s_k^g that belong to some tuple in $\mathscr{S}_l$. The mutation operator is then applied in lines 9 to 10 by first mutating the strategy parameters and then the candidate solutions afterwards. While there is no established methodology for choosing the mutation operator, in [**30**], Beyer suggests that any operator should satisfy three requirements for a successful implementation of ES: *scalability* (the ability to tune the strength of the mutation), *reachability* (the ability to reach any other state (x, s) within a finite number of steps) and *unbiasedness*. Scalability is achieved by allowing the mutation of the object states, x, to be dependent on s. For N-dimensional real-valued search spaces, the parameter s is generally used for controlling the variance of the mutation and in this regard, theoretical and practical considerations lead to a common mutation operator given by

$$\texttt{s\_mutation}(s_l)^j = s_l^j \exp\left(\tau N_j\right), \quad j = 1, \ldots, N, \tag{70}$$

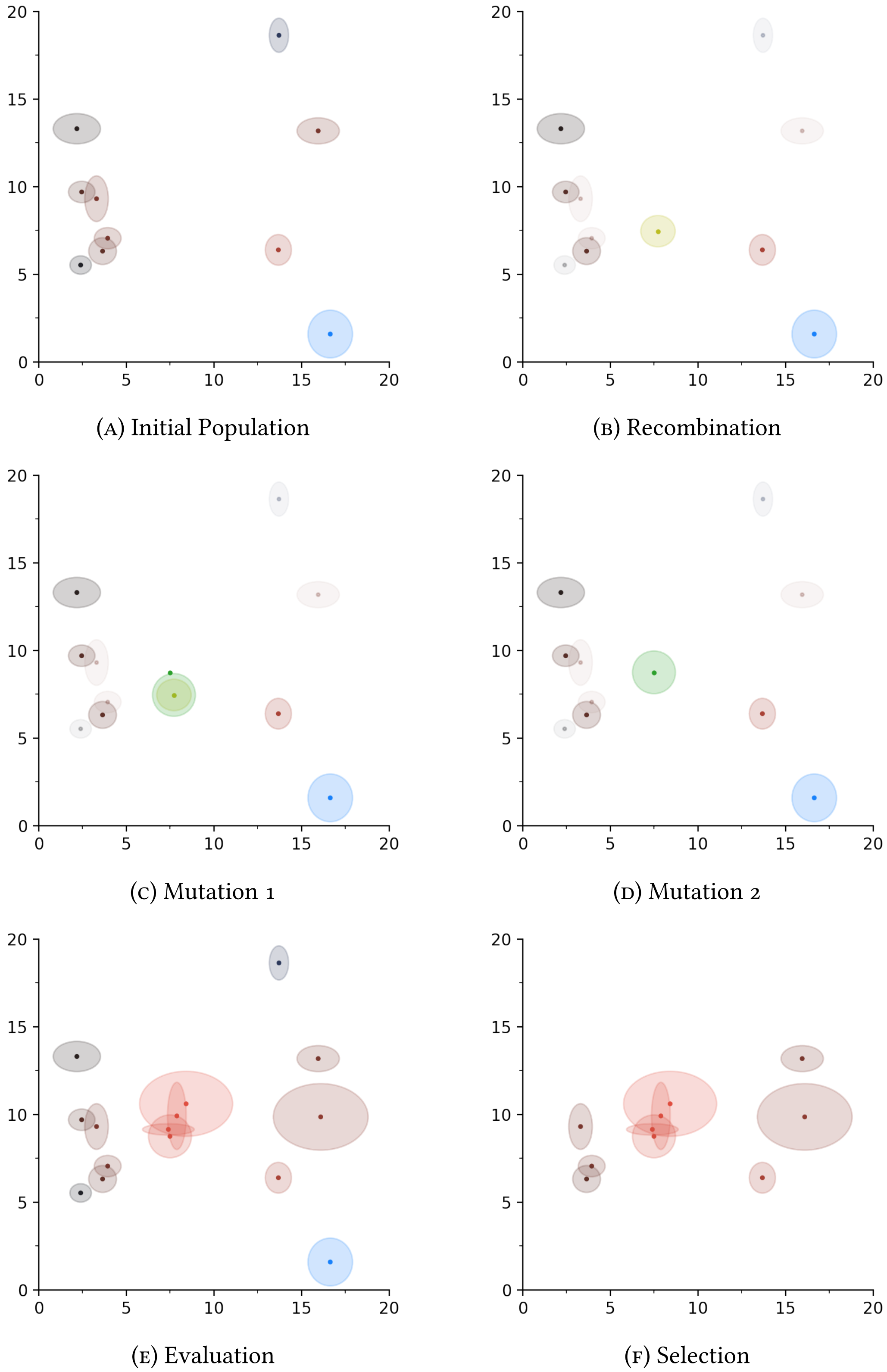
where τ is the learning-rate parameter and N_j is a standard normally distributed random variable. Given the current parental state, the unbiasedness requirement simply means that the mutation procedure should not introduce any bias, and following the so-called *maximum entropy principle*, this requirement immediately leads to mutation operators given by

$$\texttt{x\_mutation}(x_l, \tilde{s}_l)^j = x_l^j + \tilde{s}_l^j Z_j, \quad j = 1, \ldots, N \tag{71}$$

with Z_j a standard normal random variable independent from the variables used to mutate s. However, Yao et al. suggest in [**116**] and [**84**] that, more generally, allowing Z_j to have other kinds of stable distributions improves convergence speed and deals better with problems where several local extrema exist, dealing at once as well with a better handling of the reachability requirement.

The creation of the offspring is finalized in Line 11 by evaluating the objective function in the mutated objects x and the offspring population is gathered in Line 13. The last step of the main loop is achieved in Line 14, where the desired selection (plus or comma) takes place and the new parental population is created.

Figure 2 illustrates one iteration of a $(10/5+5)$ evolutionary strategy in a two-dimensional real space for the function $f(x, y) = \sqrt{\max\{10 - (x-10)^2/2 - (y-10)^2/2, 0\}}$. Mutations for the parameters occur as in equations (70) and (71) with N_j independent standard normal variables. Figure 2A represents the state of the population at the beginning of the iteration, where, in the notation of Algorithm 3, the x_k^g are shown as the center of the ellipses, the s_k^g as their axes and, using a blue-black-red scale, each point and ellipse is colored according to the value of f at x_k^g. Figure 2B shows the first step in the creation of a single offspring: after randomly selecting 5 individuals from the original population (marked by the 5 darkest ellipses), the olive-colored ellipse is created after applying the recombination operator. In this case, recombination is given by the arithmetic mean, so that the olive point and the axes of the olive ellipse are the arithmetic means of the other 5 points and the ellipses' axes respectively. After recombination, mutation takes places, which is represented in Figure 2C. Here, the axes of the olive ellipse are mutated according to equation (70), generating the green ellipse. Sampling from the normal distribution centered in the olive point and variance given by the green ellipse, the green dot is created. Conclusion of the offspring individual's creation is depicted in Figure 2D, where the olive point and ellipse are deleted and the green ellipse is "associated" with the green point. After all offspring individuals have been created and their values according to f have been computed, the parent and offspring population are merged, which is shown in Figure 2E. Finally, the *plus* selection operator is used to discard the individuals with the lowest values of f, which finishes the iteration.



(A) Initial Population

(B) Recombination

(C) Mutation 1

(D) Mutation 2

(E) Evaluation

(F) Selection

FIGURE 2. Illustration of a $(10/5+5)$ ES.

The presentation of the evolutionary strategy given so far raises two questions. First, in the case of real search spaces with the mutations defined in (71), Algorithm 3 does not consider possible constraints imposed on the search space. A solution to this is the incorporation of restriction-handling techniques, like the inclusion of penalty functions, reparation of offspring, multiobjective optimization, etc., and the algorithm has to be adapted accordingly to the technique used (see [**81**] for an overview of several constraint-handling techniques in the context of ES).

The second question concerns the convergence of the algorithm. While theoretical results exist ensuring the almost sure convergence of the iterations (see, e.g., [**102**]), the assumptions used in the statements of such results are usually quite restrictive or require a deep knowledge of the explicit form of the objective function, which leads to difficulties at the moment of the implementation. Despite this, evolutionary algorithms have been tested in a wide set of scenarios, proving to be effective tools for solving optimization problems.

3.7. ES for the optimal dividend bands problem

It is now of interest to see how competitive evolutionary strategies for the numerical determination of optimal band levels are in the present context. Following the considerations from the previous section, in Section 3.8 we will use a $(\mu_{\text{ES}}/\rho_{\text{ES}} + \lambda_{\text{ES}})$ evolutionary algorithm to find the optimal band strategy for three distinct claim distributions: mixtures of Erlang distributions, a shifted-Pareto and a mixture of shifted-Pareto and Erlang distributions. As discussed earlier, only very few instances of explicit non-barrier optimal band strategies have been identified. In what has become a classical example by now, Azcue and Muler [**22**] identified a 2-band strategy for a case with Erlang(2,1) claims. Adding to this, in [**28**] Berdel managed to expand this work by developing an algorithm for identifying non-barrier band strategies in the case of a mixture of Erlang distributions and some more general phase-type distributions. Our selection of mixtures of Erlang distributions for testing the ES was therefore made to compare its efficacy against an established baseline. Further, as can be seen from Section 3.8.1, the lack of explicit formulas for the scale function in the case of Pareto claims imposes the need of numerical approximations to the evaluation of V_π. As shown in Loeffen [**89**], for any choice of parameters, a barrier strategy is the optimal one for a Cramér-Lundberg model with shifted-Pareto claims. The second choice of claim distribution for the present work was then made to test the ES in a numerically-driven situation and test its respective efficacy. Finally, the mixture of Erlang and shifted-Pareto claim distribution was used as a means of testing the algorithm in uncharted territory.

Algorithm 4 displays the ES-algorithm adapted for the dividend-bands optimization, where $\odot$ stands for the element-wise multiplication operator, and $\bar{1}_k$ and $\bar{0}_k$ are the k-dimensional vectors of ones and zeros, respectively.

Algorithm 4 is a $(\mu_{\text{ES}}/\mu_{\text{ES}} + \lambda_{\text{ES}})$-ES based on the basic strategy described in [**31**] for a search in a real unconstrained object space. We chose to use $\mu_{\text{ES}} = \rho_{\text{ES}}$ since this facilitates implementation and, as seen in [**31**] after the study of optimization problems in real spaces, this provides the best performance.

Initialization is carried out in Lines 4 to 6, where the function `initialize` stands for random initialization of the candidate levels. Recombination of the parental population is done in Lines 8 and 9 using the arithmetic mean. In Lines 10, 12 and 13 the call of the function `random_normal`(k) represents the creation of an independent vector of dimension k of standard normal random variables. Lines 14 to 17 show the implementation of the mutation operator, where each parent individual produces one offspring individual using log-normal multiplicative mutations for the exogenous parameters and normal mutations for the object parameters. The coefficients $1/\sqrt{4m-2}$ and $R\cdot\bar{1}_{2m-1}/\sqrt{2\sqrt{4m-2}}$ are learning rates which depend on the dimension of the search space and are based on both theoretical and empirical investigations. After the offspring has been created, repairing is carried out to ensure that

Input : Initial capital u_0, upper bound on number of bands M, number of generations G and variance bound ε and value function $V_\cdot$.
Output: Levels $B^* = (b_0^*, a_1^*, \ldots, b_{M-1}^*)$ of the best M-band strategy.

1 **begin**
2   **for** $m := 1$ **to** M **do**
3     $g := 0$;
4   `initialize(`$\{B_{0,k} \mid k = 1, \ldots, \mu_{\mathrm{ES}}\}$`);`
5     $s_{0,k} := 1.0,\ k = 1, \ldots, \mu_{\mathrm{ES}}$;
6     $\mathscr{P}^{(0)} := \{(B_{0,k}, s_{0,k}, V_{B_{0,k}}(u_0)) \mid k = 1, \ldots, \mu_{\mathrm{ES}}\}$;
7   **repeat**
8       $\overline{s} = \frac{1}{\mu_{\mathrm{ES}}} \sum_{i=1}^{\mu_{\mathrm{ES}}} s_{g,i}$;
9       $\overline{B} = \frac{1}{\mu_{\mathrm{ES}}} \sum_{i=1}^{\mu_{\mathrm{ES}}} B_{g,i}$;
10       $R :=$ `random_normal(`1`);`
11   **for** $l := 1$ **to** λ_{ES} **do**
12         $s_R :=$ `random_normal(`$2m-1$`);`
13         $B_R :=$ `random_normal(`$2m-1$`);`
14         $\tilde{s}'_l :=\ \overline{s} \odot \exp\left(s_R/\sqrt{4m-2} + R \cdot \overline{1}_{2m-1}/\sqrt{2\sqrt{4m-2}}\right)$;
15         $\tilde{B}'_l :=\ \max(\tilde{s}'_l \odot B_R + \overline{B}, 0_{2m-1})$;
16         $\tilde{s}''_l :=$ `sort(`$\tilde{s}'_l$`,order(`$\tilde{B}'_l$`));`
17         $\tilde{B}''_l :=$ `sort(`$\tilde{B}'_l$`);`
18         $\tilde{V}_l := V_{\tilde{B}''_l}(u_0)$;
19   **end**
20       $\mathscr{O}^{(g)} := \{(\tilde{B}''_l, \tilde{s}''_l, \tilde{V}_l) \mid l = 1, \ldots, \lambda_{\mathrm{ES}}\}$;
21       $\mathscr{P}^{(g+1)} :=$ `selection(`$\mathscr{P}^{(g)}$`,`$\mathscr{O}^{(g)}$`,`μ_{ES}`);`
22       $g := g + 1$
23   **until** $g = G$ *or* $\max(s_{g,0}) < \varepsilon$;
24   **end**
25 **end**

ALGORITHM 4: ES-algorithm for optimal dividends

the levels satisfy the condition $0 \leq b_0 < a_1 < \cdots < b_{m-1}$. Line 16 sorts the s parameters according to the increasing order of B, while Line 17 sorts the object parameters in increasing order. Finally, the function in Line 21 performs *plus* selection and outputs the population $\mathscr{P}^{(g+1)} := \{(B_{g+1,k}, s_{g+1,k}, V_{B_{g+1,k}}(u_0)) \mid k = 1, \ldots, \mu_{\mathrm{ES}}\}$ ordered in decreasing order according to the value of V, so when the terminal condition in Line 23 is evaluated, $s_{g,0}$ holds the variances of the levels with the best fit.

Notice that the algorithm requires a value for the initial capital. While in principle this is a technical condition for the evaluation of V_π, caution should be taken: in case the optimal band strategy π^* is finite with levels $b_0^* < a_1^* < \cdots < b_{m-1}^*$, for $b_i^* \leq u \leq a_{i+1}$, any other band strategy π with first $i+1$ bands given by $b_0^* \leq a_1^* < \cdots < b_i^*$ and $u_0 < a_{i+1}$ will satisfy $V_{\pi^*}(u_0) = V_\pi(u_0)$. Therefore, unless we can ensure $b_{m-1}^* \leq u_0$, any such π will be the output of any optimization algorithm for which the initial capital is fixed. Following Lemma 3.3.1 in [**105**], the inequality $b_{m-1}^* \leq u_0$ can be guaranteed by taking $u_0 = p\lambda/(\delta(\lambda + \delta))$, which is the value that we use for all the iterations of the algorithm.

At this point, it is worthwhile to mention a key difference between the search method employed by this algorithm and the iterative algorithm discussed in, for example, [**22**] or [**23**]: given that the dimension of the search space has to be kept constant during the procedure, one has to fix in advance the number of bands for which the ES will try to identify the optimal levels. By observing that one can "collapse" levels in a band strategy, the n-bands strategies can be thought of as m-band strategies for $n \leq m$ and hence the algorithm would identify at once the best levels for all n-band strategies for $n \leq m$. If the optimal strategy is finite, one could then set m large enough and use the ES to find the optimal levels. However, this approach requires the evaluation of V_π for several bands and as explained in Section 3.8.1 below, this is not efficient. Hence, a more efficient approach is instead to consecutively compute the the best 1-band, 2-band, 3-band, etc. strategies until collapsing of the levels is observed and then verify, by means of the HJB equation, that the proposed solution is in fact the optimal band strategy. Finally, efficiency is improved by skipping the search for the optimal 1-band strategy and set $b_0 := \sup\{x \geq 0 \mid W_\delta'(x) = \inf_{y\geq 0} W_\delta'(y)\}$ in all searches.

3.8. Numerical results

We evaluate the performance of the procedures shown in the two previous sections by finding optimal band strategies for three study cases: claims distributed as a mixture of Erlang distributions, the case for a pure (shifted) Pareto distribution and a mixture between Erlang and Pareto distributions. The mixture of Erlang distributions is chosen because there are already explicit results available (see [**22, 28**]) so that we can benchmark our algorithms. Given that no explicit expressions exist for the scale function when the claims follow a Pareto distribution, the second case was chosen to test the algorithms in a purely numerical situation (and in the case of a Pareto distribution, it is known that the optimal strategy is a barrier strategy, see [**89**]). Finally, the mixture of Erlang and Pareto distributions was carried out to study the problem in a new context.

3.8.1. Objective function evaluation. In the case where f_Y comes from a mixture of Erlang distributions, the Laplace transform $\widehat{W_\delta}$ of the scale function is given in terms of a rational function, so explicit expressions in terms of the roots of the Lundberg equation can be found for W_δ and f_D. These expressions are then used for computing the value of V_π.

In the other two cases, numerical inversion of $\widehat{W_\delta}$ and (60) have to be carried out to find the values of f_D. Since the evaluation of this function is needed at several points, we opted for using a piece-wise linear approximation for W_δ, W_δ', f_{D^0} and its partial derivatives. The approximation was carried out in the following way: from the remarks of Section 3.7, it can easily be seen that in the case where the optimal band strategy is finite, it is enough to restrict the domain of f_D and W_δ to $[0, p\lambda/(\delta(\lambda+\delta))]^3$ and $[0, p\lambda/(\delta(\lambda+\delta))]$ respectively in order to find the optimal band levels. Hence, the approximation was done by evaluating 10,000 equidistant points in the interval $[0, p\lambda/(\delta(\lambda+\delta))]$ (including boundaries) and linear interpolation in between. The functions f_{D^0} and f_D were then computed using Equations (60) and (62).

The integrals appearing in (58), (65), (66) and (67) were mainly evaluated through numerical methods by using the `numpy`, `scipy` and `mpmath` libraries for Python 3. Exceptions to this were the Erlang distributions appearing in example (a) in Case I, where the equations were simple enough to be computed explicitly, and examples (b) and (c) of the same case for the evaluation of the ES. The reason for doing this only for the ES instead of for both methods was in part due to the poor convergence rate of the ES when evaluating the integrals numerically, and the computational effort that was necessary to obtain explicit forms for equations

(65), (66) and (67), necessary for solving the gradient (see Section 3.8.2 for further explanation into this issue).

Equations (68) and (69) were solved by means of the MINPACK's hybrd and hybrj algorithms implemented in the `scipy` library through the `fsolve` function. Although there are no explicit results supporting the fact, the examples found in the literature indicate that the values for the different band values $b_1, b_2, \ldots$ are close to local minima of W_δ' found after the optimal b_0. This fact helped providing starting points for both algorithms.

Finally, we would like to comment on our selection of μ_{ES} and λ_{ES}. While the choice of these parameters was mainly heuristic, the following two points served as a rule of thumb for our choices:

- In our experiments, a small or a very large value of μ_{ES} provided slower convergence, since in the first case the mutation step was based on very few cases, entailing little diversity in the offspring; while in the second case the outcome of the recombination step was negatively affected by the worst outcomes. For most of our trials, we assigned small multiples of 10 to μ_{ES} and just kept the first one that produced suitable results.
- Most of the steps in the algorithm can be implemented in a vectorized way, which justifies our selection of the `numpy` and `scipy` libraries for developing the tasks. However, since the band levels needed to be individually ordered in the offspring to be able to evaluate the value function, one cannot avoid an implicit for-loop in Lines 16 and 17 of Algorithm 4, which impedes vectorization of this step and significantly increases the complexity. Since, on the other hand, a large value of λ_{ES} improves diversity in each generation and decreases the possibility of reaching local extrema, one faces a trade-off in terms of performance when choosing its value. In our experiments, we took λ_{ES} to be of an order similar to μ_{ES}, which produced satisfactory results.

We now apply both numerical procedures introduced in this paper to the concrete examples. Below, the reporting times mean the clock time used to produce the results and do not include the time used for verification of the solution in an interval through the HJB equation, which is the same for both procedures.

3.8.2. Case I: Erlang mixture claims. The following three examples are considered (for which we know the explicit result already from [**22**] and [**28**] for the first two):

(a) An $\mathrm{Erlang}(2, 1)$ distribution with parameters $\lambda = 10$, $\delta = 0.1$ and $\eta = 0.07$.
(b) A mixture of the distributions $\mathrm{Erlang}(2, 10)$, $\mathrm{Erlang}(3, 1)$ and $\mathrm{Erlang}(4, 0.1)$ with weights $0.025, 0.225$ and 0.75 respectively, and parameters $\lambda = 1$, $\delta = 0.1$ and $\eta = 0.405$.
(c) A mixture of the distributions $\mathrm{Erlang}(2, 10)$, $\mathrm{Erlang}(3, 1.06775)$, $\mathrm{Erlang}(4, 0.2325)$ and $\mathrm{Erlang}(5, 0.05)$ with weights $0.005, 0.045, 0.225$ and 0.725 respectively, and parameters $\lambda = 1$, $\delta = 0.1$ and $\eta = 0.4$.

For the first two distributions, a $(30/30 + 60)$-ES was used in both cases, with bound on the variance equal to 0.01. For reasons that will be explained later, a combination of two ES's (a $(30/30 + 60)$-ES and a $(1/1 + 1)$-ES) was used for the third distribution, with same bound in the variance. The number of iterations vary from distribution to distribution.

For the Erlang(2,1) distribution, Table 1 shows that we indeed find the optimal two-band strategy established in [**22**].

One can observe that the gradient-based approach is very fast, while the ES algorithm takes considerably longer time, but also arrives at the correct solution.

Table 2 shows the results for the first mixture of distributions, where a 3-band strategy is optimal.

	Time	Iterations	b_0	a_1	b_1
Evolutionary Strategy	~ 1000s	1000	0	1.8064	10.2158
Gradient-Based	< 1s	–	0	1.8030	10.2161

TABLE 1. Results for the $\mathrm{Erlang}(2, 1)$ distribution.

	Time	Iterations	b_0	a_1	b_1	a_2	b_2
Evolutionary Strategy	~ 1h	5000	0.2617	0.4668	3.5249	25.7390	34.7857
Gradient	~ 1h	–	0.2615	1.5230	3.5246	25.5763	34.7696

TABLE 2. Results for the first mixture of Erlang distributions.

Several remarks for this example are in order. First, there is a clear discrepancy between the a_1-values obtained by the two methods. Comparing the value functions of both strategies shows that the strategy found by the ES provides higher values. However, the difference between the value functions is of the order of 10^{-4} while the norm of the gradient at that point is of the order 10^{-5} and the iterations do not reduce this value significantly, which shows why the gradient method has an early stop. Note that for this example, Berdel [**28**] already studied the optimal dividend strategy, and her results are very similar to the ones in Table 2, with only the values of a_1 and a_2 differing. The ES parameters above provide a larger value function, but again the difference is only of order 10^{-4}. The explanation is that the step size in [**28**] for solving the inf and sup in Lines 7 and 8 in Algorithm 1 was set to be 10^{-4}, whereas a smaller step size would have been needed to arrive at the above result.

As stated before, for the evolutionary strategy, the integrals in equations (58), (65), (66) and (67) were not evaluated numerically but instead were computed exactly, by means of symbolical calculus in Mathematica[1]. Now, for $k = 2$, and $u \geq b_2$, (58) can be more explicitly

[1]The reason for using Mathematica for this situation was the easiness with which symbolic expressions can be handled in the Wolfram Language. While existing libraries in Python such as SymPy can perform similar tasks, the performance of Mathematica in effecting this particular set of calculations outweighed the potential advantage of using the same programming language throughout.

written as

$$
\begin{aligned}
V_\pi(u) &= u - b_2 + \frac{W_\delta(h_2)}{W'_\delta(h_2)} \\
&+ \int_0^{b_0} f_D(a_2 - y, h_2, h_2) \frac{W_\delta(y)}{W'_\delta(b_0)}\, dy \\
&+ \int_{b_0}^{a_1} f_D(a_2 - y, h_2, h_2) \left(y - b_0 + \frac{W_\delta(b_0)}{W'_\delta(b_0)} \right) dy \\
&+ \int_{a_1}^{b_1} f_D(a_2 - y, h_2, h_2) \frac{W_\delta(y - a_1)}{W'_\delta(h_1)}\, dy \\
&+ \int_{a_1}^{b_1} f_D(a_2 - y, h_2, h_2) \int_0^{b_0} f_D(a_1 - z, y - a_1, h_1) \frac{W_\delta(z)}{W'_\delta(b_0)}\, dz\, dy \\
&+ \int_{a_1}^{b_1} f_D(a_2 - y, h_2, h_2) \int_0^{b_0} f_D(a_1 - z, y - a_1, h_1) \left(y - b_0 + \frac{W_\delta(b_0)}{W'_\delta(b_0)} \right) dz\, dy \\
&+ \int_{b_1}^{a_2} f_D(a_2 - y, h_2, h_2) \left(y - b_1 + V_{\pi^1}(b_1) \right) dy
\end{aligned}
$$

where $h_i = b_i - a_i$ and V_{π^1} is the 2-band strategy obtained after deleting the last band from V_π. For mixtures of Erlang distributions, the scale function can be written as a linear combination of complex exponential functions with as many terms as roots of the Lundberg equation, assuming all of them are different. For the present case, there are 10 different roots. By means of formula (59), it follows that $f_D(y, u, b)$ can be written as a sum of approximately 90 different terms involving y with coefficients dependent on u and b. Following this line of thought, the single integrals from the paragraph above have, in rough terms, 900 terms, while the first double integral has around 810,000 (in theory, further reductions that decrease these numbers considerably could in principle be possible, but the amount of terms implies that the human or computational effort for carrying out such operations is beyond reason). The computational effort for explicitly computing the integrals above was of around 1 hour, which implies that the total time for the ES was of around 2 hours, which doubled the computational time of the gradient method, but provided a slightly more accurate result. Although this procedure could also be carried out to test the computational time of the gradient method (which by virtue of the other cases would be expected to be smaller), we observe that the computational time in for obtaining explicit expressions for (65) and (66) would be at least doubled, matching the current time of the ES.

Finally, using the intuition that the number of bands in the optimal strategy is related to the number of modes of the claim distribution, we were interested to establish a case where a 4-band strategy is optimal, and the second mixture of Erlang distributions indeed leads to such an optimal 4-band strategy. The resulting optimal bands are given in Table 3.

Time	Iterations	b_0	a_1	b_1	a_2	b_2	a_3	b_3
$\sim$ 3h	6000	0.2562,	1.0543	3.1988	10.6647	19.5499	127.9288	171.6044

TABLE 3. Results for the second mixture of Erlang distributions.

These values were computed using only the ES technique. As before, the procedure was carried out in two steps, first using a $(30/30 + 60)$-ES for computing the values of the first three bands and consecutively using these values to reduce the problem to a two-dimensional optimization exercise, where a $(1/1 + 1)$-ES was used for computing the final two values.

The reported time is for the combination of both procedures. The reason for proceeding in a two-step fashion was due to the observation that, as shown by these examples, the optimal values of of the b_i's are usually located in the vicinity of the local minima of W'_δ. The three smallest local minima are found in the interval $(0, 20)$, which does not present any numerical complication. However, the last one is found at 172.7545, which due to the nature of the scale function, produces exponentials with very large exponents at the moment of evaluating the value function, creating considerable numerical instabilities. To solve this issue, the value function and evolutionary strategy were re-implemented using arbitrary-precision floating-point-arithmetic, which decreased the speed at the moment of evaluating the value function. Since the $(1/1+1)$-ES is the strategy that requires the least evaluations of the objective function, it was chosen for obtaining the final values of a_3 and b_3. Figure 3 illustrates that this 4-band strategy is indeed the optimal strategy. Concretely, Figures 3A,3B and 3C show that none of the first three strategies (with 1, 2 or 3 bands) is optimal, as the HJB equation attains positive values. Figure 3D shows that this does not happen for the 4-band strategy and, moreover, Figure 3E reveals that whenever the derivative exists, it is at least 1, so the solution is optimal.

3.8.3. Case II: Pure Pareto claims. Following [**89**], the optimal strategy will be a barrier strategy when claims have a shifted Pareto distribution with density function

$$f_Y(y) = \alpha x_0^{-1}(1 + x_0^{-1}y)^{-\alpha-1}, \quad y > 0,$$

and Laplace transform

$$\hat{f}_Y(s) = \alpha x_0^\alpha s^\alpha e^{sx_0}\Gamma(-\alpha, sx_0), \quad s > 0,$$

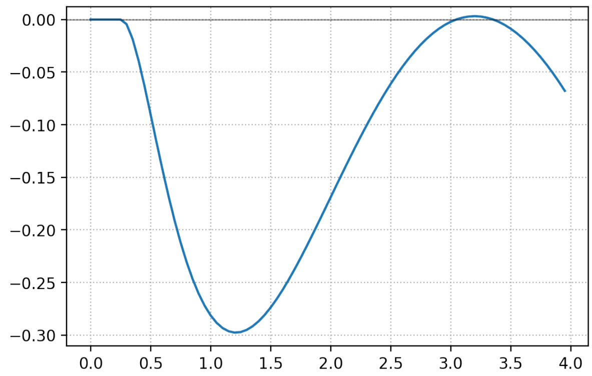
with Γ the upper incomplete Gamma function and $\alpha, x_0 > 0$. For the case at hand, we considered $x_0 = 1$ and $\alpha = 1.5$, so that the claims have finite expectation and infinite variance. Moreover, the parameters of the Cramér-Lundberg process were taken to be $\lambda = 10$, $\delta = 0.1$ and $\eta = 0.1$. The derivatives of the scale function were computed through their Laplace transforms and all of these were inverted using the de Hoog, Knight and Stones algorithm implemented in the library `mpmath`. The results are given in Table 4.

	Time	Iterations	b_0
Evolutionary Strategy	~ 1000s	100	2.71036
Gradient-Based	< 1s	-	2.71036

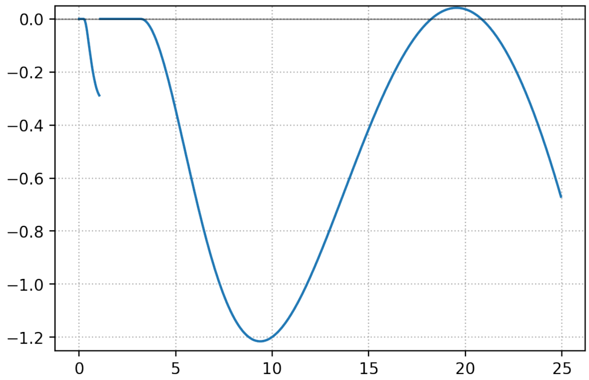
TABLE 4. Results for the Pareto distribution.

Indeed, one arrives at an optimal barrier strategy, where for the evolutionary algorithm we only used 100 iterations to arrive at a running time that is comparable to the ones of the Erlang case, and the result is already well-aligned with the one of the gradient-based method.

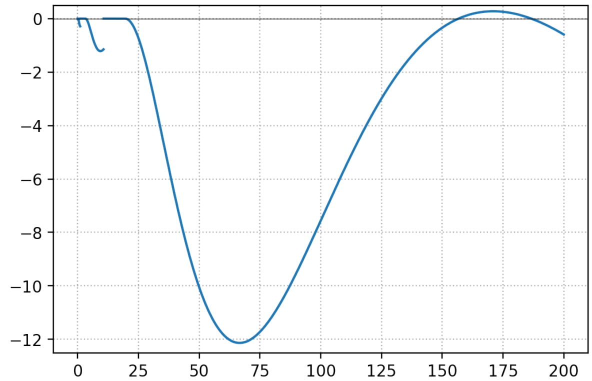
3.8.4. Case III: Erlang and Pareto mixture claims. Finally, let us consider a mixture of an $\mathrm{Erlang}(2, 1)$ distribution and a shifted Pareto distribution ($\alpha = 1.5$, $x_0 = 1$) with weights 0.8 and 0.2 respectively. The parameters of the Pareto distribution were chosen to match the mean of the Erlang component, while the weights were chosen to avoid a monotonicity of W'_δ. The other parameters are again $\lambda = 1$, $\delta = 0.1$ and $\eta = 0.1$. A $(150/150 + 100)$-ES was used and a 2-band strategy was found to be optimal. The results are shown in Table 5.



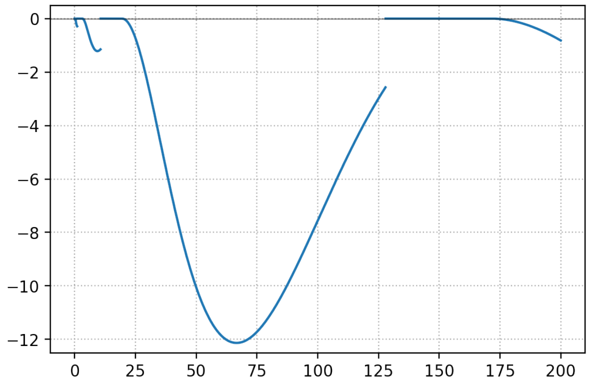
(A) HJB equation for the 1-band strategy



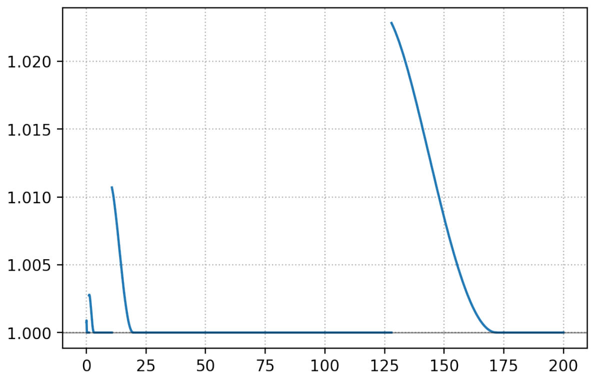
(B) HJB equation for the 2-band strategy



(C) HJB equation for the 3-band strategy



(D) HJB equation for the 4-band strategy



(E) Derivative of the value function

FIGURE 3. Plots of the l.h.s. of the HJB equation for the strategy with 1, 2, 3, and 4 bands based on the a_i's and b_i's from Table 3 as well as the derivative of the value function in the points where it exists.

	Time	Iterations	b_0	a_1	b_1
Evolutionary Strategy	~ 8h	1000	0	0.1524	3.5115
Gradient-Based	~ 1h	-	0	0.0053	3.8877

TABLE 5. Results for a mixture of an Erlang and a Pareto distribution.

In this case, the discrepancy between the results is more significant than in the other cases, with the gradient method providing a better solution. Due to the time it took for each evaluation, the ES was stopped after 1000 iterations, which meant that convergence was not fully achieved. The results from the gradient method also help to explain the poor performance of the ES: observe that the value of a_1 is rather close to zero, with an order of magnitude of 10^{-3}. At the initialization of the algorithm, there is no knowledge of what the final order of magnitude will be, and during this experiment the initial values for the exogenous parameters were set to be 1, so, due to the projection into the zero for negative values, it takes many evaluations before the desired order of magnitude is achieved. While experiments with the other cases showed that the performance of the ES in this case could have probably been improved by, for example, reducing the values of μ_{ES} and λ_{ES}, we preferred to not investigate further into this, since the evaluation time was high and the gradient method had already provided satisfactory results.

3.9. Conclusions

In this paper we added two numerical alternatives to identify optimal dividend bands in the classical optimal dividend problem of risk theory. We illustrate that both of them are efficient, and their scope and applicability goes beyond the one of the previously discussed methods in the literature. The gradient-based method can be particularly efficient. The second algorithm based on evolutionary strategies is satisfactory as well, and whereas in terms of computation times it can not compete with the gradient-based method for the complexity of this concrete problem, its range of applicability is even wider. In fact, ES algorithms can be an interesting competitor whenever an objective function can be efficiently evaluated, and it is known to work particularly well for higher-dimensional optimization problems, in which case the gradient alternative can be hard to explicitly compute or implement. We rederived optimal bands for some known cases, established new ones and also derived results for cases that were beyond the scope of previously available methods.

The focus of this paper was on optimal dividend strategies in the Cramér-Lundberg model. However, since the equations used to derive the necessary functions for these two algorithms were obtained by means of Gerber-Shiu functions, one can in principle easily extend the range of applications to the case where a diffusion is added to the surplus process or even to the case where the surplus process is modelled by a generally spectrally-negative Lévy process satisfying the safety loading condition. Since evolutionary algorithms can be applied in rather general settings, it will also be interesting to see in future research other applications of this method in risk theory, particularly also in optimization problems with constraints, which may be handled with an introduction of a penalty term in the objective function (see e.g. [**81**]).

CHAPTER 4

Dividend corridors and a ruin constraint

This chapter is based on the following article:

H. Albrecher, B. Garcia Flores and C. Hipp. Dividend corridors and a ruin constraint. *Preprint submitted for publication.*

Abstract. We propose a new class of dividend payment strategies for which one can easily control an infinite-time-horizon ruin probability constraint for an insurance company. When the risk process evolves as a spectrally negative Lévy process, we investigate analytical properties of these strategies and propose two numerical methods for finding explicit expressions for the optimal parameters. Numerical experiments show that the performance of these strategies is outstanding and, in some cases, even comparable to the overall-unconstrained optimal dividend strategy to maximize expected aggregate discounted dividend payments, despite the ruin constraint.

4.1. Introduction

Consider an insurance company whose surplus process evolves according to a spectrally-negative Lévy process. We assume that this process satisfies the safety loading condition, reflecting the idea that, in expectation, the company charges more premiums than the amount of claims to be paid. Under this assumption, however, the process also possesses the unrealistic property that, with probability one, it will diverge to infinity. One way to avoid this issue is to consider dividend payments to shareholders. Since the introduction of this idea in the seminal work of [**54**], there have been a lot of research activities on establishing optimal strategies for distributing dividends under various objective functions and constraints. For instance, it was established that for the maximization of the expected sum of discounted dividend payments until ruin a *band strategy* is often optimal (see e.g., [**63**], [**89**], [**22**] and [**21**] as well as [**18**] and [**10**] for surveys).

While band strategies maximize expected aggregate dividend payments for a rather general set of assumptions, they also lead to the undesirable property that, with probability one, the surplus process will eventually become negative, i.e., the company will get ruined. Hence, while the introduction of dividend payments makes the model more realistic, the optimal solution is typically unacceptable in practice. In response to this, a growing body of literature examined the trade-off between profitability and safety (avoiding or delaying ruin). [**111**] and [**90**] examined the dividend problem with a penalty for early ruin, see also [**86**]. For a discrete-time model, [**73**] was the first to approach the optimal dividend problem under a ruin constraint, which turns out challenging in view of the resulting *time-inconsistency* of the stochastic control problem. [**67**] studied the problem of optimizing dividend payments in finite time with a constraint on the probability of ruin for the case of the diffusion, providing a solution in terms of a complicated set of differential equations. Similarly, [**72**] provided a solution to the case where the constraint is a bound on the Laplace transform of the time of ruin, under the assumption that the density of the Lévy measure is completely monotone. Recently, [**109**] used a game-theoretic approach to reformulate the idea of optimality and provided a solution to the case of the diffusion.

