Advances in Consumption-Investment Problems with Applications to Pension

Morten Tolver Kronborg

PhD Thesis

Supervisors:	Mogens Steffensen, University of Copenhagen
	Søren Fiig Jarner, ATP and University of Copenhagen
Co-supervisor:	Michael Preisel, ATP
Submitted:	March 28, 2014.

Department of Mathematical Sciences
Faculty of Science
University of Copenhagen

Author: Morten Tolver Kronborg
Dyssegårdsvej 72A
2870 Dyssegård
Denmark
mortentolverkronborg@gmail.com

Assessment Committee: Professor Holger Kraft,
Goethe University,
Frankfurt, Germany

Peter Holm Nielsen,
Chief Actuary,
PFA Pension, Denmark

Professor Rolf Poulsen,
University of Copenhagen,
Denmark (Chairman)

Contents

Preface iii

Summary v

Sammenfatning vii

List of papers ix

1 Introduction 1

2 Optimal consumption and investment with labor income and European/American capital guarantee 13

3 Optimal consumption, investment and life insurance with surrender option guarantee 35

4 Inconsistent investment and consumption problems 59

5 Why you should care about investment costs: A risk-adjusted utility approach 95

6 Entrance times of random walks: With applications to pension fund modeling 109

Bibliography 141

Preface

This thesis has been prepared in fulfillment of the requirements for the PhD degree at the Department of Mathematical Sciences, Faculty of Science, University of Copenhagen, Denmark. The project was funded by ATP and the Danish Agency for Sciences, Technology and Innovation under the industrial PhD program.

The work has been carried out under the supervision of Professor Mogens Steffensen, University of Copenhagen, and Søren Fiig Jarner, Chief Scientific Officer, ATP, and Adjoint Professor, University of Copenhagen, in the period from January 1, 2011 to March 31, 2014 (including 13 weeks of paternity leave).

The main body of the thesis consists of an introduction to the overall work, 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. There are minor notational discrepancy among the five chapters but it is unlikely to cause any confusion.

Acknowledgments

At first, I would like to thank ATP, especially Søren Fiig Jarner, Michael Preisel, and Chresten Dengsøe, for setting up and supporting the entire project. Second, I would like to thank my supervisors Søren Fiig Jarner and Mogens Steffensen, and my co-supervisor Michael Preisel, for many great discussions and advices during the PhD program, and also Søren and Mogens for co-authoring different parts of the thesis. I also want to show appreciation to my other colleagues at Quantitative Analysis, ATP, who have shown great interest in my work, and who have always been helpful when asking.

My gratitude extends to my colleagues at the Department of Mathematical Sciences for many both academic and non-academic moments. To the few of you I have been sharing offices with I would like to say; it has been a real pleasure.

I would also like to thank Professor Xunyu Zhou for facilitating my 3 month visit at Nomura Centre for Mathematical Finance, University of Oxford, England, and Dr Hanqing Jin for hosting and supervising me during my stay.

To my dear mother, who has taught me so much in life, and who has always motivated me in all the right ways to do my very best, I am deeply grateful.

Many special thanks go to my beloved wife, Christina, and our daughter, Ella, for their ceaseless love and support.

Finally, I want to dedicate this thesis to my dad who passed away while this work was taking place. You have always been a great inspiration, and I would have been very proud to show you my work.

Morten Tolver Kronborg
Dyssegård, March 2014

Summary

This thesis consists, in excess to an introductory chapter, of five papers within the area of consumption and investment decision making. The unifying topic is optimization of a stochastic wealth process. The underlying investment market is for all five papers the celebrated Black-Scholes market. What differentiates the papers are different wealth dynamics, optimization objectives, and possibly restrictions on the wealth process.

The first paper considers an investor, endowed with deterministic labor income, who searches to maximize expected cumulated utility from consumption and terminal wealth, while being restricted to keep wealth above a given barrier. The problem is solved by use of the Martingale Method for stochastic optimization problems in complete markets. The solution becomes an OBPI (option based portfolio insurance) strategy where the option bought to protect against losses in the unrestricted portfolio is an American put option.

The second paper extends the ideas of the first paper in two ways. First, we consider an investor with an uncertain lifetime, thereby also including utility from bequest, and introduce the possibility to invest in a life-insurance market. Second, the wealth restriction is allowed to depend, in a general way, on the wealth process. This allows for an analysis of the widespread pension saving product where a minimum rate of return on pension contributions is guaranteed.

The first two papers are summarized in Section 1.2, and the full papers are presented in Chapter 2 and 3, respectively.

The third paper, summarized in Section 1.3 and presented in Chapter 4, contributes to the new area of consistent optimization within classes of inconsistent problems. More formally, a class of non-linear objectives, for which the *Bellman Optimality Principle* does not hold, is considered. The two key examples treated are the mean-variance and mean-standard deviation problems, including both consumption, labor income, and terminal wealth, for an investor without pre-commitment. As explained in the paper the term "without pre-commitment" refers to the fact that we look for optimal strategies in the sense of Nash equilibrium strategies.

The forth paper of this thesis takes a utility approach to quantify the impact of investment costs. Concretely, we consider a power utility investor searching to maximize expected utility from terminal wealth. The impact of investment costs is split into a direct loss and an indirect loss. The direct loss is due to paying investment costs, and the indirect loss is due to lost investment opportunities, caused by the investors risk aversion. The indirect loss is measured by an indifference compensation ratio, defined as the minimum relative increase in the initial wealth the investor demands in compensation to accept incurring investment costs of a certain size. The magnitude of the indirect loss turns out to be between the same as and half of the expected direct loss, i.e. surprisingly big. Finally, a related analysis allows us to conclude that the size of the investment costs is of far more importance than the specific choose of investment strategy. The paper is summarized in Section 1.4, and the full paper is presented in Chapter 5.

Finally, the last paper, summarized in Section 1.5 and presented in Chapter 6, considers entrance times of random walks, i.e. the time of first entry to the negative axis. Partition sum formulas are given for entrance time probabilities, the n^{th} derivative of the generating function, and the n^{th} falling factorial entrance time moment. Similar formulas for the characteristic function of the position of the random walk both conditioned on entry and conditioned on no entry are also established. All quantities are also considered for the stationary process. The theoretical results are applied to analyze the widespread with-profit pension product. More precisely, exact (computable) formulas for the bonus time probabilities, the expected bonus size, and the expected funding ratio given no bonus are presented. Moreover, to conduct a mean-variance analysis for a fund in stationarity we devise a simple and effective exact simulation algorithm for sampling from the stationary distribution of a regenerative Markov chain.

Sammenfatning

Denne afhandling består, i tillæg til et introducerende kapitel, af fem artikler inden for feltet omhandlende forbrugs- og investeringsbeslutninger. Det gennemgående emne er optimering af stokastiske formueprocessor. Det underlæggende investeringsmarked er for alle fem artikler det anderkendte Black-Scholes marked. Artiklerne adskiller sig fra hinanden ved forskellige formuedynamikker, optimeringskriterier og mulige formuerestriktioner.

I den første artikel betragtes en investor, med en deterministisk lønindkomst, der søger at optimere den forventede akkumulerede nytte af forbrug og slutformue, under restriktionen at formuen altid skal være større end en givet barrie. Problemet løses ved brug af Martingal Metoden for stokastiske optimeringsproblemer i komplette markeder. Løsningen består af en optionsbaseret porteføljestrategi, hvor optionen der købes som forsikring mod tab i den urestringerede portefølje er en amerikansk put option.

Den anden artikel bygger videre på ideerne fra den første artikel på to måder. For det første betragtes en investor med en stokastisk levetid. Som følge heraf introduceres nytte af testamentering samt et livsforsikringsmarked. For det andet tillades formuen at være restringeret af en generel proces der afhænger af formuehistorikken. Dette muliggør en analyse af det udbredte pensionsprodukt der indeholder minimums garantiforrentninger af pensionsbidragene.

En opsummering af de to første artikler findes i sektion 1.2, og artiklerne er præsenteret i deres fulde længde i hhv. kapitel 2 og 3.

Den tredje artikel, der opsummeres i sektion 1.3 og præsenteres i kapitel 4, bidrager til det nye felt omhandlende konsistent optimering af inkonsistente problemer. Formelt betragtes en klasse af ikke-lineære optimeringsobjekter hvor *Bellmans Optimal Princip* ikke er gældende. De to behandlede hovedeksempler er middelværdi-varians og middelværdi-spredning, med lønindkomst, forbrug og slutformue, for en investor uden præ-præferencer. Som forklaret i artiklen refererer udtrykket "præ-præferencer" til det at vi søger efter optimale strategier i form af Nash ligevægtsstrategier.

Den fjerde artikel i afhandlingen kvantificerer betydningen af investeringsomkostninger ved brug af nytteteori. Konkret betragtes en potensnytte-investor der søger at maksimere den forventede nytte af slutformuen. Tabet som følge af investeringsomkostninger splittes op i et direkte og et indirekte tab. Det direkte tab er de faktisk betalte omkostninger, og det indirekte tab er knyttet til de, som konsekvens af investorens risikoaversion, tabte investeringsmuligheder. Det indirekte tab måles ved brug af et indifferent kompensations mål givet som størrelsen af den kompensation investoren forlanger for at acceptere investeringsomkostninger af en vis størrelse. Vi får at det indirekte tab er mellem den samme og halvdelen af det forventede direkte tab (altså overraskende stor). Endeligt fås ved en relateret analyse, at størrelsen af investeringsomkostningerne er af markant større betydning relativt til valg af investeringsstrategi. Artiklen opsummeres i sektion 1.4, og den fulde artikel præsenteres i kapitel 5.

Den sidste artikel, der opsummeres i sektion 1.5 og præsenteres i kapitel 6, betragter indgangstidspunkter for random walks (tidspunkt for første gang processen bliver negativ). Partitionssumsformler for indgangstidspunktssandsynlighederne, den n^{th} afledte af den genererende funktion, og det n^{th} faldende fakultetsmoment for indgangstidspunkterne præsenteres. Lignende formler udledes for den karakteristiske funktion af positionen af random walk processen både betinget med indgang og ingen indgang. Alle størrelser betragtes også for den stationære proces. De teoretiske resultater bruges til at analysere det udbredte med bonus pensions produkt. Mere præcist præsenteres eksakte formler for bonustidspunktsandsynlighederne, den forventede bonus, og den forventede fundinggrad givet ingen bonus. Derudover opfindes en simpel men effektiv eksakt simuleringsalgoritme til sampling fra den stationære fordeling af den regenerende Markov kæde. Dette bruges til at udføre en middelværdi-varians analyse af en fond i stationaritet.

List of papers

This thesis is based on five papers:

- Morten Tolver Kronborg (2012) Optimal consumption and investment with labor income and European/American capital guarantee.
Submitted for publication.

- Morten Tolver Kronborg and Mogens Steffensen (2013) Optimal consumption, investment and life insurance with surrender option guarantee.
Accepted for publication in *Scandinavian Actuarial Journal*.
Available at: http://dx.doi.org/10.1080/03461238.2013.775964

- Morten Tolver Kronborg and Mogens Steffensen (2013) Inconsistent investment and consumption problems.
Submitted for publication.

- Søren Fiig Jarner and Morten Tolver Kronborg (2013) Why you should care about investment costs: A risk-adjusted utility approach.
Submitted for publication.

- Søren Fiig Jarner and Morten Tolver Kronborg (2013) Entrance times of random walks: With applications to pension fund modeling.
Submitted for publication.

1. Introduction

This introductory chapter gives an overview of the contributions of this thesis. To this end the chapter is divided into 6 sections. Section 1.1 briefly presents the legacy of Merton (1969, 1971) and the general market model which forms the foundation for the work done in this thesis. Section 1.2 summarizes two papers about binding capital restrictions. Section 1.3 presents the main results from a paper which contributes to the new area of consistent optimization within classes of inconsistent problems. Section 1.4 presents a paper taking a risk-adjusted utility approach to quantify the impact of investment costs. Finally, Section 1.5 sums up the content of a paper containing several results for discrete-time random walks with continuous innovations together with an application to a collective with-profit pension scheme.

1.1 General inspiration for this thesis

To a large extent this thesis builds upon the celebrated pioneering work of Merton (1969, 1971). Within a continuous-time model Merton considered an investor with the objective to maximize the time-additive utility from consumption and terminal wealth: In its simplest form we have

$$\sup_{(c,\pi)\in\mathcal{A}'} E\left[\int_0^T u(c(t))dt + u(X(T))\right], \tag{1.1}$$

where $T > 0$ is the investment horizon, u a utility function, $\mathcal{A}'$ the set of admissible strategies, and (c, π) the consumption-investment strategy affecting the wealth process $(X(t)_{t\in[0,T]})$ with initial wealth $X(0) = x_0$. The approach used by Merton is called *dynamic programming* which turns the stochastic problem (1.1) into a deterministic optimization problem and a deterministic linear differential equation, known as the *Hamilton-Jacobi-Bellman equation*. In the special case of a Black-Scholes investment market and HARA (hyperbolic absolute risk aversion) utility functions, Merton found closed-form solutions for the optimal investment-consumption strategy in feedback forms.

For this thesis the Black-Scholes market is also the market model of choice. In principle, a financial market model should contain more than one stock. However, it is well known from the Mutual Fund Theorem by Merton (1971) that for a HARA utility investor, as we shall consider, the optimal asset allocation will always be between an optimal portfolio of the stocks weighted by their sharp ratios, and the risk-free bank account (at least for diffusion processes with deterministic coefficients). Also, the main focus of this thesis is to extend the classic optimization problem of Merton's in new directions. A possible extension of the market model at the same time will blur the picture by allowing only for semi-explicit results and more cumbersome notation, thereby blurring the points to be made in this thesis.

1.2 Capital guarantees

In reality, the number of portfolio problems which can be solved explicitly by the dynamic programming approach used by Merton are very limited. In fact, even the numerical tractability is very limited.

Therefore, in Chapters 2 and 3 we consider another approach to solve such problems, called the *Martingale method*. The method, developed by Cvitanic et al. (1987), Cox and Huang (1989) and Cox and Huang (1991), builds upon the martingale theory for complete markets. It is a more direct approach which splits the problem into a time-static optimization problem for the optimal consumption process and the terminal wealth process, and the problem of constructing a self-financing strategy replicating the optimal terminal wealth. To demonstrate the simplicity of the martingale method we consider briefly Merton's problem given by (1.1). Define the inverse of the derivative of the utility function u as the function $I : (0, \infty] \to [0, \infty)$ and define the adjusted state price deflator

$$H(t) = \Lambda(t)e^{-rt},$$

where Λ is the Girsanov kernel from the measure transformation from the real probability measure P to the equivalent pricing measure Q. Only two central results and one simple observation are needed to solve the problem.

1. The completeness of the market model ensures that the budget constraint,

$$E^Q\left[\int_0^T e^{-rt}c(t)dt + e^{-rT}X(T)\right] \leq x_0, \tag{1.2}$$

is true for all admissible strategies (see. e.g Korn (1997b) Theorem 7).

2. There exists a unique constant $\xi^* > 0$ such that (see. e.g. Korn (1997b) Proposition 15 (27))

$$E^Q\left[\int_0^T e^{-rt}I(\xi^* H(t))dt + e^{-rT}I(\xi^* H(T))\right] = x_0. \tag{1.3}$$

3. From the concavity of u, and since $u'(I(z)) = z$, we obtain

$$u(x) \leq u(I(z)) - z(I(z) - x), \ \forall x \geq 0, z > 0. \tag{1.4}$$

Now, by use of (1.2)–(1.4) it follows easily that for an arbitrary strategy (c, π)

$$\begin{aligned} &E\left[\int_0^T u(c(t))dt + u(X(T))\right] \\ &\leq E\left[\int_0^T u(I(\xi^* H(t)))dt + u(I(\xi^* H(T)))\right]. \end{aligned}$$

We get directly the *candidate* optimal strategy

$$c^*(t) = I(\xi^* H(t)), \tag{1.5}$$
$$X^*(T) = I(\xi^* H(T)). \tag{1.6}$$

The replicating investment strategy, yielding the optimal wealth process, can be found by use of the martingale representation theorem (see e.g. Karatzas and Shreve (1991)). The calculation of an exact expression for the consumption-investment strategies from (1.5) and (1.6) is the hard part. However, if no closed-form solution is available, numerical computations can be obtained from (1.5), (1.6) and (1.2) by Monte Carlo simulations.

The Martingale method is convenient for solving the problems considered in Chapter 2 and 3. We use the fact that the optimal strategy can be represented in the forms given by (1.5) and (1.6) to solve expanded versions of Merton's problem, including binding capital constraints.

Consider an investor endowed with a deterministic labor income ℓ and define the financial value of the income stream by $g(t) = \int_t^T e^{-r(s-t)}\ell(s)ds$ and the total wealth process by

$X(t) + g(t)$. Solving Merton's problem for such an investor is simple since it turns out that the problem is equivalent to the problem without labor income but with an enlarged initial wealth equal to $x_0 + g(0)$. In Chapter 2 we consider Merton's problem (1.1) with the addition of a possible time-preference (discount) parameter β, deterministic labor income at rate ℓ, and (most importantly) a capital restriction allowing only for strategies with corresponding wealth processes being greater or equal to a certain barrier K at all times. Moreover, we consider a CRRA (constant relative risk aversion) utility function. The problem is set as

$$\sup_{(c,\pi)\in\mathcal{A}'} E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + e^{-\int_0^T \beta(s)ds} u(X(T))\right]. \tag{1.7}$$

under the capital constraint

$$X(t) \geq K(t),\ \forall t \in [0, T].$$

Note that if $K(t) = Ae^{-r(T-t)} - g(t)$ for some constant $A \leq e^{rT}(x_0 + g(0))$, the capital constraint reduces to the terminal constraint $X(T) \geq A$. This problem, including a terminal constraint (only), was first solved by Teplá (2001). The solution consists of an OBPI (option based portfolio insurance). The investor optimally invests a fraction of the initial total wealth in the unrestricted portfolio and uses the remainder to buy a European put option with strike price A on that portfolio. The split between money spent on the unrestricted portfolio and the put option is determined by the initial budget constraint. The important lesson to learn is that a rate of return guarantee of r on a part of the initial wealth does not significantly increase the complication since the capital restriction reduces to a terminal restriction. To get a glimpse of the strategy Figure 1.1 illustrates for a CRRA investor the two components of the OBPI, the unrestricted portfolio and the put option, for a stock market outcome where the free optimal solution to Merton's problem does not fulfill the terminal capital constraint.

On the other hand, a rate of return guarantee less than the risk free rate, $\widetilde{r} < r$, on a part of the initial wealth (possible the entire) complicates things much more. That is, $K(t) = Ae^{-\widetilde{r}(T-t)} - g(t)$ for some constant $A \leq e^{\widetilde{r}T}(x_0 + g(0))$. In that case the constraint does not reduce to a terminal constraint. In a setting excluding consumption and labor income El-Karoui et al. (2005) show that the optimal strategy is, again, an OBPI strategy. However, opposed to Merton's problem with or without a terminal capital constraint, including labor income to this setting is not trivial. Being endowed with labor income or being endowed with an enlarged (by the financial value of future labor income) initial wealth are not equivalent for this case. In fact, the latter is preferable. This indicates that adding deterministic labor income to the setting considered by El-Karoui et al. (2005) is not as easy as it may seem. The strategy actually differs from the case excluding deterministic labor income. The main contribution of Chapter 2 is to show that the OBPI approach taken by El-Karoui et al. (2005) can be generalized to the case including consumption and labor income, thereby solving the problem stated in (1.7). The solution becomes, as in Teplá (2001), an OBPI strategy. However, for this case the option bought to protect against losses in the unrestricted portfolio is an American put option. The investor optimally invests a fraction of the initial total wealth in the unrestricted portfolio and uses the remainder to buy an American put option. However, in order for the strategy to fulfill the capital constraint at all times the strike price of the put option becomes $K(t) + g(t)$, i.e. the strike price becomes time dependent and involves the financial value of the future labor income. Moreover, in order for the strategy to be admissible, the American put option has to be sold whenever it is optimal to do so. This involves a continuous re-calibration of the split between the money spent on the unrestricted portfolio and the money spent on buying insurance. In other words, the budget constraint for the unrestricted portfolio must be adjusted whenever the capital constraint is active. A great part of the work done in Chapter 2 is to show that this strategy is in fact admissible. This done, using the expression for the unrestricted strategy (1.5) and (1.6), using the concavity of the utility function, and the product property of CRRA utility, it becomes elegant to show that the strategy dominates all admissible strategies

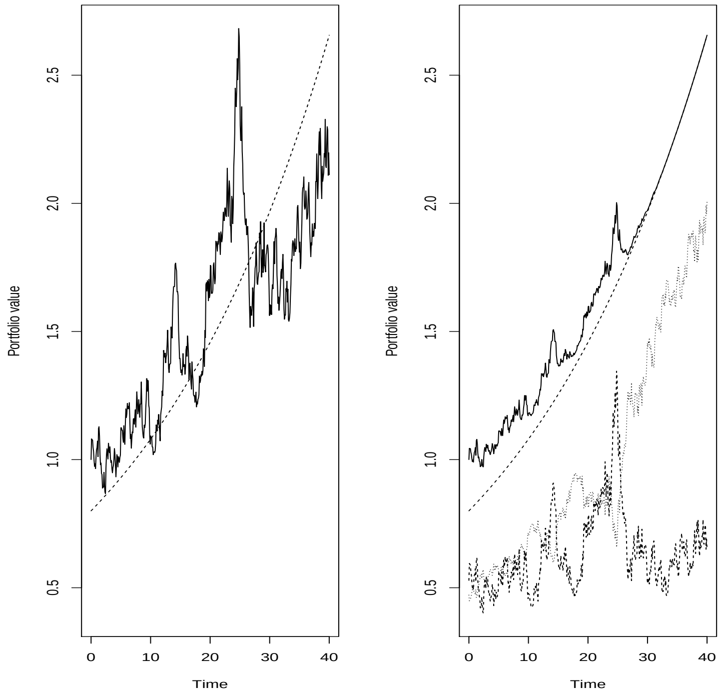

Figure 1.1: The initial wealth is one unit, and the smooth dashed curve illustrates the capital restriction $K(t) = 0.8x_0e^{rt}$, which is a terminal constraint only. Left: The unrestricted optimal Merton portfolio. Right: The optimal OBPI (solid curve), the corresponding down-scaled unrestricted portfolio (dotted curve), and the corresponding put option value (dashed curve) (which ends up "in-the-money").

in terms of expected utility. One special example of problem (1.7) considered in Chapter 2 is the case $K = 0$. In that case the wealth process has to be non-negative at all times. In the present of labor income, this becomes a real restriction since then the investor can not borrow against future labor income. Numerical illustrations for this example are presented in Chapter 2.

In Chapter 3 the setting of Chapter 2 is expanded. First, we consider an investor with an uncertain lifetime, thereby also including utility from bequest. To make the market complete we introduce the possibility to invest in a life-insurance market. Second, the capital restriction is allowed to depend, in a general way, on the wealth process, i.e. the constraint becomes stochastic. This is done in order to model the real-life example of a pension savings account where a minimum guarantee on the rate of return is promised by the pension company. Money not spent becomes pension contributions and we interpret the wealth process as the *reserve*. Traditionally, the reserve is maintain by the pension company, and upon death before retirement the pension saver receives the life-insurance sum (bequest) and not the reserve itself. This is also the case in Chapter 3. Note that a bequest larger (smaller) than the reserve corresponds to buying life-insurance (annuities). The problem can be written as

$$\sup_{(c,\theta,D)\in\mathcal{A}'} E\Big[\int_0^T e^{-\int_0^t(\beta(s)+\mu(s))ds}\Big(u(c(t)) + K_1\mu(t)u(D(t))\Big)dt + K_2e^{-\int_0^T(\beta(s)+\mu(s))ds}u(X(T))\Big], \tag{1.8}$$

under the capital guarantee restriction

$$X(t) \geq k(t, Z(t)),\ \forall t \in [0, T],$$

where μ is the mortality rate, D the sum to be paid out upon death before the time of retirement T, and $Z(t) := \int_0^t h(s, X(s))ds$ for deterministic functions k and h (K_1 and K_2 are weight). Of special interest arethe two examples

$$k(t, z) = 0, \tag{1.9}$$

and

$$k(t, Z(t)) = x_0 e^{\int_0^t \left(r^{(g)} + \mu(y)\right)dy} + \int_0^t e^{\int_s^t \left(r^{(g)} + \mu(y)\right)dy} \left[\ell(s) - c(s) - \mu(s)D(s)\right] ds, \tag{1.10}$$

where $r^{(g)} < r$ is the guaranteed rate of return on the investments, and μD the fair price for the life-insurance. The first case (1.9) is simply motivated by the fact that by law the reserve is restricted to be non-negative at all times. The latter case (1.10) imitates the real life example of a pension fund guaranteeing a minimum rate of return r^g on the pension contributions. Clearly, the constraint is binding at every intermediate time point in the saving period. We call it a "surrender option guarantee". In most countries the pension saver holds the right, by law, to stop paying premiums and leave the company (to surrender). Therefore the guarantee must be fulfilled at all times.

