Optimal Stopping and Policyholder Behaviour in Life Insurance

Kamille Sofie Tågholt Gad

PhD Thesis

This thesis has been submitted to the PhD School of the Faculty of Science, University of Copenhagen.

Supervisor: Jesper Lund Pedersen, University of Copenhagen
Co-supervisor: Mogens Steffensen, University of Copenhagen

Submitted: February 13, 2015.

DEPARTMENT OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Author: Kamille Sofie Tågholt Gad
Studsgaardsgade 42, 2
2100 København Ø
Denmark
kamillesofie@gmail.com

Assessment Committee: Associate Professor Marcus C. Christiansen,
Heriot-Watt University,
Edinburgh, Scotland

Professor Goran Peskir,
The University of Manchester,
Manchester, United Kingdom

Associate Professor Jeffrey F. Collamore,
University of Copenhagen,
Denmark (Chairman)

ISBN: 978-87-7078-961-5

Preface

This thesis has been prepared in fulfilment of the requirements for the Ph.D. degree at the Department of Mathematical Sciences, Faculty of Science, University of Copenhagen, Denmark. The work has been carried out under the supervision of Jesper Lund Pedersen and Professor Mogens Steffensen, University of Copenhagen in the period from May 1st 2011 to February 13th 2015 (including 9.5 months maternity leave).

The main body of the thesis consists of an introduction to the material in the thesis, and five chapters on different but related topics. The five chapters are written as individual academic papers, and are thus self-contained and can be read independently. This final version contains the following changes compared to the version submitted for assessment: Chapter 2 includes trivial proof-reading changes from the journal; Chapter 5 includes added references and numerical examples, and suggestions from review led to the inclusion of the new Section 5.6; Chapter 6 includes proof-reading changes

I would like to thank my two advisors Jesper Lund Pedersen and Mogens Steffensen for support, ideas and discussions during the work. A special thanks to Mogens, Edlund A/S and the IT University for setting up the Actulus project and letting me participate. Also I would like to thank Jesper, Mogens, Jeppe Juhl and Jeppe Woetmann Nielsen for co-authoring different parts of the thesis.

Furthermore, I would like to thank David Brillinger and Michael Sørensen for setting up my visit at the Statistics Department at University of California, Berkeley. Also thanks to David and my office and class mates (Kai, Hyo Jeong, Hye Kyung and Dave) at U. C. Berkeley for great discussions both academic and non-academic. This extends to my colleges at the Actulus project and in the IE group in Copenhagen. Especially my office mates (André, Sima, Ninna, Mai-Britt and Rune).

Finally, I would like to thank my family for their support. My parents for presenting to me the joy of mathematics, and for their support which makes our everyday work. My daughter, Filippa, for all her laughs which can brighten up the most stressful day. And especially my beloved husband, Esben, for love, proof-reading and for, while holding on to his own dreams, always supporting and engaging in mine.

Kamille Sofie Tågholt Gad
Copenhagen, April 2015

Summary

This thesis consists of an introductory chapter and five papers. The papers are each concerning questions within the topics life insurance, optimal stopping or the interplay between these. Each paper is presented in a chapter, and thus each of the chapters are self-contained and may be read alone. Below, I give a brief overview of the results of each of the chapters. A more thorough overview is presented in Chapter 1.

In Chapter 2 we consider a general geometric Lévy process and solve the non-linear optimal stopping problem of maximizing the variance at the stopping time. For solving this problem we solve an auxiliary quadratic optimal stopping problem. We show that the solution to maximizing variance depends on whether randomized stopping times are included in the set of stopping times we maximize over. For some problems the inclusion of randomized stopping times increase the value function and for some it does not. Even when the value function is not affected by inclusion of randomized stopping times, a solution may be easier to identify when they are.

In Chapter 3 we consider the non-linear optimal stopping problem of maximizing the mean minus a positive constant times the variance at the stopping time. First we solve the problem for spectrally negative geometric Lévy process. We derive both static and dynamic solutions which are excess boundary stopping times. Afterwards we solve the problem for a Cramér-Lundberg process with exponential upwards jumps. We derive a statically optimal stopping time which is a hitting time of an interval, and we derive a dynamically optimal stopping time which is an excess boundary stopping time. Finally, we derive optimal stopping times to the optimal stopping problem of minimizing the variance conditioned on a lower bound on the mean.

In Chapter 4 we consider the American put in a Black-Scholes market. We suggest a model for irrational exercises. We model the exercise by a stochastic intensity which depends on the profitability. Our model contains a single parameter which express how strongly the exercise intensity is affected by the profitability. This parameter we denote *the rationality parameter*. We give sufficient conditions and a probabilistic proof that when the rationality parameter increases to infinity the corresponding prices converge to to classical arbitrage-free price. We conclude the chapter with partial differential equations for valuation under irrational exercise, and we discuss relations to the penalty method.

In Chapter 5 is related to Chapter 4, but in Chapter 5 we consider modelling the time of surrender in a classical life insurance model. We suggest a model where the probability of surrender at any time depends on the profitability. We measure the profitability as the difference between the value of the insurance contract and the surrender value. The value of the insurance contract may be determined as a solution to a differential equation much similar to the Thiele differential equation. As in Chapter 4 the model contains a rationality parameter which express how strongly the surrender probability is affected by the profitability. Again we derive a probabilistic proof of the intuitive convergence result that when the rationality parameter increases to infinity, the value of the life-insurance contract converge to the value corresponding to if the policyholder surrendered at the optimal time.

In Chapter 6 we add stochastic retirement to a classical finite state life insurance model. We do this by splitting the *active* state in a *premium paying* state and a *retired* state. We derive formulas for scaling the benefits reasonably according to the time of retirement. Then we show how to calculate the reserves and expected cash. Afterwards we describe a way to add to the model that policyholders might change their benefit structure upon retirement. We determine formulas for calculating reserves and cash flows in this model too. Finally, we conclude with a numerical investigation of the implication stochastic retirement has on reserves and cash flows.

Sammenfatning på dansk

Denne afhandling består af et introducerende kapitel og fem artikler. Artiklerne beskæftiger sig med spørgsmål indenfor emnerne livsforsikring, optimale stoppetider og samspillet mellem disse. Hver artikel er præsenteret i et kapitel, og kapitlerne kan derfor alle læses enkeltstående. Nedenfor giver jeg et meget overordnet overblik over resultaterne fra hvert kapitel. Et mere grundigt overblik præsenteres i Kapitel 1.

I Kapitel 2 betragter vi en generel geometrisk Lévy proces og løser det ikke-lineære optimale stoppetidsproblem om at maksimere variansen på stoppetidspunktet. For at løse dette problem løser vi først et hjælpeproblem. Hjælpeproblemet er et klassisk, kvadratisk optimal stoppetidsproblem. Vi finder at løsningen til problemet med at maksimere varians afhænger af hvorvidt randomiserede stoppetiden er inkluderet i den mængde af stoppetider vi maksimerer over. For nogle processer vil inklusionen af randomiserede stoppetider hæve værdifunktionen og for andre processer vil det ikke. Selv når værdifunktionen ikke påvirkes af at de randomiserede stoppetider er inkluderet, er problemet nogle gange lettere at løse når de er.

I Kapitel 3 betragter vi det ikke-lineære optimale stoppetidsproblem som går ud på at maksimere middelværdien minus en konstant gange variansen på stoppetidspunktet. Først løser vi problemet for spektralt negative geometriske Lévy processer. Vi udleder både statiske og dynamiske løsninger som er givet ved første gang processen krydser over en grænse. Derefter løser vi problemet for en Cramér-Lundberg proces med exponentialfordelte spring opad. Vi udleder en statisk optimal stoppetid givet ved første gang processen rammer et interval, og vi udleder en dynamisk optimal stoppetid givet ved første gang processen kommer over en grænse. Til sidst udleder vi optimale stoppetider for det optimale stoppetidsproblem som går ud på at minimere variansen givet en nedre grænse på middelværdien.

I Kapitel 4 betragter vi en amerikansk put option i et Black-Scholes marked. Vi foreslår en model for irrationel indløsning af optionen. Vi modellerer indløsningstidspunktet ved hjælp af en stokastisk intensitet som afhænger af hvor profitabelt det er at indløse. Vores model indeholder en enkelt parameter som udtrykker hvor stærkt indløsningsintensiteten påvirkes af hvor profitabelt det er at indløse. Denne parameter kalder vi *rationalitetsparameteren.* Vi giver tilstrækkelige betingelser og et sandsynlighedsteoretisk bevis for at når rationalitetsparameteren konvergerer mod uendelig, så vil de tilsvarende priser konvergere mod den arbitrage-fri pris. Vi afslutter kapitlet med partielle differentialligninger for prisen i en model med irrationel indløsning, og vi diskuterer forbindelsen til *penalty method*.

Kapitel 5 er relateret til Kapitel 4, men i Kapitel 5 betragter vi modellering af genkøbstidspunktet i en klassisk livsforsikringsmodel. Vi foreslår en model hvor sandsynligheden for at genkøbe på et vilkårligt tidspunkt afhænger af hvor profitabelt det er. Vi måler hvor profitabelt det er som forskellen mellem værdien af forsikringskontrakten og genkøbsværdien. Værdien af forsikringskontrakten kan nu bestemmes som løsningen til en differentialligning som minder meget om Thieles differentialligning. Som I Kapitel 4 indeholder vores model en rationalitetsparameter som udtrykker hvor stærkt genkøbssandsynligheden påvirkes af hvor profitabelt det er at genkøbe. Igen giver vi et sandsynlighedsteoretisk bevis for det intuitive resultat at når rationalitetsparameteren vokser mod uendelig, så vil værdien af forsikringskontrakten konvergere værdien svarende til at forsikringstageren genkøbte på det optimale tidspunkt.

I Kapitel 6 tilføjer vi et stokastisk pensioneringstidspunkt til den klassiske endelig tilstands livsforsikrings model. Vi gør dette ved at splitte *aktiv* tilstanden i en *præmiebetalende* tilstand og en *pensioneret* tilstand. Vi udleder formler til at skalere ydelser rimeligt efter pensioneringstidspunktet. Derefter viser vi hvordan reserver og forventede betalingsstrømme kan beregnes. Derefter beskriver vi hvordan vi i modellen kan tilføje at forsikringstagerne kan ændre deres ydelsesstruktur ved pensionering. Vi bestemmer også formler til at beregne reserver og forventede betalingsstrømme i denne model. Til sidst slutter vi af med numerisk at undersøge påvirkningen på reserver og forventede betalingsstrømme fra at tilføje stokastisk pensionering.

List of papers

This thesis is based on five papers:

- Kamille Sofie Tågholt Gad and Jesper Lund Pedersen,
  Variance Optimal Stopping for Geometric Lévy Processes.
  Advances in Applied Probability **47**, 1-18 (2015).

- Kamille Sofie Tågholt Gad (2015),
  Mean-Variance Optimal Stopping for some Geometric Lévy Processes.
  Submitted for publication.

- Kamille Sofie Tågholt Gad and Jesper Lund Pedersen (2015),
  Rationality Parameter for Exercising American Put.
  Submitted for publication.

- Kamille Sofie Tågholt Gad, Jeppe Juhl and Mogens Steffensen (2014),
  Reserve-Dependent Surrender.
  Submitted for publication.

- Kamille Sofie Tågholt Gad and Jeppe Woetmann Nielsen ,
  Reserves and Cash Flows under Stochastic Retirement.
  Scandinavian Actuarial Journal (2015) DOI:10.1080/03461238.2015.1028432

Contents

Preface i

Summary iii

Sammenfatning på dansk v

List of papers vii

1 Introduction 1
- 1.1 Non-Linear Optimal Stopping for Geometric Lévy Processes . 2
- 1.2 Behaviour Modelled by Stopping Times 6

2 Variance Optimal Stopping for Geometric Lévy Processes 11
- 2.1 Introduction 11
- 2.2 The Quadratic Optimal Stopping Problem 14
- 2.3 The Variance Optimal Stopping Problem 18
- 2.4 Randomized Stopping 21
- 2.5 Variance Problem Without Randomized Stopping Revisited . . 24

3 Mean-Variance Optimal Stopping for some Geometric Lévy Processes 37
- 3.1 Introduction 37
- 3.2 Problem Formulation and Approach 39
- 3.3 Spectrally Negative Lévy Processes 40
- 3.4 A Cramér-Lundberg Process 45
- 3.5 Dynamic Optimization 54

4 Rationality Parameter for Exercising American Put 59
- 4.1 Introduction 59
- 4.2 Rationality Parameter for Exercising 60
- 4.3 Valuation Equations 67

5 Reserve-Dependent Surrender 71
5.1 Introduction . . . 71
5.2 Standard Setup . . . 75
5.3 Reserve Dependent Surrender . . . 76
5.4 Reserve Dependent Policyholder Behaviour . . . 78
5.5 Approximation of the Worst Case Reserve . . . 80
5.6 Advanced Surrender Values . . . 83
5.7 Numerical Examples . . . 86
5.8 Conclusion . . . 94

Appendices 97
5.A Proof of Convergence . . . 97

6 Reserves and Cash Flows under Stochastic Retirement 105
6.1 Introduction . . . 105
6.2 Stochastic Retirement in a Simple Model . . . 108
6.3 Stochastic Retirement in a Complex Model . . . 113
6.4 Benefit Conversion . . . 122
6.5 Numerical Results and Discussion . . . 126

Appendices 133
6.A Market Reserve in the Simple Model . . . 133
6.B Market Reserve in the Complex Model . . . 134
6.C Market Reserve with Benefit Conversion . . . 136

1. Introduction

This chapter gives an overview of the contributions of the thesis and it explains to what extent the forthcoming chapters are connected. The introduction here contains no references. References are found in the introductions of the individual chapters.

The topics addressed in this thesis are centred around stopping times. Stopping times is a key term within applications of probability theory in finance and insurance. When an agent faces the choice of taking a specific action at a time of his own choice, stopping times may formalize which strategies the agent can possibly follow.

Though stopping times may be seen as a unifying element in the thesis, the individual papers cover very different problems. There is a long way from the optimal stopping problems addressed in Chapter 2 and Chapter 3 to the modelling of retirement in Chapter 6. However, Chapter 4 and Chapter 5 present some of the ways optimal stopping problems may be reasonably included in life-insurance and financial modelling.

When we use a stopping time to describe the time an agent takes an action, the stopping time may be based on different types of strategies. In Chapter 2 and Chapter 3 we consider problems where we search for stopping times which are optimal relative to some objective. The other extreme we consider is Chapter 6, where the agent is a life-insurance policyholder with a choice of when to retire. Here we model the stopping time as unaffected by what is optimal. We model it by an intensity and a retirement probability which only depends on time and the state of the policyholder. A strategy in between these two extremes is used in Chapter 4 and Chapter 5. Here we consider an agent which is respectively a life-insurance policyholder with a choice to surrender his contract, and a holder of an American put with the choice to exercise his option. In both cases we model the stopping time of the agent with a stochastic intensity which depends on how profitable the action is at the given time.

1.1 Non-Linear Optimal Stopping for Geometric Lévy Processes

In classical optimal stopping problems we consider a stochastic process, X, adapted to a filtration, $\mathcal{F}$. We let $\mathcal{T}$ denote the class of stopping times with respect to $\mathcal{F}$ and then we search for a stopping time $\tau^* \in \mathcal{T}$ such that

$$\sup_{\tau\in\mathcal{T}} \mathbb{E}[G(X_\tau)] = \mathbb{E}[G(X_{\tau^*})], \tag{1.1}$$

where G is some measurable function which we denote the gain function. Classical optimal stopping problems have been thoroughly studied. Though solving specific problems still require work, we have fairly general results stating that if there is a solution, then typically there is a solution given as a hitting time.

In Chapter 2 and Chapter 3 we have studied two optimal stopping problems where the objective we wish to maximize depends on the variance of the process. These problems fall outside the scope of classical optimal stopping as they may not be represented in the form of (1.1), and thus we cannot directly rely on classical results. Instead the problems may be represented as

$$\sup_{\tau\in\mathcal{T}} H(\mathbb{E}[X_\tau], \mathbb{E}[X_\tau^2]) = H(\mathbb{E}[X_{\tau^*}], \mathbb{E}[X_{\tau^*}^2]), \tag{1.2}$$

where $H : \mathbb{R} \times [0, \infty) \to \mathbb{R}$ is a second order polynomial. Specifically, H is non-linear, and for this reason the problems we study are denoted *non-linear optimal stopping problems*. If H was linear, the problem (1.2) would fall under the scope of problem (1.1).

The two non-linear optimal stopping problems studied in this thesis are the following. In Chapter 2 we study the problem

$$\sup_{\tau\in\mathcal{T}} \mathbb{V}[X_\tau], \tag{1.3}$$

and in Chapter 3 we study the problem

$$\sup_{\tau\in\mathcal{T}} \left(\mathbb{E}[X_\tau] - c\mathbb{V}[X_\tau]\right), \tag{1.4}$$

where $c > 0$ and $\mathbb{V}$ denotes variance. Both these problems have been solved when X is a geometric Brownian motion and $\mathcal{T}$ is all stopping times with respect to the completion of the filtration generated from X. In both cases it was shown that whenever the problems are well posed and they have a solution, there is an optimal stopping time which is an excess boundary time, that is, a stopping time on the form $\tau_a^+ = \inf\{t \geq 0 : \ X_t \geq a\}$.

In this thesis, we have searched to expand the results to the case where X is a geometric Lévy process. For both problems our approach resembles the approach taken for geometric Brownian motions. It relies in both cases on identifying families of embedded classical optimal stopping problems which we first solve. We then search for a solution to the non-linear optimal stopping problem among the solutions to the family of embedded problems. With this approach it is not surprising if we find hitting time solutions as we often do for classical optimal stopping problems.

However, for some geometric Lévy processes the jumps of the process immediately complicates the above mentioned technique for various reasons. For some processes the problems (1.3) and (1.4) become easier to solve if we expand the class of stopping times to a class $\hat{\mathcal{T}}$ which includes what is denoted *randomized stopping times*. We define randomized stopping times by augmenting the filtration with the information of a random variable which is uniformly distributed on $[0, 1]$ and which is independent of X. From an application viewpoint it is probably not a problem to expand the class of stopping times to include randomized stopping times. The randomized stopping times are typically not considered in classical optimal stopping problems because they do not change the value function. However, as we show in Chapter 2, the randomized stopping times may have a significant impact on the non-linear optimal stopping problems.

Below we present in more details results and the structure of the proofs for problem (1.3) and (1.4).

In Chapter 2 we consider problem (1.3) of maximizing the variance. We denote this problem *the variance problem*. In an applied context, variance is a possible measure for risk, and when we maximize variance we find an upper boundary for this risk. However, the variance problem is also interesting because there are few papers on non-linear optimal stopping. Solving this problem is a step towards a better understanding of non-linear optimal stopping problems.

Let $\psi(\beta) = \log \mathbb{E}[X^\beta]$ be the Laplace exponent of the underlying Lévy process. For $\psi(2) > 0$ the possible variance is either zero or unbounded, and the problem is primarily interesting for processes with $\psi(2) < 0$, whereas $\psi(2) = 0$ is a boundary case treated separately.

The variance problem is solved by noticing that if there exists some constant $c > 0$ and some stopping time $\tau^* \in \mathcal{T}$ such that

$$\sup_{\tau \in \mathcal{T}} \mathbb{E}[(X_\tau - c)^2] = \mathbb{E}[(X_{\tau^*} - c)^2] \tag{1.5}$$

and

$$\mathbb{E}[X_{\tau^*}] = c \tag{1.6}$$

then τ^* solves the variance problem. We first solve the classical optimal stopping problem of (1.5), and derive that for any c there exists an optimal stopping boundary y such that both τ_y^+ and the stopping time $\tau_y^{++} = \inf\{t \geq 0 : X_t > y\}$ are optimal for (1.5). However, whereas for any geometric Brownian motion with $\psi(2) < 0$ there is a combination of c and τ^* which fulfill both (1.5) and (1.6), this is not the case for geometric Lévy processes.

We show that 0 being regular for $(0, \infty)$ ensures that the variance problem has an excess boundary solution which fulfills both (1.5) and (1.6). It is sometimes possible to use the approach above even when 0 is irregular for $(0, \infty)$. We deduce two equations for when we may use the above method for finding an excess boundary solution.

When there is no excess boundary time fulfilling both (1.5) and (1.6), the variance problem becomes significantly easier by maximizing over the class of randomized stopping times instead. This way we are able to find more solutions to the problem (1.5) and we derive that there will be one solving both (1.5) and (1.6).

We return to the original variance problem where randomization is not allowed. When there is no combination of τ^* and c which fulfill both (1.5) and (1.6), then the situation depends on the jump structure and the drift of the Lévy process. For compound Poisson processes the randomized stopping times can be mimicked by a stopping time from $\mathcal{T}$. The processes which are not compound Poisson processes have no stopping time in $\mathcal{T}$ giving as high a variance as the one obtained by the randomized solution. If the jump part is not a Poisson process, we may find a sequence of stopping times from $\mathcal{T}$ for which the variance converges to the variance of the randomized solution. But if the jump part is a compound Poisson process then there is a gap between the value function of the variance problem with and without randomized stopping times.

In Chapter 3 we consider the problem (1.4) of maximizing expectation minus a constant times the variance. We denote this problem *the mean-variance problem*. The mean-variance problem has the interpretation as to maximize some objective while minimizing risk.

We approach the mean-variance problem by noticing that

$$\sup_{\tau:\mathbb{E}_x[X_\tau]} (\mathbb{E}_x[X_\tau] - c\mathbb{V}_x[X_\tau]) = \sup_{M \geq x} \left(M + cM^2 - c \inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2] \right).$$

Thus, we may solve the mean-variance problem if we solve all the inner conditional problems. The inner conditional problem we solve by the Lagrange approach searching for a λ such that there exists a τ^*, which is optimal for

$$\sup_\tau \mathbb{E}_x[\lambda X_\tau - X_\tau^2] \tag{1.7}$$

and fulfills $\mathbb{E}_x[X_{\tau^*}] = M$. This way we for spectrally negative Lévy processes derive excess boundary times which are optimal for the mean-variance problem, and we derive an implicit expression for the bound. From the derivation of the optimal stopping time to the mean-variance problem, we deduce the solution to the optimal stopping problem given by

$$\inf_{\tau:\mathbb{E}[X_\tau]\geq M} \mathbb{V}[X_\tau].$$

This problem we denote *the conditional variance problem* and again we find an optimal stopping time which is an excess boundary time.

For processes with upwards jumps the mean-variance problem becomes more involved. This is mainly because the classical optimal stopping prolems (1.7) becomes difficult to solve. For the spectrally negative Lévy processes there is no difference between hitting times of an interval above the starting value of the process and an excess boundary times of the intervals lower bound. However, these two stopping times are not the same if the process may jump across the interval. From a first glance at the classical optimal stopping problem (1.7) we guess that this problem is solved by a hitting time for an interval, and these are in general difficult to work with for geometric Lévy processes.

We study a Cramér-Lundberg process with upwards exponentially distributed jumps. For this process we have explicit formulas for the distribution of the process value upon hitting times of intervals. By first expanding the class of stopping times we maximize over to the randomized stopping times, we find an optimal stopping time for each of the inner conditional problems. From these, we find a stopping time from $\mathcal{T}$, which is optimal for the mean-variance problem. As anticipated the solution we find for the Cramér-Lundberg process is a hitting time of an interval. This may be problem for applicational purposes of the mean-variance problem, since the punishment of risk was only supposed to protect against large downside risk, but apparently for these processes also prevent large values.

The randomized stopping times do not have an impact on the solution of the mean-variance problem for the Cramér-Lundberg process. However, for the conditional variance problem for the Cramér-Lundberg process, then for some starting values they do. This is not a problem for applicational purposes, but it is a remarkable feature.

One of the draw-downs to the mean-variance problem is that it is not time-consistent, and thus the stopping regions of the optimal stopping times depend on the starting value. As has been suggested in the study of geometric Brownian motions we may create time consistency by instead considering the corresponding dynamic problem. For the mean-variance problem this corresponds to searching for a stopping time τ^* such that there is no other stopping time, σ with

$$\mathbb{P}_x(\mathbb{E}_{X_{\tau^*}}[X_\sigma] - c\mathbb{V}_{X_{\tau^*}}[X_\sigma] > X_{\tau^*}) > 0,$$

where the subscript refers to the starting value of the process. For the dynamic problem we find that the mean-variance problem for the studied processes have excess boundary times as optimal stopping times.

1.2 Behaviour Modelled by Stopping Times

This thesis has been written as part of the Actulus project. The Actulus project focusses on solving Solvency II issues for pension funds. One of the challenges pension funds face regarding the forthcoming Solvency II regulations is the enhanced requirement for modelling of policyholder behaviour. In paragraph 79 of the Solvency II directive it is stated that *"... Any assumptions made by insurance and reinsurance undertakings with respect to the likelihood that policy holders will exercise contractual options, including lapses and surrenders, shall be realistic and based on current and credible information... "*. This requirement has motivated the studies in the last three chapters of the thesis. These chapters all concern modelling the exercise of contractual options more realistically by use of stopping times. One chapter concerns a problem from finance and the two others are problems directly relevant in the process of meeting the Solvency II requirements for pension funds.

Thus, the last three Chapters of this thesis concern contracts where an agent has an option to take an action at a time of his own choice. We model the agents behaviour by a stopping time which describes if and when he takes the action, and we investigate the impact this has on market valuation and cash flows of the contracts. The content of Chapter 4 and Chapter 5 is described in Section 1.2.1, and the content of Chapter 6 is described in Section 1.2.2.

1.2.1 Modelling Surrender and Exercise of an American Put by Rational Behaviour

In Chapter 4 the agent is a holder of an American put and the choice he faces is the right to exercise the put. In Chapter 5 the agent is a life insurance policyholder and the choice he faces is the right to surrender his insurance contract and in return receive a surrender sum. The two options immediately seem to have a lot in common. However, in traditional financial modelling the holder of the American put is assumed to behave as if he is solving an optimal stopping problem wanting to maximize the expected pay-off of the put under some pricing measure. And in traditional actuarial modelling the policyholder with the surrender option is assumed to behave unaffected by what is optimal. It is probably true that the exercise of the put has a stronger dependence on what is optimal, than the exercise of the surrender option has. Nevertheless, in both cases a more realistic modelling of the agents behaviour is probably somewhere in between these two extremes.

In between the extremes exist all kinds of models where the time of intervention is modelled by a stochastic intensity, but where the intensity not only depends on time, but also some stochastic factors. Various external factors appear obvious. However, rather than letting the intensity depend on external factors, one could let the intensity depend on internal factors relevant to the specific contract. We propose in Chapter 4 and Chapter 5 models where the intensity depends on how profitable it is to take the action. In both situations we measure this profitability as the difference between the pay-off upon intervention and the value of the contract if no immediate intervention is done. The value when no intervention is done should however take into account the possibility of future intervention.

In both Chapter 4 and Chapter 5 we introduce an intervention function, which gives the intensity of intervention as a function of the profitability. We find differential equations which if only they have a unique solution, give the value function of the contract. The differential equations resemble respectively the familiar Black-Scholes PDE and the Thiele differential equation with the only difference that they have a non-linear term representing intervention. The definition of the value of the contract and the intensity of intervention is circular, but we show how existence and uniqueness of a solution to the differential equation ensures that this is not a problem.

In both cases we think that a reasonable choice for the intensity functions f is a function which is positive and increasing. Instead of just considering one intensity function, we may consider a parametrized family of intensity functions. We may choose the parametrized family such that the parameter expresses how much the agent is affected by what is optimal. Two examples of such families are:

$$f_{\psi,\theta}(x) = \psi e^{\theta x}, \tag{1.8}$$
$$f_{\theta}(x) = \theta 1_{(x\geq 0)}, \tag{1.9}$$

where ψ, θ are parameters. For (1.8) ψ controls the overall tendency to intervene, whereas θ controls how profitability creates deviations from the overall tendency. For (1.9) θ controls both. In both cases we may think of θ

as a rationality parameter, and each parameter value gives a contract value which we may denote V_θ

In both Chapter 4 and Chapter 5 we give a probabilistic proof and clarify sufficient conditions for a convergence result that may seem intuitively clear: If the tendency to intervene tends to zero when the gain from intervention is negative, and if it tends to infinity whenever the gain from intervention is positive, we reach in the limit at the value based on optimal behaviour.

What is meant by this mathematically is that if we let the rationality parameter increase to infinity and the corresponding intensity functions fulfill some convergence requirements as given in Chapter 4 and Chapter 5, then $V_\theta \to V_A$ pointwise, where V_A denotes the contract value if agents behave optimally. This further motivates the name of the rationality parameter.

The convergence result gives a way to approximate the worst-case reserves. Particularly for the intensity function (1.9) this approximation method is known in finance as the penalty method. Our contribution in Chapter 4 is twofold. First we show how the values obtained for a fixed parameter value may be thought of as the price in a model where the agent is affected by what is profitable without behaving completely optimal though. Secondly we give a probabilistic proof that the convergence holds, by breaking down the sub-optimal behaviour in too early versus too late intervention and various degrees of severity of the time of intervention. For the surrender option the penalty method has, to the knowledge of the authors not been used before, and thus the contributions of Chapter 5 is threefold as it is both the suggested modelling of rational behaviour, the approximation method of the worst case reserve, and the probabilistic proof.

The fact that the method works for the American put gives us reason to believe that the method might work for the surrender option in a model with stochastic interest rate as well. However, one should be aware that the extra stochasticity in the model of the American put has lead us to extra restrictions on the parametrized family of intensity functions to ensure convergence. We would most likely need these enhanced restrictions for the surrender model as well if we want to add a stochastic interest rate.

In the model studied in Chapter 4 and Chapter 5 the agents are affected by what would be optimal behaviour with regards to expected value of discounted future payments. In line with the reasoning done by Markowitz, which motivated the study of Chapter 3, we may consider if it is reasonable to have the agent affected by the risk of the future payments. Within some financial applications such modelling might be reasonable. However, for life insurance policyholders a reasonable valuation of the policyholders risk is not straight forward. This is primarily because the intend of the life insurance contract is to hedge a stochastic, personal need for money. Thus, the stochasticity of future payment is not in itself undesirable.

1.2.2 Stochastic Retirement

In Chapter 6 the agent is a life insurance policyholder and the choices he faces are when to retire and what benefit structure he wants. Traditionally the time of retirement has been modelled as a deterministic time point, and the benefit structure is assumed to be settled upon the settlement of the contract. However, in reality policyholders are often allowed to change both the time of retirement and convert between different structures of the benefits. In Chapter 6 we address some of the challenges from modelling the time of retirement and the structure of the benefit as stochastic. Combined we call it *stochastic retirement*.

In classical models the state of the policyholder is described by a finite state Markov chain. Usually the state of premium paying and retired are the same. We introduce a stochastic time of retirement by letting the state of premium paying and retired be two different states in the Markov model. We assume that all transition probabilities are deterministic and known, but unlike the other transitions, we let retirement happen with positive probability at predefined time points.

In Chapter 6 we consider two setups. One is a simple life-death model with a premium, a pension sum, and a life annuity. The other setup is an expanded model where the policyholder in excess may become disabled, reactivate, and he may convert to free policy. His product is also expanded as it includes a disability annuity.

In a model with stochastic retirement it is reasonable to have the size of the benefits depend on the time of retirement. As is common for modelling of the policyholder behaviours surrender and conversion to free policy we choose the scaling of the benefits such that the risk sum upon retirement is zero under the technical basis. We formalize benefit scaling by first defining reference benefits corresponding to the benefits of a model with no policyholder options. We then define time dependent scaling factors to be multiplied on the reference benefits upon each policyholder behaviour. In each case we show that the factors may be calculated from traditional Thiele differential equations based on the reference benefits. In the simple model the scalings are given uniquely when we use the condition that the contract is divided into partial product and within each the benefits are scaled proportionally in a way such that the the principle of zero risk sum is fulfilled. However, in the complex model we need an extra specification which may be thought of as the expected retirement date. This specification serves to fix the relation between the downscaling of disability coverage relative to downscaling of retirement benefits upon conversion to free policy.

With the benefit scaling determined it is no problem to derive a Thiele differential equation for the market values. Calculation of expected discounted cash flows are more involved. We find that the expected discounted cash flow may be calculated in a tractable way through transition probabilities and some modified transition probabilities. The method we use has previously been used for a model containing only free policy scaling, and we show how the method may be expanded to a situation where we have multiple types of behaviour induced scalings.

The final topic covered in Chapter 6 is benefit conversion. We model this by assuming that upon retirement the policy holder may choose to convert a proportion of the reserve to each other benefit type. And we assume he is allowed to buy benefits on the technical basis. We define the conversion proportion as a stochastic process which we assume to be independent on everything else in the model. The model does not immediately fit in the traditional Markov model we have worked with since payments now are not given directly from the state process (they are not measurable with respect to the filtration generated from the state-process). We construct a new model which has payments given directly from the state-process and which is equivalent to our benefit conversion model in the sense that it produces the same market values and the same expected cash flows. However it should be stressed that the equivalent model only produces equivalent expectations and may not be used for risk analysis, which require other distributional properties of the cash flow.

We conclude Chapter 6 with numerical examples which illustrate a significant impact on both expected discounted cash flows and market values.

2. Variance Optimal Stopping for Geometric Lévy Processes

Kamille Sofie Tågholt Gad, Jesper Lund Pedersen

Abstract

The main result of this paper is the solution to the optimal stopping problem of maximizing the variance of a geometric Lévy process. We call this problem the variance problem. We show that, for some geometric Lévy processes, we achieve higher variances by allowing randomized stopping. Furthermore, for some geometric Lévy processes, the problem has a solution only if randomized stopping is allowed. When randomized stopping is allowed, we give a solution to the variance problem. We identify the Lévy processes for which the allowance of randomized stopping times increases the maximum variance. When it does, we also solve the variance problem without randomized stopping.

Keywords: Variance criterion; variance optimal stopping; geometric Lévy processes; quadratic optimal stopping.

2.1 Introduction

In this paper we solve the optimal stopping problem of maximizing the variance of a geometric Lévy process. We call this problem the variance problem. It is distinguished from classical optimal stopping problems in that we maximize the variance and not the expectation. The nonlinear structure of the variance moves the problem outside the scope of classical optimal stopping problems, and, thus, we cannot directly rely on results from, e.g. [7] and [9].

As in Markowitz mean-variance analysis [3] we identify the variance of a stock price with a risk. In the context of risk management, where an investor wishes to sell an asset, the variance problem provides the worst-case scenario, that is, the value function is an upper bound for the risk (variance) for any strategy and the optimal strategy is the strategy at highest risk.

The results in this paper extend the results of [6], in which the variance problem is solved for various diffusions. However, we face different technical issues in working with geometric Lévy processes. Whereas the optimal stopping times for the diffusions in [6] are hitting times, this is not the case for all geometric Lévy processes. For some geometric Lévy processes, the solution is of another kind. Furthermore, for some of these processes, we achieve higher variances by allowing randomized stopping. For some processes the variance problem has a solution only if randomized stopping is allowed. This is in contrast to classical optimal stopping problems and variance problems for diffusions where randomized stopping times do not change the value function.

Mathematically, the problem addressed in this paper is the following. Let X be a Lévy process, let $\mathcal{F}$ be the augmented natural filtration satisfying the usual conditions, and let $\mathcal{T}$ denote the set of stopping times with respect to $\mathcal{F}$ (all terms defined as in [2]). The main problem we consider is to find a stopping time $\tau^* \in \mathcal{T}$ such that

$$\sup_{\tau\in\mathcal{T}} \mathbb{V}[e^{X_\tau}] = \mathbb{V}[e^{X_{\tau^*}}]. \tag{2.1}$$

We call this problem the variance problem.

Initially, we identify the Lévy processes for which the variance problem is trivial to solve. If X is a deterministic process then $\mathcal{T}$ contains only almost surely (a.s.) deterministic times and the variance for any stopping time is 0. Let $\psi(\beta) = \log(\mathbb{E}[e^{\beta X_1}])$ denote the Laplace exponent. If $\psi(2) > 0$ and the Lévy process is nondeterministic, then $\psi(2) > 2\psi(1)$ by Jensen's inequality. Therefore, $\mathbb{V}[e^{X_t}] = \mathbb{E}[e^{2X_1}]^t - \mathbb{E}[e^{X_1}]^{2t} = e^{\psi(2)t} - e^{2\psi(1)t} \to \infty$ for $t \to \infty$ and the variance problem is unbounded.

In the following we consider Lévy processes with $\psi(2) < 0$. Lévy processes with $\psi(2) = 0$ are considered separately (see Theorem 2.3.1).

In [6], the variance problem was solved for various diffusions. This was achieved using a method of embedding the problem into the following classical optimal stopping problem, which we call the quadratic problem

$$\sup_{\tau\in\mathcal{T}} \mathbb{E}[(e^{X_\tau} - c)^2] = \mathbb{E}[(e^{X_{\tau^*}} - c)^2], \quad c > 0. \tag{2.2}$$

The solution to the variance problem of this paper is also based on this embedding method, and the solution to the quadratic problem is presented in Theorem 2.2.1 below. As shown in [6], it holds that if a solution τ^* to (2.2) solves

$$\mathbb{E}[e^{X_{\tau^*}}] = c \tag{2.3}$$

then it is also an optimal stopping time for variance problem (2.1).

The processes studied in [6] all have a combination of $\tau^* \in \mathcal{T}$ and $c \in \mathbb{R}$ that solves both (2.2) and (2.3). But some Lévy processes do not. Let $\bar{X}_\infty = \sup_{t\geq 0} X_t$. The problem of finding a combination of τ^* and c that solves both

(2.2) and (2.3) arises from possible discontinuities in the distribution of $\bar{X}_\infty$. Discontinuities exist exactly when 0 is irregular for $(0, \infty)$ (see Lemma 2.2.2), and continuity of the distribution of $\bar{X}_\infty$ ensures that the variance problem has an excess boundary time solution. It is sometimes possible to use the embedding method even when the distribution of $\bar{X}_\infty$ has discontinuities. We derive two equations that each give sufficient conditions that the embedding method can be used to find an excess boundary time solution (see Theorem 2.3.1).

When there is no excess boundary time that solves both (2.2) and (2.3), we solve the variance problem by introducing randomized stopping times. The concept is to allow stopping decisions to depend not only on the Lévy process, but also on a random variable independent of the Lévy process. As argued in [9], this may be powerful when solving optimal stopping problems with constraints, because it sometimes easily gives a wider class of solutions to the unconstrained problem. We see in Theorem 2.4.2 that the class of randomized optimal stopping times for the quadratic problem (2.2) is so wide that one of them also solves (2.3). Thus, for any Lévy process, it is possible to solve the variance problem with the embedding method if we maximize over the randomized stopping times.

We return to the original variance problem, where randomization is not allowed. In the case there is no combination of τ^* and c that solves both (2.2) and (2.3), the situation depends on the jump structure and the drift of the Lévy process. For compound Poisson processes, the randomized stopping times can be mimicked because the processes stay for a positive time at 0 before the first jump. This positive time is independent of the rest of the behaviour of the process and in Theorem 2.5.2 we show how this is used as the independent random information needed. The process which are not compound Poisson processes moves from 0 right away. In Theorem 2.5.3 we show that, for these processes, it holds that if there is no excess boundary time solution then the randomized solutions cannot be mimicked and there is no stopping time in $\mathcal{T}$ giving as high a variance as that obtained by the randomized solution. If the jump part is not a compound Poisson process then the filtration grows sufficiently fast that we may find a sequence of stopping times from $\mathcal{T}$ for which the variance converges to the variance of the randomized solution. But if the jump part is a compound Poisson processes then the filtration does not generate sufficient information and there is a gap between the value function of the variance problem with and without randomized stopping times (see Theorem 2.5.3).

2.2 The Quadratic Optimal Stopping Problem

In this section we solve the quadratic optimal stopping problem (2.2) for Lévy processes with $\psi(2) < 0$. This problem has some resemblance to the optimal stopping problem presented in [1] and is solved by similar method.

The quadratic problem is a classical optimal stopping problem for a Lévy process with gain function $G(x) = (e^x - c)^2$. As G is continuous, and Lévy processes are Feller processes, then the state space can be divided into a stopping region and a continuation region (see [7]), with an optimal stopping time being the first time the process reaches the stopping region. To get a first idea of the stopping region, note that, from Jensen's inequality, a Lévy process with $\psi(2) < 0$ has $\mathbb{E}[X_1] < 0$ and, thus, it converges to $-\infty$ when t goes to ∞ (see [2, Theorem 7.2]). Hence, when maximizing $\mathbb{E}[(e^{X_\tau} - c)^2]$, it is clear that the value c^2 may be obtained by never stopping the process. Therefore, it is never optimal to stop the process if $(e^{X_t} - c)^2 < c^2$ and the stopping region has to be above $\log(2c)$.

We use the following notation for the excess boundary times. For $y \in \mathbb{R}$, define

$$\tau_y^+ = \inf\{t \geq 0 | X_t \geq y\} \text{ and } \tau_y^{++} = \inf\{t \geq 0 | X_t > y\}.$$

Recall that

$$\bar{X}_\infty = \sup_{t \geq 0} X_t.$$

When we solve the quadratic and the variance problems, we repeatedly need the following fluctuation identity, which is a minor generalization of [1, Lemma 1]. If $x, y, \beta \in \mathbb{R}$ and $\mathbb{E}[e^{\beta \bar{X}_\infty}] < \infty$, then

$$\mathbb{E}_x\left[e^{\beta X_{\tau_y^+}} 1_{(\bar{X}_\infty \geq y)}\right] = e^{\beta x} \frac{\mathbb{E}[e^{\beta \bar{X}_\infty} 1_{(\bar{X}_\infty \geq y - x)}]}{\mathbb{E}[e^{\beta \bar{X}_\infty}]}, \tag{2.4}$$

where the subscript on the expectation refers to the starting value of the process. The proof follows in the same way as the proof of [1, Lemma 1].

As in [2], we say that we have continuous fit if the value function is continuous at the boundary of the stopping region, and we say that we have smooth fit if the value function is differentiable at the boundary of the stopping region.

Theorem 2.2.1. *Let X be a Lévy process with $\psi(2) < 0$. Then $\tau_{y_c}^+$ and $\tau_{y_c}^{++}$ are both optimal stopping times of the quadratic problem* (2.2), *where $y_c = \log(2c\mathbb{E}[e^{2\bar{X}_\infty}]/\mathbb{E}[e^{\bar{X}_\infty}])$.*

a) If X is spectrally negative then $y_c = \log(2c(\phi(0) - 1)/(\phi(0) - 2))$, where ϕ is the right inverse of ψ.

b) There is always continuous fit at y_c, and there is smooth fit at y_c exactly if the distribution of $\bar{X}_\infty$ is continuous at 0.

Proof. We choose $\tau^* = \tau_{y_c}^+$ as a candidate for an optimal stopping time and define the corresponding function $v^*(x) = \mathbb{E}_x[G(X_{\tau^*})]$. By [5, Lemma 1], $\psi(2) < 0$ implies that $\mathbb{E}[e^{2\bar{X}_\infty}] < \infty$ and $\mathbb{E}[e^{\bar{X}_\infty}] < \infty$, and y_c and $v^*(x)$ are well defined. Then we use the well-known sufficient conditions (see [2, Lemma 9.1]) that τ^* is an optimal stopping time if the following conditions holds:

i) $\mathbb{P}_x($ there exists $\lim_{t\to\infty} G(X_t) < \infty) = 1$,

ii) $v^*(x) \geq G(x)$ for all $x \in \mathbb{R}$,

iii) $(v^*(X_t))_{t\geq 0}$ is a right continuous supermartingale.

We show that each of the three conditions are fulfilled.

i) Whenever $\psi(2) < 0$, the Lévy process converges to $-\infty$ when t goes to ∞ and, thus, the requirement is fulfilled.

ii) It follows from (2.4) and the definition of y_c that

$$\begin{aligned}
v^*(x) &= \mathbb{E}_x[(e^{X_{\tau_{y_c}^+}} - c)^2] \\
&= c^2 - 2c\mathbb{E}_x[e^{X_{\tau_{y_c}^+}} 1_{(\bar{X}_\infty \geq y_c)}] + \mathbb{E}_x[e^{2X_{\tau_{y_c}^+}} 1_{(\bar{X}_\infty \geq y_c)}] \\
&= c^2 - 2ce^x \frac{\mathbb{E}[e^{\bar{X}_\infty} 1_{(\bar{X}_\infty \geq y_c - x)}]}{\mathbb{E}[e^{\bar{X}_\infty}]} + e^{2x} \frac{\mathbb{E}[e^{2\bar{X}_\infty} 1_{(\bar{X}_\infty \geq y_c - x)}]}{\mathbb{E}[e^{2\bar{X}_\infty}]} && (2.5) \\
&= (c - e^x)^2 + 2ce^x \left(1 - \frac{\mathbb{E}[e^{\bar{X}_\infty} 1_{(\bar{X}_\infty \geq y_c - x)}]}{\mathbb{E}[e^{\bar{X}_\infty}]}\right) - e^{2x} \left(1 - \frac{\mathbb{E}[e^{2\bar{X}_\infty} 1_{(\bar{X}_\infty \geq y_c - x)}]}{\mathbb{E}[e^{2\bar{X}_\infty}]}\right) \\
&= G(x) + e^x \left(e^{y_c} \frac{\mathbb{E}[e^{\bar{X}_\infty} 1_{(\bar{X}_\infty < y_c - x)}]}{\mathbb{E}[e^{2\bar{X}_\infty}]} - e^x \frac{\mathbb{E}[e^{2\bar{X}_\infty} 1_{(\bar{X}_\infty < y_c - x)}]}{\mathbb{E}[e^{2\bar{X}_\infty}]}\right) \\
&\geq G(x) + e^x \left(e^{y_c} \frac{\mathbb{E}[e^{\bar{X}_\infty} 1_{(\bar{X}_\infty < y_c - x)}]}{\mathbb{E}[e^{2\bar{X}_\infty}]} - e^x \frac{e^{y_c - x}\mathbb{E}[e^{\bar{X}_\infty} 1_{(\bar{X}_\infty < y_c - x)}]}{\mathbb{E}[e^{2\bar{X}_\infty}]}\right) && (2.6) \\
&= G(x).
\end{aligned}$$

iii) Let Y be an independent copy of X, and let $\bar{X}_{[t,\infty)} = \sup_{s\in[t,\infty)} X_s$. Note that $\bar{X}_{[t,\infty)}$ is equal in law to $X_t + \bar{Y}_\infty$. Then use (2.5) and the definition of y_c to get:

$$\begin{aligned}
\mathbb{E}[v^*(X_t+x)] &= \mathbb{E}[c^2 - 2ce^{x+X_t}\tfrac{\mathbb{E}[e^{\bar{Y}_\infty}1_{(\bar{Y}_\infty \geq y_c-(x+X_t))}|X]}{\mathbb{E}[e^{\bar{Y}_\infty}]} + e^{2(x+X_t)}\tfrac{\mathbb{E}[e^{2\bar{Y}_\infty}1_{(\bar{Y}_\infty \geq y_c-(x+X_t))}|X]}{\mathbb{E}[e^{2\bar{Y}_\infty}]}] \\
&= c^2 - \tfrac{2c}{\mathbb{E}[e^{\bar{X}_\infty}]}\mathbb{E}[e^{\bar{Y}_\infty+x+X_t}1_{(\bar{Y}_\infty+x+X_t\geq y_c)}] + \tfrac{1}{\mathbb{E}[e^{2\bar{X}_\infty}]}\mathbb{E}[e^{2\bar{Y}_\infty+x+X_t}1_{(\bar{Y}_\infty+x+X_t\geq y_c)}] \\
&= c^2 - \tfrac{2c}{\mathbb{E}[e^{\bar{X}_\infty}]}e^{y_c}\left(\mathbb{E}_x[e^{\bar{X}_{[t,\infty)}-y_c}1_{(\bar{X}_{[t,\infty)}\geq y_c)}] - \mathbb{E}_x[e^{2(\bar{X}_{[t,\infty)}-y_c)}1_{(\bar{X}_{[t,\infty)}\geq y_c)}]\right) \\
&\leq c^2 - \tfrac{2c}{\mathbb{E}[e^{\bar{X}_\infty}]}e^{y_c}\left(\mathbb{E}_x[e^{\bar{X}_\infty-y_c}1_{(\bar{X}_\infty\geq y_c)}] - \mathbb{E}_x[e^{2(\bar{X}_\infty-y_c)}1_{(\bar{X}_\infty\geq y_c)}]\right) \\
&= v^*(x).
\end{aligned}$$

Hence, for $s \leq t$ it follows that $\mathbb{E}[v^*(X_t)|\mathcal{F}_s] = \mathbb{E}[v^*((X_t - X_s) + X_s)|\mathcal{F}_s] \leq v^*(X_s)$. Thus, $v^*(X_t)$ is a supermartingale. We then need to prove that $v^*(X_t)$ is right continuous in t. As X_t is right continuous in t, it is sufficient to prove that $x \mapsto v^*(x)$ is continuous. From (2.5), it follows that the jump size of v^* at x is

$$\mathbb{P}(\bar{X}_\infty = y_c - x)\left(-2ce^x\tfrac{e^{y_c-x}}{\mathbb{E}[e^{\bar{X}_\infty}]} + e^{2x}\tfrac{e^{2(y_c-x)}}{\mathbb{E}[e^{2\bar{X}_\infty}]}\right) = \tfrac{\mathbb{P}(\bar{X}_\infty = y_c-x)e^{y_c}}{\mathbb{E}[e^{2\bar{X}_\infty}]}\left(-2c\tfrac{\mathbb{E}[e^{2\bar{X}_\infty}]}{\mathbb{E}[e^{\bar{X}_\infty}]} + e^{y_c}\right) = 0.$$

Thus, v^* is continuous and, hence, $\tau_{y_c}^+$ is an optimal stopping time for the quadratic problem.