Rather than directly addressing the control problem of maximizing expected discounted dividend payments under a ruin constraint whose general solution seems out of reach, [**74, 76**] started to study particular candidate strategies with an intuitive structure that allow a balancing of profitability and safety with a bottom-up approach, see also [**75**]. The contribution of the present paper is a considerable extension and deepening of the latter approach. We define a sort of corridor dividend strategies for which, using scale functions and fluctuation theory, the ruin probability can be easily controlled and at the same time the expected dividend payments can be maximized locally. While we do not prove optimality of such strategies for the general dividend problem under a ruin constraint, the numerical illustrations at the end of the manuscript show that this kind of strategies perform exceptionally well, sometimes even leading to comparable efficiency to the best overall un-constrained (band) strategy, but respecting the pre-given ruin constraint. In order to make the numerical optimization of the involved parameters work, we implement and adapt two numerical schemes to the present problem: a recursive approach inspired by Newtonian optimization techniques and an evolutionary algorithm. The resulting optimized strategies can serve as new benchmarks for both intuition and numerics for the general problem of maximizing dividends under a ruin constraint. For instance, when we apply corridor strategies to the diffusion case and finite time horizon problem studied in [**67**] and pay the remaining surplus as a final dividend lump sum at the end of the time horizon as done in that paper, the optimal corridor strategy in fact outperforms the numerical solution given in [**67**] for the same problem, cf. Section 4.6.

The rest of the paper is organized as follows: in Section 4.2 we establish the basic assumptions for the surplus model and introduce some notation. Section 4.3 introduces the corridor payment strategies and derives formulas to compute the value associated with them. Section 4.4 examines analytical properties of the value function associated with the strategies. Since the value function eventually needs to be evaluated and optimal parameters need to be determined, Section 4.5 introduces several numerical techniques that can be used for this purpose. Section 4.6 presents the numerical results for a selected number of surplus processes commonly studied in the literature. Finally, Section 4.7 concludes and provides some directions for future research.

4.2. The model

Consider a spectrally negative Lévy risk process $(C_t)_{t\geq 0}$ for the surplus process of an insurance portfolio with initial surplus level $C_0 = u$. In the following we will formulate the results first for the general case and then go into more detail for two special cases of interest, namely the case of a diffusion approximation

(72) $$C_t = u + \mu t + \sigma B_t, \quad t \geq 0,$$

where $\mu > 0$ is a constant drift, $\sigma > 0$ and $(B_t)_{t\geq 0}$ denotes a standard Brownian motion, and the Cramér-Lundberg process

(73) $$C_t = u + c\,t - \sum_{i=1}^{N_t} X_i, \quad t \geq 0,$$

where $(N_t)_{t\geq 0}$ is a homogeneous Poisson process with rate $\lambda > 0$, X_i are the individual claim sizes modelled by i.i.d. random variables with cumulative distribution function F_X and finite mean, and $c > \lambda E(X_i)$ is the premium collected per time unit. Denote by $\tau := \inf\{t \geq 0 :$

$C_t < 0\}$ the time of the ruin, by $\psi(u) := \mathbb{P}(\tau < \infty)$ the ruin probability of this risk process and by

$$\phi(u) = 1 - \psi(u)$$

the corresponding survival probability. Let $\kappa(\theta) := \log \mathbb{E}e^{\theta(C_1 - C_0)}$ denote the Laplace exponent of the Lévy process, which has the form

$$\kappa(\theta) = -a\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{(-\infty,0)} (e^{\theta x} - 1 - \theta x 1_{\{x>-1\}})\Pi(dx)$$

with Π the Lévy measure and $\Phi_\delta = \psi^{-1}(\delta) > 0$ (see for instance [**83**]). The various results obtained throughout the manuscript will be expressed in terms of the scale function of C, which for $x \geq 0$ and any $\delta \geq 0$ is defined as the function $W_\delta(x)$ satisfying the identity

$$\int_0^\infty e^{-\theta x} W_\delta(x)\, dx = \frac{1}{\kappa(\theta) - \delta}, \quad \theta > \Phi_\delta.$$

We also define $W_\delta(x) = 0$ whenever $x < 0$. Assuming that $\kappa'(0)$ is finite, it is then well-known that

$$\phi(u) = \kappa'(0) W_0(u)$$

(see e.g. [**83**] or [**17**, Ch.IX]). For the sake of convenience, we will assume that Π accepts a continuous density, so that $\Pi(dx) = f_\Pi(x)dx$ and that this density is sufficiently smooth to ensure that $W_\delta \in C^2(0, \infty)$.

Dividends are now paid out according to a strategy $D = (D_t)_{t\geq 0}$, where D_t represents the aggregate dividends up to time t. The surplus process after dividends is given by

$$C_t^D = C_t - D_t$$

and the expected value of the aggregate discounted dividend payments until ruin are given by

$$V^D(u) = \mathbb{E}\left(\int_0^{\tau^D} e^{-\delta t} dD_t\right)$$

where

$$\tau^D := \inf\{t > 0 : C_t^D < 0\}$$

is the time of the ruin of the resulting surplus process with dividends.

Consider now the following dividend payment strategy: For a fixed $n \in \mathbb{N}$, there is a sequence of surplus levels $a_1, a_2, a_3, \ldots, a_n$. Assume for the moment that $u < a_1$. When the risk process reaches a_i ($i = 1, \ldots, n$) for the first time (which we denote by τ_i), there is a lump sum dividend payment of $a_i - b_i$ down to a barrier level b_i. Then continuous dividend payments start according to a horizontal barrier strategy with barrier b_i until the surplus process goes below the lower limit $l_i \leq b_i$ for the first time (denoted by $\tau_i^d \geq \tau_i$), at which point the barrier in b_i is dissolved. Dividend payments only continue later in case the surplus process reaches the level $a_{i+1} > l_i$ before ruin, which happens at time τ_{i+1}, given that $\tau_{i+1} < \tau^D$. In that case the next lump sum $a_{i+1} - b_{i+1}$ is paid, followed by dividends according to a horizontal barrier strategy at b_{i+1} until the process goes below l_{i+1} etc. Once the last dividend barrier b_n is dissolved, the surplus process survives according to the classical survival probability (without dividends) with initial surplus level $C^D_{\tau_n^d}$. Note that this formulation of the strategy includes the case of a pure lump sum payment (in case $b_i = l_i$ and infinite variation) as well as the case of a pure 'horizontal dividend corridor' without a lump sum payment at the beginning ($a_i = b_i$), and we will be looking for the optimal values of $0 \leq l_i \leq b_i \leq a_i$, $i = 1, \ldots, n$. Figure 1 depicts a sample path of such a strategy for a Cramér-Lundberg process, in which ruin occurs before a_3 is reached.

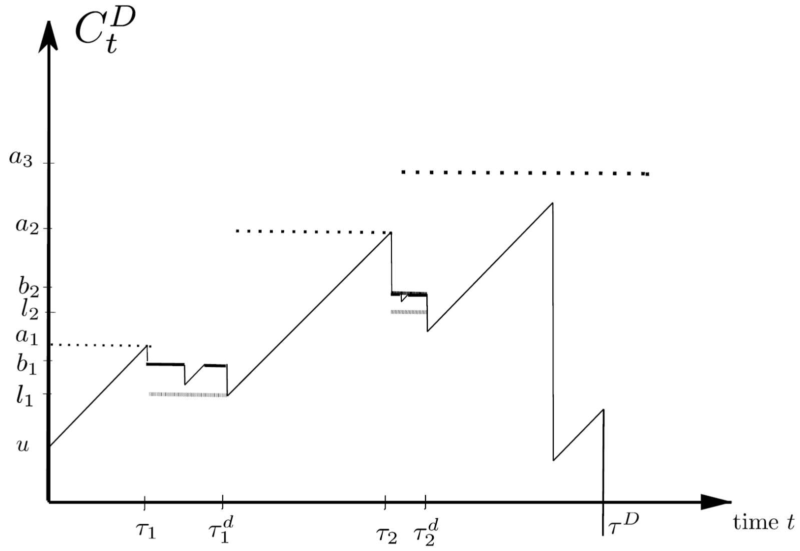


FIGURE 1. A sample path for a Cramér-Lundberg process with a dividend strategy as described above

For a fixed set of levels, we may allow $u > a_1$ by making a lump-sum payment down to b_1 and proceed as described above.

4.3. Some results

Denote by $\phi^D(u) := \mathbb{P}(\tau^D = \infty)$ the survival probability of the resulting risk process, and recall that $\phi(u)$ is the classical survival probability of the risk process without any dividend payments.

THEOREM 4.3.1. *For $u \leq a_1$, we have*

$$\phi^D(u) = \kappa'(0)\, W_0(u) \prod_{k=1}^{n} A(a_k, b_k, l_k) \tag{74}$$

with

$$A(a_k, b_k, l_k) := \frac{\mathbb{E}(W_0(C^D_{\tau^d_k}))}{W_0(a_k)}\,.$$

Note that $C^D_{\tau^d_k}$ is the surplus value at the time of the first undershoot of level l_k after paying dividends at barrier b_k, which occurs at the stopping time τ^d_k. We naturally have $\phi(x) = 0$ for $x < 0$, i.e. if the undershoot at the time of dissolving the k-th corridor leads to a negative surplus value, the company is ruined.

PROOF. A simple iterative application of the strong Markov property of C gives

$$\phi^D(u) = \frac{\phi(u)}{\phi(a_1)} \left(\prod_{k=1}^{n-1} \mathbb{P}(\tau_{k+1} < \tau^D | \tau_k < \tau^D) \right) \mathbb{E}(\phi(C^D_{\tau^d_n})),$$

which, using the strong Markov property again, can also be expressed as

$$\begin{aligned}\phi^D(u) &= \frac{\phi(u)}{\phi(a_1)}\left(\prod_{k=1}^{n-1}\frac{\mathbb{E}(\phi(C^D_{\tau^d_k}))}{\phi(a_{k+1})}\right)\mathbb{E}(\phi(C^D_{\tau^d_n}))\\ &= \phi(u)\prod_{k=1}^{n}\frac{\mathbb{E}(\phi(C^D_{\tau^d_k}))}{\phi(a_k)}.\end{aligned}$$

Note that $\phi(u)/\phi(a_1) = W_0(u)/W_0(a_1)$ is the probability that the surplus process C_t reaches surplus level a_1 before ruin, when starting at a lower surplus level $u < a_1$. We can hence rewrite the above expression with scale functions as

$$\phi^D(u) = \kappa'(0)\,W_0(u)\prod_{k=1}^{n}\frac{\mathbb{E}(W_0(C^D_{\tau^d_k}))}{W_0(a_k)},$$

establishing the result. □

While simple, Equation (74) expresses the probability of ruin implicitly in terms of expectations. We would like to obtain more formulas expressions for A, for which we make use of the concept of Gerber-Shiu measures (c.f. [**83**, Ch.X]). Recall the Gerber-Shiu measure of the process, K^δ, which for any $\omega : \mathbb{R}^2 \to [0,\infty)$ such that $\omega(0,\cdot) = 0$, allows us to write

(75) $$\mathbb{E}(e^{-\delta\tau}\omega(-C_\tau, C_{\tau-})) = \int_{(0,\infty)^2}\omega(y,z)K^\delta(dy,dz)$$

with C_τ being the severity of ruin and $C_{\tau-}$ the surplus just before ruin. An explicit expression for K^δ can be given in terms of the Lévy measure and the scale function of the process, i.e.,

(76) $$K^\delta(dy,dz) = (e^{-\Phi_\delta z}W_\delta(u) - W_\delta(u-z))f_\Pi(-y-z)dydz.$$

Similarly, the discounted probability of ruin by creeping can be computed through the formula

(77) $$\mathbb{E}(e^{-\delta\tau}1_{\{C_\tau = C_{\tau-} = 0\}}) = \frac{\sigma^2}{2}\left(W'_\delta(u) - \Phi_\delta W_\delta(u)\right),$$

where the right-hand side is understood as zero whenever $\sigma = 0$. Expectations of discounted penalties of the form $\psi(u) := \mathbb{E}(e^{-\delta\tau}g(C_\tau))$ can therefore be evaluated through equations (76) and (77) for any function g. This is almost what we need, however, to obtain an explicit form for A, we need to compute this expectation assuming dividends have been paid according to a barrier strategy. Denoting by C^b_τ the severity of ruin after dividends have been paid according to a barrier strategy at level b, this means that we require to compute expectations of the form $\psi(u;b) := \mathbb{E}(e^{-\delta\tau}g(C^b_\tau))$. Luckily, this can be easily computed through the dividends-penalty identity (cf. [**64**]),

(78) $$\psi(u;b) = \psi(u) - \frac{W_\delta(u)}{W'_\delta(b)}\psi'(b).$$

Using all these equations, we can reach an explicit expression for the function $A(a,b,l)$.

Proposition 4.3.2. *The function $A(a,b,l)$ can be written as*

$$\begin{aligned}A(a,b,l) &= \frac{\sigma^2}{2}\frac{W_0(l)}{W_0(a)}\left(W'_0(b-l) - \frac{W_0(b-l)W''_0(b-l)}{W'_0(b-l)}\right)\\ &\quad + \int_0^l\int_0^\infty \frac{W_0(l-y)}{W_0(a)}\left(\frac{W_0(b-l)W'_0(b-l-z)}{W'_0(b-l)} - W_0(b-l-z)\right)f_\Pi(-y-z)dzdy.\end{aligned}$$

PROOF. Observe that $C^D_{\tau^d_k} - l_k$ is equal in distribution to the severity of ruin after dividends have been paid according to a barrier strategy at level $b_k - l_k$ of a process with initial capital $b_k - l_k$, so Equations (76), (77) and (78) fully characterize its distribution. Hence, by setting $g = W_0$ in the definition of ψ and combining (76), (77) and (78) one obtains

$$\begin{aligned} A(a,b,l) = {} & \frac{\sigma^2}{2}\frac{W_0(l)}{W_0(a)}\left(W_0'(b-l) - \frac{W_0(b-l)W_0''(b-l)}{W_0'(b-l)}\right) \\ & + \int_0^l\int_0^\infty \frac{W_0(l-y)}{W_0(a)}\left(\frac{W_0(b-l)W_0'(b-l-z)}{W_0'(b-l)} - W_0(b-l-z)\right) f_\Pi(-y-z)dzdy, \end{aligned}$$

where we have omitted the dependence on the index k for simplicity of exposition. □

Let us now turn to the expected value of the sum of the discounted dividend payments. Recall from classical risk theory that the expected discounted dividend payments according to a horizontal dividend barrier strategy at b when starting at an initial surplus level $u < b$ is simply given by $W_\delta(u)/W_\delta'(b)$. This quantity will be a building block of our more complex dividend payment strategy in this paper.

THEOREM 4.3.3. *For $u \le a_1$, the value function $V^D(u)$ of the expected discounted dividend payments can be written as*

$$V^D(u) = W_\delta(u)\sum_{k=1}^{n} B(a_k,b_k,l_k)\prod_{i=1}^{k-1} G(a_i,b_i,l_i) \tag{79}$$

with

$$B(a,b,l) = \frac{a-b+W_\delta(b-l)/W_\delta'(b-l)}{W_\delta(a)}$$

and

$$\begin{aligned} G(a,b,l) = {} & \frac{\sigma^2}{2}\frac{W_\delta(l)}{W_\delta(a)}\left(W_\delta'(b-l) - \frac{W_\delta(b-l)W_\delta''(b-l)}{W_\delta'(b-l)}\right) \\ & + \int_0^l\int_0^\infty \frac{W_\delta(l-y)}{W_\delta(a)}\left(\frac{W_\delta(b-l)W_\delta'(b-l-z)}{W_\delta'(b-l)} - W_\delta(b-l-z)\right) f_\Pi(-y-z)dzdy. \end{aligned}$$

PROOF. Consider the scenario in which the $(k-1)$-th corridor has just been dissolved. Once we reach a_k, there will be a lump sum payment $a_k - b_k$ and then dividend payments will start according to a horizontal barrier strategy at barrier b_k until the k-th corridor is dissolved. Denote by D_k the expected present value of all dividend payments at the k-th corridor. By construction of the dividend payment strategy, D_k can also be viewed as the expected discounted dividends according to a barrier strategy collected until ruin in a risk model with initial surplus level $b_k - l_k$ and also barrier level $b_k - l_k$, since the event of ruin in that model will exactly correspond to C^D undershooting l_k for the first time. Correspondingly, $D_k = W_\delta(b_k - l_k)/W_\delta'(b_k - l_k)$. Using the strong Markov property, we can write

$$\begin{aligned} V^D(u) = {} & \frac{W_\delta(u)}{W_\delta(a_1)}\Big[(a_1-b_1) + D_1 + \mathbb{E}(e^{-\delta(\tau_2-\tau_1)}1_{\{\tau_2<\tau^D\}})\Big[(a_2-b_2)+D_2 \\ & + \mathbb{E}(e^{-\delta(\tau_3-\tau_2)}1_{\{\tau_3<\tau^D\}})\Big[(a_3-b_3)+D_3+\cdots\Big]\Big]\Big], \end{aligned} \tag{80}$$

so that we obtain

$$V^D(u) = \frac{W_\delta(u)}{W_\delta(a_1)}\sum_{k=1}^{n}\left(a_k - b_k + \frac{W_\delta(b_k-l_k)}{W_\delta'(b_k-l_k)}\right)\prod_{i=2}^{k}\mathbb{E}(e^{-\delta(\tau_i-\tau_{i-1})}1_{\{\tau_i<\tau^D\}}), \tag{81}$$

with the usual convention $\prod_{i=2}^{1} \cdot = 1$. By the strong Markov property, $\tau_{i-1}^d - \tau_{i-1}$ can be seen as the time to ruin of a process with initial surplus $b_i - l_i$ which pays dividends according to a barrier strategy at level $b_i - l_i$. Moreover, on $1_{\{\tau_i < \tau^D\}}$ and given $C^D_{\tau^d_{i-1}}$, $\tau^d_{i-1} - \tau_{i-1}$ is independent of $\tau_i - \tau^d_{i-1}$, and due to

$$\mathbb{E}\left(e^{-\delta(\tau_i - \tau^d_{i-1})} 1_{\{\tau_i < \tau^D\}} \mid C^D_{\tau^d_{i-1}}\right) = \frac{W_\delta\left(C^D_{\tau^d_{i-1}}\right)}{W_\delta(a_i)},$$

we obtain

$$\mathbb{E}(e^{-\delta(\tau_i - \tau_{i-1})} 1_{\{\tau_i < \tau^D\}}) = \mathbb{E}\left(e^{-\delta(\tau^d_{i-1} - \tau_{i-1})} \frac{W_\delta\left(C^D_{\tau^d_{i-1}}\right)}{W_\delta(a_i)} 1_{\{\tau_i < \tau^D\}}\right).$$

Hence, (81) can be rewritten as

$$V^D(u) = W_\delta(u) \sum_{k=1}^{n} B(a_k, b_k, l_k) \prod_{i=1}^{k-1} G(a_i, b_i, l_i)$$

with

$$B(a, b, l) = \frac{a - b + W_\delta(b-l)/W_\delta'(b-l)}{W_\delta(a)}$$

and

$$G(a_i, b_i, l_i) = \frac{1}{W_\delta(a_i)} \mathbb{E}\left(e^{-\delta(\tau_i^d - \tau_i)} W_\delta\left(C^D_{\tau^d_i}\right) 1_{\{\tau_{i+1} < \tau^D\}}\right).$$

As in the computation of the explicit formula for A, we can use the Gerber-Shiu measure and the dividends-penalty identity to give an explicit form for G, thus obtaining

$$\begin{aligned} G(a, b, l) = {} & \frac{\sigma^2}{2} \frac{W_\delta(l)}{W_\delta(a)} \left(W_\delta'(b-l) - \frac{W_\delta(b-l) W_\delta''(b-l)}{W_\delta'(b-l)}\right) \\ & + \int_0^l \int_0^\infty \frac{W_\delta(l-y)}{W_\delta(a)} \left(\frac{W_\delta(b-l) W_\delta'(b-l-z)}{W_\delta'(b-l)} - W_\delta(b-l-z)\right) f_\Pi(-y-z) dz dy. \end{aligned}$$

as desired. □

A few comments are in order: Equation (79) shows that we can express V^D in a recursive way as follows: define the sequence $c_1^D, \ldots, c_n^D$ by

$$c_n^D = B(a_n, b_n, l_n)$$

and, for $1 \leq j \leq n-1$,

$$c_j^D = B(a_j, b_j, l_j) + c_{j+1}^D G(a_j, b_j, l_j).$$

With these definitions, we have, for any $1 \leq j \leq n-1$,

$$V^D(u) = W_\delta(u) \sum_{k=1}^{j-1} B(a_k, b_k, l_k) \prod_{i=1}^{k-1} G(a_i, b_i, l_i) + W_\delta(u) c_j^D \prod_{i=1}^{j-1} G(a_i, b_i, l_i). \tag{82}$$

In particular, $V^D(u) = W_\delta(u) c_1^D$. The advantage of defining the c_{j+1}^D's in this way is that, once c_{j+1}^D is known, c_j^D depends only on a_j, b_j and l_j, a fact that can be exploited in a constrained optimization setting (cf. Section 4.6).

Now, while the expressions for A and G are explicit, they are rather complicated. One reason for this is found by examining the proof of the previous theorem. Observe that, while computing G, one encounters the expectation $\mathbb{E}(e^{-\delta(\tau_i-\tau_{i-1})}1_{\{\tau_i<\tau^D\}})$. Rewriting this as

$$\mathbb{E}(e^{-\delta(\tau_i-\tau_{i-1})}1_{\{\tau_i<\tau^D\}}) = \mathbb{E}(e^{-\delta(\tau_{i-1}^d-\tau_{i-1})} \cdot e^{-\delta(\tau_i-\tau_{i-1}^d)}1_{\{\tau_i<\tau^D\}}), \tag{83}$$

we see that a reason for the involved expressions is that, in general, the two product terms in (83) are not independent. Exceptions are the diffusion case and the Cramér-Lundberg model with exponential claims, where a decomposition into a product of two expectations is feasible, which significantly simplifies the analysis. To that end, note that for general spectrally negative Lévy processes, a formula for the joint Laplace transform of $\tau_k^d - \tau_k$ and $C^D_{\tau_k^d}$ for any $k = 1, \dots, n$ is available:

(84)
$$\mathbb{E}(e^{-\delta(\tau_k^d-\tau_k)-\theta(l_k-C^D_{\tau_k^d})}) = Z_\delta(b_k-l_k,\theta) + W_\delta(b_k-l_k)\,\frac{W_\delta(b_k-l_k)\,(\kappa(\theta)-\delta)-\theta Z_\delta(b_k-l_k,\theta)}{W_\delta'(b_k-l_k)},$$

where $Z_\delta(x,\theta)$ denotes the second scale function defined by

$$Z_\delta(x,\theta) = e^{-\theta x}\left(1-(\kappa(\theta)-\delta)\int_0^x e^{-\theta y}W_\delta(y)dy\right), \quad x \geq 0$$

(see [**78**]). For $\theta = 0$, one obtains the familiar simpler version

$$Z_\delta(x,0) = 1+\delta\int_0^x W_\delta(y)dy, \quad x \geq 0,$$

which was for instance used in [**83**, Ch.8.2]. Formula (84) originally goes back to [**19**]. For the concrete form used here, see [**5**, Eq.25]. In particular,

$$\begin{aligned}\mathbb{E}(e^{-\delta(\tau_k^d-\tau_k)}) &= Z_\delta(b_k-l_k,0) - \delta\,\frac{(W_\delta(b_k-l_k))^2}{W_\delta'(b_k-l_k)}\\ &= 1+\delta\int_0^{b_k-l_k} W_\delta(y)dy - \delta\,\frac{(W_\delta(b_k-l_k))^2}{W_\delta'(b_k-l_k)},\end{aligned} \tag{85}$$

which can now help to simplify the form of G, see below.

The formulas presented so far assume $u \leq a_1$. However, by the description of the strategy given at the end of Section 4.2, in the case $u > a_1$, one can simply replace a_1 by u in (74) and (79) to obtain the formulas for ϕ^D and V^D.

4.3.1. The diffusion case. Let us now look at the special case of a diffusion approximation

$$C_t = u + \mu t + \sigma B_t, \quad t \geq 0$$

in more detail, where $\mu > 0$ is a constant drift, $\sigma > 0$ is the volatility and $(B_t)_{t\geq 0}$ denotes a standard Brownian motion. In this case $C^D_{\tau_k^d} = l_k$ (deterministically), so that (74) simplifies to

$$\phi^D(u) = \kappa'(0)\,W_0(u)\prod_{k=1}^n \frac{W_0(l_k)}{W_0(a_k)}. \tag{86}$$

It is well-known that the Laplace exponent for this diffusion case is simply given by

$$\kappa(\theta) = \theta\mu + \frac{1}{2}\theta^2\sigma^2,$$

and correspondingly the (first) scale function is

$$W_\delta(x) = \frac{1}{\sqrt{\mu^2 + 2\delta\sigma^2}}\,(e^{\theta_1 x} - e^{\theta_2 x}), \quad x \geq 0, \tag{87}$$

where $\theta_1 \geq 0$ and $\theta_2 < 0$ are the two roots of the quadratic equation

$$\frac{1}{2}\sigma^2 z^2 + \mu z - \delta = 0. \tag{88}$$

See e.g. [**83**] for details. With the resulting

$$W_0(u) = (1 - e^{-(2\mu/\sigma^2)u})/\mu$$

and $\kappa'(0) = \mu$ we hence obtain the survival probability

$$\phi^D(u) = (1 - e^{-(2\mu/\sigma^2)u}) \prod_{k=1}^{n} \frac{1 - e^{-(2\mu/\sigma^2)l_k}}{1 - e^{-(2\mu/\sigma^2)a_k}}. \tag{89}$$

For numerical purposes later on, we note that in view of (74), in the diffusion case

$$A(a,l) = \frac{1 - e^{-(2\mu/\sigma^2)l}}{1 - e^{-(2\mu/\sigma^2)a}}$$

(note that $A(a,b,l)$ does not depend on b here, so that we suppress it in the notation).

For the expected discounted dividends, we note that we are in one of the exceptions where we can factor (84) as

$$\mathbb{E}(e^{-\delta(\tau_i - \tau_{i-1})} 1_{\{\tau_i < \tau^D\}}) = \mathbb{E}(e^{-\delta(\tau_{i-1}^d - \tau_{i-1})}) \cdot \mathbb{E}(e^{-\delta(\tau_i - \tau_{i-1}^d)} 1_{\{\tau_i < \tau^D\}}),$$

which helps seeing that G and B are given by

$$\begin{aligned} G(a,b,l) &= \frac{\sigma^2}{2}\frac{W_\delta(l)}{W_\delta(a)}\left(W_\delta'(b-l) - \frac{W_\delta(b-l)W_\delta''(b-l)}{W_\delta'(b-l)}\right), \\ B(a,b,l) &= \frac{a - b + W_\delta(b-l)/W_\delta'(b-l)}{W_\delta(a)}. \end{aligned}$$

Equivalently, combining (85) and (87) we obtain after algebraic manipulations,

$$G(a,b,l) = \frac{(\theta_1 - \theta_2)(e^{\theta_1 l} - e^{\theta_2 l})e^{(\theta_1+\theta_2)b}}{(e^{\theta_1 a} - e^{\theta_2 a})(\theta_1 e^{\theta_1 b + \theta_2 l} - \theta_2 e^{\theta_2 b + \theta_1 l})},$$

and

$$B(a,b,l) = \sqrt{\mu^2 + 2\delta\sigma^2}\,\frac{a - b + (e^{\theta_1(b-l)} - e^{\theta_2(b-l)})/(\theta_1 e^{\theta_1(b-l)} - \theta_2 e^{\theta_2(b-l)})}{e^{\theta_1 a} - e^{\theta_2 a}}.$$

4.3.2. The Cramér-Lundberg model with exponential claims. Consider now the Cramér-Lundberg model (73) with a homogeneous Poisson process of intensity $\lambda > 0$ and exponential claims with parameter $\alpha > 0$. In that case, we have

$$\kappa(\theta) = c\theta - \frac{\lambda\theta}{\theta + \alpha}$$

and (under the positive saftely loading condition $c > \lambda/\alpha$) the scale function is given by

$$W_\delta(x) = \frac{(\alpha + \Phi_\delta)e^{\Phi_\delta x} - (\alpha - R_\delta)e^{-R_\delta x}}{c(\Phi_\delta + R_\delta)}, \quad x \geq 0, \tag{90}$$

where $\Phi_\delta \geq 0$ and $-R_\delta < 0$ are the two roots of the quadratic equation

$$c\rho^2 + (c\alpha - \lambda - \delta)\rho - \alpha\delta = 0, \tag{91}$$

see e.g. [5]. As an immediate consequence we have

$$W_0(u) = \frac{\alpha - (\alpha - R_0)e^{-R_0 u}}{c\,R_0} = \frac{\alpha - \frac{\lambda}{c}e^{-(\alpha-\lambda/c)u}}{c\alpha - \lambda}, \quad u \geq 0,$$

and with $\kappa'(0) = c - \lambda/\alpha$ the classical formula

$$\phi(u) = 1 - \frac{\lambda}{\alpha c}e^{-(\alpha-\lambda/c)u}.$$

For our dividend model, observe that

$$\begin{aligned} \mathbb{E}(W_0(C^D_{\tau^d_k})) &= \int_0^{l_k} \left(\frac{\alpha - \frac{\lambda}{c}e^{-(\alpha-\lambda/c)(l_k-y)}}{c\alpha - \lambda} \right) \alpha\, e^{-\alpha y}\, dy \\ &= \frac{\alpha - \alpha e^{(\lambda/c-\alpha)l_k}}{c\alpha - \lambda}, \end{aligned} \tag{92}$$

which leads to the survival probability

$$\phi^D(u) = \left(1 - \frac{\lambda}{\alpha c}e^{-(\alpha-\lambda/c)u}\right) \prod_{k=1}^n \frac{1 - e^{-(\alpha-\lambda/c)l_k}}{1 - \frac{\lambda}{\alpha c}e^{-(\alpha-\lambda/c)a_k}}. \tag{93}$$

For numerical purposes later on, we note that in view of (74) in this case

$$A(a,l) = \frac{1 - e^{-(\alpha-\lambda/c)l}}{1 - \frac{\lambda}{\alpha c}e^{-(\alpha-\lambda/c)a}}.$$

Due to the lack-of-memory property of the exponential distribution, we can again decompose (84) as a product

$$\mathbb{E}(e^{-\delta(\tau_i-\tau_{i-1})}1_{\{\tau_i<\tau^D\}}) = \mathbb{E}(e^{-\delta(\tau^d_{i-1}-\tau_{i-1})}) \cdot \mathbb{E}(e^{-\delta(\tau_i-\tau^d_{i-1})}1_{\{\tau_i<\tau^D\}}),$$

and, after some algebra, obtain

$$G(a,b,l) = \frac{\alpha\lambda(\Phi_\delta - R_\delta)\left(e^{l(\alpha+\Phi_\delta)} - e^{l(\alpha+R_\delta)}\right)e^{b(\Phi_\delta+R_\delta)-l(\alpha+\Phi_\delta+R_\delta)}}{c\left(e^{a\Phi_\delta}(\alpha+\Phi_\delta) - e^{aR_\delta}(\alpha+R_\delta)\right)\left(\Phi_\delta(\alpha+\Phi_\delta)e^{\Phi_\delta(b-l)} - R_\delta(\alpha+R_\delta)e^{R_\delta(b-l)}\right)}.$$

In addition,

$$B(a,b,l) = \frac{c(\Phi_\delta+R_\delta)((\alpha+\Phi_\delta)(1+\Phi_\delta(a-b))e^{\Phi_\delta(b-l)} - (\alpha+R_\delta)(1+R_\delta(a-b))e^{R_\delta(b-l)})}{(\Phi_\delta(\alpha+\Phi_\delta)e^{\Phi_\delta(b-l)} + R_\delta(\alpha-R_\delta)e^{-R_\delta(b-l)})((\alpha+\Phi_\delta)e^{\Phi_\delta a} - (\alpha-R_\delta)e^{-R_\delta a})}.$$

4.4. Properties of the strategy with optimal parameters

Let $\mathscr{D}_n$ denote the family of dividend strategies with n corridors as defined in Section 4.2. Observe that $\mathscr{D}_n$ can naturally be identified with the set

$$\{(\bar{a},\bar{b},\bar{l}) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \times \mathbb{R}^n_+ \mid l_k \leq b_k \leq a_k, \text{ and } l_k < a_{k+1},\ k = 1,\dots,n\},$$

where, for convenience, we set $a_{n+1} = \infty$.

For a given initial surplus $u \geq 0$ and survival constraint $0 \leq \varphi < 1$, let

$$V^*_{n,\varphi}(u) = \sup\left\{V^D(u) \mid D \in \mathscr{D}_n,\ \phi^D(u) \geq \varphi\right\}, \tag{94}$$

where we define $V^*_{n,\varphi}(u) = 0$ whenever the set is empty, e.g., whenever $\varphi > \phi(u)$. If clear from the context, we will drop the dependence of $V^*_{n,\varphi}$ on φ and simply write V^*_n in (94).

In the following, for a strategy $D \in \mathscr{D}_n$ we will denote by $a^D_i, b^D_i, l^D_i, i = 1,\dots,n$ the levels of D.

We immediately observe that, by continuity of the functions involved, $V^*_{n,\varphi}$ is locally bounded and on the set $\varphi < \phi(u)$, $V^*_{n,\varphi}(u)$ is left-continuous as a function of φ for fixed

u. Hence, for each $u > 0$, there exists a strategy $D \in \mathscr{D}_n$ such that $V_n^*(u) = V^D(u)$ and $\phi^D(u) = \varphi$, so the inequality on the right-hand side of (94) could be replaced by an equality. However, in terms of continuity, a stronger result holds.

Proposition 4.4.1. *The mapping* $(u, \varphi) \mapsto V_{n,\varphi}^*(u)$ *is continuous on the set*

$$S = \{(u, \varphi) \in \mathbb{R}^2 \mid u > 0,\ 0 \le \varphi < \phi(u)\}.$$

Proof. Define $f : \mathbb{R}_+ \times \mathscr{D}_n \to [0,1)$ and $F : \mathbb{R}_+ \times \mathscr{D}_n \to \mathbb{R}_+ \times \mathbb{R}^{3n}$ by $f(u, \bar{a}, \bar{b}, \bar{l}) = \phi^D(u)$ and

$$F(u, \bar{a}, \bar{b}, \bar{l}) = (u, a_1, \ldots, a_{n-1}, \bar{b}, \bar{l}, f(u, \bar{a}, \bar{b}, \bar{l})).$$

The D in the definition of f is the strategy associated with $(\bar{a}, \bar{b}, \bar{l})$. Now, recall that W_0 is a continuous bijection, so by Equation (74), it follows that F is a one-to-one continuous open function. Let $T \subset \mathbb{R}_+ \times \mathbb{R}^{3n-1} \times [0,1)$ be the image of F and define $\tilde{V} : T \mapsto \mathbb{R}_+$ as $\tilde{V}(u, z, \varphi) = V^D(u)$, where $(u, D) = F^{-1}(u, z, \varphi)$. From Equation (79), we see that $\tilde{V}$ is continuous and, moreover, for $(u, \varphi) \in S$, we have

$$V_{n,\varphi}^*(u) = \sup\left\{\tilde{V}(u, z, \varphi) \mid (u, z, \varphi) \in T\right\}. \tag{95}$$

The set on the right hand side of (95) is not empty as $\phi^D(u) \to \phi(u)$ whenever the levels of the strategy go to infinity. It is then clear from this last equation and the continuity of $\tilde{V}$, that $V_{n,\cdot}^*$ is jointly continuous in (u, φ). □

Next, we see that in terms of the number of bands, it is better to allow as many bands as possible.

Proposition 4.4.2. *For every* $n \in \mathbb{N}$ *and* $\phi(u) > \varphi$*, we have* $V_n^*(u) \le V_{n+1}^*(u)$*.*

Remark 4.4.3. This proposition is necessary since, in general, $\mathscr{D}_n \not\subset \mathscr{D}_{n+1}$. Indeed, in order to see an n-corridor strategy $D \in \mathscr{D}_n$ as an $(n+1)$-corridor strategy, one is required to introduce a new corridor somewhere (equivalent to including the three remaining parameters to go from $\mathbb{R}_+^{3n}$ to $\mathbb{R}_+^{3(n+1)}$) and by doing this, one could end up changing the probability of ruin. Note, however, that for Lévy processes with unbounded variation, one can still identify $\mathscr{D}_n$ with a subset $\mathscr{D}_{n+1}$ by introducing the “empty corridor” at the end (that is, the corridor for which $a_{n+1} = b_{n+1} = l_{n+1}$), so Proposition 4.4.2 is immediate in this case. In the case of bounded variation, one cannot just add more corridors without care, as adding a new corridor to a previously defined strategy, whether empty or not, strictly decreases the survival probability. So in order to increase the probability to the minimal level, the previous corridors have to be shrunk or “moved vertically up” – thus potentially decreasing the amount of dividends being paid.

Proof. Given $n \in \mathbb{N}$, let $D \in \mathscr{D}_n$ be such that $V_n^*(u) = V_D(u)$. By continuity of the scale functions and its derivatives, for a given $\varepsilon > 0$, there exists an $\epsilon > 0$ such that $V_n^*(u) \le V^D(u) + \varepsilon$, where D is strategy with surplus levels $a_k^D = a_k^{D_n}, b_k^D = b_k^{D_n}, l_k^D = l_k^{D_n}$, $k = 1, \ldots, n-1$ and remaining surplus levels equal to $a_n^* + \epsilon, b_n^* + \epsilon$ and $l_n^* + \epsilon$. Now, since $\phi^D(u) > \phi^{D_n}(u) \ge \phi_{\min}$ and $A(l+1, l+1, l) \to 1$ as $l \to \infty$, we can find $\tilde{l} > a_n^D + \epsilon$ such that $A(\tilde{l}+1, \tilde{l}+1, \tilde{l}) > \phi^{D_n}(u)/\phi^D(u)$ and $\tilde{l} > l_n^D$. Letting $D' \in \mathscr{D}_{n+1}$ denote the dividend strategy with the same first n corridors equal to those of D and extra corridor composed by the levels $a_{n+1} = b_{n+1} = \tilde{l} + 1$ and $l_{n+1} = \tilde{l}$, we clearly have $V^D(u) \le V^{D'}(u)$. Hence, $V_n^*(u) \le V_{n+1}(u) + \varepsilon \le V_{n+1}^*(u) + \varepsilon$ and, in particular, $V_n^*(u) \le V_{n+1}^*(u) + \varepsilon$. Since the value of $V_{n+1}^*(u)$ is independent of $\varepsilon > 0$, we can let $\varepsilon \downarrow 0$ in the previous inequality, proving the proposition. □

Remark 4.4.4. The cases considered in Section 4.6 seem to indicate that the inequality in Proposition 4.4.2 is in fact strict, so one is in principle always obliged to add more bands to improve the amount of dividends being paid.

Proposition 4.4.5. *There exists* $\vartheta^* \geq 0$ *and* $D \in \mathscr{D}_n$ *such that* $a_1^D = \vartheta^*$ *and* $V_n^*(u) = V^D(u)$ *for all* $u \geq a_n^*$. *In particular,* $V_n^{*\prime}(u) = 1$ *for* $u \geq a_n^*$.

In other terms, the meaning of this proposition is that the strategies maximizing the value function can eventually be taken to be "constant", in the sense that for high values of the initial capital, all but their first levels can be taken to be the same.