The solution to the problem consists of an even more complicated OBPI strategy on the unrestricted optimal portfolio. This time the strike price of the American put option becomes $k(s, Z(s)) + g(s)$, i.e. a stochastic strike price. The fact that the guarantee, and consequently the strike price, depends on the wealth process through an integral over the wealth process, and not on present wealth, the approach taken in Chapter 2 becomes applicable once more. However, since the strike price of the put option becomes path dependent, in the same kind of manner as for an American-Asian put option, extra terms arise when using Itô's calculus to show that the OBPI strategy is admissible. For further details see Chapter 3, Appendix 3.6. To get a sense of how a minimum guarantee rate of return on the pension contributions affects the pension at retirement, see Table 3.1. Here it is demonstrated, for a guaranteed rate of return equal to half the risk-free rate, that the portfolio insurance incorporated in the OBPI strategy turns out very useful in the event of a very unfortunate investment market, while the loss incurred by holding a portfolio insurance out of the money, in the event of a fortunate investment market, is less perceptible.

1.3 Inconsistent problems

Traditional stochastic control problems in finance, such as Merton's problem (1.1), are aimed at linear objectives. For such problems the iterated expectation property holds. Consequently, we say that the *Bellman Optimality Principle* holds: Consider an investor with wealth process $(X(t))_{t\in[0,T]}$ and initial wealth $X(0) = x_0$. Suppose that we find the optimal strategy u^* for a time-0 problem $\mathcal{P}_{0,x_0}$ and suppose that we use this strategy on the time interval $[0, t]$, $t < T$. Then at time t the strategy u^* will *still* be optimal for the time-t problem $\mathcal{P}_{t,X^{u^*}(t)}$ (here we have specified that the wealth process at time t depends on the strategy used before time t).

In recent years a small amount of literature on inconsistent problems has emerged. One very relevant and easy to understand example is the endogenous habit formation problem considered by Kryger and Steffensen (2010). Again we consider a Black-Scholes market and an investor with time horizon $[0, T]$ and an investment strategy π denoting the fraction of wealth to invest in risky stocks. The non-linear object of consideration is

$$\inf_{\pi} E\left(\frac{1}{2}(X^\pi(T) - x_0\beta(t))^2 \,\middle|\, X(t) = x_t\right). \tag{1.11}$$

Letting $\beta(t) = e^{\widetilde{r}t}$, for some target rate of return $\widetilde{r}$, the interpretation of the problem is clear. For the time-0 version of this problem, the goal is to earn a rate of return over the interval $[0, T]$ equal to $\widetilde{r}$, and the performance of any investment strategy π is measured by the quadratic variation of the terminal wealth process to this goal. This is a solvable problem which can be handled by dynamic programming or the martingale method (see Korn (1997a) and Zhou and Li (2000)), and we refer to the solution as the *pre-commitment* solution. The problem seems at first glance reasonable since a target for the rate of return on investments is something many investors have a strong opinion about. Now the question is whenever this objective is actually the real objective for the investor? The not so attractive feature of the objective is that if the investor obtains large returns from investments in the period (say) $[0, t], \; t < T$, the investor becomes less ambitious about the rate of return for the remainder of the period, $(t, T]$. In practice, if the investor becomes rich enough, the optimal pre-commitment strategy for the objective (1.11) tells the investor to stop taking risky positions. An objective which may seem more realistic is the objective to earn a certain rate of return at any time. In discrete time, the investor chooses his investment strategy bearing in mind that the next day, with the investment horizon one day shorter, the target will also be to earn the target rate of return over this horizon. In other words, the investor does not become less ambitious regarding his investment profit just because he did well last year (week or day, etc.). The intuition can be preserved when considering the continuous-time case. As explained by Björk and Murgoci (2009) the approach corresponds to looking for a Nash equilibrium strategy for the problem. We refer to the approach as *without pre-commitment*.

This way of thinking has lead to a literature founded by Basak and Chabakauri (2010) who considers the dynamic asset allocation problem for a portfolio investor searching to maximize the mean-variance objective

$$E_{t,x}\left[X^{\pi}(T)\right] - \frac{\gamma}{2} Var_{t,x}\left[X^{\pi}(T)\right], \tag{1.12}$$

for a constant γ. By applying a so-called total variance formula they obtain an extension of the classical Hamilton-Jacobi-Bellman equation for solving the problem for an investor without pre-commitment. Björk and Murgoci (2009) extend the class of standard solvable problems to the class of objectives

$$E_{t,x}\left[\int_t^T C(x, s, X^u(s), u(s))ds + \phi(x, X^u(T))\right] + G\big(x, E_{t,x}\left[X^u(T)\right]\big), \tag{1.13}$$

for some function G. In (1.13) time inconsistency enters at two points: First, the present state x appears in C, ϕ and G, and second, the function G is allowed to be non-linear in the conditional expectation. The central example of Björk and Murgoci (2009) is the mean-variance problem (1.12) of Basak and Chabakauri (2010). One unfortunate feature of the problem is that the optimal amount of money to invest in stocks is constant. In responds, Björk et al. (2012) extend the problem to a more complicated problem allowing for a state dependent risk aversion. The motivation is the key example

$$E_{t,x}\left[X^{\pi}(T)\right] - \frac{\gamma}{2x} Var_{t,x}\left[X^{\pi}(T)\right]. \tag{1.14}$$

In Chapter 4 we expand the existing literature to a larger class of problems also taking into account consumption and possible labor income. Loosely speaking, the class of stochastic problems we consider is, for any $(t, x) \in [0, T) \times \mathbb{R}$, to maximize

$$f(t, x, y^{c,\pi}(t, x), z^{c,\pi}(t, x)), \tag{1.15}$$

where $f \in \mathcal{C}^{1,2,2,2}$, $\mathcal{A}$ is the class of admissible strategies, and

$$y^{c,\pi}(t,x) = E\left[\int_t^T e^{-\rho(s-t)} c(s) ds + e^{-\rho(T-t)} X^{c,\pi}(T) \,\middle|\, X(t) = x \right],$$

$$z^{c,\pi}(t,x) = E\left[\left(\int_t^T e^{-\rho(s-t)} c(s) ds + e^{-\rho(T-t)} X^{c,\pi}(T) \right)^2 \,\middle|\, X(t) = x \right],$$

with ρ being a constant discount rate, possibly different from the interest rate r. Compared to the general problem (1.13) considered by Björk and Murgoci (2009), and for state dependent risk aversion by Björk et al. (2012), the class of problems (1.15) differs in three ways. First, in excess for the reasons listed for problem (1.13), the inconsistency also arises from taking a non-linear function of the expected (utility of) consumption. Second, the class contains problems, such as the mean-standard deviation problem, which are not contained in (1.13). Third, we allow for a capital injection in the form of a deterministic labor income. This leads to some mathematical difficulties but we manage to establish a verification theorem containing a Bellman-type set of differential equations for determination of the optimal strategies for an investor without pre-commitment. The two main problems considered in Chapter 4 are the mean-variance and mean-standard-deviation problems given for $(t,x) \in [0,T) \times \mathbb{R}$ by

$$f(t,x,y,z) = y - \frac{\psi(t,x)}{2}\left(z - y^2\right), \tag{1.16}$$

$$f(t,x,y,z) = y - \psi(t,x)\left(z - y^2\right)^{\frac{1}{2}}, \tag{1.17}$$

respectively. Here $\psi \in \mathcal{C}^{1,2}$ is a function which is allowed to depend on time and wealth.

As an example we consider (1.16) for the case of time and state dependent risk aversion where the investor's risk aversion is hyperbolic in present wealth plus the financial value of future labor income net of consumption. Solving the mean-variance problem we restrict the class of admissible strategies to strategies for which the corresponding wealth plus the financial value of future labor income net of consumption stays positive. This conforms with the well-known and often required constraint that wealth plus human capital must stay positive at all times. Consequently, terminal wealth becomes positive. The optimal investment strategy becomes linear in the investor's wealth plus financial value of future labor income net of consumption. The optimal consumption rate becomes a deterministic bang-bang strategy. The deterministic function determining when it is optimal to consume the maximum or minimum allowed is given by a system of non-linear differential equations for which we have no explicit solution. It turns out that it is not always optimal to either consume the maximum or minimum allowed at all times. For some investors we find that it is optimal to consume the maximum allowed for a period of time, and then to consume the minimum allowed for the remainder of the period.

Finally, we show that the mean-standard-deviation problem (1.17) admits the trivial strategy of zero risky investment. The problem and its solution helps us gain some intuition about inconsistent optimization problems for investors without pre-commitment. We conclude that over an infinitesimal time interval, dt, standard deviation is of the order $\sqrt{dt}$, which means that the punishment is so hard that any risk taking is unattractive. For a very intuitive explanation, which also could serve as a first introducing to optimization of inconsistent problems for investors without pre-commitment, see Subsection 4.5.2 of Chapter 4.

1.4 Investment costs

Introducing proportional investment costs for risky investments, ν, to the Black-Scholes market implies, of course, a loss for the investor. If, again, we let π denote the investment strategy in terms of the fraction of wealth to invest in stocks, the loss is often simply quantified by the *direct loss* in the median rate of return given by $\pi\nu$.

However, there seems to be a lack of more serious approaches to quantify the impact of investment costs. Therefore, we return in Chapter 5 to the original problem of Merton, and consider (1.1) for an investor with CRRA preferences for terminal wealth (no consumption). This opens up the possibility to do a *risk-adjusted evaluation of investment costs*. By use of a Value at Risk measure Guillén et al. (2013) also perform a risk-adjusted evaluation of investment costs.

The main assumption essential for the analysis in Chapter 5 is that no investor can beat the market. Something many expensive mutual fund managers claim when asked by customers about the high fees they demand. In other words, we assume a zero correlation between cost ratios and expected investment returns. Something showed to be true by a vast amount of literature (see Chapter 5, Subsection 5.1, for references).

The point of view taken in Chapter 5 is the situation of declining investment costs. One should bear in mind the situation of an investor who has handed over his savings to a mutual fund that on his behalf invests the money. For the service the mutual fund charges a fee. Under the assumption of the existence of another fund charging a lower fee and performing (in expectation) equally well, we quantify the impact of paying the higher fee. We call an investor taking risk aversion into account, by use of a concave strictly increasing utility function, when deciding upon an investment strategy a *sophisticated investor*. The first main point of Chapter 5 is:

- The sophisticated investor experiences a greater change in the rate of return induced by a change in investment costs compared to the direct loss.
- This is so since the risk aversion of the investor optimally dictates a part of the money the investor saves by paying less investment costs to be invested in stocks.

The evaluation is based on an *indifference compensation ratio* defined as the minimum relative increase in the initial wealth the investor demands in compensation to accept incurring the higher investment costs. Let ν_1 and ν_2 denote the higher and lower cost ratio, respectively. Formally, we define the indifference compensation ratio $ICR_{(\nu_1,\nu_2)}$ by the relation

$$\begin{aligned}&\sup_{\pi} E\left[u\left(X^{(\nu_2,\pi)}(T)\right)\middle|X^{(\nu_2,\pi)}(0)=x_0\right]\\ &=\sup_{\pi} E\left[u\left(X^{(\nu_1,\pi)}(T)\right)\middle|X^{(\nu_1,\pi)}(0)=x_0\left(1+ICR_{(\nu_1,\nu_2)}\right)\right].\end{aligned} \tag{1.18}$$

We show that for a CRRA function (μ) the measure is equivalent to the relative change in certainty equivalents. We compare this measure to the present value of the investment costs the investor is expected to pay over the investment horizon, evaluated as a fraction of the initial wealth. We interpret the latter value as the *direct effect* of investment costs, and the difference between the direct effect and the ICR-value, as the *indirect effect* of investment costs. For realistic market parameters the conclusion is rather surprising. For an investor, who on one hand has the possibility to invest through a mutual fund being charged 0.6% in investment costs, but who invest through a mutual fund charging him investment costs at a ratio equal to 1.4%, with an optimally stock allocation of 60%, the indirect effect is between 50–100% of the direct effect, dependent on the investment horizon (longer horizon, higher indirect effect).

Finally, the approach allows us to answer a related question: Is the investment strategy or the size of the investment costs of most importance? Concretely, we study an investor facing high investment costs and an optimal investment strategy (w.r.t. his risk profile), and ask which investment strategies the investor is willing to accept if he at the same time is offered lower investment costs. The conclusion, the second main point of Chapter 5, is very clear

- Independent of the time horizon, the particular allocation towards risky stocks seems to be of very little importance compared to the size of investment costs.

To demonstrate the magnitude consider again the investor who invests through the expensive mutual fund charging him high investment costs at rate 1.4%, with an optimally stock allocation of 60%. For realistic market parameters we calculate that for the investor who shifts to the cheaper fund offering to manage his savings charging him a lower cost rate off 0.6%, any investment strategy allocating between 28% and 129% of his wealth in stocks makes him equally or better of. This examples quantifies why collective investment scheme with a common strategy for all members, thereby maintaining costs low, could be preferable for many investors.

1.5 A collective with-profit pension scheme

In the last chapter of this thesis, Chapter 6, we derive several results for discrete-time random walks with continuous innovations. Of special interest is the entrance time (time of return to origin) distribution. We derive an expansion formula for the entrance time probabilities, a formula for the n^{th} derivative of the generating function, a formula for the n^{th} moment of the waiting times, and formulas for the conditional undershoot process given the first entrance time. All quantities are also considered for the stationary process.

First, the analysis of the discrete-time random walk with continuous innovations is triggered by the desire to analyze the investment decision for a so called with-profit collective pension scheme, a very widespread pension product in e.g. the Nordic countries and the Netherlands.

In Chapter 6 each contribution is split into a part giving a guaranteed payment and a part invested in a, possibly leveraged, investment portfolio. Define the funding ratio F by

$$F(t) = \frac{A(t)}{R(t)},$$

where $A(t)$ denotes the total assets of the fund and $R(t)$ the financial value of the guaranteed benefits, known as the *reserve*. In our model both contributions and benefits are paid and withdrawn, respectively, such that the funding ratio is not affected, i.e. the product is fair. The difference between the total assets and the reserve, $A(t) - R(t)$, is known as the *bonus potential*. It is the bonus potential which is invested in stocks (possible leveraged). The product is with-profit in the sense that all guaranteed payments are increased, known as bonus, when the funding ratio exceeds a given threshold κ. The decision to attribute bonus or not is taken at a set of equidistant discrete set of time points $0 = t_0 < t_1 < \ldots$. We assume for simplicity that contributions and benefits also fall at these times. After (possible) bonus attribution the funding ratio becomes $\min\{F(t_i), \kappa\}$.

A simple version of the pension product, to be considered as an example in Chapter 6, is the case of a unit contribution made at time $t = 0$ for a benefit paid out in its entirety at time $t = T$ (retirement). In this case the benefit at retirement becomes

$$\frac{F(T)}{F(0)} e^{rT} \prod_{i=1}^{T} \left(1 + r_i^B\right), \tag{1.19}$$

where r_i^B is the possible bonus attribution at time t_i.

It follows from Preisel et al. (2010) that in a Black-Scholes market model the funding ratio process becomes a (discrete-time) Markov chain with dynamics

$$F_i = \min\left\{(F_{i-1} - 1)e^{\left(C\mu - \frac{1}{2}C^2\sigma_Z^2\right)\Delta + C\sigma_Z\sqrt{\Delta}U_i} + 1, \kappa\right\}, \tag{1.20}$$

where the U_i's are i.i.d. standard normal variables, μ and σ_Z the expected excess return and volatility on stocks, respectively, and C the fraction of the bonus potential invested in stocks. The funding ratio process (1.20) has also been studied by Preisel et al. (2010) and Kryger (2010) using a number of analytical approximations. In contrast, Chapter 6 relies almost exclusively on exact results. We consider the following transformation

$$Y_n = -\log\left(\frac{F_n - 1}{\kappa - 1}\right) \text{ for } n \in \mathbb{N}_0. \tag{1.21}$$

This turns the funding ratio process (1.20) into the one-sided random walk

$$Y_n = (Y_{n-1} + X_n)^+, \tag{1.22}$$

where the X_n's are i.i.d. normally distributed with mean $-\left(C\mu - \frac{1}{2}C^2\sigma_Z^2\right)\Delta$ and variance $C^2\sigma_Z^2\Delta$. Note that $F_n = \kappa$ corresponds to $Y_n = 0$, while funding ratios close to one correspond to high values of Y. In other words, bonus is attributed whenever the Y-process returns to zero. Along with Y we also consider the (unrestricted) random walk

$$S_0 = 0 \text{ and } S_n = S_{n-1} + X_n \text{ for } n \in \mathbb{N}, \tag{1.23}$$

with the same X_n's as in (1.22).

Now with the model of inspiration in place we return to a general discrete-time random walk with continuous innovations. First we consider the entrance time probabilities

$$\tau_n = P(\tau_- = n) = P(S_1 > 0, \ldots, S_{n-1} > 0, S_n \leq 0). \tag{1.24}$$

A central concept in Chapter 6 is the set of integer partitions of a given order $n \geq 1$ given by

$$\mathcal{D}_n = \left\{(\sigma_1, \ldots, \sigma_n)\middle|\sigma_1 \in \mathbb{N}_0, \ldots, \sigma_n \in \mathbb{N}_0, \sum_{i=1}^{n} i\sigma_i = n\right\},$$

where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$.

First, we show by use of the generating function, $\tau(s) = \sum_{n=1}^{\infty} \tau_n s^n$, and the Theorem of Sparre-Andersen (see Theorem 6.2.1) which expresses the generating function in terms of the marginal probabilities $P(S_n \leq 0)$, that the first entrance time probabilities (1.24) can be computed as a partition sum over the marginal probabilities of the random walk, see Theorem 6.2.3. The Theorem is also implicit in the paper by Spitzer (1956), but in Chapter 6 we give a simpler self-contained proof.

Second, we derive an expression for the n'th derivative of the generating function for the entrance times, $\tau^{(n)}(s)$, see Theorem 6.2.5, which we subsequently use to derive a formula for the factorial moments, $E((\tau_-)_n)$, see Theorem 6.2.6. Again the results consist of a partition sum over marginal probabilities of the random walk. Both results are new and presented for the first time in Chapter 6.

Third, we derive formulas for the characteristic function of the position of the random walk conditioned on entrance at time n, $E\left(e^{i\zeta S_n}|\tau_- = n\right)$, see Theorem 6.2.8, and on entrance after time n, $E(e^{i\zeta S_n}|\tau_- > n)$, see Theorem 6.2.9. The results consist of partition sum formulas of partial expectations for the characteristic function of the random walk. The first of these results is known, but the proof offered by Chapter 6, which builds upon Theorem VII 4.1 of Asmussen (2003) which is a generalization of the Sparre-Andersen theorem to combined generating and characteristic functions (see Theorem 6.2.7), is new. The second result is new and presented for the first time in Chapter 6.

Returning to the pension model our prime interest is the stationary process. The stationary process has the interpretation of an average scenario. We therefore define by

$$\mathcal{T}_1 = \inf\{n \geq 1 : F_n = \kappa\} = \inf\{n \geq 1 : Y_n = 0\}, \tag{1.25}$$

and, recursively, for $k \geq 2$

$$\mathcal{T}_k = \inf\{n \geq 1 : F_{\mathcal{T}_{k-1}+n} = \kappa\} = \inf\{n \geq 1 : Y_{\mathcal{T}_{k-1}+n} = 0\}, \tag{1.26}$$

the time between consecutive bonus attributions. The results of Theorem 6.2.3–6.2.9 is connected to the case where $F_0 = \kappa$. In Subsection 6.4 the case $F_0 \sim \pi$, i.e. the stationary process, is under consideration. Using the results for the $F_0 = \kappa$ case more or less explicit closed form solutions, for the first entrance time probabilities and the characteristic function of the position of the random walk conditioned on entrance at time n, are derived for the stationary process. The key to analyzing the pension model is (1.21). For $F_0 = \kappa$ the events $(\mathcal{T}_1 = n)$ and $(\tau_- = n)$ are identical and on this event the bonus percentage becomes

$$r_n^B = \frac{\kappa - 1}{\kappa}\left(e^{-S_n} - 1\right).$$

Likewise on $F_0 = \kappa$ the events $(\mathcal{T}_1 > n)$ and $(\tau_- > n)$ are identical and on this event

$$F_n = (\kappa - 1)e^{-S_n} + 1.$$

By these quantities we analyze the first bonus time probabilities (the time between consecutive bonus attributions), the probability distribution of the number of bonuses, the expected bonus size, and the expected funding ratio. Furthermore, these results allow us to analyze the mean and variance of the pension benefit (1.19). For an average scenario (in stationarity) we develop an exact simulation approach and show that the expected pension benefit admits an interior solution in the investment strategy C. At some point, taking more risk results in a lower expected payoff. This is in great contrast to the prevailing idea that taking more risk should be compensated by higher expected returns, but the interpretation is clear. High allocation towards risky assets implies a high conditional expected bonus, but comes with the risk of being trapped at low funding ratios for a long time (rare bonuses). The analysis of Chapter 6 presents the optimal trade-off between the two effects.

2. Optimal consumption and investment with labor income and European/American capital guarantee

Abstract: We present the optimal consumption and investment strategy for an investor, endowed with labor income, searching to maximize utility from consumption and terminal wealth when he is restricted to fulfill a binding capital constraint of a European (constraint on terminal wealth) or an American (constraint on the wealth process) type. In both cases the optimal strategy is proven to be of the option based portfolio insurance type. The optimal strategy combines a long position in the optimal unrestricted allocation with a put option. In the American case, where the investor is restricted to fulfill a capital guarantee at every intermediate time point over the interval of optimization, we prove that the investor optimally changes his budget constraint for the unrestricted allocation whenever the constraint is active. The strategy is explained in a step by step manner and graphically illustrations are presented in order to support the intuition and to compare the restricted optimal strategy with the unrestricted optimal counterpart.

Keywords: Stochastic control; martingale method; option based portfolio insurance; American put option; human capital; borrowing constraint; CRRA utility.

2.1 Introduction

We study the optimal consumption and investment decision problem for an investor, endowed with deterministic labor income, who faces an American capital guarantee restriction. An American capital guarantee is a capital constraint which is valid at all time points in the time interval of consideration. Likewise, we refer to at capital guarantee only valid at the terminal time point as a European capital guarantee. The names are inherited from the option framework, i.e. from American and European options. The problem is set in a continuous-time frictionless Black-Scholes market and the investor measures his tolerance towards risk by a constant relative risk aversion utility function.

The corresponding free consumption and investment decision problem, where the investor is not restricted to fulfill any capital guarantee, is a very well understood problem, see Merton (1969) and Merton (1971). Traditionally, such problems have been solved using stochastic dynamic programming. However, stochastic dynamic programming becomes difficult when adding a capital constraint since the value function, in addition to a stochastic differential equation, must satisfy some boundary condition whenever the wealth is at the boundary. This leaves us with a non-linear Hamilton-Jacobi-Bellman equation for which the existence of a solution has to be established.

Consequently, this article uses instead the martingale methodology developed by Cvitanic et al. (1987), Cox and Huang (1989) and Cox and Huang (1991). We adopt the ideas from El-Karoui et al. (2005) and use the put option based approach to derive the optimal consumption and investment strategy for an investor who faces an American capital guarantee restriction defined as a deterministic function. It turns out that the optimal strategy consists of investing a part of the initial total wealth[1] in the corresponding optimal unrestricted allocation and the

[1]Initial wealth plus the financial value of future labor income.

remainder in a put option written on the optimal unrestricted portfolio. In order for this strategy to be admissible we show that the investor must adjust the portfolio continuously whenever the constraint is active. This result is similar to the one of El-Karoui et al. (2005), although they do not allow for consumption and labor income. The main contribution of this article is that we show that the put option based approach can also be carried out to find the optimal strategy in the case including consumption and labor income. Clearly the introduction of consumption leads to further mathematical difficulties, but in fact nor is the introduction of deterministic labor income trivial. One important thing to note is that, in contrast to the unrestricted problem of Merton, being endowed with labor income or being endowed with an initial wealth enlarged by the financial value of future labor income is *not* equivalent when considering an American capital guarantee restriction. In the case of an American capital guarantee the latter case is preferable. This indicates that adding deterministic labor income to the setting considered in El-Karoui et al. (2005) is not as easy as it may seem. The strategy actually differs from the case excluding deterministic labor income. We show that adding deterministic labor income can be handled by buying an insurance, a put option, with a strike price substantially different from the one obtained by El-Karoui et al. (2005). To ease the understanding of the optimal strategy derived the analysis also contains a step by step explanation of the optimal strategy together with a numerically graph-based explanation. In the special case of a restriction to borrow against future labor income (an American capital guarantee equal to zero), we compare numerically the restricted solution with the unrestricted counterpart.