We can prove that $\tau_{y_c}^{++}$ is also an optimal stopping time in the same way as for $\tau_{y_c}^+$. The only difference is that we need the following modification of (2.4): Whenever $x, y, \beta \in \mathbb{R}$ and $\mathbb{E}[e^{\beta\bar{X}_\infty}] < \infty$, we have

$$\mathbb{E}_x\left[e^{\beta X_{\tau_y^{++}}}1_{(\bar{X}_\infty > y)}\right] = e^{\beta x}\frac{\mathbb{E}[e^{\beta\bar{X}_\infty}1_{(\bar{X}_\infty > y-x)}]}{\mathbb{E}[e^{\beta\bar{X}_\infty}]}. \tag{2.7}$$

a) When X is spectrally negative, $\bar{X}_\infty \sim \text{Exp}(\phi(0))$ and it follows that $y_c = \log(2c(\phi(0)-1)/(\phi(0)-2))$.

b) Finally, we present the statements about the continuous fit and the smooth fit. As v^* is continuous, we have continuous fit. It follows from ii) that there is smooth fit at y_c exactly if $h(x) = e^{y_c}\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty < y_c - x)}] - e^x\mathbb{E}[e^{2\bar{X}_\infty}1_{(\bar{X}_\infty < y_c-x)}]$ is differentiable in y_c. For $\varepsilon > 0$, we have $h(y_c+\varepsilon) - h(y_c) = 0$ and

$$h(y_c) - h(y_c-\varepsilon) = e^{y_c}\mathbb{E}[e^{\bar{X}_\infty}(1-e^{\bar{X}_\infty})1_{(\bar{X}_\infty<\varepsilon)}](e^{y_c} - e^{y_c-\varepsilon})(-\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty<\varepsilon)}]).$$

Thus, $\frac{1}{\varepsilon}(h(y_c) - h(y_c-\varepsilon)) \to e^{-y_c}\mathbb{P}(\bar{X}_\infty = 0)$, and, hence, we have smooth fit at y_c exactly if the distribution of $\bar{X}_\infty$ is continuous at 0. This completes the proof. □

Remark 2.2.1. *Recall that* 0 *is regular for* $(0,\infty)$ *if* $\tau_0^{++}=0$ *a.s., and* 0 *is irregular for* $(0,\infty)$ *if* $\tau_0^{++}>0$ *a.s.. By [2, Theorem 6.5], the latter is a subclass of Lévy processes with bounded variation, and it contains compound Poisson processes and processes with strictly negative drift. Finally, by [8, Lemma 49.6] we obtain* $\tau_y^{+}=\tau_y^{++}$ *a.s. for* $y>0$*, if* X *is not a compound Poisson process.*

Lemma 2.2.2. *Let* X *be a Lévy process with* $\psi(2)<0$.

a) If 0 *is regular for* $(0,\infty)$ *then the distribution of* $\bar{X}_\infty$ *is continuous.*

b) If 0 *is irregular for* $(0,\infty)$ *then the distribution of* $\bar{X}_\infty$ *has discontinuity points that includes* 0*. If, additionally,* X *is not a compound Poisson process then* 0 *is the only discontinuity point for the distribution of* $\bar{X}_\infty$.

Proof. a) Assume that 0 is regular for $(0,\infty)$. Clearly, $\mathbb{P}(\bar{X}_\infty=0)=0$. From Remark 2.2.1, it follows that $\tau_y^{+}=\tau_y^{++}$ a.s. for $y>0$ and, thus, for $y>0$,

$$\mathbb{P}(\bar{X}_\infty=y)=\mathbb{P}(\tau_y^{+}<\infty)-\mathbb{P}(\tau_y^{++}<\infty)=0.$$

Recall that $\psi(2)<0$ implies that X converges to $-\infty$. Thus, $\bar{X}_\infty<\infty$, and it follows that the distribution of $\bar{X}_\infty$ is continuous.

b) Assume that 0 is irregular for $(0,\infty)$. We prove the statement by contradiction and, thus, assume that $\mathbb{P}(\bar{X}_\infty=0)=0$. Then $\tau_0^{++}<\infty$ a.s.. Let $Y_1=X_{\tau_0^{++}}$ and $X^{(1)}=X$, and, for $n\in\{2,3,\ldots\}$, let $Y_n=X^{(n)}_{\tau_0^{++}}$ and $X^{(n)}_t=X^{(n-1)}_{\tau_0^{++}+t}-X^{(n-1)}_{\tau_0^{++}}$. It follows from induction, the strong Markov property, and the fact that $\tau_0^{++}<\infty$ a.s. that, for all $n\in\mathbb{N}$, the Lévy processes $X^{(n)}$ are identically distributed and a.s. well defined. Let $\tau_0^{++}(X^{(n)})=\inf\{t\geq 0: X_t^{(n)}>0\}$. Then the sequences $(Y_n)_{n\in\mathbb{N}}$ and $(\tau_0^{++}(X^{(n)}))_{n\in\mathbb{N}}$ are both independent and identically distributed (i.i.d.). For all $N\in\mathbb{N}$,

$$X_{\sum_{n=1}^{N}\tau_0^{++}(X^{(N)})}=\sum_{n=1}^{N}Y_n.$$

From the law of large numbers, it follows that $\sum_{n=1}^{N}\tau_0^{++}(X^{(n)})\to\infty$ a.s. when $N\to\infty$. Since $Y_n\geq 0$ for $n\in\mathbb{N}$, then, for X, there will a.s. exist arbitrarily large $t\in(0,\infty)$ with $X_t\geq 0$. But this is in contradiction to $X_t\to-\infty$ a.s. when $t\to\infty$. $\square$

Lemma 2.2.3. *Let* X *be a Lévy process with* $\psi(2)<0$ *which is not a compound Poisson process. If* $y_c=0$ *from Theorem 2.2.1 and* τ *is an optimal stopping time for the quadratic problem such that* $\mathbb{P}(\tau>0)>0$*, then* $\tau\geq\tau_0^{++}$.

Proof. If 0 is irregular for $(0,\infty)$, the statement follows easily. If 0 is irregular for $(0,\infty)$, we prove the statement by contradiction and assume that $\mathbb{P}(\tau < \tau_0^{++}) > 0$. First note, by Blumenthal's 0-1 law, that $\tau > 0$ a.s. It holds that $v^*(x) = \mathbb{E}_x[G(X_{\tau_0^{++}})]$ and, by (2.6), it follows that $G(x) < v^*(x)$ for $x < 0$ because $\mathbb{P}(\bar{X}_\infty = 0) > 0$. By [8, Exercise 50.4], the assumption that $\mathbb{P}(\tau < \tau_0^{++}) > 0$ implies that $\mathbb{P}(X_\tau < 0) > 0$. Hence, $\mathbb{P}(v^*(X_\tau) > G(X_\tau)) > 0$ and it follows that $\mathbb{E}[G(X_\tau)] < \mathbb{E}[v^*(X_\tau)] \leq v^*(x)$ since $v^*(X_t)$ is a supermartingale, leading to a contradiction with the fact that τ is optimal. □

Example 2.2.1. *Let X be a Lévy process such that $-X$ is given by the Cramér-Lundberg model with exponential jumps, that is, $X_t = -dt + \sum_{n=1}^{N_t} Z_n$, where $d > 0$, N is a Poisson process with parameter $\lambda > 0$, and $(Z_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of random variables independent of N with $Z_1 \sim Exp(\alpha)$ for some $\alpha > 0$. From [4, Chapter 4.2], it follows that $\mathbb{P}(\bar{X}_\infty = 0) = 1 - \frac{\lambda}{\alpha d}$ and, for $(0,\infty)$, the density of $\bar{X}_\infty$ is $f(x) = \frac{\lambda}{\alpha d}(\alpha - \frac{\lambda}{d})e^{-(\alpha-\frac{\lambda}{d})x}$. Therefore, $\psi(2) < 0$ corresponds to $\alpha > 2 + \frac{\lambda}{d}$. When this condition is fulfilled, we find that, for $\beta = \alpha - \frac{\lambda}{d}$ then $\mathbb{E}[e^{\bar{X}_\infty}] = \frac{\alpha}{\lambda d}\left(\frac{2\beta-1}{\beta-1}\right)$ and $\mathbb{E}[e^{2\bar{X}_\infty}] = \frac{\alpha}{\lambda d}\left(\frac{2\beta-2}{\beta-2}\right)$. From Theorem 2.2.1 we find that the optimal stopping point is $y_c = \log(4c(\beta-1)^2/((\beta-2)(2\beta-1)))$. By Remark 2.2.1, it follows that $\tau_{y_c}^+ = \tau_{y_c}^{++}$.*

Example 2.2.2. *Let X be a compound Poisson process given by $X_t = \sum_{n=1}^{N_t} Z_n$, where N is a Poisson process with parameter $\lambda > 0$, and $(Z_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of random variables which is independent of N and for which $\mathbb{P}(Z_1 = \alpha) = 1 - \mathbb{P}(Z_1 = -\alpha) = p$ for some $p \in (0,1)$. When $p < \frac{1}{2}$, the distribution of $\frac{1}{\alpha}\bar{X}_\infty$ is geometric with $\mathbb{P}(\bar{X}_\infty \geq k\alpha) = (p/(1-p))^k$ and, when $p \geq 1/2$, then $\bar{X}_\infty = \infty$ a.s. Therefore, assume that $p < 1/2$, and we then obtain $\mathbb{E}[e^{\bar{X}_\infty}] = \frac{1-2p}{1-p-e^\alpha p}$ and $\mathbb{E}[e^{2\bar{X}_\infty}] = \frac{1-2p}{1-p-e^{2\alpha}p}$. By Theorem 2.2.1, it follows that the optimal stopping point is $y_c = \log(2c(1-p-e^\alpha p)/(1-p-e^{2\alpha}p))$. Both $\tau_{y_c}^+$ and $\tau_{y_c}^{++}$ are solutions.*

2.3 The Variance Optimal Stopping Problem

In this section we solve the variance problem (2.1) for those processes where the method of embedding can be applied. Recall that if there exists some stopping time τ^* and some constant c such that both (2.2) and (2.3) are fulfilled, then τ^* solves the variance problem. Indeed, for all τ, we have

$$\mathbb{V}[e^{X_\tau}] \leq \mathbb{E}[(e^{X_{\tau^*}} - c)^2] - (\mathbb{E}[e^{X_\tau}] - c)^2 \tag{2.8}$$

$$= \mathbb{V}[e^{X_{\tau^*}}] - (\mathbb{E}[e^{X_\tau}] - \mathbb{E}[e^{X_{\tau^*}}])^2. \tag{2.9}$$

The existence of a combination of τ^* and c that fulfills both of the two requirements is not certain. We show that it depends on whether at least one of the following equations has a solution:

$$\frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^y = \frac{\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty \geq y)}]}{\mathbb{E}[e^{\bar{X}_\infty}]}, \tag{2.10}$$

$$\frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^y = \frac{\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty > y)}]}{\mathbb{E}[e^{\bar{X}_\infty}]}. \tag{2.11}$$

We call these the embedding equations.

Theorem 2.3.1. *Let X be a Lévy process with $\psi(2) < 0$. If the embedding equation (2.10) has a solution $\hat{y}$ then $\tau^+_{\hat{y}}$ is an optimal stopping time for variance problem (2.1). If embedding equation (2.11) has a solution $\hat{y}$, then $\tau^{++}_{\hat{y}}$ is an optimal stopping time for variance problem (2.1).*

a) Assume that 0 is regular for $(0,\infty)$. Then the embedding equations coincide and there exists a solution $\hat{y}$. If, additionally, X is spectrally negative then

$$\hat{y} = \tfrac{1}{\phi(0)} \log\left(2\tfrac{\phi(0)-1}{\phi(0)-2}\right).$$

b) Assume that 0 is irregular for $(0,\infty)$. Then if

$$\mathbb{E}[e^{\bar{X}_\infty}]^2 > 2\mathbb{E}[e^{2\bar{X}_\infty}]\,\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty > 0)}], \tag{2.12}$$

the embedding equations have no solutions. Assume, additionally, that X is not a compound Poisson process and that (2.12) is not satisfied, then at least one of the embedding equations have a solution.

c) Assume that $\psi(2) = 0$. Then, for all $\tau \in \mathcal{T}$, we have $\mathbb{V}[e^{X_\tau}] < 1$, but $\mathbb{V}[e^{X_t}] \to 1$ as $t \to \infty$.

Proof. From Theorem 2.2.1, both $\tau^+_{\hat{y}}$ and $\tau^{++}_{\hat{y}}$ are solutions to the quadratic problem with parameter $c = \frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^{\hat{y}}$. Thus, the left-hand side of the embedding equations give the parameter value of c needed for $\tau^+_{\hat{y}}$ and $\tau^{++}_{\hat{y}}$ to solve the quadratic problem. From (2.4) we deduce that the right-hand side of the embedding equations give respectively the values $\mathbb{E}[e^{X_{\tau^+_{\hat{y}}}}]$ and $\mathbb{E}[e^{X_{\tau^{++}_{\hat{y}}}}]$. This proves that, when $\hat{y}$ solves as least one of the embedding equations, then $\tau^+_{\hat{y}}$ and $\tau^{++}_{\hat{y}}$ respectively solve the variance problem.

Next, we investigate the existence of a solution to embedding equation (2.10). First, note that $\mathbb{E}[e^{\bar{X}_\infty}] \leq \mathbb{E}[e^{2\bar{X}_\infty}]$ and, thus, for $y = 0$, the left-hand side is less than or equal to $\frac{1}{2}$, whereas the right-hand side is 1. For $y \to \infty$, the left-hand side increases continuously to ∞, whereas the right-hand side converges to 0.

a) As 0 is regular for $(0, \infty)$, then, by Lemma 2.2.2, the distribution of $\bar{X}_\infty$ is continuous. Hence, the right-hand side of (2.10) is also continuous and, hence, the embedding equation (2.10) has a solution. As the distribution of $\bar{X}_\infty$ is continuous, the embedding equations coincides and, thus, they both have a solution. In the special case where X is a spectrally negative Lévy process, we recall that $\bar{X}_\infty$ is exponentially distributed with parameter $\phi(0)$ and, hence, the result is straight forward.

b) The left-hand side of the embedding equations are equal. As a function of y, the right-hand side of (2.10) is left continuous, and the right-hand side of (2.11) is the right continuous version. Thus, if the right-hand side of (2.11) is smaller than the left-hand side of (2.11) for $y = 0$ then the embedding equations have no solutions. This corresponds to (2.12). The right-hand sides of (2.10) and (2.11) have discontinuities only at points where the distribution of $\bar{X}_\infty$ has discontinuities. If X is not a compound Poisson process then 0 is the only discontinuity point for the distribution of $\bar{X}_\infty$ and, thus, if (2.12) does not hold then the embedding equations must have a solution.

c) Consider a $\tau \in \mathcal{T}$. If $\tau = \infty$ a.s., then $\mathbb{V}[e^{X_\tau}] = 0$. If $\mathbb{P}(\tau < \infty) > 0$ then $\mathbb{E}[e^{X_\tau}] > 0$ and, thus,

$$\mathbb{V}[e^{X_\tau}] < \mathbb{E}[e^{2X_\tau}] \leq \mathbb{E}[e^{2X_\tau} 1_{(\tau<\infty)}] \leq \lim_{t\to\infty} \mathbb{E}[e^{2X_{\tau\wedge t}}] = 1.$$

Next, recall that $\psi(2) = 0$ implies that e^{2X_t} is a martingale and recall that X is not deterministic. Thus, $\mathbb{V}[e^{X_1}] > 0$ and we obtain $\psi(1) = \mathbb{E}[e^{X_1}] < 1$. Therefore, $\mathbb{V}[e^{X_t}] = (e^{\psi(2)t} - e^{\psi(1)t}) \to 1$ as $t \to \infty$. □

Remark 2.3.1. *Let $Y_t = X_t + x$ be a Lévy process starting at x. Then, for any stopping time τ,*

$$\mathbb{V}[e^{Y_\tau}] = \mathbb{V}[e^{X_\tau + x}] = e^{2x}\mathbb{V}[e^{X_\tau}].$$

Therefore, a stopping time is optimal for the variance problem for process Y, if it is optimal for the variance problem for process X. Whenever the variance problem for X is solved by a hitting time, then so is the variance problem for Y. However, the stopping region will be shifted by the starting value x.

Remark 2.3.2. *When X is a spectrally negative process, it is usually possible to calculate $\phi(0)$ and, therefore, we may determine the optimal stopping time for the quadratic problem given in Theorem 2.2.1. Furthermore, as the* *distribution of* $\bar{X}_\infty$ *is Exp*$(\phi(0))$, *we can solve the embedding equations and find a solution in order to determine the optimal stopping time for the variance given in Theorem 2.3.1. However, if* X *is not spectrally negative, we need to determine* $\mathbb{E}[e^{\bar{X}_\infty}]$ *and* $\mathbb{E}[e^{2\bar{X}_\infty}]$ *in order to apply Theorem 2.2.1, and to solve the embedding equations in order to apply Theorem 2.3.1. These are, in most cases, impossible to calculate as the distribution of* $\bar{X}_\infty$ *is often not known.*

Example 2.3.1. *Consider the negative Cramér-Lundberg Lévy process given in Example 2.2.1. Again, assume that* $\alpha > 2 + \frac{\lambda}{d}$ *and recall the distribution of* $\bar{X}_\infty$ *given in Example 2.2.1. For* $y > 0$, *we compute* $\mathbb{E}[e^{\bar{X}_\infty} 1_{(\bar{X}_\infty \geq y)}] = \frac{\alpha}{\lambda d}\left(\frac{\beta}{\beta-1} e^{-(\beta-1)y}\right)$ *and write embedding equation* (2.10) *as*

$$e^{-\beta y} = \frac{(2\beta-1)^2(\beta-2)}{4\beta(\beta-1)^2}. \tag{2.13}$$

This equation has a solution, and it follows from Theorem 2.3.1 that the variance problem has solutions $\tau_{\hat{y}}^{+}$ *and* $\tau_{\hat{y}}^{++}$ *with* $\hat{y}$ *solving* (2.13). *For example, if* $\alpha = 3$, $\lambda = 2$ *and* $d = 4$ *then* $\hat{y} = (2/5)\log(45/16) \approx 0.413630$, *see Figure 2.1.*

Example 2.3.2. *Consider the compound Poisson process given in Example 2.2.2 with* $p = 1/11$ *and* $\alpha = \log(2)$. *Recall from Section 2.2 that* $\bar{X}_\infty$ *has a geometric distribution with* $\mathbb{P}(\bar{X}_\infty \geq k \cdot \log(2)) = (\frac{1}{10})^k$. *Thus, the left-hand side of* (2.12) *equals* $27/64$, *whereas the right-hand side equals* $9/40$. *As* $9/40 < 27/64$, *it follows that the embedding equations have no solution and, therefore, Theorem 2.3.1 does not give a solution for the variance problem for this process. See Figure 2.1.*

2.4 Randomized Stopping

Theorem 2.3.1 only offers a solution to the variance problem if one of the embedding equations has a solution. In this section we introduce randomized stopping to overcome this problem. When the embedding equations have no solution an immediate complication arises because it is not easy to find a constant c and stopping time τ^* such that both (2.2) and (2.3) are solved, and so the embedding method cannot be applied. By taking the supremum over a wider class of stopping times we overcome this problem.

We introduce randomized stopping times for an optimal stopping problem by expanding the filtration without removing the Markov property of the process. We create this expansion by introducing a random variable U, which is defined on the same probability space as X, and which is uniformly distributed on $[0,1]$ and independent of X. We note that this may require that we augment the probability space on which X is defined. We then define the new filtration $(\hat{\mathcal{F}}_t)_{t\geq 0}$ by $\hat{\mathcal{F}}_t = \sigma(U) \vee \mathcal{F}_t$. We let $\hat{\mathcal{T}}$ be the set of stopping times with respect to the new filtration $\hat{\mathcal{F}}_t$, and we refer to these as *randomized stopping times.* Randomized stopping times are discussed in a discrete-time setup in [9], where it was shown that the value function of a classical optimal stopping problem does not change when randomized stopping times are introduced. This result carries over to the quadratic problem, but it does not carry over to the variance problem. The verification theorem in the proof of Theorem 2.2.1 is based on the optional sampling theorem, and, hence, it also holds for a classical optimal stopping problem with randomized stopping. This is due to the fact that the Lévy process remains a Markov process with the augmented filtration. In particular, the stopping times of Theorem 2.2.1 also solve the quadratic problem with randomized stopping.

Theorem 2.4.1. *Let X be a Lévy process with $\psi(2) < 0$, and consider the quadratic optimal stopping problem with randomized stopping. Let U be a random variable uniformly distributed on $[0,1]$ and independent of X. Let $y_c = \log(2c\mathbb{E}[e^{2\bar{X}_\infty}]/\mathbb{E}[e^{\bar{X}_\infty}])$, $p \in [0,1]$ and $Y = 1_{(U<p)}$. Then $\tau^* = Y\tau_{y_c}^{+} + (1-Y)\tau_{y_c}^{++}$ is also an optimal stopping time.*

Proof. The result follows from the fact that τ^* is a stopping time with respect to the augmented filtration, and the expected gain at τ^* equals the expected gain at $\tau_{y_c}^{+}$ and $\tau_{y_c}^{++}$. □

By Remark 2.2.1, we see that, for some Lévy processes, $\tau_y^{+} = \tau_y^{++}$ for all $y \geq 0$ and the stopping times τ^* in Theorem 2.4.1 do not introduce a new class of optimal stopping times for the quadratic problem. However, if 0 is irregular for $(0,\infty)$ then, for some discontinuity points y of the distribution of $\bar{X}_\infty$, we have that $\tau_y^{+} < \tau_y^{++}$. In this case the introduction of randomized stopping times creates a new family of optimal stopping times for the quadratic problem such that $\tau_{y_c}^{+} \leq \tau^* \leq \tau_{y_c}^{++}$. With the wider family of stopping times we can solve the variance problem in all cases.

Theorem 2.4.2. *Let X be a Lévy process with $\psi(2) < 0$.*

a) If at least one of embedding equations (2.10) *and* (2.11) *has a solution, then the optimal stopping times given in Theorem 2.3.1 are also optimal for the variance problem with randomized stopping.*

b) If embedding equations (2.10) *and* (2.11) *have no solutions, let*

$$\hat{y} \equiv \inf\{y \geq 0 : \tfrac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^y > \tfrac{\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty \geq y)}]}{\mathbb{E}[e^{\bar{X}_\infty}]}\}. \tag{2.14}$$

Let U be uniformly distributed on $[0,1]$ and independent of X, and let $Y = 1_{(U<p)}$, where

$$p = \frac{\frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^{\hat{y}} - \mathbb{E}[e^{X_{\tau_{\hat{y}}^{++}}}]}{\mathbb{E}[e^{X_{\tau_{\hat{y}}^{+}}}] - \mathbb{E}[e^{X_{\tau_{\hat{y}}^{++}}}]}. \tag{2.15}$$

Then $\tau^* = Y\tau_{\hat{y}}^+ + (1-Y)\tau_{\hat{y}}^{++}$ is optimal for the variance problem with randomized stopping. Moreover, the distribution of $\bar{X}_\infty$ has a discontinuity in $\hat{y}$, and if 0 is irregular for $(0,\infty)$ and X is not compound Poisson process, then $\hat{y} = 0$.

Proof. a) The optimal stopping times given in Theorem 2.3.1 solve the corresponding problems with randomized stopping. This result follows in the same way as in the proof of Theorem 2.3.1.

b) From the proof of Theorem 2.3.1 b), it follows that if the embedding equations (2.10) and (2.11) have no solution, then the distribution of $\bar{X}_\infty$ has a discontinuity in the value $\hat{y}$, defined in (2.14). The inequality in (2.14) holds for every $y > \hat{y}$. Thus, from (2.7), it follows that

$$\tfrac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^y > \mathbb{E}[e^{X_{\tau_y^{++}}}]$$

for every $y > \hat{y}$. As both sides are right continuous in y and as the embedding equations have no solution, then the same holds for $y = \hat{y}$. On the other hand, the inequality in (2.14) does not hold for any $y < \hat{y}$. As $\mathbb{E}[e^{X_{\tau_{\hat{y}}^{+}}}]$ is the left continuous version of $\mathbb{E}[e^{X_{\tau_{\hat{y}}^{++}}}]$ then

$$\tfrac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^{\hat{y}} \leq \mathbb{E}[e^{X_{\tau_{\hat{y}}^{+}}}].$$

Hence $p \in [0,1]$ and it follows from Theorem 2.4.1 that τ^* is an optimal stopping time for the quadratic problem with parameter $c = \frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^{\hat{y}}$. Then we need only to prove that $\mathbb{E}[e^{X_{\tau^*}}] = c$:

$$\begin{aligned}
\mathbb{E}[e^{X_{\tau^*}}] &= p\mathbb{E}\left[e^{X_{\tau_{\hat{y}}^{+}}}\right] + (1-p)\mathbb{E}\left[e^{X_{\tau_{\hat{y}}^{++}}}\right] \\
&= \left(\frac{\frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^{\hat{y}} - \mathbb{E}[e^{X_{\tau_{\hat{y}}^{++}}}]}{\mathbb{E}[e^{X_{\tau_{\hat{y}}^{+}}}] - \mathbb{E}[e^{X_{\tau_{\hat{y}}^{++}}}]}\right)\mathbb{E}\left[e^{X_{\tau_{\hat{y}}^{+}}}\right] + \left(\frac{\mathbb{E}[e^{X_{\tau_{\hat{y}}^{+}}}] - \frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^{\hat{y}}}{\mathbb{E}[e^{X_{\tau_{\hat{y}}^{+}}}] - \mathbb{E}[e^{X_{\tau_{\hat{y}}^{++}}}]}\right)\mathbb{E}\left[e^{X_{\tau_{\hat{y}}^{++}}}\right] \\
&= \frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]}e^{\hat{y}} = c.
\end{aligned}$$

If the Lévy process is not a compound Poisson process then, by Lemma 2.2.2, the distribution of $\bar{X}_\infty$ has a discontinuity point only at 0. Therefore, $\hat{y} = 0$ and the optimal stopping time is a combination of $\tau_0^+ = 0$ and τ_0^{++}. □

Example 2.4.1. *Let X denote the compound Poisson process considered in Example 2.3.2 with $p = 1/11$ and $\alpha = \log(2)$. Recall that we cannot solve the variance problem for this process by Theorem 2.3.1. From* (2.15) *we calculate $p = 7/32$. Let Y be a random variable independent of X and with $\mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = p$. By Theorem 2.4.2, it follows that $\tau^* = Y\tau_0^+ + (1 - Y)\tau_0^{++} = (1 - Y)\tau_0^{++}$ is an optimal stopping time for the variance problem with randomized stopping. See Figure 2.2. The variance for this stopping time is*

$$\mathbb{V}[e^{X_{\tau^*}}] = p\mathbb{E}[e^{2X_{\tau_0^+}}]+(1-p)\mathbb{E}[e^{2X_{\tau_0^{++}}}]-(p\mathbb{E}[e^{X_{\tau_0^+}}]+(1-p)\mathbb{E}[e^{X_{\tau_0^{++}}}])^2 = 0.390625.$$

We determine the best stopping time of the form τ_y^+ and τ_y^{++} for $y \geq 0$. It is sufficient to consider stopping times of the form $\tau^+_{k\log(2)}$ for $k \in \mathbb{N}$. For these, $\mathbb{V}[e^{X_{\tau^+_{k\log(2)}}}] = 0.4^k - 0.04^k$. Inspection of this function reveals that the maximal value is obtained for $k = 1$ with $\mathbb{V}[e^{X_{\tau^+_{\log(2)}}}] = 0.36$. This is a smaller variance than the one we got from our solution τ^* from Theorem 2.4.2. However, the stopping time τ^* is not a stopping time with respect to the filtration generated from X.

2.5 Variance Problem Without Randomized Stopping Revisited

As demonstrated in Example 2, the embedding equations do not always have a solution. In this section we want to solve the variance problem without randomized stopping when the embedding equations have no solution. At first, one may hope that some new approach reveals an excess boundary solution for the variance problem when the embedding equations have no solution. However, Theorem 2.5.1 below reveals this is not possible.

Theorem 2.5.1. *Let X be a Lévy process with $\psi(2) < 0$. Assume that the embedding equations have no solutions. Then the variance problem with randomized stopping does not have an optimal stopping time of the form τ_y^+ or τ_y^{++} for any $y \in \mathbb{R}$.*

Proof. Let τ^* be the optimal stopping time for the variance problem given in Theorem 2.4.2, and let $\hat{y}$ be given by (2.14). It follows from (2.9) that for some τ to solve the variance problem, we need $\mathbb{E}[e^{X_\tau}] = \mathbb{E}[e^{X_{\tau^*}}]$. However, from the proof of Theorem 2.4.2, it follows that $\mathbb{E}[e^{X_{\tau^*}}] = \frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]} e^{\hat{y}}$ and

$$\mathbb{E}[e^{X_{\tau_{\hat{y}}^+}}] > \frac{\mathbb{E}[e^{\bar{X}_\infty}]}{2\mathbb{E}[e^{2\bar{X}_\infty}]} e^{\hat{y}} > \mathbb{E}[e^{X_{\tau_{\hat{y}}^{++}}}].$$

The inequalities are strict because the embedding equations have no solution. From (2.4) and (2.7), it follows that $\mathbb{E}[e^{X_{\tau^*}}] \neq \mathbb{E}[e^{X_{\tau_y^+}}]$ and $\mathbb{E}[e^{X_{\tau^*}}] \neq \mathbb{E}[e^{X_{\tau_y^{++}}}]$ for all y. Thus, there will not be any solutions of the form τ_y^+ or τ_y^{++} for $y \in \mathbb{R}$. □

2.5.1 Compound Poisson Processes

Let X be a compound Poisson process with jump intensity λ. That is, the Poisson process N that counts the number of jumps has parameter λ. Assume the embedding equations have no solution and consider the optimal stopping time of the variance problem with randomized stopping, given in Theorem 2.4.2. From Theorem 2.5.2 below, it follows that one may mimic τ^* by a stopping time in $\mathcal{T}$. The idea is based on the observation that $\mathcal{F}$ contains information about the process N, and N is independent of both $X_{\tau_{\hat{y}}^+}$ and $X_{\tau_{\hat{y}}^{++}}$. Hence, we may use the process N to choose between the two stopping times τ_y^+ and τ_y^{++} instead of using the random variable Y.

Theorem 2.5.2. *Let X be a compound Poisson process with $\psi(2) < 0$ and jump intensity $\lambda > 0$. Let T be the first jump time of X. Assume that the embedding equations have no solution, and let $\hat{y}$ and p be given as in Theorem 2.4.2. Then*

$$\tilde{\tau}^* = \tilde{Y}\tilde{\tau}_{\hat{y}}^+ + (1 - \tilde{Y})\tau_{\hat{y}}^{++}$$

is an optimal stopping time for the variance problem without randomized stopping where $\tilde{\tau}_{\hat{y}}^+ = t_0 \vee \tau_{\hat{y}}^+$, $t_0 = \frac{-1}{\lambda}\log(p)$ and $\tilde{Y} = 1_{(t_0<T)}$.

Proof. First note that τ^* is a stopping time of $\mathcal{T}$. For every $t < t_0$ we have that

$$\begin{aligned}\{\tilde{\tau}^* \leq t\} &= \{\tilde{\tau}_{\hat{y}}^+ \leq t,\ \tilde{Y} = 1\} \cup \{\tau_{\hat{y}}^{++} \leq t,\ \tilde{Y} = 0\} \\ &= \{\tau_{\hat{y}}^{++} \leq t,\ \tilde{Y} = 0\} = \{\tau_{\hat{y}}^{++} \leq t\} \in \mathcal{F}_t,\end{aligned}$$

and, for $t \geq t_0$

$$\{\tilde{\tau}^* \leq t\} = (\{\tau_{\hat{y}}^+ \vee t_0 \leq t\} \cap \{\tilde{Y} = 1\}) \cup (\{\tau_{\hat{y}}^{++} \leq t\} \cap \{\tilde{Y} = 0\}) \in \mathcal{F}_t.$$

Thus, $\tau^* \in \mathcal{T}$.

Recall that, on the event $\{\tilde{Y}=1\}$, $X_{\tilde{\tau}^+_{\hat{y}}}$ and $X_{\tau^+_{\hat{y}}}$ have the same distribution. Let τ^* be the optimal randomized stopping time given in part $b)$ of Theorem 2.4.2. Then, for $\beta = 1, 2$ we have

$$\begin{aligned}
\mathbb{E}[e^{\beta X_{\tilde{\tau}^*}}] &= p \cdot \mathbb{E}[e^{\beta X_{\tilde{\tau}^+_{\hat{y}}}} | \tilde{Y}=1] + (1-p) \cdot \mathbb{E}[e^{\beta X_{\tau^{++}_{\hat{y}}}} | \tilde{Y}=0] \\
&= p \cdot \mathbb{E}[e^{\beta X_{\tau^+_{\hat{y}}}} | \tilde{Y}=1] + (1-p) \cdot \mathbb{E}[e^{\beta X_{\tau^{++}_{\hat{y}}}} | \tilde{Y}=0] \\
&= p \cdot \mathbb{E}[e^{\beta X_{\tau^+_{\hat{y}}}}] + (1-p) \cdot \mathbb{E}[e^{\beta X_{\tau^{++}_{\hat{y}}}}] \\
&= \mathbb{E}[e^{\beta X_{\tau^*}}].
\end{aligned}$$

Hence, we see that $\mathbb{V}[e^{X_{\tilde{\tau}^*}}] = \mathbb{V}[e^{X_{\tau^*}}]$ and we conclude that $\tilde{\tau}^*$ is an optimal stopping time of the variance problem without randomized stopping. □

Example 2.5.1. *Let X denote the compound Poisson process considered in Example 2.3.2 with $p = 1/11$ and $\alpha = \log(2)$. Recall that, for this process, we cannot use Theorem 2.3.1 to solve the variance problem. In Theorem 2.4.2, we obtain a randomized solution, τ^*, and we have seen that no stopping time of the form τ^+_y or τ^{++}_y gives as high a variance as τ^*. However, by use of Theorem 2.5.2, we may actually find a stopping time of $\mathcal{T}$, giving the same variance as the randomized solution τ^*. Let $\tilde{Y} = 1_{\{N_{t_0}=0\}}$ and $t_0 = -\log(25/32)$, and define $\tilde{\tau}^* = \tilde{Y} t_0 + (1-\tilde{Y})\tau^{++}_0$. Then it follows that $\tilde{\tau}^* \in \mathcal{T}$ and $\tilde{\tau}^*$ solve the variance problem without randomized stopping with as high a variance as that obtained by τ^*.*

2.5.2 Lévy Processes which are not Compound Poisson Processes

In this section we consider Lévy processes of bounded variation with $\psi(2) < 0$ which are not compound Poisson processes.

Theorem 2.5.3. *Let X be a Lévy process of bounded variation with $\psi(2) < 0$ which is not a compound Poisson process, and let d denote the drift of X. Assume that the embedding equations have no solution. Let $\tau^* \in \hat{\mathcal{T}}$ be the randomized optimal stopping time given in Theorem 2.4.2. Then $\mathbb{V}[e^{X_\tau}] < \mathbb{V}[e^{X_{\tau^*}}]$ for all $\tau \in \mathcal{T}$.*

a) If $X_t - dt$ is not a compound Poisson process then $\sup_{\tau\in\mathcal{T}} \mathbb{V}[e^{X_\tau}] = \mathbb{V}[e^{X_{\tau^*}}]$.

b) If $X_t - dt$ is a compound Poisson process then $\sup_{\tau\in\mathcal{T}} \mathbb{V}[e^{X_\tau}] < \mathbb{V}[e^{X_{\tau^*}}]$.

Proof. For $\tau = 0$, we have the inequality $\mathbb{V}[e^{X_\tau}] < \mathbb{V}[e^{X_{\tau^*}}]$, and, by Blumenthal's 0-1 law, it is enough to consider $\tau > 0$ in $\mathcal{T}$. Assume that $\mathbb{V}[e^{X_\tau}] = \mathbb{V}[e^{X_{\tau^*}}]$. It follows from Theorem 2.4.2 that $\hat{y} = 0$, and from (2.8) and (2.9) that τ solves the quadratic problem with parameter $\mathbb{E}[e^{X_{\tau^*}}]$ and $\mathbb{E}[e^{X_\tau}] = \mathbb{E}[e^{X_{\tau^*}}]$. By Lemma 2.2.3 we obtain $\tau \geq \tau_0^{++}$ and, hence, $\mathbb{E}[e^{X_\tau}] \leq \mathbb{E}[e^{X_{\tau_0^{++}}}]$, as e^{X_t} is a supermartingale. However, from the proof of Theorem 2.5.1, it follows that $\mathbb{E}[e^{X_{\tau_0^{++}}}] < \mathbb{E}[e^{X_{\tau^*}}]$. Therefore, $\mathbb{E}[e^{X_\tau}] < \mathbb{E}[e^{X_{\tau^*}}]$ and we cannot have $\mathbb{V}[e^{X_\tau}] = \mathbb{V}[e^{X_{\tau^*}}]$.

a) Let α_t^q be the q-fractile of X_t. That is, $\alpha_t^q = \inf\{\alpha \in \mathbb{R} : \mathbb{P}(X_t \leq \alpha) > q\}$. It follows from [8, Theorem 27.4] that the distribution of X_t is continuous and, hence, for every $t > 0$ and $q \in (0, 1)$, we have $\mathbb{P}(X_t \leq \alpha_t^q) = \mathbb{P}(X_t < \alpha_t^q) = q$.

Next, we choose a sequence of stopping times, $(\tau_n)_{n\in\mathbb{N}} \subset \mathcal{T}$, in the following way:

$$\tau_n = \inf\{t > 1/n | X_t > 0\} \cdot 1_{(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))} + \frac{1}{n} 1_{(X_{1/n} \notin (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}.$$

That is, if the process at time $1/n$ is not between the fractiles $\alpha_{1/n}^{(1-p)/2}$ and $\alpha_{1/n}^{(1+p)/2}$, then the process is stopped. On the other hand, if the process at time $1/n$ is between the two fractiles, then the process is stopped at the first time after $1/n$ when it gets above 0. We show that the variance at the stopping time τ_n approximates the variance at the randomized stopping time given in Theorem 2.4.2.

It holds that

$$\mathbb{E}[e^{X_{\tau_n}} 1_{(X_{1/n} \notin (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}] = \mathbb{E}[e^{X_{1/n}}] - \mathbb{E}[e^{X_{1/n}} 1_{(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}]$$

is bounded from below by

$$\mathbb{E}[e^{X_{1/n}}] - e^{\alpha_{1/n}^{(1+p)/2}} \mathbb{P}(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2})) = \mathbb{E}[e^{X_{1/n}}] - e^{\alpha_{1/n}^{(1+p)/2}} p,$$

and from above by

$$\mathbb{E}[e^{X_{1/n}}] - e^{\alpha_{1/n}^{(1-p)/2}} \mathbb{P}(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2})) = \mathbb{E}[e^{X_{1/n}}] - e^{\alpha_{1/n}^{(1-p)/2}} p.$$

As $\mathbb{E}[e^{X_{1/n}}]$ converges to 1 as n tends to ∞ and as X_t converges to 0 in probability as t converges to 0, it follows that both $\alpha_{1/n}^{(1-p)/2}$ and $\alpha_{1/n}^{(1+p)/2}$ converge to 0 as n tends to infinity. Hence, $\mathbb{E}[e^{X_{\tau_n}} 1_{(X_{1/n} \notin (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}]$ converges to $1 - p$ as n tends to ∞.

Next, let $H(x) = \mathbb{E}_x[e^{X_{\tau_0^{++}}}]$. Then

$$\mathbb{E}[e^{X_{\tau_n}} 1_{(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}] = \mathbb{E}[H(X_{1/n}) 1_{(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}]. \quad (2.16)$$

We claim that, for every $q \in (0,1)$, there exists some $t^q > 0$ such that, for $t < t^q$, we have $\alpha_t^q \leq 0$. We prove the statement by contradiction. Thus, assume there is a $q \in (0,1)$ and a sequence $t_n \to 0$ such that, for every n, $\alpha_{t_n}^q > 0$. Then

$$\mathbb{P}(\limsup_{n\to\infty} X_{t_n} > 0) = \lim_{n\to\infty} \mathbb{P}(\bigcup_{k=n}^{\infty} \{X_{t_k} > 0\}) \geq \lim_{n\to\infty} (1-q) = 1-q.$$

However, this is in contradiction with 0 being irregular for $(0,\infty)$. As H is nondecreasing, we conclude that (2.16) is bounded from below by

$$\mathbb{E}[H(\alpha_{1/n}^{(1-p)/2}) 1_{(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}] = H(\alpha_{1/n}^{(1-p)/2})p,$$

and from above by

$$\mathbb{E}[H(\alpha_{1/n}^{(1+p)/2}) 1_{(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}] = H(\alpha_{1/n}^{(1+p)/2})p.$$

Both terms converge to $H(0)p$ as n tends to ∞. This is because it from (2.4) follows that H is continuous and increasing on $(-\infty, 0]$ and because there exists some $t^q > 0$ such that for $t < t^q$ we have $\alpha_t^q \leq 0$. Therefore, $\mathbb{E}[e^{X_{\tau_n}} 1_{(X_{1/n} \in (\alpha_{1/n}^{(1-p)/2}, \alpha_{1/n}^{(1+p)/2}))}]$ converges to $H(0)p$ as n tends to ∞. Taken together, these results imply that, when n tends to ∞,

$$\mathbb{E}[e^{X_{\tau_n}}] \to (1-p) + H(0)p = (1-p)\mathbb{E}[e^{X_{\tau_0^+}}] + p\mathbb{E}[e^{X_{\tau_0^{++}}}] = \mathbb{E}[e^{X_{\tau^*}}].$$

Similarly, it follows that $\mathbb{E}[e^{2X_{\tau_n}}] \to \mathbb{E}[e^{2X_{\tau^*}}]$, and hence, $\mathbb{V}[e^{X_{\tau_n}}] \to \mathbb{V}[e^{X_{\tau^*}}]$.

b) We prove the statement by contradiction. Assume that there exists a sequence, $(\tau_n)_{n\in\mathbb{N}} \subset \mathcal{T}$ such that $\lim_{n\to\infty} \mathbb{V}[e^{X_{\tau_n}}] = \mathbb{V}[e^{X_{\tau^*}}]$. The proof is quite lengthy and, therefore, it is broken into parts to clarify the structure.