Proof. Notice that (see, e.g., Equation (3.15) in [**20**])

$$\lim_{x\to\infty} \frac{W_\delta(x)}{W_\delta'(x)} = \frac{1}{\Phi(\delta)},$$

so there exists $x^* > 0$ and $c > 0$ such that $W_\delta(x) \leq cW_\delta'(x)$ for all $x \geq x^*$.

Let b_1^*, l_1^* and a_k^*, b_k^*, l_k^*, $k = 2, \dots, n$ be a set of points that maximize the value of the function T given by

$$T(l_1, b_1, l_2, b_2, \dots, a_n) = -b_1 + \frac{W_\delta(b_1 - l_1)}{W_\delta'(b_1 - l_1)} + \sum_{k=2}^{n} B(a_k, b_k, l_k) \prod_{j=1}^{k-1} G(a_j, b_j, l_j)$$

subject to the constraints $l_1 \leq b_1$, $l_k \leq b_k \leq a_k$, $l_k \leq a_{k+1}$ and

$$H(b_1, l_1) \prod_{k=2}^{n} A(a_k, b_k, l_k) \geq \varphi,$$

where H is given by $H(b, l) = W_0(1)A(1, b, l)$. Let M be this maximal value. We claim that we can take $\vartheta^* = \max(b_1^*, x^*, c - M)$ and D be given by $a_1^D = \vartheta^*$ and remaining levels given by the a_k^*, b_k^*, l_k^* in the same ordering. Indeed, clearly $D \in \mathscr{D}_n$ and by definition $a_1^D = \vartheta^*$. Now, let $u \geq \vartheta^*$ and $D' \in \mathscr{D}_n$ be such that $\phi^{D'}(u) \geq \varphi$. We need to show that $V^{D'}(u) \leq V^D(u)$. We can assume at the outset that $u \leq a_1^{D'}$, since otherwise we can replace $a_1^{D'}$ by u, obtaining the same value for the strategy. Now, notice that

$$V^{D'}(u) = \frac{W_\delta(u)}{W_\delta(a_1^{D'})}(a_1^{D'} + C)$$

and

$$V^D(u) = u + M,$$

where $C = T(l_1^{D'}, b_1^{D'}, l_2^{D'}, b_2^{D'}, \dots, a_n^{D'})$. Consider the mapping $x \mapsto (x + M)/W_\delta(x)$. This function has derivative given by

$$\frac{W_\delta(x) - W_\delta'(x)(x + M)}{W_\delta(x)^2}.$$

Hence, for $x \geq \vartheta^*$, we have

$$W_\delta(x) - W_\delta'(x)(x + M) \leq W_\delta'(x)(c - M - x) \leq 0,$$

implying that the mapping is decreasing on $[\vartheta^*, \infty)$. Since $\vartheta^* \leq u \leq a_1^{D'}$, we obtain

$$V^{D'}(u) = \frac{W_\delta(u)}{W_\delta(a_1^{D'})}(a_1^{D'} + C) \leq \frac{W_\delta(u)}{W_\delta(a_1^{D'})}(a_1^{D'} + M) \leq \frac{W_\delta(u)}{W_\delta(u)}(u + M) = V^D(u),$$

finishing the proof. $\square$

We observed before that for each $u > 0$ such that $\phi(u) > \varphi$, there exists a strategy D such that $V_n^*(u) = V^D(u)$. However, as stated in the proof of the previous proposition, if in this case we have $a_1^D \leq u$, then any other strategy D' which has $b_1^D \leq a_1^{D'} \leq u$ and the remaining levels the same as D will also satisfy $V_n^*(u) = V^{D'}(u)$ and so for large values of initial surplus, there will not be a unique strategy. Moreover, this might even be the case for $a_1^D > u$. Therefore, in the following, we will rely on the following assumption.

Assumption. For each $u > 0$, there is a unique strategy $D \in \mathscr{D}_n$ satisfying $a_1^D \geq u$, $\phi^D(u) = \varphi$ and $V_n^*(u) = V^D(u)$. This strategy will be denoted by $D_n^*(u)$ and its levels by $a_{n,k}^*, b_{n,k}^*, l_{n,k}^*$.

With this notation, we have the following:

Proposition 4.4.6. *Assume $V_n^* < V_{n+1}^*$ for every $n \in \mathbb{N}$. Then, the function D_n^* is continuous on $(\phi^{-1}(\varphi), \infty)$.*

Proof. Given $u > \phi^{-1}(\varphi)$, let $(u_m)_{m\geq 0}$ be a sequence in $(\phi^{-1}(\varphi), \infty)$ converging to u. Without loss of generality, we may assume that there exists an $\alpha > 0$ such that $\phi(u-\alpha) > \varphi$, and the entire sequence is contained in an interval $[u-\alpha, u+\alpha]$. In the notation of the proof of Proposition 4.4.1, let

$$E = ([u-\alpha, u+\alpha] \times \mathbb{R}^{3n}) \cap F^{-1}(\mathbb{R}_+ \times \mathbb{R}^{3n-1} \times \{\varphi\}).$$

Observe that E is simply the set of pairs (v, D), where $v \in [u-\alpha, u+\alpha]$ and D is a strategy such that $\phi^D(v) = \varphi$. The description given in the previous equation simply shows that E is closed and, by construction, the pairs $(u_m, D_n^*(u_m))$ belong to E. We claim that, further than that, there exists an $M > 0$ such that the ball B_M in $\mathbb{R}^{3n+1}$ of radius M centred in the origin contains the pairs $(u_m, D_n^*(u_m))$, $m \geq 0$, thus showing that these pairs are contained in the compact set $K = E \cap B_M$. Arguing by contradiction, suppose this is not the case. Since the u_m's are clearly bounded, there has to exist at least one coordinate of $D_n^*(u_m)$ that is not bounded. By the description of $\mathscr{D}_n$, it follows that there has to exist at least one k such $(a_{n,k}^*(u_m))_{m\geq 0}$ is unbounded. Let

$$J_U = \{k \in \{1, \dots, n\} \mid (a_{n,k}^*(u_m))_{m\geq 0} \text{ is unbounded}\}$$

be the set of indices producing unbounded sequences of the a's and $J_B = \{1, \dots, n\} \setminus J_U$. Observe that if $k \in J_B$, then also $(b_{n,k}^*(u_m))_{m\geq 0}$ and $(l_{n,k}^*(u_m))_{m\geq 0}$ are bounded. By passing to a subsequence if necessary, we can assume that

- If $k \in J_U$, then $a_{n,k}^*(u_m) \to \infty$ as $m \to \infty$.
- If $k \in J_B$, then there exist a_k, b_k and l_k such that $a_{n,k}^*(u_m) \to a_k$, $b_{n,k}^*(u_m) \to b_k$ and $l_{n,k}^*(u_m) \to l_k$.

Now, since

$$B(a_{n,k}^*(u_m), b_{n,k}^*(u_m), l_{n,k}^*(u_m)) \leq \frac{a_{n,k}^*(u_m)}{W_\delta(a_{n,k}^*(u_m))} + \frac{W_\delta(b_{n,k}^*(u_m) - l_{n,k}^*(u_m))}{W_\delta(a_{n,k}^*(u_m)) W_\delta'(b_{n,k}^*(u_m) - l_{n,k}^*(u_m))},$$

and

$$\lim_{m\to\infty} \frac{a_{n,k}^*(u_m)}{W_\delta(a_{n,k}^*(u_m))} = 0 \quad \text{and} \quad \lim_{m\to\infty} \frac{W_\delta(b_{n,k}^*(u_m) - l_{n,k}^*(u_m))}{W_\delta(a_{n,k}^*(u_m)) W_\delta'(b_{n,k}^*(u_m) - l_{n,k}^*(u_m))} = 0,$$

then $B(a_{n,k}^*(u_m), b_{n,k}^*(u_m), l_{n,k}^*(u_m)) \to 0$ as $m \to \infty$. The first limit is zero because

$$\lim_{m\to\infty} e^{-\Phi(\delta)x} W_\delta(x) = 1.$$

For the second we have two cases: the sequence $((b^*_{n,k}(u_m) - l^*_{n,k}(u_m))$ is bounded or unbounded. If it is bounded, the quotient

$$W_\delta(b^*_{n,k}(u_m) - l^*_{n,k}(u_m))/W'_\delta(b^*_{n,k}(u_m) - l^*_{n,k}(u_m))$$

is bounded by continuity and strict positivity of W'_δ, which implies that the limit is zero. If it is unbounded, we simply notice that $W_\delta(a^*_{n,k}(u_m))/W_\delta(b^*_{n,k}(u_m) - l^*_{n,k}(u_m)) \leq 1$ and since W'_δ diverges to infinity, the limit is again zero. Since G is bounded by 1, we see that, on the one hand

$$\lim_{m\to\infty} V^*_n(u_m) = V^*_n(u),$$

while on the other,

$$\begin{aligned}\lim_{m\to\infty} V^*_n(u_m) &= \lim_{m\to\infty} W_\delta(u) \sum_{k=1}^{n} B(a^*_{n,k}(u_m), b^*_{n,k}(u_m), l^*_{n,k}(u_m)) \prod_{i=1}^{k-1} G(a^*_{n,i}(u_m), b^*_{n,i}(u_m), l^*_{n,i}(u_m))\\ &\leq W_\delta(u) \sum_{k\in J_B}^{n} B(a_k, b_k, l_k) \prod_{\substack{i=1\\ i\in J_B}}^{k-1} G(a_i, b_i, l_i).\end{aligned}$$

It is not hard to see that, if p is the cardinality of J_B, then $p \leq n-1$ and the a_k's, b_k's and l_k's form a strategy $D \in \mathscr{D}_p$ with $\phi^D(u) \geq \varphi$. Therefore, $V^*_n(u) \leq V^*_p(u)$, which is clearly a contradiction to the hypothesis of the proposition. Thus, there exists an $M > 0$ such that $\mid D^*_n(u_m) \mid \leq M$ and $D^*_n(u_m) \in K$ for every $m \geq 0$.

Continuity of D^*_n is readily proven: if the D^*_n were not continuous at u, we would be able to find an $\varepsilon > 0$ and a subsequence, which we can assume to be the original sequence, such that $\mid D^*_n(u_m) - D^*_n(u) \mid > \varepsilon$. By compactness of K, there would exist a subsequence $(u_{m_r})_{r\geq 0}$ such that $((u_{m_r}, D^*_n(u_{m_r})))_{r\geq 0}$ converged to a point in K, say (v, D). Since $u_{m_r} \to u$, we have $v = u$ and, moreover,

$$V^*_n(u) = \lim_{r\to\infty} V^*_n(u_{m_r}) = \lim_{r\to\infty} V^{D^*_n(u_{m_r})}(u_{m_r}) = V^D(u).$$

Since $\phi^D(u) = \varphi$, uniqueness of the strategies would imply that $D = D^*_n(u)$, which would not be possible for large enough r according to the choice of $(u_m)_{m\geq 0}$. Hence, D is continuous. □

The argument to show boundedness of D^*_n in the previous proof is simply a formalization of the intuitive idea that the best strategy cannot have arbitrarily high levels/corridors, as this would imply longer waiting times between dividend payments.

Remark 4.4.7. We close this section with an (informal) discussion on the potential overall-optimality of the strategy. In the stochastic control literature, optimality results are usually derived by formulating a dynamic programming principle associated with the problem. While in general, due to time-inconsistency, these principles cannot be formulated for control problems with ruin probability considerations, in the current setting there is some evidence pointing to a weaker version of this to indeed hold: in the case of the diffusion, from Equations (74) and (79), we obtain

$$a^*_{n,i+1}(u,\varphi) = a^*_{n-1,i}(u_1,\varphi_1), \quad i = 1,\ldots,n-1, \tag{96}$$

and similarly for the $b^*_{n,i+1}$'s and $l^*_{n,i+1}$'s, where $u_1 = l^*_{n,1}(u,\varphi)$, $\varphi_1 = \varphi W_0(a^*_{n,1}(u,\varphi))/W_0(u)$ and the second argument now refers to the survival probability under consideration. The meaning of (96) is that, with the knowledge of $a^*_{n,1}(u)$, $b^*_{n,1}(u)$ and $l^*_{n,1}(u)$, one could determine the remaining levels of the best n-corridor strategy with survival probability φ by now considering instead the best $(n-1)$-corridor strategy with initial surplus u_1 and φ_1. Furthermore, iterating (96), we obtain

$$a^*_{n,i+1}(u,\varphi) = a^*_{n-i,1}(u_i,\varphi_i), \quad i = 1,\dots,n-1, \tag{97}$$

where, recursively, we have $u_i = l^*_{n-i+1,1}(u_{i-1})$ and $\varphi_i = \varphi_{i-1}W_0(a^*_{n-i+1,1}(u_{i-1}))/W_0(u_{i-1})$, $i = 2,\dots,n$. The recursive nature of these equations hints towards the existence of a type of dynamic programming principle leading to the value of $V^*_{n,\varphi}$. Indeed, observe that Equation (96) is, in a way, equivalent to the equation

$$V^*_{n,\varphi}(u) = \max_{D\in\mathscr{D}_n} \mathbb{E}_u\left[\int_0^{\tau^D\wedge\tau_1^d} e^{-\delta t}\, dD_t + e^{-\delta(\tau^D\wedge\tau_1^d)}V^*_{n-1}\left(C^D_{\tau^D\wedge\tau_1^d}, \frac{W_0(a_1^D)}{W_0(u)}\varphi\right)\right], \tag{98}$$

and Equation (97) could then be phrased as

(99)

$$V^*_{n,\varphi}(u) = \max_{D\in\mathscr{D}_n} \mathbb{E}_u\left[\int_0^{\tau^D\wedge\tau_i^d} e^{-\delta t}\, dD_t + e^{-\delta(\tau^D\wedge\tau_i^d)}V^*_{n-i}\left(C^D_{\tau^D\wedge\tau_i^d}, \frac{W_0(a_1^D)\cdots W_0(a_i^D)}{W_0(u)\cdots W_0(l_{i-1}^D)}\varphi\right)\right].$$

In both cases one needs to be able to keep track of the survival probability to account for "how much probability has been consumed" and one cannot just plug in arbitrary stopping times, allowing for a classical a dynamic programming principle. For example, we have

$$\begin{aligned}
V^*_{n,\varphi}(u) &= \max_{D\in\mathscr{D}_n} \mathbb{E}_u\left[\int_0^{\tau^D\wedge\tau_1} e^{-\delta t}\, dD_t + e^{-\delta(\tau^D\wedge\tau_1)}V^*_{n-1}\left(C^D_{\tau^D\wedge\tau_1-}, \frac{W_0(a_1^D)}{W_0(u)}\varphi\right)\right]\\
&= \max_{D\in\mathscr{D}_n} \mathbb{E}_u\left[e^{-\delta(\tau^D\wedge\tau_1)}V^*_{n-1}\left(C^D_{\tau^D\wedge\tau_1-}, \frac{W_0(a_1^D)}{W_0(u)}\varphi\right)\right],
\end{aligned}$$

where the difference is that one now has to account for the value of the process just before the lump-sum payment at a_1, i.e., $C^D_{\tau^D\wedge\tau_1-}$.

These equations, however, suggest the idea of a dynamic programming principle being satisfied *locally* as follows: for any stopping time τ with $\tau^d_{i-1}\le\tau\le\tau_i$ for $i\in\{1,\dots,n\}$ (and $\tau_0^d = 0$), we have

$$V^*_{n,\varphi}(u) = \max_{D\in\mathscr{D}_n} \mathbb{E}_u\left[\int_0^{\tau^D\wedge\tau} e^{-\delta t}\, dD_t + e^{-\delta(\tau^D\wedge\tau)}V^*_{n-i}\left(C^D_{\tau^D\wedge\tau-}, \frac{W_0(a_1^D)\cdots W_0(C^D_{\tau^D\wedge\tau-})}{W_0(u)\cdots W_0(l_{i-1}^D)}\varphi\right)\right],$$

while for $\tau_i < \tau \le \tau_i^d$,

$$V^*_{n,\varphi}(u) = \max_{D\in\mathscr{D}_n} \mathbb{E}_u\left[\int_0^{\tau^D\wedge\tau} e^{-\delta t}\, dD_t + e^{-\delta(\tau^D\wedge\tau)}V^*_{n-i}\left(C^D_{\tau^D\wedge\tau}, \frac{W_0(a_1^D)\cdots W_0(a_i^D)}{W_0(u)\cdots W_0(l_{i-1}^D)}\varphi\right)\right].$$

Although intuitive, we do not attempt to formalize this approach here, as it is beyond the scope of the present paper.

4.5. Optimization of barrier levels

From the previous considerations, we know that for each $n\in\mathbb{N}$ there exists a strategy $D^*\in\mathscr{D}_n$ such that $V^*_n(u) = V^{D^*}(u)$. We are now interested in identifying this strategy, which is equivalent to finding surplus values $a_k, b_k, l_k,\ k = 1,\dots,n$, which maximize $V^D(u)$

subject to the constraint $\phi^D(u) \geq \varphi$. Recall that the objective function and the constraint are of the form

$$V^D(u) = W_\delta(u) \sum_{k=1}^{n} B(a_k, b_k, l_k) \prod_{i=1}^{k-1} G(a_i, b_i, l_i)$$

and

$$\phi^D(u) = \kappa'(0)\, W_0(u) \prod_{k=1}^{n} A(a_k, b_k, l_k),$$

respectively. We will pursue two different approaches in the sequel.

4.5.1. A gradient-inspired method. Despite the possibly high-dimensional nature of this optimization problem, the particular structure of the above equations makes a classical Lagrange method look feasible. In what follows, we fix $n \in \mathbb{N}$.

It is clear that the constraint is "adversarial" to the objective function, meaning that the surplus values that maximize V will at the same time minimize ϕ^D. Given that the constraint is imposed in terms of an inequality, it follows that the optimal strategy D^* will as well satisfy $\phi^{D^*}(u) = \phi_{\min}$. Using this observation, we consider hence the function

$$L(a_1, \ldots, a_n, b_1, \ldots, b_n, l_1, \ldots, l_n, \Lambda) = V(u) - \Lambda(\phi^D(u) - \phi_{\min}).$$

The normal equations then turn out to be[1]

$$\begin{aligned}
&W_\delta(u) D_i B(a_m, b_m, l_m) \prod_{j=1}^{m-1} G(a_j, b_j, l_j) \\
&\quad + W_\delta(u) \sum_{k=m+1}^{n} B(a_k, b_k, l_k) D_i G(a_m, b_m, l_m) \prod_{\substack{j=1 \\ j \neq m}}^{k-1} G(a_j, b_j, l_j) \\
&\quad - \Lambda \kappa'(0)\, W_0(u) D_i A(a_m, b_m, l_m) \prod_{\substack{k=1 \\ k \neq m}}^{n} A(a_k, b_k, l_k) = 0, \quad m = 1, \ldots, n,\ i = 1, 2, 3 \\
&\kappa'(0)\, W_0(u) \prod_{k=1}^{n} A(a_k, b_k, l_k) - \phi_{\min} = 0.
\end{aligned} \tag{100}$$

Despite the relatively easy form of these equations, even for the simplest cases no exact solution can be provided in terms of elementary functions. The following example illustrates this point:

Example 4.5.1. Let us consider $n = 1$, i.e. there is only one level a, at which a lump sum is paid down to a barrier $b \leq a$, and after that dividend payments at this barrier take place until the barrier is dissolved when the surplus value undershoots l. Recalling the formulas for W and A in the diffusion case, the equations

$$W_\delta''(b - l) = 0, \qquad \kappa'(0)\, W_0(u) A(a, l) = \phi_{\min}$$

[1]Here $D_i A$ designs the partial derivative of A with respect to its i-th argument and similarly $D_i B$ and $D_i G$. We interchangeably use the notation A_a, A_b and A_l to denote the partial derivatives of A.

allow us to write a and b in terms of l, obtaining

$$
\begin{aligned}
b &= l - \frac{\log(\theta_1^2) - \log(\theta_2^2)}{\theta_1 - \theta_2}, \\
a &= -\frac{\sigma^2}{2\mu}\left(\log\left(\frac{\phi_{\min}}{\kappa'(0)\,W_0(u)} + e^{-(2\mu/\sigma^2)l} - 1\right) - \log\left(\frac{\phi_{\min}}{\kappa'(0)\,W_0(u)}\right)\right).
\end{aligned}
$$

Equations (100) reduce into

$$
\frac{A_a(a,l)}{A_l(a,l)} = \frac{B_a(a,b,l)}{B_l(a,b,l)},
$$

so, by letting $c = 2\mu/\sigma^2$ and $d = \phi_{\min}\kappa'(0)^{-1}\,W_0(u)^{-1}$, we obtain

$$
\begin{aligned}
(101)\quad & (d-1)e^{cl}\left(1 - \rho\left(d - 1 + e^{-cl}\right)^{\alpha}\right) \\
& \qquad + \left(\gamma - l - \frac{1}{c}\log\left(d - 1 + e^{-cl}\right)\right)\left(\theta_1 - \theta_2\rho\left(d - 1 + e^{-cl}\right)^{\alpha}\right) = 0
\end{aligned}
$$

with

$$
\alpha = (\theta_1 - \theta_2)/c, \quad \rho = d^{-\alpha}, \quad \xi = \frac{\log(\theta_2^2) - \log(\theta_1^2)}{\theta_1 - \theta_2}, \quad \gamma = \frac{1}{c}\log(d) + \frac{W_\delta(\xi)}{W_\delta'(\xi)} + \xi.
$$

By making the change of variable $y = e^{-cl}$, after some algebraic manipulations, (101) becomes

$$
(d-1)\left(1 - \rho\,(d-1+y)^{\alpha}\right) + y\left(\gamma + \frac{1}{c}\log\left(\frac{y}{d-1+y}\right)\right)\left(\theta_1 - \theta_2\rho\,(d-1+y)^{\alpha}\right) = 0.
$$

While easily solved by a numerical optimizer, the solution cannot be expressed in terms of elementary functions. □

The last equation in the previous example accepts two possible solutions for y, although only one making $l \le b \le a$. Situations like this arise similarly for a higher number of corridors. While for small n this might be something easy to deal with, for large n this issue might introduce a complexity problem in the numerical solution of the equations resulting from (100). These considerations motivate pursuing different alternatives for obtaining the optimal corridor levels, one of which we explain now.

In what follows we will focus only in the case of the diffusion. Hence, A does not depend on its second argument and we actually have $A(a,l) = W_0(l)/W_0(a)$. Motivated by the constraint $\phi^D(u) = \varphi$, for fixed a_k, we introduce the change of variable $s_k = A(a_k, l_k) = W_0(l_k)/W_0(a_k)$, so that $l_k = W_0^{-1}(s_k W_0(a_k))$, $k = 1, \dots, n$. The constraint can now be phrased in terms of the s_k's, where we have $\kappa'(0)W_0(u)\prod_k s_k = \varphi$. Assume for the moment that the optimal s_k's are known, which we denote by $s^*_{n,1}, \dots, s^*_{n,n}$ in accordance with the notation of Section 4.4. Since a_n, b_n and l_n appear only as arguments of B in the the last term of the sum in (79), the optimal levels of the last corridor, $a^*_{n,n}$, $b^*_{n,n}$ and $l^*_{n,n}$, should also maximize the mapping

$$
(a,b,l) \mapsto B(a,b,l)
$$

subject to the constraint $s^*_{n,n} = W_0(l)/W_0(a)$. Since the inverse of W_0 can be explicitly computed from the formula given in Section 4.3.1, this is a two-dimensional optimization problem, which can be easily solved by standard optimization techniques. Assume we have found the optimal levels for the last corridor and, motivated by (82), let $c_n = B(a^*_{n,n}, b^*_{n,n}, l^*_{n,n})$. We can move backwards one step and repeat a similar procedure by observing that a_{n-1}, b_{n-1}

and l_{n-1} appear only in the last two terms of the sum in (79) so, after dividing by a common factor, we notice that $a^*_{n,n-1}$, $b^*_{n,n-1}$ and $l^*_{n,n-1}$ should maximize the mapping

$$(a,b,l) \mapsto B(a,b,l) + c_n G(a,b,l)$$

subject to the constraint $s^*_{n,n-1} = W_0(l)/W_0(a)$. We can then set

$$c_{n-1} = B(a^*_{n,n-1}, b^*_{n,n-1}, l^*_{n,n-1}) + c_n G(a^*_{n,n-1}, b^*_{n,n-1}, l^*_{n,n-1})$$

and repeat. Since $V^*_n(u) = W_\delta(u)c_1$, we observe that by proceeding in this fashion, the optimal strategy can be obtained as a result of n consecutive 2-dimensional problems.

Having described an approach to finding the $a^*_{n,k}$'s, $b^*_{n,k}$'s and $l^*_{n,k}$'s given the $s^*_{n,k}$'s, we are only left with the question of finding the appropriate $s^*_{n,k}$'s. Since the sequential optimization part is relatively fast for a given set of the s_k's, optimal or not, we can use a greedy method to solve this last issue. The overall procedure is summarized in Algorithm 5.

The idea of the algorithm is to start with an arbitrary set of levels $s_1, \dots, s_n$ satisfying the constraint $\kappa'(0)W_\delta(u)\prod_k s_k = \varphi$ and the value of the c_k's associated with them (Lines 2 and 3). Once these initial values are set, the algorithm iterates through all couples (m_1, m_2) of indices m_1 and m_2 in the set $\{1, \dots, n\}$. The idea is to alter the values of s_{m_1} and s_{m_2} in a multiplicative way so that the constraint is still satisfied, obtaining a new set of s-levels, which we denote by $\tilde{s}_1, \dots, \tilde{s}_n$ (Lines 8 to 10). If after computing the optimal corridor levels associated with the $\tilde{s}_k$'s (Lines 11 to 18) we observe that the alteration leads to an improvement in $V^D(u)$, we discard the current s-levels and replace them by their tilded versions, moving to the next iteration. Observe that for $m = \min(m_1, m_2)$ one can already notice an improvement if $\tilde{c}_m > c_m$, so it is not necessary to check every value. Once the overall procedure has been repeated L times, we compute the corridor levels associated with the latest version of the s_k's, which are an approximation to the optimal levels $a^*_{n,k}$, $b^*_{n,k}$ and $l^*_{n,k}$, $k = 1, \dots, n$.

The optimization function P appearing in Lines 11, 15, 26 and 29 is to be understood as any procedure that maximizes the function given in the first argument subject to A being equal to the second argument of P. As explicit formulas are available for the inverse of W_0, we can replace the third arguments of B and G, and let P be any *unconstrained* maximization algorithm.

A few comments are in order: while the algorithm can be used for any value of n, the complexity of its main iterating procedure scales quadratically with the number of corridors. Combined with the optimization procedure, this will typically lead to an excessive computation time. To address this issue, one can restrict the set of couples (m_1, m_2) that are considered for improvement. This, however, requires a good set of initial values for the s_k's. A rule of thumb used during the implementation was to compute the optimal $s^*_{n,k}$'s for a small n and for computing the optimal levels for, say, $N > n$ corridors, initialize the s_k's as $s_k = s^*_{n,k}$, $k = 1, \dots, n$ and $s_k = 1$, $k = n+1, \dots, N$. For these initial values, one would restrict m_1 to the set $\{1, \dots, n\}$ and m_2 to the set $\{n+1, \dots, N\}$. Conversely, one could initialize the s_k's as $s_{N-n+k} = s^*_{n,k}$, $k = 1, \dots, n$ and $s_k = 1$, $k = 1, \dots, N-n$, and restrict m_1 to the set $\{1, \dots, N-n\}$ and m_2 to $\{N-n+1, \dots, N\}$. While both approaches yielded similar results, the latter performed slightly better, with exceptionally good results obtained by taking $N = n+1$ and repeating the procedure several times (see also Section 4.6.1 for further insights into the initialization procedure).

Remark 4.5.2. Recall from (96) and (97) that

$$a^*_{n,i+1}(u,\varphi) = a^*_{n-i,1}(u_i, \varphi_i), \quad i = 1, \dots, n-1$$

where $u_i = l^*_{n-i+1,1}(u_{i-1})$ and $\varphi_i = \varphi_{i-1}W_0(a^*_{n-i+1,1}(u_{i-1}))/W_0(u_{i-1})$, $i = 2, \dots, n$. The meaning of these equations is that, to solve the overall optimization problem, we "only" need

Input : Loops L, increment r, optimization function P.
Output: Approximation to optimal levels $a^*_{n,k}$, $b^*_{n,k}$ and $l^*_{n,k}$, $k = 1, \ldots, n$.

```
begin
    initialize({(s_k) | k = 1, ..., n});
    (c_1, ..., c_n) := computeC(s_1, ..., s_n);
    l := 1;
    while l < L do
        for (m_1, m_2) in {1, ..., n}^2 do
            m := min(m_1, m_2);
            s̃_k := s_k,  k ≠ m_1, m_2;
            s̃_{m_1} := r s_{m_1};
            s̃_{m_2} := s_{m_2}/r;
            ã_n, b̃_n := P(B, s̃_n);
            l̃_n := W_0^{-1}(s̃_n W_0(ã_n));
            c̃_n := B(ã_n, b̃_n, l̃_n);
            for j := n − 1 to m do
                ã_j, b̃_j := P(B + c̃_{j+1} G, s̃_j);
                l̃_j := W_0^{-1}(s̃_j W_0(ã_j));
                c̃_j := B(ã_j, b̃_j, l̃_j) + c̃_{j+1} G(ã_j, b̃_j, l̃_j);
            end
            if c̃_m > c_m then
                s_k = s̃_k,  k = 1, ..., n;
                (c_1, ..., c_n) := computeC(s_1, ..., s_n);
            end
        end
        l = l + 1;
    end
    a*_{n,n}, b*_{n,n} := P(B, s_n);
    l*_{n,n} := W_0^{-1}(s_n W_0(a*_{n,n}));
    for j := n − 1 to 1 do
        a*_{n,j}, b*_{n,j} := P(B + c_{j+1} G, s_j);
        l*_{n,j} := W_0^{-1}(s_j W_0(a*_{n,j}));
    end
    V = c_1;
end
```

ALGORITHM 5: Corridor level optimization algorithm

to learn to determine the optimal first corridor of any strategy. While it might seem that this leads to a more efficient optimization algorithm, the use of (96) and (97) makes implicit use of full knowledge of the functions $a^*_{i,1}$, $b^*_{i,1}$ and $l^*_{i,1}$ for each $i = 1, \ldots, n-1$ at all values u and φ, which seems infeasible to numerically achieve for a large amount of corridors[2]. One could still use this idea by optimizing over the values of $\varphi_1, \ldots, \varphi_{n-1}, u_{n-1}$, and determining

[2]Recall that there are no explicit solutions for these functions, so this needs to be done numerically.

the remaining quantities by using the optimality properties of $a^*_{n,i}$, $b^*_{n,i}$ and $l^*_{n,i}$. However, this leads precisely to the implementation described in Algorithm 5.

The considerations described so far work equally well for the case of the Cramér-Lundberg model with exponential claims, as evidenced by the results in Section 4.3.2. For more general processes, we lose the ability of expressing l purely in terms of a, as A usually depends on its second argument. If we want to keep the advantage of transforming the constrained optimization problem into a sequence of n unconstrained ones, we can shift the focus by observing that A can be written in the form $A(a, b, l) = I(b, l)/W_0(a)$ for some function I. Hence, for $0 < s < 1$ we can write $a = W_0^{-1}(I(b, l)/s)$ and then replace the first argument in B and G by this function. The limitation here is that the inverse of W_0 is in general not readily available even if explicit formulas for W_0 exist, so that further methods need to be used in this case. We explain another one in the next section.

4.5.2. An evolutionary strategy. Evolutionary strategies (ES) have been applied with some success in reinsurance problems where the evaluation of the function to optimize is only possible through numerical procedures due to the non-existence of explicit algebraic expressions, see e.g. [**103**] and [**100**]. Within the context of optimization of dividend strategies, ES were recently systematically used in [**4**] for tackling the classical (un-constrained) version of the current problem, and we refer to there for a broader discussion of the corresponding algorithms and background. In this work, we explore a suitably adapted strategy for our purposes as well as a penalized version of the algorithm.

As outlined in [**4**], basic ES's are designed for unconstrained search spaces, so in order to enforce the survival probability condition, some adaptations are needed to the way $V^D(u)$ and $\phi^D(u)$ are evaluated. We begin by identifying the current set of strategies of the form (79) as a subset of $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ in the natural way. A point $(a, b, l) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ satisfies the constraints on our dividend problem if and only if:

(i) For every $k = 1, \dots, n$, $0 \le l_k \le b_k \le a_k$,
(ii) For every $k = 1, \dots, n-1$, $a_k \le a_{k+1}$ and
(iii) $\phi^D(u) = \kappa'(0)\, W_0(u) \prod_{k=1}^n A(a_k, b_k, l_k) = \varphi$.

Here $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n)$ and $l = (l_1, \dots, l_n)$. Now, mutations occur coordinate by coordinate by adding a normally distributed error to recombinations of the parental population. If we want to maintain this procedure, then, whenever $a_k - l_k$ is small, constraint (i) will be violated with high probability after this addition, which restricts the way in which mutations can be carried out. As for problems with high values for n and φ, $a_k - l_k$ will be small, and additive mutations of this way are infeasible.

We begin by assuming that A does not depend on b, which as explained in the previous section, happens in the case of the diffusion and the Cramér-Lundberg model with exponential jumps. Thus, by strict monotonicity of W_0, for every $a > 0$, the mapping $l \mapsto A(a, l)$ is strictly increasing. With this in mind, we can use constraint (iii) to come around the limitation from the previous paragraph. Indeed, by defining the changes of variable in the l-space

$$q_1 = A(a_1, l_1)$$

and for $k = 1, \dots, n-1$

$$q_{k+1} = q_k A(a_{k+1}, l_{k+1}),$$

the previous set of constraints is converted to

(i') For every $k = 1, \dots, n-1$, $q_{k+1} \le q_k$ and $q_1 < 1$,
(ii') For every $k = 1, \dots, n-1$, $a_k \le a_{k+1}$ and
(iii') $q_n = \kappa'(0)^{-1}\, W_0(u)^{-1} \varphi$.

Denoting the product $\kappa'(0)^{-1}\, W_0(u)^{-1}\varphi$ by d, we see that these changes of variable allow us to convert condition (iii) into a univariate condition, making clear the reduction to a search in a $3n-1$ dimensional space. Constraint (i') can be summarized by the chain of inequalities $d = q_n < q_{n-1} < \cdots < q_1 < 1$. For high values of φ, the value of d will be close to 1, so, once again, with high probability, any addition of a normally distributed error to the parameters q_k will make them not satisfy constraint (i'). Hence, we apply the final transformation

$$y_k = \Phi^{-1}\left(\frac{q_k - d}{1-d}\right)$$

with Φ the c.d.f. of the normal distribution. With the latter, we see that the only constraints in y_k and a_k are

(i") For every $k = 1, \ldots, n-2$, $y_{k+1} \le y_k$,
(ii") For every $k = 1, \ldots, n-1$, $a_k \le a_{k+1}$.

The new set of constraints (i") and (ii") can easily be handled by sorting the values within the vector and is handled by the ES, similarly to [**4**].

One should keep in mind that this approach is only viable when A does not depend on its second argument. While one can still use strict monotonicity of W_0 to theoretically argue that the map $a \mapsto A(a,b,l)$ is invertible for each fixed b and l, and proceed in a similar manner, in this situation one often runs into the problem that the inverse of W_0 cannot be explicitly identified, thus limiting the applicability of the approach. In these cases, we suggest a more straightforward procedure applying a penalty function to the value function. More specifically, an adaptive penalty function is used in the algorithm, as e.g. described in [**94**]. Then, (79) is replaced by the function

$$\tilde{V}(u) = V(u) - \xi_t 1_{\{\phi^D(u) < \varphi\}}, \tag{102}$$

where $1_{\{\phi^D(u)<\varphi\}}$ denotes the indicator function of the set $\{\phi^D(u) < \varphi\}$ and ξ_t is a parameter that depends on the generation t. This parameter is initially chosen larger than the overall optimal strategy, so that levels a_k, b_k, l_k for which $\phi^D(u) < \varphi$ produce a negative value. Moreover, it is updated for every generation according to the rule

$$\xi_{t+1} = \begin{cases} c_1\xi_t & \text{if best candidate satisfies } \phi^D(u) > \varphi \text{ for } k \text{ generations,} \\ c_2\xi_t & \text{if best candidate satisfies } \phi^D(u) < \varphi \text{ for } k \text{ generations,} \\ \xi_t & \text{otherwise,} \end{cases}$$

where $c_1 < 1$, $c_2 > 1$ and k are predetermined parameters. We require $c_1 c_2 \neq 1$ to avoid circularity.

While the optimal set of levels satisfies $\phi^D(u) = \varphi$, we cannot replace the indicator function in (102) by $1_{\{\phi^D(u)=\varphi\}}$, since, by the nature of the algorithm, with probability zero the new candidates will be in the set $\{\phi^D(u) = \varphi\}$, so that using this indicator function would instead produce the overall best strategy of the form (79) without regard to survival probability.

4.6. Numerical Results

We examine the performance of the strategy for four Lévy processes: the diffusion, the Cramér-Lundberg model with exponential and Erlang claims, as well as the perturbed Cramér-Lundberg model with exponential claims. More specifically, for the diffusion model

$$C_t = u + \mu t + \sigma B_t, \quad t \ge 0,$$

we consider $\mu = 1$, $\sigma = 1$, $\varphi = 0.95$ and $\delta = 0.03$ as well as $\mu = 0.04$, $\sigma = \sqrt{0.02}$, $\varphi = 0.95$ and $\delta = 0.02$. For the case of the Cramér-Lundberg process

$$C_t = u + c\,t - \sum_{i=1}^{N_t} X_i, \quad t \geq 0,$$

we choose $X_i \sim \mathrm{Exp}(1000)$, a Poisson intensity $\lambda = 5 \times 10^5$, $c = 501$, $\delta = 0.03$ and $\varphi = 0.95$; as well as $X_i \sim \mathrm{Erlang}(2, 1)$, a Poisson intensity $\lambda = 10$, $c = 21.4$, $\delta = 0.03$ and $\varphi = 0.1$. Finally, for the perturbed Cramér-Lundberg process,

$$C_t = u + \mu t + \sigma B_t - \sum_{i=1}^{N_t} X_{i,}, \quad t \geq 0,$$

we consider $\mu = 1$, $\sigma = 1$, $\varphi = 0.7$, $X_i \sim \mathrm{Exp}(2)$, a Poisson intensity $\lambda = 1$ and $\delta = 0.03$.