A reasonable question is whether a wealth constraint is relevant for an individual endowed with *deterministic* labor income. As mentioned, for the unrestricted problem the investor optimally sells his future labor income in the market and acts as if he had an enlarged initial wealth. Consequently, the investor's terminal wealth becomes positive, but the investor's wealth may become negative at some point in time before terminal time. If labor income really is deterministic one could argue that this is perfectly acceptable. The motivation of generally introducing an American capital guarantee, and specifically to analyze the case of a restriction to borrow against future labor income, is however manifold. Consider for example the lending business. In reality you can borrow a great amount of money from the bank to buy a house if you are the receiver of a (possible stochastic) labor income. The reason is that you invest your loan in the house and thereby maintain a positive wealth. Of course the value of the house could drop, but since the loan comes with an amortization plan[2] the wealth stays positive (with a high probability). You can also get a mortgage loan from the bank to finance consumption. On the other hand it is almost impossible to get a loan from the bank to finance pure consumption if you can not provide security for the loan[3]. Why is that? Of course this is partly because the bank consider your labor income stochastic (e.g. you might lose your job), but there are many other reasons. Most of them are connected to obvious reasons as moral hazards, adverse selection and administration problems. Or one could say that the bank just does not want to take on that risk. This paper is motivated by the real life situation *many* individuals face. As an individual you do not consider your labor income as stochastic. At least you do not act as if you do. This is partly because it is very hard to model labor income since the list of relevant stochastic variables to include in the modulation should be very long. The modeling should, among other things, include future political decisions, possible world wars, personal accidents (you may become disable), your mortality (you may die) and the demand of tomatoes (if you make a living of selling tomatoes). In addition many people actually have a very stabile and secure labor income, and thereby have a very strong belief about how their life long labor income would be. If this was not true, why are more people not buying unemployment insurance? Summing up, the question many individuals face is: You consider your labor income deterministic and you act responsibly as a good citizen. Unfortunately the bank does not agree with you and consequently the bank

[2]In addition most banks require the investor himself to provide (say) 5 percent funding.

[3]In some countries this is actually possible, but since expenses are incredible high we assume that the investor does not consider this as an opportunity.

is only willing to issue a loan if you invest in such a way that they do not stand the chance of ending up with a loss (your total assets has to be positive at all times). Given this situation how should you optimally consume and invest? This is one interpretation of the question of consideration in this article.

Furthermore, the problem studied in this article is, putting consumption equal to zero, highly relevant for many financial institutions. The most obvious example is pension companies which are forced by law (by the Financial Supervisory Authority (FSA)) to fulfill certain capital requirements at all times. In such a framework labor income has the interpretation as pension contributions and the American capital guarantee, introduced by the FSA, puts restrictions on the pension company's actions.

The paper by Lakner and Nygren (2006) consider a similar problem. They maximize expected utility from consumption and terminal wealth when both the consumption rate and terminal wealth must not fall below a given minimum consumption level and a minimum wealth level, respectively. This is handled by dividing initial wealth into two parts, x_1 and x_2, and then solving the constrained pure consumption problem with initial wealth x_1 and the constrained pure portfolio problem with initial wealth x_2. The superposition of the actions for the two constrained subproblems is then shown to be the optimal strategy provided that x_1 and x_2 are chosen such that the "marginal expected utilities" from the two constrained subproblems are identical. In this article we do not consider a minimum level of consumption constraint since we do not want the consequence of the introduction of an American capital guarantee to be blurred by other constraints. One could, using the results obtained in this article, apply the approach of Lakner and Nygren (2006) to solve such a problem.

Using different approaches, the restriction to borrow against future labor income has been analyzed in different settings with spanned stochastic income rate. He and Pagès (1993) use duality theory to show the existence of an optimal strategy and find that the presence of a liquidity constraint has a smoothing effect on the optimal consumption. El-Karoui and Jeanblanc (1998) introduce the concept of super strategies to show that the fair-value, defined as the minimal investment required to replicate a consumption investment strategy with a super strategy, consists of investing a part of the initial wealth in a free portfolio and to use the remaining part to buy an American put option on this underlying asset. A setting with un-spanned stochastic income is analyzed by Duffie and Zariphopoulou (1993), Duffie et al. (1997), Koo (1998) and Munk (2000) who also gives a very complete overview of the field as a whole.

The paper is organized as follows: In Section 2.2 we present the economic model, the concept of admissible strategies and other preliminaries. In Section 2.3 we solve the unrestricted problem using the martingale technique. As an inspirational example we solve, in Section 2.4, the optimal consumption and investment decision problem when the investor is restricted to fulfill a binding terminal capital guarantee (any terminal guarantee strictly greater than zero), also known as a European capital guarantee. Using the building blocks derived in Section 2.3 and 2.4, and the theory of American put options in a Black-Scholes market, we construct in the main section, Section 2.5, an admissible put option based portfolio insurance strategy. We show that this strategy is optimal for the consumption and investment decision problem where the investor is restricted to fulfill an American capital guarantee. Finally, in Subsection 2.5.3 graphical illustrations of the optimal restricted strategy together with a comparison with the optimal unrestricted strategy are presented for the special case of a restriction to borrow against future labor income.

2.2 Setup

Consider a frictionless Black-Scholes market consisting of a bank account, B, with risk-free short rate, r, and a risky stock, S, with dynamics given by

$$\begin{aligned} dB(t) &= rB(t)dt,\ B(0) = 1, \\ dS(t) &= \alpha S(t)dt + \sigma S(t)dW(t),\ S(0) = s_0 > 0. \end{aligned}$$

Here $\alpha, \sigma, r > 0$ are constants and it is assumed that $\alpha > r$. The process W is a standard Brownian motion on an abstract probability space $(\Omega, \mathcal{F}, P)$ equipped with the filtration $\mathbb{F}^W = \left(\mathcal{F}^W(t)\right)_{t\in[0,T]}$ given by the P-augmentation of the filtration $\sigma\{W(s);\, 0 \leq s \leq t\}$, $\forall t \in [0, T]$.

We consider an investor endowed with a continuous deterministic labor income of rate ℓ, an initial amount of money x_0 and a time horizon of interest $[0, T]$, $T > 0$. Over the time interval, $[0, T]$, the investor has the opportunity to continuously consume a fraction of his wealth and to invest his wealth in the Black-Scholes market. More specific, let c denote the rate of consumption and let π denote the fraction of wealth to be invested in the risky asset, S. For a given strategy, (c, π), the wealth process of the investor, $(X(t))_{t\in[0,T]}$, is given by the dynamics

$$\begin{aligned} dX(t) &= [(r + \pi(t)(\alpha - r))X(t) + \ell(t) - c(t)]dt + \sigma\pi(t)X(t)dW(t),\ t \in [0, T), \\ X(0) &= x_0 \geq 0. \end{aligned} \tag{2.1}$$

We assume the consumption rate $c(t)$, $t \in [0, T]$, to be a non-negative, $\mathcal{F}^W(t)$-adapted process with $\int_0^T c(t)dt < \infty$ a.s., and the investment strategy $\pi(t)$, $t \in [0, T]$, to be an $\mathcal{F}^W(t)$-adapted process such that (2.1) has a unique solution $X(t)$, $t \in [0, T]$, satisfying $\int_0^T (\pi(t)X(t))^2 dt < \infty$ a.s.. We have, $\forall t \in [0, T]$, the representation

$$X(t) = x_0 e^{rt} + \int_0^t e^{r(t-s)}[\pi(s)X(s)(\alpha - r) + \ell(s) - c(s)]ds + \int_0^t e^{r(t-s)}\sigma\pi(s)X(s)dW(s).$$

It is well known that in the Black-Scholes market the equivalent martingale measure, Q, is unique and given by the Radon-Nikodym derivative

$$\frac{dQ(t)}{dP(t)} = \Lambda(t) := \exp\left(-\left(\frac{\alpha - r}{\sigma}\right)W(t) - \frac{1}{2}\left(\frac{\alpha - r}{\sigma}\right)^2 t\right),\ t \in [0, T],$$

and that the process W^Q given by $W^Q(t) = W(t) + \frac{\alpha - r}{\sigma}t$, $t \in [0, T]$, is a standard Brownian motion under Q. We get that the investor's wealth can be represented, $\forall t \in [0, T]$, by

$$X(t) = x_0 e^{rt} + \int_0^t e^{r(t-s)}(\ell(s) - c(s))ds + \int_0^t e^{r(t-s)}\sigma\pi(s)X(s)dW^Q(s). \tag{2.2}$$

Definition 2.2.1. *Define the set of admissible strategies, denoted by $\mathcal{A}$, as the consumption and investment strategies for which the corresponding wealth process given by (2.2) is well-defined,*

$$X(t) + g(t) \geq 0,\ \forall t \in [0, T], \tag{2.3}$$

where g is the time t financial value of future labor income defined by $g(t) := \int_t^T e^{-r(s-t)}\ell(s)ds$, and

$$E^Q\left[\int_0^T e^{-rs}\sigma\pi(s)X(s)dW^Q(s)\right] = 0. \tag{2.4}$$

The technical condition (2.4) is equivalent to the condition that under Q the process

$$\int_0^t e^{-rs}\sigma\pi(s)X(s)dW^Q(s),\ t\in[0,T],$$

is a martingale (in general it it only a local martingale, and also a supermartingale if (2.3) is fulfilled, see e.g. Karatzas and Shreve (1998)). From this we conclude that (2.4) insures that $(c,\pi)\in\mathcal{A}$ if and only if $X(T)\geq 0$ and, $\forall t\in[0,T]$,

$$X(t)+g(t)=E^Q\left[\int_t^T e^{-r(s-t)}c(s)ds+e^{-r(T-t)}X(T)\middle|\,\mathcal{F}^W(t)\right].$$

At time zero this means the strategies have to fulfill the *budget constraint*

$$x_0+g(0)=E^Q\left[\int_0^T e^{-rt}c(t)dt+e^{-rT}X(T)\right]. \tag{2.5}$$

For latter use we state the following remark.

Remark 2.2.1. Define

$$Z(t):=\int_0^t e^{-rs}(c(s)-\ell(s))ds+e^{-rt}X(t),\ t\in[0,T]. \tag{2.6}$$

By (2.2) we have that condition (2.4) is fulfilled if and only if Z is a martingale under Q. The natural interpretation is that under Q the discounted wealth plus discounted consumption excess of labor income should be a martingale. It is useful to note that if Z is a martingale under Q the dynamics of X can be represented in the form

$$dX(t)=[(rX(t)+\ell(t)-c(t)]dt+\phi(t)dW^Q(t),\ t\in(0,T], \tag{2.7}$$

for some $\mathcal{F}^W(t)$-adapted process ϕ. Moreover, if the dynamics of X can be represented in the form given by (2.7) for some $\mathcal{F}^W(t)$-adapted process ϕ satisfying $\phi(t)\in\pounds^2$, $\forall t\in[0,T]$, then Z is a martingale under Q.

Finally, by condition (2.3) we allow the wealth process to become negative, as long as it does not exceed (in absolute value) the financial value of future labor income. Note that condition (2.3) puts a lower boundary on the wealth process and therefore rules out doubling strategies (see e.g. Karatzas and Shreve (1998)).

2.3 The unrestricted control problem

For the rest of this paper we assume that the investor's preferences can be stated in terms of a constant relative risk aversion (CRRA) utility function $u:[0,\infty)\to[-\infty,\infty)$ in the form

$$u(x)=\begin{cases}\frac{x^\gamma}{\gamma}, & \text{if } x>0\\ \lim_{x\searrow 0}\frac{x^\gamma}{\gamma}, & \text{if } x=0,\end{cases}$$

for some $\gamma\in(-\infty,1)\backslash\{0\}$. In the unrestricted control problem, known as Merton's problem (Merton (1969) and Merton (1971)), the investor seeks to find the strategy fulfilling the supremum

$$\sup_{(c,\pi)\in\mathcal{A}'} E\left[\int_0^T e^{-\int_0^t\beta(s)ds}u(c(t))dt+e^{-\int_0^T\beta(s)ds}u(X(T))\right], \tag{2.8}$$

where $X(0) = x_0$ and β is a deterministic function representing the investor's time preferences. The supremum in (2.8) is taken over the subset of the admissible strategies, referred to as the *feasible* strategies, defined by

$$\mathcal{A}' := \left\{ (c, \pi) \in \mathcal{A} \;\middle|\; E\left[\int_0^T e^{-\int_0^t \beta(s)ds} \min(0, u(c(t)))dt + e^{-\int_0^T \beta(s)ds} \min(0, u(X(T))) \right] > -\infty \right\}, \tag{2.9}$$

i.e. it is allowed to draw an infinite utility from a strategy $(c, \pi) \in \mathcal{A}'$, but only if the expectation over the negative parts of the utility function is finite. Clearly, for $\gamma \in (0, 1)$ we have that A and A' coincide.

The unrestricted control problem (2.8) can be solved using either the conventional Hamilton-Jacobi-Bellman (HJB) method (see e.g. Fleming and Richel (1975)) or the newer martingale method (see Cvitanic et al. (1987), Cox and Huang (1989) and Cox and Huang (1991)). Despite the fact that this is most easily done using the HJB-method we now sketch the solution to (2.8) in terms of the martingale method. We do so because the solution to the corresponding American capital guarantee version of (2.8), presented in Section 2.5, is based on terms derived from the martingale method.

Define the inverse of the derivative of u as the function $I : (0, \infty] \to [0, \infty)$ and define the adjusted state price deflator

$$H(t) := \Lambda(t) e^{\int_0^t (\beta(s) - r)ds}.$$

Since I is continuous and decreasing and maps $(0, \infty]$ onto $[0, \infty)$ we get the useful result that there exist a unique constant $\xi^* > 0$ such that (see. e.g. Karatzas and Shreve (1998))

$$E^Q \left[\int_0^T e^{-rt} I(\xi^* H(t))dt + e^{-rT} I(\xi^* H(T)) \right] = x_0 + g(0). \tag{2.10}$$

Finally, note that from the concavity of u and $u'(I(z)) = z$ we get

$$u(x) \le u(I(z)) - z(I(z) - x), \; \forall x \ge 0, z > 0. \tag{2.11}$$

Now, take an arbitrary strategy $(c, \pi) \in \mathcal{A}'$ with corresponding wealth process $(X(t))_{t \in [0,T]}$ as given. Using (2.11), the budget constaint (2.5) and finally (2.10) we get that

$$\begin{aligned} & E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + e^{-\int_0^T \beta(s)ds} u(X(T)) \right] \\ \le & E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u(I(\xi^* H(t)))dt + e^{-\int_0^T \beta(s)ds} u(I(\xi^* H(T))) \right]. \end{aligned}$$

Since (c, π) was arbitrarily chosen we conclude that the *candidate* optimal strategy (c^*, π^*) is given by

$$c^*(t) = I(\xi^* H(t)), \tag{2.12}$$

$$X^*(T) = I(\xi^* H(T)), \tag{2.13}$$

where $X^*(T)$ is the corresponding terminal wealth obtained from using the strategy (c^*, π^*). By use of the martingale representation theorem (see e.g. Karatzas and Shreve (1991)) we obtain (after some rather cumbersome calculations)

$$c^*(t) = \frac{X(t) + g(t)}{f(t)}, \tag{2.14}$$

$$\pi^*(t) X^*(t) = \frac{1}{\gamma} \frac{\alpha - r}{\sigma^2} (X^*(t) + g(t)), \tag{2.15}$$

where

$$f(t) = \int_t^T e^{-\int_t^s [\widetilde{r}(y)+\beta(y)]dy} ds + e^{-\int_t^T [\widetilde{r}(y)+\beta(y)]dy},$$

with $\widetilde{r}(y) = \frac{\gamma}{\gamma}\beta(y) - \frac{\gamma}{\gamma}r - \frac{1}{2}\frac{\gamma}{(1-\gamma)^2}\left(\frac{\alpha-r}{\sigma}\right)^2$. Inserting the candidate optimal strategy (c^*, π^*) in (2.1) we observe that $X^*(t) + g(t),\ t \in [0,T]$, referred to as the *total wealth*, follows a time inhomogeneous geometric Brownian motion[4]. More specifically we get, $\forall t \in [0,T]$, that

$$X^*(t) + g(t) = (x_0 + g(0)) \exp\left\{ \left(r + \frac{\frac{1}{2} - \gamma}{(1-\gamma)^2} \left(\frac{\alpha - r}{\sigma} \right)^2 \right) t - \int_0^t \frac{1}{f(s)} ds + \frac{1}{\gamma} \frac{\alpha - r}{\sigma} W(t) \right\}. \tag{2.16}$$

In particular, since f is bounded away from zero, $\forall t \in [0,T]$, we have that $X^*(t),\ t \in [0,T]$, is well-defined, and clearly (2.3) is fulfilled. From (2.14) we also conclude that the candidate optimal consumption rate is non-negative. Moreover, since $X^*(t) + g(t)$ is lognormally distributed, $\forall t \in [0,T]$, one can easily show that

$$E^Q\left[\int_0^T (\sigma\pi^*(t)X^*(t))^2 dt\right] < \infty, \tag{2.17}$$

which ensures that (2.4) is fulfilled. In fact, by use of the martingale representation theorem we have that the candidate optimal investment strategy can be written as

$$\pi^*(t)X^*(t) = \frac{e^{rt}\psi(t)}{\sigma},\ t \in [0,T],$$

where ψ is the unique integrand from the martingale representation theorem, i.e.

$$\int_0^t \psi(s)dW^Q(s),\ t \in [0,T],$$

is a martingale under Q. From this it follows immediately that (2.4) is fulfilled. Thus, the candidate optimal strategy (c^*, π^*) is admissible. Finally, since x_0 and $g(0)$ cannot both be zero we get, from (2.14) and (2.16), that $c^*(t) > 0,\ \forall t \in [0,T]$, and $X^*(T) > 0$, and we note that the additional condition given in (2.9) is fulfilled, i.e. (c^*, π^*) is feasible. We conclude that (c^*, π^*), given by (2.14) and (2.15), is the optimal strategy for the unrestricted problem (2.8).

Remark 2.3.1. It is well known that the control problem (2.8) with and without deterministic labor income coincide in the sense that an investor with an initial wealth equal to $x_0 + g(0)$, but no labor income, optimally consumes and invests the same amount of money as the investor we consider. Define $Y(t) := X(t) + g(t)$ and the obvious notation $Y^*(t) := X^*(t) + g(t), \forall t \in [0,T]$. In Section 2.4 and 2.5 we exploit the fact that an investor with wealth process $(Y(t))_{t\in[0,T]}$, i.e. no labor income, starting with an initial amount of money $y_0 := x_0 + g(0)$, can replicate the wealth process $(Y^*(t))_{t\in[0,T]}$, by following the consumption and investment strategy given, $\forall t \in [0,T]$, by

$$c^*(t) = \frac{Y^*(t)}{f(t)}, \tag{2.18}$$

$$\pi_y^* = \frac{1}{\gamma}\frac{\alpha - r}{\sigma^2}. \tag{2.19}$$

From (2.16) is follows that $Y^*(t), \forall t \in [0,T]$, is proportional to the initial amount of money y_0. We see that starting out with the initial amount of money $\lambda y_0, \lambda \in \mathbb{R}_+$, and no labor income, one can replicate $\lambda Y^*(t), \forall t \in [0,T]$, by following the strategy $(\lambda c^*, \pi_y^*)$.

[4] In particular $X^*(t) + g(t),\ t \in [0,T]$, is a geometric Brownian motion if β is constant.

Note that the optimal strategy (c^*, π^*) guarantees that $X^*(t) > -g(t), \forall t \in [0, T]$. This is in some sense a direct consequence of CRRA utility. Loosely speaking, the marginal increase in utility is going to infinity as terminal wealth is going to zero, i.e. a terminal wealth equal to zero cannot be optimal. To be able to avoid this in all scenarios the investor simply ensures that $X^*(t) > -g(t), \forall t \in [0, T]$. As mentioned in Section 2.1, one should also notice that the formulation of problem (2.8) assumes that the bank allows the investor to have a negative wealth at any given time point $t \in (0, T)$. This will in reality, due to reasons discussed in the introduction, likely not be the case. Furthermore, most investors tend to have some sort of preference or demand for a capital guarantee. The most obvious example is that pension companies often are required by authorities (FSA) to fulfill a certain American capital guarantee. Jensen and Sørensen (2001) quantify that an individual facing a (European) mandatory minimum interest rate guarantee, as an integrated part of the pension contract, is forced to invest in an embedded put option. They argue that the investor may face a constraint on the feasible set of portfolio choices which may induce a significant utility loss.

Including an American capital guarantee to problem (2.8) is the main problem of interest in this article. Inspired by El-Karoui et al. (2005) we, as an inspirational example, start out by solving the easier case including a binding European capital guarantee. This problem was first solved by Teplá (2001).

2.4 The European capital guarantee control problem

Consider the problem

$$\sup_{(c,\pi)\in\mathcal{A}'} E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + e^{-\int_0^T \beta(s)ds} u(X(T))\right], \tag{2.20}$$

under the capital constraint $X(T) \geq K$. In Section 2.3 we saw that $K \leq 0$ is a non-binding constraint. Note that in order for the European capital guarantee problem to be solvable and non-trivial[5] we must demand that

$$K < e^{rT}(x_0 + g(0)). \tag{2.21}$$

Proposition 2.4.1 presents the optimal strategy for problem (2.20). The strategy is based on a so-called option based portfolio insurance (OBPI) written on a 'budget-adjusted' version of the optimal portfolio derived for the unrestricted problem. First, we need some notation: Denote by $P^e_X(t, T, K)$ the time t value of a European put option with strike price K and maturity T written on a portfolio $(X(s))_{s\in[t,T]}$. As in Section 2.3 $X^*(t)$, $c^*(t)$ and $\pi^*(t)$, $t \in [0, T]$, denote the optimal wealth, the optimal consumption strategy and the optimal investment strategy for the unrestricted problem (2.8), respectively.

Proposition 2.4.1. *Consider the strategy* $(\widehat{c}, \lambda Y^*)$ *where,* $\forall t \in [0, T]$*,*

$$\widehat{c}(t) := \frac{\lambda Y^*(t)}{f(t)} = \lambda c^*(t),$$
$$Y^*(t) := X^*(t) + g(t),$$

combined with a position in a European put option with strike price K and maturity T written on the portfolio $(\lambda Y^*(s))_{s\in[t,T]}$, where λ is determined by the budget constraint

$$\widehat{X}^{(\lambda)}(0) := \lambda(x_0 + g(0)) + P^e_{\lambda Y^*}(0, T, K) - g(0) = x_0. \tag{2.22}$$

[5] In the case of equality in (2.21) the investor has no choice but to invest all his wealth including labor income in the risk-free short rate to be sure to fulfil the capital guarantee.

Observing that $Y^*(T) = X^*(T)$, it follows that the terminal value of the corresponding option based portfolio insurance becomes

$$\widehat{X}^{(\lambda)}(T) := \lambda X^*(T) + (K - \lambda X^*(T))^+ = \max\left(\lambda X^*(T), K\right), \tag{2.23}$$

i.e. the European capital guarantee is fulfilled. We have that

1. *The strategy is affordable, i.e. there exist a unique $\lambda \in (0,1)$ such that the budget constraint (2.22) is fulfilled.*
2. *The strategy is optimal for the European capital guarantee control problem given by (2.20).*

Note that, clearly, the strategy in Proposition 2.4.1 is feasible since the unrestricted strategy (c^*, π^*) given in Section 2.3 is feasible.

Remark 2.4.1. To clarify, following the strategy $\left(\widehat{c}, \widehat{X}^{(\lambda)}\right)$ given by Proposition 2.4.1 means that the investor should:

1. Take a loan of size $g(0)$ and reserve the labor income to pay back the loan over the time interval $[0, T]$.
2. Reserve the initial amount of money $\lambda(x_0 + g(0))$ to follow the consumption and investment strategy $(\widehat{c}, \pi)$ given by

$$\widehat{c}(t) := \lambda c^*(t),$$
$$\pi := \pi_y^*.$$

   Doing this the investor replicates the portfolio $\lambda Y^*(t), \forall t \in [0,T]$, i.e. the investor replicates the terminal value $\lambda Y^*(T) = \lambda X^*(T)$ (see Remark 2.3.1).
3. Use the remaining initial amount of money $(1-\lambda)(x_0 + g(0))$ to buy a European put option with strike price K and time to maturity T written on the portfolio $(\lambda Y^*(t))_{t \in [0,T]}$.

The European put option is likely not to be sold in the market, but since the market is complete and frictionless such options can be replicated dynamically by a Delta-hedge, i.e. the optimal investment strategy $\widehat{\pi}(t),\ t \in [0,T]$, is given by

$$\widehat{\pi}(t)\widehat{X}^{(\lambda)}(t) = \left(1 + \frac{\partial}{\partial y} P^e_{\lambda Y^*}(t, T, K)\right) \pi_y^* \lambda Y^*(t).$$

Since the European put option has an opposite exposure to changes in the underlying portfolio $\lambda Y^*(t),\ t \in [0,T]$, we note that the total amount of money to invest in the risky asset, S, becomes smaller when we introduce a European capital guarantee to the control problem (2.8).