Part i) It follows from (2.9) that $\lim_{n\to\infty} \mathbb{E}[e^{X_{\tau_n}}] = \mathbb{E}[e^{X_{\tau^*}}] = c$. Let $\tau \in \mathcal{T}$ and $y < 0$ be given. Then

$$\begin{aligned}\mathbb{E}[e^{X_\tau}] &= \mathbb{E}[e^{X_\tau} 1_{(\tau < \tau_0^{++})} 1_{(X_\tau > y)}] + \mathbb{E}[e^{X_\tau} 1_{(\tau < \tau_0^{++})} 1_{(X_\tau \leq y)}] + \mathbb{E}[e^{X_\tau} 1_{(\tau \geq \tau_0^{++})}] \\ &\leq \mathbb{P}(\tau < \tau_0^{++}, X_\tau > y) + e^y + \mathbb{E}[e^{X_\tau} 1_{(\tau \geq \tau_0^{++})}] \\ &\leq \mathbb{P}(\tau < \tau_0^{++}, X_\tau > y) + e^y + \mathbb{E}[e^{X_{\tau_0^{++}}}].\end{aligned}$$

Thus, to obtain $\lim_{n\to\infty}\mathbb{E}[e^{X_{\tau_n}}]=c$, then, for any $\varepsilon>0$ and $y<0$, we need, for large enough n, $\mathbb{P}(\tau_n<\tau_0^{++},X_{\tau_n}>y)+e^y-(c-\mathbb{E}[e^{X_{\tau_0^{++}}}])\geq-\varepsilon$. As X is not a compound Poisson process, it follows from Theorem 2.4.2 that $\hat{y}=0$. Thus, $\mathbb{E}[e^{X_{\tau_0^{++}}}]<c$, and we may choose ε and y small enough that $(c-\mathbb{E}[e^{X_{\tau_0^{++}}}])-\varepsilon-e^y>0$. We conclude that there exists $p>0$, $y<0$, and $N\in\mathbb{N}$ such that for all $n>N$,

$$\mathbb{P}(\tau_n<\tau_0^{++},X_{\tau_n}>y)\geq p. \tag{2.17}$$

Part ii) Let T denote the first jump time of the process. Assume that $\mathbb{P}(\tau<T)>0$ for some $\tau\in\mathcal{T}$. With t^* as in Lemma 2.5.4 a) below, it follows that

$$\begin{aligned}\mathbb{E}[e^{X_\tau}]&=\mathbb{E}[e^{X_\tau}1_{(T\leq t^*)}]+\mathbb{E}[e^{X_\tau}1_{(T>t^*)}]\\&=\mathbb{E}[e^{X_\tau}1_{(T\leq t^*)}]+e^{-dt^*}\mathbb{P}(T>t^*)\\&\geq e^{-(d+\lambda)t^*}.\end{aligned}$$

Now, for each n, where $\mathbb{P}(\tau_n<T)>0$, let t_n^* be as in Lemma 2.5.4. If there are infinitely many such n then to have $\lim_{n\to\infty}\mathbb{E}[e^{X_{\tau_n}}]=c$, it must hold that, for any $\varepsilon>0$ then for large enough n, $\mathbb{E}[e^{X_{\tau_n}}]<c+\varepsilon$ and, thus, $e^{-(d+\lambda)t_n^*}<c+\varepsilon$. Hence, we need, from a certain step, $t_n^*\geq\frac{-\log(c+\varepsilon)}{d+\lambda}$. By Lemma 2.5.4 b), it follows that if $\mathbb{P}(\tau_n<T)>0$ then $\mathbb{P}(\tau_n<T,\tau_n<t_n^*)=0$. Therefore, for large enough n it must hold that

$$\begin{aligned}\mathbb{P}(\tau_n<T,\ X_{\tau_n}>\tfrac{\log(c+\varepsilon)d}{d+\lambda})&=\mathbb{P}(\tau_n<T,\ \tau_n<\tfrac{\log(c+\varepsilon)}{d+\lambda})\\&\leq\mathbb{P}(\tau_n<T,\ \tau_n<t_n^*)=0.\end{aligned}$$

Let $u_1=\frac{\log(c+\varepsilon)d}{d+\lambda}$. Note that $d<0$ else 0 would be regular for $(0,\infty)$ and we get that

$$\exists u_1<0,N\in\mathbb{N}\ \forall n>N:\ \mathbb{P}(\tau_n<T,\ X_{\tau_n}>u_1)=0. \tag{2.18}$$

Part iii) Let $Y=\sup\{X_t|\ t\in[T,\tau_0^{++})\}1_{(T<\tau_0^{++})}$. We want to show that $Y<0$ almost surely. Let $T_n=\inf\{t>T:\ X_t>\frac{-1}{n}\}$. Note that this is an increasing sequence of stopping times bounded by τ_0^{++}. Hence, the sequence of stopping times will a.s. converge to some random time $\tilde{T}$, with $\tilde{T}\leq\tau_0^{++}$ a.s.. From the quasi-left-continuity we obtain $X_{\tilde{T}}=\lim_{n\to\infty}X_{T_n}\geq\lim_{n\to\infty}(\frac{-1}{n})=0$ and, hence, $\tilde{T}=\sigma_0^+$. It then follows that

$$\begin{aligned}\mathbb{P}(Y=0)&=\mathbb{P}(T_n<\tau_0^{++}\forall n\in\mathbb{N})\\&=\mathbb{P}(X_{T_n}<0\forall n)\\&\leq\mathbb{P}(X_{\tilde{T}}=0)\\&=\mathbb{P}(X_{\sigma_0^+}=0)=0.\end{aligned}$$

Hence, there exists some $u_2 < 0$ such that $\mathbb{P}(Y > u_2) < \frac{1}{2}p$, and this implies that, for all $p > 0$, there exist $u_2 < 0$ and $N \in \mathbb{N}$ such that for all $n > N$,

$$\mathbb{P}(\tau_n \in [T, \tau_0^{++}),\ X_{\tau_n} > u_2) < \tfrac{1}{2}p. \tag{2.19}$$

Part iv) Combining (2.18) and (2.19) we obtain, with $u = \max\{u_1, u_2\}$,

$$\begin{aligned}\mathbb{P}(\tau_n < \tau_0^{++},\ X_{\tau_n} > u) =&\mathbb{P}(\tau_n \in [T, \tau_0^{++}),\ X_{\tau_n} > u)\\ &+ \mathbb{P}(\tau_n < T,\ X_{\tau_n} > u))\\ <&\tfrac{1}{2}p.\end{aligned}$$

Combining this with (2.17) it follows that, for all $p > 0$, there exists $y, u < 0$ and $N \in \mathbb{N}$ such that for all $n > N$,

$$\mathbb{P}(\tau_n < \tau_0^{++},\ X_{\tau_n} \in (y, u]) > \tfrac{1}{2}p. \tag{2.20}$$

Part v) First, we recall from the proof of Theorem 2.2.1 that

$$G(x) = v^*(x) - \tfrac{e^{2x}}{\mathbb{E}[e^{2\bar{X}_\infty}]}(e^{-x}\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty < -x)}] - \mathbb{E}[e^{2\bar{X}_\infty}1_{(\bar{X}_\infty < -x)}]).$$

Define

$$D(x_1, x_2) = \tfrac{e^{2x_1}}{\mathbb{E}[e^{2\bar{X}_\infty}]}(e^{-x_2}\mathbb{E}[e^{\bar{X}_\infty}1_{(\bar{X}_\infty < -x_2)}] - \mathbb{E}[e^{2\bar{X}_\infty}1_{(\bar{X}_\infty < -x_2)}]),$$

and note that when $x_1, x_2 \in (y, u]$,

$$D(x_1, x_2) \geq D(y, u) \geq \tfrac{e^y}{\mathbb{E}[e^{2\bar{X}_\infty}]}(e^{-u} - 1)\mathbb{P}(\bar{X}_\infty = 0) > 0. \tag{2.21}$$

Therefore, it follows that, for every $\tau \in \mathcal{T}$,

$$\begin{aligned}&\mathbb{E}[(e^{X_\tau} - c)^2 1_{(\tau < \tau_0^{++}\ ,\ X_\tau \in (y,u])}]\\ &= \mathbb{E}[G(X_\tau) 1_{(\tau < \tau_0^{++}\ ,\ X_\tau \in (y,u])}]\\ &= \mathbb{E}[(v^*(X_\tau) - D(X_\tau, X_\tau)) 1_{(\tau < \tau_0^{++}\ ,\ X_\tau \in (y,u])}]\\ &\leq \mathbb{E}[(v^*(X_\tau) - D(y, u)) 1_{(\tau < \tau_0^{++}\ ,\ X_\tau \in (y,u])}]\\ &= \mathbb{E}[v^*(X_\tau) 1_{(\tau < \tau_0^{++}\ ,\ X_\tau \in (y,u])}] - D(y, u)\mathbb{P}(\tau < \tau_0^{++}\ ,\ X_\tau \in (y, u])\\ &\leq \mathbb{E}[G(X_{\tau_0^{++}}) 1_{(\tau < \tau_0^{++}\ ,\ X_\tau \in (y,u])}] - D(y, u)\tfrac{1}{2}p.\end{aligned}$$

Hence, we obtain

$$
\begin{aligned}
\mathbb{V}[e^{X_\tau}] &= \mathbb{E}[(e^{X_\tau}-c)^2] - (c-\mathbb{E}[e^{X_\tau}])^2 \\
&= \mathbb{E}[(e^{X_\tau}-c)^2 1_{(\tau<\tau_0^{++},\ X_\tau\in(y,u])}] \\
&\quad + \mathbb{E}[(e^{X_\tau}-c)^2 1_{(\{\tau\geq\tau_0^{++}\}\ \cup\ \{X_\tau\notin(y,u]\})}] - (c-\mathbb{E}[e^{X_\tau}])^2 \\
&\leq \mathbb{E}[(e^{X_{\tau_0^{++}}}-c)^2 1_{(\tau<\tau_0^{++}\ ,\ X_\tau\in(y,u])}] - D(y,u)\tfrac{1}{2}p \\
&\quad + \mathbb{E}[(e^{X_\tau}-c)^2 1_{(\{\tau\geq\tau_0^{++}\}\ \cup\ \{X_\tau\notin(y,u]\})}] - (c-\mathbb{E}[e^{X_\tau}])^2 \\
&= \mathbb{E}[(e^{X_{\hat\tau}}-c)^2)] - D(y,u)\tfrac{1}{2}p - (c-\mathbb{E}[e^{X_\tau}])^2 \\
&\leq \mathbb{V}[e^{X_{\tau_0^{++}}}] - D(y,u)\tfrac{1}{2}p,
\end{aligned}
$$

where $\hat\tau = \tau 1_{(\tau<\tau_0^{++}\ ,\ X_\tau\in(y,u])} + \tau_0^{++}1_{(\{\tau\geq\tau_0^{++}\}\ \cup\ \{X_\tau\notin(y,u]\})}$. We conclude that

$$
\sup_{\tau\in\mathcal{T}^*} \mathbb{V}[e^{X_\tau}] \leq \mathbb{V}[e^{X_{\tau_0^{++}}}] - D(y,u)\tfrac{1}{2}p,
$$

where y and u are as given in (2.17) and (2.20). From (2.21) we have $D(y,u) > 0$; thus, the statement of the theorem follows. This completes the proof of Theorem 2.5.3. □

The following lemma, used in the proof of Theorem 2.5.3, is intuitive. Before the first jump of a compound Poison process, the process has not created any other information than the fact that there have been no jumps. Hence, if a stopping time with respect to $\mathcal{F}$ has a positive probability of stopping before the first jump then, given the stopping time occurs before the first jump, it is deterministic.

The lemma is the key reason why there is a gap between the variances at the stopping times for the solutions to the variance problem with and without randomized stopping, and that is the reason we have included a proof of it.

Lemma 2.5.4. *Let X be a compound Poisson process with negative drift. Let T denote the first jump time of the process, let τ be a stopping time with respect to $\mathcal{F}$, and assume $\mathbb{P}(\tau < T) > 0$. Then there exists some $t^* \geq 0$ such that:*

a) $\{T > t^*\} \cap \{\tau = t^*\} \overset{a.s.}{=} \{T > t^*\}$.

b) $\{T > \tau\} \cap \{\tau = t^*\} \overset{a.s.}{=} \{T > \tau\}$.

Proof. a) Let $\mathcal{F}_t^* = \sigma(X_s | s \in [0,t])$. Then all sets in $\mathcal{F}_t^*$ will either contain $\{T > t\}$ or be contained in $\{T > t\}^c$. As τ is a stopping time with respect to $\mathcal{F}$, then, for every $t \geq 0$, $\{\tau \leq t\}$ is a.s. equal to some set which is measurable with respect to $\mathcal{F}_t^*$. Thus, for all $t \geq 0$ it follows that $\{T > t\} \cap \{\tau \leq t\} \overset{a.s.}{=} \{T > t\}$ or $\{T > t\} \cap \{\tau \leq t\} \overset{a.s.}{=} \emptyset$. As $\mathbb{P}(\tau < T) > 0$, we cannot have $\{T > t\} \cap \{\tau \leq t\} \overset{a.s.}{=} \emptyset$ for all $t \geq 0$. Hence, we define $t^* = \inf\{t \geq 0 : \{T > t\} \cap \{\tau \leq t\} \overset{a.s.}{=} \{T > t\}\}$ and there exists a sequence $t_n \downarrow t^*$ such that $\{T > t_n\} \cap \{\tau \leq t_n\} \overset{a.s.}{=} \{T > t_n\}$ for every n. Hence, it follows that $\{T > t^*\} \cap \{\tau \leq t^*\} \supseteq \bigcap_{k=1}^{\infty} (\{\tau \leq t_k\} \cap \{T > t_k\}) \overset{a.s.}{=} \{T > t^*\}$. Thus, we obtain $\{T > t^*\} \cap \{\tau \leq t^*\} \overset{a.s.}{=} \{T > t^*\}$. Besides, $\{T > t^*\} \cap \{\tau < t^*\} \subseteq \bigcup_{n=1}^{\infty} (\{T > t^*\} \cap \{\tau \leq t^* - \frac{1}{n}\}) \overset{a.s.}{=} \emptyset$ and part a) follows.

b) Define a new process $Y_t = X_{\tau+t} - X_\tau$, and let T^Y denote the time of the first jump of Y. Thus, if $T > \tau$ then $T = T^Y + \tau$. Note that $\mathbb{P}(T^Y > t^*) > 0$ because Y has the same distribution as X. As T^Y is independent of $\mathcal{F}_\tau$ and using part a) we find that

$$\begin{aligned}
\mathbb{P}(\tau < T, \tau \neq t^*)\mathbb{P}(T^Y > t^*) &= \mathbb{P}(\tau < T, \tau \neq t^*, T^Y > t^*) \\
&= \mathbb{P}(\tau < T, \tau \neq t^*, T > t^* + \tau) \\
&= \mathbb{P}(\tau < T, \tau \neq t^*, T > t^* + \tau, \tau = t^*) = 0.
\end{aligned}$$

It must therefore hold that $\mathbb{P}(\tau < T, \tau \neq t^*) = 0$ and, thus, part b) follows. □

Example 2.5.2. *In this example we show that there exist compound Poisson processes with negative drift and $\psi(2) < 0$ for which the embedding equations have no solution. Let X be the compound Poisson process of Example 2.3.2 with $p = 1/11$ and $\alpha = \log(2)$. For this process, we have $\psi(2) < 0$ and the embedding equations have no solution. We define a compound Poisson processes with negative drift by $Y_t = X_t - dt$, where $d > 0$ is a constant. Then Y also has $\psi(2) < 0$. When d converges to 0, then Y converges a.s. to X dominated by X. Therefore, by choosing sufficiently small d, the embedding equations for Y have also no solution.*

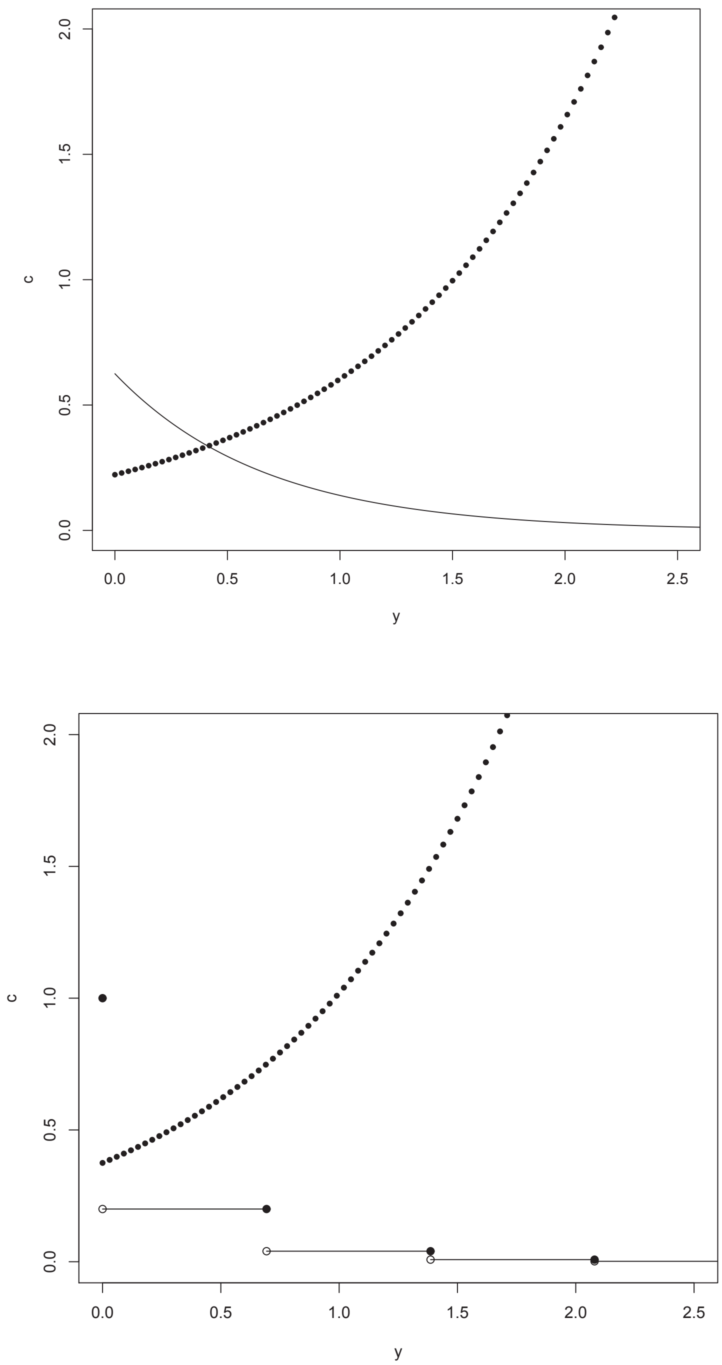

Figure 2.1: The dotted lines indicate the (y, c) values for which τ_y^+ solves the quadratic problem with parameter c, and the solid lines indicates the (y, c) values for which $c = \mathbb{E}[\exp(X_{\tau_y^+})]$. When the two lines intersect, the y-value of the intersection gives a value for which τ_y^+ solves the variance problem. Top: Cramér-Lundberg process of Example 2.3.1 with $\alpha = 3$, $d = 4$ and $\lambda = 2$. Bottom: Compound Poisson process of Example 2.3.2 with $p = 1/11$ and $\alpha = \log(2)$.

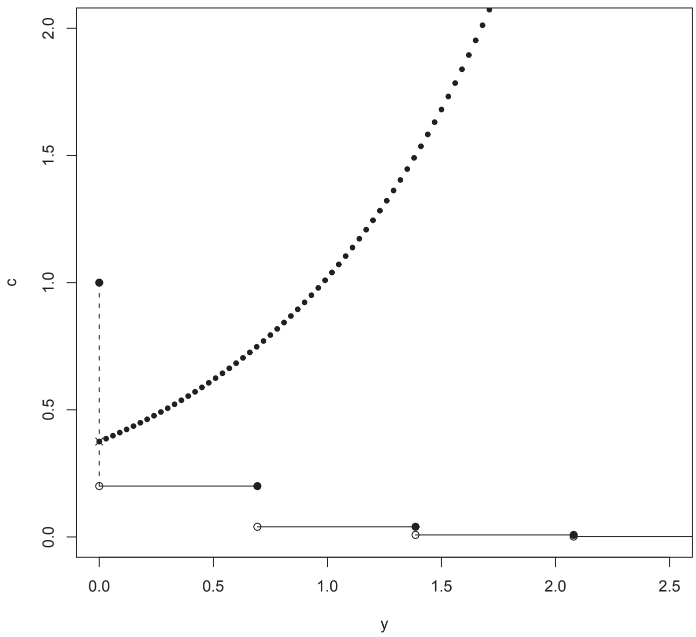

Figure 2.2: An illustration of Example 2.4.1 with $p = 1/11$ and $\alpha = \log(2)$. The dotted line indicates the (y, c) values for which τ_y^+ and τ_y^{++} solves the quadratic problem with parameter c, and the solid line indicates the (y, c) values for $c = \mathbb{E}[e^{X_{\tau_y^+}}]$. The dashed line and the cross illustrate how, by randomizing between τ_0^+ and τ_0^{++}, we may obtain a (τ, c) such that both requirements are fulfilled.

References for Chapter 2

[1] Alili, L. and Kyprianou, A. E. (2005). Some Remarks on First Passage of Lévy Processes, the American Put and Pasting Principles. *Ann. Appl. Probab.* **15,** 2062–2080.

[2] Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications.* Springer.

[3] Markowitz, H. M. (1952). Portfolio selection. *J Financ* **7,** 77–91.

[4] Mikosch, T. (2004). *Non-Life Insurance Mathematics.* Springer.

[5] Mordecki, E. (2002). Optimal stopping and perpetual options for Lévy processes. *Financ Stoch* **6,** 473–493.

[6] Pedersen, J. L. (2011). Explicit solutions to some optimal variance stopping problems. *Stoch* **83,** 505–518.

[7] Peskir, G. and Shiryaev, A. N. (2006). *Optimal Stopping and Free-Boundary Problems.* Birkhäuser.

[8] Sato, K.-I. (2002). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press.

[9] Shiryaev, A. N. (1978). *Optimal Stopping Rules.* Springer.

3. Mean-Variance Optimal Stopping for some Geometric Lévy Processes

Kamille Sofie Tågholt Gad

Abstract

Given a geometric Lévy process, X, we consider the problem of maximizing

$$\sup_{\tau} \left(\mathbb{E}[X_\tau] - c\mathbb{V}[X_\tau]\right).$$

This we denote *the mean-variance problem.* In the first part we study geometric spectrally negative Lévy processes and the main result is derivation of optimal stopping times to the mean-variance problem, both statically and dynamically. In both cases the solution is an excess boundary time. The results extend the results of [7], and the method relies on solving the same auxiliary optimal stopping problems.

In the second part the main result is an optimal stopping time for the mean-variance problem for a particular Cramér-Lundberg process with exponential upwards jumps. The solution for this process is interesting as the static solution is no longer an excess boundary time, but instead a hitting time of an interval. The dynamic solution remains an excess boundary time. At last, we derive the remarkable result that for some starting values randomized stopping times are optimal for the problem of minimizing the variance conditional on a lower bound on the mean.

Keywords: Mean-variance criterion; Optimal stopping; Geometric Lévy processes; Quadratic optimal stopping.

3.1 Introduction

Given a geometric Lévy process, X, we study the optimal stopping problem

$$\sup_{\tau \in \mathcal{T}} \left(\mathbb{E}[X_\tau] - c\mathbb{V}[X_\tau]\right), \tag{3.1}$$

where $c > 0$ is a constant and $\mathcal{T}$ is the stopping times generated from X, for which the variance of the process evaluated at that time is finite. This problem we denote *the mean-variance problem.* The idea of maximizing mean while minimizing variance is familiar in economic and financial application and is dating back to Markowitz [4], however not in the optimal stopping setting. In [7], a thorough presentation of the origin of the mean-variance problem within economic and finance is found.

The mean-variance problem is a non-linear optimal stopping problem. Hence it falls outside the scope of classical optimal stopping problems, and we cannot directly rely on results from [8] and [10]. Solving non-linear optimal stopping problems is a rather new field. The mean-variance optimal stopping problem is solved for geometric Brownian motions in [7]. We extend these results to some geometric Lévy processes. From the study of maximizing variance in [6] and [2] it is known that the jumps may have a significant influence on the solution to non-linear optimal stopping problems. We find that this is also the case for the mean-variance problem.

In the first part of the paper, the mean-variance problem is solved for spectrally negative Lévy processes. That is, geometric Lévy processes which does not have upwards jumps. The proof follows the main lines of the approach taken in [7] and is thus based on investigating an auxiliary classical quadratic optimal stopping problem. We find that the solution to the mean-variance problem is an excess boundary time and we find an implicit expression for the boundary. From the derivation of the optimal stopping time to the mean-variance problem we deduce the solution to the optimal stopping problem given by

$$\inf_{\tau \in \mathcal{T}:\mathbb{E}[X_\tau]\geq M} \mathbb{V}[X_\tau]. \tag{3.2}$$

This problem we denote *the conditional variance problem*, and we find an optimal stopping time which is an excess boundary time.

In Section 3.4 we solve the mean-variance problem for a particular Cramér-Lundberg process with exponential jumps upwards. This process is interesting because it is possible to carry through the computations, and the solutions are interesting. In the proof the class we maximize over is expanded to include randomized stopping times as defined in [1] and [10] for discrete time problems. The inclusion of randomized stopping times eases the search for an optimal stopping time. This is because our approach for solving the mean-variance problem relies on solving classical optimal stopping problems with constraints. As mentioned for discrete time problems in [10], such problems are eased by allowing randomized stopping times. Maximizing over this expanded class of stopping times we still find an optimal stopping time to the mean-variance problem from the original class of stopping times. It is interesting that the optimal stopping time we find is a hitting time for an interval. Even more remarkably, the conditional variance problem for some starting values has optimal stopping times which are randomized stopping times.

The fact that the mean-variance problem is solved by the hitting time of an interval is likely problematic for many applications. When maximizing expectation while punishing variance it is for application purposes meant to prevent risk from loses and not risk from high gains. In the solutions we find, we do not stop if we get a high gain, but wait for the process to fall down again because we know it eventually will. This is a disadvantage to the application of the mean-variance optimization and it comes from the variance having too much power in the extremes.

The fact that the conditional variance problem in some cases is solved by randomized stopping times is not a problem for applicational purposes, but we find it remarkable that random choices may be an advantage when minimizing variance conditioned on a minimum expectation.

For the mean-variance problem the stopping boundaries we find depend on the starting value of the process, and thus the problem is time inconsistent. A time consistent version of the problem is introduced in [7] by what is denoted *dynamic optimization* in contrast to the traditional optimal stopping problem of (3.1) which is denoted *static optimization.* In Section 3.5 we give dynamically optimal stopping times to the mean variance problem. For both the studied cases these optimal stopping times are excess boundary times with fixed boundaries.

3.2 Problem Formulation and Approach

Let $Y = (Y_t)_{t\geq 0}$ be a Lévy process and let $X_t = e^{Y_t}$ for $t \geq 0$. Let $\mathcal{F}$ be the augmented natural filtration satisfying the usual conditions and let $\mathcal{T}$ denote the set of stopping times with respect to $\mathcal{F}$ (all terms defined as in [3]). Given a constant $c > 0$ we study the optimal stopping problem of (3.1). When we solve the problem in the traditional static sense we wish to determine V and find a $\tau^* \in \mathcal{T}$ such that

$$V(x) \equiv \sup_{\tau\in\mathcal{T}}(\mathbb{E}_x[X_\tau] - c\mathbb{V}_x[X_\tau]) = \mathbb{E}_x[X_{\tau^*}] - c\mathbb{V}_x[X_{\tau^*}], \tag{3.3}$$

where $\mathbb{E}_x$ and $\mathbb{V}_x$ denote respectively mean and variance given the process starts in x. For Brownian motions the solution is known from [7] to be found among excess boundary times. Given a process, Y, we use the following notation for hitting times. For any $y < a$ and A an interval, then

$$\tau_a^+ \equiv \inf\{t \geq 0 : \ Y_t \geq a\}$$
$$\tau_A \equiv \inf\{t \geq 0 : \ Y_t \in A\}.$$

The present paper is built on the Lagrange approach from [7]. We describe it here. It is first noted that

$$V(x) = \sup_{\tau:\mathbb{E}_x[X_\tau]\geq x} (\mathbb{E}_x[X_\tau] - c\mathbb{V}_x[X_\tau])$$
$$= \sup_{M\geq x} (M + cM^2 - c \inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2]). \tag{3.4}$$

Then it is noted that for $x \leq M$ if there is a $\lambda > 0$ and a $\tau^* \in \mathcal{T}$ such that

$$\sup_{\tau\in\mathcal{T}} \mathbb{E}_x[\lambda X_\tau - X_\tau^2] = \mathbb{E}_x[\lambda X_{\tau^*} - X_{\tau^*}^2], \tag{3.5}$$

and

$$M = \mathbb{E}_x[X_{\tau^*}], \tag{3.6}$$

then τ^* also fulfil

$$\inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2] = \mathbb{E}_x[X_{\tau^*}^2]. \tag{3.7}$$

This we can insert in (3.4). We may also use the result of (3.7) to solve the conditional variance problem since

$$\inf_{\tau:\mathbb{E}_x[X_\tau]\geq M} \mathbb{V}_x[X_\tau] = \inf_{\beta\geq M} \left(\inf_{\tau:\mathbb{E}_x[X_\tau]=\beta} \mathbb{E}_x[X_\tau^2] - \beta^2 \right). \tag{3.8}$$

3.3 Spectrally Negative Lévy Processes

Assume Y is spectrally negative. That is, Y has no upwards jumps. Spectrally negative Lévy processes has the advantage that their value upon an excess boundary time is given by the boundary whenever the stopping time is attained. Whenever $y < a$ then $Y_{\tau_a^+} 1_{(\tau_a^+<\infty)} = a 1_{(\tau_a^+<\infty)}$. Besides, the spectrally negative processes have the advantage that the distribution of the maximum value of the process is known. Let the Laplace exponent of Y be given by:

$$\psi(\theta) = -\log(\mathbb{E}_0[e^{\theta Y_1}]),$$

and define ϕ as the right inverse of ψ. Then it is known from e.g. [3] that the distribution of the maximum is exponential with parameter $\phi(0)$.

Our main result is Theorem 3.3.1.

Theorem 3.3.1. *Assume Y is a spectrally negative Lévy process and define $\phi(0)$ from Y. Let $X = e^Y$ and consider the mean-variance problem of* (3.3).

a) If $\phi(0) = 0$, then $\mathbb{E}_x[X_{\tau_b^+}] - c\mathbb{V}_x[X_{\tau_b^+}] \to \infty$ when $b \to \infty$.

b) If $\phi(0) \in (0,1)$, then $\tau^+_{b(x)}$ is an optimal stopping time for (3.3), where $b(x)$ is the solution to

$$b(x)^{\phi(0)-1} + 2cx^{\phi(0)} - c\tfrac{2-\phi(0)}{1-\phi(0)} b(x)^{\phi(0)} = 0.$$

Let $b^* \equiv (1-\phi(0))/(\phi(0)c)$. Specifically it is optimal to stop right away if $x < b^*$.

c) If $\phi(0) \geq 1$, then it is optimal to stop right away.

Before we turn to the proof of Theorem 3.3.1, we solve the classical optimal stopping problem of (3.5) for sufficiently many combinations of (x, λ) that it is possible to solve the problem of (3.6), and from this (3.7). These preliminary problems are solved in Proposition 3.3.2 and Lemma 3.3.3.

Proposition 3.3.2. *Given $\lambda > 0$ and a spectrally negative Lévy process Y, define $\phi(0)$ from Y and let $X = e^Y$. Let*

$$b^*_\lambda = \lambda \frac{1-\phi(0)}{2-\phi(0)}. \tag{3.9}$$

If $\phi(0) \in (0,1)$ and $e^{Y_0} < b^*$, then $\tau^+_{b^*_\lambda}$ is an optimal stopping time for (3.5).

Proof. Let $G(x) = \lambda x - x^2$. Then G is continuous and $\mathbb{E}_x[\sup_{t>0} G(X_t)] \leq \lambda\frac{\lambda}{2} - \frac{\lambda^2}{2^2} = \frac{\lambda^2}{4} < \infty$. Thus, we know from [8] that the problem of (3.5) has a solution which is a hitting time. Let D denote the stopping region of the optimal hitting time. Since Y is spectrally negative, then for any $x < \inf D$ it holds that $\tau_D = \tau^+_{\inf D}$. Notice that G is a second order concave polynomial with supremum attained in $\lambda/2$. Thus, it is reasonable to guess that $-\infty < \inf D < \lambda/2$. For $x < \lambda/2$ we search for the best stopping time of the form τ^+_b, where $b \in (x, \lambda/2)$. For $x \leq b < \lambda/2$

$$\begin{aligned}
\mathbb{E}_x[\lambda X_{\tau^+_b} - X^2_{\tau^+_b}] &= (\lambda b - b^2)\mathbb{P}_x(\tau^+_b < \infty) \\
&= x^{\phi(0)}(\lambda b^{1-\phi(0)} - b^{2-\phi(0)}) \\
&\equiv x^{\phi(0)} H(b).
\end{aligned}$$

H is C^2 and by differentiation we find that b^*_λ is the only critical point. $H''(b^*_\lambda) < 0$ and thus b^*_λ is a unique supremum.

Now, we wish to show that $b^*_\lambda = \inf D$. Assume $x < b^*_\lambda$, then

$$\mathbb{E}_x[G(X_{\tau^+_{b^*_\lambda}})] = x^{\phi(0)} H_\lambda(b^*_\lambda) > x^{\phi(0)} H_\lambda(x) = \lambda x - x^2 = \mathbb{E}_x[G(X_0)].$$

This shows that when starting in $X_0 = x$ lower than b^*_λ, the stopping time $\tau^+_{b^*_\lambda}$ is strictly better than stopping right away. Thus, for all $x < b^*_\lambda$ we have $x \notin D$, and thus $-\infty < x < b^*_\lambda$. We may now conclude that for every $x < b^*_\lambda$ the solution is the hitting time of an upper boundary. We have found that b^*_λ is the best choice for a boundary, so $\tau^+_{b^*_\lambda}$ is an optimal stopping time for $x < b^*_\lambda$. □

Lemma 3.3.3. *Given x, M with $M > x$ and a spectrally negative Lévy process, Y, with $x = e^{Y_0}$. Define $\phi(0)$ from Y and let $X = e^Y$. Assume $\phi(0) \in (0, 1)$, let b^*_λ be given from* (3.9) *and choose*

$$\lambda(M, x) = (Mx^{-\phi(0)})^{\frac{1}{1-\phi(0)}} \frac{2 - \phi(0)}{1 - \phi(0)}.$$

Then $\tau^*_{b^*_{\lambda(M,x)}}$ is an optimal stopping time for (3.7) *and*

$$\inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X^2_\tau] = M^{\frac{2-\phi(0)}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}}.$$

Proof. First we search for an optimal stopping time $\tau^*_\lambda = \tau^*_{\lambda(M,x)}$ to the problem (3.5) with $\lambda = \lambda(M, x)$ such that also (3.6) is fulfilled. From Proposition 3.3.2 it follows that if $x < b^*_{\lambda(M,x)}$ then $\tau^+_{b^*_{\lambda(M,x)}}$ is optimal for (3.5). And when $x < b^*_{\lambda(M,x)}$ then (3.6) corresponds to

$$M = \mathbb{E}_x[X_{\tau^+_{b^*_{\lambda(M,x)}}}] = x^{\phi(0)} (\frac{1 - \phi(0)}{2 - \phi(0)})^{(1-\phi(0))} (\lambda(M, x))^{(1-\phi(0))},$$

and thus

$$\lambda(M, x) = (Mx^{-\phi(0)})^{\frac{1}{1-\phi(0)}} \frac{2 - \phi(0)}{1 - \phi(0)}. \tag{3.10}$$

With this $\lambda(M, x)$ then

$$\begin{aligned} b^*_{\lambda(M,x)} &= \lambda(M, x) \frac{1 - \phi(0)}{2 - \phi(0)} \\ &= (Mx^{-\phi(0)})^{\frac{1}{1-\phi(0)}} \\ &\geq (x^{1-\phi(0)})^{\frac{1}{1-\phi(0)}} = x, \end{aligned} \tag{3.11}$$

and thus λ from (3.10) can be used to obtain the desired property (3.6). Thus, from [7] as repeated in Section 3.2, $\tau^+_{b^*_{\lambda(M,y)}}$ is optimal for (3.7). By use of (3.11) we get

$$\begin{aligned}
\inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2] &= \mathbb{E}_x[X^2_{\tau^+_{b^*_{\lambda(M,y)}}}] \\
&= (b^*_{\lambda(M,y)})^2 \mathbb{P}_x(\tau^+_{b^*_{\lambda(M,x)}} < \infty) \\
&= (b^*_{\lambda(M,y)})^2 e^{-\phi(0)(\log(b^*_{\lambda(M,y)})-\log(x))} \\
&= M^{\frac{2-\phi(0)}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}}.
\end{aligned}$$

□

We are now ready to prove Theorem 3.3.1

Proof of Theorem 3.3.1.
a) If $\phi(0) = 0$, then $\mathbb{P}_x(\tau_b^+ < \infty) = 1$ for all x, b. The process is creeping over any upper boundary, and thus for $x < b$

$$\mathbb{E}_x[X_{\tau_b^+}] - c\mathbb{V}[X_{\tau_b^+}] = b$$

Thus, by choosing an arbitrary large upper bound we may obtain an arbitrarily high value.

c) If $\phi(0) \geq 1$, then $\psi(1) \leq 0$. In general $(e^{Y_t - \psi(1)t})_{t\geq 0}$ is a martingale, and thus, since $\psi(1) \leq 0$, then $(X_t)_{t\geq 0} = (e^{Y_t})_{t\geq 0}$ is a positive supermartingale. Thus, from Fubini and the Optional Sampling Theorem,

$$\mathbb{E}_x[X_\tau] = \mathbb{E}_x[\liminf_{N\to\infty} X_{\tau\wedge N}] \leq \liminf_{N\to\infty} \mathbb{E}_x[X_{\tau\wedge N}] \leq \liminf_{N\to\infty} \mathbb{E}_x[X_0] = x$$

Thus, no stopping time will give a better mean value than stopping right away, and thus it is optimal to stop at once.

b) Combining (3.4) and the result of Lemma 3.3.3 we get

$$\begin{aligned}
V(x) &= \sup_{M\geq x} (M + cM^2 - c(M^{\frac{2-\phi(0)}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}})) \vee x \\
&= \sup_{M\geq x} F_x(M) \vee x = \sup_{M\geq x} F_x(M), \qquad (3.12)
\end{aligned}$$

where

$$F_x(M) = M + cM^2 - c(M^{\frac{2-\phi(0)}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}})$$

The latter equality in (3.12) follows because $F_x(x) = x$. Notice that $\frac{2-\phi(0)}{1-\phi(0)} >$

2 and therefore $F_x(M) \to -\infty$ as $M \to \infty$.

$$\begin{aligned} F_x'(M) &= 1 + 2cM - c\tfrac{2-\phi(0)}{1-\phi(0)} M^{\frac{1}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}} \\ F_x''(M) &= 2c - c\tfrac{2-\phi(0)}{1-\phi(0)} \tfrac{1}{1-\phi(0)} M^{\frac{\phi(0)}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}} \\ F_x'(0) &= 1 \end{aligned}$$

$F_x'(0) > 0$ and $F_x''(M)$ is increasing for $M > 0$. Thus, F_x has a unique maximum for $M > 0$ and it is attained in the unique $M > 0$ with $F_x'(M) = 0$. The maximum of $F_x(M)$ for $M > 0$ is attained for $M > x$ if $F_x'(x) > 0$.

$$F_x'(x) = 1 + 2cx - c\tfrac{2-\phi(0)}{1-\phi(0)} x^{\frac{1}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}} = 1 - 2x\tfrac{\phi(0)}{1-\phi(0)}.$$

So maximum is attained above x if

$$x < \frac{1-\phi(0)}{\phi(0)c}. \tag{3.13}$$

The optimal stopping time is the solution to Lemma 3.3.3 where M is the one which maximizes $F_x(M)$, and thus under the condition (3.13) M is given from $0 = F_x'(M)$, which corresponds to

$$0 = 1 + 2cM - c\tfrac{2-\phi(0)}{1-\phi(0)} M^{\frac{1}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}}. \tag{3.14}$$

From (3.11) we have that given M, the optimal stopping time of Lemma 3.3.3 is an excess boundary time with a boundary $b(x)$ given by

$$b(x) = (Mx^{-\phi(0)})^{\frac{1}{1-\phi(0)}}$$

Since the relation between $b(x)$ and M is injective, we may rewrite the condition of (3.14) to a condition on $b(x)$ instead. We then get that the optimal boundary $b(x)$ is the solution to

$$0 = 1 + 2cx^{\phi(0)} b(x)^{1-\phi(0)} - c\tfrac{2-\phi(0)}{1-\phi(0)} b(x).$$

However, only if (3.13) is fulfilled. Otherwise it is optimal to stop at once.

□

Proposition 3.3.4. *Given a spectrally negative Lévy process, Y, with $\phi(0) \in (0,1)$, let $X = e^Y$. Then $\tau^*_{b^*_{\lambda(M,x)}}$ of Lemma 3.3.3 is also optimal for the conditional variance problem* (3.2).

Proof. Let

$$F(\beta) = \inf_{\tau:\mathbb{E}_x[X_\tau]=\beta} (\mathbb{E}[X_\tau^2] - \beta^2) = \beta^{\frac{2-\phi(0)}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}} - \beta^2.$$

Then, for $\beta \geq x$

$$\begin{aligned} F'(\beta) &= \tfrac{2-\phi(0)}{1-\phi(0)} \beta^{\frac{1}{1-\phi(0)}} x^{\frac{-\phi(0)}{1-\phi(0)}} - 2\beta \\ &\geq \tfrac{2-\phi(0)}{1-\phi(0)} \beta^{\frac{1}{1-\phi(0)}} \beta^{\frac{-\phi(0)}{1-\phi(0)}} - 2\beta > 0. \end{aligned}$$

Thus, the infimum over $[M, \infty)$ is obtained for $\beta = M$. Therefore

$$\inf_{\tau:\mathbb{E}_x[X_\tau]\geq M} \mathbb{V}_x[X_\tau] = \inf_{\beta \geq M} \left(\inf_{\tau:\mathbb{E}_x[X_\tau]=\beta} (\mathbb{E}[X_\tau^2] - \beta^2) \right) = \inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}[X_\tau^2] - M^2,$$

and the result foliows. □

3.4 A Cramér-Lundberg Process

We consider the process

$$Y_t = y - dt + \sum_{n=1}^{N_t} Z_n, \tag{3.15}$$

where, $d > 0$ is a constant, $(Z_n)_{n\geq 1}$ is a sequence of i.i.d. random variables with $Z_1 \sim \exp(\alpha)$ with $\alpha = 3$, and where N_t is a Poisson process with intensity ν, and $\frac{\nu}{d} = \frac{5}{2}$. Assume N and $(Z_n)_{n\geq 1}$ are independent and let $(X_t)_{t\geq 0} = (e^{Y_t})_{t\geq 0}$.

For this process we derive that the optimal stopping time for the problem (3.3) is the hitting time of an interval. It becomes convenient to define the constants q, $F_1(q)$, $F_2(q)$ and κ, where

$$q = \frac{9 - \sqrt{21}}{20} \approx 0.221.$$

Let $\bar{X}_\infty = \sup_{t\geq 0} X_t$. From [5] Chapter 4.2 it follows that $\mathbb{P}(\bar{X}_\infty = 0) = \frac{1}{6}$ and that $\bar{X}_\infty$ has density $f(x) = \frac{5}{12} e^{-0.5x}$ on $(0, \infty)$. With this we easily calculate for $\lambda > 0$ and $x < \lambda q$ that

$$\begin{aligned} \mathbb{E}_x[X_{\tau_{[\lambda q, \lambda/2]}}] &= \sqrt{x}\sqrt{\lambda}(\tfrac{5}{4} q^{\frac{1}{2}} - \tfrac{5}{3} q^{\frac{5}{2}}) \equiv \sqrt{x}\sqrt{\lambda} F_1(q) \\ \mathbb{E}_x[X^2_{\tau_{[\lambda q, \lambda/2]}}] &= \sqrt{x}\lambda^{\frac{3}{2}}(\tfrac{5}{2} q^{\frac{3}{2}} - \tfrac{10}{3} q^{\frac{5}{2}}) \equiv \sqrt{x}\lambda^{\frac{3}{2}} F_2(q) \\ \kappa &= \frac{F_2(q)}{(F_1(q))^3} \approx 1.105. \end{aligned}$$

Theorem 3.4.1. *Let Y be given from* (3.15) *and let $X = e^Y$. Let*

$$b^* = \frac{3\kappa}{w_+^2 - 1} \approx 0.426,$$

where w_+^2 is the larger root of

$$0 = 2w^3 + (3 - 9\kappa)w^2 + 9\kappa - 1,$$

and let

$$\lambda^*(x) = \frac{x(1 + \sqrt{1 + \frac{3\kappa}{cx}})}{9\kappa F_1(q)^2}.$$

Then the following is an optimal stopping time for mean-variance problem of (3.3)*:*

$$\tau^*(x) = \begin{cases} \tau_{[\lambda^*(x)q, \lambda^*(x)/2)]} & cx \leq b^*, \\ 0 & cx \geq b^*. \end{cases}$$

Specifically it is optimal to stop right away if and only if $cx \leq b^*$.

In order to prove the theorem it is useful first to solve two auxiliary problems.

Proposition 3.4.2. *Let Y be given from* (3.15) *and let $X = e^Y$. Then both $\tau_{[\lambda q, \lambda/2]}$ and $\tau_{(\lambda q, \lambda/2]}$ are optimal for* (3.5).

Proof. Let $G(x) = \lambda x - x^2$. G is a concave second order polynomial with maximum attained in $x = \lambda/2$. X has no downwards jumps and $X_t \to -\infty$ almost surely as $t \to \infty$. Thus, if X starts above $\lambda/2$, then it almost surely reaches $\lambda/2$ at some time, but if it starts below $\lambda/2$ then it is not certain that it reaches $\lambda/2$. For this reason we guess there is a solution of the form $\tau_{[b,\lambda/2]}$ for some b. Let $G^Y(y) = \lambda e^y - e^{2y}$. Then

$$V_\lambda(x) \equiv \sup_{\tau \in \mathcal{T}} \mathbb{E}_x[G(X_\tau)] = \sup_{\tau \in \mathcal{T}} \mathbb{E}_{\log(x)}[G^Y(Y_\tau)].$$

We approach the problem based on Y instead of the problem X. Let $\tau_A^Y = \inf\{t \geq 0 : \ Y_t \in A\}$, for any interval A and note that

$$\mathbb{E}_x[G(X_{\tau_{[b,\lambda/2]}})] = \mathbb{E}_{\log(x)}[G^Y(Y_{\tau^Y_{[\log(b),\log(\lambda/2)]}})]$$

For $y < \log(b) < \log(\frac{\lambda}{2})$ we have

$$\begin{aligned}\mathbb{E}_y[G^Y(Y_{\tau^Y_{[\log(b),\log(\lambda/2)]}})] &= \lambda\mathbb{E}_y[e^{Y_{\tau^Y_{[\log(b),\log(\lambda/2)]}}}] - \mathbb{E}_y[e^{2Y_{\tau^Y_{[\log(b),\log(\lambda/2)]}}}] \\ &= \tfrac{5}{6}e^{y/2}(\tfrac{3}{2}\lambda b^{\frac{1}{2}} - 3b^{\frac{3}{2}} + 2\lambda^{-1}b^{\frac{5}{2}}) \\ &\equiv R(b)e^{y/2}.\end{aligned}$$

Let

$$\tilde{R}(\bar{b}) \equiv R(\bar{b}^2) = \tfrac{5}{6}(\tfrac{3}{2}\lambda\bar{b} - 3\bar{b}^3 + 2\lambda^{-1}\bar{b}^5)$$

Then optimizing $R(b)$ for $b \in (0, \lambda/2]$ corresponds to optimizing $\tilde{R}(\bar{b})$ for $\bar{b} \in (0, \sqrt{\lambda/2}]$ with $b = \bar{b}^2$.

$$\tilde{R}'(\bar{b}) = \tfrac{5}{6}(\tfrac{3}{2}\lambda - 9\bar{b}^2 + 10\lambda^{-1}\bar{b}^4)$$

and thus the critical points of $\tilde{R}$ are $\bar{b} = \pm\sqrt{\lambda(9 \pm \sqrt{21})/20}$. We wish to optimize $\tilde{R}$ over $(0, \sqrt{\lambda/2}]$, and since $-\sqrt{\lambda(9 \pm \sqrt{21})/20} < 0$ and $\sqrt{\lambda(9 + \sqrt{21})/20} > \sqrt{\lambda/2}$, and $\tilde{R}(b) \to \infty$ for $b \to \infty$, then $\tilde{R}$ is maximized over $(0, \sqrt{\lambda/2}]$ by $\bar{b} = \sqrt{\lambda(9 - \sqrt{21})/20} = \sqrt{\lambda q}$. And thus, R is maximized over $(0, \lambda/2]$ by $a = \log(\lambda q)$.