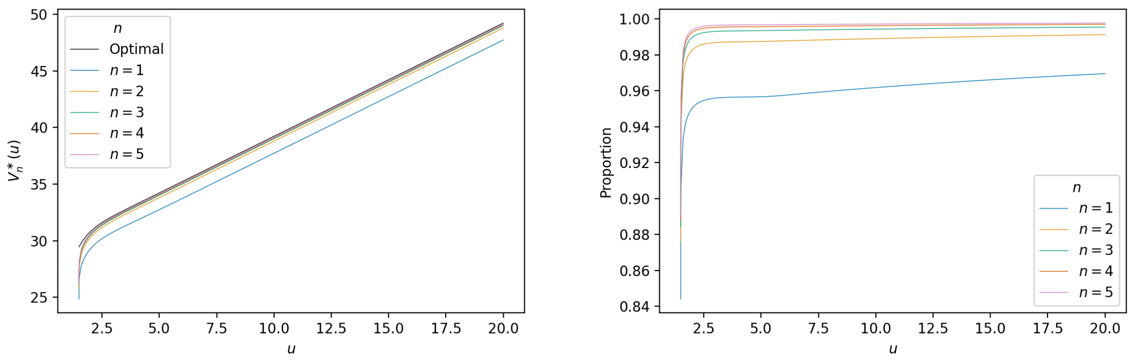
While the concrete choice of parameters for the perturbed Cramér-Lundberg process and the first set of parameters for the diffusion is somewhat arbitrary, the second set for the diffusion was chosen in accordance with the parameters used in [**67**]. The Cramér-Lundberg process with exponential claims is chosen with a high Poisson intensity λ and a small expectation in such a way that it approximates the diffusion case with $\mu = 1$ and $\sigma = 1$. The parameter choice for the Cramér-Lundberg process with Erlang claims is taken from [**22**], for which the overall (unconstrained) optimal dividend strategy is known to be a two-band strategy (as opposed to a barrier strategy like in the other three cases). The relevant quantities for the diffusion and Cramér-Lundberg process with exponential claims are evaluated through the formulas obtained in Section 4.3.1 and 4.3.2, while the formulas for the Cramér-Lundberg process with Erlang claims and perturbed Cramér-Lundberg process are obtained through the more general formulas from earlier in Section 4.3. The results are shown in Figures 2–7. In all these cases, and for each relevant u, there was only one optimal strategy D making $V_n^*(u) = V^D(u)$, which allows us to present the information in terms of the functions $a_{n,k}^*$, $b_{n,k}^*$ and $l_{n,k}^*$ as in Figures 5, 6 and 7. The results in Figure 5 are shown only in the case of the diffusion, since the resulting plots for the other processes are similar and we decided to omit them for the sake of brevity.

We start by observing that the performance of the strategy seems to depend rather strongly on the type of process and the concrete parameters. The values for different n converge rather fast and uniformly to a limiting function $\lim_{n\to\infty} V_n^*$, which is why we only display the results for small n. In Figure 2B, we see that even for values of u close to $\phi^{-1}(0.95) \approx 1.497866$, the process achieves about 90% of the value of the unconstrained optimal strategy (which for the diffusion is a barrier strategy). This is quite remarkable, given that the optimal barrier strategy has a survival probability of zero. Figure 3 shows the corresponding plots for the second set of parameters in the diffusion case. Here the convergence to the solution for large n is slower, and only about 80% of the unconstrained optimal value is achieved, which is nevertheless still noteworthy. Recall that this figure shows the results obtained for the parameters that were also used in the numerical experiments of [**67**] who considered the optimal dividend problem up to a finite time horizon T with a ruin probability constraint and a potential lump sum payout of the remaining surplus at T. Within the current framework one can not directly compare the results from Figure 3 to those obtained in that paper, but a small adaptation adding a lump sum dividend payment at T makes the comparison possible: using D_n^*, we can compute

$$V_n(u, T) = \mathbb{E}\left(\int_0^{\tau\wedge T} e^{-\delta t} dD_t^* + e^{-\delta(\tau\wedge T)} C_{\tau\wedge T}^{D^*}\right) \tag{103}$$

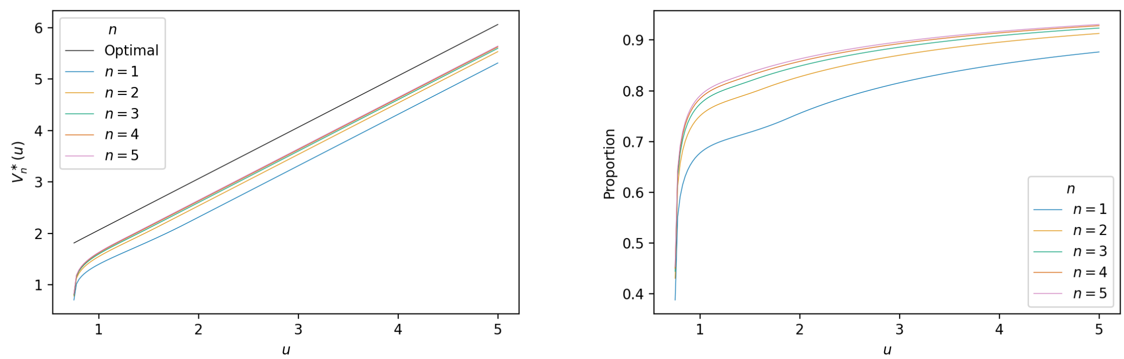
where $\tau = \inf\{t > 0 : C^{D^*} < 0\}$ and D^* is the stochastic process associated with using the strategy represented by $D_n^*(u)$. For $n = 10$, $T = 10$ and $u = 1$, we computed the expectation in (103) through MC simulation using a sample of 100,000 simulations and approximating the diffusion through a random walk with drift and a step size of 10^{-5} time units. These simulations provide an approximating value of 1.158832, which is in fact considerably higher than the value 0.20879 reported in [**67**].

Figure 4 shows that for the Cramér-Lundberg process with Erlang claims, the performance is similar. Recall from Proposition 4.4.5 that for any $n \geq 1$, $\lim_{u\to\infty} V_n^*(u)/\bar{V}(u) = 1$, where $\bar{V}$ denotes the unconstrained optimal value function. Figures 5, 6 and 7 show the evolution



(A) V_n^* as a function of initial surplus. (B) V_n^* relative to the best strategy.

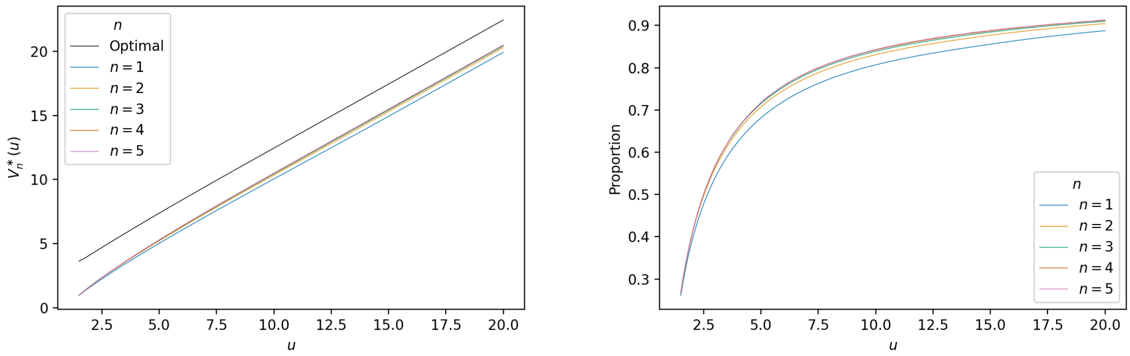
FIGURE 2. $V_n^*(u)$ for $n = 1, 2, 3, 4$ and 5 in absolute terms (left), and relative to the unconstrained optimal dividend strategy (that is also given on the left) for the diffusion with parameters $\mu = 1$, $\sigma = 1$, $\varphi = 0.95$ and $\delta = 0.03$ (right).



(A) V_n^* as a function of initial surplus. (B) V_n^* relative to the best strategy.

FIGURE 3. $V_n^*(u)$ for $n = 1, 2, 3, 4$ and 5 in absolute terms (left) and relative to the unconstrained optimal dividend strategy (that is also given on the left) for the diffusion with parameters $\mu = 0.04$, $\sigma = \sqrt{0.02}$, $\varphi = 0.95$ and $\delta = 0.02$ (right).

of the optimal strategies D_n^* as a function of u and n for the diffusion and Cramér-Lundberg process with exponential claims, respectively. Due to the imposition $a_1^D \geq u$ made before Proposition 4.4.6 to ensure uniqueness of the strategies, for values of u large enough we will



(A) V_n^* as a function of initial surplus. (B) V_n^* relative to the best strategy.

FIGURE 4. $V_n^*(u)$ for $n = 1, 2, 3, 4$ and 5 in absolute terms (left), and relative to the unconstrained optimal dividend strategy (that is also given on the left) for the Cramér-Lundberg process with Erlang claims (right).

have $a_{n,1}^*(u) = u$, which is reflected in all the plots of Figure 5. Moreover, the conclusion of Proposition 4.4.5 can be traced in Figures 5B and 5C, where, again, for large enough u, all levels but $a_{n,1}^*$ become constant. Finally, it seems that we have the general trends $a_{n+1,k}^* \leq a_{n,k}^*$ and $b_{n+1,k}^* \leq b_{n,k}^*$ for $k \leq n$, as well as $a_{n,k}^* \leq a_{n,k+1}^*$, $b_{n,k}^* \leq b_{n,k+1}^*$ and $l_{n,k}^* \leq l_{n,k+1}^*$, which simply means that $a_{n,k}^*$ and $b_{n,k}^*$ are decreasing in n, but increasing in k, while $l_{n,k}^*$ is only increasing in k as shown in Figures 6 and 7. While this can not be proven explicitly with the current means, it is somewhat intuitive, at least for the last observation: after a corridor $a_{n,k}^*, b_{n,k}^*$ and $l_{n,k}^*$, it seems optimal to wait for a level higher than $a_{n,k}^*$ to start the new corridor $a_{n+1,k}^*$, for if $a_{n+1,k}^* \leq a_{n,k}^*$ we could have exchanged the order of the corridors, which would imply paying dividends earlier and hence on average increasing the amount of dividends paid without changing the overall survival probability.

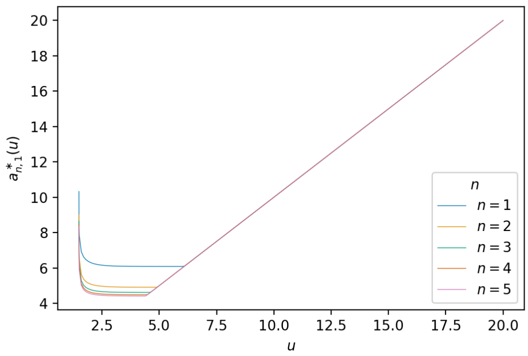
Finally, we would like to comment two further details about the implementation: first, all processes except the diffusion exhibited local maxima around the points where $b_k = l_k$ for some index k. Hence, for these particular cases, the algorithms described in Section 4.5 were not applied exactly as described there but with the extra condition that $l_k < rb_k$ for some $0 < r < 1$, generally $r \approx 0.95$, which seemed to produce more adequate results.

Second, while the gradient equations derived at the beginning of Section 4.5 are hard to deal with, they produce an interesting equation:

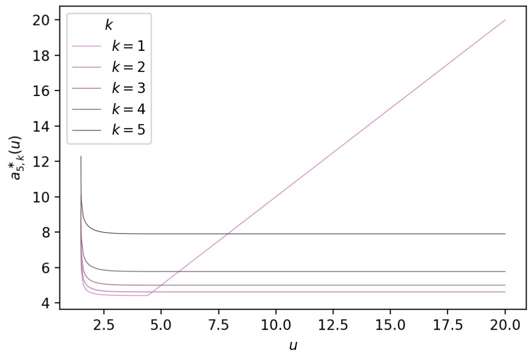
$$W_\delta(b_n - l_n) = 0.$$

Combined with the form of B, one can then deduce that the optimal distance between $b_{n,n}^*$ and $l_{n,n}^*$ equals the barrier level of the optimal barrier strategy with initial capital $a_{n,n}^* - l_{n,n}^*$ (which in most cases equals the overall unconstrained dividend strategy). This observation was used, for example, to solve the first step in Algorithm 5 or as another check for convergence of the ES.

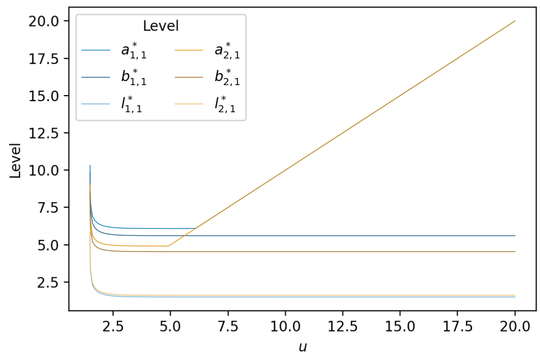
4.6.1. Asymptotic behaviour of barrier levels. We will restrict ourselves now to the case of the diffusion and fix the parameters to $\mu = 0.04$, $\sigma = \sqrt{0.02}$, $\varphi = 0.95$, $\delta = 0.02$ and initial surplus $u = 2$. Figure 8 displays the optimal barrier levels for large values of n (80, 100, and 150) computed by means of Algorithm 5. Together with Figures 6 and 7, the results show a sort of common behavior both at the "middle" levels as well as in the last barriers. Figure 9 displays the distances between $a_{n,k}^*$, $b_{n,k}^*$ and $l_{n,k}^*$ for $n = 30$ and $n = 80$ for the last 30 bands,



(A) $a^*_{n,1}$ as function of initial surplus.



(B) $a^*_{5,k}$ as function of initial surplus.



(C) Comparison of levels for $n = 1, 2$.

FIGURE 5. Plot of the change of the optimal levels $a^*_{n,k}$, $b^*_{n,k}$, $l^*_{n,k}$ for different n's and k's as a function of initial surplus for diffusion with parameters $\mu = 1$, $\sigma = 1$, $\varphi = 0.95$ and $\delta = 0.03$.

as well as the distances between $a^*_{n,n}$ and $a^*_{n,k}$, with a shift in the indices of $n = 80$ to match the last bands. We notice that except for small values of k, the distances are extremely close (one even does not see the difference visually).

These observations suggest that the distances between the barrier levels converge in the final barriers. Explicitly, the results suggest that the limits

$$
\begin{aligned}
&\lim_{n\to\infty} a^*_{n,n-M+1} - l^*_{n,n-M+1},\\
&\lim_{n\to\infty} b^*_{n,n-M+1} - l^*_{n,n-M+1},\\
&\lim_{n\to\infty} a^*_{n,n} - a^*_{n,n-M+1},
\end{aligned}
\tag{104}
$$

exist for $M \in \mathbb{N}$ (and that in turn several other limits pertaining to the distances among the levels also exist, e.g., the existence of $\lim_{n\to\infty} a^*_{n,n-M+2} - a^*_{n,n-M+1}$). In the following, we assume that the limits in (104) indeed exist and denote them by ζ_M, η_M and ν_M respectively.

In fact, from the Lagrange equations it was observed already earlier that $b^*_{n,n} - l^*_{n,n}$ should always equal the (surplus-independent) optimal barrier level for the diffusion, giving the exact value for η_1. Motivated by this, we use the Lagrange equations to deduce further properties of ζ_M, η_M and ν_M.

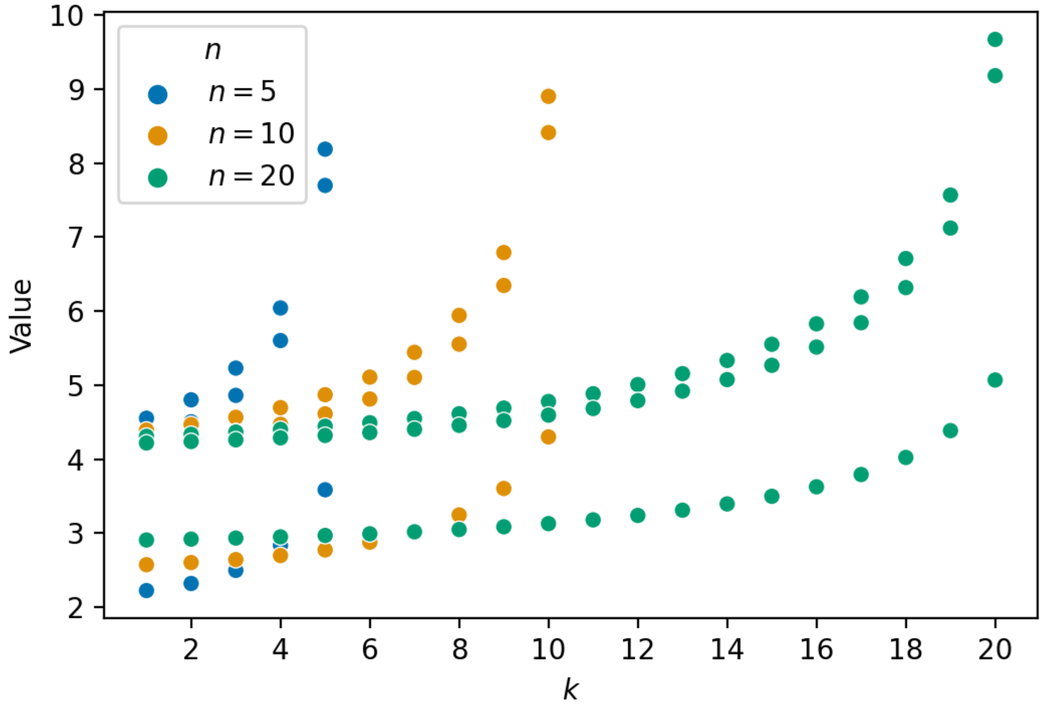


FIGURE 6. Plot of the change of the optimal levels $a^*_{n,k}$, $b^*_{n,k}$, $l^*_{n,k}$ for different n's and k's as a function of k for the Cramér-Lundberg process with exponential claims and initial surplus $u = 2$.

Recall the normal equations (100), as well as the definitions of A, B and G, here specialized for the diffusion:

$$\begin{aligned} A(a,l) &= W_0(l)/W_0(a), \\ B(a,b,l) &= \frac{a-b+W_\delta(b-l)/W'_\delta(b-l)}{W_\delta(a)}, \\ G(a,b,l) &= \frac{\sigma^2}{2}\frac{W_\delta(l)}{W_\delta(a)}\left(W'_\delta(b-l) - \frac{W_\delta(b-l)W''_\delta(b-l)}{W'_\delta(b-l)}\right). \end{aligned}$$

For $m = n$ we obtain, from (100) (with $i = 1$ and $i = 3$),

$$\frac{D_1B(a^*_{n,n}, b^*_{n,n}, l^*_{n,n})}{D_3B(a^*_{n,n}, b^*_{n,n}, l^*_{n,n})} - \frac{D_1A(a^*_{n,n}, , l^*_{n,n})}{D_2A(a^*_{n,n}, l^*_{n,n})} = 0$$

Using the explicit forms of A and B we obtain

$$\frac{-W_\delta(a^*_{n,n}) + W'_\delta(a^*_{n,n})\left(a^*_{n,n} - b^*_{n,n} + \frac{W_\delta(\eta_1)}{W_{\delta'(\eta_1)}}\right)}{W_\delta(a^*_{n,n})} + \frac{W_0(l^*_{n,n})W'_0(a^*_{n,n})}{W'_0(l^*_{n,n})W_0(a^*_{n,n})} = 0.$$

If we assume that $a^*_{n,n} \to \infty$ as $n \to \infty$, we obtain

$$-1 + \theta_1\left(\zeta_1 - \eta_1 + \frac{W_\delta(\eta_1)}{W'_\delta(\eta_1)}\right) + e^{-2\mu\zeta_1/\sigma^2} = 0. \tag{105}$$

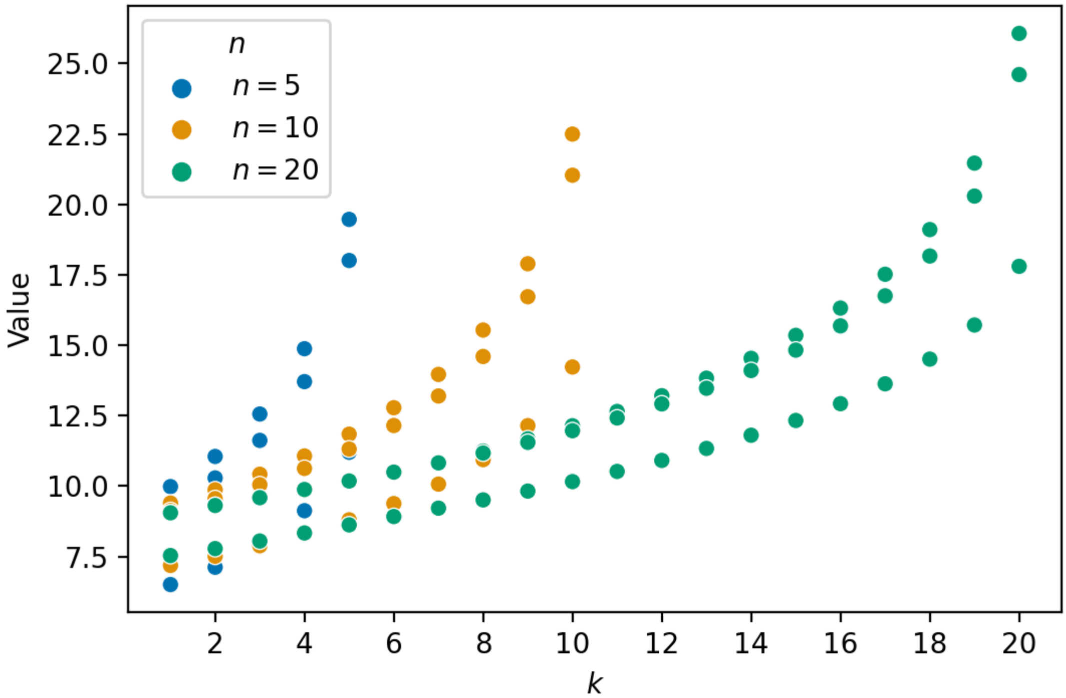


FIGURE 7. Plot of the change of the optimal levels $a^*_{n,k}$, $b^*_{n,k}$, $l^*_{n,k}$ for different n's and k's as a function of k for the perturbed CL process with initial surplus $u = 2$.

This is an implicit equation for ζ_1 which can be easily (numerically) solved given the value of η_1. While the derivation of this equation relied on the (somewhat mild) assumptions made along the way, the value obtained from solving (105) is indeed pleasantly close to the value obtained after computing $a^*_{80,80} - l^*_{80,80}$ with the previous numerical means (1.137077040 against 1.137077050).

Following this line of thought, one could try to obtain simpler equations for ζ_M, η_M and ν_M for $M \geq 1$. Indeed, assuming that (100) is always satisfied, we obtain for $(r, s) \in \{(1,1),(3,2)\}$ and for $m = n - M + 1$,

(106)

$$\Lambda\kappa'(0)W_0(u)\varphi = \frac{W_\delta(u)D_rB_{n-M+1}\prod_{j=1}^{n-M}G_j + W_\delta(u)D_rG_{n-M+1}\sum_{k=n-M+2}^{n}B_k\prod_{\substack{j=1\\ j\neq n-M+1}}^{k-1}G_j}{D_sA_{n-M+1}A_{n-M+1}}$$

with

$$\begin{aligned} B_k &= B(a^*_{n,k}, b^*_{n,k}, l^*_{n,k}), \\ D_rB_k &= D_rB(a^*_{n,k}, b^*_{n,k}, l^*_{n,k}) \end{aligned}$$

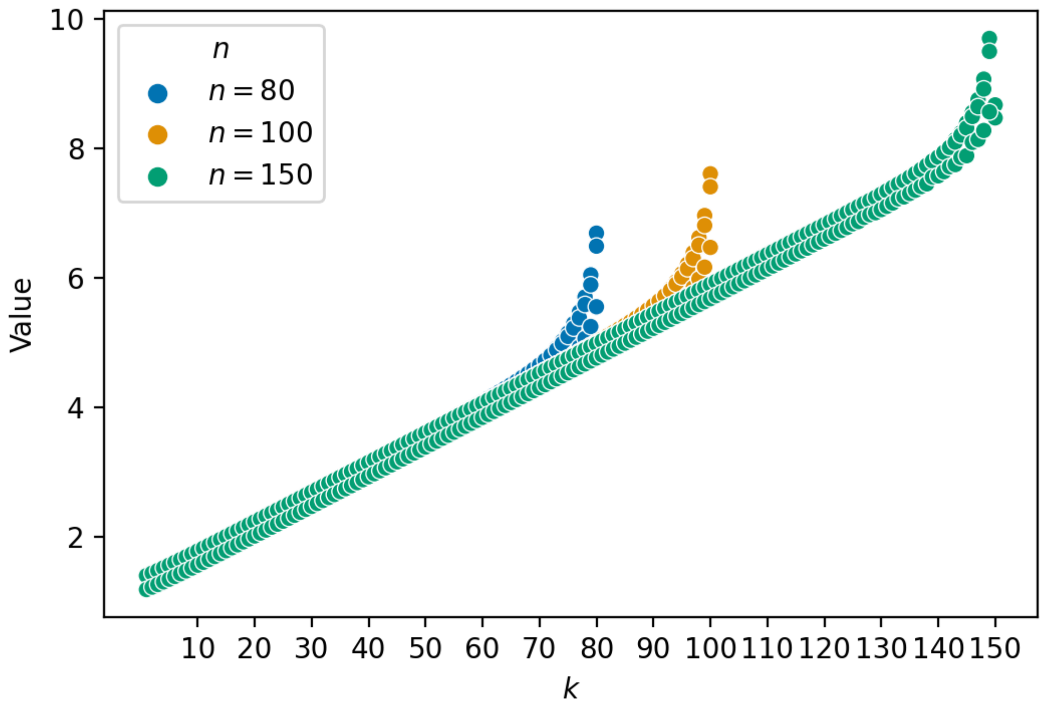


FIGURE 8. Plot of the change of the optimal levels $a^*_{n,k}$, $b^*_{n,k}$, $l^*_{n,k}$ for large n's and k's as a function of k for the diffusion with parameters $\mu = 0.04$, $\sigma = \sqrt{0.02}$, $\varphi = 0.95$, $\delta = 0.02$ and initial surplus $u = 2$.

and similarly for A and G. Plugging (106) in (100) with $m = n - M$, we obtain, after cancellation of common factors,

(107)

$$D_r B_{n-M} + D_r G_{n-M} \sum_{k=n-M+1}^{n} B_k \prod_{j=n-M+1}^{k-1} G_j -$$

$$\frac{D_s A_{n-M} A_{n-M+1}}{D_s A_{n-M+1} A_{n-M}} \left(D_r B_{n-M+1} G_{n-M} + D_r G_{n-M+1} G_{n-M} \sum_{k=n-M+2}^{n} B_k \prod_{j=n-M+2}^{k-1} G_j \right) = 0.$$

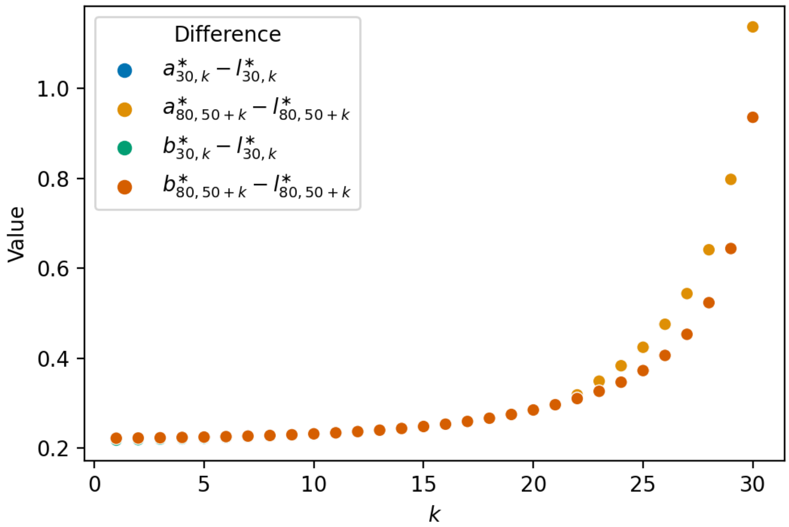
Now, observe that the left hand side of (107) converges to zero as $n \to \infty$. However, after multiplication by $W_\delta(a^*_{n,n})$, we can take the limit, obtaining a new set of equations:

(108)

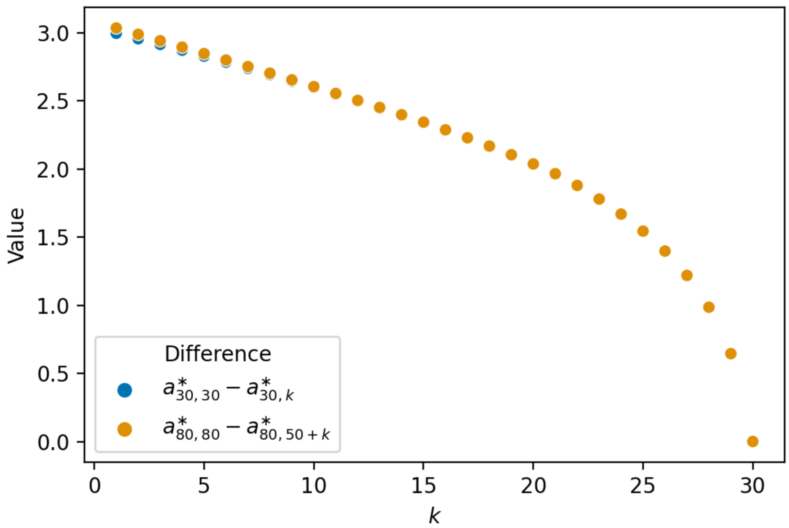
$$\tilde{B}_1(\zeta_{M+1}, \eta_{M+1}, \nu_{M+1}) + \tilde{G}_1(\zeta_{M+1}, \eta_{M+1}) \sum_{k=1}^{M} \tilde{B}(\zeta_k, \eta_k, \nu_k) \prod_{j=k+1}^{M} \tilde{G}(\zeta_j, \eta_j)$$

$$- e^{-2\mu(\nu_{M+1}-\nu_M)/\sigma^2} \tilde{B}_1(\zeta_M, \eta_M, \nu_M) \tilde{G}_1(\zeta_{M+1}, \eta_{M+1})$$

$$- e^{-2\mu(\nu_{M+1}-\nu_M)/\sigma^2} \tilde{G}_1(\zeta_M, \eta_M) \tilde{G}(\zeta_{M+1}, \eta_{M+1}) \sum_{k=1}^{M-1} \tilde{B}(\zeta_k, \eta_k, \nu_k) \prod_{j=k+1}^{M-1} \tilde{G}(\zeta_j, \eta_j) = 0,$$



(A) Differences $a^*_{n,k} - l^*_{n,k}$ and $b^*_{n,k} - l^*_{n,k}$ for $n = 30$ and $n = 80$.



(B) Differences $a^*_{n,n} - a^*_{n,k}$ for $n = 30$ and $n = 80$.

FIGURE 9. Plot of the differences between the optimal levels of the last bands. For $n = 80$, there is a shift by of 50 to match the indices with those of $n = 30$.

$$\tilde{B}_2(\eta_{M+1}, \nu_{M+1}) + \tilde{G}_2(\zeta_{M+1}, \eta_{M+1}) \sum_{k=1}^{M} \tilde{B}(\zeta_k, \eta_k, \nu_k) \prod_{j=k+1}^{M} \tilde{G}(\zeta_j, \eta_j) = 0 \tag{109}$$

(110)

$$\tilde{B}_3(\eta_{M+1},\nu_{M+1}) + \tilde{G}_3(\zeta_{M+1},\eta_{M+1})\sum_{k=1}^{M}\tilde{B}(\zeta_k,\eta_k,\nu_k)\prod_{j=k+1}^{M}\tilde{G}(\zeta_j,\eta_j)$$
$$- e^{-2\mu(\zeta_{M+1}+\nu_{M+1}-\zeta_M-\nu_M)/\sigma^2}\tilde{B}_3(\eta_M,\nu_M)\tilde{G}_1(\zeta_{M+1},\eta_{M+1})$$
$$- e^{-2\mu(\zeta_{M+1}+\nu_{M+1}-\zeta_M-\nu_M)/\sigma^2}\tilde{G}_3(\zeta_M,\eta_M)\tilde{G}(\zeta_{M+1},\eta_{M+1})\sum_{k=1}^{M-1}\tilde{B}(\zeta_k,\eta_k,\nu_k)\prod_{j=k+1}^{M-1}\tilde{G}(\zeta_j,\eta_j) = 0$$

with

$$\begin{aligned}
\tilde{B}(\zeta,\eta,\nu) &= e^{\theta_1\nu}\left(\zeta-\eta+\frac{W_\delta(\eta)}{W_\delta'(\eta)}\right),\\
\tilde{B}_1(\zeta,\eta,\nu) &= e^{\theta_1\nu}-\theta_1\tilde{B}(\zeta,\eta,\nu),\\
\tilde{B}_2(\eta,\nu) &= -e^{\theta_1\nu}\frac{W_\delta(\eta)W_\delta''(\eta)}{W_\delta'(\eta)^2},\\
\tilde{B}_3(\eta,\nu) &= -\tilde{B}_2(\eta,\nu)-e^{\theta_1\nu}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{G}(\zeta,\eta) &= \frac{\sigma^2}{2}e^{-\theta_1\zeta}\left(W_\delta'(\eta)-\frac{W_\delta(\eta)W_\delta''(\eta)}{W_\delta'(\eta)}\right),\\
\tilde{G}_1(\zeta,\eta,) &= -\theta_1\tilde{G}(\eta,\nu),\\
\tilde{G}_2(\zeta,\eta) &= \frac{\sigma^2}{2}e^{-\theta_1\zeta}\left(\frac{W_\delta(\eta)W_\delta''(\eta)^2}{W_\delta'(\eta)^2}-\frac{W_\delta(\eta)W_\delta'''(\eta)}{W_\delta'(\eta)}\right),\\
\tilde{G}_3(\zeta,\eta) &= \theta_1\tilde{G}(\eta,\nu)-\tilde{G}_2(\eta,\nu).
\end{aligned}$$

These functions can be obtained from W_δ, B, G and their derivatives. Hence, we have

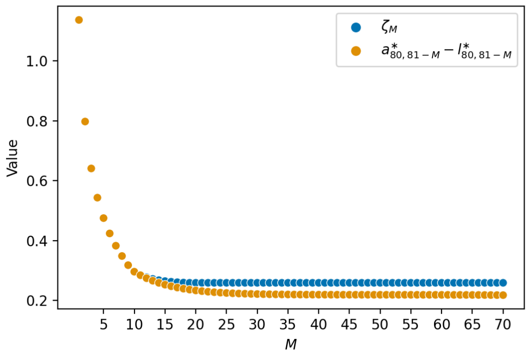
$$\tilde{B}(\zeta,\eta,\nu) = \lim_{x\to\infty} W_\delta(x+\nu)B(x,x+\eta-\zeta,x-\zeta)$$

and similarly for $\tilde{B}_1$, $\tilde{B}_2$ and $\tilde{B}_3$ using D_1B, D_2B and D_3B respectively. Likewise, we have

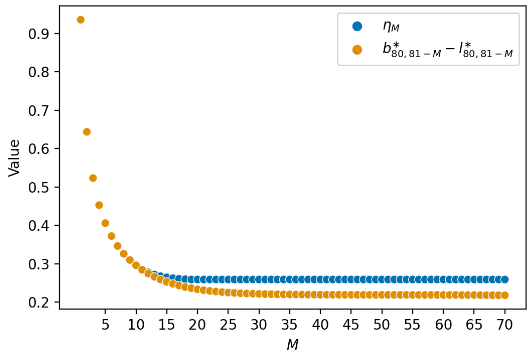
$$\tilde{G}(\zeta,\eta) = \lim_{x\to\infty} G(x+\zeta,x+\eta,x)$$

and similarly for $\tilde{G}_1$, $\tilde{G}_2$ and $\tilde{G}_3$.
Observe that Equations (108) to (110) represent an improvement over the Lagrange equations as by taking the limit we reduce the dimensionality of the problem by eliminating the Lagrange multiplier. Moreover, these equations allow for a truly recursive algorithm since they show that the values of ζ_{M+1}, η_{M+1} and ν_{M+1} depend only on the previous values, and ζ_1 and η_1 can be obtained independently from the equations displayed before.

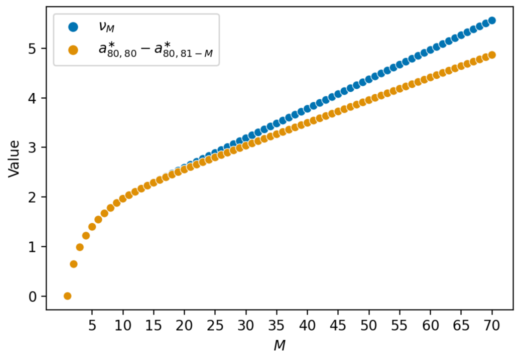
Derivation of (108), (109) and (110) relied on the fact that the gradient equations (100) are always satisfied, which will happen if and only if the optimal strategy is in the interior of $\mathscr{D}_n$. However, as seen by the numerical examples considered before, this is in general not the case as, for example, with the current parameters, one has $a^*_{n,1} = b^*_{n,1}$ for n large enough. This will be reflected in a way that there will exist a minimal M_1 such (108) to (110) will not have a "sensitive" solution (e.g., they will only have a solution with negative values). Numerical experiments for the diffusion show, however, that for $M \geq M_1$ one can simply assume $\eta_M = \zeta_M$ and replace (108) and (109) by the equation obtained after adding their left hand sides (which can be derived from a Lagrange equation after assuming $a^*_{n,n-M+1} = b^*_{n,n-M+1}$), thus obtaining a system of two equations with two unknowns. Curiously enough, the case of the diffusion also shows that there might exist a (minimal) $M_2 \geq M_1$ such that $\eta_M = \zeta_M = \eta_{M_2}$ for all $M \geq M_2$. In this case, one is only left with Equation (110) to obtain the successive values of ν_M, which produces the "linear" behavior observed for the "middle" barriers in Figure 9. Figure 10 shows the results of comparing the results from this procedure with the differences between barrier levels for $n = 80$, where one observes that for these parameters one has $M_1 = 10$ and $M_2 = 19$.



(A) Comparison between $a^*_{80,81-M} - l^*_{80,81-M}$ and ζ_M for $M = 1, \dots, 30$.



(B) Comparison between $b^*_{80,81-M} - l^*_{80,81-M}$ and η_M for $M = 1, \dots, 30$.



(C) Comparison between $a^*_{80,80} - a^*_{80,81-M}$ and ν_M for $M = 1, \dots, 30$.

FIGURE 10. Comparison between the limits in (104) and the distances between barrier levels for $n = 80$.

Now, while the divergence of $a^*_{n,n}$ to infinity implies that a limit strategy does not exist, it might still be useful to compute the values of ζ_k, η_k and ν_k for $k = 1, \dots, M$ for a large M: as explained before (cf. Section 4.5), for large n, the algorithms have difficulties finding the optimal levels unless an appropriate set of values is provided for initialization. Since the previous figures indicate that convergence of the limits happens relatively fast, one might want to use the following pseudo-algorithm (Algorithm 6) to approximate the values of $a^*_{n,k}$, $b^*_{n,k}$, $l^*_{n,k}$, $k = 1, \dots, n$ for large n. The idea is to suppose that k is large enough so that convergence of the limits in (104) is already achieved (or close enough to be achieved) and hence one only needs to optimize over $3k+1$ variables instead of $3n$. Moreover, the constraint on the survival probability might be used to find a suitable value for a_n and hence the `initialize` and `improve` functions on Lines 4 and 12 can be thought as determined by this condition. This

Input : Large n and $M < n$ such that the problem with $k := M - n$ barriers can easily be solved (for example, $k = 30$).
Output: Approximation to optimal levels $a^*_{n,k}$, $b^*_{n,k}$ and $l^*_{n,k}$, $k = 1, \dots, n$.

begin
  $(\zeta_1, \eta_1, \nu_1, \dots, \zeta_M, \eta_M, \nu_M) :=$ `computeLimits(`M`);`
  `initialize(`$\{(a_k, b_k, l_k) \mid k = 1, \dots, n - M\}$`);`
  `initialize(`a_n`);`
  for $j := n - M + 1$ **to** n **do**
    $a_j := a_n - \nu_{n-j+1}$;
    $b_j := a_j - \zeta_{n-j+1} + \eta_{n-j+1}$;
    $l_j := a_j - \zeta_{n-j+1}$;
  end
  while *not convergence* **do**
    `improve(`$\{(a_k, b_k, l_k) \mid k = 1, \dots, n - M\}$`);`
    `improve(`a_n`);`
    for $j := n - M + 1$ **to** n **do**
      $a_j := a_n - \nu_{n-j+1}$;
      $b_j := a_j - \zeta_{n-j+1} + \eta_{n-j+1}$;
      $l_j := a_j - \zeta_{n-j+1}$;
    end
  end
end

ALGORITHM 6: Corridor level optimization pseudo-algorithm

pseudo-algorithm is not restricted to work only with the barrier levels but can be adapted to the step-wise survival probabilities instead (the $s^*_{n,k}$'s described in Section 4.5).