Proof of affordability. First note that $\widehat{X}^{(\lambda)}(T) = \max\left(\lambda X^*(T), K\right)$ is non-decreasing in λ with values in $[K, \infty)$. For $\lambda_1 < \lambda_2$ we get

$$0 \le \widehat{X}^{(\lambda_2)}(T) - \widehat{X}^{(\lambda_1)}(T) \le (\lambda_2 - \lambda_1) X^*(T).$$

Discounting and taking the expectation under Q we get by use of (2.5) that

$$\begin{aligned} 0 \le \widehat{X}^{(\lambda_2)}(0) - \widehat{X}^{(\lambda_1)}(0) &\le (\lambda_2 - \lambda_1) E^Q\left[e^{-rT} X^*(T)\right] \\ &= (\lambda_2 - \lambda_1)\left(x_0 + g(0) - E^Q\left[\int_0^T e^{-rt} c^*(t) dt\right]\right) \\ &< (\lambda_2 - \lambda_1)(x_0 + g(0)). \end{aligned}$$

We conclude that $\widehat{X}^{(\lambda)}(0)$ is an increasing and Lipschitz function with respect to λ (Lipschitz constant equal to $x_0 + g(0)$). No-arbitrage gives us $\widehat{X}^{(\lambda)}(0) \in \left[e^{-rT}K - g(0), \infty\right)$. Since $K < e^{rT}(x_0 + g(0))$ and $\widehat{X}^{(1)}(0) > x_0$ we conclude that there exist a unique $\lambda \in (0,1)$ fulfilling (2.22). $\square$

Before we prove the optimality of the strategy in Proposition (2.4.1) we present the following well-known result (see e.g. Karatzas and Shreve (1998)):

Lemma 2.4.2. *Since (c^*, X^*) is the optimal strategy for the unrestricted problem (2.8) we have that*

$$\frac{\partial}{\partial \epsilon} E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u\Big(\epsilon c^*(t) + (1-\epsilon)c(t)\Big)dt + e^{-\int_0^T \beta(s)ds} u\Big(\epsilon X^*(T) + (1-\epsilon)X(T)\Big)\right]_{|\epsilon=1} = 0$$

$$\Longrightarrow E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u'(c^*(t))(c^*(t) - c(t))dt + e^{-\int_0^T \beta(s)ds} u'(X^*(T))(X^*(T) - X(T))\right] = 0,$$

for any feasible strategy (c, X) with $X(0) = x_0$.

We are now ready to prove that the put option based strategy defined in Proposition 2.4.1 is optimal among all feasible strategies which satisfy the European capital guarantee.

Proof of optimality. Let $(c, \pi) \in \mathcal{A}'$, with corresponding wealth process $(X(t))_{t\in[0,T]}$, be any given feasible strategy satisfying $X(0) = x_0$ and $X(T) \geq K$. Using the concavity of u we get that

$$\begin{aligned}
&\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + e^{-\int_0^T \beta(s)ds} u(X(T)) \\
&- \left(\int_0^T e^{-\int_0^t \beta(s)ds} u\left(\widehat{c}(t)\right)dt + e^{-\int_0^T \beta(s)ds} u\left(\widehat{X}^{(\lambda)}(T)\right)\right) \\
=\; &\int_0^T e^{-\int_0^t \beta(s)ds} \left(u(c(t)) - u\left(\widehat{c}(t)\right)\right)dt + e^{-\int_0^T \beta(s)ds}\left(u(X(T)) - u\left(\widehat{X}^{(\lambda)}(T)\right)\right) \\
\leq\; &\int_0^T e^{-\int_0^t \beta(s)ds} u'\left(\widehat{c}(t)\right)\left(c(t) - \widehat{c}(t)\right)dt + e^{-\int_0^T \beta(s)ds} u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
:=\; &(*).
\end{aligned}$$

Since (c, π) was arbitrarily chosen we end the proof by showing that $E[(*)] \leq 0$.

A special feature of CRRA utility is that $u'(xy) = u'(x)u'(y)$. Using this feature we get

$$u'\left(\widehat{c}(t)\right)\left(c(t) - \widehat{c}(t)\right) = u'(\lambda)u'(c^*(t))\left(c(t) - c^*(t)\right) + u'(\lambda)u'(c^*(t))\left(c^*(t) - \widehat{c}(t)\right). \tag{2.24}$$

By use of (2.23) and noting that u' is a decreasing function we obtain

$$\begin{aligned}
&u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&= \left(u'(\lambda)u'(X^*(T)) \wedge u'(K)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&= u'(\lambda)u'(X^*(T))\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) - \left(u'(\lambda)u'(X^*(T)) - u'(K)\right)^+ \left(X(T) - K\right),
\end{aligned}$$

where we in the last equality have used that $\widehat{X}^{(\lambda)}(T) = K$ on the set $\{(T, \omega) : u'(\lambda)u'(X^*(T)) \geq u'(K)\}$. Since by assumption $X(T) \geq K$ this reduces to

$$\begin{aligned}
&u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&\leq u'(\lambda)u'(X^*(T))\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&= u'(\lambda)u'(X^*(T))\left(X(T) - X^*(T)\right) + u'(\lambda)u'(X^*(T))\left(X^*(T) - \widehat{X}^{(\lambda)}(T)\right).
\end{aligned} \tag{2.25}$$

Finally, combining (2.24) and (2.25) and using Lemma 2.4.2 twice give us $E[(*)] \leq 0$. □

2.5 The American capital guarantee control problem

Consider the problem

$$\sup_{(c,\pi)\in\mathcal{A}'} E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + e^{-\int_0^T \beta(s)ds} u(X(T))\right], \tag{2.26}$$

under the capital constraint $X(t) \geq K(t)$, $\forall t \in [0,T]$, where K is a deterministic function of time. In Section 2.3 we saw that $K(t) \leq -g(t)$, $\forall t \in [0,T]$, is a non-binding constraint. Note that $K = 0$ corresponds to the case where the investor is restricted from borrowing against future labor income. In order for the American capital guarantee problem to be solvable and non-trivial[6] we must demand

$$K(t) < e^{rt}\left(x_0 + \int_0^t e^{-rs}\ell(s)ds\right), \ \forall t \in (0,T]. \tag{2.27}$$

Remark 2.5.2 presents the case without terminal consumption. Equivalent to the unrestricted problem we get that the optimal strategies in the cases with and without terminal consumption are very closely related.

2.5.1 An admissible American put option based portfolio

In Subsection 2.5.2 we present the optimal strategy for the problem (2.26). In turns out that the optimal strategy is very similar to the optimal strategy for the European capital guarantee problem presented in Proposition 2.4.1. Again, the optimal strategy consists of a position in a budget-adjusted version of the optimal portfolio derived for the unrestricted problem plus a put option written on that portfolio. This time though, to fulfil the American capital guarantee the OBPI strategy must consist of an American put option. However, the introduction of an American put option is not uncomplicated. In turns out that the part of the strategy which consist of the American put option is not admissible by default. Loosely speaking, to make the strategy admissible we have to continuously adjust the budget whenever the constraint is active. In the following this will be made clear.

Still, we denote by $X^*(t)$, $c^*(t)$ and $\pi^*(t)$, $t \in [0,T]$, the optimal wealth, the optimal consumption strategy and the optimal investment strategy for the unrestricted problem (2.8), respectively. Observe that the portfolio $Y^*(t) = X^*(t) + g(t),\ t \in [0,T]$, has dynamics

$$\begin{aligned} dY^*(t) &= \left(r + \frac{1}{\gamma}\left(\frac{\alpha - r}{\sigma}\right)^2 - \frac{1}{f(t)}\right)Y^*(t)dt + \frac{1}{\gamma}\frac{\alpha - r}{\sigma}Y^*(t)dW(t) \\ &= \left(r - \frac{1}{f(t)}\right)Y^*(t)dt + \frac{1}{\gamma}\frac{\alpha - r}{\sigma}Y^*(t)dW^Q(t),\ t \in [0,T), \\ Y^*(0) &= y_0, \end{aligned} \tag{2.28}$$

where $y_0 = x_0 + g(0)$. Denote by $P^a_y(t,T,K+g)$ the time t value of an American put option with strike price $K(s)+g(s), \forall s \in [t,T]$, and maturity T written on a portfolio Y, where $(Y(s))_{s\in[t,T]}$ is the solution to (2.28) which *equals y at time t*. By definition the price of such a put option is given by

$$P^a_y(t,T,K+g) := \sup_{\tau\in\mathcal{T}_{t,T}} E^Q\left[e^{-r(\tau-t)}(K(\tau)+g(\tau)-Y(\tau))^+ \,\middle|\, Y(t) = y\right], \ \forall t \in [0,T],$$

[6] In the case of equality in (2.27) the investor has no choice but to invest all his wealth including labor income in the risk-free short rate to be sure to fulfil the capital guarantee.

where $\mathcal{T}_{t,T}$ is the set of stopping times taking values in the interval $[t,T]$.

By analogy with the European case we introduce the American put option based portfolio

$$X^{(\lambda)}(t) := \lambda Y^*(t) + P^a_{\lambda Y^*}(t,T,K+g) - g(t),\ \forall t \in [0,T], \tag{2.29}$$

where $\lambda \in (0,1)$ is determined by the budget constraint

$$X^{(\lambda)}(0) = \lambda(x_0 + g(0)) + P^a_{\lambda y_0}(0,T,K+g) - g(0) = x_0. \tag{2.30}$$

By definition of an American put option $P^a_{\lambda Y^*}(t,T,K+g) \geq (K(t)+g(t)-\lambda Y^*(t))^+,\ \forall t \in [0,T]$. We conclude that $X^{(\lambda)}$ fulfills the American capital guarantee since, $\forall t \in [0,T]$, we have

$$\begin{aligned} X^{(\lambda)}(t) &= \lambda Y^*(t) + P^a_{\lambda Y^*}(t,T,K+g) - g(t) \\ &\geq \lambda Y^*(t) + (K(t) + g(t) - \lambda Y^*(t))^+ - g(t) \\ &\geq K(t). \end{aligned} \tag{2.31}$$

Now, recall some basis properties for American put options in a Black-Scholes market (see e.g. Karatzas and Shreve (1998)):

$$\begin{aligned} P^a_y(t,T,K+g) &= K(t) + g(t) - y, && \forall (t,y) \in \mathcal{C}^c, \\ \mathcal{L}P^a_y(t,T,K+g) &= rP^a_y(t,T,K+g), && \forall (t,y) \in \mathcal{C}, \\ \frac{\partial}{\partial y} P^a_y(t,T,K+g) &= -1, && \forall (t,y) \in \mathcal{C}^c, \end{aligned}$$

where (see (2.28) and (2.19))

$$\mathcal{L} := \frac{\partial}{\partial t} + \left(r - \frac{1}{f(t)}\right) y \frac{\partial}{\partial y} + \frac{1}{2}\sigma^2 (\pi^*_y)^2 y^2 \frac{\partial^2}{\partial y^2},$$

and

$$\mathcal{C} := \{(t,y) : P^a_y(t,T,K+g) > (K(t) + g(t) - y)^+\},$$

defines the continuation region. By $\mathcal{C}^c$ we mean the complementary of $\mathcal{C}$, i.e. the stopping region. The continuation region can be described via the exercise boundary

$$b(t) = \sup\left\{y : P^a_y(t,T,K+g) = (K(t) + g(t) - y)^+\right\}.$$

Note, that b is a deterministic function of time. We get

$$\mathcal{C} = \{(t,y) : y > b(t)\}.$$

Introducing the function A by

$$A(t,y) := y + P^a_y(t,T,K+g) - g(t),$$

we can write (2.29) as

$$X^{(\lambda)}(t) = A(t, \lambda Y^*(t)),\ \forall t \in [0,T].$$

From the properties of $P^a_y(t,T,K+g)$ we deduce that

$$\begin{aligned} A(t,y) &= K(t), && \forall (t,y) \in \mathcal{C}^c, \\ \mathcal{L}A(t,y) &= \left(r - \frac{1}{f(t)}\right) y + rP^a_y(t,T,K+g) - (-\ell(t) + rg(t)) \\ &= rA(t,y) + \ell(t) - \frac{y}{f(t)}, && \forall (t,y) \in \mathcal{C}, && (2.32) \\ \mathcal{L}A(t,y) &= \frac{\partial}{\partial t} K(t), && \forall (t,y) \in \mathcal{C}^c, && (2.33) \\ \frac{\partial}{\partial y} A(t,y) &= 0. && \forall (t,y) \in \mathcal{C}^c. && (2.34) \end{aligned}$$

Since $Y^*(t)$ is linear in its initial value, $\forall t \in [0,T]$, we have that $\lambda Y^*(t)$ has the same dynamics as $Y^*(t)$, $\forall t \in [0,T]$. By use of (2.32), (2.33) and (2.18) we get that[7]

$$\begin{aligned} dA(t,\lambda Y^*(t)) &= \frac{\partial}{\partial y}A(t,\lambda Y^*(t))\sigma\pi_y^*\lambda Y^*(t)dW^Q(t) + \mathcal{L}A(t,\lambda Y^*(t))dt \\ &= \frac{\partial}{\partial y}A(t,\lambda Y^*(t))\sigma\pi_y^*\lambda Y^*(t)dW^Q(t) \\ &\quad + [rA(t,\lambda Y^*(t)) + \ell(t) - \lambda c^*(t)]1_{(\lambda Y^*(t)>b(t))}dt + \frac{\partial}{\partial t}K(t)1_{(\lambda Y^*(t)\leq b(t))}dt. \end{aligned} \tag{2.35}$$

From the condition (2.7) we see that the strategy $\left(\lambda c^*, X^{(\lambda)}\right)$ is *admissible up to the hitting time of the exercise boundary* b. If we allow λ to be a function of time, we can choose $(\lambda(t), t \in [0,T])$ to be constant whenever $\lambda(t)Y^*(t) > b(t)$ and increasing at the boundary such that $\lambda(t)Y^*(t) \geq b(t), \forall t \in [0,T]$. More precise, define

$$\lambda(t) = \lambda_0 \vee \sup_{s\leq t}\left(\frac{b(s)}{Y^*(s)}\right), \ \forall t \in [0,T], \tag{2.36}$$

where λ_0 is determined by the budget constraint (2.30).

Proposition 2.5.1. *The strategy* $\left(\lambda c^*, X^{(\lambda)}\right)$ *where,* $\forall t \in [0,T]$,

$$X^{(\lambda)}(t) := \lambda(t)Y^*(t) + P^a_{\lambda Y^*}(t,T,K) - g(t)$$

with $\lambda(t), t \in [0,T]$, *being the function defined by (2.36) and (2.30), is admissible*[8] .

Proof. Using Itô's formula, (2.35) and that λ increases only at the boundary, we get that

$$\begin{aligned} dA(t,\lambda(t)Y^*(t)) &= [dA(t,\lambda Y^*(t))]_{|\lambda=\lambda(t)} + Y^*(t)\frac{\partial}{\partial y}A(t,\lambda(t)Y^*(t))d\lambda(t) \\ &= \frac{\partial}{\partial y}A(t,\lambda(t)Y^*(t))\sigma\pi_y^*\lambda(t)Y^*(t)dW^Q(t) \\ &\quad + [rA(t,\lambda(t)Y^*(t)) + \ell(t) - \lambda(t)c^*(t)]1_{(\lambda(t)Y^*(t)>b(t))}dt + \frac{\partial}{\partial t}K(t)1_{(\lambda(t)Y^*(t)\leq b(t))}dt \\ &\quad + Y^*(t)\frac{\partial}{\partial y}A(t,\lambda(t)Y^*(t))1_{(\lambda(t)Y^*(t)=b(t))}d\lambda(t). \end{aligned}$$

Since by (2.34) $\frac{\partial}{\partial y}A(t,\lambda(t)Y^*(t)) = 0$ on the set $\{(t,\omega) : \lambda(t)Y^*(t) = b(t)\}$ this reduces to

$$\begin{aligned} dA(t,\lambda(t)Y^*(t)) &= \frac{\partial}{\partial y}A(t,\lambda(t)Y^*(t))\sigma\pi_y^*\lambda(t)Y^*(t)dW^Q(t) + [rA(t,\lambda(t)Y^*(t)) + \ell(t) - \lambda(t)c^*(t)]dt \\ &\quad + \left(\frac{\partial}{\partial t}K(t) - (rK(t) + \ell(t) - \lambda(t)c^*(t))\right)1_{(\lambda(t)Y^*(t)\leq b(t))}dt. \end{aligned}$$

Finally, since $\{(t,\omega) : \lambda(t)Y^*(t) \leq b(t)\} = \left\{(t,\omega) : \lambda(t) = \frac{b(t)}{Y^*(t)}\right\}$ has a zero $dt \otimes dP$-measure we conclude that

$$dA(t,\lambda(t)Y^*(t)) = [rA(t,\lambda(t)Y^*(t)) + \ell(t) - \lambda(t)c^*(t)]dt + \frac{\partial}{\partial y}A(t,\lambda(t)Y^*(t))\sigma\pi_y^*\lambda(t)Y^*(t)dW^Q(t),$$

i.e. by (2.7) the strategy is admissible. □

[7] $\frac{\partial}{\partial y}$ now means differentiating w.r.t. the second variable.

[8] It is implicit given that $P^a_{\lambda Y^*}(t,T,K)$ now refers to an American put option written on the portfolio $(\lambda(s)Y^*(s))_{s\in[t,T]}$.

2.5.2 The optimal strategy

As mentioned, the put option based strategy introduced in Subsection 2.5.1 turns out to be optimal for the American capital guarantee problem (2.26).

Theorem 2.5.2. *Consider the strategy $(\widehat{c}, \lambda Y^*)$ where, $\forall t \in [0,T]$,*

$$\widehat{c}(t) := \frac{\lambda(t)Y^*(t)}{f(t)} = \lambda(t)c^*(t),$$
$$Y^*(t) := X^*(t) + g(t),$$

combined with a position in an American put option with strike price $K(s) + g(s), \forall s \in [t,T]$, and maturity T written on the portfolio $(\lambda(s)Y^*(s))_{s\in[t,T]}$, where $\lambda(s), s \in [t,T]$, is the function defined by (2.36) and (2.30). By (2.31) the value of the corresponding option based portfolio insurance becomes

$$\widehat{X}^{(\lambda)}(t) := \lambda(t)Y^*(t) + P^a_{\lambda Y^*}(t,T,K+g) - g(t) \geq K(t), \tag{2.37}$$

i.e. the American capital guarantee is fulfilled. Observing that $Y^*(T) = X^*(T)$ the terminal value becomes

$$\widehat{X}^{(\lambda)}(T) = \lambda(T)X^*(T) + (K(T) - \lambda(T)X^*(T))^+ = \max(\lambda(T)X^*(T), K(T)). \tag{2.38}$$

We have that

- *The strategy is optimal for the American capital guarantee control problem given by (2.26).*

Note that, clearly, the strategy in Theorem 2.5.2 is feasible since the unrestricted strategy (c^*, π^*) given in Section 2.3 is feasible.

Remark 2.5.1. To clarify, following the strategy $\left(\widehat{c}, \widehat{X}^{(\lambda)}\right)$ given by Theorem 2.5.2 (heuristically) means that the investor should:

1. Take a loan of size $g(0)$ and reserve the labor income to pay back the loan over the time interval $[0,T]$.
2. Reserve the initial amount of money $\lambda_0(x_0+g(0))$ to follow the consumption and investment strategy $(\widehat{c}, \pi)$ given by

$$\widehat{c}(t) := \lambda_0 c^*(t), \tag{2.39}$$
$$\pi := \pi^*_y. \tag{2.40}$$

   Doing this the investor replicates the portfolio $\lambda_0 Y^*(t), \forall t \in [0,T]$ (see Remark 2.3.1).
3. Use the remaining initial amount of money $(1-\lambda_0)(x_0+g(0))$ to buy an American put option with strike price $K(s) + g(s), \forall s \in [0,T]$, and time to maturity T written on the portfolio $(\lambda_0 Y^*(t))_{t\in[0,T]}$.
4. The first time (say τ) $\lambda_0 Y^*$ drops below the optimal exercise boundary b, sell the American put option. Not taking the loan we made in 1. into account, the portfolio will then be worth $K(\tau) + g(\tau)$[9]. Reserve now the amount of money $\lambda(\tau)(K(\tau) + g(\tau))$ to follow the consumption and investment strategy $(\widehat{c}, \pi)$ given by

$$\widehat{c}(t) := \lambda(\tau)c^*(t),$$
$$\pi := \pi^*_y.$$

   Doing this the investor replicates the portfolio $\lambda(\tau)Y^*(t), \forall t \in [\tau,T]$. Use the remaining initial amount of money $(1-\lambda(\tau))(K(\tau)+g(\tau))$ to buy an American put option with strike price $K(t)+g(t), \forall t \in [\tau,T]$, and maturity T written on the portfolio $\lambda(\tau)Y^*(t), \forall t \in [\tau,T]$.

[9]Taken the loan we made in 1. into account the portfolio is worth $K(\tau)$, i.e. we stand at the capital guarantee boundary.

5. In case the new portfolio $\lambda(\tau)Y^*(t)$ drops below the optimale exercise boundary b at some point in time, $t \in (\tau, T]$, repeat step 4.

Identically to the European case, the American put option is likely not to be sold in the market, but since the market is complete and frictionless such options can be replicated dynamically by a Delta-hedge. i.e. the optimal investment strategy $\widehat{\pi}(t),\ t \in [0,T]$, is given by

$$\widehat{\pi}(t)\widehat{X}^{(\lambda)}(t) = \left(1 + \frac{\partial}{\partial y} P^a_{\lambda Y^*}(t,T,K+g)\right) \pi^*_y \lambda(t) Y^*(t).$$

One should notice, that at the boundary no risk is taken, i.e. the position in the American put option offsets the position in the underlying portfolio such that the total exposure to changes in the risky asset, S, becomes zero. Clearly, at the boundary it is optimal to consume a certain time-dependent part of the labor income, thereby leaving room for a risky position in S immediately after hitting the capital boundary. This is numerically illustrated in Subsection 2.5.3.

Proof of optimality. Let $(c,\pi) \in \mathcal{A}'$, with corresponding wealth process $(X(t))_{t\in[0,T]}$, be any given feasible strategy satisfying $X(0) = x_0$ and $X(t) \geq K(t),\ \forall t \in [0,T]$. Since u is concave we get

$$\begin{aligned}
&\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + e^{-\int_0^T \beta(s)ds} u(X(T)) \\
&\quad - \left(\int_0^T e^{-\int_0^t \beta(s)ds} u\left(\widehat{c}(t)\right) dt + e^{-\int_0^T \beta(s)ds} u\left(\widehat{X}^{(\lambda)}(T)\right)\right) \\
&= \int_0^T e^{-\int_0^t \beta(s)ds} \left(u(c(t)) - u\left(\widehat{c}(t)\right)\right) dt + e^{-\int_0^T \beta(s)ds} \left(u(X(T)) - u\left(\widehat{X}^{(\lambda)}(T)\right)\right) \\
&\leq \int_0^T e^{-\int_0^t \beta(s)ds} u'\left(\widehat{c}(t)\right)\left(c(t) - \widehat{c}(t)\right) dt + e^{-\int_0^T \beta(s)ds} u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&=: (*).
\end{aligned}$$

Since (c,π) was arbitrarily chosen we end the proof by showing that $E[(*)] \leq 0$.