We guess that $\tau^Y_{[\log(\lambda q),\log(\frac{\lambda}{2})]}$ is optimal and compute the corresponding value function $\bar{V}_\lambda(y) \equiv \mathbb{E}_{\log(x)}[G^Y(Y_{\tau^Y_{[\log(\lambda q),\log(\lambda/2)]}})]$

$$\bar{V}_\lambda(y) = \begin{cases} R(\log(\lambda q))e^{y/2} & \text{if } y < \log(\lambda q), \\ G^Y(y) & \text{if } y \in [\log(\lambda q), \log(\frac{\lambda}{2})], \\ G^Y(\log(\frac{\lambda}{2})) = \frac{\lambda^2}{4} & \text{if } y > \log(\frac{\lambda}{2}). \end{cases} \tag{3.16}$$

Notice that $\bar{V}_\lambda \geq G$, which is the first crucial requirement for $\bar{V}_\lambda$ to be the value function.

For F differentiable from the right in y, and the following integral well defined let

$$AF(y) = -dF'(y) + \int_0^\infty (F(y+z) - F(y))\nu\alpha e^{-\alpha z}dz.$$

$\bar{V}_\lambda(y)$ is constant for $y \geq \log(\frac{\lambda}{2})$ and thus $A\bar{V}_\lambda(y) = 0$ for $y \geq \log(\frac{\lambda}{2})$.

For $y \in [\log(\lambda q), \log(\frac{\lambda}{2})]$ it holds that

$$\begin{aligned} A\bar{V}_\lambda(y) &= -d\bar{V}_\lambda'(y) + \int_0^\infty (\bar{V}_\lambda(y+z) - \bar{V}_\lambda(y))\nu\alpha e^{-\alpha z} dz \\ &= -\tfrac{2\nu}{5}(\lambda e^y - 2e^{2y}) + \int_0^{\log(\frac{\lambda}{2})-y} G^Y(y+z)\nu\alpha e^{-\alpha z} dz \\ &+ \int_{\log(\frac{\lambda}{2})-y}^\infty G^Y(\log(\tfrac{\lambda}{2}))\nu\alpha e^{-\alpha z} dz - (\lambda e^y - e^{2y})\nu \\ &= \tfrac{\nu e^y}{10\lambda}[\lambda^2 - 12\lambda e^y + 20e^{2y}]. \end{aligned}$$

This is non-positive for $e^y \in [\lambda/10, \lambda/2]$. Notice that $[\lambda q, \lambda/2] \subset [\lambda/10; \lambda/2]$. We then have $A\bar{V}_\lambda(y) \leq 0$ for $y \in [\log(\lambda q), \log(\lambda/2)]$.

For $y < \log(\lambda q)$ it holds that

$$\begin{aligned} A\bar{V}_\lambda(y) = &-\tfrac{2}{5}\nu R(\log(\lambda q))\tfrac{1}{2}e^{y/2} + \int_0^{\log(\lambda q)-y} R(\log(\lambda q))e^{(y+z)/2}\nu\alpha e^{-\alpha z} dz \\ &+ \int_{\log(\lambda q)-y}^{\log(\frac{\lambda}{2})-y} (\lambda e^{y+x} - e^{2(y+z)})\nu\alpha e^{-\alpha z} dz \\ &+ \int_{\log(\frac{\lambda}{2})-y}^\infty \tfrac{\lambda^2}{4}\nu\alpha e^{-\alpha z} dz - R(\bar{a})\nu e^{y/2} = 0. \end{aligned}$$

Then notice that

$$\begin{aligned} \bar{V}_\lambda(Y_t) &= \bar{V}_\lambda(y) - d\int_0^t \bar{V}_\lambda'(Y_s)1_{(Y_s \neq \bar{a})} ds + \int_0^t (\bar{V}_\lambda(Y_s) - \bar{V}_\lambda(Y_{s-}))dN_s \\ &= \bar{V}_\lambda(y) + \int_0^t A\bar{V}_\lambda(Y_{s-})ds + M_t, \end{aligned}$$

where

$$M_t = \int_0^t (\bar{V}_\lambda(Y_s) - \bar{V}_\lambda(Y_{s-}))dN_s - \int_0^t \int_0^\infty (\bar{V}_\lambda(Y_{s-}+z) - \bar{V}_\lambda(Y_{s-}))\nu\alpha e^{\alpha z} dz ds$$

is a martingale since $\bar{V}_\lambda$ is bounded (follows from e.g. [11]). Then for any $\tau \in \mathcal{T}$

$$\begin{aligned} \mathbb{E}_y[G^Y(Y_\tau)] &= \mathbb{E}_y[\liminf_{t\to\infty} G^Y(Y_{\tau\wedge t})] \\ &\leq \mathbb{E}_y[\liminf_{t\to\infty} \bar{V}_\lambda(Y_{\tau\wedge t})] \\ &\leq \liminf_{t\to\infty} \mathbb{E}_y[\bar{V}_\lambda(Y_{\tau\wedge t})] \\ &= \liminf_{t\to\infty} \mathbb{E}_y[\bar{V}_\lambda(y) + \int_0^{\tau\wedge t} A\bar{V}_\lambda(Y_{s-})ds + M_{\tau\wedge t}] \\ &\leq \bar{V}_\lambda(y) + \liminf_{t\to\infty} \mathbb{E}_y[M_{\tau\wedge t}] = \bar{V}_\lambda(y). \end{aligned}$$

Since $\bar{V}_\lambda$ is the value from choosing $\tau^Y_{[\log(\lambda q),\log(\lambda/2)]}$, then $\bar{V}_\lambda$ is the value function for (**??**). $V_\lambda(x) = \bar{V}_\lambda(\log(x))$ and $\tau_{[\lambda q,\lambda/2]}$ is optimal for (3.5).

When the process, X, starts in $x < \lambda q$ then $\tau_{[\lambda q,\lambda/2]} \overset{a.s.}{=} \tau_{(\lambda q,\lambda/2]}$ (see e.g. [9] Lemma 49.6), and continuous fit is easily verified from (3.16)

$$\mathbb{E}_{\lambda q}[G(X_{\tau_{[\lambda q,\lambda/2]}})] = \bar{V}_\lambda(\lambda q) = \lim_{x\uparrow\lambda q} \bar{V}_\lambda(x) = \mathbb{E}_{\lambda q}[G(X_{\tau_{(\lambda q,\lambda/2]}})].$$

Thus the two stopping times $\tau_{(\lambda q,\lambda/2]}$ and $\tau_{[\lambda q,\lambda/2]}$ are equally good. □

Lemma 3.4.3. *Let Y be given from* (3.15) *and let $X = e^Y$. Consider the problem of* (3.7).

For $M \geq \frac{F_1(q)}{\sqrt{q}}x$ let $\lambda(M,x) = M^2x^{-1}F_2(q)^{-2}$. Then $\tau_{[\lambda(M,x)q,\lambda(M,x)/2)}$ is an optimal stopping time and

$$\inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2] = x^{-1}M^3\kappa.$$

For $M \in (x, \frac{F_1(q)}{\sqrt{q}}x)$, let U be a random variable independent of X and taking values in $\{0,1\}$ with

$$\mathbb{P}(U=1) = p(M) = \frac{\frac{F_1(q)}{\sqrt{q}} - Mx^{-1}}{\frac{F_1(q)}{\sqrt{q}} - 1}. \tag{3.17}$$

Then $(1-U)\tau_{(x,x(2q)^{-1}]}$ is an optimal stopping time and

$$\inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2] = x^2p(M) + x^2q^{-2/3}F_2(q)(1-p(M)).$$

Remark 3.4.1. *$(1-U)\tau_{(x,x(2q)^{-1}]}$ is a randomized stopping time as known from e.g. [1] and [10].*

Proof. Let τ^*_λ be optimal to the classical optimal stopping problem of Proposition 3.4.2. Given, $x < M$ we wish to find a $\lambda(M,x)$ such that (3.6) is fulfilled. From Proposition 3.4.2 it follows that $\tau^*_\lambda = \tau_{[\lambda q,\lambda/2]}$ is optimal to the quadratic classical optimal stopping problem (3.5). Thus, (3.6) can be obtained if $x < \lambda(M,x)q$ and

$$M = \sqrt{x}\sqrt{\lambda(M,x)}F_1(q) \Leftrightarrow$$
$$\lambda(M,x) = \frac{M^2}{x(F_1(q))^2}. \tag{3.18}$$

Notice that the requirement $x \leq \lambda(M,x)q$ corresponds to

$$M \geq \frac{F_1(q)}{\sqrt{q}}x \approx 1.17x.$$

Thus, if $M \geq \frac{F_1(q)}{\sqrt{q}}x$, then $\lambda(M,x)$ of (3.18) and $\tau^*_\lambda = \tau_{[\lambda(M,x)q,\lambda(M,x)/2]}$ fulfil the two requirements.

For $M \in (x, \frac{F_1(q)}{\sqrt{q}}x)$ take $\lambda(M,x) = x/q$, then both $\tau_{[x,x(2q)^{-1})]} = 0$ and $\tau_{(x,x(2q)^{-1}]}$ are optimal for (3.5). Let $p(M)$ and U be given from the statement of the lemma. Define $\tau^*_{\lambda(M,x)} = 1_{(U=0)}\tau_{(x,x(2q)^{-1}]}$. Then

$$\begin{aligned}\mathbb{E}_x[G(X_{\tau^*_{\lambda(M,x)}})] &= \mathbb{E}_x[G(X_0)]p(M) + \mathbb{E}_x[G(X_{\tau_{(x,x(2q)^{-1}]}})](1-p(M)) \\ &= V_{\lambda(M,x)}(x)p(M) + V_{\lambda(M,x)}(x)(1-p(M)) = V_{\lambda(M,x)}(x),\end{aligned}$$

so $\tau^*_{\lambda(M,x)}$ is optimal for (3.5). Besides

$$\begin{aligned}\mathbb{E}_x[X_{\tau^*_{\lambda(M,x)}}] &= \mathbb{E}_x[X_0]p(M) + \mathbb{E}_x[X_{\tau_{(x,x(2q)^{-1}]}}](1-p(M)) \\ &= x\frac{\frac{F_1(q)}{\sqrt{q}} - Mx^{-1}}{\frac{F_1(q)}{\sqrt{q}} - 1} + \frac{F_1(q)}{\sqrt{q}}x\frac{Mx^{-1}-1}{\frac{F_1(q)}{\sqrt{q}} - 1} = M.\end{aligned}$$

Thus $\tau^*_{\lambda(M,x)}$ fulfill both of the two requirements and thus is optimal for (3.7). For $M \geq \frac{F_1(q)}{\sqrt{q}}x$

$$\begin{aligned}\inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2] &= \mathbb{E}_x[X^2_{\tau^*_{\lambda(M,x)}}] \\ &= \sqrt{x}(\lambda(M,x))^{\frac{3}{2}}F_2(q) \\ &= x^{-1}\frac{M^3F_2(q)}{(F_1(q))^3} \equiv x^{-1}M^3\kappa.\end{aligned}$$

and for $M \in (x, \frac{F_1(q)}{\sqrt{q}}x)$ then

$$\begin{aligned}\inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2] &= \mathbb{E}_x[X^2_{\tau^*_{\lambda(M,x)}}] \\ &= \mathbb{E}_x[X_0^2]p(M) + \mathbb{E}_x[X^2_{\tau_{(x,x(2q)^{-1}]}}](1-p(M)) \\ &= x^2p(M) + (\sqrt{x}\lambda(M,x)^{3/2}F_2(q))(1-p(M)) \\ &= x^2p(M) + x^2q^{-3/2}F_2(q)(1-p(M))\end{aligned}$$

□

Proof of Theorem 3.4.1:

We have

$$
\begin{aligned}
V(x) &= \sup_{M \geq x} (M + cM^2 - c \inf_{\tau:\mathbb{E}_x[X_\tau]=M} \mathbb{E}_x[X_\tau^2]) \\
&= \sup_{M \geq \frac{F_1(q)}{\sqrt{q}}x} \left(M + cM^2 - cx^{-1}\kappa M^3\right) \vee x \\
&\vee \sup_{M \in (x, \frac{F_1(q)}{\sqrt{q}}x)} \left(\mathbb{E}_x[X_{\tau^*(y,M)}] - c\mathbb{V}_x[X_{\tau^*(M,x)}]\right),
\end{aligned}
$$

where $\tau^*(M, x)$ is the randomized optimal stopping time of Lemma 3.4.3 for when $M \in (x, \frac{F_1(q)}{\sqrt{q}})$. However, it holds in general that the mixture of two stopping times cannot solve the mean-variance problem. If we have two stopping times τ_1 and τ_2 and let $U \in \{0, 1\}$ be a random variable independent of X and with $\mathbb{P}(U = 1) = p \in (0, 1)$, and define $\tau = \tau_1 U + \tau_2(1 - U)$, then

$$
\begin{aligned}
\mathbb{E}[X_\tau] - c\mathbb{V}[X_\tau] = (\mathbb{E}[X_{\tau_1}] - c\mathbb{V}[X_{\tau_1}])\, p + (\mathbb{E}[X_{\tau_2}] - c\mathbb{V}[X_{\tau_1}])\,(1 - p) \\
- cp(1 - p)\left(\mathbb{E}[X_{\tau_1}]^2 + \mathbb{E}[X_{\tau_2}]^2\right).
\end{aligned}
$$

Thus either τ_1 or τ_2 is strictly better than τ. Let

$$
H(M) = M + cM^2 - cx^{-1}\frac{F_2(q)}{(F_1(q))^3}M^3.
$$

Then

$$
\begin{aligned}
\mathbb{E}_x[X_{\tau^*(y,M)}] - c\mathbb{V}_x[X_{\tau^*(y,M)}] &\leq p(M)x + (1 - p(M))H(\tfrac{F_1(q)}{\sqrt{q}}x) \\
&\leq \max\{x, H(\tfrac{F_1(q)}{\sqrt{q}}x)\},
\end{aligned}
$$

and then

$$
V(x) = \sup_{M \geq \frac{F_1(q)}{\sqrt{q}}x} H(M) \vee x.
$$

H' is a concave second order polynomial. The larger root of $H'(M) = 0$ is

$$
M_+ = x\frac{1}{3\kappa}\left(1 + \sqrt{1 + \tfrac{3}{c}x^{-1}\kappa}\right) \tag{3.19}
$$

$$
\tag{3.20}
$$

and the smaller root is negative. $H(M)$ converges to $-\infty$ when M converges to infinity. Thus, if $M_+ \leq \frac{F_1(q)}{\sqrt{q}}x$ then the supremum of $H(M)$ over $M \geq$

$x\frac{F_1(q)}{\sqrt{q}}$ is $H(\frac{F_1(q)}{\sqrt{q}}x)$ and if $M_+ \geq \frac{F_1(q)}{\sqrt{q}}x$ then the supremum of $H(M)$ over $M \geq x\frac{F_1(q)}{\sqrt{q}}$ is $H(M_+)$. From the expression of M_+ in (3.19) we find that

$$\begin{aligned} &M_+ > \tfrac{F_1(q)}{\sqrt{q}}x \Leftrightarrow \\ &cx < \left(3\tfrac{F_2(q)}{qF_1(q)} - 2\tfrac{F_1(q)}{\sqrt{q}}\right)^{-1} (\approx 0.46). \end{aligned} \tag{3.21}$$

Besides

$$\begin{aligned} H(M_+) &= \frac{x}{3\kappa}(1+\sqrt{1+\tfrac{3\kappa}{cx}x^{-1}}) + c(\frac{x}{3\kappa}(1+\sqrt{1+\tfrac{3\kappa}{cx}}))^2 \\ &\quad - cx^{-1}\kappa(\frac{x}{3\kappa}(1+\sqrt{1+\tfrac{3\kappa}{cx}}))^3 \\ &= (\frac{x}{3\kappa})^2[3\kappa x^{-1} + \tfrac{2}{3}c + (2\kappa x^{-1} + \tfrac{2}{3}c)\sqrt{1+\tfrac{3\kappa}{cx}}]. \end{aligned}$$

Now, we must only relate the two values $H(M_+)$ and $H(x\frac{F_1(q)}{\sqrt{q}})$ to x. First compare x to $H(x\frac{F_1(q)}{\sqrt{q}})$.

$$\begin{aligned} &x \leq H(x\tfrac{F_1(q)}{\sqrt{q}}) \Leftrightarrow \\ &x \leq (x\tfrac{F_1(q)}{\sqrt{q}}) + c(x\tfrac{F_1(q)}{\sqrt{q}})^2 - c\kappa x^{-1}(x\tfrac{F_1(q)}{\sqrt{q}})^3 \Leftrightarrow \\ &cx \leq \frac{\frac{F_1(q)}{\sqrt{q}} - 1}{\frac{F_2(q)}{q\sqrt{q}} - (\frac{F_1(q)}{\sqrt{q}})^2}(\approx 0.424) \end{aligned} \tag{3.22}$$

Then compare x to $H(M_+)$.

$$\begin{aligned} x &\leq H(M_+) \\ 9\kappa^2 &\leq 3\kappa + \tfrac{2}{3}xc + (2\kappa + \tfrac{2}{3}xc)\sqrt{1+\tfrac{3\kappa}{cx}} \end{aligned} \tag{3.23}$$

Let $w = \sqrt{1+3\kappa/(cx)}$. Then $w > 1$ and (3.23) corresponds to

$$0 \leq 2w^3 + (3-9\kappa)w^2 + 9\kappa - 1 \tag{3.24}$$

Solving this numerically it is found that (3.23) corresponds approximately to $cx \notin (0.426, 2.59)$. Combining this with (3.22) and (3.21) we get

$$V(x) = \begin{cases} H(M_+) & cx \leq b^*, \\ x & cx \geq b^*. \end{cases}$$

where $b^* = \frac{3\kappa}{w_+^2 - 1} \approx 0.426$ and w_+ is the highest root in (3.24). Therefore, we find the optimal stopping times $\tau^*(x)$

$$\tau^*(x) = \begin{cases} \tau_{[\bar\lambda(x)q, \bar\lambda(x)/2)]} & cx \le b^*, \\ 0 & cx \ge b^*, \end{cases}$$

with

$$\bar\lambda(x) = \lambda(M_+, x) = \frac{M_+^2}{xF_1(q)^2} = \frac{x(1 + \sqrt{1 + \frac{3\kappa}{cx}})}{9\kappa F_1(q)^2}.$$

□

From Proposition 3.4.2 we deduce

Proposition 3.4.4. *Let Y be given from* (3.15) *and let $X = e^Y$. Then the optimal stopping times of Lemma 3.4.3 are also optimal for the conditional variance problem* (3.2).

Proof. Let

$$F(\beta) = \inf_{\tau : \mathbb{E}_x[X_\tau] = \beta} (\mathbb{E}[X_\tau^2] - \beta^2)$$

Then, for $\beta \ge x\frac{F_1(q)}{\sqrt{q}}$ we have $F(\beta) = x^{-1}\beta^3\kappa - \beta^2$ and

$$F'(\beta) = 3x^{-1}\beta^2\kappa - 2\beta \ge \beta(3\kappa\frac{F_1(q)}{\sqrt{q}} - 2) > 0.$$

For $\beta \in [x, x\frac{F_1(q)}{\sqrt{q}})$ then $F(\beta) = x^2 p(\beta) + x^2 q^{-3/2} F_2(q)(1 - p(\beta)) - \beta^2$ and

$$\begin{aligned} F'(\beta) &= -x\frac{1}{\frac{F_1(q)}{\sqrt{q}} - 1} + x\frac{q^{-3/2}F_2(q)}{\frac{F_1(q)}{\sqrt{q}} - 1} - 2\beta \\ &\ge \frac{x}{\frac{F_1(q)}{\sqrt{q}} - 1}\left(q^{-3/2}F_2(q) - 1 - 2(\frac{F_1(q)}{\sqrt{q}} - 1)\right) > 0. \end{aligned}$$

It is easily verified that F is continuous on $[x, \infty)$, and thus the infimum of F over $[M, \infty)$ is obtained for $\beta = M$. Therefore

$$\inf_{\tau : \mathbb{E}_x[X_\tau] \ge M} \mathbb{V}_x[X_\tau] = \inf_{\beta \ge M}\left(\inf_{\tau : \mathbb{E}_x[X_\tau] = \beta}(\mathbb{E}[X_\tau^2] - \beta^2)\right) = \inf_{\tau : \mathbb{E}_x[X_\tau] = M} \mathbb{E}[X_\tau^2] - M^2,$$

and it follows that the optimal stopping time of Lemma 3.3.3 is also optimal for the conditional variance problem. □

It is interesting that for $M \in (e^y, \frac{F_1(q)}{\sqrt{q}})$ the optimal stopping time we derive is a randomized stopping time. The randomized solutions we know from maximizing variance [2], but it is remarkable that even though we saw in the proof of Theorem 3.4.1 that the randomization between two stopping times can never be optimal for the mean-variance problem, they may still be optimal for the conditional variance problem when the traditional solutions do not work.

It would be interesting to expand the solution to the mean-variance problem to include general geometric Lévy processes with upwards jumps. However, this is particularly difficult because we believe that whenever the problem has an optimal stopping time, this is in general a hitting time of an interval. If it is always possible to find a solution to (3.5) and (3.6) by including randomized stopping times when necessary, then the optimal stopping time to the mean-variance problem for geometric spectrally negative Lévy processes is found among solutions to the classical quadratic problem of (3.5). Thus, the solution is found among solutions to an optimal stopping problem with a gain function which is a concave second order polynomial. It seems likely that the solution to this auxiliary classical optimal stopping problem has a stopping region as an interval around the maximum.

Another reason to expect the solution to be an interval is illustrated through the following example. Let X be a random variable and define another random variable $Z = X1_A + x1_{A^c}$, where x is a real number. Consider the problem of choosing x so to maximize $E[Z] - cVar[Z]$ for some given c. The best choice is never to choose x as high as possible. For high enough values of x the cost from the variance will be higher than the gain from the mean. This is seen from

$$\begin{aligned} E[Z] - cVar[Z] = &\mathbb{E}[X]P(A) - c\mathbb{E}[X^2]P(A) + c\mathbb{E}[X]^2P(A)^2 \\ &- x^2cP(A^c)P(A) + xP(A^c)(1 + c2\mathbb{E}[X]P(A)). \end{aligned}$$

The processes for which the mean-variance problem is interesting all converge to 0. Thus it is always possible to get a lower value by waiting. The calculations above motivates that if the gain is set on some event A, the ideal stopping time on A^c is not to get as high a gain as possible.

3.5 Dynamic Optimization

One of the drawbacks of the mean-variance problem is that it is not time consistent. For this reason it is in [7] suggested to instead look at the corresponding dynamic optimal stopping problem. For the corresponding dynamic problem we search for a stopping time $\tau* \in \mathcal{T}$ such that there is no other stopping time $\sigma \in \mathcal{T}$ with

$$\mathbb{P}_x(\mathbb{E}_{X_{\tau^*}}[X_\sigma] - c\mathbb{V}_{X_{\tau^*}}[X_\sigma] > X_{\tau^*}) > 0. \tag{3.25}$$

When the solution to the traditional (static) optimal stopping problem is known, it is easy to deduce the solution to the corresponding dynamic problem. For the mean-variance problem we get the following result

Corollary 3.5.1. *Assume Y is spectrally negative and define $\phi(0)$ from Y. Consider the dynamic mean-variance problem* (3.25)*.*

a) If $\phi(0) = 0$, then it is never optimal to stop.

b) If $\phi(0) \in (0, 1)$, then it is optimal to stop by $\tau^+_{b^*_c}$, where b^*_c is as defined in Theorem 3.3.1

c) If $\phi(0) \geq 1$, then it is optimal to stop right away.

Assume Y is given from (3.15) *and let $X = e^Y$ with $X_0 = x = e^y$. Consider the dynamic mean-variance problem* (3.25)*. Let b_x be given as in Theorem 3.4.1, then $\tau^+_{[b_x,\infty]}$ is an optimal stopping time.*

References for Chapter 3

[1] BAXTER J. R. AND CHACON, R. V. (1977). Compactness of Stopping Times. *Probab Theory Relat Fields* **40,** 169–181.

[2] GAD, K. S. T. AND PEDERSEN, J. L. (2015). Variance Optimal Stopping for Geometric Lévy Processes. *Adv Appl Probab* To appear.

[3] KYPRIANOU, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications.* Springer.

[4] MARKOWITZ, H. M. (1952). Portfolio selection. *J Financ* **7,** 77–91.

[5] MIKOSCH, T. (2004). *Non-Life Insurance Mathematics.* Springer.

[6] PEDERSEN, J. L. (2011). Explicit solutions to some optimal variance stopping problems. *Stoch* **83,** 505–518.

[7] PEDERSEN, J. L. AND PESKIR, G. (2012). Optimal Mean-Variance Selling strategies. *Research Report No. 12, 2012, Probab. Statist. Group Manchester* .

[8] PESKIR, G. AND SHIRYAEV, A. N. (2006). *Optimal Stopping and Free-Boundary Problems.* Birkhäuser.

[9] SATO, K. I. (2002). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press.

[10] SHIRYAEV, A. N. (1978). *Optimal Stopping Rules.* Springer.

[11] ØKSENDAL, B. AND SULEM, A. (2007). *Applied Stochastic Control of Jump Difusions.* Springer.

4. Rationality Parameter for Exercising American Put

Kamille Sofie Tågholt Gad, Jesper Lund Pedersen

Abstract

In this paper, irrational exercise behavior of the buyer of an American put is characterized by a single parameter. We model irrational exercise rules as the first jump time of a point processes with stochastic intensity. By the rationality parameter, we parameterize a family of stochastic intensities that depends on the value of the put itself. We then give a probabilistic proof that the value of the put using the irrational exercise rule converges to the arbitrage-free price as the rationality parameter converges to infinity. Another application of this result, is the penalty method for approximating the price of an American put.

Keywords: Behavioural modelling; Optimal stopping; Partial differential equation; Penalty method.

4.1 Introduction

The buyer of an American put can exercise at any time of his choice within the time of the contract. The arbitrage-based theory for the pricing of the American put is formulated as an optimal stopping problem (see Karatzas & Shreve [8]), where the optimal stopping time is an optimal exercise rule for the buyer of the American put. The buyer's exercise behaviour is called optimal (in this paper) if he follows the rules of the optimal exercise strategy. Empirical studies in Diz & Finucane [5] and Poteshman & Serbin [10] show that there are a large number of irrational exercises. The irrational exercises may have various reasons. For example, the irrationality may be due to that the buyer does not have the correct input for the model, he does not monitor his position sufficiently, or he holds the American put as part of a hedge where it might not be optimal to apply the optimal exercise rule. The irrational exercises might then tend to cause over-valuation of the American put.

In the present paper we develop a methodology that takes irrational exercise behaviour into consideration. We use an intensity-based model in which the exercise rule is modelled as the first jump of a point process with stochastic intensity. We let the exercise intensity depend on the endogenous parameter of how profitable it is to exercise. This profitability we measure as the difference between the pay-off and the value of the put if it is not exercised yet. In line with the game-theoretical approach of irrational decision making (see e.g. Chen et al. [3]), we characterize the rationality of the buyer of the American put by a parameter. This parameter measures how strongly the exercise intensity is affected by the profitability, and for that reason we denote it *the rationality parameter*. The main contribution of the present paper is a probabilistic proof of the following convergence result: Under mild and reasonable restrictions the price of the put under the intensity-based model converges to the classical arbitrage-free price when the rationality parameter converges to infinity. This proof breaks down the price of the put under the intensity-based model into the arbitrage-free price and losses coming from respectively exercising when it is not optimal and not exercising when it is optimal.

An intensity based approach has been used for valuation of executive stock options by e.g. Jennergren & Naslund [7] and by Carr & Linetsky [2]. In the latter paper the intensity of exercising depends on time and the underlying stock. Dai et al. [4] model the mortgagor's prepayment in mortgage loans and the issuer's call in the American warrant as an event risk where the intensity of prepayment or calling depends on the value and may be view as one of the examples in this paper. Moreover, as also pointed out by Dai et al. [4] the valuation equations may be viewed as the penalty method (see Forsyth & Vetzal [6]) for approximating the value of the American put.

The paper is structured in the following way. In Section 4.2, we introduce the rationality parameter for exercising the put and show the convergence result that the value of the put under irrational exercise conferges to the arbitrage-free price when the rationality parameter converges to infinity. In Section 4.3, we derive valuation equations for the put under the exercise strategies considered in Section 4.2 and we set up sufficient conditions for the strategies in Section 4.2 to be well defined.

4.2 Rationality Parameter for Exercising

We assume a Black-Scholes market where the underlying stock price satisfies the stochastic differential equation (under the risk-neutral probability measure)

$$dS_u = rS_u\, du + \sigma S_u\, dW_u$$

for $u \geq t$ with $S_t = s$ under $\mathbb{P}_{t,s}$. Here r is a constant interest rate, $\sigma > 0$ is a constant volatility, and W is a Brownian motion.

Consider an American put with strike price K and maturity date T, written on the stock and thus the payoff process is given by $(K - S_t)^+$. The arbitrage-free value, P^A, of the American put is given as an optimal stopping problem

$$\begin{aligned} P^A(t,s) &= \sup_{t \leq \tau \leq T} \mathbb{E}_{t,s}[e^{-r(\tau - t)}(K - S_\tau)^+] \\ &= \mathbb{E}_{t,s}[e^{-r(\tau^* - t)}(K - S_{\tau^*})^+] \end{aligned}$$

where the supremum is taken over all exercise rules (stopping times) with values in $[t, T]$. Furthermore, there is an optimal exercise rule τ^* for which the supremum is attained. This optimal exercise rule has a stopping boundary $b(\cdot)$ such that it is given by

$$\tau^* = \inf\{t \leq u \leq T : S_u \leq b(u)\}.$$

We define irrational exercise rules τ, as the minimum of first jump time of a point process with stochastic intensity $(\mu_u)_{t \leq u \leq T}$ (see Bremaud [1]) and the terminal time T. The value of the American put exercising at time τ is given by

$$P(t,s;\tau) = \mathbb{E}_{t,s}[e^{-r(\tau - t)}(K - S_\tau)^+].$$

We wish to model that despite the holder of the American put does not exercise according to the optimal exercise strategy, then whether he exercises the option is at any time affected by how profitable it is to exercise. The profitability we measure as the difference between the pay-off and the value of the put if he does not exercise immediately. Thus, for exercise strategy τ the profitability at time t is given by $(K - S_t)^+ - P(t, S_t, \tau)$. The relation between the profitability and the stochastic exercise intensity is given by an intensity function, $f : [-K, K] \to [0, \infty)$ by

$$\mu_u = f\big((K - S_u)^+ - P(u, S_u; \tau)\big).$$

For modelling purposes it is reasonable to require the function f to be increasing. Conditions on f for the circular definition to by well defined are given is Section 4.3.

We now introduce a single, strictly positive parameter, θ, that measures how rational the behaviour of the holder of the put is. This is done in the following way: We let θ be the index of a family of intensity functions, f^θ and thereby a family of exercise strategies, τ^θ. We wish the corresponding values of the put converge to the arbitrage-free price when the parameter θ converges to infinity. This gives us the definition

Definition 4.2.1. *Let $(\tau^\theta)_{\theta>0}$ be a family of irrational exercise rules indexed by $\theta > 0$ and denote the corresponding values by*

$$P^\theta(t,s) = P(t,s;\tau^\theta).$$

We say θ is a rationality parameter for exercising if $P^\theta(t,s) \to P^A(t,s)$ for $\theta \to \infty$.

In the case that the exercise intensity depends of the value of the put itself,

Theorem 4.2.1 below is the main result of this paper. It gives sufficient conditions for an index of a family of intensity functions to be a rationality parameter. The proof consists of a probabilistic analysis of the exercise strategies. The key idea is to define events that categorize how profitable an exercise strategy turn out to be upon exercise. Given some tolerance, $\varepsilon_1 > 0$, we use the following definition

$$\begin{aligned} \{\tau \text{ good}\} &= \{(K - S_\tau)^+ - P(\tau, S_\tau; \tau) \geq 0\} \\ \{\tau \text{ ok}\} &= \{(K - S_\tau)^+ - P(\tau, S_\tau;) \in [-\varepsilon_1, 0)\} \\ \{\tau \text{ bad}\} &= \{(K - S_\tau)^+ - P(\tau, S_\tau) < -\varepsilon_1\}. \end{aligned}$$

Theorem 4.2.1. *Let $(f^\theta)_{\theta>0}$ be a family of positive, deterministic intensity functions and for each $\theta > 0$, let a stochastic intensity process be given by*

$$\mu_u^\theta = f^\theta\big((K - S_u)^+ - P^\theta(u, S_u)\big)$$

where $P^\theta(t,s) = P(t,s;\tau^\theta)$ and τ^θ is the strategy of the put exercised at the first jump of the point process with intensity μ^θ.

Let $\nu_\theta(x) = 1_{(x<0)} \sup_{y \leq x} f^\theta(y) + 1_{(x \geq 0)} \sup_{y \geq x} f^\theta(y)$ and suppose that

- *$\nu_\theta(0+) \to \infty$ as $\theta \to \infty$.*
- *There exists a function $\varepsilon : (0,\infty) \to (0,\infty)$ such that $\nu_\theta(-\varepsilon(\theta)) \to 0$ and $\varepsilon(\theta)\nu_\theta(0-) \to 0$ as $\theta \to \infty$.*

Then θ is a rationality parameter, that is, for every $(t,s) \in [0,T] \times \mathbb{R}^+$ we have that $P^\theta(t,s) \to P^A(t,s)$ as $\theta \to \infty$.

Remark 4.2.1. *If we include the natural restriction that that f^θ is increasing, then $f^\theta = \nu_\theta$.*

Proof. 1. Let τ_n^θ be the sequence of jump times of the point process with the stochastic intensity process μ^θ. Note that $\tau_1^\theta = \tau^\theta$. Let $\hat{\tau}^\theta$ be the minimum of T and the first jump time after the rational exercise rule τ^*, that is,

$$\hat{\tau}^\theta = \sum_{i=1}^{\infty} \tau_i^\theta 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)}$$

with $\tau_0^\theta = t$ under $\mathbb{P}_{t,s}$. The value of the put exercising at this jump time

$$\begin{aligned} P(t,s;\hat{\tau}^\theta) &= \mathbb{E}_{t,s}\big[e^{-r(\hat{\tau}^\theta - t)}(K - S_{\hat{\tau}^\theta})^+\big] \\ &= \sum_{i=1}^{\infty} \mathbb{E}_{t,s}\big[e^{-r(\tau_i^\theta - t)}(K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \le \tau_i^\theta)}\big]. \end{aligned} \tag{4.1}$$

The strategy $\hat{\tau}^\theta$ corresponds to the strategy τ^θ where we have removed the chance of the holder of the put exercising to early. Studying $\hat{\tau}^\theta$ we may separate the loss from exercising to early from the loss of exercising too late.

We first study the loss of exercising too early. The overall idea of this part is the following. The starting point is the representation of $P(t,s;\hat{\tau}^\theta)$ from (4.1). Through the representation (4.1) it follows that the strategy $\hat{\tau}^\theta$ corresponds to the strategy τ^θ, except each time there is a jump in the point process before τ^* the holder of the put regrets and do not exercise. At each time of regret the holder looses some value if the exercise time was good, and he gains at most ε_1 if the exercise time was ok, and if the exercise time was bad then he gains more that ε_1. As the exercise intensity in times which are ok or bad are sufficiently restricted, then the expected value one gains from exercising at a time which is ok may be made arbitrarily small by using a small ε_1. No matter when the exercise is one cannot gain more than K on an exercise, and given an arbitrary ε_1 the intensity for exercising at bad times can be made uniformly arbitrarily small by choosing a large θ. Then the gain from regretting the exercises when τ is bad can be made arbitrarily small.

2. We verify the following inequality. For given $\varepsilon_1 > 0$ and $n \in \mathbb{N}$ then

$$\begin{aligned} P^\theta(t,s) \ge{}& \mathbb{E}_{t,s}[e^{-r(\tau_n^\theta - t)} P^\theta(\tau_n^\theta, S_{\tau_n^\theta}) 1_{((\tau_n^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}] \\ &+ \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)}(K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \le \tau_i^\theta)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1})}] \\ &- \varepsilon_1 \sum_{i=1}^{n} \mathbb{P}_{t,s}(\tau_i^\theta < \tau^*, (\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1}, \tau_i^\theta \text{ ok}). \end{aligned} \tag{4.2}$$

We show this by induction. For $n = 1$ we have that

$$\begin{aligned}
P^\theta(t,s) &= \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau_1^\theta < \tau^*)}] + \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau^* \le \tau_1^\theta)}] \\
&\ge \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau_1^\theta < \tau^*)} 1_{(\tau_1^\theta \text{ good})}] \\
&\quad + \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau_1^\theta < \tau^*)} 1_{(\tau_1^\theta \text{ ok})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau^* \le \tau_1^\theta)}] \\
&\ge \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)} P^\theta(\tau_1^\theta, S_{\tau_1^\theta}) 1_{(\tau_1^\theta < \tau^*)} 1_{(\tau_1^\theta \text{ good})}] \\
&\quad + \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(P^\theta(\tau_1^\theta, S_{\tau_1^\theta}) - \varepsilon_1) 1_{(\tau_1^\theta < \tau^*)} 1_{(\tau_1^\theta \text{ ok})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau^* \le \tau_1^\theta)}] \\
&= \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)} P^\theta(\tau_1^\theta, S_{\tau_1^\theta}) 1_{(\tau_1^\theta < \tau^*)} 1_{(\tau_1^\theta \text{ good or ok})}] \\
&\quad - \varepsilon_1 \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)} 1_{(\tau_1^\theta < \tau^*)} 1_{(\tau_1^\theta \text{ ok})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau^* \le \tau_1^\theta)}] \\
&\ge \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)} P^\theta(\tau_1^\theta, S_{\tau_1^\theta}) 1_{(\tau_1^\theta < \tau^*)} 1_{(\tau_1^\theta \text{ good or ok})}] \\
&\quad - \varepsilon_1 \mathbb{P}_{t,s}(\tau_1^\theta < \tau^*, \tau_1^\theta \text{ ok}) \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_1^\theta - t)}(K - S_{\tau_1^\theta})^+ 1_{(\tau^* \le \tau_1^\theta)}]
\end{aligned}$$

We assume that the inequality is true for n. Then we have that

$$\begin{aligned}
&\mathbb{E}_{t,s}[e^{-r(\tau_n^\theta - t)} P^\theta(\tau_n^\theta, S_{\tau_n^\theta}) 1_{(\tau_n^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}] \\
&\quad = \mathbb{E}_{t,s}[e^{-r(\tau_n^\theta - t)} \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - \tau_n^\theta)}(K - S_{\tau_{n+1}^\theta})^+ | \tau_n^\theta, S_{\tau_n^\theta}] 1_{(\tau_n^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}] \\
&\quad = \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)}(K - S_{\tau_{n+1}^\theta})^+ 1_{(\tau_n^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}] \\
&\quad \ge \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)}(K - S_{\tau_{n+1}^\theta})^+ 1_{(\tau_{n+1}^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})} 1_{(\tau_{n+1}^\theta \text{ good})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)}(K - S_{\tau_{n+1}^\theta})^+ 1_{(\tau_{n+1}^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})} 1_{(\tau_{n+1}^\theta \text{ ok})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)}(K - S_{\tau_{n+1}^\theta})^+ 1_{(\tau_n^\theta < \tau^* \le \tau_{n+1}^\theta)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}] \\
&\quad \ge \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)} P^\theta(\tau_{n+1}^\theta, S_{\tau_{n+1}^\theta}) 1_{(\tau_{n+1}^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})} 1_{(\tau_{n+1}^\theta \text{ good})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)}(P^\theta(\tau_{n+1}^\theta, S_{\tau_{n+1}^\theta}) - \varepsilon_1) 1_{(\tau_{n+1}^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})} 1_{(\tau_{n+1}^\theta \text{ ok})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)}(K - S_{\tau_{n+1}^\theta})^+ 1_{(\tau_n^\theta < \tau^* \le \tau_{n+1}^\theta)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}] \\
&\quad = \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)} P^\theta(\tau_{n+1}^\theta, S_{\tau_{n+1}^\theta}) 1_{(\tau_{n+1}^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n+1})}] \\
&\qquad - \varepsilon_1 \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)} 1_{(\tau_{n+1}^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})} 1_{(\tau_{n+1}^\theta \text{ ok})}] \\
&\qquad + \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)}(K - S_{\tau_{n+1}^\theta})^+ 1_{(\tau_n^\theta < \tau^* \le \tau_{n+1}^\theta)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}].
\end{aligned}$$

Thus, using the induction assumption we find

$$
\begin{aligned}
P^\theta(t,s) \geq\ & \mathbb{E}_{t,s}[e^{-r(\tau_n^\theta - t)} P^\theta(\tau_n^\theta, S_{\tau_n^\theta}) 1_{(\tau_n^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n})}] \\
& + \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1})}] \\
& - \varepsilon_1 \sum_{i=1}^{n} \mathbb{P}_{t,s}(\tau_i^\theta < \tau^*, (\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1}, \tau_i^\theta \text{ ok}) \\
\geq\ & \mathbb{E}_{t,s}[e^{-r(\tau_{n+1}^\theta - t)} P^\theta(\tau_{n+1}^\theta, S_{\tau_{n+1}^\theta}) 1_{(\tau_{n+1}^\theta < \tau^*)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,n+1})}] \\
& + \sum_{i=1}^{n+1} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1})}] \\
& - \varepsilon_1 \sum_{i=1}^{n+1} \mathbb{P}_{t,s}(\tau_i^\theta < \tau^*, (\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1}, \tau_i^\theta \text{ ok}).
\end{aligned}
$$

Hence we have shown inequality (4.2).

3. Now we investigate the second term in (4.2)

$$
\begin{aligned}
& \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)} 1_{((\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1})}] \\
&= \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)}] \\
&\quad - \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)} 1_{(\exists j \in \{1,\dots,i-1\}: \tau_j^\theta \text{ bad})}] \\
&\geq \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)}] \\
&\quad - K \sum_{i=1}^{n} \mathbb{P}_{t,s}(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta, \exists j \in \{1,\dots,i-1\} : \tau_j^\theta \text{ bad}) \\
&\geq \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)}] - K \mathbb{P}_{t,s}(\exists i \in \mathbb{N} : \tau_i^\theta < \tau^*, \tau_i^\theta \text{ bad}) \\
&\geq \sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)} (K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \leq \tau_i^\theta)}] - K(1 - e^{-(T-t)\nu_\theta(-\varepsilon_1)}).
\end{aligned}
$$

Note that given $\varepsilon_1 > 0$ the latter term can be made arbitrarily small by choosing θ large. This means that with large θ there is small probability that the option with price $P(t,s;\hat{\tau}^\theta)$ has an exercise time which contains regrets of bad jump times.