4.7. Conclusions and further remarks

In this paper we proposed a new kind of dividend strategies which naturally generalize classical barrier strategies, but have the advantage of being adjusted to control for survival probability. While the performance of these strategies turns out to be process- and parameter-dependent, their nature has an easy interpretation and – as observed in the illustrations – typically only a few parameters are needed to reach a remarkable resulting survival probability while not losing much efficiency. It is rather surprising that the performance of these strategies turns out to be typically outstanding, giving comparable overall expected dividends payments obtained in the unconstrained problem while respecting restrictive ruin probability constraints. Correspondingly, these strategies can serve as benchmarks for further studies of the constrained optimization problem.

Much like in the passage from barrier to band strategies, one can think of a further generalization of the strategies proposed here: given sequences of surplus levels $a_1, a_2, a_3, \dots, a_n$ and $l_1, l_2, l_3, \dots, l_n$ with $l_n \leq \min(a_n, a_{n+1})$ and stopping times τ_k as defined in Section 4.2, at time τ_k, one can proceed to pay dividends according to an r-band strategy up until the time when the process controlled in this way reaches the lower limit l_k. It is easy to see that in this case the formula for V is similar to (80), but D_k is replaced by the relevant value. We could go yet one step further and allow the number of bands at each corridor to vary, however, since locally in the time interval $[\tau_k, \tau^d_k]$ the process behaves exactly like a controlled process for a normal band strategy, it seems a priori better to keep the number of bands constant and equal to the number of bands of the band strategy that produces the overall best dividend-payment strategy. Since the dimensionality and complexity of the formulas for this generalization increases greatly with the number of bands considered, we preferred in the present paper to adhere to the simpler case of barrier strategies in each corridor. A further difficulty arises when trying to generalize the results of Section 4.4, as it is well known that for general band strategies, the value function is not necessarily continuous.

It will also be interesting in future research to investigate further theoretical properties for the proposed strategies. For example, it is not clear whether the weaker form of the dynamic programming principle as discussed in Remark 4.4.7 guides the value of V_n^* or $\lim_{n\to\infty} V_n^*$, which could help to prove optimality properties of this kind of strategies.

CHAPTER 5

Optimal reinsurance from an optimal transport perspective

This chapter is based on the following article:

B. Acciaio, H. Albrecher and B. Garcia Flores. Optimal reinsurance from an optimal transport perspective. *Preprint.*

Abstract. We regard the optimal reinsurance problem as an iterated optimal transport problem between a (known) initial and an (unknown) resulting risk exposure of the insurer. We also provide conditions that allow to characterize the support of optimal treaties, and show how this can be used to deduce the shape of the optimal contract, reducing the task to a finite-dimensional optimization problem, for which standard techniques can be applied. The proposed approach provides a general framework that encompasses many reinsurance problems, which we illustrate in several concrete examples, providing alternative proofs of classical optimal reinsurance results as well as establishing new optimality results, some of which contain optimal treaties that involve external randomness.

5.1. Introduction

The identification of optimal reinsurance forms is a classical problem of actuarial risk theory. Starting with pioneering work of de Finetti [**53**], Borch [**37, 38**] and Arrow [**13**], with varying objective functions, constraints and choice of involved contract parties, the topic has been a rich source of interesting mathematical problems and is still a very active field of research, see for instance [**44, 1**] for an overview. Among the many conceptual and influential contributions to the topic over the last years, we mention here Kaluszka [**79**], Cai & Tan [**42**], Balbas et al. [**25, 26**] as well as Cheung et al. [**45, 48, 47**]. For game-theoretic approaches to equilibria and efficient solutions, see e.g. [**16, 33, 117, 35**]. For a situation with several reinsurers, see Boonen & Ghossoub [**34**]. An interesting link of optimal reinsurance problems to the Neyman-Pearson lemma of statistical hypothesis testing was established by Lo [**88**], encompassing earlier contributions such as [**118**]; see [**46**] for a recent considerable generalization of this approach. For a backward-forward optimization procedure in a rather general setting, we refer to Boonen & Jiang [**36**].

An intuitive practical constraint when looking for an optimal reinsurance contract is its deterministic nature, i.e., the reinsured amount is identified deterministically once the claim size of the first line insurance company is known. At the same time, one may imagine scenarios where a randomized contract (a contract where beside the original claim size, an additional exogeneous random mechanism is used to determine the eventual reinsured amount) leads in fact to a better solution of the original optimization problem. That is, once the realizations of the original claim sizes are available, that defined additional random mechanism is then used to determine the eventual reinsured and retained amount. For instance, the proportionality factor in a proportional reinsurance contract, or the retention in an Excess-of-Loss contract, may be determined by the outcome of a random experiment with a defined distribution, which might practically be realized with the help of a lottery, or a random number simulator that all involved parties agree upon and that could be applied in the presence of a notary. While it is not straightforward to overcome the psychological barriers of such a contract formulation, if the objective and constraints of the optimization problem match the true goals of the involved parties, such a randomized solution should be preferable in case it dominates all the deterministic solutions (see [3] for an extensive discussion). Gajek & Zagrodny [**60**] were the first to point out such a randomized solution in a particular setup with a discrete loss distribution, where the goal was to minimize the ruin probability of the insurer for a given budget constraint. In [3], [15] and [114] it was then shown how random reinsurance treaties can be optimal in more general situations, with problem formulations that are closer to actuarial practice. In fact, even if an insurer prefers to restrict the risk management to purely deterministic reinsurance contracts, there often is an additional random element present in any case, such as the reinsurance counterparty default risk, see e.g. [**14, 41**]. From a purely mathematical perspective, Guerra & Centeno [**69**] looked for optimal reinsurance contracts in a specific situation, found solutions of a potentially randomized form for their particular one-dimensional model setup, and showed that there always exists an optimal nonrandom treaty, so that their randomization feature could be interpreted merely as a mathematical tool. In a certain way, parts of the present work can also be seen as a generalization of their approach to arbitrary risk measures in a multi-dimensional setting, where we allow for constraints on the solutions.

When it comes to interpreting the existence of a reinsurance contract for a portfolio of insurance contracts, one can view its effect as a reshaping of the loss distribution of the first-line insurer (into a "safer" one) for a particular cost, namely the reinsurance premium. From this perspective, one may then ask the question what the cheapest way to "couple" the original risk distribution μ and the reinsured risk distribution ν is. Allocating costs to this coupling corresponds to a transport problem between the marginals μ, ν. Optimal transport (OT) theory is an extremely active field of research, that has played a key role in many areas of applied mathematics ranging from PDEs, image processing, inverse problems, sampling, optimization, finance and economics to machine learning. We refer to the monographs by Villani [**112**], Santambrogio [**104**] and Ambrosio et al. [**11**] for an overview of the theory, as well as to [**98**] for the computational development, and to [**61, 71**] for applications in economics and finance. In our framework, when restricting our attention to deterministic reinsurance contracts, we are facing an OT problem in the so-called "Monge formulation" [**95**], where the joint distribution of μ, ν is supported on the graph of a function. This means that the final risk exposure of the insurer is simply a (deterministic) function of the initial exposure (a Monge map, in transport terms). On the other hand, allowing for randomized reinsurance treaties corresponds to the so-called "Kantorovich formulation" of the OT problem [**80**]. In that case we look for a coupling of μ and ν that minimizes the transportation cost without imposing it to be supported on the graph of a function, so that the original risk alone does not necessarily already determine the reinsured amount. It is therefore natural to try to connect optimal reinsurance to the optimal transport problem more closely, which is the goal of this paper.

The main characteristic of classical OT is that the cost functional is linear in the measure. When establishing a link to optimal reinsurance problems, a first obstacle is that even for simple risk measures, such as the variance, linearity does not hold. Consequently, in order to apply OT techniques, we need to linearize the reinsurance problem. The easiest way to that end is taking derivatives, but already at this point one runs into challenges since there are several notions of differentiability in infinite dimensions (e.g. the Fréchet or Gateaux type; and even for the latter one might need to impose linearity, continuity, etc.). We cannot ask for a very strong notion of differentiability since this implies continuity on opens sets (in the space of signed measures), which is hardly satisfied for situations of practical interest (e.g., even not for the expectation operator). For our purposes, it will turn out to be sufficient (and lead to interesting results already) that the directional derivatives exist and that they are convex linear (i.e., they satisfy the definition of a convex function with an equality instead of an inequality). If we suppose that these derivatives are integral operators, then, for simple constraints, it turns out that we can often characterize the support of optimal treaties up to a finite set of parameters; see Section 5.3. In the case of more general constraints, it will be possible to recast the problem as an iterated OT problem, which will allow some further insight and deductions; see Section 5.4 for details.

Organization of the paper. The rest of the paper is organized as follows. In Section 5.2 we define the general optimal reinsurance problem considered in this paper and introduce some notation. Section 5.3 then considers in more detail the situation with finitely many constraints. Here we show how this situation can be cast into a setup analogous to Lagrange optimization of a multivariate function under constraints. In Section 5.4 we cast the optimal reinsurance problem into the framework of an (iterated) optimal transport problem, which involves a local linearization of the optimization problem. The established link of the two fields then allows a characterization of reinsurance problems with purely deterministic optimal treaties. In Section 5.5 we then apply the results to a number of examples, some of which rederive classical optimality results by alternative means, while others lead to new results, and we compare these to existing literature. Some technical derivations in these examples are deferred to the appendix. Finally, Section 5.6 concludes.

5.2. Random Reinsurance Treaties

This section is devoted to introducing the setting and notations, and to defining the reinsurance problem studied in this paper.

5.2.1. Definitions and preliminaries. For a measurable set $A \subseteq \mathbb{R}^d$, we denote by $\mathscr{P}(A)$ the set of probability measures on the Borel sets of A. The push-forward measure of $\eta \in \mathscr{P}(\mathbb{R}^d)$ through a measurable map $f : \mathbb{R}^d \to \mathbb{R}^d$, denoted by $f_\#\eta$, is the probability measure such that $f_\#\eta(A) = \eta(f^{-1}(A))$, for any Borel set A. For $\mu \in \mathscr{P}(\mathbb{R}^n)$, we denote by F_μ its cummulative distribution function. If ν is another probability measure on $\in \mathscr{P}(\mathbb{R}^n)$, denoting by $\pi_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ the projection into the i-th coordinate, $i = 1, 2$, we can introduce the set of couplings (or transport plans) between two given marginal distributions μ and ν as

$$\Pi(\mu, \nu) = \{\eta \in \mathscr{P}(\mathbb{R}^n \times \mathbb{R}^n) \mid \pi_{1\#}\eta = \mu, \pi_{2\#}\eta = \nu\}. \tag{111}$$

We will also use the notation $\Pi(\mu, .) = \cup_{\nu \in \mathscr{P}(\mathbb{R}^n)}\Pi(\mu, \nu)$ for couplings where only the first marginal is fixed. A key role in defining our problem will be played by the subset

$$\Pi_{\leq}(\mu, .) = \{\eta \in \Pi(\mu, \cdot) \mid \eta(\mathcal{A}_R) = 1\},$$

where $\mathcal{A}_R = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid 0 \leq y_i \leq x_i, i = 1, \dots, n\}$. We denote by $\mathbb{R}^n_+$ the subset of $\mathbb{R}^n$ where all the coordinates are non-negative. Finally, for $\mu, \nu \in \mathscr{P}(\mathbb{R})$, we write $\nu \prec_1 \mu$ if $\nu([x, \infty)]) \leq \mu([x, \infty)])$ for every $x \in \mathbb{R}$.

5.2.2. Problem setting. From now on, we consider a fixed probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and a non-negative random vector $X = (X_1, \dots, X_n)$ defined on it, representing a portfolio of n risks from one or more insurers which are sought to be partially reinsured. We denote the distribution of each X_i by μ_i, the joint distribution of X by μ, and we assume that each X_i has a finite first moment, that is, $\mu_i \in \mathscr{P}_1(\mathbb{R}_+)$ and $\mu \in \mathscr{P}_1(\mathbb{R}^n_+)$.

In the classical formulation of optimal reinsurance problems, the random vector X is defined in some function space $\mathscr{L}$ (normally an L^p-space) and the objective is to find the best contract according to a pre-specified *risk measure*, i.e., a functional $\mathcal{P} : \mathscr{L} \to \overline{\mathbb{R}}$ representing the risk carried by the treaty. In this setting, a reinsurance contract is given by a sequence of functions $f_i : \mathbb{R}_+^n \to \mathbb{R}_+$ such that

$$0 \leq f_i(x_1, \ldots, x_n) \leq x_i \text{ for } (x_1, \ldots, x_n) \in \mathbb{R}_+^n \text{ and } i = 1, \ldots, n. \tag{112}$$

These functions help determine the split between the insurer and the reinsurer for the i-th contract in the portfolio in such a way that the reinsured amount is given by $f_i(X)$ and the retained amount is $X_i - f_i(X)$. Condition (112) ensures that these quantities are non-negative[1]. The optimal reinsurance problem is usually accompanied by a set of constraints (beliefs, conditions or requirements) that either of the involved parties might require. For example, if the reinsurance premiums are calculated according to an expectation principle, the first-line insurer might ask for $\mathbb{E}[f_i(X)] \leq c_i$ for some $c_i > 0$, ensuring that reinsurance is not too expensive. All such constraints can be represented by a set $\mathcal{S} \subset \mathscr{L}$, in the sense that a contract $(f_1, \ldots, f_n)$ is allowed if and only if $(f_1(X), \ldots, f_n(X)) \in \mathcal{S}$. In this example, $\mathcal{S}$ could then simply be

$$\mathcal{S} = \{(f_1(X), \ldots, f_n(X)) \in \mathscr{L} \mid \mathbb{E}[f_i(X)] \leq c_i, i = 1, \ldots, n\}.$$

The problem formulation would then be: find a sequence of functions $f_1^*, \ldots, f_n^*$ such that

$$\mathcal{P}(f_1^*(X), \ldots, f_n^*(X)) = \min_{\substack{0 \leq f_i(x) \leq x_i \\ (f_1(X), \ldots, f_n(X)) \in \mathcal{S}}} \mathcal{P}(f_1(X), \ldots, f_n(X)).$$

It is clear that with such generality, this problem is hard to solve and, moreover, the existence of solutions is not necessarily guaranteed. Instead of restricting ourselves to a specific kind of risk measures (convex risk measures are the usual choice in the literature), we reformulate the setting, echoing the passage from the Monge formulation of OT to the Kantorovich formulation: inspired by [**69**], we define a reinsurance treaty as the joint distribution of the initial risk(s) of the insurer(s) and the respective reinsured amounts.

Definition 5.2.1. *A (random) reinsurance treaty for the portfolio X is a probability measure $\eta \in \mathscr{P}(\mathbb{R}_+^n \times \mathbb{R}_+^n)$ such that:*

(i) the first marginal of η equals μ, i.e. $\pi_{1\#}\eta = \mu$;
(ii) $0 \leq y_i \leq x_i$ η-a.s. for all $i = 1, \ldots, n$, i.e. $\eta(\mathcal{A}_R) = 1$.

We denote the space of reinsurance treaties as $\mathscr{M}$ and endow it with the weak topology. In transport terms, $\mathscr{M} = \Pi_{\leq}(\mu, \cdot)$.

The weak topology appearing in the definition corresponds to the smallest topology making the functionals $\eta \mapsto \int f d\eta$ continuous, where f is any bounded and continuous function on $\mathbb{R}^n \times \mathbb{R}^n$. This definition can be seen as the multi-dimensional generalization of the one in [**69**], but different from the one found in [**70**]. Compared to the latter, we allow for dependencies between the reinsured amounts and all the risks in the portfolio, which makes $\mathscr{M}$ a compact space, instead of relatively compact. Also, we endow the space with a seemingly larger topology; however, these two topologies agree on $\mathscr{M}$.

Remark 5.2.2. Random treaties as in the definition above can be understood as reinsurance contracts in which the risk carried by the cedent and the reinsurer possibly has a degree of randomness external to the risks represented in X. Indeed, for any reinsurance treaty for the

[1]Note that in this general formulation, the reinsured amount in contract i is allowed to depend on the claim sizes of the other contracts as well, which in classical contracts is usually not the case, but we are interested here to explore whether that can lead to improved solutions.

portfolio $X \in \mathbb{R}^n_+$, there exists a random vector $R \in \mathbb{R}^n_+$ representing the part of the risks carried by the reinsurer, while $X - R \in \mathbb{R}^n_+$ is the retained amount (deductible) that stays with the first-line insurer(s). Then the η in Definition 5.2.1 is the distribution of the random vector (X, R). Observe that much as with the functions $f_1, \ldots, f_n$ from before, reinsurance treaties also specify how contracts are settled: given a realization of the claims $X = x$, one uses η to determine the conditional distribution of R given $X = x$. One then uses this conditional distribution to sample a value for R, say r, thus obtaining the retained amount r and the deductible $x - r$. The fact that η is supported in $\mathcal{A}_R$ thus guarantees that both r and $x - r$ are non-negative.

The risk measure from before is now a functional $\mathcal{P} : \mathscr{M} \to \overline{\mathbb{R}}$ and the constraints $\mathcal{S}$ are then a subset of $\mathscr{M}$. The optimal reinsurance problem is then to find $\eta^* \in \mathcal{S}$ such that

$$\mathcal{P}(\eta^*) = \min_{\eta \in \mathcal{S}} \mathcal{P}(\eta). \tag{113}$$

Any η^* satisfying (113) is called an optimal reinsurance contract.

We notice the following:

Proposition 5.2.3. *If $\mathcal{P}$ is lower semi-continuous and $\mathcal{S}$ is closed, then an optimal treaty η^* always exists.*

Proof. The proof carries over from the proof of from Proposition 1 in [**69**] showing that $\mathscr{M}$ is compact (observe that the proof of that proposition shows only that $\mathscr{M}$ is sequentially compact; however, since $\mathscr{M}$ is metrizable, the two notions agree). Since $\mathcal{S}$ is closed, it is compact as well, ensuring that $\mathcal{P}$ attains a minimum. $\square$

It follows that under rather mild assumptions on the functional $\mathcal{P}$ and the set of constraints $\mathcal{S}$, the existence of optimal treaties is guaranteed. However, in this broad framework, very little information can be obtained about the structure of these treaties. To address that, in this work we explore the idea of (locally) linearizing the problem, a technique often employed within the context of finite-dimensional optimization. In the subsequent sections, we will follow this idea and delve into a specific set of assumptions about $\mathcal{P}$ or $\mathcal{S}$ that allow us to reach more concrete conclusions about optimal contracts, while trying to preserve a level of generality suitable for a wide range of applications.

5.3. Optimal reinsurance with finite-dimensional constraints

As mentioned above, we want to use the idea of linearization to deduce further properties about optimal reinsurance treaties. In order to do this, it is natural to then impose smoothness properties on $\mathcal{P}$. In this section we study the effects of doing this when $\mathcal{S}$ has some particular form.

We first introduce some notation. For a function $f : \Omega_c \to V$ from any convex subset Ω_c of a vector space U into a normed space V, we let $df(u; h)$ denote the directional derivative of f at $u \in \Omega_c$ in the direction of $h \in U$. This derivative is given by the limit

$$df(u; h) = \lim_{t \to 0^+} \frac{f(u + th) - f(u)}{t},$$

whenever the terms on the right are defined and the limit exists. Notice that this derivative is defined as the right limit at zero, and by stating that $df(u; h)$ exists we make no assumption about the existence or value of the left limit (thus distinguishing df from the Gateaux derivative). If $V = \mathbb{R}^N$ and $v \in V$, we understand the inequalities $f(u) \leq v$, $f(u) < v$, etc. as holding component-wise.[2] In the following, whenever we apply these concepts to elements

[2]Notice that with this convention, $a \not\leq b$ does not imply $b < a$ unless $n = 1$.

in $\mathscr{M}$, we do so by regarding $\mathscr{M}$ as a subset of $\mathscr{S}$, the space of finite signed measures on $\mathbb{R}^n \times \mathbb{R}^n$ endowed with the total-variation norm. With this convention, we have implicitly defined two topologies on $\mathscr{M}$: the weak topology as defined in Section 5.2 and the topology inherited from $\mathscr{S}$. While we will always consider on $\mathscr{M}$ the weak topology, in what follows we will be mainly interested in the geometric properties of $\mathscr{M}$, in particular its convexity, and the only fact that we will require is that an optimal reinsurance contract exists. In this sense, we will assume from now on that $\mathcal{P}$ is lower semi-continuous and, throughout this section, that the set $\mathcal{S}$ is given as

$$\mathcal{S} = \{\eta \in \mathscr{M} \mid \mathcal{G}(\eta) \leq 0\} \tag{114}$$

for a lower semi-continuous function $\mathcal{G} = (g_1, \ldots, g_m) : \mathscr{M} \to \mathbb{R}^m$. That is, we allow for m constraints expressed as inequalities $g_i(\mu) \leq 0$, $i = 1, \ldots, m$.

Remark 5.3.1. Observe that when speaking of the directional derivatives of $\mathcal{P}$, it is meaningless to try to obtain expressions of the form $d\mathcal{P}(\eta; \vartheta)$ for $\eta, \vartheta \in \mathscr{M}$ since for any $t > 0$, $\eta + t\vartheta$ will not be a probability measure, so the expression $\mathcal{P}(\eta + t\vartheta)$ in the definition of $d\mathcal{P}(\eta; \vartheta)$ would not make sense. However, by exploiting the convexity of $\mathscr{M}$, we observe that for $0 < t < 1$, $\eta + t(\vartheta - \eta) = (1-t)\eta + t\vartheta \in \mathscr{M}$ and it is meaningful to inquire about the existence and properties of $d\mathcal{P}(\eta; \vartheta - \eta)$, which we will do in the sequel. These directional derivatives have the simple and natural interpretation of the instant change in $\mathcal{P}$ when, standing in η, we move in the direction of ϑ. In this context, the use of directional derivatives resembles the notion of Gateaux-differentiability employed by Deprez and Gerber in **[56]** for solving problems within optimal reinsurance and optimal cooperation. In their context, there is a risk functional $\widehat{\mathcal{P}}$ acting on random variables (instead of their distributions) and one obtains the derivatives by limits of the form $(\widehat{\mathcal{P}}(Y + t(Z - Y)) - \widehat{\mathcal{P}}(Y))/t$. This is, however, different from our approach, because when Y and Z are distributed according to η and ϑ respectively, $(1-t)Y + tZ$ will in general not be distributed according to $(1-t)\eta + t\vartheta$.

Before proceeding, we would like to emphasize that the following approach in particular applies to constraints involving the Value-at-Risk (which, compared to other common risk measures, exhibits poor mathematical properties), for which other approaches often fail.

Proposition 5.3.2. *Let $\mathcal{S}$ be given by $\mathcal{S} = \{\eta \in \mathscr{M} \mid \mathcal{G}(\eta) \leq 0\}$ for a function $\mathcal{G} : \mathscr{M} \to \mathbb{R}^m$, and η^* be an optimal reinsurance contract, i.e. satisfying* (113). *Let $\mathcal{D} \subset \mathscr{M}$ be given by*

$$\mathcal{D} = \{\eta \in \mathscr{M} \mid d\mathcal{P}(\eta^*; \eta - \eta^*) \text{ and } d\mathcal{G}(\eta^*; \eta - \eta^*) \text{ exist}\}.$$

Suppose there exists a subset $\mathcal{C}$ of $\mathcal{D}$ satisfying: $\eta^* \in \mathcal{C}$, $\mathcal{C}$ is convex, and if $\eta_1, \eta_2 \in \mathcal{C}$ then

$$d\mathcal{P}(\eta^*; (1-t)\eta_1 + t\eta_2 - \eta^*) = (1-t)d\mathcal{P}(\eta^*; \eta_1 - \eta^*) + t d\mathcal{P}(\eta^*; \eta_2 - \eta^*) \quad \text{for } 0 \leq t \leq 1, \tag{115}$$

and similarly for $d\mathcal{G}$. Then, there exist $r^* \in \mathbb{R}_+$ and $\lambda^* \in \mathbb{R}_+^m$ such that $\lambda^* \cdot \mathcal{G}(\eta^*) = 0$ and

$$r^* d\mathcal{P}(\eta^*; \eta - \eta^*) + \lambda^* \cdot d\mathcal{G}(\eta^*; \eta - \eta^*) \geq 0 \quad \text{for every } \eta \in \mathcal{C}. \tag{116}$$

If $\mathcal{G}$ is constant on $\mathcal{C}$ or there exists $\eta \in \mathcal{C}$ such that $\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) < 0$, then r^* in (116) *is positive.*

Remark 5.3.3. Before showing a proof for the proposition, we would like to provide an intuitive explanation of its meaning. Recall the standard optimization procedure for differentiable functions on $\mathbb{R}^N$ with smooth constraints, where one forms the Lagrangian and solves for its gradient while finding the multipliers that make the solutions satisfy the constraints. Proposition 5.3.2 generalizes this procedure and shows that, in the current scenario, one can still do something similar (observe that the left-hand side of Equation (116) would correspond to the derivative of the Lagrangian of the associated functionals). However, due to the looseness of our assumptions, the conclusion is much weaker. Indeed, (116) is only an inequality instead of an equality. This is due to the fact that $\mathcal{P}$ and $\mathcal{G}$ are, in principle, defined on $\mathscr{M}$ only rather than on the much larger space $\mathscr{S}$ of finite signed measures. Next, observe that we have to resort to the use of directional derivatives, as opposed to a stronger concept of differentiability. Using directional derivatives we cannot ensure linearity in the second argument of $d\mathcal{P}$ or $d\mathcal{G}$, so that we are then forced to operate on a smaller subset of $\mathscr{M}$. It is here that the set $\mathcal{C}$ plays a relevant role. One can think of this set as a "large enough set in which both $\mathcal{P}$ and $\mathcal{G}$ are smooth" (see also Remark 5.5.4 on how "large enough" could be understood). Since, for arbitrary functionals, $\mathcal{C}$ will be a strict subset of $\mathscr{M}$, the information provided by (116) will not hold for all possible reinsurance contracts. Finally, notice that as opposed to the usual Lagrangian in finite dimensions, there is a factor r^* multiplying $\mathcal{P}$. The appearance of this factor is related to the regularity constraint $\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \mu - \eta^*) < 0$, which essentially eliminates the possibility of a wedge at boundary points. In particular, this also prevents us from imposing *hard* equality constraints by redefining $\mathcal{G}$ to include negative among its components. Despite these shortcomings, we will see that Proposition 5.3.2 is still strong enough to derive important properties of η^*.

Proof. Define the sets

$$A = \{(r, \lambda) \in \mathbb{R} \times \mathbb{R}^m \mid r \geq d\mathcal{P}(\eta^*; \eta - \eta^*), \lambda \geq \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) \text{ for some } \eta \in \mathcal{C}\},$$
$$B = \{(r, \lambda) \in \mathbb{R} \times \mathbb{R}^m \mid r < 0, \lambda < 0\}.$$

Observe that A is non-empty, since $\mathcal{C}$ is non-empty, and convex thanks to (115). On the other hand, B is clearly convex, non-empty and open. We claim that A and B are disjoint. Arguing by contradiction, suppose there is some $r < 0$, $\lambda < 0$ and $\eta \in \mathcal{C}$ such that

$$d\mathcal{P}(\eta^*; \eta - \eta^*) \leq r \quad \text{and} \qquad \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) \leq \lambda.$$

Notice that the first inequality implies $\eta \neq \eta^*$. Since

$$\lim_{t \to 0^+} \mathcal{P}(\eta^* + t(\eta - \eta^*)) - \mathcal{P}(\eta^*) - td\mathcal{P}(\eta^*; \eta - \eta^*) = 0$$

and

$$\lim_{t \to 0^+} \mathcal{G}(\eta^* + t(\eta - \eta^*)) - \mathcal{G}(\eta^*) - td\mathcal{G}(\eta^*; \eta - \eta^*) = 0,$$

it follows that there exists $0 < s < 1$ such that

$$\mathcal{G}(\eta^* + s(\eta - \eta^*)) < 0 \text{ and } \mathcal{P}(\eta^* + s(\eta - \eta^*)) - \mathcal{P}(\eta^*) < 0,$$

so that $\eta^* + s(\eta - \eta^*) \in \mathcal{S}$ and $\mathcal{P}(\eta^* + s(\eta - \eta^*)) < \mathcal{P}(\eta^*)$, contradicting the optimality of η^*. Therefore, $A \cap B = \varnothing$. From the separation theorem for convex sets (see e.g. Theorem 3.4 in [**101**]), there exists a (non-zero) continuous linear functional Λ^* on $\mathbb{R} \times \mathbb{R}^m$ and $\gamma \in \mathbb{R}$ such that

$$\Lambda^*(r_1, \lambda_1) < \gamma \leq \Lambda^*(r_2, \lambda_2), \quad (r_1, \lambda_1) \in B, (r_2, \lambda_2) \in A$$

and $\Lambda^*(r, \lambda) = r^* r + \lambda^* \cdot \lambda$ for some $r^* \in \mathbb{R}$ and $\lambda^* \in \mathbb{R}^m$. Since $(0, 0) \in A \cap \overline{B}$, $\gamma = 0$. Thus $r^* r + \lambda^* \cdot \lambda < 0$ for every $r < 0$ and $\lambda < 0$, which can hold if and only if $r^* \geq 0$ and $\lambda^* \geq 0$.

Now, for every $\eta \in \mathcal{C}$, $(d\mathcal{P}(\eta^*; \eta - \eta^*), \mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*)) \in A$, and therefore

$$r^* d\mathcal{P}(\eta^*; \eta - \eta^*) + \lambda^* \cdot (\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*)) \geq 0, \quad \eta \in \mathcal{C}. \tag{117}$$

In particular, by choosing $\eta = \eta^*$, we obtain $\lambda^* \cdot \mathcal{G}(\eta^*) \geq 0$. However, the inequalities $\mathcal{G}(\eta^*) \leq 0$ and $\lambda^* \geq 0$ together imply $\lambda^* \cdot \mathcal{G}(\eta^*) \leq 0$, so that necessarily $\lambda^* \cdot \mathcal{G}(\eta^*) = 0$. Hence, (117) becomes equation (116). If there exists $\eta \in \mathcal{C}$ such that $\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) < 0$, we cannot have $r^* = 0$, for otherwise we would have $\lambda^* \neq 0$ and $\lambda^* \cdot d\mathcal{G}(\eta^*; \eta - \eta^*) < 0$, contradicting (116). In this situation we can therefore replace Λ^* by Λ^*/r^*, and proceed in the same manner. Finally, we observe that this argument does not hold if $\mathcal{G}$ is constant on $\mathcal{C}$. However, in this scenario, $d\mathcal{G}(\eta^*;\eta-\eta^*)=0$ for every $\eta\in\mathcal{C}$ and

$$d\mathcal{P}(\eta^*;\eta-\eta^*)=\lim_{t\to 0^+}\frac{\mathcal{P}((1-t)\eta^*+t\eta)-\mathcal{P}(\eta^*)}{t}\geq 0$$

for every $\eta\in\mathcal{C}$ by convexity of $\mathcal{C}$. Thus, (116) holds in any case for any $\lambda^*\geq 0$. □

In several situations, $\mathcal{P}$ and $\mathcal{G}$ are regular enough so that one can naturally extend them to a larger subspace of $\mathscr{S}$, making the results stronger. For example, the functionals of Examples 5.5.10 and 5.5.11 below can be naturally extended to the subspace $U\subset\mathscr{S}$ given by

$$U=\left\{\eta\in\mathscr{S}\mid\int_{\mathbb{R}^{2+}}x^2|\eta|(dx)<\infty\right\},$$

where $|\eta|$ denotes the total variation of η. In this set, the Gateaux derivatives of $\mathcal{P}$ and $\mathcal{G}$ are defined everywhere and are linear. Since Proposition 5.3.2 is formulated in terms of the directional derivatives, it is not necessary to extend $\mathcal{P}$ and $\mathcal{G}$ to a larger space, but we can focus instead on an extension of their derivatives. This is relevant when, e.g., $\mathcal{P}$ is related to the Value-at-Risk, since in this case we do not have to specify an extension of $\mathcal{P}$ to signed measures, but instead we can focus on extensions of $d\mathcal{P}$. This idea is used in the following result.

Proposition 5.3.4. *In the setup of Proposition 5.3.2, assume further that there exists a subspace $U\subset\mathscr{S}$ and linear mappings $\widehat{\mathcal{P}}:U\to\mathbb{R}$ and $\widehat{G}:U\to\mathbb{R}^m$ such that:*

(i) U contains the point masses of points in $\mathcal{A}_R$ and $\mathscr{M}\subset U$,

(ii) $d\mathcal{P}(\eta^*;\cdot)$ and $d\mathcal{G}(\eta^*;\cdot)$ are the restrictions to $\mathcal{C}-\eta^*$ of $\widehat{\mathcal{P}}$ and $\widehat{G}$, respectively, and

(iii) $\widehat{\mathcal{P}}$ and $\widehat{G}$ are integral operators with continuous kernels, i.e., there exist continuous functions $p_{\eta^*}:\mathbb{R}^n_+\times\mathbb{R}^n_+\to\mathbb{R}$ and $g_{\eta^*}:\mathbb{R}^n_+\times\mathbb{R}^n_+\to\mathbb{R}^m$ such that

$$\widehat{\mathcal{P}}(\eta)=\int_{\mathbb{R}^n_+\times\mathbb{R}^n_+}p_{\eta^*}(x,y)\eta(dx,dy)\quad\textit{and}\quad\widehat{G}(\eta)=\int_{\mathbb{R}^n_+\times\mathbb{R}^n_+}g_{\eta^*}(x,y)\eta(dx,dy).$$

Then, for every $(x,y)\in\operatorname{Supp}(\eta^*)$, there exists a closed set $I\subset[0,x]$ with $y\in I$ and

$$r^*p_{\eta^*}(x,y)+\lambda^*\cdot g_{\eta^*}(x,y)=\min_{t\in I}r^*p_{\eta^*}(x,t)+\lambda^*\cdot g_{\eta^*}(x,t).\tag{118}$$

Here, for $x,y\in\mathbb{R}^n$, $[x,y]$ represents the closed box $[x_1,y_1]\times\cdots\times[x_n,y_n]$. Similarly, we denote by $]x,y[$ the product of the open intervals $]x_1,y_1[\times\cdots\times]x_n,y_n[$. Observe that the above assumptions implicitly require that the integrals are finite for all $\eta\in U$, in particular for all $\eta\in\mathscr{M}$.

Proposition 5.3.4 converts the information about *measures* conveyed by (116) to information about *points* in $\mathbb{R}^n\times\mathbb{R}^n$, by describing the support of optimal contracts in terms of the functions p_{η^*} and g_{η^*}. Essentially, one would like to use (116) evaluated at point masses $1_{(x,y)}$ to accomplish this. However, since for any η, $\mathscr{M}$ will not contain all the point masses, one needs to be more creative.

Proof. Let (x,y) be an arbitrary point in $\mathcal{A}_R$. For $t\in\mathbb{R}^n$, define the measures $\eta_{x,y,t,\varepsilon}$ by

$$\eta_{x,y,t,\varepsilon}(A)=\eta^*(A)-\eta^*(A\cap B_\varepsilon(x,y))+\eta^*((A-(0,t))\cap B_\varepsilon(x,y)),\quad\text{for }A\in\mathbb{R}^n_+\times\mathbb{R}^n_+,$$

where $A-(0,t)$ denotes the translation of A by $-(0,t)\in\mathbb{R}^n_+\times\mathbb{R}^n_+$, and $B_\varepsilon(x,y)$ is the open ball around (x,y) of radius $\varepsilon>0$. The proof will be divided into four steps:

Step 1. Let $(x,y)\in\mathcal{A}_R$ be such that $0<y<x$. Define δ by

$$\delta=\min\left\{d((x,y),\partial\mathcal{A}_R),d((x,y),\partial\mathcal{A}_R-(0,t))\right\}.$$

Then $\delta > 0$ *and* $\eta_{x,y,t,\varepsilon} \in \mathscr{M}$ *for every* $-y < t < x - y$ *and* $0 \leq \varepsilon < \delta$.

Here, $d((x,y),B)$ denotes the distance of (x,y) from the set B. The idea of this choice is that δ is the largest number for which the last two terms in the definition of $\eta_{x,y,t,\varepsilon}(\mathcal{A}_R)$ cancel out, thus producing an element of $\mathscr{M}$. Now, strict positivity of δ follows from the choice of y and t: since $0 < y < x$, $(x,y) \notin \partial\mathcal{A}_R$, and since $(x,y) \in \mathcal{A}_R - (0,t)$ if and only if $(x, y+t) \in \mathcal{A}_R$, it follows that, for $-y < t < x - y$, $(x, y+t) \notin \partial\mathcal{A}_R$, so $(x,y) \notin \partial\mathcal{A}_R - (0,t)$. As both $\partial\mathcal{A}_R$ and $\partial\mathcal{A}_R - (0,t)$ are closed, we obtain $\delta > 0$. Next, we need to check that $\eta_{x,y,t,\varepsilon} \in \mathscr{M}$. Since $\eta_{x,y,t,\varepsilon}$ is a (positive) measure in $\mathbb{R}^n_+ \times \mathbb{R}^n_+$, we just need to show that it is a probability measure giving full measure to $\mathcal{A}_R$ with $\pi_{1\#}\eta_{x,y,t,\varepsilon} = \mu$. We have:

- $\eta_{x,y,t,\varepsilon}(\mathcal{A}_R) = 1$ and $\eta_{x,y,t,\varepsilon}(\mathcal{A}_R^c) = 0$, since $B_\varepsilon(x,y) \subset \mathcal{A}_R \cap (\mathcal{A}_R - (0,t))$ for $0 \leq \varepsilon < \delta$;
- For measurable $A \subset \mathbb{R}^n_+$, $A \times \mathbb{R}^n - (0,t) = A \times \mathbb{R}^n$, so

$$\eta_{x,y,t,\varepsilon}(A \times \mathbb{R}^n) = \eta^*(A \times \mathbb{R}^n) = \mu(A).$$

We conclude that $\eta_{x,y,t,\varepsilon} \in \mathscr{M}$.

Step 2. If $(x,y) \in \operatorname{Supp}(\eta^*)$ *is such that* $0 < y < x$, *then, for every* $-y < t < x - y$ *and continuous function* $f : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}^n$, *we have*

$$\lim_{\varepsilon \to 0} \int f d\eta_\varepsilon = f(x, y+t) - f(x,y), \tag{119}$$

where

$$\eta_\varepsilon = \frac{\eta_{x,y,t,\varepsilon} - \eta^*}{\eta^*(B_\varepsilon(x,y))}.$$

Let $\vartheta_\varepsilon, \pi_\varepsilon \in \mathscr{P}(\mathbb{R}^n_+ \times \mathbb{R}^n_+)$ be given by

$$\vartheta_\varepsilon(A) = \frac{\eta^*(A \cap B_\varepsilon(x,y))}{\eta^*(B_\varepsilon(x,y))}$$

and

$$\pi_\varepsilon(A) = \frac{\eta^*((A - (0,t)) \cap B_\varepsilon(x,y))}{\eta^*(B_\varepsilon(x,y))},$$

so that $\eta_\varepsilon = \pi_\varepsilon - \vartheta_\varepsilon$. To prove (119) it is then enough to show that $\int f d\vartheta_\varepsilon \to f(x,y)$ and $\int f d\pi_\varepsilon \to f(x, y+t)$. For this, note that

$$\left| \int f d\vartheta_\varepsilon - f(x,y) \right| = \left| \frac{1}{\eta^*(B_\varepsilon(x,y))} \int_{B_\varepsilon(x,y)} f d\eta^* - f(x_0, y_0) \right| \to 0 \qquad \text{as } \varepsilon \to 0$$

by the Lebesgue differentiation theorem. The other convergence is proved similarly.