Again by the CRRA property $u'(xy) = u'(x)u'(y)$ we have

$$u'\left(\widehat{c}(t)\right)\left(c(t) - \widehat{c}(t)\right) = u'(\lambda(t))u'(c^*(t))\left(c(t) - \widehat{c}(t)\right). \tag{2.41}$$

By use of (2.38) and noting that u' is a decreasing function we get that

$$\begin{aligned}
&u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&= \left(u'(\lambda(T))u'(X^*(T)) \wedge u'(K(T))\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&= u'(\lambda(T))u'(X^*(T))\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) - \left(u'(\lambda(T))u'(X^*(T)) - u'(K(T))\right)^+ \left(X(T) - K(T)\right),
\end{aligned}$$

where the last equality is established by using that $\widehat{X}^{(\lambda)}(T) = K(T)$ on the set $\{(T,\omega) : u'(\lambda(T)X^*(T)) \geq u'(K(T))\}$. Since by assumption $X(t) \geq K(t),\ \forall t \in [0,T]$, this reduces to

$$u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \leq u'(\lambda(T))u'(X^*(T))\left(X(T) - \widehat{X}^{(\lambda)}(T)\right). \tag{2.42}$$

Inserting (2.41) and (2.42) and then (2.12) and (2.13) we get

$$
\begin{aligned}
E[(*)] &:= E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u'\left(\widehat{c}(t)\right)\left(c(t)-\widehat{c}(t)\right)dt\right] + E\left[e^{-\int_0^T \beta(s)ds} u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T)-\widehat{X}^{(\lambda)}(T)\right)\right] \\
&\leq E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u'(\lambda(t)) u'(c^*(t))\left(c(t)-\widehat{c}(t)\right)dt\right] \\
&\quad + E\left[e^{-\int_0^T \beta(s)ds} u'(\lambda(T)) u'(X^*(T))\left(X(T)-\widehat{X}^{(\lambda)}(T)\right)\right] \\
&= E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u'(\lambda(t)) \xi^* H(t)\left(c(t)-\widehat{c}(t)\right)dt\right] \\
&\quad + \left[e^{-\int_0^T \beta(s)ds} u'(\lambda(T)) \xi^* H(T)\left(X(T)-\widehat{X}^{(\lambda)}(T)\right)\right] \\
&= \xi^* E^Q\left[\int_0^T e^{-rt} u'(\lambda(t))\left(c(t)-\widehat{c}(t)\right)dt + e^{-rT} u'(\lambda(T))\left(X(T)-\widehat{X}^{(\lambda)}(T)\right)\right].
\end{aligned}
$$

Since $u'(\lambda(t))$ is a decreasing function[10] we can use the integration by parts formula

$$
\begin{aligned}
E[(*)] \leq \xi^* \Bigg(& E^Q\Bigg[\underbrace{\int_0^T e^{-rt} u'(\lambda(t))\left(c(t)-\widehat{c}(t)\right)dt + \int_0^T u'(\lambda(t)) d\left(e^{-rt}\left(X(t)-\widehat{X}^{(\lambda)}(t)\right)\right)}_{(\star)}\Bigg] \\
& + E^Q\Bigg[\underbrace{\int_0^T e^{-rt}\left(X(t)-\widehat{X}^{(\lambda)}(t)\right) du'(\lambda(t))}_{(\star\star)}\Bigg]\Bigg).
\end{aligned}
\tag{2.43}
$$

The second part of (2.43) can be rewritten as

$$
\begin{aligned}
E^Q[(\star\star)] &= E^Q\left[\int_0^T e^{-rt}(X(t)-K(t)) du'(\lambda(t))\right] \\
&\quad + E^Q\left[\int_0^T e^{-rt}\left(K(t)-\widehat{X}^{(\lambda)}(t)\right) du'(\lambda(t))\right].
\end{aligned}
$$

The first term is non-positive since per definition $X(t) \geq K(t)$, $\forall t \in [0,T]$, and $du'(\lambda(t)) \leq 0$, $\forall t \in [0,T]$ (u' is decreasing and λ is increasing). The second term equals zero since $du'(\lambda(t)) \neq 0$ only on the set $\{(t,\omega) : \widehat{X}^{(\lambda)}(t) = K(t)\}$. We conclude that $E^Q[(\star\star)] \leq 0$. The first part of (2.43) can be rewritten as

$$
\begin{aligned}
E^Q[(\star)] &= E^Q\left[\int_0^T u'(\lambda(t)) d\left(\int_0^t e^{-rs}(c(s)-\widehat{c}(s))ds\right)\right] \\
&\quad + E^Q\left[\int_0^T u'(\lambda(t)) d\left(e^{-rt}\left(X(t)-\widehat{X}^{(\lambda)}(t)\right)\right)\right] \\
&= E^Q\left[\int_0^T u'(\lambda(t)) d\left(\int_0^t e^{-rs}(c(s)-\ell(s))ds + e^{-rt}X(t)\right)\right] \\
&\quad - E^Q\left[\int_0^T u'(\lambda(t)) d\left(\int_0^t e^{-rs}(\widehat{c}(s)-\ell(s))ds + e^{-rt}\widehat{X}^{(\lambda)}(t)\right)\right].
\end{aligned}
\tag{2.44}
$$

[10]This ensures the that the stochastic integral in (2.43) is well-defined.

Since both strategies are admissible we note that according to (2.6), equation (2.44) consists of two stochastic integrals with respect to Q-martingales. Since $u'(\lambda(t)) \leq u'(\lambda_0), \forall t \in [0,T]$, we get

$$E^Q[(\star)] = 0.$$

Finally, we can conclude that

$$E[(*)] = E^Q[(\star)] + E^Q[(\star\star)] \leq 0.$$

□

Remark 2.5.2. As a final remark we want to emphasize that the case where the investor assigns zero utility to terminal wealth can be handled equivalent to Section 2.3–2.5. To be clear, consider the unrestricted problem

$$\sup_{(c,\pi)\in\mathcal{A}'} E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt\right], \tag{2.45}$$

where the investors wealth follows the dynamics given by (2.1) with initial wealth equal to x_0. It is well known that the solution to (2.45) is equivalent to the unrestricted solution given by (2.14) and (2.15), the only difference being that the deterministic function f is now given by

$$f(t) = \int_t^T e^{-\int_t^s [\widetilde{r}(y)+\beta(y)]dy} ds, \tag{2.46}$$

with $\widetilde{r}(y) = \frac{\gamma}{\gamma}\beta(y) - \frac{\gamma}{\gamma}r - \frac{1}{2}\frac{\gamma}{(1-\gamma)^2}\left(\frac{\alpha-r}{\sigma}\right)^2$. Now f converges to zero as time goes to T. Consequently, since by (2.14)

$$c^*(t) = \frac{X(t)+g(t)}{f(t)},$$

we notice that the optimal terminal wealth will be zero (since the investor assigns no utility to the terminal wealth he simply consumes everything before time T). The solution to (2.45) under a European or an American capital guarantee is equivalent to the solutions presented in Proposition 2.4.1 respectively Theorem 2.5.2, the only difference being that f is now given by (2.46).

2.5.3 Numerical illustrations

In this subsection we illustrate the optimal strategy presented in Theorem 2.5.2 in the case with a restriction to borrow against future labor income, i.e. $K(t) = 0,\ \forall t \in [0,T]$. Naturally, we compare the strategy to the unrestricted strategy derived in Section 2.3. More specifically, consider an investor with time horizon $T = 10$, time and utility preferences $(\beta, \gamma) = (0.01, -2)$, endowed with zero initial wealth and a labor income rate starting out at $\ell(0) = 40000$ followed by a monthly half a percent raise. The investor faces a Black-Scholes market with parameters $(\alpha, r, \sigma) = (0.12, 0.04, 0.2)$ resulting in $\pi_y^* = \frac{1}{\gamma}\frac{\alpha-r}{\sigma^2} = \frac{2}{3}$. The time interval has been discretizied equidistantly into 120 time points, i.e. monthly time points. By this we mean that the American put option has been evaluated as a discrete time option such that the guarantee is only valid monthly. Likewise, λ and therefore the optimal consumption and investment strategy has only been changed monthly. Saying that, it is important to stress that in-between the monthly time points, the amount of money to consume and invest, respectively, have been re-balanced continuously[11].

[11]We have used (2.16). Note that $X(t)^* + g(t),\ t \in [0,T]$, is a geometric Brownian motion since we have chosen β to be constant.

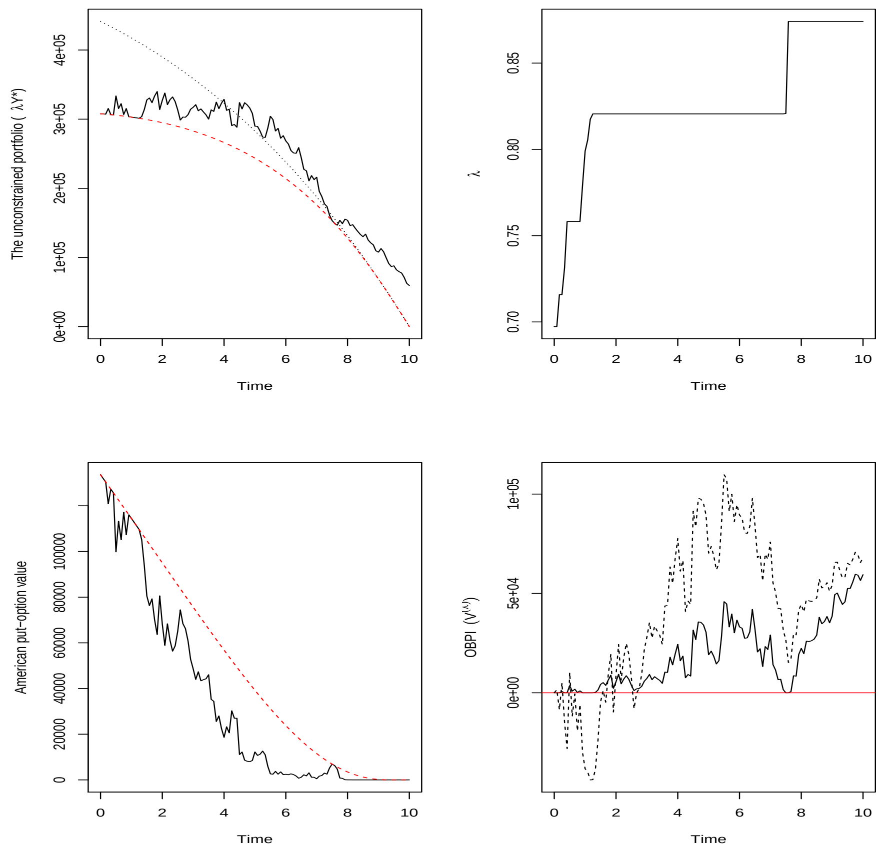

Figure 2.1: Upper left: The underlying portfolio $\lambda(t)Y^*(t)$ (solid curve), together with the strike price $g(t)$ (dotted curve) and the optimal exercise boundary $b(t)$ of the American put option (dashed curve). Upper right: $\lambda(t)$. Lower left: The American put option (solid curve) together with the optimal exercise value (dashed curve). Lower right: The optimal portfolio $\widehat{X}^{(\lambda)}(t)$ (solid curve) together with the corresponding optimal unrestricted portfolio $X^*(t)$ (dashed curve) and the guarantee ($K = 0$).

Figure 2.1 illustrates the optimal portfolio together with its building blocks (the underlying portfolio and the associated American put option). The investor starts out by splitting the initial borrowed amount of money $g(0) = 441361.8$ into two portions: One of size $\lambda(0)g(0) = 307759.9$ reserved to finance the investment and consumption strategy given by (2.39) and (2.40) (i.e. a hedge of $\lambda_0 Y^*$), and one of size $(1-\lambda(0))g(0) = 133601.9$ used to buy (or delta-hedge) an American put option with strike $g(t)$ and maturity T written on the underlying portfolio $\lambda_0 Y^*$. Looking at Figure 2.1 upper left corner we see that initially this corresponds to choosing an underlying portfolio of size equal to the exercise boundary of the corresponding American put option. One can realize that this is always the case when the initial amount of money equals the capital guarantee $K(0)$.

As seen in Figure 2.1, as time goes by the optimal portfolio hits the boundary. For this specific example we observe (after evaluating a large number of sample paths) that the optimal portfolio hits the boundary within the (say) first 3 years with a rather large probability, and that it hits the boundary in the (say) last 2 years before terminal time T with a negligible probability. Realizations like this can often be found to be a consequence of the model design. In this example the reason is partly due to the fact that the investor is endowed with an increasing labor income and no initial wealth. Looking at the optimal consumption strategy (2.39) one observes that this labor income design causes the investor, who holds a small portfolio, to save a little fraction of the labor income in the beginning of the period, whereas he chooses to save a large fraction of the labor income in the years just before terminal time T (see Figure 2.2).

In Section 2.5.1 we found how to change λ dynamically in order for the put option based strategy to be admissible. In Figure 2.1 upper right corner we see how $\lambda(t),\ t \in [0,T]$, increases when the underlying process $(\lambda Y^*)_{t\in[0,T]}$ hits the exercise boundary of the American put option. Consequently, the American put option never exceeds its inner value as seen in Figure 2.1 lower left corner.

Finally, Figure 2.1 lower right corner shows the optimal portfolio plotted against the corresponding unrestricted optimal portfolio. We see how the optimal portfolio stays above or at zero during the time period while the unrestricted optimal portfolio becomes negative over several distinct time intervals. The fact that the optimal portfolio increases drastically away from zero the last two years before terminal time T is primary due to the fact that the investor's optimal consumption strategy leaves room for saving labor income in the final years whenever wealth is sufficient low (see Figure 2.2).

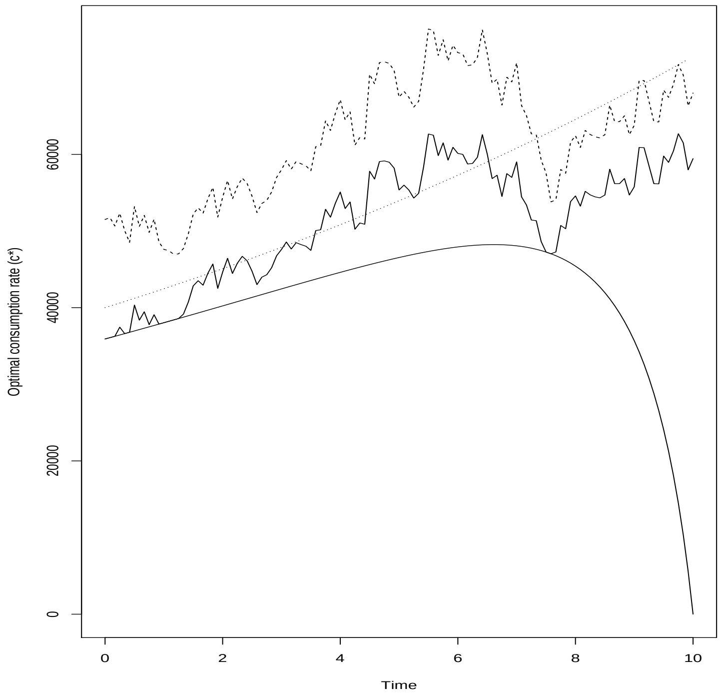

Figure 2.2: The optimal consumption rate (solid curve) together with the optimal unrestricted consumption rate (dashed curve), the deterministic labor income rate (dotted curve) and the deterministic optimal boundary consumption rate (solid smooth curve).

In Figure 2.2 we see how the optimal consumption rate fluctuates around the labor income rate. To begin with it is optimal to save a certain fraction of the labor income. As investments start to pay off it eventually becomes optimal to consume more than the labor income ($t \in [2.5, 6.5]$). After 6.5 years the investor loses money on his investments (for this specific sample path) and it therefore becomes optimal for him to start saving money again. One can say that Figure 2.2 simply shows the investor's preferences for consuming now versus consuming later. Comparing to the optimal unrestricted consumption rate, we see that the restriction to borrow against future labor income has a smoothing effect on consumption (as shown by He and Pagès (1993)). The reason that the optimal consumption rate at the boundary goes to zero is clear: As time approaches T while wealth is equal to zero it becomes infinitely valuable to save money since the marginal utility of terminal wealth goes to infinity as terminal wealth goes to zero.

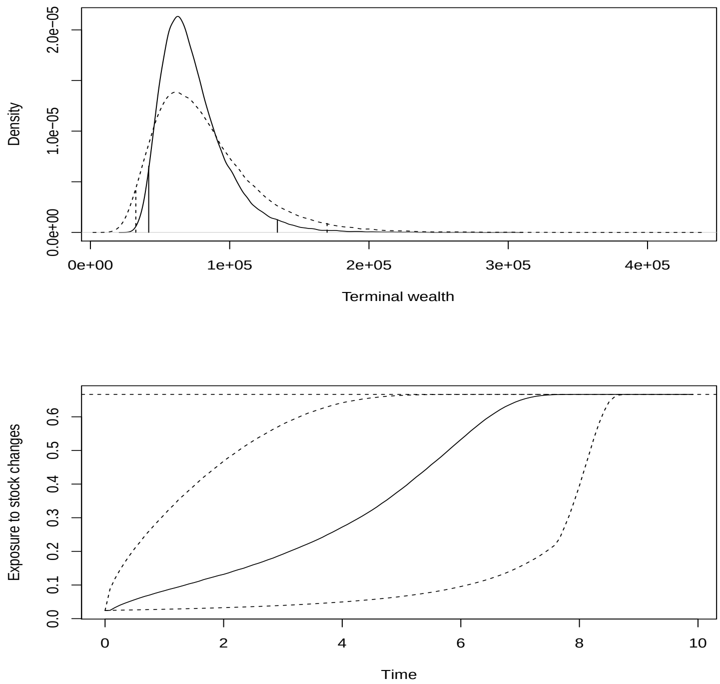

Figure 2.3: Upper window: The density of the optimal terminal wealth (solid curve) together with the density of the optimal unrestricted terminal wealth (dashed curve). Vertical lines indicate the 95 percent confidence intervals. Lower window: The optimal fraction of total wealth invested in the risky stock. The solid curve represents the median while the dashed curves indicate the 95 percent confidence interval. The line $y = \frac{2}{3}$ represents the unrestricted optimal fraction to invest in the risky stock.

In Figure 2.3 upper window we compare the optimal terminal wealth with the unrestricted optimal terminal wealth of Merton. Not surprisingly, the restriction to borrow against future labor income results in a much more narrow distribution for the terminal wealth. The optimal terminal wealth has 95 percent of its density mass within the interval $[41854, 134252]$, whereas for the unrestricted problem this interval is $[32590, 169928]$. Figure 2.3 lower window illustrates the risk taken in the optimal portfolio. The risk taken is defined as the fraction of total wealth (wealth plus the financial value of future labor income) invested in the risky stock. We see that the median forms a 's-curve' towards the unrestricted Merton fraction $\left(\pi = \frac{2}{3}\right)$. This is due to the fact that eventually, the optimal portfolio will increase away from the boundary $(K = 0)$ and the American put option will become very cheap, resulting in a strategy almost identical to the one of Merton. It should be stressed that we could easily find another example where this will not be the case. The very wide 95 percent confidence interval illustrates how the optimal investment strategy can be very different for different sample paths. It also illustrates how much the optimal investment strategy differs from the unrestricted strategy of Merton. Even in the case where the investor gets a very large and steady payoff from his risky investments he do not invest nearly as aggressively as Merton before after 5 years (halfway). Reversely, in the case of very large and steady loses on the investments the investor places after 5 years only about 6 percent of total wealth in the risky stock.

2.6 Conclusion

We have solved the classic consumption and investment control problem of Merton in the presence of labor income and an American capital guarantee restriction. The solution obtained is an option based portfolio insurance (OBPI) combining a long position in the optimal unrestricted allocation with an American put option. More precisely, the investor optimally invests in a budget adjusted version of the unrestricted portfolio and buys an insurance on that portfolio in the form of an American put option. Initially the budget for the unrestricted portfolio is adjusted such that this is affordable. The introduction of consumption and labor income to such a setting is, to the author's knowledge, not considered before, and the main contribution from this article was to show, by use of the martingale approach to arbitrage-free theory in finance, that the OBPI approach can be carried out to find the optimal strategy in the presence of consumption and labor income. The introduction of a deterministic labor income by itself, in the presence of an American capital guarantee, is not trivial. Unlike the unrestricted problem of Merton, being endowed with a deterministic labor income or being endowed with an initial wealth enlarged by the financial value of future labor income is not equivalent. We have shown that the introduction of deterministic labor income can be handled by choosing an appropriate strike value for the American put option used to insure the investors portfolio from falling below the capital restriction. The strategy, which has to be re-balanced whenever the constraint is active, is rather complicated. Another contribution offered by this article is an in-depth explanation and analysis of the optimal strategy supported by graphical illustrations.

3. Optimal consumption, investment and life insurance with surrender option guarantee

Abstract: We consider an investor, with an uncertain life time, endowed with deterministic labor income, who has the possibility to continuously invest in a Black-Scholes market and to buy life insurance or annuities. We solve the optimal consumption, investment and life insurance problem when the investor is restricted to fulfill an American capital guarantee. By allowing the guarantee to depend, in a very general way, on the past we include, among other possibilities, the interesting case of a minimum rate of return guarantee, commonly offered by pension companies. The optimal strategies turn out to be on option based portfolio insurance form, but since the capital guarantee is valid at every intermediate point in time, re-calibration is needed whenever the constraint is active.

Keywords: Stochastic control; martingale method; life insurance; rate guarantee; option based portfolio insurance; CRRA utility.

3.1 Introduction

We unify two directions of generalizing the classical Merton's consumption-investment problem with labor income, namely the introduction of life insurance decisions and the introduction of capital constraints. This unification is particularly relevant since the savings in a life and pension contract often contain restrictions that are special examples of our capital constraint. We solve the joint problem where the individual with *an uncertain life time* can buy life insurance or annuities and, at the same time, there is a general *capital constraint* on his savings.

The *uncertain life time* is well-motivated but it turns a deterministic income rate *while being alive* into a stochastic income rate. Access to life insurance and annuities completes the market and makes the financial value of labor income unique.

The *capital constraint* allowed for in this paper is sufficiently general to cope with two well-motivated situations:

1) One possible capital constraint is that the savings are not allowed to become negative. We can speak of this as a no-borrowing constraint but it is important to stress that it is a constraint on wealth and not on investment position. In a life insurance context one would speak of such a constraint as a non-negative reserve constraint. The intuition is that it should always be the pension institution that owes money to the policy holder and not the other way around. This is a standard institutional/regulatory constraint that serves to separate, institutionally, saving business from lending business.

2) Another possible capital constraint is that the savings earn a minimum return. We speak of this as an interest rate guarantee. This is a delicate capital guarantee since it protects the dynamics of wealth rather than wealth itself. Thus, formalized as a constraint on the capital itself, this constraint depends on the trajectory of the state variables. Such a guarantee appears in many different types of life insurance contracts, ranging from participating life insurance with a so-called surrender option with a guaranteed sum upon surrender to unit-link life insurance where you can add on different types of options on the return.

In both situations one may ask why is the constraint there at all? The separation of saving business from lending business perhaps serves to protect the pension institution from unspanned policy holder credit risk. We stress the word *unspanned* here because one could argue that *spanned* life insurance risk is also credit risk in the context of borrowing against future income. Credit risk is a matter for the lending business who would require some material collateral, like a house or something, in addition to a 'planned' working life. But there is no such unspanned policy holder credit risk within our model, so who asked for a constraint? Also the guarantee on return may seem odd: Who asked for that? The titular question posed by Jensen and Sørensen (2001) means almost the same. Well, the answer 'the social planner' applies to both questions. If the society prefers that savings are non-negative, or if the society prefers that policy holders (via their delegated pension institutions) invest as safely as it is dictated by an interest rate guarantee, then the constraints can appear through the regulatory environment. Thus, these constraints should not necessarily be interpreted as a twist on the policy holders preferences within the model but as a societal constraint from outside the model/preferences. Formalistically, the two interpretations are the same, though. It is interesting to discuss if and when such constraints really are meaningful from the point of view of the social planner. However, this is beyond the scope of this paper. Furthermore, during these decades where certain markets show a tendency from classical interest rate guarantee participating life insurance towards modern unit-link products, the policy holders tend to actually buy these options on the return. So, could it be something that the people want, after all? Prospect theory may help motivating guarantees, see e.g. Døskeland and Nordahl (2008). We do not incorporate prospect theory in our preferences but take the guarantees to be exogenously given.

Life insurance was introduced to Merton's classical consumption-investment problem first already by Richard (1975) but Yaari (1965) had even earlier results in discrete time. It took three decades before the actuarial academic community realized the importance of the fundamental patterns of thinking by Richard (1975). Since then, many articles generalize the results by Richard in various directions. They include Pliska and Ye (2007) who allow for an unbounded lifetime; Huang and Milevsky (2008) who allow for unspanned labor income; Huang et al. (2008) who separate the breadwinner income process from the family consumption process; Steffensen and Kraft (2008) who generalize to a multistate Markov chain framework typically used by actuaries for modeling a series of life history events; Nielsen and Steffensen (2008) who work with constraints on the insurance sum paid out upon death; Bruhn and Steffensen (2011) who generalize to a multiperson household, with focus on a married couple with economically and/or probabilistically dependent members; Kwak et al. (2011) who also consider a household but focus on generation issues.

Our contribution compared to the list of papers in the former paragraph is that we add capital constraints. They are particularly interesting since the life insurance and pension savings contracts typically contain borrowing or return constraints set by either the insurance company directly or by the regulatory authorities indirectly.

Capital constraints have been studied in a series of papers. They were first studied as a European (i.e. terminal) capital constraint in Teplá (2001) and Jensen and Sørensen (2001) who, interestingly, motivated the study of capital constraints from guarantees in pension contracts. Later results of El-Karoui et al. (2005) concern an American (i.e. continuous) capital guaranteed. The three papers on capital constraints mentioned above all deal with the investment problem exclusively, disregarding labor income and consumption. Kronborg (2011) generalizes the results of El-Karoui et al. (2005) to include spanned labor income and optimal consumption.

Our contribution compared to the list of papers in the former paragraph is that we add life insurance risk and decisions and allow for more general path dependent constraints. This combination makes it possible to study realistic decision processes and product designs including relevant regulatory and/or institutional constraints in the pension savings market.