Next we investigate the third term in (4.2)

$$\begin{aligned}\sum_{i=1}^{n} &\mathbb{P}_{t,s}(\tau_i^\theta < \tau^*, (\tau_j^\theta \text{ good or ok})_{j=1,\dots,i-1}, \tau_i^\theta \text{ ok}) \\ &\le \mathbb{E}_{t,s}(\sharp\{i \in \mathbb{N} : \tau_i^\theta < \tau^*, \tau_i^\theta \text{ ok}\}) \\ &\le (T-t)\nu_\theta(0-).\end{aligned}$$

The latter inequality follows from that the ok jump times at most occur with intensity $\nu_\theta(0-)$ in the time until T. This shows that the expected number of regrets of ok stopping times for the exercise time of the option with price $P(t,s;\hat{\tau}^\theta)$ is uniformly bounded with respect to ε_1. Therefore the contribution from here can be made arbitrarily small by making ε_1 small, as ε_1 is then an upper bound for the contribution for every regret of an ok stopping time. Combined we get:

$$\begin{aligned}P^\theta(t,s) \ge &\sum_{i=1}^{n} \mathbb{E}_{t,s}[e^{-r(\tau_i^\theta - t)}(K - S_{\tau_i^\theta})^+ 1_{(\tau_{i-1}^\theta < \tau^* \le \tau_i^\theta)}] \\ &- K(1 - e^{-(T-t)\nu_\theta(-\varepsilon_1)}) - \varepsilon_1(T-t))\nu_\theta(0-).\end{aligned}$$

As this holds for all $n \in \mathbb{N}$ we find

$$P^\theta(t,s) - P(t,s;\hat{\tau}^\theta) \ge -K(1 - e^{-(T-t)\nu_\theta(-\varepsilon_1)}) - \varepsilon_1(T-t)\nu_\theta(0-)$$

4. We now investigate the losses from too late exercise. Let

$$\sigma_{\varepsilon_2} = \inf\{u \ge \tau^* : |(K - S_u)^+ e^{-r(u-t)} - (K - S_{\tau^*})^+ e^{-r(\tau^* - t)}| \ge \varepsilon_2\},$$

and let

$$\begin{aligned}C_{\varepsilon_2} &= \{u \in [\tau^*, \sigma_{\varepsilon_2}] | (K - S_u)^+ - P^A(u, S_u) = 0\} \\ &= \{u \in [\tau^*, \sigma_{\varepsilon_2}] | S_u \le y_u\}.\end{aligned}$$

Let $\mathcal{L}$ denote the Lebesgue measure. As the rational exercise boundary $u \mapsto b(u)$ is increasing and S_u is a geometric Brownian motion, then $\mathcal{L}(C_{\varepsilon_2}) > 0$ almost surely for every $\varepsilon_2 > 0$. Hence for every $\varepsilon_2, \varepsilon_3 > 0$ there exists a $\delta \ge 0$

such that $\mathbb{P}_{t,s}(\mathcal{L}(C_{\varepsilon_2}) > \delta) > 1 - \varepsilon_3$. Now we get

$$\begin{aligned}
P^A(t,s) - P(t,s;\hat{\tau}^\theta) &= \mathbb{E}_{t,s}[e^{-r(\tau^*-t)}(K-S_{\tau^*})^+ - e^{-r(\hat{\tau}^\theta-t)}(K-S_{\hat{\tau}^\theta})^+] \\
&= \mathbb{E}_{t,s}[(e^{-r(\tau^*-t)}(K-S_{\tau^*})^+ - e^{-r(\hat{\tau}^\theta-t)}(K-S_{\hat{\tau}^\theta})^+)1_{\hat{\tau}^\theta \le \sigma_{\varepsilon_2}}] \\
&\quad + \mathbb{E}_{t,s}[(e^{-r(\tau^*-t)}(K-S_{\tau^*})^+ - e^{-r(\hat{\tau}^\theta-t)}(K-S_{\hat{\tau}^\theta})^+)1_{\hat{\tau}^\theta > \sigma_{\varepsilon_2}}] \\
&\le \varepsilon_2 + K\mathbb{P}_{t,s}(\hat{\tau}^\theta > \sigma_{\varepsilon_2}) \\
&= \varepsilon_2 + K(\mathbb{P}_{t,s}(\hat{\tau}^\theta > \sigma_{\varepsilon_2}, \mathcal{L}(C_{\varepsilon_2}) > \delta) + \mathbb{P}_{t,s}(\hat{\tau}^\theta > \sigma_{\varepsilon_2}, \mathcal{L}(C_{\varepsilon_2}) \le \delta)) \\
&\le \varepsilon_2 + K(\mathbb{P}_{t,s}(\hat{\tau}^\theta > \sigma_{\varepsilon_2} | \mathcal{L}(C_{\varepsilon_2}) > \delta) + \mathbb{P}_{t,s}(\mathcal{L}(C_{\varepsilon_2}) \le \delta)) \\
&\le \varepsilon_2 + K(e^{-\delta\nu_\theta(0+)} + \varepsilon_3).
\end{aligned}$$

Thus

$$P^A(t,s) - P^\theta(t,s) \le \varepsilon_2 + K(e^{-\delta\nu_\theta(0+)} + \varepsilon_3) + K(1 - e^{-(T-t)\nu_\theta(-\varepsilon_1)}) + \varepsilon_1(T-t)\nu_\theta(0-).$$

Choose $\varepsilon_1 : \mathbb{R}^+ \to \mathbb{R}^+$ such that $f^\theta(-\varepsilon_1(\theta)) \to 0$ as $\theta \to \infty$ and such that $\varepsilon_1(\theta) f^\theta(0+) \to 0$ as $\theta \to \infty$. Then we find that $P^\theta(t,s) \to P^A(t,s)$, when $\theta \to \infty$. □

4.3 Valuation Equations

In this section, we use the set-up in the previous section to obtaining valuation equations for the put under irrational exercise.

Consider an irrational exercise rule, τ which is given as the first jump time of a point process with stochastic intensity $\mu_u = \lambda(u, S_u)$ for some positive, deterministic, measurable function λ. As in the previous section we have that the value of the put is given by the risk-neutral expectation

$$\begin{aligned}
P^\mu(t,s) &= \mathbb{E}_{t,s}[e^{-r(\tau-t)}(K-S_\tau)^+] \\
&= e^{-r(T-t)}\mathbb{E}[(K-S_T)^+ 1_{(\tau \ge T)}] + \mathbb{E}_{t,s}[e^{-r\tau}(K-S_\tau)^+ 1_{(\tau < T)}]
\end{aligned}$$

By [9, Proposition 3.1], the expectation can be re-written to

$$\begin{aligned}
P^\mu(t,s) &= e^{-r(T-t)}\mathbb{E}_{t,s}\Big[\exp\Big(-\int_t^T \lambda(u,S_u)\,du\Big)(K-S_T)^+\Big] \\
&\quad + \int_t^T e^{-r(u-t)}\mathbb{E}_{t,s}\Big[\lambda(u,S_u)\exp\Big(\int_t^u \lambda(v,S_v)\,dv\Big)(K-S_u)^+\Big]\,du.
\end{aligned}$$

By the Feynman-Kac (see [8]), the value of the put P^μ is the solution for the partial differential equation

$$\begin{aligned}
P_t^\mu(t,s) + rsP_s^\mu(t,s) + \tfrac{1}{2}\sigma^2 s^2 P_{ss}^\mu(t,s) \\
= rP^\mu(t,s) + \lambda(t,s)\big((K-s)^+ - P^\mu(t,s)\big). \qquad (4.3)
\end{aligned}$$

with $P^{\mu}(T,s) = (K-s)^+$, whenever this partial differential equation has a unique solution.

Now, given some intensity function $f : [-K, K] \to [0, \infty)$ consider the partial differential equation

$$\begin{aligned} p_t(t,s) + rsp_s(t,s) + \tfrac{1}{2}\sigma^2 s^2 p_{ss}(t,s) \\ = rp(t,s) + f\big((K-s)^+ - p(t,s)\big)\big((K-s)^+ - p(t,s)\big). \end{aligned} \tag{4.4}$$

with $p(T,s) = (K-s)^+$. If this partial differential equation has a unique solution, then use this solution to define $\lambda(t,s) = f((K-s)^+ - p(t,s))$. Consider (4.3) for this function λ. Now the solution p of (4.4) must fulfill (4.3). Thus, if the solution to (4.3) is unique, then from the above arguments we get the following. p is the value of a put which is exercised by a strategy which is given by the the minimum of T and first jump time of a point process with intensity $\mu_u = \lambda(t, S_u)$. By the construction of λ we have $\mu_u = f((K-S_u)^+ - p(u, S_u))$. Thus, the existence and uniqueness of the solutions to (4.3) and (4.4) ensures that the strategies of Section 4.2 are well defined.

Finally, we suggest two simple specifications of the function f^{θ} given in Theorem 2.2. In the first example the function is specified as follows

$$f^{\theta}(x) = \theta e^{\theta^2 x}.$$

In the second example the function is specified by

$$f^{\theta}(x) = \begin{cases} \theta & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases}$$

Related to this second function is the penalty method found in recent computational finance literature (see e.g. [6]) which approximate the rational value of the American by semi-linear PDE (4.3).

References for Chapter 4

[1] Brémaud, P. (1981). *Point processes and queues.* Springer.

[2] Carr, P. and Linetsky, T. (2000). The valuation of executive stock options in an intensity-based framework. *Eur Financ Rev* **4,** 211–230.

[3] Chen, H.-C., Friedman, J.W. and Thisse, J. F. (1997). Boundedly rational Nash equilibrium: A probabilistic choice approach. *Games Econ Behav* **18,** 32–54.

[4] Dai, M., Kwok, Y.K. and You, H. (2007). Intensity-based framework and penalty formulation of optimal stopping problems. *J Econ Dyn Control* **31,** 3860–3880.

[5] Diz, F. and Finucane, T. (1993). The rationality of early exercise decisions: Evidence from the S&P 100 index options market. *Rev Financ Stud* **6,** 765–798.

[6] Forsyth, P.A. and Vetzal, K.R. (2002). Quadratic convergence for valuing American options using a penalty method. *SIAM J Sci Comput* **23,** 2096–2123.

[7] Jennergren, L. and Naslund, B. (1993). A comment on the valution of executive stock options and the FASB proposal. *The Account Rev* **68,** 179–183.

[8] Karatzas, I. and Shreve, S.E. (1998). *Methods of mathematical finance.* Springer.

[9] Lando, D. (1998). On Cox processes and credit risky securities. *Rev Derivatives Res* **2,** 99–120.

[10] Poteshman, A.M. and Serbin, V. (2003). Clearly irrational financial market behavior: evidence from the early exercise of exchange traded stock options. *J Financ* **58,** 37–70.

5. Reserve-Dependent Surrender

Kamille Sofie Tågholt Gad, Jeppe Juhl, Mogens Steffensen

Abstract

We study the modelling and valuation of surrender and other behavioural options in life insurance and pension. We place ourselves in between the two extremes of completely arbitrary intervention and optimal intervention by the policyholder. We present a method that is based on differential equations and that can be used to approximate contract values when policyholders exhibit optimal behaviour. This presentation includes a specification of sufficient conditions for both consistency of the model and convergence of the contract values. When not going to the limit in the approximation we obtain a technique for balancing off arbitrary and optimal behaviour in a simple, intuitive way. This leads to our suggestions for intervention models where one single parameter reflects the extent of rationality among policyholders. When expenses are taken into account we loose the duality between the policyholder's valuation of the contract and the market reserve. We include this in our model, and we give an upper bound for the difference between the market reserve when the policyholder exhibit optimal behaviour and the worst case market reserve from the pension fund point of view. In a series of numerical examples we illustrate the impact of the rationality parameter on the contract values.

Keywords: Behavioural option; ordinary differential equation; penalty method; optimal stopping; Solvency II.

5.1 Introduction

Modern solvency and accounting rules (Solvency II and IFRS) require that expected policyholder behaviour is taken into account. In Article 79 of the Solvency II Directive [17] it is stated that

> *"... Any assumptions made by insurance and reinsurance undertakings with respect to the likelihood that policy holders will exercise contractual options, including lapses and surrenders, shall be realistic and based on current and credible information... "*

It is further stated in Article 26 of Commission Delegated Regulation [16] for Solvency II that

> *"... That analysis shall take into account all of the following:*
>
> *(a) how beneficial the exercise of the options was and will be to the policy holders under circumstances at the time of exercising the option;*
>
> *(b) the influence of past and future economic conditions; ... "*

In IFRS in B63 [12] it is stated that

> *"... The measurement of an insurance contract shall reflect, on an expected value basis, the entity's view of how the policyholders in the portfolio that contains the contract will exercise options available to them, and the risk adjustment shall reflect the entity's view of how the actual behaviour of the policyholders in the portfolio of contracts may differ from the expected behaviour... "*

Thus, the expected policyholder behaviour is supposed to take into account both the economic conditions under which the behaviour takes place as well as the extent to which intervention is to the benefit of the policyholder. These factors may change over time. Therefore, one should properly speak of dynamic behaviour models when formalizing these effects in the actuarial valuation formulas. Changing economic conditions could e.g. be a changing level of interest rates, and one idea would be to let the intensity or probability of intervention depend on the current (possibly stochastic) level of interest rates. How beneficial an action is can be formalized by the gain from intervention. Determining the gain may be a delicate issue since both intervening and not intervening opens up for new intervention options in the future that also have to be taken into account. E.g., not to surrender typically opens up for surrendering later, and transcription into free policy changes the effect of the surrender option. This challenge calls for a recursive solution such that the gain is always measuring correctly the tendencies of intervening in the future. We disregard the economic condition by assuming deterministic interest rates and focus on the latter idea of a recursive formula to deal with the benefit of intervention. One motivation for this focus is that, perhaps, the external economic conditions are supposed to approximate to the internal benefit.

There exists a range of approaches to modelling of behavioural risk. One extreme is to say that intervention occurs in a completely arbitrary way, like insurance risk. We hereby mean that we model the behaviour as independent of everything else in our model than the state of the policyholder and the time measured through calendar time, the policyholder's age, time since initiation, or time to (deterministic) retirement. Specifically, the behaviour depends on neither the contract the policyholder holds nor the interest rate. With this approach it is tractable to study various aspects beyond just adding surrender to a survival model. Buchardt et al. [3] studied the formalistic interaction between semi-Markov modelling of insurance risk and behavioural risk, including duration dependence of mortality and payments in the disability state and recognizing duration dependence of free policy payments. A simpler exposition is found in Buchardt and Møller [2]. Henriksen et al. [11] also combine surrender and free policy options and study the impact on reserving from different simplifying assumptions about the dependence between insurance risk and behaviour risk.

Another extreme is to say that intervention occurs in a completely rational, optimal way. We hereby mean that the policyholder, who is assumed to have the same information as the pension fund has, intervenes according to a strategy that maximizes the value of her contract. If we assume duality between the policyholder's and the pension fund's valuation of the contract, then this approach gives worst case reserves for the pension fund. The approach was taken in Steffensen [18], who derived general variational inequalities that characterize the reserve in case of a multi-state Markov model for insurance risk and a multi-state model for behavioural risk. In the surrender case, this is known as American option pricing of surrender risk. Other early references based on this approach to surrender risk are Grosen and Jørgensen [10] and Bacinello [1].

In between these extremes exist all different kinds of models where intervention is modelled by an intensity, but where the intensity not only depends on time but also some stochastic factors. The dependence on the interest rate appears obvious and is thoroughly examined by De Giovanni [9], who calculate reserves by solving partial differential equations numerically. There exists a large amount of literature examining relevant explanatory variables but since these studies appear somewhat marginal to our approach we refer to Eling and Kiesenbauer [5] and references therein for a comprehensive literature overview. See also Gatzert [8] and references therein for an overview on approaches to policyholder options.

Rather than letting the intensity depend on external factors, one could let the intensity depend on internal factors relevant to the specific policy. That could e.g. be to take the difference between the surrender value and (some notion of the) reserve as a measure of how beneficial an intervention is. If the reserve compared with the surrender value does not take future intervention options into account, the calculation can be split up in two standard exercises: First, calculate the reserve without intervention and then plug this reserve into the intensity for a calculation including surrender. If the reserve compared with the surrender value does take future intervention into account, the (usually) linear Thiele differential equation characterizing the reserve becomes in general a non-linear differential equation. Reserve-dependent changes to Thiele's differential equation has been studied before by Christiansen et al. [4]. However, whereas the changes to Thiele's differential equation studied in [4] are linear changes in the benefits and costs, then the change in our model comes from the intensity of the surrender event that contains a non-linear function of the reserve itself. The rationale for this paper is to take a thorough look at this non-linear differential equation in order to motivate it, interpret it, generalize it, and solve it numerically. Also, we present a probabilistic proof of and clarify sufficient conditions for a convergence result that may seem intuitively clear: If the tendency to intervene tends to zero whenever the gain from intervention is negative and tends to infinity whenever the gain from intervention is positive, we reach in the limit at the reserve based on completely rational behaviour. We establish sufficient convergence of intensities to reach such a conclusion. When expenses are not taken into account we find it reasonable to assume duality between the policyholder's and the pension fund's valuation of the contract. Thus, the reserve based on completely rational behaviour equals the reserve based on the policyholder behaving according to the worst case strategy from the pension fund point of view. If expenses are included in the model we lose this duality. We show how expenses may be taken into account in our model, and we give an upper bound for the difference between the reserve based on completely rational behaviour and the reserve based on the policyholder behaving according to the worst case strategy from the pension fund point of view. Thus, our approach to intervention option pricing has two purposes: First, it represents in itself a relevant approach in between the two extremes that, certainly, takes into account the extent to which intervention is to the benefit of the policyholder. Second, for simple parametric forms of the intensity, our calculation may approximate the largest possible liability. As such it can be used as a worst-case or stress calculation with respect to surrender risk.

The idea of approximating the maximum value by a series of solutions to differential equation has been known as the penalty method. In computational finance it has been used as an approximation method for American option pricing. In Forsyth and Vetzal [6] the penalty method is compared with alternative techniques for pricing of the American put option. In Gad and Pedersen [7] the modelling of non-rational option holder behaviour is studied in a way similar to what is done here. The contribution of the present paper is three-fold: First, we introduce, to the knowledge of the authors, for the first time the penalty method in intervention option pricing in insurance. Second, we prove sufficient conditions for the convergence to hold. Third, we do not only think of the intensity model as a means of approximating the largest value, but as a highly relevant approach to general intervention option pricing, useful in accounting and solvency. The approach balances arbitrariness and benefit in a simple form, and in some examples we catch the notion of rationality in one single parameter.

5.2 Standard Setup

Consider a model with a policyholder who is either alive (active) or dead. We assume the state of the policyholder is governed by a state process with a deterministic, continuous mortality intensity, $\mu(t)$, see Figure 5.1. Let I be the process indicating whether the policyholder is alive, and let N be the process counting the numbers of deaths of the policyholder. The policyholder

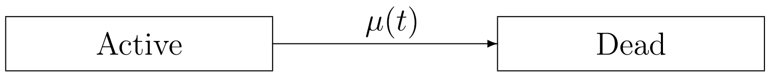

Figure 5.1: Standard survival model.

is assumed to have the following simple contract. She pays a deterministic premium with continuous intensity $\pi(t)$ until a terminal time, n, as long as she is alive. If she is alive at time n she receives a deterministic pension sum $b_a(n)$, and if she dies before then upon death she gets a deterministic death sum, $b_{ad}(t)$. Thus, the accumulated payments in the time interval $[0,t]$ is given by the following process of accumulated payments:

$$B(t) = B(0) - \int_0^t \pi(u)I(u)du + \int_0^t b_{ad}(u)dN(u) + I(n)b_a(n)1_{(t\geq n)},$$

for $t \in [0,n]$. We assume that the market offers a deterministic, continuous interest rate, $r(t)$. We introduce the reserve corresponding to the policyholder being active as the conditional expected present value of future payments,

$$V(t) = \mathbb{E}\left[\int_t^n e^{-\int_t^u r(\tau)d\tau} dB(u) \middle| I(t) = 1\right].$$

We then know, e.g. from [14], that the reserve, V, is continuously differentiable on $[0,n)$ and that it is the solution to Thiele's differential equation,

$$V'(t) = r(t)V(t) + \pi(t) - \mu(t)(b^{ad}(t) - V(t)), \tag{5.1}$$

with $V(n-) = b_a(n)$.

We now add to our model the possibility that the policyholder surrenders. That is, we add the possibility that the policyholder terminates her contract and instead receives a deterministic, continuous surrender value, $G(t)$. This can, e.g., be added to the model by assuming that the policyholder at any time surrenders with some deterministic, continuous intensity, $\nu(t)$, see Figure 5.2. We use the term active for when the policy is in force.

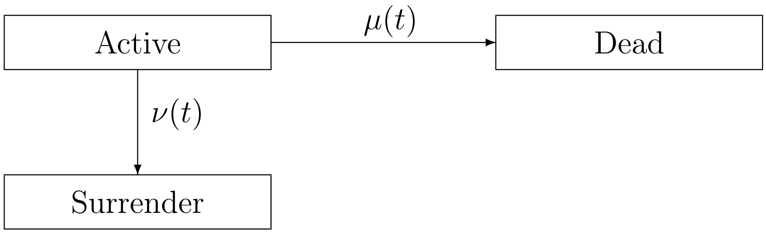

Figure 5.2: Standard surrender model.

Mathematically, the state of surrender is in this model not different from the state of death, except that the associated payments are different. The reserve, V_ν, is continuously differentiable and solves the following Thiele's differential equation, see e.g. [14],

$$V_\nu'(t) = r(t)V_\nu(t) + \pi(t) - \mu(t)\left(b^{ad}(t) - V_\nu(t)\right) - \nu(t)\left(G(t) - V_\nu(t)\right), \quad (5.2)$$

with $V_\nu(n-) = b_a(n)$.

For V_v to be continuously differentiable we need, in general, that ν is continuous as assumed above. However, what is really needed is that $\nu(t)\left(G(t) - V_\nu(t)\right)$ is continuous and this can be obtained even when ν is discontinuous and properly defined at the point where $G(t) = V_\nu(t)$.

The surrender value G can be anything exogenously given. In practice it is, typically, a technical value of the same payment stream based on technical assumptions on interest rates and intensities that we denote by (r^*, μ^*). In that case, the surrender value is the technical reserve V^* that solves (5.1) with (r, μ) replaced by (r^*, μ^*).

5.3 Reserve Dependent Surrender

The forthcoming Solvency II regulations requires that the traditional modelling of surrender is revisited. Thus, we need to investigate and model what influences the policyholders choice to surrender and we need to be able to calculate the reserves in the more advanced models. In the present section we suggest a way to do this, and discuss our method.

In a more realistic model of surrender we want to be able to express both that surrender is likely influenced by how profitable it is, but also that it is still random. On one hand, we wish surrender to be influenced by how profitable it is, because surrender is a decision the policyholder makes. On the other hand we also have multiple reasons for surrender being random. Randomness is natural because the policyholder most likely lacks information to decide what is profitable. Even if she had all the information that the pension fund has and were able to use it, then her preferences may differ seemingly randomly from the model set up by the pension fund because of the policyholders personal preference and economical situation. She might shift her job and get an offer from a new pension fund or she might need cash for a divorce.

We can obtain randomness in our model by keeping the surrender modelled by an intensity. Further, we model that the policyholders decision depends on how profitable it is by letting the surrender intensity depend on how profitable it is for the policyholder to surrender. If she surrenders at time t she gains $G(t)$, but she loses the rest of the contract including her right to exercise later. Hence, she loses $V_\nu(t)$. Therefore, we denote by $G(t) - V_\nu(t)$ her profit from surrendering at time t. We would like the surrender intensity to be non-negative and increasing in this profit. At first glance this modelling seems to have a problem that the definition of the surrender intensity is circular. However, Theorem 5.3.1 below gives sufficient conditions for this circular definition not to be a problem.

Theorem 5.3.1. *For some given non-negative function, h, consider the following differential equation in the function U:*

$$U'(t) = r(t)U(t) + \pi(t) - \mu(t)(b_{ad}(t) - U(t)) - h(t, U(t))(G(t) - U(t)), \quad (5.3)$$

with $U(n-) = \Delta B(n)$. Suppose (5.3) has a unique solution, U, and define a surrender intensity by $\nu(t) \equiv h(t, U(t))$. Then U is the reserve when the policyholder chooses to surrender at time t with intensity $\nu(t)$.

Proof. The possible problem in this model is the circular definition of the surrender intensity. However, the existence and uniqueness of the solution to both (5.2) and (5.3) ensures that this does not become a problem.

The process ν defined by $\nu(t) \equiv h(t, U(t))$ is uniquely determined from (5.3) and the reserve is then uniquely determined from (5.2). It follows from the definition of U that U solves (5.2), and then from the uniqueness of the solution to (5.2) it follows that the reserve is given by U. □

Once we have decided on a policyholder with a specific policy and a function h, and thereby also ν, then for this single policyholder, our model does not differ from a model with a deterministic time dependent surrender intensity as what we had in the classical model of (5.2). However, when we use the model for pricing a portfolio of insurance contracts for a group of policyholders, then the model assigns different surrender intensities to each policyholder. Thereby, the reserves in general become higher than if we had used a constant surrender intensity or a specific time dependent surrender intensity for the whole portfolio.

The relation between the surrender intensity and the profitability may be chosen in many different ways. Two examples we investigate are:

$$\nu(t) = h(t, V_\nu(t)) = \psi \exp\{\theta(G(t) - V_\nu(t))\}, \tag{5.4}$$

$$\nu(t) = h(t, V_\nu(t)) = \theta 1_{(G(t) - V_\nu(t) > 0)}, \tag{5.5}$$

where ψ, $\theta > 0$ are constants. For equation (5.4), ψ tells about the overall tendency to surrender, whereas θ tells about how profitability creates deviations from this tendency. For equation (5.5), θ controls both. In both cases we speak of θ as the rationality parameter. Other intensity functions can be chosen and one should choose a functional form which matches with data. The only mathematical requirement is that the function h has to make it possible to use Theorem 5.3.1.

One immediate drawback of our model is that we most often do not have an explicit solution for the differential equation (5.3). This implies that we do not have an explicit expression for the reserve. However, we do have algorithms available for numerical solutions to ordinary differential equations.

5.4 Reserve Dependent Policyholder Behaviour

The idea of modelling behaviour by profit dependent intensities may be used for other applications as well. Within life insurance the policyholder's choice to convert into free policy (paid-up policy) has some resemblance with the surrender choice. Thereby we may find it reasonable to expand our model with the possibility of conversion into free policy in the same way as we added surrender. Figure 5.3 displays a simple model where ν_{af} denotes the intensity of conversion into free policy, ν_{as} denotes the intensity of surrender when active and ν_{fs} denotes the intensity of surrender after converted into free policy. Here the term active is used when the policyholder is paying premiums.

If all transition intensities are known explicitly, this model is studied in [11]. When a policyholder converts into free policy the payments are reduced

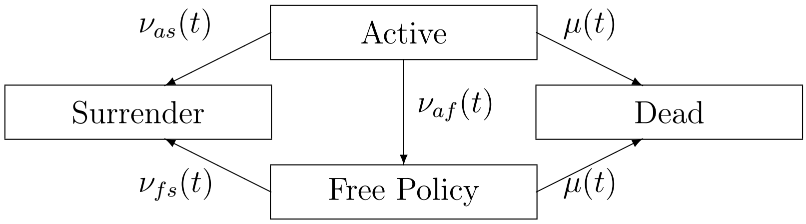

Figure 5.3: Free policy and surrender model.

depending of the time of conversion. Let $b_{fd}(t, u)$ denote the death sum at time t if converted into free policy at time u, let $b_f(n, u)$ denote the terminal payment at time n if converted into free policy at time u, and let $G_f(t, u)$ denote the surrender value at time t when converted into free policy at time u. For the reserves we let $V_a(t)$ denote the reserve at time t if the policyholder is active, and let $V_f(t, u)$ denote the reserve at time t if the policyholder is in the free policy state and converted to free policy at time u. Now, we assume that the intensities are reserve dependent and given in the form

$$\begin{aligned} \nu_{as}(t) &= h_{as}(t, V_a(t)), \\ \nu_{af}(t) &= h_{af}(t, V_a(t)), \\ \nu_{fs}(t, u) &= h_{fs}(t, u, V_f(t, u)). \end{aligned}$$

Then the reserves are given from the following differential equations:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} V_a(t) &= r(t)V_a(t) + \pi(t) - \mu(t)(b_{ad}(t) - V_a(t)) \\ &\quad - h_{as}(t, V_a(t))(G(t) - V_a(t)) - h_{af}(t, V_a(t))(V_f(t, t) - V_a(t)), \\ V_a(n-) &= b_a(n), \\ \frac{\partial}{\partial t} V_f(t, u) &= r(t)V_f(t, u) - \mu(t)(b^*_{fd}(t, u) - V_f(t, u)) \\ &\quad - h_{fs}(t, u, V_f(t, u))(G^f(t, u) - V_f(t, u)), \\ V_f(n-, u) &= b_f(n, u). \end{aligned}$$

The only requirement is that the system of differential equations has a unique solution. However, the differential equations from above are heavy to work with, as we need to solve a new differential equation for each value of $V_f(t, t)$. When modelling the free policy option, this problem is usually overcome by introducing a scaling function, f, that describes the reduction of payments as a result of the conversion to free policy. Thus, $b_{fd}(t, u) = f(u)b_{ad}(t)$, $b_f(n, u) = f(u)b_a(n)$ and $G_f(t, u) = f(u)G_f(t)$. Assume the transition intensity ν_{fs} does not depend on the time of transition to free policy. Then the prospective reserve, $V_f^*(t)$, from the free policy state based on the payments $G_f(t)$, $b_{ad}(t)$ and $b_a(n)$ does not depend on this transition time either, and we get $V_f(t,u) = f(u)V_f^*(t)$ with

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V_f^*(t) &= r(t)V_f^*(t) - \mu(t)(b_{ad}(t) - V_f^*(t)) \\ &\quad - \nu_{fs}(t)(G^f(t) - V_f^*(t)), \\ V_f^*(n-) &= b_a(n).\end{aligned}$$

This makes $V_f(t,u)$ a lot easier to calculate. For more on the determination of the reference payments and scaling function, see [11]. Note however that if ν_{fs} cannot depend on the time of transition to free policy, u, then it cannot depend on $G_f(t,u) - V_f(t,u)$ either and this is a large disadvantage.

To get profit dependent choices we may use

$$\begin{aligned}\nu_{as}(t) &= h_{as}(t, V_a(t)) = \psi_{as}e^{\theta_{as}(G(t)-V_a(t))}, \\ \nu_{af}(t) &= h_{af}(t, V_a(t)) = \psi_{af}e^{\theta_{ap}(V_f(t,t)-V_a(t))}, \\ \nu_{fs}(t,u) &= h_{fs}(t, u, V_f(t,u)) = \psi_{fs}e^{\theta_{fs}(G_f(t,u)-V_f(t,u))}.\end{aligned}$$

5.5 Approximation of the Worst Case Reserve

In the two previous sections we discussed our model with reserve dependent surrender and we found it being a reasonable model for predicting the dynamics of surrender. However, in the following section we discuss how the model may also be used for determining worst case reserves when the true dynamics of the surrender intensity is not known. This is because our model is a version of what in the literature is known as the penalty method, and a large rationality parameter gives us the worst case reserve.

Typically the technical reserve is paid out upon surrender (potentially minus expenses). In that case, if we take maximum of the technical reserve and the market reserve calculated under the assumption of no surrender, then we get a worst case reserve of either surrendering immediately or never surrender. However, a surrender strategy somewhere in between the two extremes may result in a higher market reserve. For determining the worst case reserve we consider all possible surrender strategies. To do this we construct a more general model. We assume that the transition from active to surrender is governed by a randomized stopping time, τ, with respect to the state of the policyholder, with randomized stopping times being defined as in [15]. That is, the time of surrender may depend on everything but the future time of death and the future interest rate. If the policyholder never surrenders her contract we let $\tau = n$. The model is illustrated in Figure 5.4. The class of admissible surrender strategies at time t are the variables in $[t, n]$ that are randomized stopping times with respect to the filtration generated from I. We denote this class by $\mathcal{T}_t$.

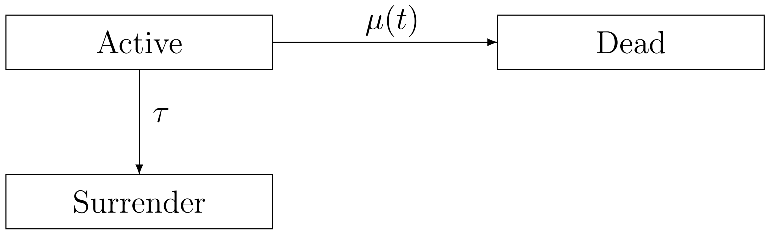

Figure 5.4: Optimal surrender model.

We hereby disregard the possibility that the policyholder has more information about her future time of death than the pension fund has. We do this despite that such knowledge could influence the policyholders decisions.

Let V_τ denote the prospective reserve if the policyholder surrenders according to the randomized stopping time τ. Assume $G(n) = 0$, assume $G(n-) \leq V(n-)$ and assume G continuous on $[0, n)$. Then from [14] it follows that V_τ is given by:

$$\begin{aligned} V_\tau(t) &= \mathbb{E}\left[\int_t^\tau e^{-\int_t^u r(x)dx} dB(u) + e^{-\int_t^\tau r(x)dx} G(\tau)I(\tau) \middle| I(t) = 1\right] \qquad (5.6) \\ &= V(t) + \mathbb{E}\left[e^{-\int_t^\tau r(x)dx} G(\tau)I(\tau) - \int_\tau^n e^{-\int_t^u r(x)dx} dB(u) \middle| I(t) = 1\right] \\ &= V(t) + \mathbb{E}\left[e^{-\int_t^\tau r(x)dx} I(\tau)G(\tau) \middle| I(t) = 1\right] \\ &- \mathbb{E}\left[e^{-\int_t^\tau r(x)dx} I(\tau)\mathbb{E}\left[\int_\tau^n e^{-\int_\tau^u r(x)dx} dB(u) \middle| \tau, I(\tau) = 1\right] \middle| I(t) = 1\right] \\ &= V(t) + \mathbb{E}\left[e^{-\int_t^\tau r(x)dx} I(\tau)\left(G(\tau) - V(\tau)\right) \middle| I(t) = 1\right]. \end{aligned}$$

Consider the worst case scenario for the pension fund, where the policy holder chooses the surrender strategy as the stopping time, τ, that maximizes V_τ. This is an optimal stopping problem. Any classical stopping time from the filtration generated by I must fulfil $\tau I(\tau) = t_0 I(t_0)$ for some deterministic $t_0 \in [t, \infty]$. The reserve is then given by:

$$\begin{aligned} V_\tau(t) &= V(t) + \mathbb{E}\left[e^{-\int_t^\tau r(x)dx} I(\tau)\left(G(\tau) - V(\tau)\right) \middle| I(t) = 1\right] \\ &= V(t) + e^{-\int_t^{t_0} r(x)+\mu(x)dx}\left(G(t_0) - V(t_0)\right). \end{aligned}$$

Thus, for the classical optimal stopping problem, without randomization allowed, it is optimal to choose t_0 as any time from the set:

$$A_t \equiv \arg\max_{u\in[t,n]} \left(e^{-\int_t^u r(x)+\mu(x)dx} (G(u) - V(u)) \right).$$

As the inner part is continuous in u on $[0, n)$ and as $G(n-) - V(n-) < G(n) - V(n)$, then A_t must have a largest element. Denote this element by u^*, i.e. let $u^*(t) \equiv \max A_t$, such that $u^*(t)$ is the latest optimal time to surrender. Let $\tau^* = u^*(t)I(u^*(t)) + n(1 - I(u^*(t)))$. We define the worst case reserve, W, by:

$$W(t) = \sup_{\tau\in\mathcal{T}_t} V_\tau(t) = V_{\tau^*}(t).$$

By a proof similar to the one of the verification theorem of Chapter 9 of [13] it may be seen that τ^* is optimal even if we allow randomized surrender strategies.

Now, assume a family of functions, h_θ, is implied. Let τ_θ denote the surrender strategy of surrendering at time u with intensity $\nu_\theta(u) = h_\theta(G(u) - V_{\nu_\theta}(u))$ and let $V_\theta = V_{\nu_\theta}$ with V_{ν_θ} as defined in Sect. 5.3. Let

$$\bar{h}_\theta(x) \equiv \sup_{y\le x} h_\theta(y),$$

and

$$\underline{h}_\theta(x) \equiv \inf_{y\ge x} h_\theta(y).$$

Now, the following holds:

Theorem 5.5.1. *Suppose that for each $\theta \ge 0$ we have that h_θ is defined in a way such that we may use Theorem 5.3.1 and suppose that the surrender value G is continuous on $[0, n)$ with $G(n-) \le V(n-)$ and $G(n) = 0$. Also assume for $x < 0$:*

$$\bar{h}_\theta(x) \to 0, \quad \theta \to \infty, \tag{5.7}$$

and for $x > 0$:

$$\underline{h}_\theta(x) \to \infty, \quad \theta \to \infty. \tag{5.8}$$

Then, for every $t \in [0, n]$:

$$V_\theta(t) \to W(t), \qquad \theta \to \infty.$$

Proof. See the appendix. □

Remark 5.5.1. *Some of the details in the proof has been omitted, but a fully detailed proof following the same reasoning for a closely related result for an American Put option may be found in [7]. The fact that the penalty method provides convergence and the rate of convergence is not new. However, we find the proof of our article and of [7] interesting. This is because they visualize how the error terms may be thought of as probabilities of economically bad choices of the policyholder times the loss the policyholder faces from her bad choices.*

5.6 Advanced Surrender Values

In Section 5.3 we made simplifying assumptions on the surrender value, G. In Section 5.6.1 and Section 5.6.2 below we investigate two ways to generalize this model. First, in Section 5.6.1 we allow the surrender value to depend on the market value of the reserve. This makes it possible for the pension fund to avoid a loss upon surrender. Second, in Section 5.6.2 we add expenses upon surrender.

5.6.1 Surrender Value Dependent on the Market Reserve

In Section 5.3, we assumed the surrender value, G, to be given exogenously from the market valuation. That is, we assumed the surrender value was based on the guaranteed benefits and the technical basis only, and not on the market basis. However, to prevent losses from surrender the pension fund may choose that the surrender value depends on the market value of the contract. Thus there may be an explicitly given function $h : \mathbb{R}_+ \times [0, n] \to \mathbb{R}_+$ such that $G(t) = h(V(t), t)$. In a similar way as argued in Section 5.3 we may expect that the policyholder finds that the profit from surrender is $h(V(t), t) - V(t)$. If the surrender intensity is given by a rationality function, f_θ, of the profit, the market value of the reserve becomes

$$\begin{aligned} V'(t) = r(t)V(t) + \pi(t) - \mu(t)\left(b^{ad}(t) - V(t)\right) \\ - f_\theta\left(h(V(t), t) - V(t)\right)\left(h(V(t), t) - V(t)\right), \end{aligned} \tag{5.9}$$

with $V(n-) = b_a(n)$. Similar to Section 5.3 we have to consider if our construction of the intensity for surrender is well defined. With arguments similar to the proof of Theorem 5.3.1 it is sufficient that there exists a unique solution to (5.9) as well as a unique solution to (5.9) with f_θ replace by $\nu(t)$ for ν sufficiently regular.

With $G(t) = h(V(t), t)$ the technical reserve, and thereby the equivalence premium, in general depends on $h(V(t), t)$. Besides, it is likely that h is chosen such that it depends on the value of the technical reserve. Then the technical and market reserves are mutually dependent in a mathematically intractable way. However, if we choose the technical surrender intensity to zero, then the technical reserve and thereby the equivalence premium may be calculated independently of the market reserve. If the surrender value is set such that the pension fund avoids losses upon surrender, then a technical surrender intensity of zero is a safe-side choice in the following sense: We define a safe-side technical basis as a technical basis that ensures the market reserve is never higher than the technical reserve. Let V^* denote the technical reserve. By differentiation it is easily verified that

$$\begin{aligned} V^*(t) - V(t) \\ = \int_t^n e^{\int_t^s r+\mu(u)+\nu^*(u)du} \Big((r - r^*)V^*(s) - (\mu(s) - \mu^*(s))(b^{ad}(s) - V^*(s)) \\ - (\nu(s) - \nu^*(s))(G(s) - V(s)) \Big) ds. \end{aligned} \tag{5.10}$$

Note that the technical surrender intensity in the exponential function conforms with the replacement of the replacement of the technical reserve with the market reserve in the surrender risk term. If $\nu^* = 0$ then the technical reserve equals the technical reserve from a model without surrender. Thus, the first line of (5.10) corresponds to $V^* - V$ if surrender were not included in the model. It follows that if the surrender value is set such that $G \leq V$, then $\nu^* = 0$ is safe-side for a model with surrender.

Note that with $\nu^* = 0$ and $G = V$, then both V^* and V equal the corresponding reserves in a model where surrender is not included.

5.6.2 Expenses upon Surrender

In Section 5.3 we assumed that $G(t) - V_\nu(t)$ is an appropriate measure for the policyholder's profit from surrendering, and that this same value reflects the loss for the pension fund if the policyholder surrenders. However, in a more realistic model, this duality is not as precise. Surrender has a cost, which is lost both for the policyholder and the pension fund. In the present section we first assume surrender has a cost $\epsilon_G(t)$ which is deterministic and independent of the market basis. E.g., this may be a constant value as is common in Denmark, or it may be some deterministically varying proportion of the technical reserve as described in [4]. The expense is paid by the policyholder upon surrender, and thus the payout to the policyholder upon surrender is $G(t) - \epsilon_G(t)$.

Since we have lost the duality between the policyholder and the pensions fund with regards to the payments of the contract, the policyholder and the pension fund find different values for the contract. We introduce $\tilde{V}$ to denote the market value of the contract from the policyholders point of view. Now the profit of the policyholder from surrendering instead becomes $G(t) - \epsilon_G(t) - \tilde{V}(t)$. With a profit-intensity function f_θ the market value from the policyholder's point of view is characterized by

$$\begin{aligned} \tilde{V}'(t) &= r(t)\tilde{V}(t) + \pi(t) + \mu(t)\left(b^{ad}(t) - \tilde{V}(t)\right) \\ &\quad - f_\theta\left(G(t) - \epsilon_G(t) - \tilde{V}(t)\right)\left(G(t) - \epsilon_G(t) - \tilde{V}(t)\right), \end{aligned} \tag{5.11}$$

whereas the market value from the pension fund's point of view is characterized by

$$\begin{aligned} V'(t) &= r(t)V(t) + \pi(t) + \mu(t)\left(b^{ad}(t) - V(t)\right) \\ &\quad - f_\theta\left(G(t) - \epsilon_G(t) - \tilde{V}(t)\right)\left(G(t) - V(t)\right), \end{aligned} \tag{5.12}$$

with $\tilde{V}(n-) = V(n-) = b_a(n)$. The surrender sum at risk reflects that, in addition to the net surrender value $G(t) - \epsilon_G(t)$, also the expense $\epsilon_G(t)$ is lost. In this model the two reserves may be calculated simultaneously or the policyholder market value may be calculated first. Let $\tilde{G}(t) = G(t) - \epsilon_G(t)$. We see from (5.11) that under the conditions of Theorem 5.5.1 we have that when the rationality parameter goes to infinity, then $\tilde{V}$ converges to the policyholder market value corresponding to optimal surrender from her point of view. However, due to the lack of duality, this value is not the worst case market value for the pension fund, which was found in Section 5.5. The pension fund market value of (5.12) no longer fits with the requirements of Theorem 5.5.1 and from Example 6 we see that V does not converge to the worst case reserve for the pension fund's point of view, when the rationality parameter increases to infinity. However, if we have an upper bound for the expenses upon surrender, then we may use the optimal policyholder market value to create an upper bound for the worst case market reserve. Let τ^*_{ph} denote the optimal surrender time from the policyholder point of view, and let τ^*_{pf} denote the worst case surrender strategy from the pension fund point of view. Assume the surrender expenses are bounded by $\epsilon_G(t) \leq C$ for some constant C. Suppose that for any rationality parameter higher that θ then τ_θ is a good approximating strategy in the sense that $\tilde{V}_{\tau_\theta}(t) \in [\tilde{V}_{\tau^*_{ph}}(t) - \varepsilon, \tilde{V}_{\tau^*_{ph}}(t)]$.

Let $V_{\epsilon_G,\tau}(t)$ denote the present value of the future surrender expenses under the strategy τ. From (5.6) we find that for any surrender strategy τ,

$$\begin{aligned} V_\tau(t) &= \mathbb{E}\left[\int_n^\tau e^{-\int_t^u r(x)dx} dB(u) + e^{-\int_t^\tau r(x)dx} G(\tau) I(\tau) \middle| I(t) = 1\right] \\ &= \mathbb{E}\left[\int_n^\tau e^{-\int_t^u r(x)dx} dB(u) + e^{-\int_t^\tau r(x)dx} (\tilde{G}(\tau) + \epsilon_G(\tau)) I(\tau) \middle| I(t) = 1\right] \\ &= \tilde{V}_\tau(t) + \mathbb{E}\left[e^{-\int_t^\tau r(x)dx} \epsilon_G(\tau) I(\tau) \middle| I(t) = 1\right] \equiv \tilde{V}_\tau(t) + V_{\epsilon_G,\tau}(t). \end{aligned}$$

Thus,

$$\begin{aligned} V_{\tau_\theta}(t) &= \tilde{V}_{\tau_\theta}(t) + V_{\epsilon_G,\tau_\theta}(t) \geq \tilde{V}_{\tau^*_{ph}}(t) - \varepsilon + V_{\epsilon_G,\tau_\theta}(t) \geq \tilde{V}_{\tau^*_{pf}}(t) - \varepsilon + V_{\epsilon_G,\tau_\theta}(t) \\ &= V_{\tau^*_{pf}}(t) - V_{\epsilon_G,\tau^*_{pf}}(t) - \varepsilon + V_{\epsilon_G,\tau_\theta}(t) = V_{\tau^*_{pf}}(t) - \varepsilon - C. \end{aligned}$$

Thus, even though a high rationality parameter does not give the worst case market value for the pension fund, then in the limit the difference is bounded by the maximum surrender expenses. The maximum surrender value is the best general fixed bound we can find since Example 6 shows that there exists a product and a scenario for which we have this difference between $V_{\tau^*_{pf}}$ and $V_{\tau^*_{ph}}$. Of course, exploiting knowledge of the product and the market interest rate makes it possible to find better bounds.

5.7 Numerical Examples

In this section we show four examples of how various surrender models impact the development of the reserves in four different interest rate situations. In each example we consider a contract with a death sum of $b^{ad} = 1,000,000$, a pension sum of $2,000,000$ and a constant premium intensity of $\pi = 16,218$. All values measured in DKK and the premium is the equivalence premium set at age 25. These numbers are chosen as they have a realistic level for a Danish pension policy. For fairly realistic numbers in EUR divide by ten. The policyholder is assumed to be 35 years old at time 0 and the time of retirement is at age 65. Time is measured in years and her mortality intensity is assumed to be given by:

$$\mu(t) = 0.0005 + 10^{5.728-10+0.038*(t+35)}.$$

This is the mortality intensity from the Danish life table G82 for females. If the policyholder surrenders her contract, she receives a surrender value given by the technical reserve. The technical reserve is based on the same payments as the contract, and on a technical interest rate intensity of $r^* = 0.05$. Interest rates are chosen high to better visualize the impact of the choice of surrender intensity. We assume no extra expenses at surrender. Thus, the surrender value is given from the differential equation:

$$G'(t) = r^* G(t) + \pi - \mu(t)(b^{ad} - G(t)), \tag{5.13}$$

with $G(n-) = 2,000,000$. We consider the following five surrender models:

$$\begin{aligned}
\text{Model a}: \quad \nu^a(t) &= 0.05 \cdot \exp\{0.000003(G(t) - V_{\theta,\psi}(t))\} \\
\text{Model b}: \quad \nu^b(t) &= 0.05 \cdot 1_{(G(t)-V_\theta(t)>0)} \\
\text{Model c}: \quad \nu^c(t) &= 0.05 \\
\text{Model d}: \quad \nu^d(t) &= 0 \\
\text{Model e}: \quad \nu^e(t) &= 5 \cdot 1_{(G(t)-V_\theta(t)>0)}.
\end{aligned}$$

The first three models are based on a surrender intensity of around 5%. The last model is a model with a rationality parameter $\theta = 5$, which has been found to be high enough for us to approximate the worst case reserve. Additionally, we consider four different developments of the interest rate, r, used for pricing market reserves. For the two first interest rate situations we compare the surrender value and the reserves for the five different surrender models. For the two last interest rate situations we compare the surrender value and the reserves for surrender Model d and Model e.

Example 1: Market interest rate is above technical interest rate Assume $r = 0.15$. The reserve developments are displayed in Figure 5.5. In this situation it is at all time points optimal for the policyholder to surrender. The worst case reserve corresponds to the surrender value. The lowest reserve is the market reserve based on no surrender, Model d. Models with a chance of surrender has reserves in between. Since there is no risk of surrendering too early, then Model b and the traditional Model c do not differ. For Model a we get a slightly higher reserve than the one for Model b and Model c, because the basic intensity 0.05 is slightly increased at all time points by the exponential factor in the intensity.