Step 3. For $(x,y) \in \operatorname{Supp}(\eta^*)$ *with* $0 < y < x$, *the statement of the proposition holds by taking*

$$I = \overline{\{t \in]0, x[\mid 0 \textit{ is a limit point of } \{\varepsilon \in [0,\delta) \mid \eta_{x,y,t,\varepsilon} \in \mathcal{C}\}\}}.$$

Define

$$E(t) = \{\varepsilon \in [0,\delta) \mid \eta_{x,y,t,\varepsilon} \in \mathcal{C}\}$$

and

$$J = \{t \in]0, x[\mid 0 \text{ is a limit point of } E(t - y)\}.$$

By assumption, we know that $\eta^* \in \mathcal{C}$, so $E(0) = [0, \delta)$, implying $y \in J$. From (116) and linearity, we obtain

$$r^* \widehat{\mathcal{P}} \left(\frac{\eta_{x,y,t-y,\varepsilon} - \eta^*}{\eta^*(B_\varepsilon(x,y))} \right) + \lambda^* \cdot \widehat{G} \left(\frac{\eta_{x,y,t-y,\varepsilon} - \eta^*}{\eta^*(B_\varepsilon(x,y))} \right) \geq 0$$

for $t \in J$ and $\varepsilon \in E(t-y)$. Letting $\varepsilon \to 0$ in the previous inequality, by Step 2 we obtain

$$r^*(p_{\eta^*}(x,t) - p_{\eta^*}(x,y)) + \lambda^* \cdot (g_{\eta^*}(x,t) - g_{\eta^*}(x,y)) \geq 0, \quad t \in J(x,y).$$

Using continuity once again, we see that this inequality is valid for every t in the closure of J, so we can take $I = \overline{J}$, obtaining (118).

Step 4. The statement of the proposition holds for arbitrary $(x,y) \in \operatorname{Supp}(\eta^*)$.

Observe that it only remains to show that the statement is true for $(x,y) \in \partial\mathcal{A}_R$. Step 1 still holds by setting $\delta = d((x_0,y_0), \partial\mathcal{A}_R - (0,t))$ for $-y < t < x - y$. Notice that now we only have $B_\varepsilon(x,y) \cap \subset \mathcal{A}_R \cap (\mathcal{A}_R - (0,t))$, however, this is enough to show that $\eta_{x,y,t,\varepsilon}(\mathcal{A}_R) = 1$ and $\eta_{x,y,t,\varepsilon}(\mathcal{A}_R^c) = 0$ by noticing that

$$\eta^*(A) = \eta^*(A \cap \mathcal{A}_R)$$

for every measurable $A \subset \mathbb{R}_+^n$, so that we anyway have $\eta_{x,y,t,\varepsilon} \in \mathscr{M}$. Steps 2 and 3 carry over verbatim to this case. □

Remark 5.3.5. The measures $\eta_{x,y,t,\varepsilon}$ in the previous proof were already used in [**69**] and again in [**70**] with the same objective of identifying the support of optimal reinsurance contracts. However, since the functionals in these papers were more specific, one was able to directly obtain a description of the sets I there in a more direct manner.

Remark 5.3.6. Note that the conclusion of Proposition 5.3.4 is somewhat redundant for our purposes: Equation (118) assumes knowledge of points in the support of η^*, which is the contract that we wish to determine. Nonetheless, the procedure can be reversed: if by some mechanism we can identify the sets I, then we will know that points in the support of η^* will be the minima of $r^* p_{\eta^*}(x,\cdot) + \lambda^* \cdot g_{\eta^*}(x,\cdot)$. In most situations, p_{η^*} and g_{η^*} depend on η^* only through a finite set of parameters, so we can compute these minima without full specification of η^*, and in common applications, there will be only finitely many minima. Hence, the task of finding optimal contracts is then reduced to identifying the minima that *do* belong to the support of η^* together with the correct parameters specifying η^*, a task that is in general easier than computing a full measure[3]. Examples 5.5.1 and 5.5.3 below illustrate the procedure of identifying the minima, reducing the problem to a two-dimensional and one-dimensional optimization task, respectively.

Finally, when the extension can be done across the entire set $\mathscr{M}$, more can be said.

Proposition 5.3.7. *In the setting of Proposition 5.3.4, assume we can take* $\mathcal{C} = \mathcal{D} = \mathscr{M}$*. Moreover, assume that the partial minimization function*

$$m(x) = \inf_{y \in [0,x]} r^* p_{\eta^*}(x,y) + \lambda^* \cdot g_{\eta^*}(x,y)$$

is measurable. Then

$$\eta^* \left(\{(x,y) \in \mathcal{A}_R \mid y \in \operatorname{argmin}_{t \in [0,x]} r^* p_{\eta^*}(x,t) + \lambda^* \cdot g_{\eta^*}(x,t)\} \right) = 1. \tag{120}$$

[3]This can again be compared with the common optimization technique in $\mathbb{R}^N$: when looking for extremes of a smooth function, one first finds the zeros of the gradient and then chooses the zeros that produce the desired extrema.

Proof. Let $h : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$ be given by $h(x,y) = r^* p_{\eta^*}(x,y) + \lambda^* \cdot g_{\eta^*}(x,y)$ and let M be the set appearing in (120). Observe first that

$$M = \{(x,y) \in \mathcal{A}_R \mid h(x,y) - m(x) = 0\},$$

so measurability of m implies that M is measurable, and (120) makes sense. The statement then follows once we show the inequality

$$\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x,y)\ \eta^*(dx,dy) \le \int_{\mathbb{R}^n_+} m(x)\ \mu(dx). \tag{121}$$

Indeed, since $\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} m(x)\ \eta^*(dx,dy) = \int_{\mathbb{R}^n_+} m(x)\ \mu(dx)$ and $h(x,y) \ge m(x)$ for every $(x,y) \in \mathcal{A}_R$, from the inequality in (121) it follows that even equality holds, thus $h(x,y) = m(x)$ η^*-a.e., which is equivalent to (120).

To prove (121), we make use of Theorem 5.5.3 in [**108**], which states that every analytic subset A of the product of two Polish spaces admits a section s that is universally measurable, i.e., s is measurable with respect to the completion of the Borel σ-algebra w.r.t. any probability measure. With $A = M$, it follows that there exists a function $s : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ such that $s(x) \in M$ and $\pi_1 \circ s(x) = x$ for every $x \in \mathbb{R}^n$ (a *section* of M), and such that s is universally measurable. In particular, s is measurable with respect to the μ-completion of $\mathbb{R}^n$. Let η be the probability measure on $\mathbb{R}^n \times \mathbb{R}^n$ given by $\eta = s_{\#}\mu$. Notice that by the μ-measurability of s, η is well defined and, moreover, it is a reinsurance contract. Hence, (116) implies

$$\begin{aligned} r^* d\mathcal{P}(\eta^*; \eta - \eta^*) + \lambda^* \cdot d\mathcal{G}(\eta^*; \eta - \eta^*) &= r^* \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} p_{\eta^*}(x,y)(\eta - \eta^*)(dx,dy) \\ &\quad + \lambda^* \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} g_{\eta^*}(x,y)(\eta - \eta^*)(dx,dy) \\ &= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x,y)(\eta - \eta^*)(dx,dy) \ge 0, \end{aligned}$$

so

$$\begin{aligned} \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x,y)\ \eta^*(dx,dy) &\le \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x,y)\ \eta(dx,dy) \\ &= \int_{\mathbb{R}^n_+} h \circ s(x)\ \mu(dx) = \int_{\mathbb{R}^n_+} m(x)\ \mu(dx), \end{aligned}$$

as desired. □

Corollary 5.3.8. *In the setting of Proposition 5.3.7, let $\mathcal{P}$ be given by*

$$\mathcal{P}(\eta) = \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} p(x,y)\eta(dx,dy),$$

for a continuous function $p : \mathbb{R}^n_+ \times \mathbb{R}^n_+ \to \mathbb{R}$, and define the set M by

$$M = \{(x,y) \in \mathcal{A}_R \mid y \in \operatorname{argmin}_{t \in [0,x]} p(x,t)\}.$$

Then, $\eta \in \mathscr{M}$ is an optimal reinsurance contract if and only if $\eta(M) = 1$.

Proof. Let $\eta \in \mathscr{M}$. If η is optimal, Proposition 5.3.7 implies $\eta(M) = 1$, since $h = p_\eta = p$ and $g_\eta = 0$ (i.e., the function p_η is the same for all optimal contracts), so we only need to prove the reverse implication. However, this is immediate from (the proof of) Proposition 5.3.7, since if $\eta(M) = 1$, then $p(x, y) = m(x)$ η-a.e., where m is the partial minimization function, so

$$\begin{aligned}\mathcal{P}(\eta) &= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} p(x, y)\ \eta(dx, dy) = \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} m(x)\ \mu(dx) \\ &= \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} p(x, y)\ \eta^*(dx, dy) = \mathcal{P}(\eta^*),\end{aligned}$$

where η^* is any optimal reinsurance contract. Hence η is optimal. $\square$

Remark 5.3.9. In Proposition 5.3.7, it is necessary to require the partial minimization function to be measurable. While this might seem like an assumption that should usually be satisfied, in general the partial minimization operation yields only a lower semi-analytic function and one cannot ensure the set M to be analytic (see Proposition 7.47 of [**29**] for further details). This observation also explains why tools from descriptive set theory need to be used in a place which is seemingly unrelated. Note, however, that these tools are mostly required to ensure the existence of the section s. In some applications, the functions p and g are smooth enough so that one can show the existence of s by more conventional methods (e.g., the implicit function theorem).

Remark 5.3.10. If for every $x \in \mathbb{R}^n$ the function $r^* p_{\eta^*}(x, \cdot) + \lambda^* \cdot g_{\eta^*}(x, \cdot)$ has exactly one minimizer, the section s from the proof of Proposition 5.3.7 is unique μ-a.e and $\eta^* = (\mathrm{Id}, \tau)_{\#}(\mu)$ for some function τ, where Id is the identity function. However, it does not necessarily follow that a deterministic reinsurance contract is optimal, since τ might fail to be Borel-measurable. This might happen when, for instance, τ does not "move measurably" from one x to the other, describing a measurable set of measure zero in $\mathbb{R}^n \times \mathbb{R}^n$ without a measurable projection into $\mathbb{R}^n$. For example, for $n = 1$, let E be a non-Lebesgue-measurable subset of $]0, \infty[$ and define

$$E' = \{(x, y) \in \mathbb{R}^2_+ \mid y = x \text{ if } x \in E \text{ and } y = 0 \text{ otherwise}\} \subset D \cup (\mathbb{R}_+ \times \{0\}),$$

where $D = \{(x, x) \in \mathbb{R}^2_+ \mid x \geq 0\}$ is the diagonal of $\mathbb{R}^2_+$. As $D \cup (\mathbb{R}_+ \times \{0\})$ has Lebesgue-measure zero, E' is Lebesgue measurable. Letting $p : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ be the distance to E', $\mathcal{P}(\eta) = \int p\ d\eta$ and $G = 0$, we see that p is continuous and the assumptions from Proposition 5.3.7 are satisfied with partial minimization function identically zero. Moreover, $r^* p_{\eta^*}(x, \cdot) + \lambda^* \cdot g_{\eta^*}(x, \cdot) = p(x, \cdot)$ has exactly one minimum, namely, at 0 or x, so that $\tau(x) = x 1_E(x)$, where 1_E is the indicator function of E. Hence, τ is not measurable, so by uniqueness, no deterministic reinsurance contract exists. Observe, however, that any contract supported in $\overline{E'}$ will be optimal, so one can easily identify optimal contracts. It is clear that this kind of pathology is not likely to appear in examples commonly happening in practice, however the fact that this kind of risk measure is considered within our assumptions points towards the generality of our setting.

Remark 5.3.11. Corollary 5.3.8 is, in a way, the best we can do in terms of fully identifying optimal reinsurance contracts. Coming back to the hypotheses and notation of Proposition 5.3.7, the same argument given in the proof of Corollary 5.3.8 shows that

$$\int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x, y)\ \eta^*(dx, dy) = \int_{\mathbb{R}^n_+ \times \mathbb{R}^n_+} h(x, y)\ \hat{\eta}(dx, dy)$$

for any $\hat{\eta} \in \mathscr{M}$ such that $\hat{\eta}(M) = 1$. This implies that an equation analogous to (116) is valid when we replace η^* by $\hat{\eta}$ on the second argument of $d\mathcal{P}$ and $d\mathcal{G}$, i.e.,

$$r^* d\mathcal{P}(\eta^*; \eta - \hat{\eta}) + \lambda^* \cdot d\mathcal{G}(\eta^*; \eta - \hat{\eta}) \geq 0, \quad \eta \in \mathscr{M}.$$

However, this condition is not sufficient for optimality when $\mathcal{G}$ is not constant.

5.4. Optimal transport

The optimal transport theory is concerned with the problem of moving mass from a defined distribution to another one while minimizing the cost of transportation. Specifically, given two Polish spaces $\mathcal{X}$ and $\mathcal{Y}$, two distributions μ and ν on them, respectively, and a cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$, the OT problem can be formulated as

$$\inf_{\eta \in \Pi(\mu,\nu)} \int c(x,y)\eta(dx,dy), \tag{122}$$

where $\Pi(\mu,\nu)$ is the set of probability measures on $\mathcal{X} \times \mathcal{Y}$ with marginals μ, ν; see [**112, 11**]. An optimizer η^* of (122) is called optimal coupling. If it is concentrated on the support of a map, that is $\eta^* = (\mathrm{Id}, \tau)_{\#}\mu$, then it is called an optimal Monge coupling, and τ is an optimal Monge map.

In this section we will reformulate the optimization problem (113) as a transport problem. We consider $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$, so that $\Pi(\mu,\nu)$ is the set of couplings defined in (111). Moreover, the first marginal μ is fixed to be the distribution of the portfolio X to be reinsured, while ν is the distribution of the risk carried by the reinsurer, to be determined in an optimal way. This will correspond to the second marginal of the optimal treaty η^* in (113). This analysis will require: (i) defining the set of feasible distributions for the risk carried by the reinsurer, i.e. the possible second marginals ν to consider in (122); (ii) identifying the appropriate cost function in (122); (iii) ensuring the joint distribution η of μ and ν to satisfy the requirements in Definition 5.2.1, so that $\eta \in \mathscr{M} = \Pi_{\leq}(\mu, \cdot)$.

In the previous section it was necessary that the set $\mathcal{S}$ was described by a finite set of inequalities (finitely many constraints). Throughout this section we would like to drop this assumption and investigate what conclusions can be drawn when the constraints are more general (that is, involving equalities or infinitely many constraints).[4] As indicated before, this is a rather hard task if we allow $\mathcal{P}$ to be an arbitrary (lower semi-continuous) functional. We make then some concessions and, still inspired by the idea of local linearization from the previous section, we assume the following:

(A1) If $\eta^* \in \mathcal{S}$ is an optimal reinsurance contract, then for every $\eta \in \mathcal{S}$ and $0 \leq t \leq 1$, we have
$$\mathcal{P}(\eta^*) \leq \mathcal{P}((1-t)\eta^* + t\eta);$$

(A2) For every $\eta \in \mathcal{S}$, $d\mathcal{P}(\eta; \cdot)$ exists for every direction in $\mathcal{S} - \eta$ and is given as an integral operator, i.e., there exists a measurable function $p_\eta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that, for every $\vartheta \in \mathcal{S}$,
$$d\mathcal{P}(\eta; \vartheta - \eta) = \int p_\eta(x,y)(\vartheta - \eta)(dx,dy).$$
Observe that this implicitly yields that the integral is finite.

Given the considerations made in the previous sections, assumption (A2) seems natural in order to linearize the problem. In contrast, assumption (A1) might seem odd, as it may appear more natural to require $\mathcal{S}$ to be convex. In our examples, this will most often be the case, but we prefer to phrase it this way to cover a larger amount of scenarios (for example, when $\mathcal{S}$ is arbitrary and $\mathcal{P}$ is concave, as is the case for Value-at-Risk in Example 5.5.11).

Assumptions (A1) and (A2) jointly imply that

$$\int p_{\eta^*}(x,y)\, \eta^*(dx,dy) = \min_{\eta \in \mathcal{S}} \int p_{\eta^*}(x,y)\, \eta(dx,dy). \tag{123}$$

[4]Recall from Remark 5.3.3 that introducing equalities through $g \leq 0$ and $-g \leq 0$ was not necessarily possible for positive r^*.

Letting q_{η^*} denote the function on $\mathbb{R}^n \times \mathbb{R}^n$ such that $q_{\eta^*}(x,y) = p_{\eta^*}(x,y)$ on $\mathcal{A}_R$ and otherwise being equal to $+\infty$, (123) can be stated as

$$\int q_{\eta^*}(x,y)\,\eta^*(dx,dy) = \min_{\nu\in\pi_2(\mathcal{S})} \mathcal{C}(\mu,\nu), \tag{124}$$

where $\pi_2(\mathcal{S}) = \{\pi_{2\#}\eta : \eta \in \mathcal{S}\}$ and

$$\mathcal{C}(\mu,\nu) = \min_{\eta\in\Pi(\mu,\nu)\cap\mathcal{S}} \int q_{\eta^*}(x,y)\,\eta(dx,dy). \tag{125}$$

Equations (124)-(125) mean that the optimal contract satisfies a double minimization property, where the inner minimum (125) is a constrained optimal transport problem (the couplings need to satisfy the constraint of belonging to $\mathcal{S}$). Note that we still face the issue from the previous section that the function p_{η^*} depends on η^* and so does the cost function in (125). This means that we are facing transport problems depending on the optimal treaty that we are looking for. The idea behind (125) is, however, similar to the one developed before in the sense that for a large set of functionals, the function q_{η^*} will depend on η^* solely through a finite set of parameters which can be thought of as fixed at the beginning. The hope is that, by means of optimal transport techniques, one can provide information about the general structure of optimal couplings η^*, for example about the geometric characterizations of their supports, and by leveraging this one can find the parameters which achieve the minimum in (124).[5] For example, if for every $\nu \in \pi_2(\mathcal{S})$, there exists an optimal Monge coupling for the OT problem (125), then η^* is also given by a deterministic reinsurance contract. This is evident by observing that for an optimal reinsurance treaty η^*, we have

$$\int q_{\eta^*}(x,y)\,\eta^*(dx,dy) = \mathcal{C}(\mu,\nu^*),$$

where $\nu^* = \pi_{2\#}\eta^*$. Now, while existence of an optimal Monge coupling is a rather scarce property in OT problems, this observation is relevant enough to cover some interesting cases in the context of optimal reinsurance. This is illustrated in the following proposition, and then applied in Examples 5.5.10 and 5.5.11 below.

Proposition 5.4.1. *Assume $\mathcal{P}$ is given by*

$$\mathcal{P}(\eta) = \mathcal{P}_1(T_{\#}\eta) + \mathcal{P}_2(\pi_{2\#}\eta), \tag{126}$$

where T is the linear operator $T : \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}^n$ given by $T(x,y) = x-y$, and the functionals $\eta \mapsto \mathcal{P}_1(T_{\#}\eta)$, $\eta \mapsto \mathcal{P}_2(\pi_{2\#}\eta)$ from $\mathscr{M}$ to $\overline{\mathbb{R}}$ satisfy condition (A2) above with functions $h_\eta, k_\eta : \mathbb{R}^n \to \mathbb{R}$, respectively. Then, if $\mathcal{P}$ satisfies (A1), it also satisfies (A2) with p_η given by

$$p_\eta(x,y) = h_\eta(x-y) + k_\eta(y). \tag{127}$$

Moreover, if $n = 1$ and

(i) *for every $\nu \in \pi_2(\mathcal{S})$, we have $\Pi(\mu,\nu)\cap\mathscr{M} \subset \mathcal{S}$,*
(ii) *h_{η^*} is strictly convex for every optimal treaty η^*, and*
(iii) *the distribution of X is continuous,*

then an optimal Monge map exists for (125) *for every $\nu \in \pi_2(\mathcal{S})$.*

Observe that, under the assumption $\Pi(\mu,\nu)\cap\mathscr{M} \subset \mathcal{S}$ for every $\nu \in \pi_2(\mathcal{S})$, (125) turns into

$$\mathcal{C}(\mu,\nu) = \min_{\eta\in\Pi(\mu,\nu)} \int q_{\eta^*}(x,y)\,\eta(dx,dy).$$

[5]Note, however, that this is not a requirement for optimality but rather a consequence. After having used the information to characterize optimal treaties, one still needs to find the one that minimizes $\mathcal{P}$.

If we dropped the requirement of the reinsurance contracts being supported in $\mathcal{A}_R$, this problem would become

$$\widehat{\mathcal{C}}(\mu,\nu) = \min_{\eta\in\Pi(\mu,\nu)} \int p_{\eta^*}(x,y)\,\eta(dx,dy), \tag{128}$$

with p_{η^*} given as in (127). With hypotheses (ii) and (iii) as in Proposition 5.4.1, we would be in the Gangbo-McCann setting (cf. [**62**]) and the minimum in (128) would be achieved by an unique optimal coupling $\pi^* \in \Pi(\mu,\nu)$ of Monge type of the form

$$\pi^* = (\mathrm{Id}, \tau^*)_{\#}\mu = (\mathrm{Id}, \mathrm{Id} - (\nabla p_{\eta^*})^{-1} \circ \nabla\varphi)_{\#}\mu, \tag{129}$$

for some p_{η^*}-concave function φ, irrespective of n. Furthermore, any map τ^* of this form would be optimal between μ and $\tau^*_{\#}\mu$. Since this would be valid for any $\nu \in \pi_2(\mathcal{S})$, the optimal contract would be deterministic. In the reinsurance setting, we could use this observation to then optimize over contracts of the form (129) such that

$$0 \le (\nabla p_{\eta^*})^{-1} \circ \nabla\varphi \le \mathrm{Id},$$

thus ensuring $\pi^* \in \mathscr{M}$. However, we do not know a priori whether any such function φ exists or whether all $\nu \in \pi_2(\mathcal{S})$ can be coupled with μ with these functions. The conclusion of Proposition 5.4.1 is that we can still guarantee this at least for the case $n = 1$.

Proof. It is easy to see that $\mathcal{P}$ satisfies (A2) and p_η is given as in (127), so we only need to show the second part of the statement. For this, fix $\nu \in \pi_2(\mathcal{S})$ and observe that by (A2), $\mathcal{C}(\mu,\nu)$ is finite. The first condition means that the constraints depend only on the second marginal of any reinsurance treaty, so that for any optimal contract η^*,

$$\begin{aligned} \min_{\eta\in\Pi(\mu,\nu)\cap\mathcal{S}} \int p_{\eta^*}(x,y)\,\eta(dx,dy) &= \min_{\eta\in\Pi(\mu,\nu)} \int p_{\eta^*}(x,y)\,\eta(dx,dy) \\ &= c(\nu) + \min_{\eta\in\Pi(\mu,\nu)} \int h_{\eta^*}(x-y)\,\eta(dx,dy), \end{aligned} \tag{130}$$

where $c(\nu) = \int_0^\infty k_{\eta^*}(y)\,\nu(dy)$. The second and third condition from the statement of the proposition are technical conditions that ensure the existence of a unique Monge optimizer for the last minimum at the end of (130), given by

$$R = F_\mu \circ F_\nu^{-1}.$$

Since $\nu \prec_1 \mu$, we have $R(x) \le x$ for every $x \ge 0$, so R is also an optimizer for $\mathcal{C}(\mu,\nu)$. Since ν was arbitrary, the proposition follows. $\square$

In what follows we use the notation $\pi_{i,j}$ for the composition $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \to \mathbb{R}$ of the projection onto the i^{th}-coordinate, $i = 1, 2$, and then onto the j-th coordinate, $j = 1, \ldots, n$.

Proposition 5.4.2. *Assume that $\mathcal{P}$ can be written as a composition $\mathcal{P} = \mathcal{F} \circ \mathcal{R}$, where $\mathcal{R} : \mathscr{M} \to \mathscr{P}(\mathbb{R})^n$ is given by*

$$\mathcal{R}(\eta) = (\pi_{2,1\,\#}\eta, \ldots, \pi_{2,n\,\#}\eta)$$

and $\mathcal{F} : \mathscr{P}(\mathbb{R})^n \to \overline{\mathbb{R}}$ is a lower semi-continuous functional. Assume moreover that there exists a topological space O such that $\mathcal{S}$ is given as $\mathcal{S} = \mathcal{U}^{-1}(E)$ for some closed $E \subset O$, where $\mathcal{U} = \mathcal{V} \circ \mathcal{R}$ and $\mathcal{V} : \mathscr{P}(\mathbb{R})^n \to O$ is continuous on the image of $\mathcal{R}$. Let C be a copula for the distribution μ of X. Then, if η^* is an optimal reinsurance treaty and $\nu_i^* = \pi_{2,i\,\#}\eta^*$, $i = 1, \ldots, n$, the treaty π whose distribution function is given by

$$F_\pi(x_1, \ldots, x_n, y_1, \ldots, y_n) = C(\min(F_{\mu_1}(x_1), F_{\nu_1^*}(y_1)), \ldots, \min(F_{\mu_n}(x_n), F_{\nu_1^*}(y_n)))$$

is also optimal. In particular, if each F_{μ_i} is continuous, an optimal deterministic reinsurance contract exists, where the components are given by the functions

$$R_i(x) = F_{\mu_i}(F^{-1}_{\nu_i^*}(x)), \quad i = 1, \dots, n.$$

Proof. This is straightforward after noticing that $\mathcal{R}(\eta^*) = \mathcal{R}(\pi)$. Observe that the distributions $\min(F_{\mu_i}, F_{\nu_i^*})$ represent the c.d.f's of the monotonic rearrangements of the individual claims with the ν_i^*'s, so by the argument given at the end of the proof of Proposition 5.4.1, π is concentrated on $\mathcal{A}_R$. □

Proposition 5.4.2 simply states that if the risk measure and the constraints depend solely on the reinsured distribution, there is *no need for randomization.* Observe moreover that the statement does not make use of the particular form of $\mathcal{P}$ other than its factorization, so the statement is valid if we change the point of view and assume that $\pi_{2\#}\eta$ represents the distribution of the deductible instead of the reinsured amount. Proposition 5.4.1 differs from Proposition 5.4.2, in the sense that Proposition 5.4.1 allows for a "mixing" between the deductible and the reinsured amount, while Proposition 5.4.2 assumes $\mathcal{P}$ is only determined by the reinsured amounts.

5.5. Examples

We explore the techniques developed in the previous sections to illustrate some examples in optimal reinsurance that can be approached with this framework; see [**1**, Ch. VIII] for a systematic survey on optimal reinsurance problems. As in Proposition 5.4.1, in the sequel the linear operator $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ represents the vector of retained risks $T(x, y) = x - y$. For any $\nu \in \mathscr{P}(\mathbb{R})$, we denote by $\overline{\nu}$ the mean of ν, that is, $\overline{\nu} = \int y\nu(dy)$.

The following three examples exemplify how the results of Section 5.3 can be used. For ease of reading, several details of the rigorous derivation are deferred to the appendix.

Example 5.5.1. In this example we deal with a problem originally considered by de Finetti in [**53**] (see also Section 8.2.6.1 in [**1**]). Here, a first-line insurer has n sub-portfolios with insurance risks $X_1, \dots, X_n$ and is looking for a reinsurance contract $R = (R_1, \dots, R_n)$ that minimizes the aggregate expected loss after reinsurance, under a constraint on the retained aggregate variance. Assume that the premium for the i-th contract is computed according to an expected value principle with safety loading β_i, so that the total loss experienced by the first line insurer is given by

$$\sum_{i=1}^{n}(X_i - R_i + (1 + \beta_i)\mathbb{E}[R_i]).$$

Assume that all the β_i's are different (e.g. because they represent different business lines) and w.l.o.g. ordered increasingly, i.e., $0 < \beta_1 < \cdots < \beta_n$. Taking expectations in the previous equation, we see that the risk measure can be chosen as

$$\mathcal{P}(\eta) = \int \sum_{i=1}^{n} \beta_i y_i \, \eta(dx, dy). \tag{131}$$

Assume further a bound on the retained variance, i.e., that for a constant c s.t. $0 < c < \mathrm{Var}(\sum_{i=1}^n X_i)$, the contract is required to satisfy $\mathrm{Var}(\sum_{i=1}^n (X_i - R_i)) \le c$. In our notation, $\mathcal{G}$ can then be written as

$$\mathcal{G}(\eta) = \int \left(\sum_{i=1}^{n} (x_i - y_i) \right)^2 - \left(\sum_{i=1}^{n} \int (x_i - y_i) \, \eta(dx, dy) \right)^2 \eta(dx, dy) - c.$$

Letting η^* be any optimal contract, we observe that the functionals $\mathcal{P}$ and $\mathcal{G}$ satisfy the conditions of Propositions 5.3.4 and 5.3.7 with functions $p_{\eta^*}, g_{\eta^*} : \mathbb{R}_+^n \times \mathbb{R}_+^n \to \mathbb{R}$ given by

$$p_{\eta^*}(x,y) = \sum_{i=1}^n \beta_i y_i,$$
$$g_{\eta^*}(x,y) = \left(\sum_{i=1}^n (x_i - y_i)\right)^2 - 2\sigma \sum_{i=1}^n (x_i - y_i),$$

with $\sigma = \sum_{i=1}^n \overline{(\pi_i \circ T)_\# \eta^*}$. Observe further that if η is the contract that specifies full reinsurance, then

$$\mathcal{G}(\eta^*) + d\mathcal{G}(\eta^*; \eta - \eta^*) = -2c < 0,$$

so it follows that there exists $\lambda^* \geq 0$ such that the support of η^* is contained in the minima of $h(x,\cdot) = p_{\eta^*}(x,\cdot) + \lambda^* g_{\eta^*}(x,\cdot)$. Note, however, that we cannot have $\lambda^* = 0$, since the minimum of $p(x,\cdot)$ occurs at $y = 0$ and this would imply that η^* is the contract for which no reinsurance takes place, violating the condition $\mathcal{G}(\eta^*) < 0$. By obtaining the point-wise minima, it can be seen that the optimal reinsurance contract is deterministic and component-wise given by

$$R_i(x) = \min\left(\left(\sum_{j=i}^n x_j - \frac{\beta_i}{2\lambda^*} - \sigma\right)_+, x_i\right), \tag{132}$$

see Appendix 5.A for details.

Remark 5.5.2. In the original problem considered by de Finetti, claims are independent and only quota-share contracts are considered for each subportfolio, i.e. with $R_i(x) = a_i x_i$. The optimal proportions are then determined as

$$a_i = \left(1 - \frac{\beta_i \mathbb{E}[X_i]}{2\lambda_{\mathrm{Fin}} \mathrm{Var}(X_i)}\right)_+$$

with $\lambda_{\mathrm{Fin}} = \frac{1}{4c}\sum_{i=1}^n \frac{(\beta_i \mathbb{E}[X_i])^2}{\mathrm{Var}(X_i)}$, cf. [53]. Observe that for some values of the β_i's, the optimal proportions are then zero, which implies no reinsurance for that subportfolio. In contrast, the overall optimal solution (132) of this problem (beyond the restriction to proportional treaties) leads to reinsurance for all subportfolios regardless of the size of risk loading (if the random variables are not almost surely bounded). See also Remark 5.5.9 for an interpretation of the optimal solution (132). As a numerical illustration, consider the case where X_1 has a $\Gamma(1/2, 1/2)$ distribution and X_2 a (shifted) Pareto distribution with p.d.f. given by

$$f_{X_2}(x) = 324\,(x+3)^{-5}, \quad x \geq 0,$$

so that $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1$ and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 2$. Let $\beta_1 = 0.1$ and $\beta_2 = 0.25$ (reflecting that relative risk loadings are typically higher for heavy-tailed risks). Assume that the first-line insurer would like to halve the retained total variance, so that the bound on the retained variance is given by $c = 2$. The optimal parameters for (132) are $\sigma = 1.8026351$ and $\lambda^* = 0.0443408$, while the optimal proportions from de Finetti's solution are $a_1 = 0.6286093$ and $a_2 = 0.0715233$. Letting η_{Fin} denote the joint distribution implied by de Finetti's solution, we therefore obtain $\mathcal{P}(\eta_{\mathrm{Fin}}) = 0.0807417$, while $\mathcal{P}(\eta^*) = 0.0232948$ for the overall optimal contract (132). Observe that while this represents an improvement of 71.14% of the objective function (131), the overall expected loss for the cedent under η_{Fin} is 2.0807417, while under η^* it is 2.0232948, so the improvement is still visible, but considerably smaller.

Example 5.5.3. Consider a variant of Example 5.5.1, where instead of fixing the variance, the cedent has a constraint on the Value-at-Risk (at some level α) of the total retained amount. Assume that X has a density. The functionals are then given by

$$\mathcal{P}(\eta) = \int \sum_{i=1}^{n} \beta_i y_i \, \eta(dx, dy),$$
$$\mathcal{G}(\eta) = \widehat{\mathrm{VaR}}_\alpha(T_{S\#}\eta) - c,$$

where $0 < \beta_1 < \cdots < \beta_n$, $c \geq 0$ and $T_S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is the linear operator defined by $T_S(x,y) = \sum_{i=1}^n (x_i - y_i)$. Observe that we can restrict ourselves to the case $0 < c < \mathrm{VaR}\left(\sum_{i=1}^n X_i\right)$ to avoid the optimal contract being the one given by full or no reinsurance. In this case Proposition 5.3.7 is not applicable and one has to resort to the (weaker) Proposition 5.3.4. Let η^* be an optimal reinsurance contract and $v^* = \widehat{\mathrm{VaR}}_\alpha(T_{S\#}\eta)$. Set

$$\mathcal{C} = \{\eta \in \mathscr{M} \mid \widehat{\mathrm{VaR}}_\alpha(T_{S\#}\eta) = v^*\}. \tag{133}$$

Observe that $\eta^* \in \mathcal{C}$, $\mathcal{C}$ is convex and for every $\eta \in \mathcal{C}$,

$$d\mathcal{P}(\eta^*; \eta - \eta^*) = \int \sum_{i=1}^{n} \beta_i y_i \, (\eta - \eta^*)(dx, dy) \text{ and } d\mathcal{G}(\eta^*; \eta - \eta^*) = 0.$$

Letting $U \subset \mathscr{S}$ be the space of measures with finite marginal expectations, the conditions of Proposition 5.3.4 are satisfied with $p(x,y) = \sum_{i=1}^n \beta_i y_i$ (independent of η^*). Setting

$$\begin{aligned} D_1 &= \left\{ (x,y) \in \mathcal{A}_R \mid \sum_{i=1}^{n} (x_i - y_i) < v^* \right\}, \\ D_2 &= \left\{ (x,y) \in \mathcal{A}_R \mid \sum_{i=1}^{n} (x_i - y_i) > v^* \right\}, \\ D_3 &= \left\{ (x,y) \in \mathcal{A}_R \mid \sum_{i=1}^{n} (x_i - y_i) = v^* \right\}, \end{aligned} \tag{134}$$

then, for every $(x,y) \in \mathrm{Supp}(\eta^*)$, we have

$$I(x,y) = \{z \in [0,x] \mid (x,z) \in \overline{D_1}\}$$

if $(x,y) \in D_1$, while $0 \in I(x,y)$ if $(x,y) \in D_2$, where $I(x,y)$ is the set over which minimization occurs in (118). For $(x,y) \in D_2$, the minimum of $p(x,\cdot)$ is achieved at $y = 0$, regardless of x. Similarly, if $(x,y) \in D_1$, then we must have $\sum_{i=1}^n x_i < v^*$ and $y = 0$, so that

$$\mathrm{Supp}(\eta^*) \subset D_3 \cup \{(x,0) \mid x \in \mathbb{R}_+^n\}.$$

Since $T_S(D_3) = \{v^*\}$, it follows that η^* has a point mass at v^*. Hence, since X is absolutely continuous and v^* is the $(1-\alpha)$-th quantile of $T_{S\#}\eta$, we obtain $T_{S\#}\eta([0,v^*]) = 1 - \alpha$. Thus

$$\eta^*(D_3) = 1 - \alpha - \mu\left(\left\{ x \in \mathbb{R}_+^n \mid \sum_{i=1}^{n} x_i < v^* \right\} \right) := m.$$

The quantity on the right-hand side of the last equation is fixed (with the knowledge of v^*), and we are therefore left with the task of assigning mass m to D_3. Now, the assumption $0 < \beta_1 < \cdots < \beta_n$ implies that $p(x,\cdot)$ has an unique minimum on $[0,x]$. Denoting this minimum by $y^*(x)$, it can be seen that the optimal way of assigning this mass is to do it through points of the form $(x, y^*(x)) \in D_3$ for the x's that produce the smallest values of $q(x) := p(x, y^*(x))$. Since

$$y^*(x) = (\min (Q_1(x), x_1), \ldots, \min (Q_{n-1}(x), x_{n-1}), Q_n(x)),$$

with

$$Q_i(x) = \left(\sum_{j=i}^{n} x_j - v^* \right)_+, \quad i = 1, \ldots, n,$$

and $y^*(x)$ is unique, it follows that η^* is deterministic and is given by the function

$$R(x) = \begin{cases} y^*(x), & \text{if } x \in E \\ 0, & \text{otherwise.} \end{cases} \tag{135}$$

Here $E = \{x \in \mathbb{R}^n_+ \mid q(x) \le d\}$ where d is chosen so that $\mu(E) = 1 - \alpha$. More explicitly, we get

$$R(x) = \left(x_1, \ldots, x_{i-1}, \sum_{j=i}^{n} x_j - v^*, 0, \ldots, 0 \right) \tag{136}$$

if $\sum_{j=1}^{i-1} \beta_j x_j + \beta_i \sum_{j=i}^{n} x_j - \beta_i v^* \le d$, $\sum_{j=i+1}^{n} x_j \le v^*$ and $\sum_{j=i}^{n} x_j > v^*$, and $R(x) = 0$ otherwise. See Appendix 5.B for details.

A couple of remarks are in order.

Remark 5.5.4. The choice of $\mathcal{C}$ as in the previous example is motivated from the similar result found in [**69**] and is convenient because it allows us to get rid of the terms coming from $\mathcal{G}$. Observe, however, that other sensible choices could have been

$$\mathcal{C}' = \{\eta \in \mathcal{D} \mid \widehat{\mathrm{VaR}}_\alpha(T_{S\#}\eta) \le v^*\}$$

or

$$\mathcal{C}'' = \{\eta \in \mathcal{D} \mid \widehat{\mathrm{VaR}}_\alpha(T_{S\#}\eta) \ge v^*\},$$

where $\mathcal{D}$ is the set appearing in the statement of Proposition 5.3.2. While using these sets one might obtain deeper information from the measures $\eta_{x,y,t,\varepsilon}$, one is also left with the task of fully describing $\mathcal{C}'$ and $\mathcal{C}''$ and, in particular, proving that they are convex and satisfy (115). Hence, when choosing $\mathcal{C}$, one needs to consider the trade-off between letting it be too big and the simplicity by which one can describe it (as otherwise one could simply take the largest convex set containing η^* and satisfying (115)).