Speaking of terminal and continuous constraints as being European and American, respectively, may seem a bit far-fetched until the resulting strategy appears: The best strategy is to follow a so-called option based portfolio insurance strategy (OBPI) where a certain part of the capital is invested in the optimal portfolio for the unconstrained problem and then put option protected downwards for the residual amount such that the guarantee is fulfilled and such that the combined position is worth the wealth. In the European case the portfolio is protected by a European put option whereas in the American case the portfolio is protected by an American put option. The underlying portfolio of the American put option is continuously updated, though, paid by the cash flow from the put option, see El-Karoui et al. (2005). A particular feature arises in the presence of income and consumption, namely, that also the strike is updated, see Kronborg (2011). This 'updating of the American strike' pattern appears also in the present paper in the presence of life insurance risk and protection. The update becomes even more delicate due to the path dependence of the capital guarantee.

Two other lines of literature should be mentioned for dealing with related problems although the market or the control fundamentally differ: Optimal annuitization has been studied in a series of papers, see Milevsky and Young (2007) and references therein. Optimal retirement timing has been studied in e.g. Farhia and Panageas (2007). We also wish to mention Dybvig and Liu (2010) since they combine the retirement decision with capital constraints similar to the ones we deal with here.

We apply the martingale method developed by Cvitanic et al. (1987), Cox and Huang (1989), Cox and Huang (1991) and Cvitanic and Karatzas (1992). This method deals efficiently with capital constraints and has been used in all the papers on capital constraints mentioned above. However, Kraft and Steffensen (2012) show how also dynamic programming applies to problems with capital constraints.

The outline of the paper is as follows: In Section 2, we present the state processes and define the natural set of admissible strategies in presence of spanned labor income. In Section 3, we formalize and solve the unconstrained problem. Section 4 contains a formalization and a solution to the constrained problem. Numerical examples of the (constrained and unconstrained) optimal wealth dynamics are illustrated. We conclude in Section 5. Finally, a rigorous proof showing the admissibility of the strategy found in Section 4 is presented in the appendix.

3.2 Setup

Consider a policyholder with a life insurance policy issued at time 0 and terminated at a deterministic point in time $0 < T < \infty$. One should think of T as the time of retirement decided beforehand by the policyholder. The policyholder's time of death is given by a non-negative random variable τ_d defined on a given probability space $(\Omega, \mathcal{F}, P)$. We assume that τ_d has a probability distribution given by a function F with density function denoted by f. We can then define the *survivor function* $\overline{F}$ by

$$\overline{F}(t) := P(\tau_d \geq t) = 1 - F(t) \text{ where } F(t) := P(\tau_d < t) = \int_0^t f(s)ds. \tag{3.1}$$

The instantaneous time-t mortality rate of the policyholder is given by the hazard function

$$\mu(t) := \lim_{\epsilon \to 0} \frac{P(t \leq \tau_d < t + \epsilon \mid \tau_d \geq t)}{\epsilon} = \lim_{\epsilon \to 0} \frac{F(t+\epsilon) - F(t)}{\epsilon} \frac{1}{\overline{F}(t)} = \frac{F'(t)}{\overline{F}(t)} = \frac{f(t)}{\overline{F}(t)} = -\frac{\partial}{\partial t} \log \overline{F}(t).$$

We get the following well-known expressions frequently used by actuaries

$$\overline{F}(t) = e^{-\int_0^t \mu(s)ds}, \tag{3.2}$$

$$f(t) = \mu(t) e^{-\int_0^t \mu(s)ds}. \tag{3.3}$$

In this paper we assume that the mortality rate is deterministic and given by a continuous function $\mu : [0, \infty) \to [0, \infty)$.

We now describe the investment and life insurance market available to the policyholder. Consider a Black-Scholes market consisting of a bank account, B, with risk free short rate, r, and a risky stock, S, with dynamics given by

$$dB(t) = rB(t)dt,\ B(0) = 1,$$
$$dS(t) = \alpha S(t)dt + \sigma S(t)dW(t),\ S(0) = s_0 > 0.$$

Here $\alpha, \sigma, r > 0$ are constants and we assume $\alpha > r$. The process W is a standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ equipped with the filtration $\mathbb{F}^W = (\mathcal{F}^W(t))_{t\in[0,T]}$ given by the P-augmentation of the filtration $\sigma\{W(s);\ 0 \le s \le t\}, \forall t \in [0,T]$. We assume that the policyholder's random time of death τ_d is stochastically independent of the filtration $\mathbb{F}^W$.

The policyholder is assumed to be endowed with a continuous deterministic labor income of rate $\ell(t) \ge 0, \forall t \in [0,T]$, as long as he is alive, and possibly an initial amount of money $x_0 \ge 0$. The policyholder should be thought of as an agent who has the possibility to invest and continuously consume from or save on a pension account. Naturally, we, as done by actuaries, refer to the policyholder's wealth, denoted by X, as the *reserve*. The reserve is maintained and invested by the pension company, but the policyholder chooses, in a continuous way, the risky asset allocation and the sum insured to be paid out upon death before time T[1]. More specifically, let $c \ge 0$ be the rate of consumption and let θ be the amount of money invested in the risky stock. Furthermore, let $D \in [0,\infty)$ be the sum insured to be paid out upon death before time T. Choosing D, the policyholder agrees to hand over the amount of money $X - D$ to the pension company upon death before time T, i.e. the pension company keeps the reserve X for themselves and pays out D as life insurance. Disregarding policy expenses the natural actuarial *risk premium* rate to pay for the life insurance D at time t is $\mu(t)(D(t) - X(t))dt$. Observe that choosing $D > X$ corresponds to buying a life insurance and choosing $D < X$ corresponds to selling a life insurance, i.e. buying an annuity. The dynamics of the reserve at the state alive[2] becomes

$$dX(t) = [rX(t) + \theta(t)(\alpha - r) + \ell(t) - c(t) - \mu(t)(D(t) - X(t))]dt + \sigma\theta(t)dW(t),\ t \in [0,T), \tag{3.4}$$
$$X(0) = x_0.$$

Once again we stress that in the case of death before the terminal time T only the sum insured, D, is paid out to the policyholder. We call the cash flow injected to the reserve given by $\ell(t) - c(t) - \mu(t)D(t),\ \forall t \in [0,T]$, the *pension contributions*.

We assume the trading strategy, θ, to be an $\mathcal{F}_t^W$-adapted process with $\int_0^T (\theta(t))^2 dt < \infty$ a.s., the consumption rate, c, to be a non-negative $\mathcal{F}_t^W$-adapted process with $\int_0^T c(t)dt < \infty$ a.s., and the insurance strategy, D, to be a non-negative $\mathcal{F}_t^W$-adapted process with $\int_0^T D(t)dt < \infty$ a.s. Note that this ensures that X is well-defined and we get the unique solution, $\forall t \in [0,T]$, known as the time-t *retrospective reserve*

$$\begin{aligned} X(t) = x_0 e^{\int_0^t (r+\mu(y))dy} &+ \int_0^t e^{\int_s^t (r+\mu(y))dy}[\theta(s)(a - r) + \ell(s) - c(s) - \mu(s)D(s)]ds \\ &+ \int_0^t e^{\int_s^t (r+\mu(y))dy}\sigma\theta(s)dW(s). \end{aligned} \tag{3.5}$$

It is well known that in the Black-Scholes market the equivalent martingale measure, Q, is unique and given by the Radon-Nikodym derivative

$$\frac{dQ(t)}{dP(t)} = \Lambda(t) := \exp\left(-\int_0^t \frac{\alpha - r}{\sigma}dW(s) - \frac{1}{2}\int_0^t \left(\frac{\alpha - r}{\sigma}\right)^2 ds\right),\ t \in [0,T], \tag{3.6}$$

[1]An alternative interpretation, also given by Nielsen and Steffensen (2008), is that the company chooses the investment and life insurance strategy on the behalf of the policyholder.

[2]For the rest of the paper we refer to *the reserve at the state alive* as simply *the reserve.*

and that the process W^Q given by

$$W^Q(t) = W(t) + \frac{\alpha - r}{\sigma} t, \ t \in [0, T], \tag{3.7}$$

is a standard Brownian motion under the martingale measure Q. We get that the time-t retrospective reserve can be represented, $\forall t \in [0, T]$, by

$$\begin{aligned} X(t) = x_0 e^{\int_0^t (r+\mu(y))dy} &+ \int_0^t e^{\int_s^t (r+\mu(y))dy} [\ell(s) - c(s) - \mu(s)D(s)]ds \\ &+ \int_0^t e^{\int_s^t (r+\mu(y))dy} \sigma\theta(s) dW^Q(s). \end{aligned} \tag{3.8}$$

Definition 3.2.1. *Define the set of admissible strategies, denoted by $\mathcal{A}$, as the consumption, investment and life insurance strategies for which the corresponding wealth process given by (3.8) is well-defined,*

$$X(t) + g(t) \geq 0, \ \forall t \in [0, T], \tag{3.9}$$

where g is the time-t actuarial value of future labor income defined by

$$g(t) := \int_t^T e^{-\int_t^s (r+\mu(y))dy} \ell(s)ds,$$

and

$$E^Q \left[\int_0^T e^{-\int_0^s (r+\mu(y))dy} \sigma\theta(s) dW^Q(s) \right] = 0. \tag{3.10}$$

The technical condition (3.10) is equivalent to the condition that under Q the process

$$\int_0^t e^{-\int_0^s (r+\mu(y))dy} \sigma\theta(s) dW^Q(s), \ t \in [0, T],$$

is a martingale (in general it it only a local martingale, and also a supermartingale if (3.9) is fulfilled, see e.g. Karatzas and Shreve (1998)). From this we conclude that (3.10) insures that $(c, \theta, D) \in \mathcal{A}$ if and only if $X(T) \geq 0$ and, $\forall t \in [0, T]$,

$$X(t) + g(t) = E^Q \left[\int_t^T e^{-\int_t^s (r+\mu(y))dy} (c(s) + \mu(s)D(s))ds + e^{-\int_t^T (r+\mu(y))dy} X(T) \,\middle|\, \mathcal{F}^W(t) \right]. \tag{3.11}$$

At time zero this means that the strategies have to fulfill the *budget constraint*

$$x_0 + g(0) = E^Q \left[\int_0^T e^{-\int_0^t (r+\mu(y))ds} (c(t) + \mu(t)D(t))dt + e^{-\int_0^T rds} X(T) \right]. \tag{3.12}$$

For latter use we state the following remark.

Remark 3.2.1. Define

$$Z(t) := -\int_0^t e^{-\int_0^s (r+\mu(y))dy} (\ell(s) - c(s) - \mu(s)D(s))ds + e^{-\int_0^t (r+\mu(y))dy} X(t), \ t \in [0, T]. \tag{3.13}$$

By (3.8) we have that condition (3.10) is fulfilled if and only if Z is a martingale under Q. The natural interpretation is that under Q the discounted reserve minus the discounted pension contributions should be a martingale. It is useful to note that if Z is a martingale under Q the dynamics of X can be represented in the form

$$dX(t) = [(r + \mu(t))X(t) + \ell(t) - c(t) - \mu(t)D(t)]dt + \phi(t)dW^Q(t),\ t \in (0, T], \tag{3.14}$$

for some $\mathcal{F}^W(t)$-adapted process ϕ. Moreover, if the dynamics of X can be represented in the form given by (3.14) for some $\mathcal{F}^W(t)$-adapted process ϕ satisfying $\phi(t) \in \mathcal{L}^2$, $\forall t \in [0, T]$, then Z is a martingale under Q.

Finally, by condition (3.9) we allow the reserve process to become negative, as long as it does not exceed (in absolute value) the actuarial value of future labor income. Note that condition (3.9) puts a lower boundary on the reserve process and therefore rules out doubling strategies (see e.g. Karatzas and Shreve (1998)).

3.3 The unrestricted pension problem

For the unrestricted pension problem the policyholder chooses his consumption, investment and life insurance strategy in order to optimize the expected utility from consumption, bequest upon death, and terminal pension. That is, he searches for the strategy fulfilling the supremum

$$\begin{aligned}\sup_{(c,\theta,D)\in\mathcal{A}'} E\Bigg[\int_0^{T\wedge\tau_d} & e^{-\int_0^t \beta(s)ds} u(c(t))dt + K_1 e^{-\int_0^{\tau_d} \beta(s)ds} u(D(\tau_d))1_{(\tau_d<T)} \\ & + K_2 e^{-\int_0^T \beta(s)ds} u(X(T))1_{(\tau_d\geq T)}\Bigg],\end{aligned} \tag{3.15}$$

where $u : \mathbb{R} \to [-\infty, \infty)$ is a utility function representing the policyholder's risk aversion and β is a deterministic function representing the policyholder's time preferences. The set $\mathcal{A}'$, called the set of *feasible* strategies, is a subset of the admissible strategies, defined by

$$\begin{aligned}\mathcal{A}' := \Bigg\{(c,\theta,D) \in \mathcal{A} \,\Bigg|\, & E\Bigg[\int_0^{T\wedge\tau_d} e^{-\int_0^t \beta(s)ds} \min(0, u(c(t)))dt \\ & + e^{-\int_0^{\tau_d} \beta(s)ds} K_1 \min(0, u(D(\tau_d)))1_{(\tau_d<T)} + e^{-\int_0^t \beta(s)ds} K_2 \min(0, u(X(T)))1_{(\tau_d\geq T)}\Bigg] > -\infty\Bigg\}.\end{aligned} \tag{3.16}$$

We see that it is allowed to draw an infinite utility from a strategy $(c, \theta, D) \in \mathcal{A}'$, but only if the expectation over the negative parts of the utility function is finite. Clearly, for a positive utility function we have that $\mathcal{A}$ and $\mathcal{A}'$ coincide. The constant weight factors $K_1 > 0$ and $K_2 > 0$ are measures of the policyholder's preferences concerning life insurance versus immediate consumption, and terminal pension versus immediate consumption, respectively. The case $K_1 = 0$ has the well-known solution given by the pioneering work of Merton (1971). The case $K_2 = 0$ and the case excluding consumption can be analyzed along the same lines as the analysis given in this paper. In order not to complicate the analysis further this will not be done since it requires special treatment. The results are given by natural modifications of the results obtained in this paper.

We want to write the optimization problem (3.15) in terms of the policyholder's instantaneous mortality rate. By the assumption that the random time of death, τ_d, is stochastically independent of the filtration $\mathbb{F}^W$, and by use of (3.1), we can rewrite the expectation in (3.15) as

$$
\begin{aligned}
&E\Bigg[1_{\{\tau_d \geq T\}} \int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + 1_{\{\tau_d < T\}} \int_0^{\tau_d} e^{-\int_0^t \beta(s)ds} u(c(t))dt \\
&\qquad + K_1 e^{-\int_0^{\tau_d} \beta(s)ds} u(D(\tau_d)) 1_{(\tau_d < T)} + K_2 e^{-\int_0^T \beta(s)ds} u(X(T)) 1_{(\tau_d \geq T)}\Bigg] \\
&= E\Bigg[\overline{F}(T) \int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))dt + \int_0^T f(y) \int_0^y e^{-\int_0^t \beta(s)ds} u(c(t))dtdy \\
&\qquad + K_1 \int_0^T e^{-\int_0^t \beta(s)ds} u(D(t)) f(t)dt + K_2 e^{-\int_0^T \beta(s)ds} u(X(T))\overline{F}(T)\Bigg].
\end{aligned} \tag{3.17}
$$

Observe that

$$
\begin{aligned}
\int_0^T f(y) \int_0^y e^{-\int_0^t \beta(s)ds} u(c(t))dtdy &= \int_0^T e^{-\int_0^t \beta(s)ds} u(c(t)) \int_t^T f(y)dydt \\
&= \int_0^T e^{-\int_0^t \beta(s)ds} u(c(t)) \left[\overline{F}(t) - \overline{F}(T)\right] dt.
\end{aligned}
$$

Plug this into (3.17) to obtain

$$
E\left[\int_0^T e^{-\int_0^t \beta(s)ds} u(c(t))\overline{F}(t)dt + K_1 \int_0^T e^{-\int_0^t \beta(s)ds} u(D(t)) f(t)dt + K_2 e^{-\int_0^T \beta(s)ds} u(X(T))\overline{F}(T)\right]. \tag{3.18}
$$

Finally, inserting (3.2) and (3.3) into (3.18) we can write the optimization problem (3.15) as

$$
\sup_{(c,\theta,D)\in\mathcal{A}'} E\left[\int_0^T e^{-\int_0^t (\beta(s)+\mu(s))ds} [u(c(t)) + K_1\mu(t)u(D(t))]dt + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds} u(X(T))\right]. \tag{3.19}
$$

For the rest of this paper we assume that the policyholder's preferences towards risk are given by a constant relative risk aversion function (CRRA) in the form

$$
u(x) = \begin{cases} \frac{x^\gamma}{\gamma}, & \text{if } x > 0, \\ \lim_{x\searrow 0} \frac{x^\gamma}{\gamma}, & \text{if } x = 0, \\ -\infty, & \text{if } x < 0, \end{cases}
$$

for some $\gamma \in (-\infty, 1)\backslash\{0\}$.

Remark 3.3.1. We conjecture that the analysis presented in this paper can be extended to the case of a general utility function. This would, following the ideas of El-Karoui and Karatzas (1995) and El-Karoui et al. (2005), require use of the Gittins index methodology. We specialize to the case of CRRA utility due to the well-known property that the maximization problem (3.15) becomes linear in initial total wealth. Even in the case of a general path dependent constraint on the reserve (see Section 3.4) this feature allows us to derive fairly explicit expressions for the optimal strategies, leaving us with a greater understanding of the impact on the strategies caused by the introduced constraints.

3.3.1 How to choose K_2 - consumption after retirement

Choosing K_1 is really a matter of personal preferences. How much do you love yourself relatively to your inheritors? However, choosing K_2 can be done in a natural way. A common choice in the literature of consumption/portfolio selection is $K_2 = 1$ by simply not allowing any weight constant. This corresponds to the case where the agent prefers a terminal wealth approximately equal to the size of accumulated consumption over the last year before the terminal time T. At least in a pension context this seems to be a very odd choose. What the policyholder (probably) wants is a terminal pension big enough for him to be able to maintain his standard of living throughout his uncertain remaining life time. At the same time, he should still take his time preferences into account. Define therefore the actuarial time-T value of a life annuity paying 1 per year continuously until the time of death as

$$\overline{a}(T) := E\left[\int_T^\infty e^{-\int_T^s r\,dy} 1_{(\tau_d \geq s)} ds \,\middle|\, \tau_d > T\right] = \int_T^\infty e^{-\int_T^s (r+\mu(y))dy} ds.$$

Disregarding policy expenses, the policyholder standing at the time of retirement, T, might use his terminal pension $X(T)$ to buy, in the life insurance market, a life annuity of rate $X(T)/\overline{a}(T)$. Assume that he chooses to do so and thereafter simply consumes the entire life annuity (leaving no room for saving)[3]. We get that

$$\begin{aligned}
&E\left[\int_T^\infty e^{-\int_0^s \beta(y)dy} u\left(\frac{X(T)}{\overline{a}(T)}\right) 1_{(\tau_d \geq s)} ds\right] \\
&= E\left[\int_T^\infty e^{-\int_0^s (\beta(y)+\mu(y))dy} u\left(\frac{X(T)}{\overline{a}(T)}\right) ds\right] \\
&= E\left[e^{-\int_0^T (\beta(y)+\mu(y))dy} \int_T^\infty e^{-\int_T^s (\beta(y)+\mu(y))dy} ds u\left(\frac{X(T)}{\overline{a}(T)}\right)\right] \\
&= E\left[e^{-\int_0^T (\beta(y)+\mu(y))dy} \overline{a}^S(T) u\left(\frac{X(T)}{\overline{a}(T)}\right)\right], \qquad (3.20)
\end{aligned}$$

where

$$\overline{a}^S(T) := E\left[\int_T^\infty e^{-\int_T^s \beta(y)dy} 1_{(\tau_d \geq s)} ds \,\middle|\, \tau_d > T\right] = \int_T^\infty e^{-\int_T^s (\beta(y)+\mu(y))dy} ds$$

defines the subjective actuarial time-T value of a life annuity paying 1 per year continuously until the time of death. Comparing the terminal term in (3.19) with (3.20) we get the natural choice

$$K_2 = \overline{a}^S(T)\overline{a}(T)^{-\gamma}. \qquad (3.21)$$

3.3.2 Solving the unrestricted problem

We now solve the unrestricted optimization problem given by (3.15). In the calculations we focus on the representation of the problem given by (3.19). The methodology used in the following, introduced by Cvitanic et al. (1987), Cox and Huang (1989), Cox and Huang (1991) and Cvitanic and Karatzas (1992), is know as the martingale method. One could, as done in Richard (1975), Pliska and Ye (2007), Nielsen and Steffensen (2008), Huang et al. (2008) and others, have used the Hamilton-Jacobi-Bellman (HJB) technique to derive the optimal consumption, investment and life insurance strategy. We choose the martingale approach since the solution to the restricted capital guarantee problem introduced in Section 3.4 is based on terms derived from the martingale method in the unrestricted case. Since the HJB technique seems to be the approach applied in the literature to solve such unrestricted optimization problems as (3.15) this subsection may also serve, in its own right, as an example of how to apply the martingale approach to a classic continuous time unrestricted optimization problem including life insurance. First we state the result in Proposition 3.3.1. The result is also obtained, in slightly different

[3]The classical annuity result of Yaari (1965) proves that consuming the entire annuity income is optimal in the absence of bequest.

forms due to different setups and notation, by e.g. Richard (1975), Pliska and Ye (2007) and Nielsen and Steffensen (2008).

Proposition 3.3.1. *The optimal strategy for the problem (3.15) is given by the feedback forms*

$$c^*(t) := \frac{X(t)+g(t)}{f(t)}, \tag{3.22}$$

$$D^*(t) := \frac{X(t)+g(t)}{f(t)} K_1^{\frac{1}{1-\gamma}}, \tag{3.23}$$

$$\theta^*(t) := \frac{1}{1-\gamma}\frac{\alpha-r}{\sigma^2}(X(t)+g(t)), \tag{3.24}$$

where

$$g(t) := \int_t^T e^{-\int_t^s (r+\mu(y))dy}\ell(s)ds, \tag{3.25}$$

$$f(t) := \int_t^T e^{-\int_t^s (\widetilde{r}(y)+\mu(y))dy}\left(1+\mu(s)K_1^{\frac{1}{1-\gamma}}\right)ds + K_2^{\frac{1}{1-\gamma}} e^{-\int_t^T (\widetilde{r}(y)+\mu(y))dy}, \tag{3.26}$$

and

$$\widetilde{r}(t) := -\frac{\gamma}{1-\gamma}r - \frac{1}{2}\frac{\gamma}{(1-\gamma)^2}\left(\frac{\alpha-r}{\sigma}\right)^2 + \frac{1}{1-\gamma}\beta(t). \tag{3.27}$$

Proof. Let $I:(0,\infty]\to[0,\infty)$ denote the inverse of the derivative of the utility function u and observe that by use of the concavity of u we have

$$u(x) \le u(I(z)) - z(I(z)-x),\ \forall x\ge 0, z>0. \tag{3.28}$$

Define the adjusted state price deflator H as

$$H(t) := \Lambda(t)e^{\int_0^t (\beta(s)-r)ds},$$

and define $\xi^*>0$ as the constant satisfying

$$\begin{aligned}\mathcal{H}(\xi^*) := E^Q\Bigg[\int_0^T & e^{-\int_0^t (r+\mu(s))ds}\left[I(\xi^*H(t)) + \mu(t)I\left(K_1^{-1}\xi^*H(t)\right)\right]dt \\ & + e^{-\int_0^T (r+\mu(s))ds} I\left(K_2^{-1}\xi^*H(T)\right)\Bigg] = x_0 + g(0).\end{aligned} \tag{3.29}$$

Such a ξ^* clearly exist since I is continuous and decreasing and maps $(0,\infty]$ onto $[0,\infty)$ (see e.g. Karatzas and Shreve (1998), Chapter 3, Lemma 6.2).

Now, take an arbitrary strategy $(c,\theta,D)\in\mathcal{A}'$ with corresponding wealth process $(X(t))_{t\in[0,T]}$ as given. Using (3.28), the budget constraint (3.12) and finally (3.29) we get that

$$\begin{aligned}& E\left[\int_0^T e^{-\int_0^t (\beta(s)+\mu(s))ds}[u(c(t)) + K_1\mu(t)u(D(t))]dt + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds}u(X(T))\right] \\ \le\ & E\Bigg[\int_0^T e^{-\int_0^t (\beta(s)+\mu(s))ds}[u(I(\xi^*H(t))) + K_1\mu(t)u(I(K_1^{-1}\xi^*H(t)))]dt \\ & \quad + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds}u(I(K_2^{-1}\xi^*H(T)))\Bigg].\end{aligned}$$

Since (c, θ, D) was an arbitrarily chosen feasible strategy we obtain the *candidate* optimal strategy (c^*, θ^*, D^*) given by

$$c^*(t) = I(\xi^* H(t)), \tag{3.30}$$
$$D^*(t) = I(K_1^{-1}\xi^* H(t)), \tag{3.31}$$
$$X^*(T) = I(K_2^{-1}\xi^* H(T)). \tag{3.32}$$

We have expressed the candidate optimal investment strategy in terms of the optimal terminal pension given by (3.32). Since (c^*, θ^*, D^*), by the definition of ξ^*, fulfills the budget constraint (3.12) it is well known that, in a complete market, there exist an investment strategy θ^* such that $X(T) = X^*(T)$ and (c^*, θ^*, D^*) is admissible (see e.g. Cvitanic and Karatzas (1992) or Karatzas and Shreve (1998)). We are now going to calculate the hedging investment strategy θ^* and to write the expressions (3.30) and (3.31) in more explicit forms.