Example 2: Market interest rate is below technical interest rate Assume $r = 0.02$. The reserve developments are displayed in Figure 5.6. In this situation it is never optimal for the policyholder to surrender. The worst case reserve corresponds to the market reserve with no surrender. In Model b and Model e the policyholder does not make the mistake of surrendering if it is not profitable, and thus, this has an equally high reserve. The surrender value is the lowest value and the reserves of Model a and the traditional Model

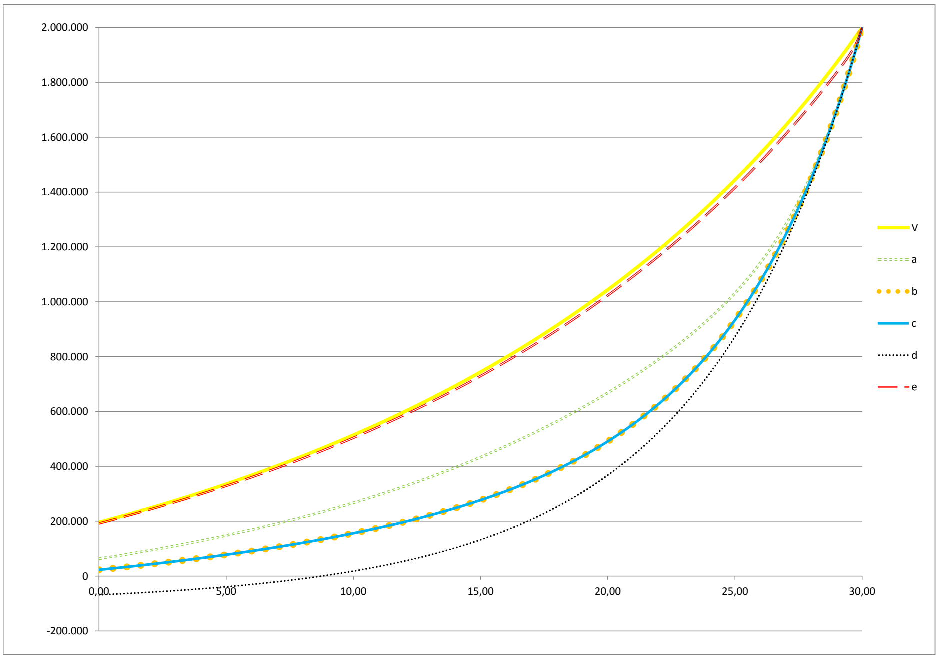

Figure 5.5: Example 1. The technical interest rate is $r^* = 0.05$. The market interest rate is $r = 0.15$. Immediate surrender is always optimal.

c are in between. Model a has a higher reserve than Model c, because the basic intensity 0.05 is slightly increased at all time points by the exponential factor in the intensity.

Example 3: Market interest rate is decreasing Assume $r(t) = 0.10 \cdot 1_{(t \le 20)} + 0.04 \cdot 1_{(t>20)}$. The qualitative feature we capture is that the interest rate crosses the guaranteed interest rate downwards. The reserve developments are displayed in Figure 5.7. In this situation it is optimal to surrender if the surrender value is higher that the market reserve in Model d with no surrender. Thus, after time $t = 20$ it is optimal to keep the contract because the technical interest rate is higher than the market interest rate. Right before time $t = 20$ the interest rate of the market is higher than the technical interest rate, but this is only for a short time, and thus it is still optimal to keep the policy in order to benefit from the technical interest rate later on. At some point before time $t = 20$ the surrender value and the market reserve of Model d intersects. Before this time it is optimal to surrender because

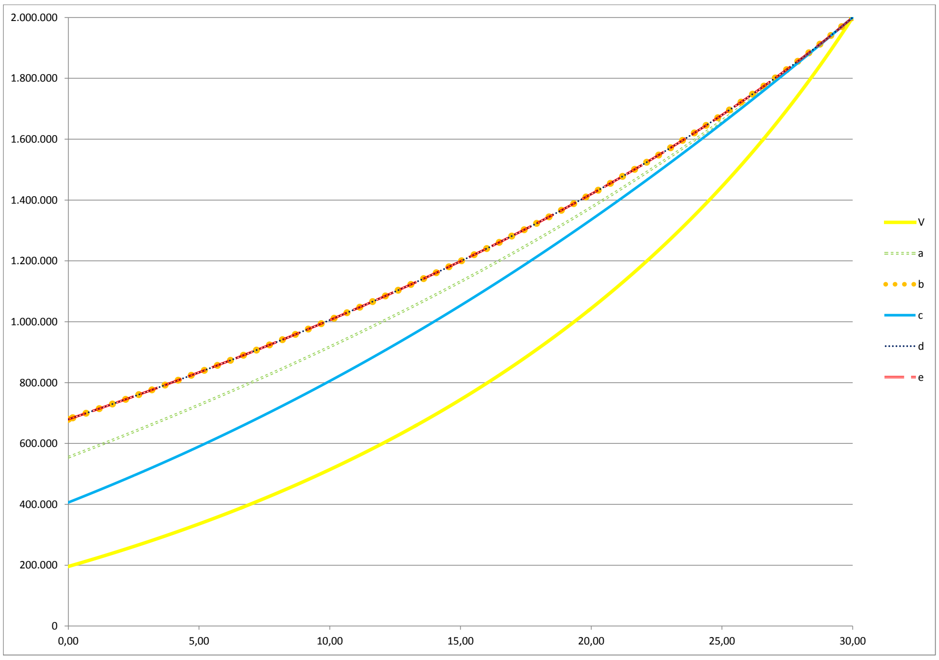

Figure 5.6: Example 2. The technical interest rate is $r^* = 0.05$. The market interest rate is $r = 0.02$. Surrender is never optimal.

the gain from the high market interest rate before time $t = 20$ is then higher than the future loss from the low market interest rate. All together the worst case reserve is given as the maximum of the surrender value and the market reserve with no surrender.

Example 4: Market interest rate is increasing Assume $r(t) = 0.01 \cdot 1_{(t \leq 20)} + 0.065 \cdot 1_{(t>20)}$. The qualitative feature we capture is that the interest rate crosses the guaranteed interest rate upwards. The reserve developments are displayed in Figure 5.8. In this situation we have that after time $t = 20$ it is optimal to surrender. Before time $t = 20$ it is optimal to plan to surrender at time $t = 20$. With this strategy the policyholder benefits from both the high market interest rate after time $t = 20$ and the technical interest rate before time $t = 20$ when the market interest rate is low. Thereby, unlike in the previous three examples, the worst case reserve is no longer the supremum of the surrender value and the market reserve with no surrender. Before time $t = 20$ the worst case reserve is higher than both of the other reserves,

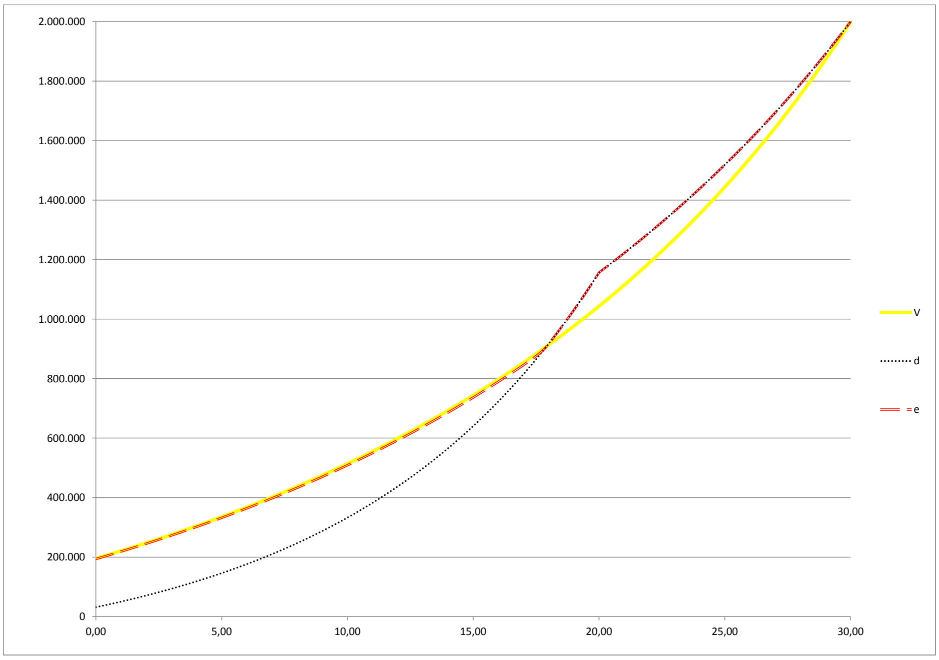

Figure 5.7: Example 3. The technical interest rate is $r^* = 0.05$. The market interest rate is $r(t) = 0.10 \cdot 1_{(t \leq 20)} + 0.04 \cdot 1_{(t>20)}$. Surrender is optimal if the surrender value is higher than the market reserve with no surrender.

because there exists a surrender strategy which is better for the policyholder than both immediate surrender and no surrender.

We recall that the reserves of Model a and Model b converge to the worst case reserve when the rationality parameter converges to infinity. Thus, if the rationality parameter is sufficiently high and the future increase in interest rate is sufficiently high, then the reserves of Model a and Model b become higher than the maximum of the surrender value and the market reserve of Model d with no surrender.

The values in Example 4 is chosen to clearly visualize the possible impact of a market interest rate that crosses the guaranteed interest rate upwards. This is not an unrealistic scenario today. In a model with a deterministic market interest rate, a natural choice for the market interest rate in Denmark is the Danish FSA yield curve. Figure 5.9 displays the Danish FSA yield curve as of March 25th 2015. It is common in Denmark to sell pension products with a technical interest rate of $0\% - 1\%$. Thus the Danish FSA yield curve gives a market rate that is currently lower than a technical interest rate of

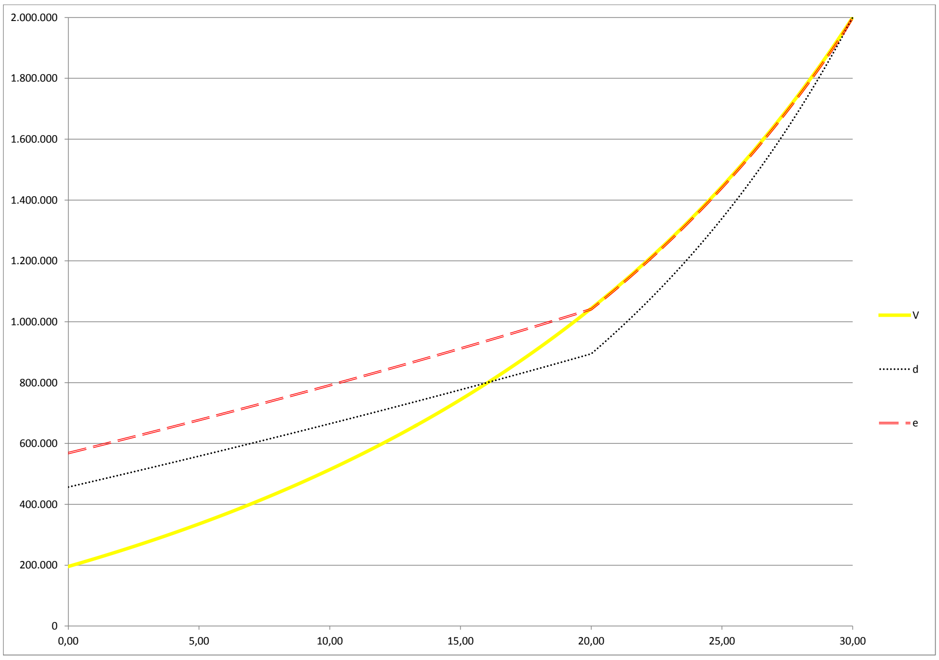

Figure 5.8: Example 4. The technical interest rate is $r^* = 0.05$. The market interest rate is $r(t) = 0.01 \cdot 1_{(t \leq 20)} + 0.065 \cdot 1_{(t>20)}$. After time $t = 20$ it is optimal to surrender. Before time $t = 20$ it is wise to plan to surrender at time $t = 20$.

0.5% or 1%, but within the next 20 years increases to above.

Example 5: The Effect of Expenses We revisit Example 1 and assume a surrender intensity function of $f_\theta(x) = 0.05e^{\theta x}$. We now investigate the impact on the market reserve from adding a surrender expense of $\epsilon_G = 2,000$ to the model. We assume G is the technical reserve. We calculate the market reserve without expenses, $V_{no\ expenses}$, from (5.2) with $\nu(t) = f_\theta(G(t) - V_{no\ expenses}(t))$, and we calculate the market reserve with expenses, $V_{expenses}$, from (5.12). Figure 5.10 displays $V_{no\ expenses} - V_{expenses}$ for various values of the rationality parameter θ.

The effect of adding expenses to the model is very sensitive to θ. For $\theta = 0$ the policyholder's surrender intensity is not affected by profitability. Thus, the market reserve from the pension fund point of view is not affected by adding surrender expenses. If $\theta = 0.000003$ the policyholders become less inclined to surrender when expenses are added to the model. This results in

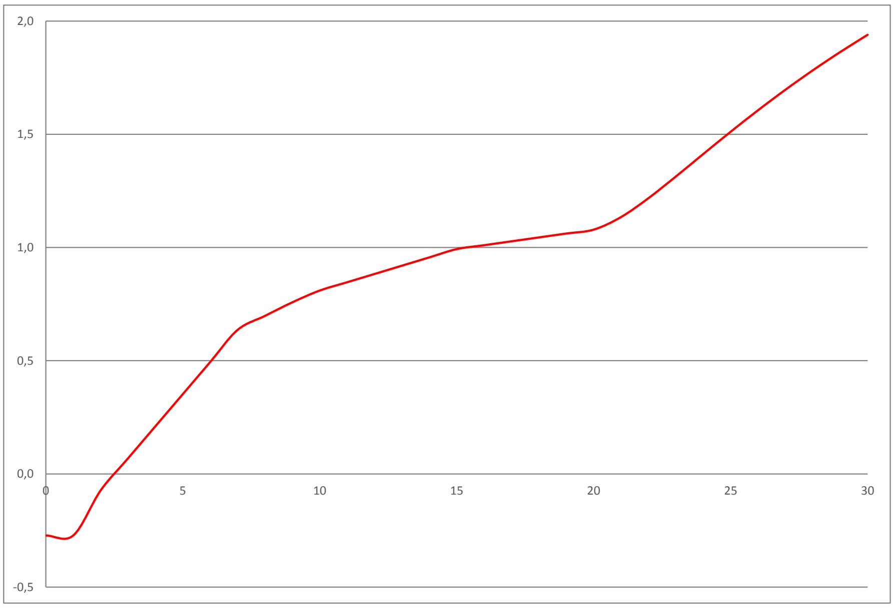

Figure 5.9: Example 4. The Danish FSA yield curve as of March 25th 2015.

a decrease in the market reserve. From numerical calculations we find that the maximal decrease over time in this model is 458. In a more extreme case with $\theta = 0.003$ the maximal decrease over time is $1,196$.

However, the impact of the rationality parameter and the expenses is more involved than the numbers above reflect. For $\theta = 0.000003$ the impact of the expenses has a broad peak, whereas and for $\theta = 0.003$, it has a narrow spike. Further, these patterns depend of course on the product and the interest rate scenario. In Example 6 below we show that for specific examples we can also have a broad peak with a high impact.

Example 6: Maximal Duality Gab In Section 5.6.2 we discussed the maximal duality gap between the market reserve based on the policyholder acting completely rational and the market reserve based on the the strategy which is worst case for the pension fund. We found that if there is an upper bound on the surrender expenses, then this is also an upper bound for the duality gap. In the present example we show that hitting this maximal duality gab is realistic.

Consider a contract where the premium is $\pi = 9.500$, the death sum is

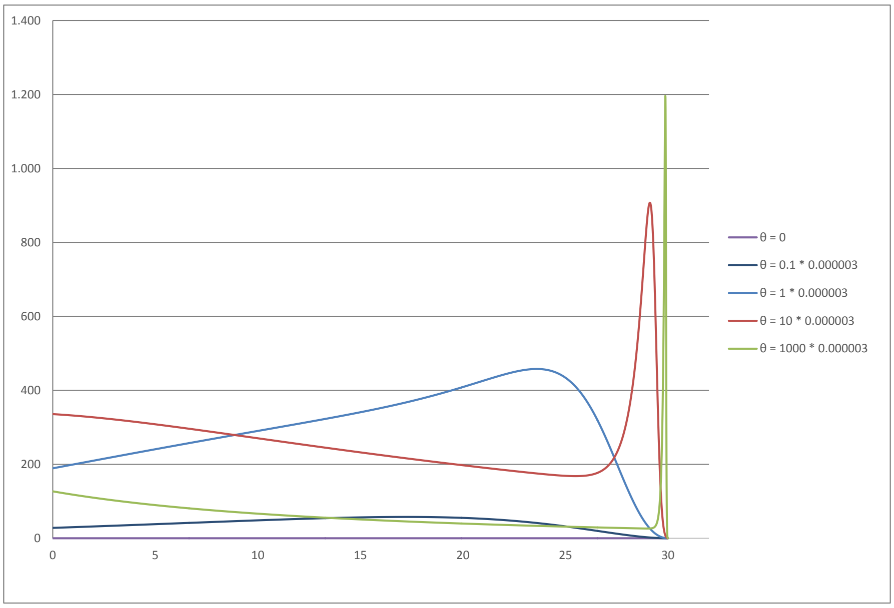

Figure 5.10: Example 5. The impact of expenses measured by $V_{no\ expenses} - V_{expenses}$ for $f_\theta(x) = 0.05e^{\theta x}$ with various rationality parameters.

$b^{ad} = 200,000$ and the pension sum is $400,000$. Upon surrender the policyholder receives the technical reserve minus an expense of $\epsilon_G = 2,000$. All values measured in DKK. The policyholder is assumed to be 50 years old and the age at retirement is 65. We assume the same mortality intensity and the same technical interest rate as in the previous examples.

Figure 5.10 displays the duality gap $V_{\tau^*_{pf}} - V_{\tau^*_{ph}}$ of Section 5.6.2. That is, the difference between the market reserve based on the policyholder surrendering completely rational the market reserve if the policyholder surrenders according to the strategy which maximizes the market reserve from the pension fund point of view. Note that $V_{no\ expenses} - V_{expenses}$ converges to this value when the rationality parameter goes to infinity. In this example $2,000$ is an upper bound for the surrender expenses and from Figure 5.11 it is seen that it is possible to experience a duality gab of this size.

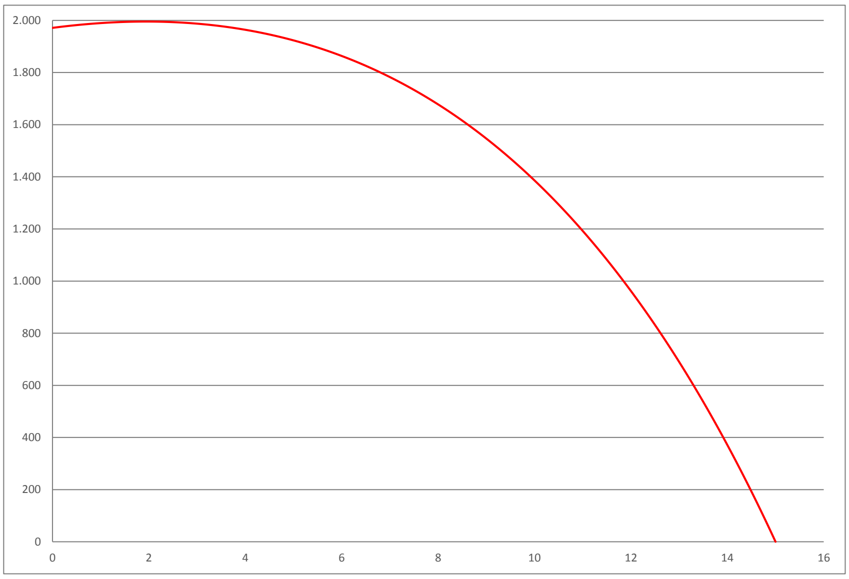

Figure 5.11: Example 6. The difference between the market reserve, $V_{\tau^*_{ph}}$, based on the policyholder being completely rational, and the market reserve, $V_{\tau^*_{pf}}$, based on the strategy which is the worst for the pension fund.

5.8 Conclusion

In this paper we have proposed a way to model surrender, as happening with an intensity that depends on how profitable it is to surrender. The model has a parameter denoted the rationality parameter that controls how strongly the behaviour of the policyholder is affected by profitability. We have identified sufficient conditions and given a probabilistic proof of the intuitive result that when the rationality parameter increases to infinity the reserve converges to the reserve based on completely rational behaviour.

If expenses upon surrender are not taken into account we have a duality between the contract value from the policyholder's point of view and from the pension fund's point of view, respectively. This duality implies that completely rational behaviour of the policyholder corresponds to the worst case scenario for the pension fund. If expenses are included in the model this duality is lost. However, we have shown that if there is an upper bound on the expenses upon surrender, the this is also an upper bound for the difference between the worst case reserve for the pension fund and the value based on completely rational behaviour.

In the model studied first we assumed that the surrender value may be calculated prior to the calculation of the market reserve. We have also investigated a model where the surrender value may depend on the market reserve. This possibly results in more involved differential equations for the market reserve and the technical reserve. However, if the surrender value is bounded from above by the market reserve, then the surrender intensity of the technical basis may be set to zero, and the surrender value may be calculated prior to the calculation of the market reserve.

In the numerical section we first considered a model where expenses were not taken into account. Through numerical examples we compared the development of the technical reserve with the development of the market reserve under various assumptions of the surrender intensity and under various interest rate scenarios. A key finding is that the common method in Denmark of taking the maximum of the technical reserve and the market reserve under an assumption of no surrender does not give an upper bound for the worst case reserve. We also see how traditional Danish expenses have little influence on the market reserves in the surrender model we suggest.

Appendix

5.A Proof of Convergence

We Proof the convergence result of Theorem 5.5.1. The proof is divided in two parts. One part associated with the risk from the ν_θ based stopping time surrendering before the optimal time u^* and another part associated with the risk from the ν_θ based stopping time surrendering after the optimal time u^*. For this reason we define an intermediate reserve, W_θ. The surrender strategy related to W_θ resembles the one related to V_θ. The only difference is that the strategy related to W_θ does not surrender before the optimal time. Mathematically we make the following definition. Let $\hat{\tau}_{\theta,t}$ be the minimum of n and a stopping time for which the policyholder surrenders at time u with intensity $\nu_\theta(u)1_{(u\geq u^*(t))}$. We may write this stopping time in a convenient way by introducing stopping times, $\hat{\tau}^i_{\theta,t}$, given recursively by $\hat{\tau}^0_{\theta,t} \equiv 0$ and $\hat{\tau}^i_{\theta,t}$ for $i \in \mathbb{N}$ is the minimum of n and of surrendering with intensity $\nu_\theta(u)1_{(u\geq \hat{\tau}^{i-1}_{\theta,t})}$. With these definitions we get:

$$(I, \hat{\tau}_{\theta,t}) \stackrel{d}{=} (I, \sum_{i=1}^{\infty} \hat{\tau}^i_{\theta,t} 1_{(\hat{\tau}^{i-1}_{\theta,t} < u^*(t) \leq \hat{\tau}^i_{\theta,t})}).$$

This identity comes from renewal theory and the memoryless property of the exponential distribution. It says that it does not matter if we set the surrender intensity to zero before the optimal time or if we make the policy holder regret her decision every time she is about to surrender before the optimal time. We denote for $s \in [t, n]$ by $W_\theta(t, s)$ the reserve at time s associated with the surrender strategy $\hat{\tau}_{\theta,t}$. Then, from the identity above we find that

$$W_\theta(t, s) = V(s) + \sum_{i=1}^{\infty} \mathbb{E}_s \left[e^{-\int_s^{\hat{\tau}^i_\theta} r(u)+\mu(u)du} (G(\hat{\tau}^i_\theta) - V(\hat{\tau}^i_\theta)) 1_{(\hat{\tau}^{i-1}_\theta < u^*(t) \leq \hat{\tau}^i_\theta)} \right].$$

It follows that if $u^*(t) = n$ then $W_\theta(t, s) = V(s) = W(s)$.

1. First we show that for every $t \in [0, n]$:

$$\liminf_{\theta\to\infty} V_\theta(t) \le \liminf_{\theta\to\infty} W_\theta(t,t).$$

To prove this we use, given $t \in [0, n]$ and $\varepsilon > 0$, the following notation about stopping times, τ:

$$\begin{aligned}\{\tau \ good\} &= \{G(\tau) - V_\theta(\tau) \ge 0\},\\ \{\tau \ ok\} &= \{G(\tau) - V_\theta(\tau) \in [-\varepsilon, 0)\},\\ \{\tau \ bad\} &= \{G(\tau) - V_\theta(\tau) < -\varepsilon\}.\end{aligned}$$

Thus, a stopping time, τ, is called good when it is profitable to surrender at the corresponding time, and it is called bad when the policy holder loses more than ε on surrendering. In the following, let $u^* \equiv u^*(t)$ and let $\hat\tau^i_\theta \equiv \hat\tau^i_{\theta,t}$. By induction one can show that for every $m \in \mathbb{N}$:

$$\begin{aligned}V_\theta(t) &= V(t) + \mathbb{E}_t\left[e^{-\int_t^{\hat\tau^1_\theta} r(u)+\mu(u)du}(G(\hat\tau^1_\theta) - V(\hat\tau^1_\theta))\right]\\ &\ge V(t)\\ &\quad + \sum_{i=1}^m \mathbb{E}_t\left[e^{-\int_t^{\hat\tau^i_\theta} r(u)+\mu(u)du}(G(\hat\tau^i_\theta) - V(\hat\tau^i_\theta))1_{(\hat\tau^{i-1}_\theta < u^* \le \hat\tau^i_\theta, \hat\tau^1_\theta, \dots, \hat\tau^{i-1}_\theta ok \ or \ good)}\right]\\ &\quad + \mathbb{E}_t\left[e^{-\int_t^{\hat\tau^{m+1}_\theta} r(u)+\mu(u)du}(G(\hat\tau^{m+1}_\theta) - V(\hat\tau^{m+1}_\theta))1_{(\hat\tau^m_\theta < u^*, \hat\tau^1_\theta, \dots, \hat\tau^m_\theta ok \ or \ good)}\right]\\ &\quad + \sum_{i=1}^m \mathbb{E}_t\left[e^{-\int_t^{\hat\tau^i_\theta} r(u)+\mu(u)du}(G(\hat\tau^i_\theta) - V(\hat\tau^i_\theta))1_{(\hat\tau^i_\theta < u^*, \hat\tau^1_\theta, \dots, \hat\tau^{i-1}_\theta ok \ or \ good, \hat\tau^i_\theta bad)}\right]\\ &\quad - \varepsilon \sum_{i=1}^m \mathbb{E}_t\left[e^{-\int_t^{\hat\tau^i_\theta} r(u)+\mu(u)du}1_{(\hat\tau^i_\theta < u^*, \hat\tau^1_\theta, \dots, \hat\tau^{i-1}_\theta ok \ or \ good, \hat\tau^i_\theta ok)}\right] \qquad (5.14)\end{aligned}$$

The idea is that the reserve V_θ corresponds to the technical reserve, V, plus the expected gain from surrender. We investigate what happens if the policyholder regrets to surrender. The impact if the policy holder regrets to surrender at the observed stopping time, $\hat\tau^1_\theta$, depends on whether this stopping time was good, ok or bad. If the stopping time is good, then we know that the value of the gain of surrender is at least as high as waiting for the next time to surrender, and if the stopping time is ok, then we know that the value of the gain of surrender is at most ε worse than waiting for the next time to surrender. In the above expression we have made these judgements for up to m surrender possibilities before the optimal time.

The sum in the first line corresponds to the case when one of the first m stopping times reaches beyond the optimal time, u^*. The terms of the second line correspond to the case when all of the first m stopping times are before the optimal time, u^*, and they have all been ok or good. In this case, the value of the gain of surrendering at the first stopping time is no higher than waiting for the $m+1$'th stopping time. The sum in the third line corresponds to the case when one of the m first stopping times is bad and is before the optimal time, u^*. The sum of the fourth line is a correction of the ε-small loses from ok stopping times.

If we display the bound relative to W_θ instead of relative to the technical reserve, V, then we get the following expression:

$$\begin{aligned}
&V_\theta(t)\\
&\geq W_\theta(t,t) - \sum_{i=1}^{\infty} \mathbb{E}_t \left[e^{-\int_t^{\hat{\tau}_\theta^i} r(u)+\mu(u)du} (G(\hat{\tau}_\theta^i) - V(\hat{\tau}_\theta^i)) 1_{(\hat{\tau}_\theta^{i-1} < u^* \leq \hat{\tau}_\theta^i, \exists j \in \{1,\dots,i-1\}:\ \hat{\tau}_\theta^j bad)} \right]\\
&\quad - \sum_{i=m+1}^{\infty} \mathbb{E}_t \left[e^{-\int_t^{\hat{\tau}_\theta^i} r(u)+\mu(u)du} (G(\hat{\tau}_\theta^i) - V(\hat{\tau}_\theta^i)) 1_{(\hat{\tau}_\theta^{i-1} < u^* \leq \hat{\tau}_\theta^i, \hat{\tau}_\theta^1,\dots,\hat{\tau}_\theta^{i-1} ok\ or\ good)} \right]\\
&\quad + \mathbb{E}_t \left[e^{-\int_t^{\hat{\tau}_\theta^{m+1}} r(u)+\mu(u)du} (G(\hat{\tau}_\theta^{m+1}) - V(\hat{\tau}_\theta^{m+1})) 1_{(\hat{\tau}_\theta^m < u^*, \hat{\tau}_\theta^1,\dots,\hat{\tau}_\theta^m ok\ or\ good)} \right]\\
&\quad + \sum_{i=1}^{m} \mathbb{E}_t \left[e^{-\int_t^{\hat{\tau}_\theta^i} r(u)+\mu(u)du} (G(\hat{\tau}_\theta^i) - V(\hat{\tau}_\theta^i)) 1_{(\hat{\tau}_\theta^i < u^*, \hat{\tau}_\theta^1,\dots,\hat{\tau}_\theta^{i-1} ok\ or\ good, \hat{\tau}_\theta^i bad)} \right]\\
&\quad - \varepsilon \sum_{i=1}^{m} \mathbb{E}_t \left[e^{-\int_t^{\hat{\tau}_\theta^i} r(u)+\mu(u)du} 1_{(\hat{\tau}_\theta^i < u^*, \hat{\tau}_\theta^1,\dots,\hat{\tau}_\theta^{i-1} ok\ or\ good, \hat{\tau}_\theta^i ok)} \right].
\end{aligned}$$

In the limit of θ, ε and m, then W_θ is the only term which does not converge to 0. To see this, notice that there exists some $K > 0$ such that for all $u \in [t,n]$:

$$G(u) - V(u), \in [-K,K], \quad \text{and} \quad G(u) - V_\theta(u) \in [-K,K].$$

That is, for any stopping time, the adjustment $G - V$ is bounded by K. Thereby we may further bound the value of V_θ by replacing each of these adjustments with $-K$ times an upper bound of the probability of the corresponding event:

$$\begin{aligned}V_\theta(t) \geq W_\theta(t,t) &- K\mathbb{P}_t(\exists j \in \mathbb{N} :\ \hat{\tau}_\theta^j\ bad, \hat{\tau}_\theta^j < u^*) - K \sum_{i=m+1}^{\infty} \mathbb{P}_t(\hat{\tau}_\theta^{i-1} < u^* \leq \hat{\tau}_\theta^i) \\ &- K\mathbb{P}_t(\hat{\tau}_\theta^m < u^*, \hat{\tau}_\theta^1, \dots, \hat{\tau}_\theta^m ok\ or\ good) - K\mathbb{P}_t(\exists j \in \mathbb{N} :\ \hat{\tau}_\theta^j\ bad, \hat{\tau}_\theta^j < u^*) \\ &- \varepsilon \sum_{i=1}^{m} \mathbb{P}_t(\hat{\tau}_\theta^i < u^*, \hat{\tau}_\theta^i ok) \\ \geq W_\theta(t,t) &- K(1 - e^{(n-t)\bar{h}_\theta(-\varepsilon)}) - K \sum_{i=m+1}^{\infty} \mathbb{P}_t(\hat{\tau}_\theta^{i-1} < u^* \leq \hat{\tau}_\theta^i) \\ &- K\mathbb{P}_t(\hat{\tau}_\theta^m < u^*) - K(1 - e^{(n-t)\bar{h}_\theta(-\varepsilon)}) - \varepsilon n.\end{aligned}$$

Given θ and ε, then this holds for every n. Thus, the second sum can be made arbitrarily small and so can $\mathbb{P}_t(\hat{\tau}_\theta^n < u^*)$, the later follows because given θ, then the intensity of surrender is bounded on $[0, n]$ and thus the distribution of the number of $\hat{\tau}_\theta^i$ before u^* is bounded by a Poisson distribution. Thereby:

$$\liminf_{\theta \to \infty} V_\theta(t) \geq \liminf_{\theta \to \infty} W_\theta(t,t).$$

We find from the calculations above that the lower bound holds because the surrender strategy of W_θ and V_θ only differs by the strategy of W_θ, regretting every surrender before the optimal time. The impact of this difference is bounded because the following main reasons: The probability of a bad stopping time converges to zero in the limit because of (5.7). The number of ok or good stopping times occurring before the optimal time is finite. Regret of a good stopping time decreases the value. Regret of an ok stopping time has an impact bounded by ε. At last, the technical calculations justify that the convergence of ε does not cancel the impact of the convergence of (5.7).

2. Consider some arbitrary $t \in [0, n]$. We wish to show that:

$$W_\theta(t,t) \to W(t), \quad \theta \to \infty.$$

Let $u^* \equiv u^*(t)$ and $\hat{\tau}_\theta = \hat{\tau}_{\theta,t}$, and notice that since the policyholders related to W_θ, W and V_θ behave similarly before time u^*, then convergence at time t corresponds to convergence at time u^*. This is seen from:

$$\begin{aligned}W(t) - W_\theta(t,t) &= \mathbb{E}_t\Big[e^{-\int_t^{u^*} r(u)+\mu(u)du}(G(u^*) - V(u^*)) \\ &\qquad -e^{-\int_t^{\hat{\tau}_\theta} r(u)+\mu(u)du}(G(\hat{\tau}_\theta) - V(\hat{\tau}_\theta))\Big] \\ &= e^{-\int_t^{u^*} r(u)+\mu(u)du}(W(u^*) - W_\theta(u^*, u^*)) \\ &= e^{-\int_t^{u^*} r(u)+\mu(u)du}(W(u^*) - V_\theta(u^*)).\end{aligned}$$

Thereby, it is sufficient to prove that $V_\theta(u^*) \to W(u^*)$ when $\theta \to \infty$. Either this holds, or there is some $\varepsilon_1 > 0$ and some sequence $(\theta_i)_{i\in\mathbb{N}}$ converging to infinity such that for all $i \in \mathbb{N}$: $V_{\theta_i}(u^*) < W(u^*) - 2\varepsilon_1$. Thereby $V_{\theta_i}(u^*) < G(u^*) - 2\varepsilon_1$.

The derivative of V_{θ_i} is uniformly bounded over i as long as $V_{\theta_i} < G$. Thus, there exists some δ_1 such that $V_{\theta_i}(u) \le G(u) - \varepsilon_1$ for $u \in [u^*, u^* + \delta_1]$. For this time interval the gain of surrender compared to waiting is at least ε_1, and thereby, for this time interval the intensity for surrender is at least $\underline{h}_{\theta_i}(\varepsilon_1)$.

As V is continuous, then, for every $\varepsilon_2 > 0$, there exists some δ_2 such that $(G(u^*) - V(u^*)) - e^{-\int_{u^*}^t r(u)+\mu(u)du}(G(t) - V(t)) \le \varepsilon_2$ for $t \in [u^*, u^* + \delta_2]$. That is, if surrender happens within time δ_2 of the optimal time then the loss of the delay is at most ε_2.

Now, let $\delta = \delta_1 \wedge \delta_2$. Then the loss of surrender according to $\hat{\tau}_{\theta_i}$ instead of at the optimal time is bounded in the following way:

$$
\begin{aligned}
&W(u^*) - V_{\theta_i}(u^*) \\
&= \mathbb{E}\left[(G(u^*) - V(u^*)) - (G(\hat{\tau}_{\theta_i}) - V(\hat{\tau}_{\theta_i}))e^{-\int_{u^*}^{\hat{\tau}_{\theta_i}} r(x)+\mu(x)dx} \,\middle|\, \hat{\tau}_{\theta_i} \le \delta \right] \mathbb{P}(\hat{\tau}_{\theta_i} \le \delta) \\
&\quad + \mathbb{E}\left[(G(u^*) - V(u^*)) - (G(\hat{\tau}_{\theta_i}) - V(\hat{\tau}_{\theta_i}))e^{-\int_{u^*}^{\hat{\tau}_{\theta_i}} r(x)+\mu(x)dx} \,\middle|\, (\hat{\tau}_{\theta_i} \le \delta)^c \right] \mathbb{P}((\hat{\tau}_{\theta_i} \le \delta)^c) \\
&\le \varepsilon_2 + \mathbb{E}\left[(G(u^*) - V(u^*)) - (G(\hat{\tau}_{\theta_i}) - V(\hat{\tau}_{\theta_i}))e^{-\int_{u^*}^{\hat{\tau}_{\theta_i}} r(x)+\mu(x)dx} \,\middle|\, (\hat{\tau}_{\theta_i} \le \delta)^c \right] e^{-\delta \underline{h}_{\theta_i}(\varepsilon_1)} \\
&\le \varepsilon_2 + 2Ke^{-\delta \underline{h}_{\theta_i}(\varepsilon_1)}.
\end{aligned}
$$

Thus $V_{\theta_i}(u^*) \to W(u^*)$ as $\theta \to \infty$, and the result follows.

References for Chapter 5

[1] Bacinello, A. R. (2003). Fair Valuation of a Guaranteed Life Insurance Participating Contract Embedding a Surrender Option. *J Risk Insur* **70(3)**, 461–487. doi: 10.1111/1539-6975.t01-1-00060

[2] Buchardt, K. and Møller, T. (2013). Life insurance cash flows with policyholder behaviour. *Preprint (submitted)*, available at http://www.math.ku.dk/ buchardt/.

[3] Buchardt, K., Møller, T. and Schmidt, K. B. (2013). Cash flows and policyholder behaviour in the semi-Markov life insurance setup. *Scand Actuar J* **6**, 765–798.

[4] Christiansen, M. C, Denuit, M. M. and Dhaene, J. (2014). Reserve-dependent benefits and costs in life and health insurance contracts. *Insur Math Econ* **57**, 132–137.

[5] Eling, M. and Kiesenbauer, D. (2014). What Policy Features Determine Life Insurance Lapse? An analysis of the German Market. *J Risk Insur* **81(2)**, 241–269. doi: 10.1111/j.1539-6975.2012.01504.x.

[6] Forsyth, P. A. and Vetzal, K. R. (2002). Quadratic convergence for valuing American options using a penalty method. *SIAM J Sci Comput* **23(6)**, 2095–2122. doi: 10.1137/S1064827500382324.

[7] Gad, K. S. T. and Pedersen, J. L. (2014). Profit dependent Exercise of the American Put, *Preprint*, available at arxiv.org/abs/1410.1287.

[8] Gatzert, N. (2009). Implicit options in life insurance: An overview, *Zeitschrift fÃĳr die gesamte Versicherungswissenschaft*, **98(2)**, 141–164.

[9] De Giovanni, D. (2010). Lapse rate modeling: a rational expectation approach. *Scand Actuar J* **2010(1)**, 56–67.

[10] Grosen, A. and Jørgensen, P. L. (2000). Fair Valuation of Life Insurance Liabilities: The Impact of Interest Rate Guarantees, Surrender Options, and Bonus Policies. *Insur Math Econ* **26(1)**, 37–57.

[11] Henriksen, L. F. B., Nielsen, J. W., Steffensen, M. and Svensson, C. (2014). Markov chain modeling of policyholder behaviour in life insurance and pension. *Eur Actuar J* **4(1)**, 1–29. doi: 10.1007/s13385-014-0091-2.

[12] IFRS Exposure Draft ED/2013/7 (2013). *Insurance contracts*, http://www.ifrs.org/Current-Projects/IASB-Projects/Insurance-Contracts/Exposure-Draft-June-2013/Documents/ED-Insurance-Contracts-June-2013.pdf.

[13] Kyprianou, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications.* Springer

[14] Møller, T. and Steffensen, M. (2007). *Market-Valuation Methods in Life and Pension Insurance.* Cambridge.

[15] Shiryayev, A. N. (1978). *Optimal Stopping Rules.* Springer-Verlag.

[16] Solvency II Commission Delegated Regulation (2014). *Commission Delegated Regulation (EU) 2015/35*, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R0035.

[17] Solvency II Directive (2009). *Directive 2009/138/EC*, http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009L0138.

[18] Steffensen, M. (2002). Intervention options in life insurance. *Insur Math Econ* **31**, 71–85.

6. Reserves and Cash Flows under Stochastic Retirement

Kamille Sofie Tågholt Gad, Jeppe Woetmann Nielsen

Abstract

Uncertain time of retirement and uncertain structure of retirement benefits are risk factors for life insurance companies. Nevertheless, classical life insurance models assume these are deterministic. In this paper we include the risk from stochastic time of retirement and stochastic benefit structure in a classical finite state Markov model for a life insurance contract. We include discontinuities in the distribution of the retirement time. First, we derive formulas for appropriate scaling of the benefits according to the time of retirement and discuss the link between the scaling and the guarantees provided. Stochastic retirement creates a need to rethink the construction of disability products for high ages and ways to handle this are discussed. We show how to calculate market reserves and how to use modified transition probabilities to calculate expected cash flows without significantly more complexity than in the traditional model. At last, we demonstrate the impact of stochastic retirement on market reserves and expected cash flow in numerical examples.

Keywords: Behavioural option; Solvency II; benefit scaling; ordinary differential equation; discontinuous transition probabilities.

6.1 Introduction

In classical life insurance models the time of retirement and the structure of the retirement benefits are typically settled at the beginning of the contract. However, in practice pension funds often allow the policyholders to both change the time of retirement and convert between different benefit structures. Conversion of benefit structures may for example be converting a pension sum into a life annuity.

Modelling retirement as a deterministic time with a deterministic benefit structure does not take the risk of these options into account and this is a source of error both in market reserves and, to an even higher degree in expected cash flows. The forthcoming Solvency II rules require that any contractual option is taken into account. In this paper we address challenges from modelling the time of retirement and the structure of the benefits as stochastic. Combined we call it *stochastic retirement*.

In classical models, the state of the policyholder is described by a finite state Markov chain. Here, the states of premium paying and retired are the same, and at some fixed time the payments change sign corresponding to the change from premiums to benefits. We introduce a stochastic time of retirement by letting the states of premium paying and retired be two different states in the Markov model. We assume that all transition probabilities are deterministic and known, but unlike the other transitions, we let retirement happen with positive probability at predefined time points.

The idea of modelling retirement by a state in the Markov model has been mentioned before, e.g. in [8] and [12]. In [8] the idea of modelling retirement as a separate state in the classical actuarial Markov model is mentioned en passant, but no calculations or discussions are done. In [12] the Markov model with a retirement state is used for calculations of demographical variables such as time to retirement. However, impacts on retirement savings are not mentioned. To the knowledge of the authors we present the first discussion of the actuarial implications of modelling stochastic retirement. We find that this subject deserves attention for several reasons. One reason is the relevance of the numerical impact of the forthcoming Solvency II regulations. Another reason is the unconventional considerations we find is needed for constructing reasonable products in a model with stochastic retirement. A third reason is the mathematical impact on the calculation of reserves, benefits and expected cash flows induced by stochastic retirement through e.g. discontinuous transition probabilities, interaction with the free policy option and our introduction of an auxiliary Markov model for describing benefit conversions.

Changing the time of retirement or the structure of the benefits is a policyholder option just like conversion to free policy or surrender. These two options are typically modelled by adding states to the model, and modelling of the intensity of exercising them have been studied thoroughly. The studies ranges from purely random decisions, as mentioned in e.g. [2] and [10], to optimal exercising strategies as mentioned in e.g. [1], [6] and [10]. In between we find models which are perhaps more realistic where policyholder actions happen randomly, but where the intensity depends on some factors such as the profit from taking the action as in [4], or the interest rate as in [5]. There is a large amount of literature on explanatory variables and we refer to [3] for further references and an overview. Many of the studies done on the modelling of the exercise of the surrender option and the free policy option would be equally relevant to do for the exercise of the options of stochastic retirement. However, that is not our focus in this paper.

In a model with stochastic time of retirement it is reasonable to have the size of the benefits depend on the time of retirement. As is common for modelling of other policyholder behaviours (see e.g. [2] and [7]), we let the benefits be affected in a way such that the risk sum of retirement is zero under the technical basis. Scaling the benefits this way makes the policyholder pay for her own retirement under the technical basis at any time. The guarantee provided by the technical basis thereby remains in force after the exercise of the option of changing time or structure of the retirement. This means that even after the exercise of one of the options the reserves bears interest with the technical interest rate, and risk premiums are based on the technical transition intensities. The impact of stochastic retirement bears some resemblance to the impact of conversion to free policy or surrender. However, whereas the free policy and the surrender options only induce a risk of reduced or earlier benefits, stochastic retirement also induces a risk of increased, postponed benefits. Its impact on market reserves relies heavily on the extent of the guarantee provided by the technical basis. It is not obvious that a pension fund wants to have deferred retirement covered by the guarantee from the technical basis. In the case where deferred retirement after some reference age of the contract is not covered by the guarantee from the technical basis, we may model the time of retirement with this reference age as an upper limit for the time of retirement. Policyholders who want to keep saving after this time may then use their retirement benefits to start a new contract under new terms.

In this paper we study, in Section 6.2, a simple life-death model with a retirement state added. We determine how to scale the benefits, we show how to set up a Thiele differential equation for the market reserves, and we determine formulas for expected cash flows. This example serves to introduce the method in a very simple setup. In Section 6.3 we study a complex model which includes both cycles (through disability and rehabilitation) and other behaviour scalings (through conversion to free policy). As is seen in [2] the combination of cycles and behaviour scaling may complicate the calculation of expected cash flows, and likewise we find that calculations of expected cash flows are eased by introduction of modified transition probabilities which incorporates expected scalings. Studying the complex model we find that one has to be careful in the description of payments and transition probabilities, as it is easy to formalize models and products which are not practically meaningful. In Section 6.4 the possibility of conversion of benefits is added to the complex model. We find that a Markov model, that has one state for each benefit structure, is equivalent to our model in the sense that it produces the same expected discounted cash flows. In Section 6.5 we demonstrate the impact of stochastic retirement on benefits, market reserves and expected cash flows by numerical examples from the setup in Section 6.2 and Section 6.4.