Remark 5.5.5. Observe that for $n = 1$, the result in Example 5.5.3 agrees with the result in Corollary 1 of [**69**] and the current framework might help to explain the resemblance of these results to the one in [**115**]: Equation (116) implies that Value-at-Risk constraints might be recast as an optimization of a Lagrangian, so both approaches lead to the same sort of optimal contracts.

Remark 5.5.6. One can easily generalize the approach from Example 5.5.3 to consider slightly more complex constraints involving two or more levels for the VaR, for example, $\mathcal{G} : \mathscr{M} \to \mathbb{R}^m$ with i-th component given by

$$g_i(\eta) = \widehat{\mathrm{VaR}}_{\alpha_i}(T_{S\#}\eta) - c_i,$$

with $1 > \alpha_1 > \cdots > \alpha_m > 0$ and $0 \le c_1 \le \cdots \le c_m$. In this case, one could choose the set $\mathcal{C}$ as

$$\mathcal{C} = \{\eta \in \mathcal{D} \mid \widehat{\mathrm{VaR}}_{\alpha_i}(T_{S\#}\eta) = v_i^*,\ i = 1, \ldots, m\},$$

where $v_i^* = \widehat{\mathrm{VaR}}_{\alpha_i}(T_{S\#}\eta)$ for an optimal reinsurance contract η^*. Alternatively, this set can also be used to minimize weighted combinations of values-at-risk at different levels, i.e., risk measures of the form $\sum_{i=1}^m \beta_i \widehat{\mathrm{VaR}}_{\alpha_i}(T_{S\#}\eta)$ under, say, a budget constraint in the premium. Such an approach is implicitly used in reinsurance practice when trying to fix quantiles of the target distribution of the cedents in the context of regulatory ruin (where different measures apply to different "degrees" of insolvency, cf. [**1**, Ch.8]).

For $n = 1$, one can consider another direction of generalization of Example 5.5.3: let ρ and ω denote non-decreasing functions with $\rho(0) = \omega(0) = 0$ and $\rho(1) = 1$. For $\delta > 0$ one can consider the risk measure given by

$$\mathcal{P}(\eta) = \int (1-\delta)(x-y)\,\eta(dx,dy) + \int_0^\infty \omega\left(\pi_{2\#}\eta(t,\infty)\right)\,dt + \delta \int_0^\infty \rho\left(\pi_{2\#}\eta(t,\infty)\right)\,dt.$$

This risk functional corresponds to minimizing the risk-adjusted liability of the cedent, a scenario considered by [**47**]. The risk incurred by the first-line insurer is measured through a distortion risk measure, which is then represented by the last integral in the definition of $\mathcal{P}$. We do not study this case here, but note that the techniques seem to extend to this case by letting $\mathcal{C}$ capture the discontinuities/points of non-differentiability of the functions ω and ρ.

Remark 5.5.7. Notice that in both examples above, the solution is not fully specified, but instead is given in terms of some unknown parameters (λ^* and σ in Example 5.5.3, and v^* in Example 5.5.1). As mentioned earlier, this is an unavoidable feature of our procedure, which arises from the dependence of p_{η^*} and g_{η^*} on η^*. However, through (132) or (136) we can obtain an expression for $\mathcal{P}(\eta^*)$ where the only unknowns are these parameters. Hence, we may instead treat them as variables and optimize over them (ensuring that the constraints are still satisfied), thus obtaining a full description of the optimal contracts.

Remark 5.5.8. At this point, one can see that Propositions 5.3.4 and 5.3.7 can be applied to several situations and are particularly well-suited whenever the risk measure or the constraints can be written by means of (functions of) integrals. As a further example, we mention that the methodology can be applied to deal with more complex situations such as the ones considered in [**79**]. There, one would like to minimize risk measures of the form

$$\mathcal{P}(\eta) = f\left(\int_{\mathbb{R}^n_+} p_1(x-y)\,\eta(dx,dy), \dots, \int_{\mathbb{R}^n_+} p_\ell(x-y)\,\eta(dx,dy)\right)$$

subject to the constraints $\mathcal{G} = (g_1, \dots, g_m)$ given by

$$g_i(\eta) = h_i\left(\int_{\mathbb{R}^n_+} q_{i,1}(y)\,\eta(dx,dy), \dots, \int_{\mathbb{R}^n_+} p_{i,\ell_i}(y)\,\eta(dx,dy)\right),$$

where all the p_i's and $q_{i,j}$'s are (multivariate) rational functions and f and the h_i's are differentiable.[6] For some particular choices of functions, one can even immediately see that the solutions are deterministic by means of the techniques developed in Section 5.4 (for example, for the case when $\mathcal{P}$ is given as the variance and the constraints depend only on the second marginal, which corresponds to the situation in Theorem 1 in [**79**]).

Loosely speaking, our results also give an intuitive explanation to the ubiquity of stop-loss contracts in optimal reinsurance problems: often, the function $p_{\eta^*} + \lambda^* \cdot g_{\eta^*}$ can be written in

[6]The setting in [**79**] allows for equalities in the constraints, and for f and the h_i's to not be differentiable at the expense of being increasing in one variable. Equalities can be handled in the same way as in Examples 5.5.1 and 5.5.3, and while we cannot get rid of the differentiability requirement, all of the examples in [**79**] seem to satisfy this as well.

the form

$$p_{\eta^*}(x,y) + \lambda^* \cdot g_{\eta^*}(x,y) = \hat{p}_{\eta^*}(x-y) + \lambda^* \cdot \hat{g}_{\eta^*}(x-y)$$

for some functions $\hat{p}_{\eta^*}$ and $\hat{g}_{\eta^*}$ such that $\hat{p}_{\eta^*} + \lambda^* \cdot \hat{g}_{\eta^*}$ has one minimum. According to Propositions 5.3.4 and 5.3.7, the optimal contract then needs to satisfy $x - y = c$ for some constant c, which together with the condition $(x,y) \in \mathcal{A}_R$, implies that $y = (x-c)_+$, which is the form of a stop loss contract.

Remark 5.5.9. While the contracts in (132) and (136) are deterministic — in the sense that knowledge of X implies knowledge of $R(X)$, — the contracts for the individual subportfolios are still random, since X_i stand-alone is not enough to fully specify $R_i(X)$. This is in contrast to [**70**], where it is enforced that, conditional on X_i, $R_i(X)$ is independent of the remaining contracts in the portfolio. While going into separate contracts with potentially different reinsurers with such marginally random contracts may be challenging in current reinsurance practice, reinsuring all these subportfolios with the same reinsurer (but possibly different safety loadings β_i, e.g. due to different business lines) may be quite feasible. Compared to alternatives, such a contract just leads to a slightly more involved (but deterministic) formula for settling the overall reinsured amount once all claim data for the considered time period are available. In some sense, a part of the risk diversification is done *in house* this way, which is quite common for certain types of aggregate reinsurance covers in practical use, see e.g. [**1**].

The following examples (re)examine some of the classical problems in optimal reinsurance through the lens of optimal transport. We use the results we developed in Section 5.4.

Example 5.5.10. Let $n = 1$ and assume that X has finite variance. Let $\mathcal{P}$ be given by

$$\mathcal{P}(\eta) = \widehat{\mathrm{Var}}(T_\#\eta) := \int x^2\, T_\#\eta(dx) - \left(\int x\, T_\#\eta(dx)\right)^2$$

and $S = \{\eta \in \mathscr{M} \mid \int y\, \pi_{2\#}\eta(dy) = c\}$ for some $c \geq 0$. This is the classical example (see e.g. [**97**]) where the objective is to minimize the retained variance of the insurer subject to a fixed reinsurance premium which is computed through the expected value principle. In this case, the optimal reinsurance contract is known to have the deterministic form

$$\eta^* = (\mathrm{Id}, R_{SL,a^*})_\#(\mu) \quad \text{for some } a^* \geq 0, \tag{137}$$

where, for $a \geq 0$, $R_{SL,a}$ is the function on $\mathbb{R}$ given by $R_{SL,a}(x) = (x-a)_+$, i.e., a stop-loss contract is optimal. Note that finiteness of the variance of X implies that the set S is closed. In order to apply the results from Section 5.4, observe that

$$\begin{aligned}\inf_{\eta \in S} \mathcal{P}(\eta) &= \inf_{\eta \in \mathscr{M}, \overline{\pi_{2\#}\eta} = c} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 \eta(dx,dy) - \left(\int_{\mathbb{R}\times\mathbb{R}} (x-y)\eta(dx,dy)\right)^2 \\ &= \inf_{\eta \in \mathscr{M}, \overline{\pi_{2\#}\eta} = c} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 \eta(dx,dy) - (\overline{\mu} - c)^2 .\end{aligned}$$

Since the second term is constant, this corresponds to the problem

$$\inf_{\eta \in \mathscr{M}, \overline{\pi_{2\#}\eta} = c} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 \eta(dx,dy) = \inf_{\substack{\nu \in \mathcal{P}(\mathbb{R}), \\ \overline{\nu} = c, \nu \prec_1 \mu}} \inf_{\substack{\eta \in \Pi(\mu,\nu), \\ \eta(y \leq x) = 1}} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 \eta(dx,dy). \tag{138}$$

From (138), we observe that the conditions of Proposition 5.4.1 are satisfied. Hence, it follows that, with

$$x \mapsto g_\nu(x) := F_\nu^{-1} \circ F_\mu(x),$$

the coupling $\pi_\nu := (\mathrm{Id}, g_\nu)_\# \mu$ is optimal for the inner problem in the right-hand side of (138), so we get

$$\inf_{\eta\in\mathscr{M}, \overline{\pi_{2\#}\eta}=c} \int_{\mathbb{R}\times\mathbb{R}} (x-y)^2 \eta(dx,dy) = \inf_{\substack{\nu\in\mathcal{P}(\mathbb{R}),\\ \overline{\nu}=c, \nu\prec_1\mu}} \int_{\mathbb{R}} (x-g_\nu(x))^2 \mu(dx). \tag{139}$$

From this we can see how the optimizer is given by the contract η^* given in (137), with $R_{SL,a^*} = g_{\nu^*}$, for ν^* minimizer in (139). Indeed, we want to minimize the integral w.r.t. μ of $(x - g_\nu(x))^2$, over functions g_ν which are non-increasing, below Id, and such that the area below them (i.e. the integral w.r.t. μ) is fixed (equal to c). Then clearly the optimal g_{ν^*} is parallel to Id, thus of the form $(x-a^*)_+$, with a^* determined by the constraint $\mathbb{E}[(X-a^*)_+] = \int y\,(g_{\nu^*\#}\mu)(dy) = \int y\,\nu^*(dy) = c$. Hence, the OT approach provides an alternative proof of this classical result.

Example 5.5.11. If we modify the previous example by setting instead

$$S = \{\eta \in \mathscr{M} \mid \widehat{\mathrm{Var}}(\pi_{2\#}\eta) = c\}$$

for some $c \geq 0$, we obtain the situation where the objective is still to minimize the retained variance, but now subject to a fixed reinsurance premium loading that is proportional to the variance (cf. [**97**]). In this case, the optimal reinsurance contract is known to be deterministic and of the form $\eta^* = (\mathrm{Id}, R_{QS,a^*})_\#(\mu)$ for some $0 \leq a^* \leq 1$, where $R_{QS,a}(x) = ax$, i.e., a so-called quota-share contract is optimal. Let $\mathcal{P}$ denote the same functional as in the previous case, and observe that it still satisfies Assumptions (A1) and (A2) above with function p_η given by

$$p_\eta(x,y) = (x-y)^2 - 2\overline{T_\#\eta}(x-y).$$

This function and S satisfy the conditions from Proposition 5.4.1, so it follows that, for $\nu \in \pi_2(\mathcal{S})$, the minimum of (124) is achieved through couplings π_ν of the form $\pi_\nu := (\mathrm{Id}, g_\nu)_\#\mu$ with

$$x \mapsto g_\nu(x) := F_\nu^{-1} \circ F_\mu(x).$$

Plugging this coupling into the definition of $\mathcal{P}$, we obtain

$$\begin{aligned}\mathcal{P}(\pi_\nu) &= \int_0^\infty (x-g_\nu(x))^2\ \mu(dx) - \left(\int_0^\infty (x-g_\nu(x))\ \mu(dx)\right)^2 \\ &= \int_0^1 \left(F_\mu^{-1}(x) - F_\nu^{-1}(x)\right)^2\ dx - \left(\int_0^1 F_\mu^{-1}(x) - F_\nu^{-1}(x)\ dx\right)^2\end{aligned}$$

and the problem is reduced to finding the optimal distribution function F_ν. We can phrase this problem in terms of functions: consider the operator $\mathcal{F} : L^2([0,1]) \to \mathbb{R}$ given by

$$\mathcal{F}(f) = \int_0^1 \left(F_\mu^{-1}(x) - f(x)\right)^2\ dx - \left(\int_0^1 F_\mu^{-1}(x) - f(x)\ dx\right)^2.$$

We want to minimize this functional subject to the constraints $0 \leq f \leq F_\mu^{-1}$, f non-decreasing and

$$\int_0^1 f(x)^2\ dx - \left(\int_0^1 f(x)\ dx\right)^2 = c. \tag{140}$$

We can first look at the problem that considers only the last constraint and examine the associated Lagrange operator, $\mathcal{L} : L^2([0,1]) \times \mathbb{R} \to \mathbb{R}$ given by

$$\mathcal{L}(f,\lambda) = \mathcal{F}(f) + \lambda \int_0^1 f(x)^2\ dx - \lambda \left(\int_0^1 f(x)\ dx\right)^2 - \lambda c.$$

For each $f \in L^2([0,1])$ and $\lambda \in \mathbb{R}$, the functional derivative of this operator with respect to h is the functional $\mathcal{L}'(f,\lambda;\cdot) : L^2([0,1]) \to \mathbb{R}$ given by

$$\begin{aligned}\mathcal{L}'(f,\lambda;h) &= 2\int_0^1 (F_\mu^{-1}(x) - f(x))h(x)\,dx - 2\int_0^1 (F_\mu^{-1}(x) - f(x))\,dx \int_0^1 h(x)\,dx \\ &\quad + 2\lambda \int_0^1 f(x)h(x)\,dx - 2\lambda \int_0^1 f(x)\,dx \int_0^1 h(x)\,dx \\ &= 2\int_0^1 \left(F_\mu^{-1}(x) + (\lambda-1)f(x) - \int_0^1 (F_\mu^{-1}(y) + (\lambda-1)f(y))\,dy\right) h(x)\,dx.\end{aligned}$$

If f is to be an extreme point of $\mathcal{L}$, then $\mathcal{L}'(f,\lambda;h)$ is to be zero for every $h \in L^2([0,1])$, which can happen if and only if

$$F_\mu^{-1}(x) + (\lambda-1)f(x) - \int_0^1 (F_\mu^{-1}(y) + (\lambda-1)f(y))\,dy = 0, \quad x \in [0,1].$$

Hence, the function is constant and $F_\mu^{-1} = (1-\lambda)f + a$ for some $a \in \mathbb{R}$. Regardless of a, the function will satisfy (140) as long as we choose λ such that $(1-\lambda)^2 = \mathrm{Var}(X)/c$. Observe then that all the other constraints will be satisfied by choosing $\lambda = 1 - \sqrt{\mathrm{Var}(X)/c}$ and any $0 \le a \le F_\mu^{-1}(0)$. The distribution function thus obtained corresponds to the reinsurance contract given by

$$R(X) = \frac{X-a}{(1-\lambda)},$$

which for $a = 0$ is the quota-share contract known to be optimal **[97]** (and any other choice of a would just lead to a deterministic ('side') payment from the reinsurer to the insurer (see e.g. **[65]**), which would be priced in the reinsurance premium in an additive way, as its variance is zero, and so would only lead to a deterministic additional exchange and serve no purpose). Again, the OT approach in this way provides an alternative proof of this classical result.

Observe that in this situation one cannot directly apply Proposition 5.4.2: given an arbitrary reinsurance treaty η, the monotonic rearrangement between μ and $\nu = T(\eta)$ leads to a function R such that $R(X)$ is distributed according to ν. However, it is in general not the case that $X - R(X)$ is distributed according to $\pi_2(\eta)$, so we cannot guarantee that $\mathrm{Var}(X - R(X)) = c$.

For the next example, we take the viewpoint of the reinsurer in the optimization problem. It will lead to a situation where introducing *external randomness* is indeed optimal.

Example 5.5.12. For simplicity of exposition, here the second marginal of a reinsurance treaty will refer to the deductible rather than the reinsured amount. For each $k = 1,\dots,n$, let $\nu_k \in \mathscr{P}(\mathbb{R}^+)$ denote a predefined distribution. For any lower semi-continuous $\mathcal{P}$, we can set

$$S = \{\eta \in \mathscr{M} : \pi_{2,k\#}\eta = \nu_k,\ k = 1,\dots,n\}.$$

One can interpret this situation as follows: n insurers ask each for a target distribution ν_k, $k = 1,\dots,n$, after reinsurance, and the reinsurer tries to minimize $\mathcal{P}$ respecting these target distributions (potentially involving the introduction of randomized treaties). Phrased in optimal transport terms, this example corresponds to a problem in the area of multi-marginal optimal transport, a generalization from the classical transport problem in which there might be more than one target measure. While we do not attempt to solve the problem in general, we point out some of the insights obtained from seeing the problem from this perspective and solve it for one particular case. Consider the case $n = 2$, μ absolutely continuous with finite second moments and $\mathcal{P}$ given as the variance of the sum of the reinsured amounts. For any $\eta \in S$, we then have

$$\mathcal{P}(\eta) = \int_{\mathbb{R}_+^4} (x_1 - y_1 + x_2 - y_2)^2 \, \eta(dx_1, dx_2, dy_1, dy_2) - (\overline{\mu_1} - \overline{\nu_1} + \overline{\mu_2} - \overline{\nu_2})^2 .$$

The second term on the right-hand side is fixed. Hence, the problem is equivalent to minimizing the functional

$$\mathcal{Q}(\eta) = \int_{\mathbb{R}_+^4} (x_1 - y_1 + x_2 - y_2)^2 \, \eta(dx_1, dx_2, dy_1, dy_2)$$

on $\Pi(\mu, \nu_1, \nu_2) \cap \mathscr{M}$, the set of couplings between μ, ν_1 and ν_2 supported on $\mathcal{A}_R$. Since X_1 and X_2 have finite variance, this minimization problem is finite. Now, for any coupling η (supported on $\mathcal{A}_R$), we can disintegrate η with respect to μ, obtaining a family of probability measures $(\vartheta_{x_1,x_2})_{(x_1,x_2)\in\mathbb{R}_+^2}$, and we can write

$$\mathcal{Q}(\eta) = \int_{\mathbb{R}_+^2} \int_{[0,x_1]\times[0,x_2]} (x_1 - y_1 + x_2 - y_2)^2 \, \vartheta_{x_1,x_2}(dy_1, dy_2) \mu(dx_1, dx_2).$$

Observe that the only dependency on the coupling is given by the inner integral, so if η is optimal, then μ-a.e., ϑ_{x_1,x_2} is optimal for the transport problem with cost $(y_1, y_2) \mapsto (y_1 + y_2 - x_1 - x_2)^2$ when we fix (x_1, x_2). This (sub-)transport problem has as solution the antimonotonic rearrangement of its marginals, and the problem is therefore transformed into finding the optimal marginals. Considering that probability measures are determined by their distribution functions, the problem can be rephrased as follows: find functions $f, g : [0,1] \times \mathbb{R}_+^2 \to \mathbb{R}$ such that if U is a uniformly distributed random variable independent of X_1 and X_2, then $f(U, X_1, X_2)$ and $g(1-U, X_1, X_2)$ have distribution ν_1 and ν_2 respectively,

$$0 \le f(u, x_1, x_2) \le x_1, \text{ and } 0 \le g(u, x_1, x_2) \le x_2, \quad (u, x_1, x_2) \in [0,1] \times \mathbb{R}_+^2,$$

and f and g minimize the integral

$$\int_{\mathbb{R}_+^2} \int_0^1 f(u, x_1, x_2) g(1-u, x_1, x_2) \, du \, \mu(dx_1, dx_2),$$

among all possible pairs of functions satisfying these constraints. The optimal coupling will then be given as the distribution of

$$(X_1, X_2, f(U, X_1, X_2), g(1-U, X_1, X_2)).$$

As can be seen from quick inspection, this task is challenging and for arbitrary μ, there is no guarantee that there even exist solutions that can be expressed in terms of elementary functions[7], so that numerical solutions have to be considered. Nevertheless, we provide an illustration of how this can lead to explicit results in specific cases.

Assume X_1 and X_2 are independent, X_1 has a lognormal distribution with p.d.f. given by

$$f_{X_1}(x) = \frac{1}{x\sqrt{2\pi \log(3)}} \exp\left(-\frac{(\log(\sqrt{3}x))^2}{2\log(3)} \right), \quad x > 0,$$

and X_2 a (shifted) Pareto distribution with p.d.f given by

$$f_{X_2}(x) = 324\,(x+3)^{-5}, \quad x \ge 0.$$

Here the parameters are chosen such that $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1$ and $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 2$. As there is no standard solution method for the multi-marginal transportation problem with

[7]Observe, however, that the previous considerations show that such functions exist.

arbitrary cost, we utilize a discretized setting (see e.g. [98]). Let $q \in]0,1[$ and define $u_i = F_{X_i}^{-1}(q)$, $i = 1,2$. For $N \in \mathbb{N}$, we introduce the variables $\tilde{X}_1$ and $\tilde{X}_2$ such that, for $i = 1$ and $i = 2$,

$$\tilde{X}_i = \frac{ku_i}{N} \text{ with probability } F_{X_i}\left(\frac{ku_i}{N}\right) - F_{X_i}\left(\frac{(k-1)u_i}{N}\right),\ k = 1,\dots,N-1$$

and

$$\tilde{X}_i = u_i \text{ with probability } 1 - F_{X_i}\left(\frac{(N-1)u_i}{N}\right).$$

These variables are simply the result of binning the X_i's into N bins of equal length up to a (high) quantile u_i and assigning probabilities according to their distributions (and putting the remaining mass into the last bin to account for the unboundedness of the distributions). Let $\tilde{Y}_1$ and $\tilde{Y}_2$ be two random variables such that

$$\tilde{Y}_1 \overset{d}{=} 0.5\tilde{X}_1, \quad \tilde{Y}_2 \overset{d}{=} \min(\tilde{X}_2, 0.5) + 0.25(\tilde{X}_2 - 0.95)_+,$$

where $\overset{d}{=}$ denotes equality in distribution. The idea behind this choice is that $\tilde{Y}_1$ could arise from applying a quota-share contract (with proportionality factor 0.5) to $\tilde{X}_1$, while $\tilde{Y}_2$ could be the retained amount from a bounded stop-loss contract on $\tilde{X}_2$ (with deductible 0.5 and layer size 0.45), where the reinsurer still takes 75% of the exceedance above that layer. So these are the target distributions $\tilde{\nu}_1$ and $\tilde{\nu}_2$ of the two insurers, and the task is now to see how the reinsurer can offer these while keeping the variance of $(\tilde{X}_1 - \tilde{Y}_1) + (\tilde{X}_2 - \tilde{Y}_2)$ (the reinsured amount) minimal, for instance in order to provide a competitive reinsurance premium.

By letting M denote the amount of distinct values taken by $\tilde{Y}_2$, in the following we will implicitly assume that these random variables and distributions are defined on the space $\Omega' = \{1,\dots,N\}^3 \times \{1,\dots,M\}$ and are such that, for example, $\tilde{Y}_1(i,j,k,l) = \tilde{Y}_1(k)$ is the k-th value taken by $\tilde{Y}_1$ in increasing order and $\tilde{\nu}_1(k) = \mathbb{P}[\tilde{Y}_1 = \tilde{Y}_1(k)]$. Similar definitions apply to the other random variables. The optimal transport problem is then equivalent to finding a 4-dimensional array P in $\mathbb{R}_+^{N\times N\times N\times M}$ such that

$$\begin{aligned}
\sum_{k,l} P_{i,j,k,l} &= \tilde{\mu}(i,j), \quad (i,j) \in \{1,\dots,N\}^2,\\
\sum_{i,j,l} P_{i,j,k,l} &= \tilde{\nu}_1(k), \quad k \in \{1,\dots,N\},\\
\sum_{i,j,k} P_{i,j,k,l} &= \tilde{\nu}_2(l), \quad l \in \{1,\dots,N\},\\
P_{i,j,k,l} &= 0 \text{ if } \tilde{Y}_1(k) > \tilde{X}_1(i) \text{ or } \tilde{Y}_2(l) > \tilde{X}_2(j),
\end{aligned} \tag{141}$$

and P minimizes the sum

$$\sum_{i,j,k,l} (\tilde{X}_1(i) - \tilde{Y}_1(k) + \tilde{X}_2(j) - \tilde{Y}_2(l))^2 P_{i,j,k,l}$$

among all arrays satisfying (141). This is a linear optimization problem. We want to use standard linear optimization techniques to solve it, so we cast these equations into standard form: we start by "flattening" P into an element $p \in \mathbb{R}^{N^3M}$ by making the (i,j,k,l) entry of P into the $i + N(j-1) + N^2(k-1) + N^3(l-1)$ entry of p. Similarly, we let the cost be represented by a vector $c \in \mathbb{R}^{N^3M}$ with $i + N(j-1) + N^2(k-1) + N^3(l-1)$ entry equal to $(\tilde{X}_1(i) - \tilde{Y}_1(k) + \tilde{X}_2(j) - \tilde{Y}_2(l))^2$. Denoting by $\mathbb{1}_N$ the column vector of dimension

N filled with ones, $\mathbb{I}_N$ the identity matrix of dimension N and $\otimes$ the Kronecker product, the $(N^2+N+M)\times N^3M$ matrix

$$A = \begin{bmatrix} \mathbb{1}_{NM}^\intercal \otimes \mathbb{I}_{N^2} \\ \mathbb{1}_M^\intercal \otimes \mathbb{I}_N \otimes \mathbb{1}_{N^2}^\intercal \\ \mathbb{I}_M \otimes \mathbb{1}_{N^3}^\intercal \end{bmatrix}$$

can be used to encode the first three constraints in (141). Indeed, if $\theta \in \mathbb{R}^{N^2+N+M}$ is given as the stacking of $\tilde{\mu}$, $\tilde{\nu}_1$ and $\tilde{\nu}_2$ (flattening first $\tilde{\mu}$, so that $\tilde{\mu}(i,j)$ is the $i+N(j-1)$ entry of θ), we see that P satisfies the first three equalities in (141) if and only if $Ap=\theta$. Finally, we can enforce the last constraint in (141) by deleting the columns and entries of A, p and c for which we have $\tilde{Y}_1(k) > \tilde{X}_1(i)$ or $\tilde{Y}_2(l) > \tilde{X}_2(j)$ (that is, we delete the $i+N(j-1)+N^2(k-1)+N^3(l-1)$ column of A if this condition is satisfied). Let B, q and d denote, respectively, the matrix and vectors obtained this way and K the number of columns that remained after this operation. The optimization problem then is of the form

$$\text{Minimize } d^\intercal q \text{ subject to } Bq=\theta \text{ and } q\geq 0.$$

Although potentially high-dimensional, this is a relatively easy linear optimization exercise. For $M=40$, we solve it using standard linear optimization packages. Specifically, we use the `linprog` routine within the `SciPy` library, which implements the HiGHS software for linear optimization. The results for some of the bivariate distributions of $(\tilde{X}_1,\tilde{X}_2,\tilde{Y}_1,\tilde{Y}_2)$ are shown in Figures 1 and 2. Let η_{Det} denote the distribution of

$$(\tilde{X}_1,\tilde{X}_2, 0.5\tilde{X}_1, \min(\tilde{X}_2, 0.5)+0.25(\tilde{X}_2-0.95)_+),$$

η^* the distribution found by optimization and denoter by $\tilde{R}_i$ the reinsured amount $\tilde{X}_i-\tilde{Y}_i$. With the dependence structure indicated by η_{Det}, $\tilde{R}_1$ and $\tilde{R}_2$ are independent and $\text{Var}_{\eta_{\text{Det}}}(\tilde{R}_1+\tilde{R}_2)=1.05314$. The variance after optimization is $\text{Var}_{\eta^*}(\tilde{R}_1+\tilde{R}_2)=0.82875$, which represents an improvement of 21.31%. As can be seen from Figures 1 and 2, this is achieved in two ways: first, the joint distributions of $(\tilde{X}_1,\tilde{Y}_1)$ and $(\tilde{X}_2,\tilde{Y}_2)$ are changed in such a way that, under η^*, $\tilde{Y}_1$ has positive probability of being close to $\tilde{X}_1$ and $\tilde{Y}_2$ has positive probability of being close to $\tilde{X}_2$ thus allowing the reinsured amounts $\tilde{R}_1$ and $\tilde{R}_2$ to take smaller values than under η_{Det}.

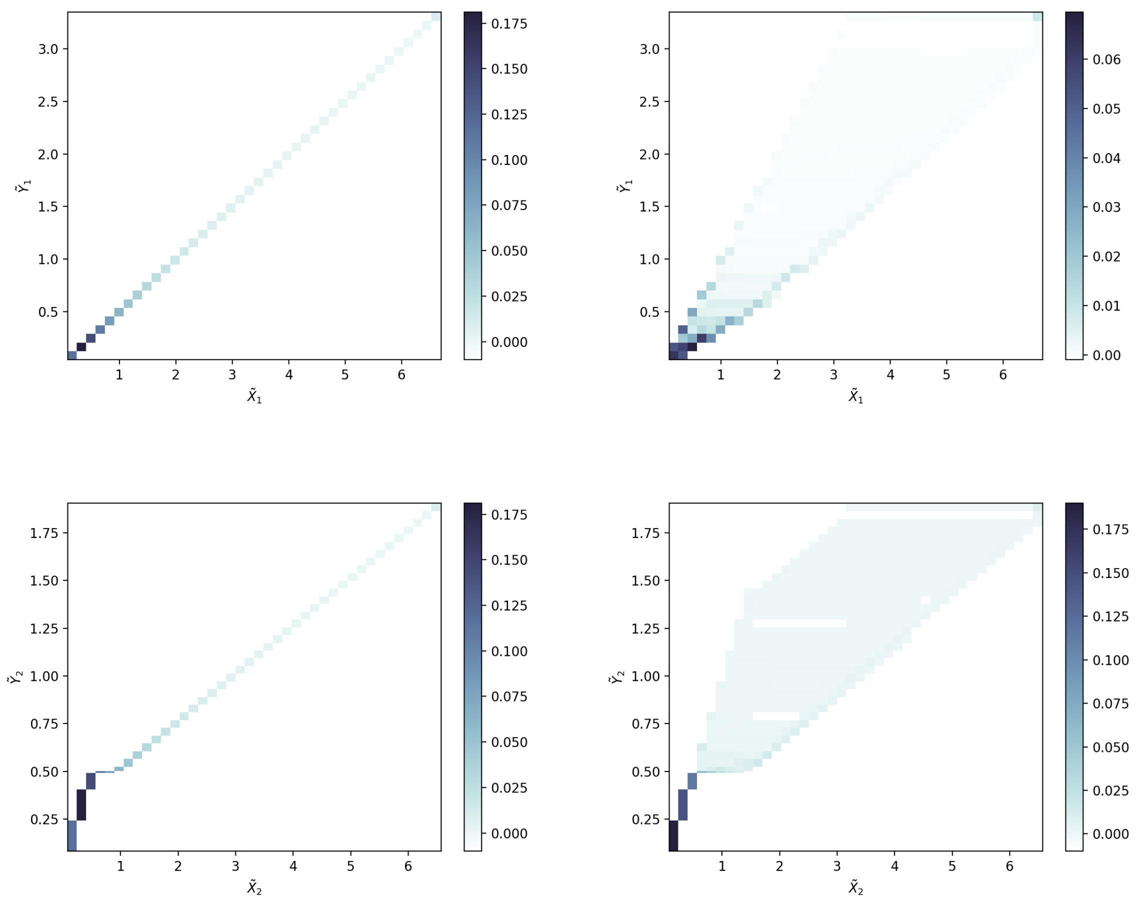


FIGURE 1. Probability mass functions for $\tilde{X}_1$ and $\tilde{Y}_1$ (upper row) and $\tilde{X}_2$ and $\tilde{Y}_2$ (lower row) under η_{Det} (left column) and η^* (right column).

Secondly, the variance is also reduced by introducing a positive dependence relationship between $\tilde{Y}_1$ and $\tilde{Y}_2$. This may be slightly counter-intuitive at first if we think of the variance as being reduced by making $\tilde{R}_1$ and $\tilde{R}_2$ counter-monotonic. While Figure 2 seems to indicate that $\tilde{R}_1$ and $\tilde{R}_2$ have a negative dependence structure, we cannot fully expect it to be counter-monotonic. Under the assumption of counter-monotonicity, small values for $\tilde{R}_1$ would be coupled with larger values of $\tilde{R}_2$, which would imply that values of $\tilde{Y}_1$ close to $\tilde{X}_1$ would be paired with values of $\tilde{Y}_2$ far away from $\tilde{X}_2$. Since $\tilde{Y}_1$ and $\tilde{Y}_2$ are bounded by $\tilde{X}_1$ and $\tilde{X}_2$, this would imply that small values for $\tilde{Y}_1$ would be coupled with larger values of $\tilde{Y}_2$. However, this argument lacks to take into account the fact that the reduction in variance can also be achieved by introducing a different dependence relationship between $\tilde{X}_1$ and $\tilde{Y}_2$, and between $\tilde{X}_2$ and $\tilde{Y}_1$.

Curiously enough, observe that the optimal joint distributions of $(\tilde{X}_1, \tilde{Y}_2)$ and $(\tilde{X}_2, \tilde{Y}_1)$ have a tendency to be concentrated in the upper left corner (as well as in the area where $\tilde{Y}_2 \leq 0.5$ for $(\tilde{X}_1, \tilde{Y}_2)$). This seems to indicate that $\tilde{Y}_2$ "uses" the extra degree of freedom in the distribution $(\tilde{X}_1, \tilde{Y}_2)$ to compensate for the behavior when $\tilde{X}_2 \leq 0.5$ (in which we necessarily must have $\tilde{Y}_2 = \tilde{X}_2$, regardless of the joint distribution) and for the constraint $\tilde{Y}_2 \leq \tilde{X}_2$.

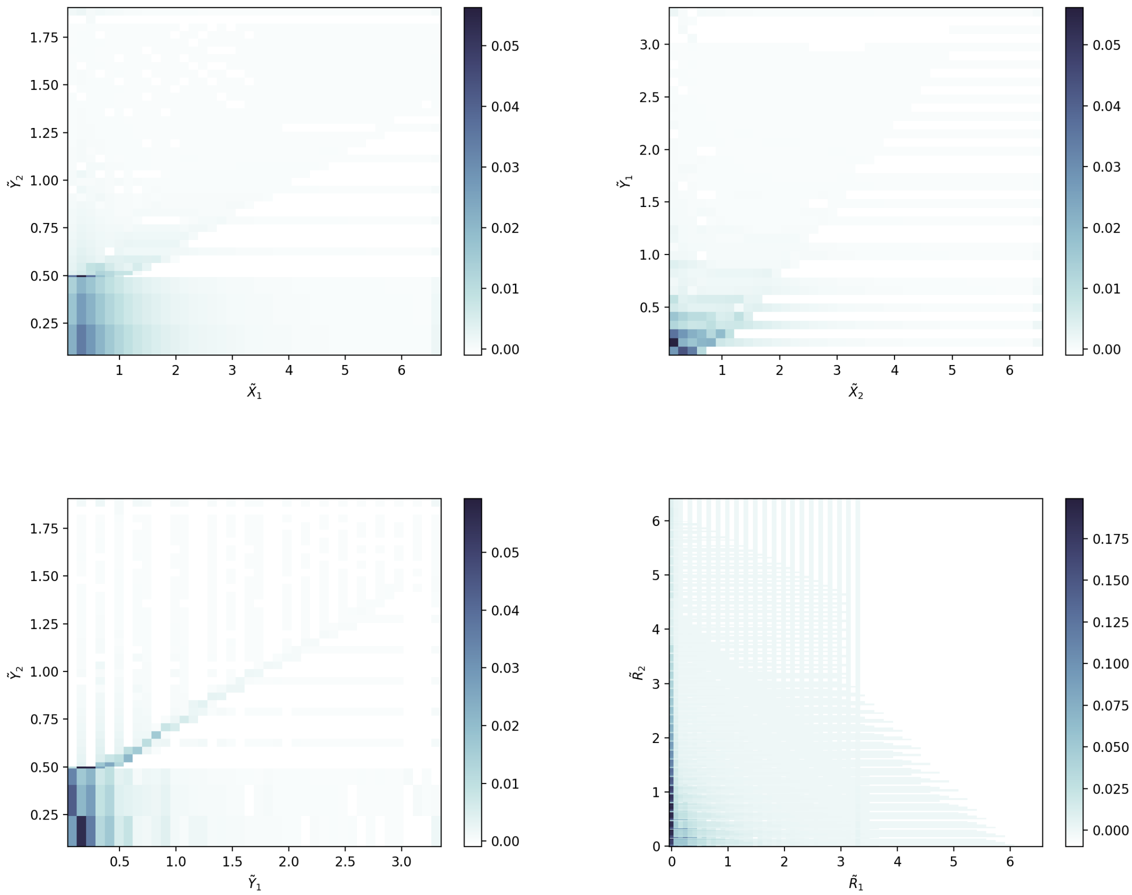


FIGURE 2. Probability mass functions for the joint distributions under η^* for $\tilde{X}_1$ and $\tilde{Y}_2$ (top left), $\tilde{X}_2$ and $\tilde{Y}_1$ (top right), $\tilde{Y}_1$ and $\tilde{Y}_2$ (bottom left) and $\tilde{R}_1$ and $\tilde{R}_2$ (bottom right).

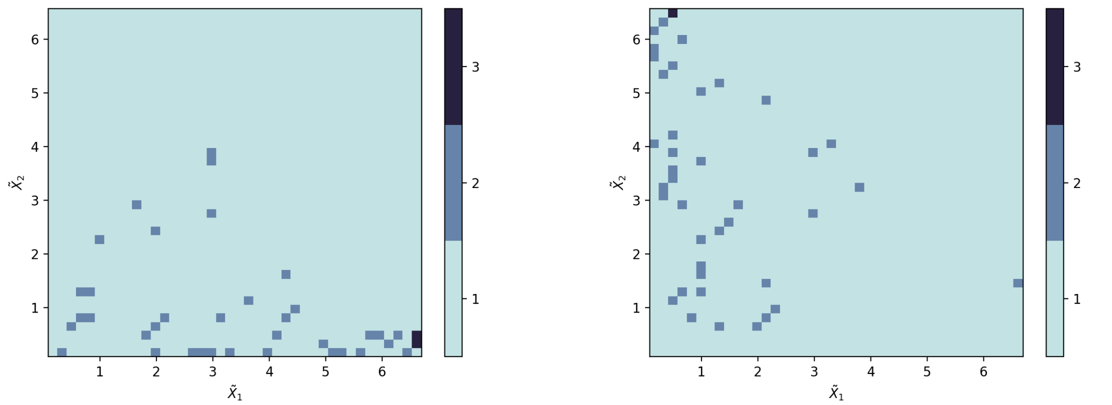


FIGURE 3. Cardinality of the supports of the conditional distributions of $\tilde{Y}_1$ (left) and $\tilde{Y}_2$ (right) under η^* given $\tilde{X}_1$ and $\tilde{X}_2$.