Observing that $I(x) = x^{-1/(1-\gamma)}$ we get from (3.29) that

$$\begin{aligned}
\mathcal{H}(\xi^*) &= E\Bigg[\int_0^T H(t) e^{-\int_0^t (\mu(s)+\beta(s))ds} (\xi^*)^{\frac{-1}{1-\gamma}} H(t)^{\frac{-1}{1-\gamma}} \left(1 + \mu(t) K_1^{\frac{1}{1-\gamma}}\right) dt \\
&\qquad + H(T) e^{-\int_0^T (\mu(s)+\beta(s))ds} K_2^{\frac{1}{1-\gamma}} (\xi^*)^{\frac{-1}{1-\gamma}} H(T)^{\frac{-1}{1-\gamma}}\Bigg] \\
&= (\xi^*)^{\frac{-1}{1-\gamma}} f(0),
\end{aligned}$$

where we have defined

$$\begin{aligned}
f(t) = E\Bigg[&\int_t^T e^{-\int_t^s (\mu(y)+\beta(y))dy} \left(\frac{H(s)}{H(t)}\right)^{\frac{-\gamma}{1-\gamma}} \left(1 + \mu(s) K_1^{\frac{1}{1-\gamma}}\right) ds \\
&+ e^{-\int_t^T (\mu(y)+\beta(y))dy} K_2^{\frac{1}{1-\gamma}} \left(\frac{H(T)}{H(t)}\right)^{\frac{-\gamma}{1-\gamma}} \,\Bigg|\, \mathcal{F}^W(t)\Bigg].
\end{aligned} \tag{3.33}$$

Since $\mathcal{H}(\xi^*) = x_0 + g(0)$ per definition, we get that

$$\xi^* = (x_0 + g(0))^{-(1-\gamma)} f(0)^{1-\gamma}.$$

Inserting this into (3.30)–(3.32) and using the budget constraint (3.12) we get the following expressions for the candidate optimal strategy

$$\begin{aligned}
c^*(t) &= \frac{X(t) + g(t)}{f(t)}, \\
D^*(t) &= \frac{X(t) + g(t)}{f(t)} K_1^{\frac{1}{1-\gamma}}, \\
X^*(T) &= \frac{X(t) + g(t)}{f(t)} \left(\frac{H(T)}{H(t)}\right)^{\frac{-1}{1-\gamma}} K_2^{\frac{1}{1-\gamma}},
\end{aligned}$$

and we recognize (3.22) and (3.23) from Proposition 3.3.1. To calculate f note that

$$E\left[e^{-\int_t^s (\mu(y)+\beta(y))dy} \left(\frac{H(s)}{H(t)}\right)^{\frac{-\gamma}{1-\gamma}} \,\middle|\, \mathcal{F}^W(t)\right] = e^{-\int_t^s \left(\mu(y) + \frac{1}{1-\gamma}\beta(y) - \frac{\gamma}{1-\gamma} r - \frac{1}{2}\frac{\gamma}{(1-\gamma)^2}\left(\frac{\alpha - r}{\sigma}\right)^2\right) dy}. \tag{3.34}$$

Now combine (3.33) and (3.34) to obtain

$$f(t) = \int_t^T e^{-\int_t^s [\widetilde{r}(y)+\mu(y)]dy} \left(1 + \mu(s) K_1^{\frac{1}{1-\gamma}}\right) ds + K_2^{\frac{1}{1-\gamma}} e^{-\int_t^T [\widetilde{r}(y)+\mu(y)]dy},$$

with $\widetilde{r}$ given by (3.27), which we identify as (3.26) from Proposition 3.3.1.

From the Girsanov Theorem we have

$$dH(t) = H(t)\left((\beta(t) - r)dt - \frac{\alpha - r}{\sigma} dW(t)\right).$$

An application of Itô's lemma gives that

$$dX^*(t) = \ldots dt + (X^*(t) + g(t))\frac{1}{1-\gamma}\frac{\alpha - r}{\sigma} dW(t). \tag{3.35}$$

If we compare (3.35) with the X-dynamics of the reserve given by (3.4) we get that

$$\theta^*(t) = \frac{1}{1-\gamma}\frac{\alpha - r}{\sigma^2}(X^*(t) + g(t)),$$

which we recognize as (3.24).

Finally, plug in (3.22)–(3.24) into (3.4) to obtain

$$\begin{aligned} d(X^*(t) + g(t)) = &\left(r + \mu(t) + \frac{1}{1-\gamma}\left(\frac{\alpha - r}{\sigma}\right)^2 - \left(1 + \mu(t)K_1^{\frac{1}{1-\gamma}}\right)\frac{1}{f(t)}\right)(X^*(t) + g(t))dt \\ &+ \frac{1}{1-\gamma}\frac{\alpha - r}{\sigma}(X^*(t) + g(t))dW(t). \end{aligned} \tag{3.36}$$

We get the solution

$$\begin{aligned} X^*(t) + g(t) = (x_0 + g(0))\exp\Bigg\{ &\int_0^t \left(r + \mu(s) + \frac{\frac{1}{2} - \gamma}{(1-\gamma)^2}\left(\frac{\alpha - r}{\sigma}\right)^2 - \frac{1 + \mu(s)K_1^{\frac{1}{1-\gamma}}}{f(s)}\right) ds \\ &+ \frac{1}{1-\gamma}\frac{\alpha - r}{\sigma} W(t)\Bigg\}. \end{aligned} \tag{3.37}$$

In particular, since f is bounded away from zero, $\forall t \in [0, T]$, we have that $X^*(t)$, $t \in [0, T]$, is well-defined, and clearly (3.9) is fulfilled. Moreover, since $X^* + g$ is lognormally distributed one can easily show that

$$E^Q\left[\int_0^T (\theta^*)^2 dt\right] < \infty,$$

which ensures that (3.10) is fulfilled. We conclude that the candidate optimal strategy (c^*, D^*, θ^*) given by (3.22)–(3.24) is admissible. Since x_0 and $g(0)$ cannot both be zero we get, from (3.22), (3.23) and (3.37), that $c^*(t), D^*(t) > 0$, $t \in [0, T]$, and $X^*(T) > 0$, and we note that the additional condition (3.16) is fulfilled, i.e. (c^*, θ^*, D^*) is feasible. We conclude that (c^*, θ^*, D^*) is the optimal strategy for the unrestricted problem (3.15). □

3.3.3 Comments on the unrestricted solution

The function g defined by (3.25) is referred to as the *actuarial present value of future labor income.* It is actuarial practice to define the reserve as a conditional expected present value of future payments. Isolating X in (3.11) leaves us with such an expression for the reserve. We get, $\forall t \in [0, T]$,

$$X(t) = E^Q\left[\int_t^T e^{-\int_t^s (r + \mu(y))dy}(c(s) - \ell(s) + \mu(s)D(s))ds + e^{-\int_t^T (r+\mu(y))dy} X(T)\,\middle|\, \mathcal{F}^W(t)\right].$$

We call the sum of the reserve and the actuarial present value of future labor income, given by (3.37), the *total reserve*. By (3.24) we have that the optimal amount invested in the stock is, comparable to Merton (1971), proportional to the total reserve and to the Sharpe ratio of the market, and dependent on the policyholder's risk aversion. The optimal consumption rate and the optimal life insurance strategy, given by (3.22) and (3.23), equals the total reserve divided by the deterministic function f given by (3.26), and for the optimal life insurance, multiplied by the weight factor $K_1^{1/(1-\gamma)}$. The function f can be interpreted as a subjective value of a unit consumption rate until terminal time T or time of death, whatever occurs first, plus a subjective actuarial present value of an endowment insurance paying $K_1^{1/(1-\gamma)}$ upon death before time T and $K_2^{1/(1-\gamma)}$ upon survival until time T. The value of f is subjective since the time dependent function $\widetilde{r}$ given by (3.27), used for discounting in the expression of f, depends on the policyholder's risk aversion and time preferences. Naturally, we conclude that the optimal consumption strategy is decreasing in K_1 and K_2. That is, a policyholder to whom bequest and/or terminal pension are more important will optimally consume less. We also observe, that a more impatient policyholder, i.e. a policyholder with a greater time preference parameter β, optimally consumes more and buys more life insurance than a less patient policyholder. The optimal consumption and life insurance strategy depends in a much more complex way on the risk aversion parameter γ. By use of (3.36) we get, as also obtained in Steffensen and Kraft (2008), that, $\forall t \in [0,T]$,

$$\begin{aligned} dc^*(t) &= \frac{d(X^*(t)+g(t))}{f(t)} - \frac{X^*(t)+g(t)}{f(t)^2} f'(t)dt \\ &= \left(\frac{r-\beta(t)+\left(\frac{\alpha-r}{\sigma}\right)^2}{1-\gamma} + \frac{1}{2}\frac{\gamma}{(1-\gamma)^2}\left(\frac{\alpha-r}{\sigma}\right)^2 \right) c^*(t)dt + \frac{1}{1-\gamma}\frac{\alpha-r}{\sigma} c^*(t)dW(t). \end{aligned}$$

We get, $\forall t \in [0,T]$, the solutions

$$\begin{aligned} c^*(t) &= c^*(0)\exp\left\{ \int_0^t \left(\frac{r-\beta(s)+\frac{1}{2}\left(\frac{\alpha-r}{\sigma}\right)^2}{1-\gamma} \right) ds + \frac{1}{1-\gamma}\frac{\alpha-r}{\sigma}W(t) \right\}, \\ D^*(t) &= D^*(0)\exp\left\{ \int_0^t \left(\frac{r-\beta(s)+\frac{1}{2}\left(\frac{\alpha-r}{\sigma}\right)^2}{1-\gamma} \right) ds + \frac{1}{1-\gamma}\frac{\alpha-r}{\sigma}W(t) \right\}. \end{aligned}$$

Note that $c^*(t)$ and $D^*(t)$ do not depend on $\mu(t)$. The optimal consumption and life insurance strategy depend only on μ through $f(0)$ and $g(0)$ (through $c^*(0)$ and $D^*(0)$), i.e. on the overall level of μ. Further discussions of the qualitative behavior of the optimal controls are found in Pliska and Ye (2007).

3.4 The restricted pension problem

From (3.37) we have for the unrestricted optimal solution that

$$X^*(t) > -g(t),\ \forall t \in [0,T],$$

where $g(t)$ is the time-t actuarial value of future labor income given by (3.25). Allowing for a negative reserve is in conflict with the usual constraint in life insurance that the reserve should be non-negative at any time. This should be the case since, in practice, at any time the policyholder holds the right to stop paying pension contributions. We say that the policyholder holds the right to surrender. This serves to separate, institutionally, the pension business from the lending business. For further discussion of this issue see the Introduction, Section 3.1. As pointed out by Nielsen and Steffensen (2008), solving the unrestricted pension problem (3.15) we assume that the policyholder sticks to the pension contribution plan stipulated in the policy.

The interpretation is that in the unrestricted pension problem given by (3.15) the policyholder decides initially on a consumption strategy, together with a life insurance strategy and an investment strategy, and is thereafter obligated to follow that specific strategy, i.e. he does not hold the right to surrender at any time.

The purpose of this paper is to find the optimal consumption, investment and life insurance strategy when the reserve is restricted to fulfill a capital guarantee at any point in time. As discussed, a capital guarantee of size zero is simply motivated by the policyholder's right to surrender at any time during the saving period. Often pension companies offer a minimum constant rate guarantee to the policyholders. In fact, such a guarantee is often found to be mandatory in compulsory (employer) pension schemes. In that case, the pension company guarantees a minimum rate of return on all pension contributions. Of course, no pension company can guarantee a constant minimum rate of return larger than the risk free short rate plus the objective mortality rate[4] since such a guarantee cannot be fulfilled with certainty. To cover a hole class of capital guarantee problems, including the two cases discussed, we allow the capital guarantee restriction to depend on the past in a very general way. Note that since future pension contributions may be stochastic, the future capital guarantee may becomes stochastic. For a comprehensive discussion and motivation of the problem see the Introduction, Section 3.1. More specific, consider the problem given by the indirect utility function

$$\sup_{(c,\theta,D)\in\mathcal{A}'} E\left[\int_0^T e^{-\int_0^t(\beta(s)+\mu(s))ds}\Big(u(c(t))+K_1\mu(t)u(D(t))\Big)dt+K_2e^{-\int_0^T(\beta(s)+\mu(s))ds}u(X(T))\right], \tag{3.38}$$

under the capital guarantee restriction

$$X(t)\geq k(t,Z(t)),\ \forall t\in[0,T], \tag{3.39}$$

where $Z(t):=\int_0^t h(s,X(s))ds$, and k and h are deterministic functions. The two guarantees discussed above are covered by

$$k(t,z)=0, \tag{3.40}$$

and

$$k(t,z)=x_0e^{\int_0^t\left(r^{(g)}+\mu(s)\right)ds}+e^{\int_0^t\left(r^{(g)}+\mu(s)\right)ds}z, \tag{3.41}$$

with $h(s,x)=e^{-\int_0^s\left(r^{(g)}+\mu(y)\right)dy}\left(\ell(s)-c(s,x)-\mu(s)D(s,x)\right)$, where $r^{(g)}\leq r$ is the minimum rate of return guarantee excess of the objective mortality rate μ. Note that we in the expression of h have indicated the possible x dependence in the consumption and life insurance strategy. We can write (3.41) as

$$k(t,Z(t))=x_0e^{\int_0^t\left(r^{(g)}+\mu(y)\right)dy}+\int_0^t e^{\int_s^t\left(r^{(g)}+\mu(y)\right)dy}\left[\ell(s)-c(s)-\mu(s)D(s)\right]ds. \tag{3.42}$$

To clarify things, one could preferably compare (3.42) with the retrospective reserve expression given by (3.5).

The solution to the restricted capital guarantee problem given by (3.38) and (3.39) turns out to be of the Option Based Portfolio Insurance (OBPI) form. OBPI strategies are also obtained in El-Karoui et al. (2005) and Kronborg (2011), and for a European capital constraint in Teplá (2001). The big differences, compared to El-Karoui et al. (2005) and Kronborg (2011), are the

[4] Remember that the mortality rate is given by a deterministic function, i.e. the pension company has no risk concerning life times.

introduction of life insurance and the possible *stochastic* capital guarantee restriction. In addition, El-Karoui et al. (2005) does not include consumption and labor income. Combining classic results from the American option theory in a Black-Scholes market and the OBPI construction technique we are able to solve the restricted optimization problem given by (3.38) and (3.39).

Before we present the results we need some notation: Still, denote by X^*, c^*, θ^* and D^* the optimal unrestricted reserve and the unrestricted optimal consumption, investment and life insurance strategy derived in Subsection 3.3.2, respectively. Let $Y^*(t) := X^*(t) + g(t)$ denote the optimal unrestricted total reserve. By (3.36) and (3.7) we have that

$$dY^*(t) = \left(r + \mu(t) - \left(1 + \mu(t)K_1^{\frac{1}{1-\gamma}}\right)\frac{1}{f(t)}\right)Y^*(t)dt + \frac{1}{1-\gamma}\frac{\alpha - r}{\sigma}Y^*(t)dW^Q(t),\ t \in (0,T), \tag{3.43}$$

$$Y^*(0) = y_0,$$

where $y_0 := x_0 + g(0)$. Denote by $P^a_{y,z}(t,T,k+g)$ the time-t value of an American put option with strike price $k(s,Z(s)) + g(s), \forall s \in [t,T]$, where $Z(t) = z$, and maturity T written on a portfolio Y, where Y is the solution to (3.43) which *equals y at time t*. By definition the price of such a put option is given, $\forall t \in [0,T]$, by

$$P^a_{y,z}(t,T,k+g) := \sup_{\tau \in \mathcal{T}_{t,T}} E^Q\left[e^{-\int_t^\tau (r+\mu(s))ds}(k(\tau,Z(\tau)) + g(\tau) - Y(\tau))^+ \middle| Y(t) = y, Z(t) = z\right],$$

where $\mathcal{T}_{t,T}$ is the set of stopping times taking values in the interval $[t,T]$.

Theorem 3.4.1. *Consider the strategy* $\left(\widehat{c},\overline{\theta},\widehat{D}\right)$ *given by,* $\forall t \in [0,T]$,

$$\widehat{c}(t) := \lambda(t,Z(t))c^*(t) = \frac{\lambda(t,Z(t))Y^*(t)}{f(t)}, \tag{3.44}$$

$$\widehat{D}(t) := \lambda(t,Z(t))D^*(t) = \frac{\lambda(t,Z(t))Y^*(t)}{f(t)}K_1^{\frac{1}{1-\gamma}}, \tag{3.45}$$

$$\overline{\theta}(t) := \lambda(t,Z(t))\theta^*(t) = \frac{1}{1-\gamma}\frac{\alpha - r}{\sigma^2}\lambda(t,Z(t))Y^*(t), \tag{3.46}$$

combined with a position in an American put option written on the portfolio $(\lambda(t)Y^*(t))_{s\in[t,T]}$ *with strike price* $k(s,Z(s)) + g(s), \forall s \in [t,T]$, *and maturity* T. *Here* λ *is the increasing function defined by*

$$\lambda(t,Z(t)) = \lambda(0,z_0) \vee \sup_{s\le t}\left(\frac{b(s,Z(s))}{Y^*(s)}\right), \tag{3.47}$$

with $\lambda(0,z_0)$ *determined by*

$$\lambda(0,z_0)(x_0 + g(0)) + P^a_{\lambda(0,z_0)y_0,z_0}(0,T,k+g) - g(0) = x_0, \tag{3.48}$$

and b *being the optimal exercise boundary for the American put option given by*

$$b(t,z) := \sup\left\{y : P^a_{y,z}(t,T,k+g) = (k(t,z) + g(t) - y)^+\right\}. \tag{3.49}$$

We have that

- *The strategy is optimal for the restricted capital guarantee problem given by (3.38) and (3.39).*

Remark 3.4.1. By (3.37) the optimal unrestricted total reserve, Y^*, is proportional to the initial total reserve y_0. Recall also that the optimal unrestricted consumption, life insurance and investment strategy, given by (3.22)–(3.24), are proportional to the total reserve. It should then be clear that starting out with reduced initial total reserve equal to $\lambda(0, z_0)(x_0 + g(0))$ and following the strategy $(\widehat{c}, \overline{\theta}, \widehat{D})$ given by (3.44)–(3.49) results in a total reserve equal to $\lambda(t, Z(t))Y^*(t), \forall t \in [0, T]$. In total we get that the optimal restricted reserve becomes

$$\widehat{X}^{(\lambda)}(t) := \lambda(t, Z(t))Y^*(t) + P^a_{\lambda Y^*,Z}(t, T, k + g) - g(t), \ t \in [0, T], \tag{3.50}$$

i.e. the optimal restricted reserve consists of a portfolio, an insurance on that portfolio and a loan equal to the actuarial value of future labor income.

Remark 3.4.2. The specific American put option is likely not to be sold in the market, but since the market is complete and frictionless such options can be replicated dynamically by a Delta-hedge, i.e. the optimal investment strategy can be written as

$$\widehat{\theta}(t) = \left(1 + \frac{\partial}{\partial y} P^a_{\lambda Y^*,Z}(t, T, k + g)\right) \overline{\theta}. \tag{3.51}$$

Since the put option has an opposite exposure to changes in the underlying portfolio $(\lambda(t)Y^*(t))_{t\in[0,T]}$ we note that the total amount of money to invest in the risky asset, S, becomes smaller when we introduce an American capital guarantee to the control problem (3.15). One should notice, that at the boundary no risk is taken, i.e. the position in the American put option offsets the position in the underlying portfolio such that the total exposure to changes in the risky asset, S, becomes zero. Clearly, at the boundary it is optimal to consume a certain time-dependent part of the labor income, thereby leaving room for a risky position in S immediately after hitting the capital boundary.

Proof of Theorem 3.4.1. First, one should check that the strategy defined by Theorem 3.4.1 fulfills the capital guarantee restriction given by (3.39). Using that the value of an American put option is always greater than or equal to its inner value this is easily done. We get

$$\begin{aligned}
\widehat{X}^{(\lambda)}(t) &:= \lambda(t, Z(t))Y^*(t) + P^a_{\lambda Y^*,Z}(t, T, k + g) - g(t) \\
&\geq \lambda(t, Z(t))Y^*(t) + [k(t, Z(t)) + g(t) - \lambda(t, Z(t))Y^*(t)]^+ - g(t) \\
&\geq k(t, Z(t)),
\end{aligned} \tag{3.52}$$

Second, one should check that the strategy defined by Theorem 3.4.1 is admissible. It turns out that the rather complex strategy $\left(\widehat{c}, \widehat{D}, \widehat{X}^{(\lambda)}\right)$ involving the stochastic increasing function λ is the natural way to make the OBPI admissible. The heuristic argument is as follows. Consider the strategy $\left(\widehat{c}, \widehat{D}, X^{(\lambda)}\right)$ where $\lambda(t, Z(t)) = \lambda_0, \forall t \in [0, T]$, i.e. where we do not adjust λ. By (3.52) such a strategy clearly fulfills the capital guarantee restriction. The strategy corresponds to, in addition to follow the consumption, life insurance and investment strategy given by (3.44)–(3.46), to hold the initially bought American put option to maturity. In some scenarios this is to throw away money since the American put option is not sold when it is optimal to do so, i.e. the strategy is not admissible (self-financing). The complicated strategy defined in Theorem 3.4.1 is exactly designed to avoid this. The function λ defined by (3.47) and (3.48) corresponds to selling the American put option and re-balancing the strategy whenever optimal to do so. For a comprehensive explanation of the nature of the OBPI strategies see Kronborg (2011). A rigorous proof showing that the strategy in Theorem 3.4.1 is admissible is given in Appendix 3.6.