6.2 Stochastic Retirement in a Simple Model

In this section we consider a simple model and a simple life insurance contract to illustrate the main lines of the implications from stochastic retirement clearly. We consider a life-death model with a retirement state as illustrated

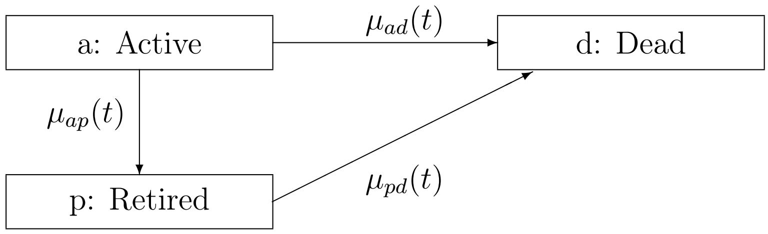

Figure 6.1: Simple retirement model.

in Figure 6.1. We denote the three states a (active), d (dead) and p (retired). The state of a policyholder over time is described by a cádlàg, finite state Markov process, $(Z_t)_{t\geq 0}$, taking values in $E = \{a, p, d\}$. In the continuous time Markov models commonly used in life insurance, it is natural to assume that the distribution of the transition times is continuous. However, there are multiple reasons to expect that there are time points at which there is a positive probability of retirement. These reasons are among others monetary advantages coming from legislative rules, and standard dates for termination of employment. Thus, even though we make retirement stochastic, we want to place ourselves in between the stochasticity of death and the usual deterministic modelling of retirement. We do this by introducing deterministic time points at which active policyholders have a positive probability of retiring. Let $\hat{P}$ denote the probability measure of the technical basis of the contract. That is, $\hat{P}$ is the measure used for setting equivalence premiums and benefits. The distribution of Z is given from the transition probabilities defined by

$$\hat{p}_{jk}(t,s) = \hat{P}(Z_s = k | Z_t = j), \tag{6.1}$$

for $j, k \in E$. We assume that for every $t \geq 0$ the transition intensities are well defined by

$$\hat{\mu}_{jk}(t) = \lim_{\varepsilon \downarrow 0} \tfrac{1}{\varepsilon} \hat{p}_{jk}(t, t+\varepsilon), \tag{6.2}$$

for $j,k \in E$ with $j \neq k$, and we assume that they are continuous and that $\hat{\mu}_{pa} = \hat{\mu}_{da} = \hat{\mu}_{dp} = 0$. For a fixed set of time points $t_1,\dots,t_n$ and probabilities $\hat{p}_1,\dots,\hat{p}_n$ we assume:

$$\hat{p}_{jk}(t_h-,t_h) = \hat{P}(Z_{t_h} = k | Z_{t_h-} = j) = \begin{cases} \hat{p}_h & \text{if } (j,k) = (a,p), \\ 0 & \text{otherwise}, \end{cases} \tag{6.3}$$

for $j,k \in E$ and we assume $\hat{p}_n = 1$, such that $t_n = T_{max}$ is a maximum age of retirement. Such a maximum is chosen to ease computations in practice when valuing contracts. For each $h = 0,\dots,n-1$ and $t \le s$ we assume $s \mapsto \hat{p}_{jk}(t,s)$ are continuous on $[t_h, t_{h+1})$ with limits towards the right endpoints of the interval. Thereby

$$\hat{p}_{jk}(t,t_h) - \hat{p}_{jk}(t,t_h-) = \begin{cases} \hat{p}_{ja}(t,t_h-)\hat{p}_h & \text{if } k = p, \\ -\hat{p}_{ja}(t,t_h-)\hat{p}_h & \text{if } k = a, \\ 0 & \text{otherwise}, \end{cases} \tag{6.4}$$

for $j,k \in E$. From the Kolmogorov equations we have that

$$\tfrac{\partial}{\partial s}\hat{p}_{jk}(t,s) = -\hat{p}_{jk}(t,s)\hat{\mu}_{k.}(s) + \sum_{l \in E, l \neq k} \hat{p}_{jl}(t,s)\hat{\mu}_{lk}(s), \tag{6.5}$$

on (t_h, t_{h+1}) for $j,k \in E$ and where $\hat{\mu}_{k.}(s) = \sum_{\{l \in E:\ l \neq k\}} \hat{\mu}_{kl}(s)$. Combined with (6.4) and $\hat{p}_{jk}(t,t) = 1_{(j=k)}$ for $j,k \in E$ this describes the transition probabilities. Except for the retirement state and the discontinuities in the transition probabilities, this model is a classical life insurance model as described in e.g. [9], [10] and [11].

6.2.1 Scaling the Benefits

We consider a contract containing a premium, π. As it is common in pension systems, upon retirement at time t there is a lump sum payment, denoted $b_{ap}(t)$, and a beginning of a life annuity payment, denoted $b_p(t)$. Since retirement is a policyholder choice, we want benefits to depend on the time of retirement to be able to reward the policyholder who retires late, and to avoid speculation. We wish to describe the dependence of benefits on the time of retirement through scaling functions. Therefore, we choose a reference retirement time, T, and consider an alternative model where the time of retirement is deterministically T. Then we determine benefits b_{ap}^T, b_p^T for this model such that the equivalence principle is fulfilled. Now, for the model with stochastic retirement we define scaling functions that scale the reference benefits according to the time of retirement. We speak of scaling functions as *retirement factors*. Let ρ_1 be the function that gives the factor for scaling the life annuity and let ρ_3 be the function that gives the factor for scaling the pension sum. The subscripts reflect the Danish tax codes.

Let $I_a(t) = 1_{(Z_t=a)}$, $I_p(t) = 1_{(Z_t=p)}$, $dN_{ap}(t) = 1_{(Z_t=p, Z_{t-}=a)}$, and let U be the process of the duration the policyholder has been retired. Then the contract has a payment stream given by:

$$\mathrm{d}B(t) \quad = \quad -\pi I_a(t)\mathrm{d}t + \rho_1(t-U_t)b_p^T I_p(t)\mathrm{d}t + \rho_3(t)b_{ap}^T \mathrm{d}N_{ap}(t), \quad (6.6)$$

and we expect $\rho_1(T) = \rho_3(T) = 1$. Let $\hat{r}$ denote the technical interest rate and let $\hat{V}_a(t)$ denote the retrospective technical reserve in the active state at time t. In the retirement state we define two prospective technical reserves: $\hat{V}_p(t,u)$ is the reserve at time t after the duration u in the state, and $\hat{V}_p^T(t)$ is a reserve for a life-annuity with the reference benefits. Then $\hat{V}_p(t,u) = \rho_1(t-u)\hat{V}_p^T(t)$. Notice that $\hat{V}_p^T$ does not depend on the time of retirement. This is a special feature of the life annuity, but it would not be a problem to extend the results of this paper to duration dependent reference benefits.

Except for the discontinuity points, $t_1, \ldots, t_n$, then $\hat{V}_a$ is continuous. Thiele's differential equation gives

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}\hat{V}_a(t) &= \pi + \left(\hat{r} + \hat{\mu}_{ad}(t)\right)\hat{V}_a(t) \\ &\quad - \hat{\mu}_{ap}(t)\left(\rho_3(t)b_{ap}^T + \rho_1(t)\hat{V}_p^T(t) - \hat{V}_a(t)\right), \end{aligned} \quad (6.7)$$

and in discontinuity points we have

$$\hat{V}_a(t_h) - \hat{V}_a(t_h-) = -\hat{p}_h\left(\rho_3(t_h)b_{ap}^T + \rho_1(t_h)\hat{V}_p^T(t_h) - \hat{V}_a(t_h)\right). \quad (6.8)$$

How to choose the scaling depends on how we understand the guarantee provided by the technical basis. As mentioned in the introduction we choose the scaling such that the risk sums of the policyholders actions are zero. That is the expression in the large parentheses in (6.7) and (6.8). Thereby, the guarantee from the technical basis covers through the exercise of the option of changing the time of retirement. This means that no matter the time of the retirement or the benefit structure, the reserve bears interest with the technical interest rate, and risk premiums are based on the technical transition probabilities. With this choice of scaling we see from (6.8) that the retrospective technical reserve of the active state, $\hat{V}_a$, becomes continuous, and it follows from (6.7) that before time T, $\hat{V}_a$ equals the classical retrospective technical reserve from the reference model with deterministic retirement at time T.

To calculate a prospective technical reserve in the active state, $\hat{V}_a^{prosp}$ we need to have a terminal boundary condition for (6.7). To get this we need to either assume an upper retirement age or make assumptions about the relation between the intensities for death and for retirement for high ages. We choose to have an upper retirement age, T_{max}, as this is easiest for computations. In this case we get from $p_n = 1$, $t_n = T_{max}$ and (6.8) the terminal condition

$$\hat{V}_a^{prosp}(T_{max}-) = \rho_3(T_{max})b_{ap}^T + \rho_1(T_{max})\hat{V}_p^T(T_{max}).$$

Since the retirement factors are chosen such that this equal $\hat{V}_a(T_{max}-)$, we find that the prospective reserve equals the retrospective, and our choice for the retirement factors ensures that the equivalence principle is fulfilled.

However, a zero risk sum for retirement still leaves us with some choices regarding how to specify the relation between the two retirement factors. The choices we face here resemble the choices one faces upon conversion to free policy. For the conversion to free policy it is often seen in the literature (see e.g. [2], [7]) that the factors of the different benefits are either kept identical or one factor is fixed at a desired level. However, in practice the saving is often divided into partial reserves according to the benefit structure, and it is natural to wish to have each of these pay for itself in a way that makes the risk terms for each of the partial reserves zero. This is in Denmark done for tax reasons. We assume that the reserve is divided in two. Let $\hat{V}_a^1$ denote the reserve for the annuity and π^1 the annuity premium, and likewise let $\hat{V}_a^3$ and π^3 denote pension sum reserve and premium.

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{V}_a^1(t) = \pi^1 + \hat{r}\hat{V}_a^1(t) + \hat{\mu}_{ad}(t)\hat{V}_a^1(t) - \hat{\mu}_{ap}(t)(\rho_1(t)\hat{V}_p^T(t) - \hat{V}_a^1(t)),$$
$$\frac{d}{dt}\hat{V}_a^3(t) = \pi^3 + \hat{r}\hat{V}_a^3(t) + \hat{\mu}_{ad}(t)\hat{V}_a^3(t) - \hat{\mu}_{ap}(t)(\rho_3(t)b_{ap}^T - \hat{V}_a^3(t)).$$

By choosing the factors such that risk sums from retiring are zero we do not get any discontinuities. As for the combined reserve we get that the retrospective reserves equal the technical reserves from the reference model with fixed retirement at time T and the equivalence principle is kept. The retirement factors become

$$\rho_1(t) = \frac{\hat{V}_a^1(t)}{\hat{V}_p^T(t)} \quad \text{and} \quad \rho_3(t) = \frac{\hat{V}_a^3(t)}{b_{ap}^T} = \frac{\hat{V}_a^3(t)}{\hat{V}_a^3(T)}. \tag{6.9}$$

Note that $\rho_1(T) = \rho_3(T) = 1$ as anticipated. Further, notice that both scaling factors may be determined from $\hat{V}_a^1$, $\hat{V}_a^3$ and $\hat{V}_p^T$, which can all be calculated from the reference benefits alone.

If we chose $\rho_1 = \rho_3$, we would find that the combined technical reserve would be unchanged by the stochastic modelling of retirement. However, money would be moved between the partial reserves upon retirement. When we instead assume that risk terms for both of the partial reserves are zero we ensure that no money is moved between the partial reserves upon retirement.

When splitting the reserve calculation in partial reserves we may assume that the policyholder has separate Markov state processes for each partial reserve. It is common in Denmark that policyholders choose to have their lump sum retirement payment paid out at a different time than the beginning of their life annuity. With separate state processes we may choose different distributions of the time of retirement for each partial reserve and allow the policyholder to retire for one benefit structure while remaining active for another.

6.2.2 Market Valuation

We saw above how our benefit scaling resulted in the technical reserves being unaffected by modelling the retirement as stochastic. However, for market values the risk terms of retirement are no longer zero and thus market values are affected by applying stochastic retirement.

We assume a market basis with a deterministic, time-dependent interest rate, r, and a distribution of Z that resembles the one under the technical measure. The only difference is that the transition intensities are replaced with μ_{ap}, μ_{ad}, μ_{pd}, and $\hat{p}_1, \ldots, \hat{p}_n$ are replaced with $p_1, \ldots, p_n$, where $p_n = 1$. the two bases agree on when the transition intensities are zero and on the time of the discontinuities in the transition probabilities. From this, transition probabilities p_{ap}, p_{ad}, p_{pd} are obtained through (6.4) and (6.5).

Let $\mathbb{E}_{a,t}$ denote expectation on the market basis given the policyholder is active at time t. Then, the prospective market reserve, $V_a(t)$, at time t given that the policyholder is active at this time is

$$V_a(t) = \mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\mathrm{d}B(s)\right].$$

Let V_p^T denote the prospective market reserve from the retirement state with reference benefits. Then, in continuity points, Thiele's differential equation for the market reserve in the active state becomes

$$\tfrac{\mathrm{d}}{\mathrm{d}t}V_a(t) = \pi + \left(r(t) + \mu_{ad}(t)\right)V_a(t) - \mu_{ap}(t)\left(\rho_3(t)b_{ap}^T + \rho_1(t)V_p^T(t) - V_a(t)\right),$$

with the terminal condition

$$V_a(T_{max}-) = \rho_3(T_{max})b_{ap}^T + \rho_1(T_{max})V_p^T(T_{max}).$$

The market reserve is discontinuous in the time points where there is a positive probability of retiring and the jumps are given by the risk term

$$V_a(t_h) - V_a(t_h-) = -p_h(\rho_3(t_h)b_{ap}^T + \rho_1(t_h)V_p^T(t_h) - V_a(t_h)).$$

As stochastic retirement makes the timing of the benefits stochastic, it is interesting to look at how the expected cash flow is affected. Accumulated expected cash flow given the policyholder is active at time t is given by

$$A_a(t,s) = \mathbb{E}_{a,t}[B(s) - B(t)]. \tag{6.10}$$

From [2] it follows that the market reserve is given by

$$V_a(t) = \int_t^\infty e^{-\int_t^s r(x)dx} \mathrm{d}A_a(t,s).$$

In Appendix 6.A we derive the market reserve. We determine an expression that allows us to immediately deduce the expected cash flow, which is given by

$$\begin{aligned} \mathrm{d}A_a(t,s) = -p_{aa}(t,s)\left(\pi \mathrm{d}s + \left(\mu_{ap}(s)\mathrm{d}s + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(s)\right)\rho_3(s)b_{ap}^T\right) \\ + p_{ap}^{\rho_1}(t,s)b_p^T \mathrm{d}s, \end{aligned} \tag{6.11}$$

where $\varepsilon_{t_h}(s) = 1_{\{s \geq t_h\}}$ and

$$\begin{aligned} p_{ap}^{\rho_1}(t,s) &= \mathbb{E}_{a,t}[I_p(s)\rho_1(s-U_s)] \qquad (6.12) \\ &= \int_t^s p_{aa}(t,\tau)\rho_1(\tau)p_{pp}(\tau,s)\left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(\tau)\right). \end{aligned}$$

The expectation in (6.12) is a kind of modified transition probability with a weight for the benefit scaling. It resembles the modified transition probabilities used in [2] for calculating cash flows in a model with a free policy option. At first it seems that this modified probability might be very time consuming to calculate, because it depends on $p_{pp}(\tau, v)$ for every value of both τ and v in $[t,\infty) \times [t,\infty)$. However, as we have closed form expressions for these probabilities, the calculation is tractable.

6.3 Stochastic Retirement in a Complex Model

We now consider a more complex model in which we add the possibility of the policyholder becomingg disabled, re-activating, or converting to free policy. We denote the states of our model a (active), p (retired), i (disabled), d (dead), and $\bar{a}$, $\bar{p}$, $\bar{i}$, $\bar{d}$, for the corresponding states after conversion to free policy. Our model is displayed in Figure 6.2. This model is interesting because we want to show the interplay between the retirement scaling, the free policy factor and disability products. Investigating the complex model we find that it is very easy to accidentally construct contracts which are not meaningful or assume policyholder behaviour which is unlikely. This is partly because of the assumption that the policyholder behaviour is stochastic and independent of everything else whereas it is actually a choice.

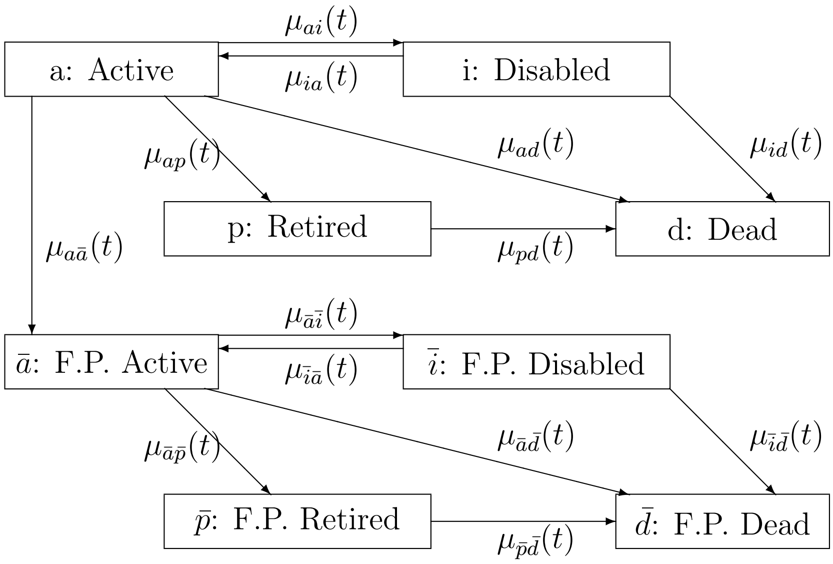

Figure 6.2: Model with disability, free policy and stochastic retirement.

As in Section 6.2 we let Z be a stochastic process that describes the state of the policyholder, such that Z takes values in $E = \{a, p, i, d, \bar{a}, \bar{i}, \bar{p}, \bar{d}\}$ and we assume Z is cádlàg and Markov. We let $\hat{P}$ denote the probability measure of the technical basis, define transition probabilities from (6.1), and assume that the intensities of (6.2) are well defined for $j, k \in E$. In accordance with Figure 6.2 we assume that only transitions given in Figure 6.2 are non-zero. As in Section 6.2 we assume that transition probabilities are continuous in the second argument on each of the intervals $[t_h, t_{h+1})$ with limits towards the left endpoint. In the points $t_1, \ldots, t_n$, for every $h = 1, \ldots, n$, we replace (6.3) with

$$\hat{p}_{jk}(t_h-, t_h) = \hat{P}(Z_{t_h} = k | Z_{t_h-} = j) = \begin{cases} \hat{p}_h & \text{if } (j,k) = (a,p), \\ \hat{p}_h^\phi & \text{if } (j,k) = (\bar{a},\bar{p}), \\ 0 & \text{otherwise}, \end{cases} \tag{6.13}$$

for $j, k \in E$ and thereby

$$\hat{p}_{jk}(t, t_h) - \hat{p}_{jk}(t, t_h-) = \begin{cases} \hat{p}_{ja}(t, t_h-)\hat{p}_h & \text{if } k = p, \\ \hat{p}_{j\bar{a}}(t, t_h-)\hat{p}_h^{\phi} & \text{if } k = \bar{p}, \\ -\hat{p}_{ja}(t, t_h-)\hat{p}_h & \text{if } k = a, \\ -\hat{p}_{j\bar{a}}(t, t_h-)\hat{p}_h^{\phi} & \text{if } k = \bar{a}, \\ 0 & \text{otherwise,} \end{cases} \tag{6.14}$$

for $j, k \in E$. Kolmogorovs differential equation (6.5) still holds on each interval (t_h, t_{h+1}).

6.3.1 A Realistic Contract

We consider a simple, but more realistic contract. However, we find that with stochastic retirement in the model even a simple contract requires careful investigation. The contract we consider consists of a premium, π, a death sum, $b_{ad}(t)$, from active to death, a disability annuity $b_i(t)$, when disabled, a pension sum of $b_{ap}(u)$ if retiring at time u, a life annuity of $b_p(u)$ if retiring at time u, and $b_{ad}(t, v)$, $b_i(t, v)$, $b_{ap}(u, v)$, $b_p(u, v)$ corresponding payments after conversion to free policy at time v. We have to be very careful to not construct a contract that does not make sense. Our motivation for scaling of benefits is to reward policyholders who retire late. However, it is not desirable to give a similar reward to a disabled policyholder who reactivates at a correspondingly high age and immediately retires. There are several ways to handle this problem. These have different levels of complexity.

One approach, the simplest, is to set $\mu_{ia}(t) = \mu_{\bar{i}\bar{a}}(t) = 0$ for $t > T$. This implies that policyholders are not able to re-activate after time T. The biggest drawback from this model is that we lose the information of whether the policyholder recovered after time T. It is likely that reactivated policyholders have lower death intensity than the disabled. However, the death intensity from the disabled state is commonly estimated from everybody in this state. If the policyholders who actually have recovered have comparable benefits to those who are still disabled, then the diversification principle ensures that the lost information is not a problem for the insurance company's calculation of the collected reserves for all policyholders.

Another approach is to force recovered policyholders after time T directly to the retirement state or to a new state specifically for late recovered disabled. In this state it would be natural to have the benefits set as if policyholder had stayed disabled. This is to have the policyholder neither gain nor loose from recovering once time T is reached. In this model we gain the possibility of managing the information on whether the policyholder has recovered and thereby calculate more precise death intensities for the single policyholder.

The last and most complicated approach we mention is to allow the policyholder to reactivate to the active state or to the free policy state. This way the policyholder can continue saving or at least have a break in receiving payments. If this model is chosen it is worth noting that we may equally well allow retired people to return to work. This modelling option has not been present in earlier models with fixed retirement date. If this model is used it is natural to let the reserve of the policyholder be unchanged upon reactivation after time T. This way the policyholders cannot speculate about whether to be declared recovered. However, there is a disadvantage that if the reserve is kept unchanged upon recovery after time T the policyholder is not able to get the same level of retirement benefits as if she had stayed disabled.

Similar problems arise if it is possible to become disabled after time T. For fixed premiums it is common in Denmark that disability benefits do not depend on when the policyholder becomes disabled. It is likely that at some point after time T the policyholders receive higher benefits from retiring than from being declared disabled. Again we are faced with options similar to the ones described above. In the following we let $\mu_{ia}(t) = \mu_{ai}(t) = \mu_{\bar{i}\bar{a}}(t) = \mu_{\bar{a}\bar{i}}(t) = 0$ for $t > T$, and assume that people who becomes disabled after time T choose to retire.

6.3.2 Scaling the Benefits

We define reference benefits from the equivalence principle when the time of retirement is deterministic, T, and when there is no conversion to free policy. The reference benefits we denote with a superscript T. We then define scaling functions that scales the benefits according to the time of retirement and conversion to free policy. Every time the policyholder takes one of the two actions, the benefits are scaled by a factor, which depends on the time of the action. Again, assume $x \in \{1, 3\}$ represents respectively the life annuity and the pension sum. For her saving for benefits of type x we use the notation that upon retirement from the active state at time u, the payments are scaled with a factor $\rho_x(u)$, upon conversion to free policy at time v the payments are scaled with a factor $\phi^x(v)$, and upon retirement at time u from the free policy state the payments are scaled with a factor $\rho_{\phi x}(u)$. We choose scaling such that risk sums for policyholder behaviour are zero. Now, it follows from (6.16) that we do not need for the factor $\rho_{\phi x}$ to depend on the time of conversion to free policy. If the policyholder converges to free policy at time v and retires at time $u \geq v$, then the reference retirement payments in total are scaled by $\phi^x(v)\rho_{\phi x}(u)$. Thus, the effect from the time of conversion to free policy and the effect from the time of retirement after conversion to free policy are multiplicative. This is natural since the time of conversion to free policy controls how long premiums are paid, and the time of retirement after conversion to free policy controls when retirement benefits are paid.

For $y, z \in E$ let $I_y(t) = 1_{(Z_t=y)}$ and $dN_{yz}(t) = 1_{(Z_{t-}=y, Z_t=z)}$, and let U be a process of the duration the policyholder has been retired, and let V be a process of the duration the policyholder has been converted to free policy. Assume the disability annuity and the death sum are paid by the partial reserve for the life annuity. Now, the contract has payment streams given by:

$$\begin{aligned} \mathrm{d}B^1(t) &= -\pi^1 I_a(t)\mathrm{d}t + b_i(t)I_i(t)\mathrm{d}t + b_{ad}(t)\mathrm{d}N_{ad}(t) + b_p^T\rho_1(t-U_t)I_p(t)\mathrm{d}t \\ &\quad + b_i(t)\phi^1(t-V_t)I_{\bar{i}}(t)\mathrm{d}t + b_{ad}(t)\phi^1(t-V_t)\mathrm{d}N_{\bar{a}\bar{d}}(t) \\ &\quad + b_p^T\rho_{\phi 1}(t-U_t)\phi^1(t-V_t)I_{\bar{p}}(t)\mathrm{d}t, \\ \mathrm{d}B^3(t) &= -\pi^3 I_a(t)\mathrm{d}t + b_{ap}^T\rho_3(t)\mathrm{d}N_{ap}(t) + b_{ap}^T\rho_{\phi 3}(t)\phi^3(t-V_t)\mathrm{d}N_{\bar{a}\bar{p}}(t). \end{aligned}$$

As for the simple model we define scalings such that for each of the partial reserves, risk sums for policyholder behaviour are zero. We use the following notation for the technical reserves: Let $\hat{V}_a^x$ denote the retrospective technical reserve of type x saving in the active state. Let $\hat{V}_i^x$ denote the prospective technical reserve of type x saving in the disabled state. Let $\hat{V}_p^T$ denote the prospective technical reserve of a life annuity with the reference benefits. Let $\hat{V}_{\bar{a}}^{x*}$ denote the prospective reserve in the free policy active state if the free policy scaling is omitted. Once converted to free policy, the policyholder cannot return to the active state. Thus, $\hat{V}_{\bar{a}}^{x*}$ is the prospective reserve of the payment stream B^{x*} with

$$\begin{aligned} \mathrm{d}B^{1*}(t) &= b_i(t)I_{\bar{i}}(t)\mathrm{d}t + b_{ad}(t)\mathrm{d}N_{\bar{a}\bar{d}}(t) + b_p^T\rho_{\phi 1}(t-U_t)I_{\bar{p}}(t)\mathrm{d}t, \\ \mathrm{d}B^{3*}(t) &= b_{ap}^T\rho_{\phi 3}(t)\mathrm{d}N_{\bar{a}\bar{p}}(t). \end{aligned}$$

Let $\hat{V}_{\bar{i}}^{x*}$ denote the prospective technical reserve for the type x saving for the free policy disabled with the payment stream B^{x*}. Since we have assumed that the disability annuity and the death sum are paid by the partial reserve for the life annuity, we let $b_i^1 = b_i$, $b_{ad}^1 = b_{ad}$ and $b_i^3 = b_{ad}^3 = 0$. Now, the Thiele differential equation for the active state and for the free policy active state for each of the partial reserves becomes

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{V}_a^x(t) &= \pi^x + \hat{r}\hat{V}_a^x(t) - \hat{\mu}_{ad}(t)\left(b_{ad}^x(t) - \hat{V}_a^x(t)\right) - \hat{\mu}_{ai}(t)\left(\hat{V}_i^x(t) - \hat{V}_a^x(t)\right) \\
&\quad - \hat{\mu}_{ap}(t)\left(1_{(x=1)}\rho_1(t)\hat{V}_p^T(t) + 1_{(x=3)}\rho_3(t)b_{ap}^T - \hat{V}_a^x(t)\right) \\
&\quad - \hat{\mu}_{a\bar{a}}(t)\left(\phi^x(t)\hat{V}_{\bar{a}}^{x*}(t) - \hat{V}_a^x(t)\right), \qquad (6.15)
\end{aligned}$$

and

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{V}_{\bar{a}}^{x*}(t) &= \hat{r}\hat{V}_{\bar{a}}^{x*}(t) - \hat{\mu}_{\bar{a}\bar{d}}(t)\left(b_{ad}^x(t) - \hat{V}_{\bar{a}}^{x*}(t)\right) - \hat{\mu}_{\bar{a}\bar{i}}(t)\left(\hat{V}_{\bar{i}}^{x*}(t) - \hat{V}_{\bar{a}}^{x*}(t)\right) \\
&\quad - \hat{\mu}_{\bar{a}\bar{p}}(t)\left(1_{(x=1)}\rho_{\phi 1}(t)\hat{V}_{\bar{p}}^T(t) + 1_{(x=3)}\rho_{\phi 3}(t)b_{ap}^T - \hat{V}_{\bar{a}}^{x*}(t)\right), \qquad (6.16)
\end{aligned}$$

with terminal conditions

$$\begin{aligned}
\hat{V}_{\bar{a}}^{1*}(T_{max}-) &= \rho_{\phi 1}(T_{max})\hat{V}_{\bar{p}}^T(T_{max}), \\
\hat{V}_{\bar{a}}^{3*}(T_{max}-) &= \rho_{\phi 3}(T_{max})b_{ap}^T.
\end{aligned}$$

We choose the scalings such that the risk sums for policyholder behaviours are zero. Thereby there are no discontinuities, and the reserves $\hat{V}_a^x$ correspond to the technical reserves from the classical model with deterministic retirement at time T. We see that in order to have zero risk term we must have scalings for retirement

$$\rho_1(t) = \frac{\hat{V}_a^1(t)}{\hat{V}_p^T(t)} \qquad \text{and} \qquad \rho_3(t) = \frac{\hat{V}_a^3(t)}{b_{ap}^T} = \frac{\hat{V}_a^3(t)}{\hat{V}_a^3(T)}.$$

Specifically we notice that $\rho_1(T) = \rho_3(T) = 1$, and that the retirement factors can be calculated exclusively from the reference benefits. The prospective technical reserves in the active state have terminal conditions

$$\begin{aligned}
\hat{V}_a^{1,prosp}(T_{max}-) &= \rho_1(T_{max})\hat{V}_p^T(T_{max}) = \hat{V}_a^1(T_{max}) \\
\hat{V}_a^{3,prosp}(T_{max}-) &= \rho_3(T_{max})b_{ap}^T = \hat{V}_a^3(T_{max}).
\end{aligned}$$

Thus, the retrospective and the prospective reserves for the active state are the same, and thus our choice of scaling implies that the equivalence principle is kept.

Scalings for conversion to free policy should fulfil

$$\phi^1(t) = \frac{\hat{V}_a^1(t)}{\hat{V}_{\bar{a}}^{1*}(t)} \qquad \text{and} \qquad \phi^3(t) = \frac{\hat{V}_a^3(t)}{\hat{V}_{\bar{a}}^{3*}(t)}. \qquad (6.17)$$

And scaling for retiring after conversion to free policy should fulfil

$$\rho_{\phi 1}(t) = \frac{\hat{V}_{\bar{a}}^{1*}(t)}{\hat{V}_p^T(t)} \qquad \text{and} \qquad \rho_{\phi 3}(t) = \frac{\hat{V}_{\bar{a}}^{3*}(t)}{b_{ap}^T} = \frac{\hat{V}_{\bar{a}}^{3*}(t)}{\hat{V}_a^3(T)}. \tag{6.18}$$

We notice that $\phi^x(t)\rho_{\phi x}(t) = \rho_x(t)$ for all x, so that one gets the same benefits for retiring directly as for converting to free policy and immediately retiring. However, the free policy scaling and the retirement scaling after conversion to free policy are not uniquely given from the formulas above, since we have that $\hat{V}_a^{x*}$ reserves were given from $\rho_{\phi x}$. The flexibility we have reflects how much disability benefits are scaled upon conversion to free policy compared to how much retirement benefits are scaled. We choose to fix $\phi^x(T) = 1$, in order for the scaling of the disability benefits to vanish when the time of conversion to free policy tends to T. With $\phi^x(T) = 1$ it follows from (6.17) that $\hat{V}_{\bar{a}}^{x*}(T) = \hat{V}_a^x(T)$, and this gives a computable boundary condition for (6.16). With (6.18) we reduce (6.16) to

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{V}_{\bar{a}}^{x*}(t) = \hat{r}\hat{V}_{\bar{a}}^{x*}(t) - \hat{\mu}_{\bar{a}\bar{d}}(t)(b_{ad}^x(t) - \hat{V}_{\bar{a}}^{x*}(t)) - \hat{\mu}_{\bar{a}\bar{i}}(t)(\hat{V}_{\bar{i}}^{x*}(t) - \hat{V}_{\bar{a}}^{x*}(t)), \tag{6.19}$$

and thus $\hat{V}_{\bar{a}}^{x*}$ is uniquely determined. With $\hat{V}_{\bar{a}}^{x*}$ given it follows from (6.17) and (6.18) that also ϕ^3 and $\rho_{\phi 3}$ are given in every timepoint.

In the present example the partial reserve for the pension sum only contains retirement benefits, and thereby it is actually not important how we fix $\phi^3(T)$. From (6.19) it follows that if $\phi^3(T) = k$, $\hat{V}_{\bar{a}}^{3*}$ is decreased with a factor k. Thus from (6.17) and (6.18) ϕ^3 is increased with a factor k and $\rho_{\phi 3}$ is decreased with a factor k. Every benefit for the partial reserve for the pension sum is either scaled with both ϕ^3 and $\rho_{\phi 3}$ or with none of them, and thus $\phi^3(T)$ has no influence.

Combined we have that ϕ^x, ρ_x and $\rho_{\phi x}$ are calculated from $\hat{V}_p^T$, $\hat{V}_a^x$ and $\hat{V}_{\bar{a}}^{x*}$. $\hat{V}_p^T$ is as in Section 6.2. $\hat{V}_a^x$ before time T corresponds to the reserve in a model with deterministic retirement, and after time T, $\hat{V}_a^x$ is given from (6.15). $\hat{V}_{\bar{a}}^{x*}$ is calculated from (6.19) with the boundary condition $\hat{V}_{\bar{a}}^{x*}(T) = \hat{V}_a^x(T)$. Notice that since it is assumed that the policyholder cannot become disabled after time T, we can determine the values of $\hat{V}_a^x$ and $\hat{V}_{\bar{a}}^{x*}$ independently of $\hat{V}_i^x$ and $\hat{V}_{\bar{i}}^{x*}$ on $[T, T_{max}]$. This is an advantage numerically, since our boundary conditions for the Thiele differential equations for $\hat{V}_a^x$ and $\hat{V}_{\bar{a}}^{x*}$ are given in T, whereas the boundary conditions for the Thiele differential equations for $\hat{V}_i^x$ and $\hat{V}_{\bar{i}}^{x*}$ are given in T_{max}. Thus, for $t < T$, $\hat{V}_a^x$ and $\hat{V}_{\bar{a}}^{x*}$ are calculated from (6.15) and (6.19) simultaneous with $\hat{V}_i^x$ and $\hat{V}_{\bar{i}}^{x*}$ being calculated from a similar differential equation.

6.3.3 Market Valuation

We have once again seen how the technical reserve in the active state is unaffected by modelling retirement as stochastic. However, under the market basis the risk terms are no longer zero, and the market values are thus affected by modelling retirement as stochastic.

As in Section 6.2.2 we assume a deterministic market interest rate, r, and we assume the distribution of Z resembles the one under the technical basis, with the only difference that the transition intensities $\hat{\mu}_{xy}$ are replaced by intensities μ_{xy} for $x, y \in \{a, i, p, d, \bar{a}, \bar{i}, \bar{d}, \bar{p}\}$ and $\hat{p}_1, \ldots, \hat{p}_n, \hat{p}_1^\phi, \ldots, \hat{p}_n^\phi$ are replaced by $p_1, \ldots, p_n, p_1^\phi, \ldots, p_n^\phi$, with $p_n = p_n^\phi = 1$. The transition probabilities are denoted p_{xy} instead of $\hat{p}_{xy}$ and they may be deduced from (6.5) and (6.14) by replacing $\hat{\mu}_{xy}$ with μ_{xy}. The two bases agree on when the transition intensities are zero, and they agree on the time of the discontinuities of the transition probabilities.

Let V_a^x and V_i^x denote the market reserves for respectively the active state and the disability state for the type x saving. Let V_p^T denote the market value of a life annuity with the reference benefits, and let $V_{\bar{a}}^{x*}$ and $V_{\bar{i}}^{x*}$ denote the market values from respectively the free policy active state and the free policy disability state of the payment steam B^{x*}. Let $B(s) = B^1(s) + B^3(s)$ and $B^*(s) = B^{1*}(s) + B^{3*}(s)$ then

$$V_a(t) = \mathbb{E}_{a,t}[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\mathrm{d}B(s)],$$
$$V_{\bar{a}}^*(t) = \mathbb{E}_{a,t}[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\mathrm{d}B^*(s)].$$

These are continuous and differentiable except in $t_1, \ldots, t_n$. By differentiating the following Thiele differential equation follows

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}V_a(t) &= \pi + r(t)V_a(t) - \mu_{ad}(t)\left(b_{ad}(t) - V_a(t)\right) - \mu_{ai}(t)\left(V_i(t) - V_a(t)\right) \\
&\quad - \mu_{ap}(t)\left(\rho_1(t)V_p^T(t) + \rho_3(t)b_{ap}^T - V_a(t)\right) \\
&\quad - \mu_{a\bar{a}}(t)\left(\phi^1(t)V_{\bar{a}}^{1*}(t) + \phi^3(t)V_{\bar{a}}^{3*}(t) - V_a(t)\right), \\
\tfrac{\mathrm{d}}{\mathrm{d}t}V_{\bar{a}}^{x*}(t) &= r(t)V_{\bar{a}}^{x*}(t) - \mu_{\bar{a}\bar{d}}(t)\left(b_{ad}^x(t) - V_{\bar{a}}^{x*}(t)\right) - \mu_{\bar{a}\bar{i}}(t)\left(V_{\bar{i}}^{x*}(t) - V_{\bar{a}}^{x*}(t)\right) \\
&\quad - \mu_{\bar{a}\bar{p}}(t)\left(1_{(x=1)}\rho_{\phi 1}(t)V_{\bar{p}}^T(t) + 1_{(x=3)}\rho_{\phi 3}(t)b_{ap}^T - V_{\bar{a}}^{x*}(t)\right),
\end{aligned}$$

with terminal conditions

$$\begin{aligned}
V_a(T_{max}-) &= \rho_1(T_{max})V_p^T(T_{max}) + \rho_3(T_{max})b_{ap}^T, \\
V_{\bar{a}}^{1*}(T_{max}-) &= \rho_{\phi 1}(T_{max})V_{\bar{p}}^T(T_{max}), \\
V_{\bar{a}}^{3*}(T_{max}-) &= \rho_{\phi 3}(T_{max})b_{ap}^T,
\end{aligned}$$

and in discontinuity points we get

$$\begin{aligned}V_a(t_h) - V_a(t_h-) &= -p_h(\rho_1(t_h)V_p^T(t_h) + \rho_3(t_h)b_{ap}^T - V_a(t_h)),\\ V_{\bar a}^{1*}(t_h) - V_{\bar a}^{1*}(t_h-) &= -p_h^{\phi}(\rho_{\phi 1}(t_h)V_p^T(t_h) - V_{\bar a}^{1*}(t_h)),\\ V_{\bar a}^{3*}(t_h) - V_{\bar a}^{3*}(t_h-) &= -p_h^{\phi}(\rho_{\phi 3}(t_h)b_{ap}^T - V_{\bar a}^{3*}(t_h)).\end{aligned}$$

As the stochastic retirement makes the time of the benefits stochastic, numerically we find the highest impact on the expected cash flows. In Appendix 6.B we derive the market reserve. We determine an expression that allows us to immediately deduce the expected cash flow, which is given by

$$\begin{aligned}\mathrm{d}A_a(t,s) = {} & p_{aa}(t,s-)\left(-\pi + \mu_{ad}(s)b_{ad}(s)\right)\mathrm{d}s + p_{ai}(t,s)b_i(s)\mathrm{d}s\\ & + p_{aa}(t,s-)\rho_3(s)b_{ap}^T\left(\mu_{ap}(s)\mathrm{d}s + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(s)\right)\\ & + p_{ap}^{\rho_1}(t,s)b_p^T\mathrm{d}s + p_{a\bar a}^{\phi^1}(t,s)b_{ad}(s)\mu_{\bar a\bar d}(s)\mathrm{d}s\\ & + p_{a\bar i}^{\phi^1}(t,s)b_i(s)\mathrm{d}s + p_{a\bar p}^{\phi^1\rho_{\phi 1}}(t,s)b_p^T\mathrm{d}s\\ & + p_{a\bar a}^{\phi^3}(t,s-)\rho_{\phi 3}(s)b_{ap}^T\left(\mu_{\bar a\bar p}(s)\mathrm{d}s + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(s)\right), \qquad (6.20)\end{aligned}$$

with

$$\begin{aligned}p_{ap}^{\rho_1}(t,s) &= \mathbb{E}_{a,t}[I_p(s)\rho_1(s-U_s)]\\ &= \int_t^s p_{aa}(t,\tau-)\rho_1(\tau)p_{pp}(\tau,s)\left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h\mathrm{d}\varepsilon_{t_h}(\tau)\right),\\ p_{a\bar i}^{\phi^1}(t,s) &= \mathbb{E}_{a,t}[I_{\bar i}(s)\phi^1(s-V_s)] = \int_t^s p_{aa}(t,\sigma)\mu_{a\bar a}(\sigma)\phi^1(\sigma)p_{\bar a\bar i}(\sigma,s)\mathrm{d}\sigma,\\ p_{a\bar p}^{\phi^1\rho_{\phi 1}}(t,s) &= \mathbb{E}_{a,t}[I_{\bar p}(s)\phi^1(s-V_s)\rho_{\phi 1}(s-U_s)]\\ &= \int_t^s\int_\sigma^s p_{aa}(t,\sigma)\phi^1(\sigma)\mu_{a\bar a}(\sigma)p_{\bar a\bar a}(\sigma,\tau-)\rho_{\phi 1}(\tau)p_{\bar p\bar p}(\tau,s)\\ &\quad \left(\mu_{\bar a\bar p}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h^{\phi}\mathrm{d}\varepsilon_{t_h}(\tau)\right)\mathrm{d}\sigma,\\ p_{a\bar a}^{\phi^x}(t,s) &= \mathbb{E}_{at}[I_{\bar a}(s)\phi^x(s-V_s)] = \int_t^s p_{aa}(t,\sigma)\mu_{a\bar a}(\sigma)\phi^x(\sigma)p_{\bar a\bar a}(\sigma,s)\mathrm{d}\sigma.\end{aligned}$$

The expected cash flow seems time consuming to calculate as we need transition intensities for all combinations of $s, u \in [t,\infty)$ with $s < u$. The modified probabilities above ease the calculations as they may be calculated from differential equations. This is done in a similar way for a similar model with deterministic retirement in [2]. For $s \notin \{t_1, \ldots, t_n\}$ the modified probabilities are smooth with:

$$\begin{aligned}
\frac{\partial}{\partial s} p^{\rho_1}_{ap}(t,s) &= p_{aa}(t,s)\mu_{ap}(s)\rho_1(s) - \mu_{pd}(s)p^{\rho_1}_{ap}(t,s),\\
\frac{\partial}{\partial s} p^{\phi^x}_{a\bar{i}}(t,s) &= -p^{\phi^x}_{a\bar{i}}(t,s)(\mu_{\bar{i}\bar{a}}(s)+\mu_{\bar{i}\bar{d}}(s)) + p^{\phi^x}_{a\bar{a}}(t,s)\mu_{\bar{a}\bar{i}}(s),\\
\frac{\partial}{\partial s} p^{\phi^1\rho_{\phi 1}}_{a\bar{p}}(t,s) &= p^{\phi^1}_{a\bar{a}}(t,s)\mu_{\bar{a}\bar{p}}(s)\rho_{\phi 1}(s) - \mu_{\bar{p}\bar{d}}(s)p^{\phi^1\rho_{\phi 1}}_{a\bar{p}}(t,s),\\
\frac{\partial}{\partial s} p^{\phi^x}_{a\bar{a}}(t,s) &= p_{aa}(t,s)\mu_{a\bar{a}}(s)\phi^x(s) - \mu_{\bar{a}\cdot}(s)p^{\phi^x}_{a\bar{a}}(t,s) + \mu_{\bar{i}\bar{a}}(s)p^{\phi^x}_{a\bar{i}}(t,s).
\end{aligned}$$

In the discontinuity points we have

$$\begin{aligned}
p^{\rho_1}_{ap}(t,t_h) - p^{\rho_1}_{ap}(t,t_h-) &= p_{aa}(t,t_h-)\rho_1(t_h)p_h,\\
p^{\phi^x}_{a\bar{i}}(t,t_h) - p^{\phi^x}_{a\bar{i}}(t,t_h-) &= 0,\\
p^{\phi^1\rho_{\phi 1}}_{a\bar{p}}(t,t_h) - p^{\phi^1\rho_{\phi 1}}_{a\bar{p}}(t,t_h-) &= p^{\phi^x}_{a\bar{a}}(t,t_h-)\rho_{\phi 1}(t_h)p^{\phi}_h,\\
p^{\phi^x}_{a\bar{a}}(t,t_h) - p^{\phi^x}_{a\bar{a}}(t,t_h-) &= p^{\phi^x}_{a\bar{a}}(t,t_h-)p^{\phi}_h.
\end{aligned}$$

The modified probabilities are no more complicated to calculate than the traditional transition probabilities. Thus, with these the expected cash flows are easily calculated.

6.4 Benefit Conversion

In Section 6.2 and Section 6.3 we have modelled the time of retirement as stochastic. In reality it is also very common that upon retirement the benefits or parts of these are converted to another structure than originally stated in the contract. This could mean that a saving originally intended for a pension sum is used for buying a life annuity. We assume that the policyholder's choice regarding this is independent of everything else in the model, except the time of retirement. E.g. the policyholder may be more inclined to convert her savings to a life annuity if she retires early rather than if she retires very late.