Similarly, $\tilde{Y}_1$ "uses" the extra degree of freedom in the distribution $(\tilde{X}_2, \tilde{Y}_1)$ to compensate for the constraint $\tilde{Y}_1 \leq \tilde{X}_1$. Finally, from Figure 1 we can observe that, under η^*, there is not a deterministic association between $\tilde{X}_1$ and $\tilde{Y}_1$ nor between $\tilde{X}_2$ and $\tilde{Y}_2$. Moreover, in contrast to Examples 5.5.1 and 5.5.3, the randomization of $\tilde{Y}_1$ for given $\tilde{X}_1$ is not only due to the realization of $\tilde{X}_2$ (and vice versa), as the conditional distributions of $\tilde{Y}_1$ and $\tilde{Y}_2$ given $(\tilde{X}_1, \tilde{X}_2)$ are not concentrated in one point, cf. Figure 3. That is, for minimizing the variance, in this example one uses an extra degree of randomness external to $(\tilde{X}_1, \tilde{X}_2)$ (like for instance a lottery). Observe, that this external randomness helps, in a way, to reach rather intuitive joint distributions of the variables involved.

5.6. Conclusion and Outlook

In this paper we provided a link between the two fields of optimal transport and optimal reinsurance, which allows for a reinterpretation of some classical optimal reinsurance results, a characterization of conditions for the optimality of deterministic treaties as well as the derivation of new results, extending some previous approaches in the literature. In a number of concrete examples we illustrated the benefits of this additional perspective on optimal reinsurance problems. We also established an example with two insurers and a reinsurer where external randomness in the contract specification effectively increases the efficiency.

While we worked out several concrete cases in detail, there are a number of directions that could be interesting for future research. In particular, we like to mention the following here. In this paper, we did not explicitly deal with the particular case of convex risk measures, an important class of risk measures on which a large part of the literature is focused (cf. [**26, 48**]). When the risk measure is given, for example, by

$$\mathcal{P}(\eta) = \int u\left(w - x - \int y\, \pi_{2\#}\eta)(dy)\right) T_{\#}\eta(dx)$$

for $u : \mathbb{R} \to \mathbb{R}$ a convex and non-decreasing function, then $\mathcal{P}$ is convex and its optimization can be addressed by means of Propositions 5.3.4, 5.3.7 or 5.4.1, depending on the nature of the constraints. Also, for general convex risk measures, it seems that the dual representation is tightly connected with the existence of the directional derivatives at optimal reinsurance contracts, and it will be interesting to connect the present approach with concepts from duality theory.

5.A. Continuation of Example 5.5.1

Recall that the function to minimize is $h(x, \cdot)$, where h is given by

$$h(x, y) = \sum_{i=1}^{n} \beta_i y_i + \lambda^* (s_X - s_Y)^2 - 2\lambda^* \sigma (s_X - s_Y).$$

and $s_X = \sum_{i=1}^n x_i$ and $s_Y = \sum_{i=1}^n y_i$. The partial derivatives of h are

$$\frac{\partial h(x, y)}{\partial y_k} = \beta_k - 2\lambda^* \left(\sum_{i=1}^{n}(x_i - y_i)\right) + 2\lambda^* \sigma, \quad k = 1, \dots, n,$$

and the condition $\beta_i \neq \beta_j$ for $i \neq j$ implies that no y in $\mathbb{R}^n_+$ solves $\partial h(x,y)/\partial y_k = 0$ for every $k = 1, \ldots, n$. Therefore we take another approach: we rewrite h in the form

$$\begin{aligned} h(x,y) &= \sum_{i=1}^n (\beta_i - \beta_n) y_i + \beta_n s_Y + \lambda^* \left(s_Y^2 + 2(\sigma - s_X) s_Y\right) + \lambda^* s_X^2 - 2\lambda^* \sigma s_X \\ &= -\sum_{i=1}^n (\beta_n - \beta_i) y_i + \lambda^* \left(s_Y - s_X + \sigma + \frac{\beta_n}{2\lambda^*}\right)^2 + \lambda^* s_X^2 - 2\lambda^* \sigma s_X \\ &\quad - \lambda^* \left(\sigma + \frac{\beta_n}{2\lambda^*} - s_X\right)^2 . \end{aligned}$$

In the last line, only the first two terms depend on y, so it suffices to find the y's minimizing the function

$$a(x,y) = -\sum_{i=1}^n (\beta_n - \beta_i) y_i + \lambda^* \left(s_Y - s_X + \sigma + \frac{\beta_n}{2\lambda^*}\right)^2 .$$

We make the following claim:

If y^* minimizes $a(x,\cdot)$ in $[0,x]$ and $y_i^* > 0$ for some $i \geq 2$, then $y_j^* = x_j$ for $j = 1, \ldots, i-1$.

Indeed, arguing by contradiction, assume there exists $j < i$ such that $y_j^* < x_j$. Define $\tilde{y}$ as $\tilde{y}_k = y_k^*$ if $k \neq i, j$, $\tilde{y}_j = y_j^* + \min(x_j - y_j^*, y_i^*)$ and $\tilde{y}_i = y_i^* - \min(x_j - y_j^*, y_i^*)$. Then

- either $\tilde{y}_j = x_j$ or $\tilde{y}_i = 0$, and
- $a(x,y^*) - a(x,\tilde{y}) = (\beta_i - \beta_j)\min(x_j - y_j^*, y_i^*) > 0$.

The last statement clearly contradicts the minimum property of y^*, so we conclude that the above claim is true. Coincidentally, this also implies that if $y_i^* = 0$ for some i, then $y_j^* = 0$ for all $j \geq i$.

From this it follows that for each fixed x, there are at most n candidate solutions of the form $(x_1, \ldots, x_{j-1}, y_j, 0, \ldots, 0)$, where y_j is the only point minimizing the mapping

$$y \mapsto -\sum_{i=1}^{j-1} (\beta_n - \beta_i) x_i - (\beta_n - \beta_j) y + \lambda^* \left(y - \sum_{i=j}^n x_i + \sigma + \frac{\beta_n}{2\lambda^*}\right)^2 \tag{142}$$

in $[0, x_j]$. For each $j = 1, \ldots, n$ we denote the associated candidate solution by $y^{(j)}$. Observe that

$$y_j^{(j)} = \begin{cases} 0, & \text{if } \sum_{i=j}^n x_i < \sigma + \frac{\beta_j}{2\lambda^*} \\ \sum_{i=j}^n x_i - \sigma - \frac{\beta_j}{2\lambda^*}, & \text{if } \sum_{i=j}^n x_i \geq \sigma + \frac{\beta_j}{2\lambda^*}, \text{ but } \sum_{i=j+1}^n x_i < \sigma + \frac{\beta_j}{2\lambda^*} \\ x_j, & \text{if } \sum_{i=j+1}^n x_i \geq \sigma + \frac{\beta_j}{2\lambda^*} \end{cases} . \tag{143}$$

Let j^* be the index defined by

$$j^* = \max\left\{ j \in \{1, \ldots, n\} \mid \sum_{i=j}^n x_i \geq \sigma + \frac{\beta_j}{2\lambda^*} \right\}$$

or $j^* = 1$ if the set on the right is empty. We claim that $y^{(j^*)}$ minimizes $a(x,\cdot)$. Indeed, notice that for $k < j^* - 1$, we have $y_r^{(k)} = y_r^{(k+1)} = x_r$ for $r = 1, \ldots, k$, $y_{k+1}^{(k)} = x_{k+1} - y_{k+1}^{(k+1)} = 0$

and $y_r^{(k)} = y_r^{(k+1)} = 0$ for $r > k+1$. Hence

$$\begin{aligned} a(x, y^{(k)}) - a(x, y^{(k+1)}) &= (\beta_n - \beta_{k+1})x_{k+1} + \lambda^* \left(- \sum_{i=k+1}^n x_i + \sigma + \frac{\beta_n}{2\lambda^*} \right)^2 \\ &\quad - \lambda^* \left(- \sum_{i=k+2}^n x_i + \sigma + \frac{\beta_n}{2\lambda^*} \right)^2 \\ &= \lambda^* x_{k+1} \left(\sum_{i=k+1}^n x_i - \sigma - \frac{\beta_{k+1}}{2\lambda^*} \right) \geq 0. \end{aligned}$$

The only difference between $y^{(j^*-1)}$ and $y^{(j^*)}$ is in the j^*-th entry. However, since $y_{j^*}^{(j^*)}$ minimizes the map in (142), we have $a(x, y^{(j^*)}) \leq a(x, y^{(j^*-1)})$. For $k > j^*$,

$$a(x, y^{(k+1)}) - a(x, y^{(k)}) = -\lambda^* x_{k+1} \left(\sum_{i=k+1}^n x_i - \sigma - \frac{\beta_{k+1}}{2\lambda^*} \right) \geq 0,$$

so we only need to show $a(x, y^{(j^*)}) \leq a(x, y^{(j^*+1)})$. However, if $\tilde{y}$ is such that $\tilde{y}_i = x_i$ for $i \leq j^*$ and zero otherwise, we can easily see that

$$a(x, y^{(j^*+1)}) - a(x, \tilde{y}) = -\lambda^* x_{j^*+1} \left(\sum_{i=j^*+1}^n x_i - \sigma - \frac{\beta_{j^*+1}}{2\lambda^*} \right) \geq 0,$$

and the minimal property of $y_{j^*}^{(j^*)}$ implies $a(x, y^{(j^*)}) \leq a(x, \tilde{y})$. Hence, $a(x, y^{(j^*)}) \leq a(x, y^{(j)})$ for every $j = 1, \ldots, n$.

Noticing that (143) can be written in the form

$$y_{j^*}^{(j^*)} = \min \left(\left(\sum_{i=j^*}^n x_i - \frac{\beta_{j^*}}{2\lambda^*} - \sigma \right)_+ , x_{j^*} \right),$$

the condition $\beta_1 < \ldots < \beta_n$ implies that the optimal reinsurance contract is as given in (132).

5.B. Continuation of Example 5.5.3

Recall the definitions of the sets $\mathcal{C}, D_1, D_2$ and D_3 given in (133)-(134). With this $\mathcal{C}$, the functionals $\mathcal{P}$ and $\mathcal{G}$ satisfy the conditions of Proposition 5.3.4 with $U \subset \mathscr{S}$ the space of measures with finite marginal expectations, $p(x, y) = \sum_{i=1}^n \beta_i y_i$ and $g = 0$. Let $(x, y) \in \mathrm{Supp}(\eta^*)$ and recall the measures

$$\mu_{x,y,t,\varepsilon}(A) = \eta^*(A) - \eta^*(A \cap B_\varepsilon(x, y)) + \eta^*((A - (0, t)) \cap B_\varepsilon(x, y)),$$

appearing in Proposition 5.3.4. We wish to identify the sets over which minimization occurs in (118). We proceed as in the proof of that proposition, i.e., we see that for a suitable chosen t, we can find a $\delta > 0$ such that $\mu_{x,y,t,\varepsilon} \in \mathcal{C}$ for $\varepsilon < \delta$. Denote the set from the conclusion of Proposition 5.3.4 by $I(x, y)$.

Assume first that $(x, y) \in D_1 \setminus \partial\mathcal{A}_R$ and let T denote the set of t's such that $(x, y) \in D_1 - (0, t)$, i.e., $T = \{-y < t < x - y \mid (x, y + t) \in D_1\}$. For $t \in T$, let

$$\delta = \min \{ d((x, y), \partial\mathcal{A}_R), d((x, y), \partial\mathcal{A}_R - (0, t)), d((x, y), \partial D_1), d((x, y), \partial D_1 - (0, t)) \}.$$

With this definition, $\delta > 0$ and for every $0 \leq \varepsilon < \delta$, we have that $\mu_{x,y,t,\varepsilon} \in \mathscr{M}$. We claim that, further, $\widehat{\mathrm{VaR}}_\alpha(T_{S\#}\mu_{x,y,t,\varepsilon}) = v^*$. Indeed, notice that $B_\varepsilon(x, y) \subset D_1 \cap (D_1 - (0, t))$ and therefore

$$T_{S\#}\mu_{x,y,t,\varepsilon}((v^*,\infty)) = \mu_{x,y,t,\varepsilon}(D_2) = \eta^*(D_2) \leq \alpha.$$

Now, let $\varepsilon' = \min\{d((x,y),\partial D_1), d((x,y),\partial D_1 - (0,t))\} - \varepsilon$. We have $\varepsilon' > 0$ by choice of ε. For $v^* - \varepsilon' < u < v^*$, let

$$D_u = \left\{(x,y) \in \mathcal{A}_R \mid \sum_{i=1}^n (x_i - y_i) > u\right\}.$$

Observe that, for these u's, we still have $B_\varepsilon(x,y) \subset D_u^c \cap (D_u^c - (0,t))$ and therefore

$$T_{S\#}\mu_{x,y,t,\varepsilon}((u,\infty)) = \mu_{x,y,t,\varepsilon}(D_u) = \eta^*(D_u) > \alpha.$$

Hence $\widehat{\mathrm{VaR}}_\alpha(T_{S\#}\mu_{x,y,t,\varepsilon}) = v^*$. In a similar manner, we see that if $t \notin T$ (but still $-y < t < x - y$), we have $\widehat{\mathrm{VaR}}_\alpha(T_{S\#}\mu_{x,y,t,\varepsilon}) > v^*$ for ε small enough, so that

$$I(x,y) = \{z \in [0,x] \mid (x,z) \in \overline{D_1}\}. \tag{144}$$

Now we consider $(x,y) \in D_1 \cap \partial\mathcal{A}_R$ and proceed similarly by defining

$$\begin{aligned}
\delta &= \min\{d((x,y),\partial\mathcal{A}_R - (0,t)), d((x,y),\partial D_1 - (0,t))\},\\
\varepsilon' &= d((x,y),\partial D_1 - (0,t)) - \varepsilon,
\end{aligned}$$

for t's such that $-x_0 < t_i \leq 0$ if $y_{0,i} = x_{0,i}$, $0 \leq t_i < x_0$ if $y_{0,i} = 0$, $-x_{0,i} < t_i < x_{0,i} - y_{0,i}$ if $0 < y_{0,i} < x_{0,i}$ and $(x,y) \in D_1 - (0,t)$. We still obtain $I(x,y)$ as in (144). Thus, for $(x,y) \in D_1$,

$$I(x,y) = \left\{z \in [0,x] \mid \sum_{i=1}^n (x_i - z_i) \leq v^*\right\}.$$

Likewise, we can show that for $(x,y) \in D_2$ we can define $T = \{-y < t < x - y \mid (x, y+t) \in D_2\}$ and obtain

$$I(x,y) \supset \left\{z \in [0,x] \mid \sum_{i=1}^n (x_i - z_i) \geq v^*\right\}.$$ [8]

In particular, $0 \in I(x,y)$ for $(x,y) \in D_2$.

Now, let (x,y) be an arbitrary point in the support of η^*. Since $\mathcal{A}_R = D_1 \cup D_2 \cup D_3$, (x,y) is in one (an only one) of these three sets. If $(x,y) \in D_1$, the previous computations, together with the conclusion of Proposition 5.3.4, show that

$$y \in \operatorname{argmin}\left\{\sum_{i=1}^n \beta_i z_i \mid z \in [0,x], \sum_{i=1}^n x_i - v^* \leq \sum_{i=1}^n z_i\right\}.$$

This is a linear optimization exercise with linear constraints and one can see that the condition $(x,y) \in D_1$ implies $\sum_{i=1}^n x_i \leq v^*$ and $y = 0$. Similarly, if $(x,y) \in D_2$, $\sum_{i=1}^n x_i > v^*$ and $y = 0$. Therefore,

$$\mathrm{Supp}(\eta^*) \subset D_3 \cup \{(x,0) \mid x \in \mathbb{R}_+^n\}. \tag{145}$$

As stated before, this implies

$$\eta^*(D_3) = 1 - \alpha - \mu\left(\left\{x \in \mathbb{R}_+^n \mid \sum_{i=1}^n x_i < v^*\right\}\right) := m$$

[8] In this case we obtain only an inclusion as opposed to an equality. This comes from the fact that for $t \notin T$ we can only ensure $\widehat{\mathrm{VaR}}_\alpha(T_{S\#}\mu_{x,y,t,\varepsilon}) \leq v^*$. As a matter of fact, for ε small enough, we will actually have an equality and therefore $I(x,y) = [0,x]$, though we do not need this information to find the solution.

and we need to optimally assign this mass to D_3. Now, observe that (145) implies

$$\mathcal{P}(\eta^*) = \int_{D_3} \sum_{i=1}^{n} \beta_i y_i \, \eta^*(dx, dy)$$

and, just as in the proof of Proposition 5.3.7, it is intuitive that if $(x, y) \in \mathrm{Supp}(\eta^*) \cap D_3$, then y should be in the minima of $p(x, \cdot)$. Before making this statement more rigorous, let us note the following:

For fixed $x \in \mathbb{R}^n$ such that $\sum_{i=1}^{n} x_j > v^*$, the minimum of $y \mapsto \sum_{i=1}^{n} \beta_i y_i$ on

$$[0, x] \cap \left\{ y \in \mathbb{R}^n_+ \mid \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i - v^* \right\}$$

is unique and is given by

$$y^*(x) = \left(\min\left(Q_1(x), x_1\right), \ldots, \min\left(Q_{n-1}(x), x_{n-1}\right), Q_n(x)\right), \tag{146}$$

where

$$Q_i(x) = \left(\sum_{j=i}^{n} x_j - v^* \right)_+, \quad i = 1, \ldots, n.$$

Indeed, let $\tilde{y}$ denote any minimum and for $i = 1, \ldots, n$, let C_i be given by

$$C_i = \left\{ z \in \mathbb{R}^n_+ \mid \sum_{j=i+1}^{n} z_j \le v^*, \sum_{j=i}^{n} z_j > v^* \right\}.$$

These sets are disjoint and x belongs to one (and only one) of them. Let i_0 denote the index such that $x \in C_{i_0}$. We claim that $\tilde{y}_j = x_j$ for $j < i_0$. For otherwise, we can let j_0 be the first index for which this does not happen or $j_0 = 1$ in case $\tilde{y}_j < x_j$ for all j's. Therefore, $\tilde{y}_j = x_j$ for $j < j_0$, while $\tilde{y}_{j_0} < x_{j_0}$. If $\tilde{y}_k = 0$ for all $k > j_0$, then the condition $\sum_{i=1}^{n} \tilde{y}_i = \sum_{i=1}^{n} x_i - v^*$ implies

$$\tilde{y}_{j_0} = \sum_{i=j_0}^{n} x_i - v^* > x_{j_0}$$

since $j_0 < i_0$. Thus, there exists $k_0 > j_0$ such that $y_{k_0} > 0$. However, in this case, the vector $\hat{y} \in \mathbb{R}^n$ given by $\hat{y}_j = \tilde{y}_j$ if $j \neq j_0, k_0$, $\hat{y}_{j_0} = \min(x_{j_0}, \tilde{y}_{j_0} + \tilde{y}_{k_0})$ and $\hat{y}_{k_0} = \max(0, \tilde{y}_{k_0} + \tilde{y}_{j_0} - x_{j_0})$ satisfies all the constraints and further $p(x, \hat{y}) < p(x, \tilde{y})$ as $\beta_{j_0} < \beta_{k_0}$. Since this is a contradiction to the optimality of $\tilde{y}$, we therefore conclude $\tilde{y}_j = x_j$ for all $j < i_0$ and

$$\sum_{j=i_0}^{n} \tilde{y}_j = \sum_{j=i_0}^{n} x_i - v^*.$$

This now implies $\tilde{y}_{i_0} = \sum_{j=i_0}^{n} x_i - v^* \le x_{i_0}$, since

$$\beta_{i_0} \tilde{y}_{i_0} = \beta_{i_0} \left(\sum_{j=i_0}^{n} x_i - v^* \right) = \beta_{i_0} \sum_{j=i_0}^{n} y_j \le \sum_{j=i_0}^{n} \beta_j y_j$$

for any other $y \in \mathbb{R}^n_+$ with $y_j = x_j$ for $j < i_0$ (which is a condition for optimality). Hence

$$\tilde{y} = \left(x_1, x_2, \ldots, x_{i_0 - 1}, \sum_{j=i_0}^{n} x_i - v^*, 0, \ldots, 0 \right),$$

which is equivalent to (146) on C_{i_0} and shows unicity of $\tilde{y}$.

From (146) it follows that the mapping $x \mapsto y^*(x)$ is continuous. The rest of the proof now follows easily: if for some $(x, y) \in \mathrm{Supp}(\eta^*) \cap D_3$, $y \neq 0$, we had $y \neq y^*(x)$, then, by continuity of $q(x) := p(x, y^*(x))$, we could find an open ball $B \subset \mathbb{R}^n_+ \times \mathbb{R}^n_+$ around (x, y) not intersecting $\mathbb{R}^n_+ \times \{0\}$ and such that $q(x') < p(x', y')$ for every $(x', y') \in B$. We could then modify η^* by moving all the mass in B to the graph of q, obtaining a measure with the same VaR and a strictly smaller value for $\mathcal{P}$. For example, we can use the following measure

$$\eta(A) = \eta^*(A \setminus B) + \mu(Q^{-1}(A) \cap B').$$

Here $Q = (\mathrm{Id}, q)$ and $B' \subset \pi_1(B)$ is an open ball around x such that $\eta^*(B) = \mu(B')$, which exists by absolute continuity of μ. It therefore follows that

$$\mathrm{Supp}(\eta^*) \subset \{(x, 0) \mid x \in \mathbb{R}^n_+\} \cup \{(x, y^*(x)) \mid x \in \mathbb{R}^n_+\}.$$

Letting $E = \{z \in \mathbb{R}^n_+ \mid q(z) \leq d\}$, where d is chosen so that $\mu(E) = 1 - \alpha$, a similar argument shows that, in case $d > 0$, there cannot be a point (x, y) in the support of η^* such that $q(x) > d$.

Bibliography

[1] H. Albrecher, J. Beirlant, and J. L. Teugels. *Reinsurance: actuarial and statistical aspects.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2017.

[2] H. Albrecher and A. Cani. Risk theory with affine dividend payment strategies. *Number Theory–Diophantine Problems, Uniform Distribution and Applications: Festschrift in Honour of Robert F. Tichy's 60th Birthday*, pages 25–60, 2017.

[3] H. Albrecher and A. Cani. On randomized reinsurance contracts. *Insurance Math. Econom.*, 84:67–78, 2019.

[4] H. Albrecher and B. Garcia Flores. Optimal dividend bands revisited: a gradient-based method and evolutionary algorithms. *Scandinavian Actuarial Journal*, pages 1–23, 2022.

[5] H. Albrecher, J. Ivanovs, and X. Zhou. Exit identities for Lévy processes observed at Poisson arrival times. *Bernoulli*, 22(3):1364–1382, 2016.

[6] H. Albrecher, S. A. Ladoucette, and J. L. Teugels. Asymptotics of the sample coefficient of variation and the sample dispersion. *J. Statist. Plann. Inference*, 140(2):358–368, 2010.

[7] H. Albrecher, C. Y. Robert, and J. L. Teugels. Joint asymptotic distributions of smallest and largest insurance claims. *Risks*, 2(3):289–314, 2014.

[8] H. Albrecher and J. L. Teugels. Asymptotic analysis of a measure of variation. *Theory of Probability and Mathematical Statistics*, 74:1–10, 2006.

[9] H. Albrecher, J. L. Teugels, and K. Scheicher. A combinatorial identity for a problem in asymptotic statistics. *Appl. Anal. Discrete Math.*, 3(1):64–68, 2009.

[10] H. Albrecher and S. Thonhauser. Optimality results for dividend problems in insurance. *RACSAM-Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 103(2):295–320, 2009.

[11] L. Ambrosio, E. Brué, and D. Semola. *Lectures on optimal transport.* Springer, 2021.

[12] M. Arendarczyk, T. J. Kozubowski, and A. K. Panorska. The Greenwood statistic, stochastic dominance, clustering and heavy tails. *Scandinavian Journal of Statistics, to appear*, 2021.

[13] K. Arrow. Uncertainty and the welfare economics of medical care. *American Economic Review*, 53:941–973, 1963.

[14] A. V. Asimit, A. M. Badescu, and K. C. Cheung. Optimal reinsurance in the presence of counterparty default risk. *Insurance: Mathematics and Economics*, 53(3):690–697, 2013.

[15] A. V. Asimit, T. J. Boonen, Y. Chi, and W. F. Chong. Risk sharing with multiple indemnity environments. *European Journal of Operational Research*, 295(2):587–603, 2021.

[16] V. Asimit and T. J. Boonen. Insurance with multiple insurers: A game-theoretic approach. *European Journal of Operational Research*, 267(2):778–790, 2018.

[17] S. Asmussen and H. Albrecher. *Ruin probabilities*, volume 14. World scientific, 2010.

[18] B. Avanzi. Strategies for dividend distribution: A review. *North American Actuarial Journal*, 13(2):217–251, 2009.

[19] F. Avram, A. E. Kyprianou, and M. R. Pistorius. Exit problems for spectrally negative Lévy processes and applications to (Canadized) Russian options. *Ann. Appl. Probab.*, 14(1):215–238, 2004.

[20] F. Avram, Z. Palmowski, and M. R. Pistorius. On the optimal dividend problem for a spectrally negative Lévy process. *The Annals of Applied Probability*, 17(1):156–180, 2007.

[21] F. Avram, Z. Palmowski, and M. R. Pistorius. On gerber–shiu functions and optimal dividend distribution for a Lévy risk process in the presence of a penalty function. *The Annals of Applied Probability*, 25(4):1868–1935, 2015.

[22] P. Azcue and N. Muler. Optimal reinsurance and dividend distribution policies in the cramér-lundberg model. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 15(2):261–308, 2005.

[23] P. Azcue and N. Muler. Optimal dividend policies for compound poisson processes: The case of bounded dividend rates. *Insurance: Mathematics and Economics*, 51(1):26–42, 2012.

[24] P. Azcue and N. Muler. *Stochastic optimization in insurance: a dynamic programming approach.* Springer, 2014.

[25] A. Balbás, B. Balbás, R. Balbás, and A. Heras. Optimal reinsurance under risk and uncertainty. *Insurance: Mathematics and Economics*, 60:61–74, 2015.

[26] A. Balbás, B. Balbás, R. Balbás, and A. Heras. Risk transference constraints in optimal reinsurance. *Insurance: Mathematics and Economics*, 103:27–40, 2022.

[27] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes: theory and applications.* John Wiley & Sons, Chichester, 2004.

[28] J. Berdel. *Optimale Bandstrategien für Dividendenzahlungen im Cramér-Lundberg-Modell.* PhD thesis, Karlsruher Instituts für Technologie, 2014.

[29] D. Bertsekas and S. E. Shreve. *Stochastic optimal control: the discrete-time case*, volume 5. Athena Scientific, 1996.

[30] H.-G. Beyer. *The theory of evolution strategies.* Springer Science & Business Media, 2001.

[31] H.-G. Beyer and H.-P. Schwefel. Evolution strategies–a comprehensive introduction. *Natural computing*, 1(1):3–52, 2002.

[32] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation.* Number 27. Cambridge university press, 1989.

[33] T. J. Boonen, K. C. Cheung, and Y. Zhang. Bowley reinsurance with asymmetric information on the insurer's risk preferences. *Scandinavian Actuarial Journal*, 2021(7):623–644, 2021.

[34] T. J. Boonen and M. Ghossoub. Optimal reinsurance with multiple reinsurers: distortion risk measures, distortion premium principles, and heterogeneous beliefs. *Insurance: Mathematics and Economics*, 101:23–37, 2021.

[35] T. J. Boonen and M. Ghossoub. Bowley vs. Pareto optima in reinsurance contracting. *European Journal of Operational Research*, 307(1):382–391, 2023.

[36] T. J. Boonen and W. Jiang. A marginal indemnity function approach to optimal reinsurance under the Vajda condition. *European Journal of Operational Research*, 303(2):928–944, 2022.

[37] K. Borch. An attempt to determine the optimum amount of stop loss reinsurance. In *Transactions of the 16th International Congress of Actuaries*, page 597–610, 1960.

[38] K. Borch. The safety loading of reinsurance premiums. *Scandinavian Actuarial Journal*, 1960(3-4):163–184, 1960.

[39] M. Brown and J. E. Cohen. Squared coefficient of variation of Taylor's law for random absolute differences. *Probab. Engrg. Inform. Sci.*, 32(4):483–494, 2018.

[40] M. Brown, J. E. Cohen, and V. H. de la Peña. Taylor's law, via ratios, for some distributions with infinite mean. *J. Appl. Probab.*, 54(3):657–669, 2017.

[41] J. Cai, C. Lemieux, and F. Liu. Optimal reinsurance with regulatory initial capital and default risk. *Insurance: Mathematics and Economics*, 57:13–24, 2014.

[42] J. Cai, K. S. Tan, C. Weng, and Y. Zhang. Optimal reinsurance under var and cte risk measures. *Insurance: Mathematics and Economics*, 43(1):185–196, 2008.

[43] J. D. Castillo, J. Daoudi, and R. Lockhart. Methods to distinguish between polynomial and exponential tails. *Scandinavian Journal of Statistics*, 41(2):382–393, 2014.

[44] M. L. Centeno and O. Simões. Optimal reinsurance. *RACSAM-Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas*, 103:387–404, 2009.

[45] K. C. Cheung. Optimal reinsurance revisited–a geometric approach. *ASTIN Bulletin*, 40(1):221–239, 2010.

[46] K. C. Cheung, W. He, and H. Wang. Multi-constrained optimal reinsurance model from the duality perspectives. *Insurance: Mathematics and Economics*, 113:199–214, 2023.

[47] K. C. Cheung and A. Lo. Characterizations of optimal reinsurance treaties: a cost-benefit approach. *Scandinavian Actuarial Journal*, 2017(1):1–28, 2017.

[48] K. C. Cheung, K. Sung, S. C. P. Yam, and S. P. Yung. Optimal reinsurance under general law-invariant risk measures. *Scandinavian Actuarial Journal*, 2014(1):72–91, 2014.

[49] G. Chistyakov, F. Götze, et al. Limit distributions of studentized means. *The Annals of Probability*, 32(1A):28–77, 2004.

[50] J. E. Cohen, R. A. Davis, and G. Samorodnitsky. Heavy-tailed distributions, correlations, kurtosis and Taylor's Law of fluctuation scaling. *Proc. A.*, 476(2244):20200610, 27, 2020.

[51] H. Cramer. *Collective risk theory.* Skandia ins. Company, 1955.

[52] H. Cramér. *On the mathematical theory of risk.* Centraltryckeriet, 1959.

[53] B. De Finetti. *Il problema dei pieni.* Istituto italiano degli attuari, 1940.

[54] 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. New York, 1957.

[55] V. De la Pena, P. Doukhan, and Y. Salhi. A dynamic taylor's law. *arXiv preprint arXiv:2010.10598*, 2020.

[56] O. Deprez and H. U. Gerber. On convex principles of premium calculation. *Insurance: Mathematics and Economics*, 4(3):179–189, 1985.

[57] 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.

[58] F. Dufresne, H. U. Gerber, and E. S. Shiu. Risk theory with the gamma process. *ASTIN Bulletin: The Journal of the IAA*, 21(2):177–192, 1991.

[59] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

[60] L. Gajek and D. Zagrodny. Reinsurance arrangements maximizing insurer's survival probability. *Journal of Risk and Insurance*, 71(3):421–435, 2004.

[61] A. Galichon. *Optimal transport methods in economics*. Princeton University Press, 2018.

[62] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113 – 161, 1996.

[63] H. U. Gerber. *Entscheidungskriterien für den zusammengesetzten Poisson-Prozess*. PhD thesis, ETH Zurich, 1969.

[64] H. U. Gerber, X. S. Lin, and H. Yang. A note on the dividends-penalty identity and the optimal dividend barrier. *Astin Bull.*, 36(2):489–503, 2006.

[65] H. U. Gerber and G. Pafumi. Utility functions: from risk theory to finance. *North American Actuarial Journal*, 2(3):74–91, 1998.

[66] H. U. Gerber and E. S. Shiu. On the time value of ruin. *North American Actuarial Journal*, 2(1):48–72, 1998.

[67] P. Grandits. An optimal consumption problem in finite time with a constraint on the ruin probability. *Finance Stoch.*, 19(4):791–847, 2015.

[68] M. Greenwood. The statistical study of infectious diseases. *Journal of the Royal Statistical Society*, 109(2):85–110, 1946.

[69] M. Guerra and M. L. Centeno. Are quantile risk measures suitable for risk-transfer decisions? *Insurance: Mathematics and Economics*, 50(3):446–461, 2012.

[70] M. Guerra and A. de Moura. Reinsurance of multiple risks with generic dependence structures. *Insurance: Mathematics and Economics*, 101:547–571, 2021.

[71] P. Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint*. CRC Press, 2017.

[72] C. Hernandez, M. Junca, and H. Moreno-Franco. A time of ruin constrained optimal dividend problem for spectrally one-sided Lévy processes. *Insurance: Mathematics and Economics*, 79:57–68, 2018.

[73] C. Hipp. Optimal dividend payment under a ruin constraint: discrete time and state space. *Blätter der DGVFM*, 26(2):255–264, 2003.

[74] C. Hipp. Company value with ruin constraint in Lundberg models. *Risks*, 6(3):73, 2018.

[75] C. Hipp. Dividend payment with ruin constraint. *Risk and Stochastics, Ragnar Norberg. Edited by P. Barrieu. Hackensack: World Scientific*, pages 1–22, 2019.

[76] C. Hipp. Optimal dividend payment in de Finetti models: Survey and new results and strategies. *Risks*, 8(3):96, 2020.

[77] F. Hubalek and E. Kyprianou. Old and new examples of scale functions for spectrally negative Lévy processes. In *Seminar on Stochastic Aanalysis, Random Fields and Applications VI*, pages 119–145. Springer, 2011.

[78] J. Ivanovs and Z. Palmowski. Occupation densities in solving exit problems for Markov additive processes and their reflections. *Stochastic Process. Appl.*, 122(9):3342–3360, 2012.

[79] M. Kaluszka. Mean-variance optimal reinsurance arrangements. *Scandinavian Actuarial Journal*, 2004(1):28–41, 2004.

[80] L. V. Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

[81] O. Kramer. A review of constraint-handling techniques for evolution strategies. *Applied Computational Intelligence and Soft Computing*, 2010, 2010.

[82] A. Kuznetsov, A. E. Kyprianou, and V. Rivero. The theory of scale functions for spectrally negative Lévy processes. *Lévy matters II*, pages 97–186, 2012.

[83] A. E. Kyprianou. *Fluctuations of Lévy processes with applications*. Universitext. Springer, Heidelberg, second edition, 2014.

[84] C.-Y. Lee and X. Yao. Evolutionary programming using mutations based on the Lévy probability distribution. *IEEE Transactions on Evolutionary Computation*, 8(1):1–13, 2004.

[85] R. Lepage, M. Woodroofe, and J. Zinn. Convergence to a stable distribution via order statistics. *The Annals of Probability*, 9(4):624–632, 1981.

[86] Z. Liang and V. R. Young. Dividends and reinsurance under a penalty for ruin. *Insurance: Mathematics and Economics*, 50(3):437–445, 2012.

[87] X. S. Lin, G. E. Willmot, and S. Drekic. The classical risk model with a constant dividend barrier: analysis of the gerber–shiu discounted penalty function. *Insurance: Mathematics and Economics*, 33(3):551–566, 2003.

[88] A. Lo. A Neyman-Pearson perspective on optimal reinsurance with constraints. *ASTIN Bulletin: The Journal of the IAA*, 47(2):467–499, 2017.

[89] R. L. Loeffen. On optimality of the barrier strategy in de Finetti's dividend problem for spectrally negative Lévy processes. *The Annals of Applied Probability*, pages 1669–1680, 2008.

[90] R. L. Loeffen and J.-F. Renaud. De Finetti's optimal dividends problem with an affine penalty function at ruin. *Insurance: Mathematics and Economics*, 46(1):98–108, 2010.

[91] F. Lundberg. Försäkringsteknisk riskutjämning.

[92] F. Lundberg. *1. Approximerad framställning af sannolikhetsfunktionen: 2. Återförsäkring af kollektivrisker. Akademisk afhandling*. Almqvist &Wiksells, 1903.

[93] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[94] Z. Michalewicz and M. Schoenauer. Evolutionary algorithms for constrained parameter optimization problems. *Evolutionary computation*, 4(1):1–32, 1996.

[95] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.

[96] M. Morales and W. Schoutens. A risk model driven by lévy processes. *Applied Stochastic Models in Business and Industry*, 19(2):147–167, 2003.

[97] M. I. Pesonen. Optimal reinsurances. *Scandinavian Actuarial Journal*, 1984(2):65–90, 1984.

[98] G. Peyré, M. Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[99] T. Rolski, H. Schmidli, V. Schmidt, and J. L. Teugels. *Stochastic processes for insurance and finance*. John Wiley & Sons, 2009.

[100] S. Román, A. M. Villegas, and J. G. Villegas. An evolutionary strategy for multiobjective reinsurance optimization. *Journal of the Operational Research Society*, 69(10):1661–1677, 2018.

[101] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, Singapore, 2nd edition, 1991.

[102] G. Rudolph. Convergence of evolutionary algorithms in general search spaces. In *Proceedings of IEEE international conference on evolutionary computation*, pages 50–54. IEEE, 1996.

[103] S. Salcedo-Sanz, L. Carro-Calvo, M. Claramunt, A. Castañer, and M. Mármol. Effectively tackling reinsurance problems by using evolutionary and swarm intelligence algorithms. *Risks*, 2(2):132–145, 2014.

[104] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

[105] H. Schmidli. Optimisation in non-life insurance. *Stochastic models*, 22(4):689–722, 2006.

[106] H. Schmidli. *Stochastic control in insurance*. Springer Science & Business Media, 2007.

[107] W. F. Sharpe. Mutual fund performance. *The Journal of Business*, 39(1):119–138, 1966.

[108] S. M. Srivastava. *A course on Borel sets*, volume 180. Springer Science & Business Media, 2008.

[109] J. A. Strini and S. Thonhauser. Time-inconsistent view on a dividend problem with penalty. *Scandinavian Actuarial Journal*, pages 1–23, 2023.

[110] K. S. Tan, C. Weng, and Y. Zhang. Var and cte criteria for optimal quota-share and stop-loss reinsurance. *North American Actuarial Journal*, 13(4):459–482, 2009.

[111] S. Thonhauser and H. Albrecher. Dividend maximization under consideration of the time value of ruin. *Insurance Math. Econom.*, 41(1):163–184, 2007.

[112] C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

[113] C. Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[114] L. Vincent, H. Albrecher, and Y. Krvavych. Structured reinsurance deals with reference to relative market performance. *Insurance: Mathematics and Economics*, 101:125–139, 2021.

[115] C.-P. Wang, D. Shyu, and H.-H. Huang. Optimal insurance design under a value-at-risk framework. *The Geneva Risk and Insurance Review*, 30:161–179, 2005.

[116] X. Yao, Y. Liu, and G. Lin. Evolutionary programming made faster. *IEEE Transactions on Evolutionary computation*, 3(2):82–102, 1999.
[117] M. B. Zhu, M. Ghossoub, and T. J. Boonen. Equilibria and efficiency in a reinsurance market. *Insurance: Mathematics and Economics*, 113:24–49, 2023.
[118] S. C. Zhuang, C. Weng, K. S. Tan, and H. Assa. Marginal indemnification function formulation for optimal reinsurance. *Insurance: Mathematics and Economics*, 67:65–76, 2016.