Recall that $x \in \{1, 3\}$ represents respectively the annuity and the pension sum. Let Y^x_t denote the proportion of the partial reserve originally intended for benefit structure x which is used for a life annuity upon retirement if this happens at time t before conversion to free policy. We let $Y^{\phi x}_t$ denote the corresponding proportion if the policyholder has converted to free policy first. Assume the remaining partial reserve is used for a pension sum. $Y^x = (Y^x_t)_{t\geq 0}$ and $Y^{\phi x} = (Y^{\phi x}_t)_{t\geq 0}$ are stochastic processes taking values in

$[0,1]$, and $(Y^x, Y^{\phi x})$ is independent of everything else in our model. Conversion of benefits is often regulated. Many regulation types governing which conversions are allowed, can be easily implemented in our model through choice of the distributions of $(Y^x, Y^{\phi x})$.

6.4.1 Scaling the Benefits

We consider the complex model of Section 6.3. The size of the benefits after conversion is not obvious. Just like for the retirement factor, setting the benefits after a conversion depends on the guarantees the policyholder has. At one extreme is the case where the policyholder is not guaranteed that she is allowed to convert. Even if the policyholder is allowed to convert, then she might not have any guarantees regarding the basis used for pricing. In this case conversion may be modelled as surrender or as full conversion to a pension sum. We study a case where the policyholder is guaranteed that upon retirement she may use her retrospective saving to buy a combination of a pension sum and a life annuity valued under the technical basis. In this case the guarantee from the technical basis stretches very far and one could argue that it may not be realistic. If for example a policyholder has a pension sum saving with a very high technical interest rate, she might not be allowed to convert the saving to a life annuity with the same high guaranteed interest rate.

Now, for the partial reserves for each of the benefit structure types $x \in \{1,3\}$, we define reference retirement benefits b_p^{Tx} and b_{ap}^{Tx}. These corresponds to the benefits if the policyholder retires at time T and chooses to have everything paid out as respectively a life annuity or a pension sum. The reference benefits are scaled depending on the time of retirement and conversion to free policy. When all the saving meant for benefit type x is used for benefit type w, then the reference benefits are scaled in the following way: If the policyholder retires from the active state at time t, the scaling $\rho_w^x(t)$ is used. And if the policyholder retires at time t after conversion to free policy at time u the scaling $\phi^x(u)\rho_w^x(t)$ is used. Here ϕ^x is the free policy scaling from the model of section 6.3. For every other benefits the scaling of Section 6.3 is used. Thereby we get the payment stream for the partial reserve intended for benefit type x to be:

$$\begin{aligned}
\mathrm{d}B^x(t) &= -\pi^x I_a(t)\mathrm{d}t + b_i^x(t)I_i(t)\mathrm{d}t + b_{ad}^x(t)\mathrm{d}N_{ad}(t) \\
&\quad + (1 - Y_t^x)b_{ap}^{Tx}\rho_3^x(t)\mathrm{d}N_{ap}(t) + Y_{t-U_t}^x b_p^{Tx}\rho_1^x(t - U_t)I_p(t)\mathrm{d}t \\
&\quad + b_i^x(t)\phi^x(t - V_t)I_{\bar{i}}(t)\mathrm{d}t + (1 - Y_t^{\phi x})b_{ap}^{Tx}\rho_{\phi 3}^x(t)\phi^x(t - V_t)\mathrm{d}N_{\bar{a}\bar{p}}(t) \\
&\quad + b_{ad}^x(t)\phi^x(t - V_t)\mathrm{d}N_{\bar{a}\bar{d}}(t) + Y_{t-U_t}^{\phi x} b_p^{Tx}\rho_{\phi 1}^x(t - U_t)\phi^x(t - V_t)I_{\bar{p}}(t)\mathrm{d}t.
\end{aligned}$$

Using the notation from the previous section, but with $\hat{V}_p^{Tx}$ being the prospective technical value of a life annuity with payments b_p^{Tx} the scalings become

$$\rho_1^x(t) = \frac{\hat{V}_a^x(t)}{\hat{V}_p^{T_x}(t)} \qquad \text{and} \qquad \rho_3^x(t) = \frac{\hat{V}_a^x(t)}{\hat{V}_a^x(T)},$$
$$\rho_{\phi 1}^x(t) = \frac{\hat{V}_{\bar{a}}^{x*}(t)}{\hat{V}_p^{T_x}(t)} \qquad \text{and} \qquad \rho_{\phi 3}^x(t) = \frac{\hat{V}_{\bar{a}}^{x*}(t)}{\hat{V}_a^x(T)},$$

Let $y_t^x = \mathbb{E}[Y_t^x]$ and $y_t^{\phi x} = \mathbb{E}[Y_t^{\phi x}]$. In Appendix 6.C we derive the market reserve. We determine an expression that allows us to immediately deduce the expected cash flow, which is given by

$$\begin{aligned}
\mathrm{d}A_a^x(t,s) &= p_{aa}(t,s-)\left(-\pi^x + \mu_{ad}(s)b_{ad}^x(s)\right)\mathrm{d}s + p_{ai}(t,s)b_i^x(s)\mathrm{d}s \\
&\quad + p_{aa}(t,s-)(1-y_s^x)b_{ap}^{Tx}\rho_3^x(s)\left(\mu_{ap}(s)\mathrm{d}s + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(s)\right) \\
&\quad + p_{ap}^{y\rho_1^x}(t,s)b_p^{Tx}\mathrm{d}s + p_{a\bar{a}}^{\phi^x}(t,s-)b_{ad}^x(s)\mu_{\bar{a}\bar{d}}(s)\mathrm{d}s \\
&\quad + p_{a\bar{i}}^{\phi^x}(t,s)b_i^x(s)\mathrm{d}s + p_{a\bar{p}}^{y\phi^x\rho_1^x}(t,s)b_p^{Tx}\mathrm{d}s \\
&\quad + p_{a\bar{a}}^{\phi^x}(t,s-)(1-y_s^{\phi x})b_{ap}^{Tx}\rho_{\phi 3}^x(s)\left(\mu_{\bar{a}\bar{p}}(s)\mathrm{d}s + \sum_{h=1}^n p_h^{\phi}\mathrm{d}\varepsilon_{t_h}(s)\right),
\end{aligned} \tag{6.21}$$

where

$$\begin{aligned}
p_{ap}^{y\rho_1^x}(t,s) &= \int_t^s \rho_1^x(\tau)y_\tau^x p_{pp}(\tau,s)p_{aa}(t,\tau-)(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(\tau)), \\
p_{a\bar{p}}^{y\phi^x\rho_1^x}(t,s) &= \int_t^s\int_\sigma^s p_{aa}(t,\sigma)\mu_{a\bar{a}}(\sigma)\phi^x(\sigma)p_{\bar{a}\bar{a}}(\sigma,\tau-)y_\tau^{\phi x}\rho_{\phi 1}^x(\tau)p_{\bar{p}\bar{p}}(\tau,s) \\
&\qquad \left(\mu_{\bar{a}\bar{p}}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h^{\phi}\mathrm{d}\varepsilon_{t_h}(\tau)\right)\mathrm{d}\sigma.
\end{aligned}$$

This is the same we would get if we had considered the model of Figure 6.3. This is a Markov model with four states a (active), i (disabled), l (life annuity), s (pension sum), d (dead) and corresponding free policy states $\bar{a}$, $\bar{i}$, $\bar{l}$, $\bar{s}$ and $\bar{d}$. Transitions between any of the states $\{a, i, d, \bar{a}, \bar{i}, \bar{d}\}$ has the intensities and probabilities as in the previous model. Transitions to and from the retirement states has intensities

$$\begin{aligned}
&\mu_{al}(t) = y_t^x\mu_{ap}(t), &&\mu_{as}(t) = (1-y_t^x)\mu_{ap}(t), &&\mu_{ld}(t) = \mu_{sd}(t) = \mu_{pd}(t), \\
&\mu_{\bar{a}\bar{l}}(t) = y_t^{\phi x}\mu_{\bar{a}\bar{p}}(t), &&\mu_{\bar{a}\bar{s}}(t) = (1-y_t^{\phi x})\mu_{\bar{a}\bar{p}}(t), &&\mu_{\bar{l}\bar{d}}(t) = \mu_{\bar{s}\bar{d}}(t) = \mu_{\bar{p}\bar{d}}(t),
\end{aligned}$$

and all other intensities to and from the retirement states are zero. Transition probabilities to the retirement states are continuous, except for the time points $t_1, \dots, t_n$. In those discontinuities we have

$$
\begin{aligned}
p_{al}(t_h-, t_h) &= p_h y^x_{t_h}, & p_{\bar a \bar l}(t_h-, t_h,) &= p^\phi_h y^{\phi x}_{t_h}, \\
p_{as}(t_h-, t_h) &= p_h(1 - y^x_{t_h}), & p_{\bar a \bar s}(t_h-, t_h) &= p^\phi_h(1 - y^{\phi x}_{t_h}).
\end{aligned}
$$

From this new model we easily deduce the following Thiele differential equa-

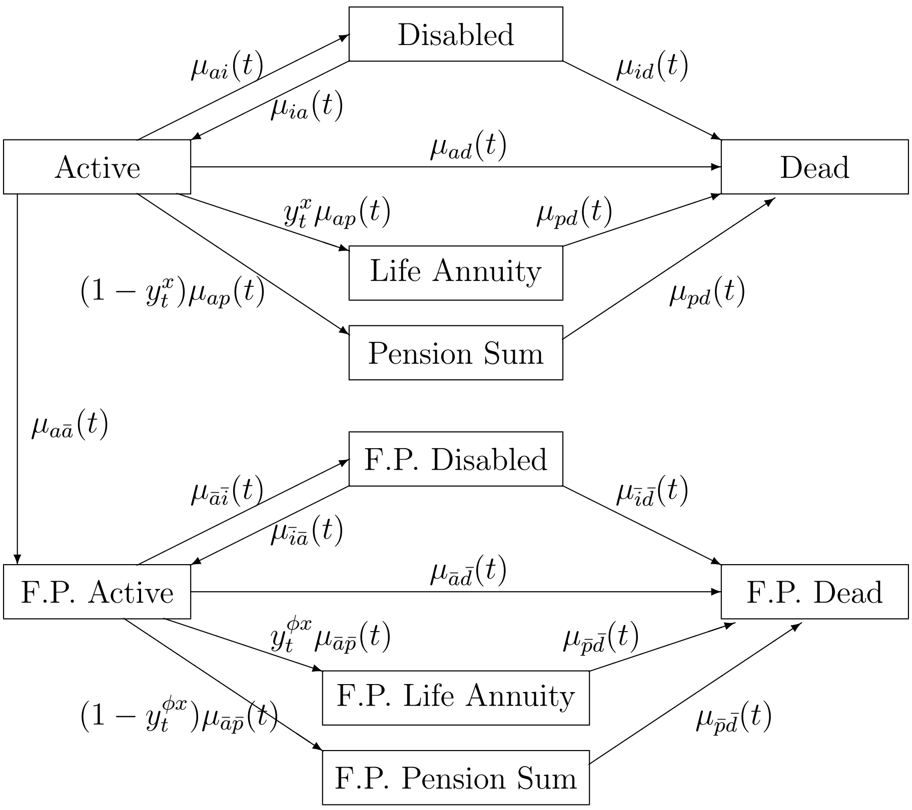

Figure 6.3: Model with disability, free policy and stochastic retirement including states for benefit conversion.

tion for differentiability points

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} V^x_a(t) &= \pi^x + (r(t) + \mu_{ad}(t))\, V^x_a(t) - \mu_{ai}(t)\,(V^x_i(t) - V^x_a(t)) \\
&\quad - \mu_{ap}(t)\left(y^x_t \rho^x_1(t) V^{Tx}_p(t) + (1 - y^x_t)\rho^x_3(t) b^x_{ap} - V^x_a(t)\right) \\
&\quad - \mu_{a\bar a}(t)\,(\phi^x(t) V^{x*}_{\bar a}(t) - V^x_a(t))\,, \\
\frac{\mathrm{d}}{\mathrm{d}t} V^{x*}_{\bar a}(t) &= (r(t) + \mu_{\bar a \bar d}(t)) V^{x*}_{\bar a}(t) - \mu_{\bar a \bar i}(t)\,(V^{x*}_{\bar i}(t) - V^{x*}_{\bar a}(t)) \\
&\quad - \mu_{\bar a \bar p}(t)\left(y^{\phi x}_t \rho^x_{\phi 1}(t) V^{Tx}_p(t) + (1 - y^{\phi x}_t)\rho^x_{\phi 3}(t) b^x_{ap} - V^{x*}_{\bar a}(t)\right),
\end{aligned}
$$

with terminal conditions

$$V_a^x(T_{max}-) = y^x_{T_{max}}\rho_1^x(T_{max})V_p^{Tx}(T_{max}) + (1-y^x_{T_{max}})\rho_3^x(T_{max})b^3_{ap},$$
$$V_{\bar{a}}^{x*}(T_{max}-) = y^{\phi x}_{T_{max}}\rho_{\phi 1}^x(T_{max})V_p^{Tx}(T_{max}) + (1-y^{\phi x}_{T_{max}})\rho_{\phi 3}^x(T_{max})b^x_{ap}.$$

In discontinuity points we get

$$V_a^x(t_h) - V_a^x(t_h-) = -p_h\left(y^x_{t_h}\rho_1^x(t_h)V_p^{Tx}(t_h) + (1-y^x_{t_h})\rho_3^x(t_h)b^x_{ap} - V_a^x(t_h)\right),$$
$$V_{\bar{a}}^{x*}(t_h) - V_{\bar{a}}^{x*}(t_h-) = -p_h^{\phi}\left(y^{\phi x}_{t_h}\rho_{\phi 1}^x(t_h)V_{\bar{p}}^{Tx}(t_h) + (1-y^{\phi x}_{t_h})\rho_{\phi 1}^x(t_h)b^x_{ap} - V_{\bar{a}}^{x*}(t_h)\right).$$

We may also deduce the differential equation by differentiating the integrated discounted expected cash flow. This approach may be used to verify the connection to the model of Figure 6.3

We may obtain a similar Thiele equation under the technical basis and we see from this that the scaling functions are such that the risk sums for retiring are zero under the technical basis, and so are the risk sums for conversion to free policy. Specifically the technical reserves in any other state than retired equals the technical reserves when the structure of the benefits is fixed. Thus the equivalence principle still holds.

It is tractable that for this more advanced model with benefit conversion reserves can be calculated from a model of the kind we are used to. However, we should be aware that the model of Figure 6.3 is constructed to give us the correct expected cash flow and thereby also the correct market reserve. This does not guarantee that the model is appropriate risk management calculations based on other distributional properties. Even if policyholders in the true model are likely to convert a proportion of their savings, the model of Figure 6.3 assumes extreme choice. In the model of Figure 6.3 policyholders always decide to have all retirement benefits paid out with the same benefit structure.

6.5 Numerical Results and Discussion

We consider here the numerical consequences of taking stochastic retirement into account. We begin in Section 6.5.1 by looking at a simple contract in the simple model, and examine the numerical consequences of expanding this to include a stochastic retirement time. This corresponds to what is done in Section 6.2.

In Section 6.5.2 we look at a more realistic contract in the complex model. In this model we look at the numerical consequences of adding first the possibility to convert to free policy, then a stochastic retirement time, and lastly stochastic benefit conversion at the time of retirement. This corresponds to what is done in Section 6.3 and Section 6.4.

6.5.1 Stochastic Retirement in the Simple Model

We consider the model and the contract from Section 6.2. We assume that the mortality from the active state is the same as from the retired state, and furthermore we assume that the mortality is the same for the technical and the market basis. For this mortality we choose the standard intensities for a female occurring in the Danish G82 risk table. This corresponds to the following intensity expression

$$\hat{\mu}_{ad}(age) = \mu_{ad}(age) = 0.0005 + 10^{5.728-10+0.038*age}$$

As mentioned in Section 6.3, modelling the time of retirement as stochastic does not have any effect on the technical reserves. On the market basis we consider three different models for the retirement transition. We name these *low retirement intensity*, *deterministic retirement*, and *high retirement intensity*, and they are given by

Retirement	$(t_1, \ldots, t_n)$	$(p_1, \ldots, p_n)$=$(p_1^\phi, \ldots, p_n^\phi)$	$\mu_{ap}(age)$
Low intensity	(62,67,72)	(0.1, 0.2, 1)	$e^{0.05*age-8}$
Deterministic	67	1	0
High intensity	(62,67,72)	(0.1, 0.2, 1)	$e^{0.1*age-8}$

This is in some regards an overly simplistic example since one would also expect the positive probability mass to be smaller in the example with smaller retirement intensity.

Figure 6.4 contains the transition probabilities based on the low and high retirement intensity. For the deterministic retirement, probability mass in the active state simply moves to the pension state at the agreed time of retirement. For the low retirement intensity, the three time points with positive probability mass of retirement are clearly seen, whereas these are not as clear in the example where policyholders have a high intensity of retirement between these time points.

We now consider two contracts; one with a high interest rate guarantee of 5%, and another with a low interest rate guarantee of 1%. For both contracts the future market yield is assumed to be constant at 3.5%. All interest rates are continuously compounded. The contracts are assumed to have a fixed premium and the same two benefits with amounts found by the equivalence principle; a life annuity, and a pension sum. 10% of the premium is used to fund the pension sum. Furthermore the contracts are calculated at the initiation date, and the policyholder is assumed to be 30 years old at this time. The premium and benefits for retirement at age 67 for the two contracts are found to be

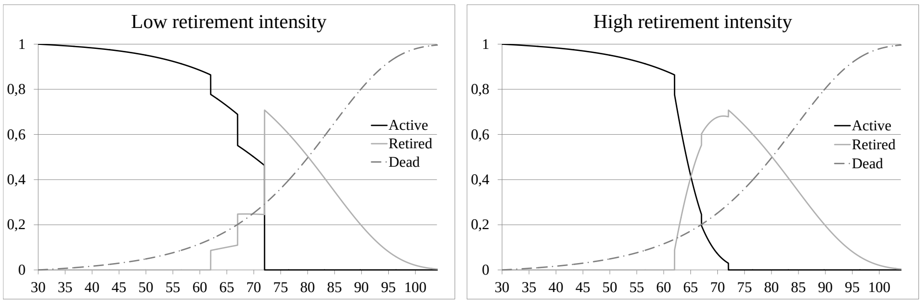

Figure 6.4: Transition probabilities based on the low and high retirement transition intensity.

	High interest ($\hat{r} = 5\%$)	Low interest ($\hat{r} = 1\%$)
Premium	10,000 €	10,000 €
ref. life annuity (b_a^{67})	108,177 €	32,121 €
ref. pension sum (b_{ap}^{67})	125,590 €	52,904 €

The dependence of the time of retirement on the benefits is described by the retirement factor. Figure 6.5 displays the scaling factors as a function of the time of retirement in each of the two cases low / high interest rate. These graphs are easily accessible tools for guiding policyholders about the impact of their choice regarding when to retire. We have calculated the market

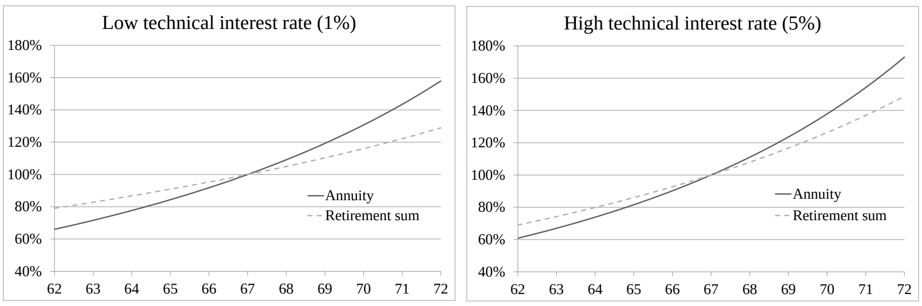

Figure 6.5: The retirement scaling factors in the simple model for respectively a low (1%), and a high (5%) guaranteed interest rate.

reserve at time zero for the two contracts using the three different models for the transition to retirement. The results were found to be

Retirement	High interest ($\hat{r} = 5\%$)	Low interest ($\hat{r} = 1\%$)
Low intensity	124,178 €	-109,425 €
Deterministic	113,205 €	-103,681 €
High intensity	107,789 €	-100,288 €

This shows that for a contract with a high interest rate guarantee compared to the market expectations the reserve is higher for the low retirement intensity than for deterministic, which is then again higher than the reserve for the high retirement intensity. This is what we would expect since a late time of retirement means that more premium is paid. The insurance company has to bear interest for this premium with an interest rate that is higher than market interest rate. The opposite is the case for the contract with an interest rate guarantee lower than market expectations.

In Figure 6.6 the cash flows for each of the three models of the retirement transition is shown for the contract with the low interest rate guarantee. These cash flows contain no discounting from the market interest rate. The cash flows for the contract with the high interest rate guarantee looks similar to these. In these figures the pension sum can easily be seen in the cash flow

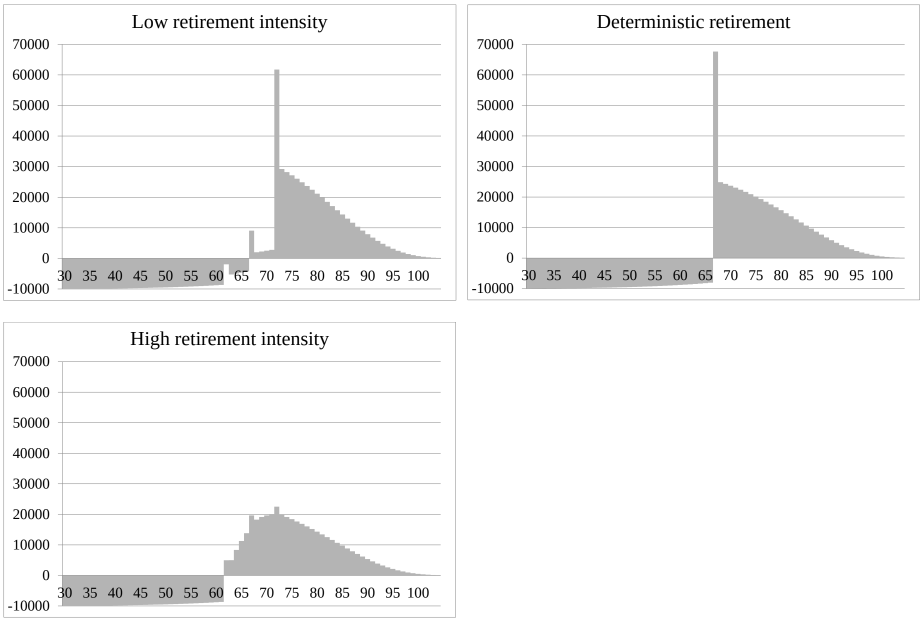

Figure 6.6: Cash flows for the contract low interest rate (1%) using the different retirement intensity.

for the low retirement intensity in each of the three time points with positive probability for retirement. The pension sum is also seen at the agreed upon time of retirement for the deterministic retirement. It is also seen that as expected a higher retirement intensity push payments forward in time, and leaves smaller expected payments after the latest retirement time.

6.5.2 Stochastic Retirement in the Complex Model

We now consider the models and the contract from Section 6.3 and 6.4. The same death intensity as in Section 6.5.1 is used for transitions to the dead states. Furthermore the transitions to retirement states are also as described in 6.5.1 both before and after conversion to free policy. The transition from the active state to the disabled state has the same intensity for the technical and market basis. We choose the standard intensity for a female occurring in the Danish G82 risk table. This corresponds to the following expression

$$\hat{\mu}_{ai}(age) = \mu_{ai}(age) = 0.0006 + 10^{4.71609-10+0.06*age}.$$

The transition from the disabled state to the active state is assumed to have a zero intensity for the technical basis, that is $\hat{\mu}_{ia}(age) = 0$, and the following intensity for the market basis $\mu_{ia}(age) = \exp\{-0.06 * age\}$. For the market basis we assume $\mu_{a\bar{a}}(age) = \exp\{-0.11 * age\}$.

As in Section 6.4, we define $y_t^3 = y_t^{\phi 3}$ as the expected proportion of the partial reserve for the retirement sum used for a life annuity upon retirement. We set this proportion to zero for the technical basis, and we choose the following expression for the market basis $y_t^3 = y_t^{\phi 3} = 0.25 + 0.5 \cdot (t-62)/(72-62)$. Furthermore we assume that there is no other benefit conversions than the one from the pension sum to a life annuity, and thus $y_t^1 = y_t^{\phi 1} = 1$.

As in Section 6.5.1, we consider two contracts. These are defined from the same guaranteed interest rates, and the market interest rate is also assumed to be the same as in Section 6.5.1. The contracts are assumed to have a fixed premium with a disability premium waiver, a fixed death sum, a fixed disability annuity and the same two benefits with amounts found by the equivalence principle; a life annuity, and a retirement sum. Again 10% of the premium is used to fund the pension sum. Furthermore the contracts are calculated at the initiation date, and the policyholder is assumed to be 30 years old at that time. The premium and benefits for the two contracts are found to be

	High interest ($\hat{r} = 5\%$)	Low interest ($\hat{r} = 1\%$)
Premium	10,000 €	10,000 €
ref. life annuity (b_p^{67})	84,827 €	21,224 €
ref. disab. annuity (b_i^{67})	30,000 €	30,000 €
ref. death sum (b_{ad}^{67})	100,000 €	100,000 €
ref. pension sum (b_{ap}^{67})	120,584 €	49,488 €

We have calculated the market reserve at time zero for the two contracts with various options included. First, we have calculated the reserves using the deterministic retirement. Next, we have added the free policy option. Then, we have included a stochastic retirement time using the low retirement intensity from Section 6.5.1. Lastly, we have added the possibility of benefit conversion at the time of retirement for the retirement sum. The resulting market reserves are found to be

	High interest ($\hat{r} = 5\%$)	Low interest ($\hat{r} = 1\%$)
Deterministic Retirement	88,121 €	-95,559 €
+ Free Policy Option	73,523 €	-78,814 €
+ Stochastic Retirement	78,462 €	-81,027 €
+ Benefit Conversion	79,720 €	-81,752 €

As expected, we see that all market reserves for the contract with the high interest rate guarantee is positive whereas all reserves for the contract with low interest rate guarantee are negative.

Including the free policy option in the calculation adds a possibility to stop premium payments and thus decrease the size of the future payments. This decreases the reserve for the contract with high interest rate guarantee, and increases the reserve for the contract with low interest rate guarantee.

As in Section 6.5.1 we see that a low stochastic retirement intensity delays the benefit payments and increase the size of the retirement benefits. This increases the reserve for the contract with high interest rate guarantee, and decreases the reserve for the contract with low interest rate guarantee.

Including benefit conversion from retirement sum to life annuity again delays the benefits and has the same effect as adding low stochastic retirement intensity. The opposite would be the case if the conversion where from life annuity to retirement sum.

We see that the market value of the contract with the high interest rate guarantee in our current economic environment is too small if one only takes into account the free policy option and not the stochastic retirement and the benefit conversion. In our setup, stochastic retirement time with a low retirement intensity increases the market value by approximately 7% while benefit conversion on just 10% of the premium increases the market value by approximately 1.5%. This means that insurance companies and pension funds not reserving for these types of options might not be setting aside enough to cover future payments. Furthermore, as illustrated in Figure 6.6, stochastic retirement has a high influence on the expected cash flow.

Appendix

6.A Market Reserve in the Simple Model

We derive the market reserve for the simple contract in Section 6.2 in a form that allows us to immediately deduce the expected cash flow of (6.11) and (6.12).

$$
\begin{aligned}
V_a(t) = \mathbb{E}_{a,t}\Bigg[& \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} \left(-I_a(s)\pi + I_p(s)\rho_1(s-U_s)b_p^T\right)\mathrm{d}s \\
& + \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} \rho_3(s) b_{ap}^T \mathrm{d}N_{ap}(s)\Bigg] \\
= & \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} \Big(p_{aa}(t,s-)\left(-\pi + \rho_3(s) b_{ap}^T \mu_{ap}(s)\right) \\
& + \mathbb{E}_{a,t}[I_p(s)\rho_1(s-U_s)] b_p^T \Big)\mathrm{d}s.
\end{aligned}
$$

We define

$$
\begin{aligned}
p_{ap}^{\rho_1}(t,s) &\equiv \mathbb{E}_{a,t}\left[I_p(s)\rho_1(s-U_s)\right] \\
&= \int_t^s \mathbb{E}_{a,t}\left[I_p(s)\rho_1(s-U_s) | s-U_s=\tau\right] \mathrm{d}P_{a,t}(s-U_s \leq \tau) \\
&= \int_t^s \mathbb{E}_{a,t}\left[I_p(s) | s-U_s=\tau\right] \rho_1(\tau) p_{aa}(t,\tau-) \\
&\quad \left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(\tau)\right) \\
&= \int_t^s p_{aa}(t,\tau-)\rho_1(\tau) p_{pp}(\tau,s) \left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(\tau)\right).
\end{aligned}
$$

Now (6.11) and (6.12) immediately follow.

6.B Market Reserve in the Complex Model

We derive the market reserve for the complex model of Section 6.3 in a form that allows us to immediately deduce the expected cash flow of (6.20).

$$\begin{aligned}
V_a(t) = \mathbb{E}_{a,t}\bigg[&\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\Big(- I_a(s)\pi + I_i(s)b_i(s) + I_p(s)b_p^T\rho_1(s-U_s) \\
&+ I_{\bar{i}}(s)b_i(s)\phi^1(s-V_s) + I_{\bar{p}}(s)b_p^T\phi^1(s-V_s)\rho_{\phi 1}(s-U_s)\Big)\mathrm{d}s \\
&+ \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} b_{ad}(s)\mathrm{d}N_{ad}(s) + \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} b_{ap}^T\rho_3(s)\mathrm{d}N_{ap}(s) \\
&+ \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} b_{ad}(s)\phi^1(s-V_s)\mathrm{d}N_{\bar{a}\bar{d}}(s) \\
&+ \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} b_{ap}^T\rho_{\phi 3}(s)\phi^3(s-V_s)\mathrm{d}N_{\bar{a}\bar{p}}(s)\bigg] \\
= \int_t^\infty & e^{-\int_t^s r(x)\mathrm{d}x}\Big(- p_{aa}(t,s-)\pi + p_{ai}(t,s)b_i(s) \\
&+ \mathbb{E}_{a,t}\left[I_p(s)\rho_1(s-U_s)\right] b_p^T + \mathbb{E}_{a,t}\left[I_{\bar{i}}(s)\phi^1(s-V_s)\right] b_i(s) \\
&+ \mathbb{E}_{a,t}\left[I_{\bar{p}}(s)\phi^1(s-V(s))\rho_{\phi 1}(s-U_s)\right] b_p^T + p_{aa}(t,s-)\mu_{ad}(s)b_{ad}(s)\Big)\mathrm{d}s \\
&+ \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\rho_3(s)b_{ap}^T p_{aa}(t,s-)\left(\mu_{ap}(s)\mathrm{d}s + \sum_{h=1}^n p_h \mathrm{d}\varepsilon_{t_h}(s)\right) \\
&+ \mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} b_{ad}(s)\phi^1(s-V_s)\mathrm{d}N_{\bar{a}\bar{d}}(s)\right] \\
&+ b_{ap}^T\mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\rho_{\phi 3}(s)\phi^3(s-V_s)\mathrm{d}N_{\bar{a}\bar{p}}(s)\right].
\end{aligned}$$

We evaluate the unresolved expectations one at a time

$$\begin{aligned}
p_{ap}^{\rho_1}(t,s) &= \mathbb{E}_{a,t}\left[I_p(s)\rho_1(s-U_s)\right] \\
&= \int_t^s \mathbb{E}_{a,t}\left[I_p(s)\rho_1(s-U_s) | s-U_s=\tau\right] \mathrm{d}P_{a,t}(s-U_s\le\tau) \\
&= \int_t^s \rho_1(\tau)\mathbb{E}_{p,\tau}\left[I_p(s)\right] p_{aa}(t,\tau-)\left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h\mathrm{d}\varepsilon_{t_h}(\tau)\right) \\
&= \int_t^s p_{aa}(t,\tau-)\rho_1(\tau)p_{pp}(\tau,s)\left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h\mathrm{d}\varepsilon_{t_h}(\tau)\right).
\end{aligned}$$

$$
\begin{aligned}
p_{a\bar{i}}^{\phi^1}(t,s) &= \mathbb{E}_{a,t}\left[I_{\bar{i}}(s)\phi^1(s-V_s)\right] \\
&= \int_t^s \mathbb{E}_{a,t}\left[I_{\bar{i}}(s)\phi^1(s-V_s)|s-V_s=\sigma\right]\mathrm{d}P_{a,t}(s-V_s\leq\sigma) \\
&= \int_t^s \phi^1(\sigma)\mathbb{E}_{\bar{a},\sigma}\left[I_{\bar{i}}(s)\right]p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\
&= \int_t^s p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\phi^1(\sigma)p_{\bar{a}\bar{i}}(\sigma,s)\mathrm{d}\sigma.
\end{aligned}
$$

$$
\begin{aligned}
p_{a\bar{p}}^{\phi^1\rho_{\phi 1}}(t,s) &= \mathbb{E}_{a,t}\left[I_{\bar{p}}(s)\phi^1(s-V_s)\rho_{\phi 1}(s-U_s)\right] \\
&= \int_t^s \mathbb{E}_{a,t}[I_{\bar{p}}(s)\phi^1(s-V_s)\rho_{\phi 1}(s-U_s)|s-V_s=\sigma]\mathrm{d}P_{a,t}(s-V_s\leq\sigma) \\
&= \int_t^s \phi^1(\sigma)\mathbb{E}_{\bar{a},\sigma}\left[I_{\bar{p}}(s)\rho_{\phi 1}(s-U_s)\right]p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\
&= \int_t^s\int_\sigma^s \mathbb{E}_{\bar{a},\sigma}\left[I_{\bar{p}}(s)\rho_{\phi 1}(s-U_s)|s-U_s=\tau\right]\mathrm{d}P_{\bar{a},\sigma}(s-U_s\leq\tau) \\
&\quad p_{aa}(t,\sigma-)\phi^1(\sigma)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\
&= \int_t^s\int_\sigma^s \rho_{\phi 1}(\tau)\mathbb{E}_{\bar{p},\tau}\left[I_{\bar{p}}(s)\right]p_{\bar{a}\bar{a}}(\sigma,\tau-)\left(\mu_{\bar{a}\bar{p}}(\tau)\mathrm{d}\tau+\sum_{h=1}^n p_h^\phi\mathrm{d}\varepsilon_{t_h}(\tau)\right) \\
&\quad p_{aa}(t,\sigma-)\phi^1(\sigma)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\
&= \int_t^s\int_\sigma^s p_{aa}(t,\sigma-)\phi^1(\sigma)\mu_{a\bar{a}}(\sigma)p_{\bar{a}\bar{a}}(\sigma,\tau)\rho_{\phi 1}(\tau)p_{\bar{p}\bar{p}}(\tau,s) \\
&\quad \left(\mu_{\bar{a}\bar{p}}(\tau)\mathrm{d}\tau+\sum_{h=1}^n p_h^\phi\mathrm{d}\varepsilon_{t_h}(\tau)\right)\mathrm{d}\sigma.
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}b_{ad}(s)\phi^1(s-V_s)\mathrm{d}N_{\bar{a}\bar{d}}(s)\right] \\
&= \mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}b_{ad}(s)\phi^1(s-V_s)I_{\bar{a}}(s-)\mu_{\bar{a}\bar{d}}(s)\mathrm{d}s\right] \\
&= \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}b_{ad}(s)\mathbb{E}_{a,t}\left[\phi^1(s-V_s)I_{\bar{a}}(s-)\right]\mu_{\bar{a}\bar{d}}(s)\mathrm{d}s.
\end{aligned}
$$

with

$$\begin{aligned} p_{a\bar{a}}^{\phi^x}(t,s-) &= \mathbb{E}_{a,t}\left[\phi^x(s-V_s)I_{\bar{a}}(s-)\right] \\ &= \int_t^s \mathbb{E}_{a,t}\left[\phi^x(s-V_s)I_{\bar{a}}(s-)|s-V_s=\sigma\right]\mathrm{d}P_{a,t}(s-V_s\le\sigma) \\ &= \int_t^s \phi^x(\sigma)\mathbb{E}_{\bar{a},\sigma}\left[I_{\bar{a}}(s-)\right]p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\ &= \int_t^s p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\phi^x(\sigma)p_{\bar{a}\bar{a}}(\sigma,s-)\mathrm{d}\sigma. \end{aligned}$$

$$\begin{aligned} &\mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\rho_{\phi 3}(s)\phi^3(s-V_s)\mathrm{d}N_{\bar{a}\bar{p}}(s)\right] \\ &= \mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\rho_{\phi 3}(s)\phi^3(s-V_s)I_{\bar{a}}(s-)\left(\mu_{\bar{a}\bar{p}}(s)\mathrm{d}s+\sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(s)\right)\right] \\ &= \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\rho_{\phi 3}(s)\mathbb{E}_{a,t}\left[\phi^3(s-V_s)I_{\bar{a}}(s-)\right]\left(\mu_{\bar{a}\bar{p}}(s)\mathrm{d}s+\sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(s)\right) \\ &= \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\rho_{\phi 3}(s)p_{a\bar{a}}^{\phi^3}(t,s-)\left(\mu_{\bar{a}\bar{p}}(s)\mathrm{d}s+\sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(s)\right), \end{aligned}$$

Now (6.20) immediately follows.

6.C Market Reserve with Benefit Conversion

We derive the market reserve for the model with benefit conversion studied in Section 6.4. We derive it in a form that allows us to immediately deduce the expected cash flow of (6.21).

$$
\begin{aligned}
V_a^x(t) = \mathbb{E}_{a,t}\Bigg[& \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}\Big(- I_a(s)\pi^x + b_i^x(s)I_i(s) \\
& + Y^x_{s-U_s} b_p^{Tx} I_p(s)\rho_1^x(s-U_s) + b_i^x(s)\phi^x(s-V_s)I_{\bar{i}}(s) \\
& + Y^{\phi x}_{s-U_s} b_p^{Tx}\phi^x(s-V_s)\rho^x_{\phi 1}(s-U_s)I_{\bar{p}}(s)\Big)\mathrm{d}s \\
& + \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} b^x_{ad}(s)\mathrm{d}N_{ad}(s) \\
& + \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y_s^x)b^{Tx}_{ap}\rho_3^x(s)\mathrm{d}N_{ap}(s) \\
& + \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x} b^x_{ad}(s)\phi^x(s-V_s)\mathrm{d}N_{\bar{a}\bar{d}}(s) \\
& + \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y_s^{\phi x})b^{Tx}_{ap}\rho^x_{\phi 3}(s)\phi^x(s-V_s)\mathrm{d}N_{\bar{a}\bar{p}}(s)\Bigg] \\
= \int_t^\infty & e^{-\int_t^s r(x)\mathrm{d}x}\Big(- p_{aa}(t,s-)\pi^x + p_{ai}(t,s)b_i^x(s) \\
& + b_p^{Tx}\mathbb{E}_{a,t}[Y^x_{s-U_s}I_p(s)\rho_1^x(s-U_s)] + p^{\phi^x}_{a\bar{i}}(t,s)b_i^x(s) \\
& + b_p^{Tx}\mathbb{E}_{a,t}[Y^x_{s-U_s}\phi^x(s-V_s)\rho^x_{\phi 1}(s-U_s)I_{\bar{p}}(s)] \\
& + p_{aa}(t,s-)\mu_{ad}(s)b^x_{ad}(s) + p^{\phi^x}_{a\bar{a}}(t,s-)\mu_{\bar{a}\bar{d}}(s)b^x_{ad}(s)\Big)\mathrm{d}s \\
& + b^{Tx}_{ap}\mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y_s^x)\rho_3^x(s)\mathrm{d}N_{ap}(s)\right] \\
& + b^{Tx}_{ap}\mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y_s^x)\rho^x_{\phi 3}(s)\phi^x(s-V_s)\mathrm{d}N_{\bar{a}\bar{p}}(s)\right].
\end{aligned}
$$

We evaluate each of the four unresolved expectations one at a time.

$$
\begin{aligned}
p_{ap}^{y\rho_1^x}(t,s) &= \mathbb{E}_{a,t}[Y^x_{s-U_s}I_p(s)\rho_1^x(s-U_s)] \\
&= \int_t^s \mathbb{E}_{a,t}[Y^x_{s-U_s}I_p(s)\rho_1^x(s-U_s)|s-U_s=\tau]\mathrm{d}P_{a,t}(s-U_s\le\tau) \\
&= \int_t^s \rho_1^x(\tau)\mathbb{E}[Y_\tau^x]\mathbb{E}_{p,\tau}[I_p(s)]p_{aa}(t,\tau-)\left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h\mathrm{d}\varepsilon_{t_h}(\tau)\right) \\
&= \int_t^s \rho_1^x(\tau)y_\tau p_{pp}(\tau,s)p_{aa}(t,\tau-)\left(\mu_{ap}(\tau)\mathrm{d}\tau + \sum_{h=1}^n p_h\mathrm{d}\varepsilon_{t_h}(\tau)\right).
\end{aligned}
$$

$$
\begin{aligned}
p_{ap}^{y\phi^x\rho^x_{\phi 1}}(t,s) &= \mathbb{E}_{a,t}[Y^{\phi x}_{s-U_s}\phi^x(s-V_s)\rho^x_{\phi 1}(s-U_s)I_{\bar{p}}(s)] \\
&= \int_t^s \mathbb{E}_{a,t}[Y^{\phi x}_{s-U_s}\phi^x(s-V_s)\rho^x_{\phi 1}(s-U_s)I_{\bar{p}}(s)|s-V_s=\sigma] \\
&\quad \mathrm{d}P_{a,t}(s-V_s\le\sigma) \\
&= \int_t^s \phi^x(\sigma)\mathbb{E}_{\bar{a},\sigma}[Y^{\phi x}_{s-U_s}\rho^x_{\phi 1}(s-U_s)I_{\bar{p}}(s)]p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\
&= \int_t^s \phi^x(\sigma)\int_\sigma^s \mathbb{E}_{\bar{a},\sigma}[Y^{\phi x}_{s-U_s}\rho^x_{\phi 1}(U_s)I_{\bar{p}}(s)|s-U_s=\tau] \\
&\quad \mathrm{d}P_{\bar{a},\sigma}(s-U_s\le\tau)p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\
&= \int_t^s\int_\sigma^s \mathbb{E}[Y^{\phi x}_\tau]\rho^x_{\phi 1}(\tau)\mathbb{E}_{\bar{p},\tau}[I_{\bar{p}}(s)]p_{\bar{a}\bar{a}}(\sigma,\tau-) \\
&\quad \left(\mu_{\bar{a}\bar{p}}(\tau)\mathrm{d}\tau+\sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(\tau)\right)\phi^x(\sigma)p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\mathrm{d}\sigma \\
&= \int_t^s\int_\sigma^s p_{aa}(t,\sigma-)\mu_{a\bar{a}}(\sigma)\phi^x(\sigma)p_{\bar{a}\bar{a}}(\sigma,\tau-)y^{\phi x}_\tau\rho^x_{\phi 1}(\tau)p_{\bar{p}\bar{p}}(\tau,s) \\
&\quad \left(\mu_{\bar{a}\bar{p}}(\tau)\mathrm{d}\tau+\sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(\tau)\right)\mathrm{d}\sigma.
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y^x_s)\rho^x_3(s)\mathrm{d}N_{ap}(s)\right] \\
&= \mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y^x_s)\rho^x_3(s)I_a(s-)\left(\mu_{ap}(s)\mathrm{d}s+\sum_{h=1}^n p_h\mathrm{d}\varepsilon_{t_h}(s)\right)\right] \\
&= \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-y^x_s)\rho^x_3(s)p_{aa}(t,s-)\left(\mu_{ap}(s)\mathrm{d}s+\sum_{h=1}^n p_h\mathrm{d}\varepsilon_{t_h}(s)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}_{a,t}\left[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y_s^{\phi x})\rho_{\phi 3}^x(s)\phi^x(s-V_s)\mathrm{d}N_{\bar{a}\bar{p}}(s)\right] \\
&= \mathbb{E}_{a,t}\Bigg[\int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-Y_s^{\phi x})\rho_{\phi 3}^x(s)\phi^x(s-V_s)I_{\bar{a}}(s-) \\
&\quad \left(\mu_{\bar{a}\bar{p}}(s)\mathrm{d}s + \sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(s)\right)\Bigg] \\
&= \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-y_t^{\phi x})\rho_{\phi 3}^x(s)\mathbb{E}_{a,t}[\phi^x(s-V_s)I_{\bar{a}}(s-)] \\
&\quad \left(\mu_{\bar{a}\bar{p}}(s)\mathrm{d}s + \sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(s)\right) \\
&= \int_t^\infty e^{-\int_t^s r(x)\mathrm{d}x}(1-y_t^{\phi x})\rho_{\phi 3}^x(s)p_{a\bar{a}}^{\phi^x}(t,s-)\left(\mu_{\bar{a}\bar{p}}(s)\mathrm{d}s + \sum_{h=1}^n p_h^\phi \mathrm{d}\varepsilon_{t_h}(s)\right).
\end{aligned}
$$

Now (6.21) immediately follows.

References for Chapter 6

[1] Bacinello, A. R. (2003). Fair Valuation of a Guaranteed Life Insurance Participating Contract Embedding a Surrender Option. *J Risk Insur* **70 (3)**, 461–487.

[2] Buchardt, K., Møller, T. and Schmidt, K. B. (2013). Cash flows and policyholder behaviour in the semi-Markov life insurance setup. *Scand Actuar J* **2014**, 1–29.

[3] Eling, M. and Kiesenbauer, D. (2014). What policy features determine life insurance lapse? An analysis of the German market. *J Risk Insur* **81(2)**, 241–269.

[4] Gad, K. S. T., Juhl, J. and Steffensen, M. (2014). Reserve-dependent surrender. *Available online at http://arxiv.org/abs/1412.1991.*

[5] Giovanni, D. D. (2010). Lapse rate modeling: A rational expectation approach. *Scand Actuar J* **2010(1)**, 56–67.

[6] Grosen, A. and Jørgensen, P. L. (2000). Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies. *Insur Math Econ* **26(1)**, 37–57.

[7] Henriksen, L. F. B. and Nielsen, J. and Steffensen, M. and Svensson, C. (2014). Markov chain modeling of policyholder behavior in life insurance and pension. *Eur Actuar J* **4(1)**, 1–29.

[8] Jones, B. L. (1994). Actuarial Calculations Using a Markov Model. *Trans Soc Actuar* **46**, 227–250.

[9] Koller, M. (2012). *Stochastic models in life insurance*. EAA Series, Springer.

[10] Møller, T. and Steffensen, M. (2007). *Market-Valuation Methods in Life and Pension Insurance*. Cambridge University Press.

[11] Norberg, R. (1991). Reserves in life and pension insurance. *Scand Actuar J* **1991(1)**, 1–24.

[12] Skoog, G. R. and Ciecka, J. E. (2010). Measuring years of inactivity, years in retirement, time to retirement within the Markov model. *Demogr* **47(3)**, 609–628.