To prove the optimality of the strategy given in Theorem 3.4.1 we consider an arbitrarily chosen feasible strategy (c, θ, D) with corresponding reserve process $(X(t)_{t\in[0,T]})$ satisfying

$X(0) = x_0$ and $X(t) \geq k(t, Z(t)), \forall t \in [0, T]$. Since u is a concave function we get that

$$\begin{aligned}
&\int_0^T e^{-\int_0^t (\beta(s)+\mu(s))ds} \left[u(c(t)) + K_1 \mu(t) u(D(t))\right] dt + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds} u(X(T)) \\
&- \left(\int_0^T e^{-\int_0^t (\beta(s)+\mu(s))ds} \left[u(\widehat{c}(t)) + K_1 \mu(t) u\left(\widehat{D}(t)\right)\right] dt + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds} u\left(\widehat{X}^{(\lambda)}(T)\right)\right) \\
&= \int_0^T e^{-\int_0^t (\beta(s)+\mu(s))ds} \left[u(c(t)) - u(\widehat{c}(t)) + K_1 \mu(t)\left(u(D(t)) - u\left(\widehat{D}(t)\right)\right)\right] dt \\
&\quad + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds} \left(u(X(T)) - u\left(\widehat{X}^{(\lambda)}(T)\right)\right) \\
&\leq \int_0^T e^{-\int_0^t (\beta(s)+\mu(s))ds} \left[u'(\widehat{c}(t))(c(t) - \widehat{c}(t)) + K_1 \mu(t) u'\left(\widehat{D}(t)\right)\left(D(t) - \widehat{D}(t)\right)\right] dt \\
&\quad + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds} u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&=: (*).
\end{aligned} \tag{3.53}$$

Since (c, θ, D) was arbitrarily chosen we simply end the proof by showing that $E[(*)] \leq 0$. By the CRRA property $u'(xy) = u'(x)u'(y)$ we have

$$u'(\widehat{c}(t))(c(t) - \widehat{c}(t)) = u'(\lambda(t, Z(t)))u'(c^*(t))(c(t) - \widehat{c}(t)), \tag{3.54}$$

$$u'\left(\widehat{D}(t)\right)\left(D(t) - \widehat{D}(t)\right) = u'(\lambda(t, Z(t)))u'(D^*(t))\left(D(t) - \widehat{D}(t)\right). \tag{3.55}$$

Observe that since $Y^*(T) = X^*(T)$ the value of the terminal capital pension becomes

$$\begin{aligned}
\widehat{X}^{(\lambda)}(T) &= \lambda(T, Z(T))X^*(T) + [k(T, Z(T)) - \lambda(T, Z(T))X^*(T)]^+ \\
&= \max[\lambda(T, Z(T))X^*(T), k(T, Z(T))].
\end{aligned} \tag{3.56}$$

By use of (3.56) and using that u' is a decreasing function we get that

$$\begin{aligned}
&u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&= \min[u'(\lambda(T, Z(T)))u'(X^*(T)), u'(k(T, Z(T)))]\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&= u'(\lambda(T, Z(T)))u'(X^*(T))\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \\
&\quad - [u'(\lambda(T, Z(T)))u'(X^*(T)) - u'(k(T), Z(T))]^+ \left(X(T) - k(T, Z(T))\right),
\end{aligned}$$

where the last equality is established by using that $\widehat{X}^{(\lambda)}(T) = k(T, Z(T))$ on the set $\{(T, \omega) : u'(\lambda(T, Z(T))X^*(T)) \geq u'(k(T, Z(T)))\}$. Since by assumption $X(t) \geq k(t, Z(t)), \forall t \in [0, T]$, we conclude that

$$u'\left(\widehat{X}^{(\lambda)}(T)\right)\left(X(T) - \widehat{X}^{(\lambda)}(T)\right) \leq u'(\lambda(T, Z(T)))u'(X^*(T))\left(X(T) - \widehat{X}^{(\lambda)}(T)\right). \tag{3.57}$$

Now insert (3.54), (3.55) and (3.57) and then (3.30)–(3.32) into (3.53) to get

$$\begin{aligned}
E[(*)] \leq E\Bigg[\int_0^T & e^{-\int_0^t (\beta(s)+\mu(s))ds} \Big[u'(\lambda(t, Z(t)))u'(c^*(t))\,(c(t) - \widehat{c}(t)) \\
& + K_1 \mu(t) u'(\lambda(t, Z(t)))u'(D^*(t)) \left(D(t) - \widehat{D}(t) \right) \Big] dt \\
& + K_2 e^{-\int_0^T (\beta(s)+\mu(s))ds} u'(\lambda(T, Z(T)))u'(X^*(T)) \left(X(T) - \widehat{X}^{(\lambda)}(T) \right) \Bigg] \\
= \xi^* E^Q \Bigg[\int_0^T & e^{-\int_0^t (r+\mu(s))ds} u'(\lambda(t, Z(t))) \left((c(t) - \widehat{c}(t)) + \mu(t) \left(D(t) - \widehat{D}(t) \right) \right) dt \\
& + e^{-\int_0^T (r+\mu(s))ds} u'(\lambda(T, Z(T))) \left(X(T) - \widehat{X}^{(\lambda)}(T) \right) \Bigg].
\end{aligned}$$

Since $u'(\lambda(t, Z(t)))$ is a decreasing function[5] we can use the integration by parts formula to get

$$\begin{aligned}
E[(*)] \leq \xi^* \Bigg(& E^Q \Bigg[\underbrace{\int_0^T e^{-\int_0^t (r+\mu(s))ds} u'(\lambda(t, Z(t))) \left((c(t) - \widehat{c}(t)) + \mu(t) \left(D(t) - \widehat{D}(t) \right) \right) dt}_{(\star)} \\
& + \underbrace{\int_0^T u'(\lambda(t, Z(t))) d \left(e^{-\int_0^t (r+\mu(s))ds} \left(X(t) - \widehat{X}^{(\lambda)}(t) \right) \right)}_{(\star\star)} \Bigg] \\
& + E^Q \Bigg[\underbrace{\int_0^T e^{-\int_0^t (r+\mu(s))ds} \left(X(t) - \widehat{X}^{(\lambda)}(t) \right) du'(\lambda(t, Z(t)))}_{(\star\star\star)} \Bigg] \Bigg).
\end{aligned} \tag{3.58}$$

The third term in (3.58) can be rewritten as

$$\begin{aligned}
E^Q[(\star\star\star)] = & E^Q \left[\int_0^T e^{-\int_0^t (r+\mu(s))ds} (X(t) - k(t, Z(t))) du'(\lambda(t, Z(t))) \right] \\
& + E^Q \left[\int_0^T e^{-\int_0^t (r+\mu(s))ds} \left(k(t, Z(t)) - \widehat{X}^{(\lambda)}(t) \right) du'(\lambda(t, Z(t))) \right].
\end{aligned}$$

The first term is non-positive since per definition $X(t) \geq k(t, Z(t)), \forall t \in [0, T]$, and $du'(\lambda(t, Z(t))) \leq 0, \forall t \in [0, T]$ (u' is decreasing and λ is increasing). The second term equals zero since $du'(\lambda(t, Z(t))) \neq 0$ only on the set $\{(t, \omega) : \widehat{X}^{(\lambda)}(t) = k(t, Z(t))\}$. We conclude that $E^Q[(\star\star\star)] \leq 0$. The two first terms of (3.58) can be rewritten as

$$E^Q[(\star) + (\star\star)] = E^Q \left[\int_0^T u'(\lambda(t, Z(t))) dM_1(t) \right] - E^Q \left[\int_0^T u'(\lambda(t, Z(t))) dM_2(t) \right], \tag{3.59}$$

where

$$\begin{aligned}
M_1(t) &:= \int_0^t e^{-\int_0^s (r+\mu(y))dy} (c(s) + \mu(s)D(s) - \ell(s)) ds + e^{-\int_0^t (r+\mu(y))dy} X(t), \\
M_2(t) &:= \int_0^t e^{-\int_0^s (r+\mu(y))dy} \left(\widehat{c}(s) + \mu(s)\widehat{D}(s) - \ell(s) \right) ds + e^{-\int_0^t (r+\mu(y))dy} \widehat{X}^{(\lambda)}(t).
\end{aligned}$$

[5]This ensures the that the stochastic integral in (3.58) is well-defined.

Since both strategies are admissible we have by (3.13) that M_1 and M_2 are martingales under the equivalent measure Q. Since $u'(\lambda(t, Z(t))) \leq u'(\lambda(0, z_0), \forall t \in [0, T]$, we get that

$$E^Q[(\star) + (\star\star)] = 0.$$

Finally, we can conclude that

$$E[(*)] = E^Q[(\star) + (\star\star)] + E^Q[(\star\star\star)] \leq 0.$$

$\square$

3.4.1 Numerical illustrations

In this subsection we illustrate the optimal strategy presented in Theorem 3.4.1 for the case of a minimum rate of return guarantee given by (3.42) with $r^{(g)} = \frac{1}{2}r$. The parameter values used for the simulations are, for the Black-Scholes market, $r = 0.01885$, $\alpha = 0.05885$ and $\sigma = 0.2$. The rather low risk free short rate and expected stock rate return should be interpreted as *inflation adjusted* parameters (subtracted 2.115 percent which is the average inflation in Denmark over the last 20 years). Adjusting for inflation allow us to directly compare the size of the terminal pension with past reserve values and labor income. The personal preference towards risk is set to $\gamma = -4$ which is in accordance with the popular choice to have 20 percent of the reserve invested in stocks after retirement (we have $\frac{1}{1-\gamma}\frac{\alpha - r}{\sigma^2} = 0.2$). The time preference is set to the natural value $\beta = r$, but one could of cause equally well had chosen the time preference parameter to be smaller or greater than the risk free short rate. However, it seems natural to have $\beta \geq r$ since this assures that one cannot, in a risk free way, obtain a greater amount of utility by simply investing in the risk free short rate and then consume at a later point in time. Finally, we have chosen to model the uncertain life time by a Gompertz-Makeham hazard rate, $u(x) = \exp((x - m)/b)/b$, with modal value and scale parameter as in Milevsky and Young (2007); $(m, b) = (88.18, 10.5)$.

The individual we consider has just turned 50 years old, has a reserve of size 200000 euro and a constant future labor income of $\ell(t) = 30000,\ t \in [50, T]$, with terminal time $T = 65$. The weight factor K_2 in (3.38) is calculated by use of (3.21), but since it seems natural that the reserve will also be invested after retirement we have chosen to use $\bar{r} = 0.2\alpha + 0.8r = 0.02685$ (since $\frac{1}{1-\gamma}\frac{\alpha - r}{\sigma^2} = 0.2$) for discounting in the expression of $\overline{a}(T)$. The weight factor K_1 in (3.38) is set to $K_1 = \int_0^{10} e^{-\int_t^{t+x}\beta(y)dy}dx \left(\overline{b}^{(10)}\right)^{-\gamma}$ with $\overline{b}^{(10)} = \int_0^{10} e^{-\int_0^s r dy} ds$. One can realize that this is in accordance with the standard Danish choice of a 10 years period certain pension to the inheritor starting at the time of death (if death occurs before terminal time T).

The remaining 15 years of optimization has been discretizied equidistantly into quarterly time points, i.e. into 60 time points. By this we mean that the guarantee is only valid quarterly and the American put option has been evaluated as a Bermuda put option with quarterly exercise dates. Likewise, λ and therefore the optimal consumption, investment and life insurance strategy has only been changed quarterly. Saying that, it is important to stress that between the quarterly time points, consumption, investments and bequest have been re-balanced continuously[6].

It seems natural to compare the restricted solution with the unrestricted counterpart. Figure 3.1 illustrates how the restricted optimal reserve always is greater or equal to the guarantee while the unrestricted reserve sometimes is greater and sometimes smaller than the 'fictive' guarantee[7]. In fact we conclude that ignoring a guarantee requirement, and behaving optimal from the no guarantee point of view, will very likely result in a reserve smaller than the guarantee.

[6] We have used (3.37).

[7] To be clear, there is of cause no guarantee belonging to the unrestricted reserve, but we can calculate the guarantee anyway and then compare the unrestricted reserve with the guarantee.

Note that the restricted strategy can, for certain outcomes, outperform the unrestricted strategy both in terms of accumulated utility and terminal pension. In fact we have a 24 percent probability that the restricted investor will obtain more utility over the interval of optimization than the unrestricted investor.

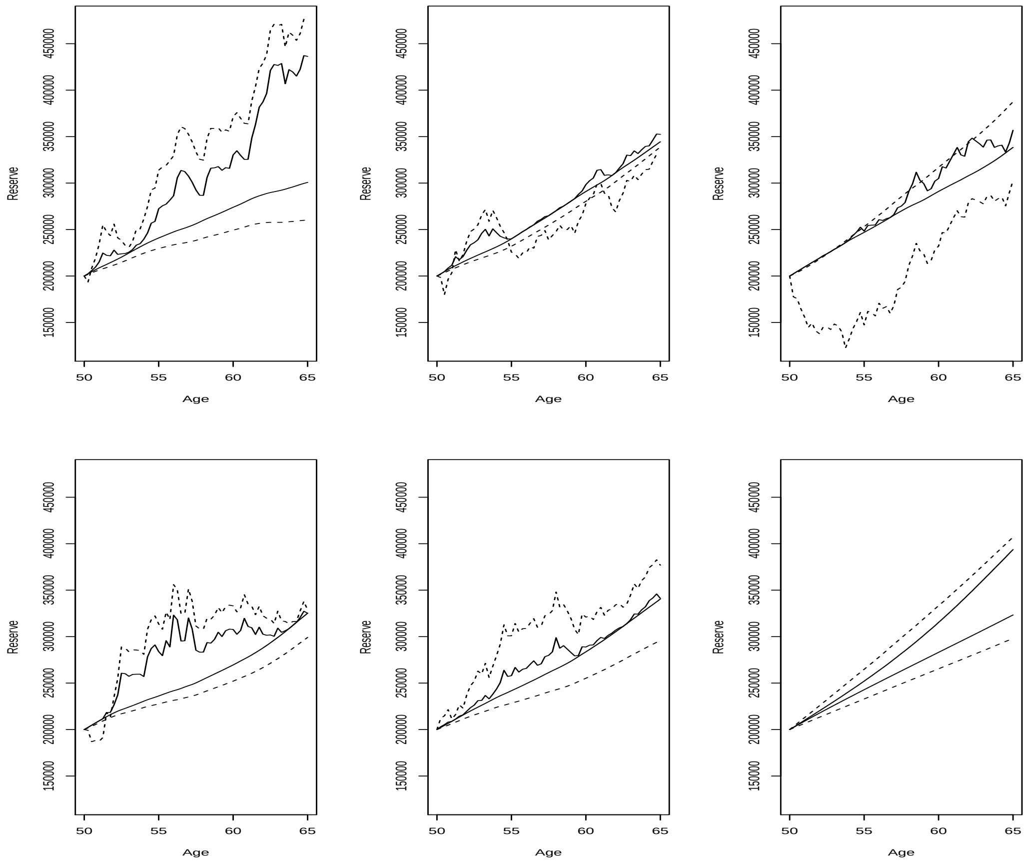

Figure 3.1: The restricted reserve (solid curves) and the unrestricted reserve (dashed curves) together with the corresponding guarantee (solid smooth curves) and fictive guarantee (dashed smooth curves), respectively. The lower right plot indicates the average scenario.

In Table 3.1 we see that the restricted optimal terminal pension has a slightly smaller median than the unrestricted one, but at the same time the restricted terminal pension has a much more narrow distribution. We also see that the pension saver with the minimum rate of return guarantee will, in contrast to the one without, more or less avoid ending up with a very small terminal pension. We want to stress that one cannot use Table 3.1 to compare the performances of two strategies since money spent before the terminal time T is not taking into account.

The actuarial value of future labor income is $g(50) = 380387$ euro leaving the 50 years old investor with a total reserve of size $Y(50) = 580387$. By simulation we calculate $\lambda(50) = 0.9041$ meaning that initially about 90 percent of the total reserve is invested in the optimal unrestricted portfolio and the remaining 10 percent is used to buy the American put option as an insurance against downfalls in the portfolio. Figure 3.2 shows how lambda increases at the time points where the insurance is in use (where it is optimal to sell the put option and re-calibrate the OBPI strategy accordingly to the new lambda value).

Fractal	0.001%	0.01%	2.5%	25%	median	75%	97.5%	99.99 %	99.999%
Restricted	10.21	10.58	11.33	12.18	13.12	14.30	17.08	22.03	23.60
Unrestricted	6.10	7.53	10.02	12.22	13.57	15.09	18.39	24.09	25.91

Table 3.1: Reserve at retirement expressed in numbers of yearly labor income (the terminal reserve divided by the labor income rate). Chosen fractals for both the restricted and unrestricted case are presented.

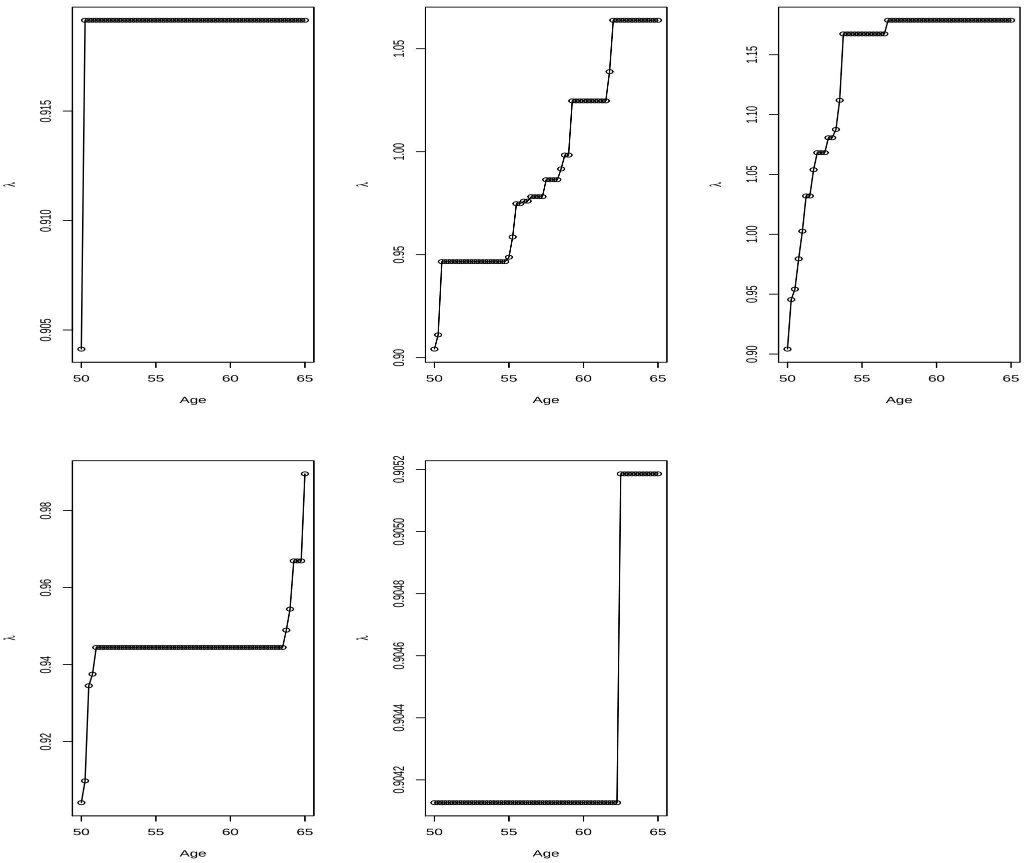

Figure 3.2: The lambda process corresponding to the reserve processes from Figure 3.1.

The optimal investment strategy is given by (3.51) and (3.46). Figure 3.3 shows the optimal fraction of the total reserve invested in stocks. The main observation to do is to note how wide the 95 percent confidence interval is. Even in the case of very big investment returns it takes about seven years (about half the optimization interval) before the insurance part becomes negligible and the restricted investor starts to invest (almost) as aggressively as the unrestricted investor. In the case of very low investment returns we see that the restricted investor gets stocked with the guarantee and cannot afford to take very much risk.

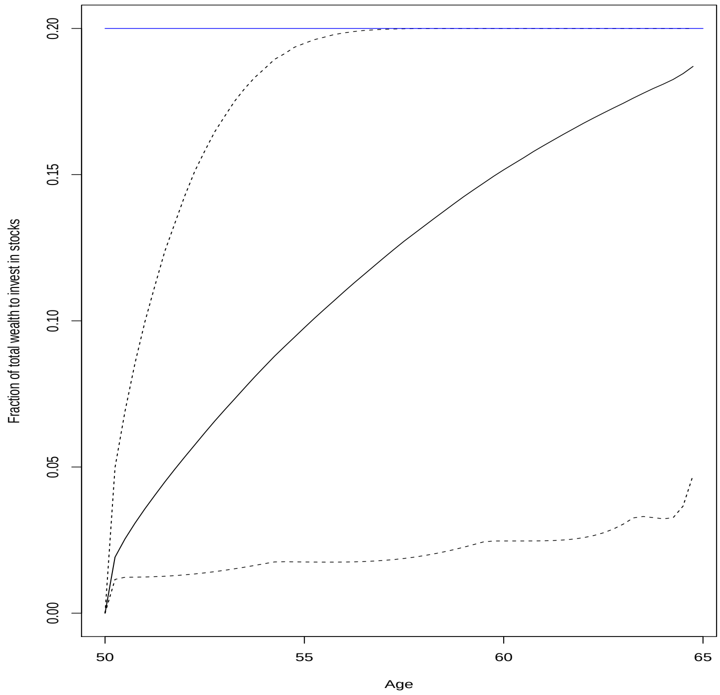

Figure 3.3: The restricted investment strategy presented as the fraction of the total reserve invested in stocks. The average scenario (solid curve) as well as the 95 percent confidence interval (dotted curves) are indicated together with the unrestricted case ($\pi = 0.2$).

At last we have in Figure 3.4 illustrated the optimal life insurance and pension contributions as a fraction of the reserve and labor income, respectively. We see that at the age 50-55 the investor stops protecting parts of his future labor income and buys annuities instead. One should note how narrow the 95 percent confidence interval is in the restricted case compared to the unrestricted case. We also see that the restricted investor saves more money than the unrestricted investor. In other words: The introduction of a minimum rate of return guarantee seems to imply bigger pensions contributions. This is somehow surprising from a logical point of view but this is simply a consequence of the OBPI strategy.

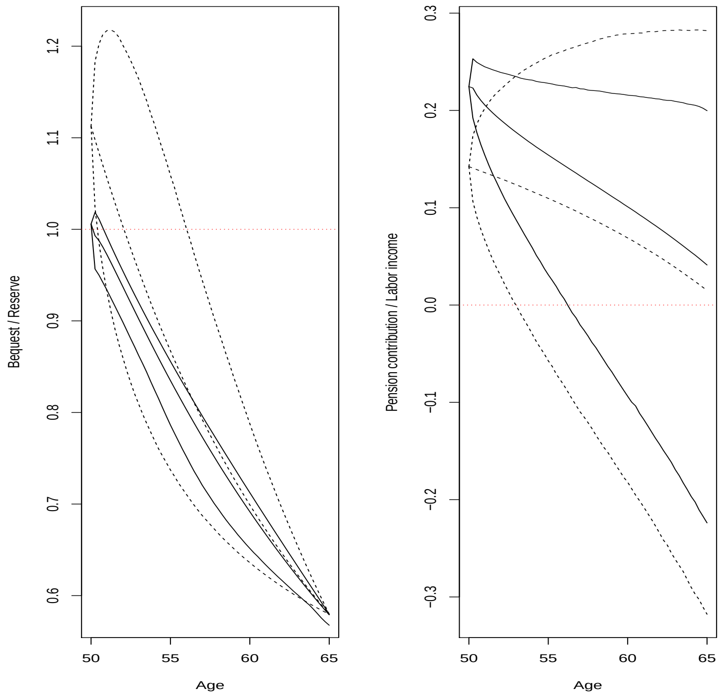

Figure 3.4: Left window: The restricted (solid curves) and unrestricted (dashed curves) life insurance strategy presented, for the average scenario, as the fraction of the reserve to be payed out upon death. The 95 percent confidence intervals are indicated. Right window: The restricted (solid curve) and unrestricted (dashed curve) pension contribution strategy presented, for the average scenario, as a fraction of the labor income rate. The 95 percent confidence intervals are indicated.

3.5 Conclusion

This paper solves the classic consumption, investment and life insurance control problem in the present of an American capital guarantee. Pension saving products often come with a minimum rate of return guarantee, which seems to be quiet a popular feature. Therefore the main focus, and the entire numerical section, has been on the solution for this specific type of capital guarantee. By use of cleaver chosen weight factors concerning the utility from consumption vs bequest and consumption vs pension the problem seems to be rather close to the real life question of how to save for retirement. The solution is an option based portfolio insurance (OBPI) strategy which is well know from the portfolio selection literature. However, to the authors knowledge this paper applies for the first time the OBPI approach to the saving for retirement problem. The encouraging realization offered by this paper is that the OBPI approach can handle the inclusion of a life insurance market as well as the introduction of a non-deterministic American guarantee, thereby allowing the capital guarantee to depend upon the size of the pension contributions.

Acknowledgments

The authors wish to thank an anonymous referee for valuable comments and Adjoint Professor Søren Fiig Jarner for valuable discussions and comments.

3.6 Appendix

Proof of admissibility. To prove that the strategy in Theorem 3.4.1 is admissible we first recall some basis properties of an American put option in a Black-Scholes market (see e.g. Karatzas and Shreve (1998)). We have

$$\begin{aligned} P^a_{y,z}(t,T,k+g) &= k(t,z)+g(t)-y, && \forall (t,y,z)\in\mathcal{C}^c,\\ \mathcal{L}P^a_{y,z}(t,T,k+g) &= (r+\mu(t))P^a_{y,z}(t,T,k+g), && \forall (t,y,z)\in\mathcal{C},\\ \frac{\partial}{\partial y}P^a_{y,z}(t,T,k+g) &= -1, && \forall (t,y,z)\in\mathcal{C}^c, \end{aligned}$$

where (see (3.43))

$$\mathcal{L} := \frac{\partial}{\partial t} + \left(r+\mu-\left(1+\mu(t)K_1^{\frac{1}{1-\gamma}}\right)\frac{1}{f(t)}\right)y\frac{\partial}{\partial y} + h\frac{\partial}{\partial z} + \frac{1}{2}\left(\frac{1}{1-\gamma}\frac{\alpha-r}{\sigma}y\right)^2\frac{\partial^2}{\partial y^2},$$

and

$$\mathcal{C} := \{(t,y,z) : P^a_{y,z}(t,T,k+g) > (k(t,z)+g(t)-y)^+\},$$

defines the continuation region. By $\mathcal{C}^c$ we mean the complementary of $\mathcal{C}$, i.e. the stopping region. The continuation region can be described via the exercise boundary b given by (3.49). We get

$$\mathcal{C} = \{(t,y,z) : y > b(t,z)\}.$$

Introducing the function A by

$$A(t,y,z) := y + P^a_{y,z}(t,T,k+g) - g(t),$$

we can write (3.50) as

$$X^{(\lambda)}(t) = A(t,\lambda(t,Z(t))Y^*(t),Z(t)).$$

From the properties of $P^a_{y,z}(t,T,k+g)$ we deduce that

$$\begin{aligned} A(t,y,z) &= k(t,z), && \forall (t,y,z)\in\mathcal{C}^c,\\ \mathcal{L}A(t,y,z) &= \left(r+\mu-\left(1+\mu(t)K_1^{\frac{1}{1-\gamma}}\right)\frac{1}{f(t)}\right)y + (r+\mu(t))P^a_{y,z}(t,T,k+g)\\ &\quad -(-\ell(t)+(r+\mu(t))g(t))\\ &= (r+\mu(t))A(t,y,z)+\ell(t)-\left(1+\mu(t)K_1^{\frac{1}{1-\gamma}}\right)\frac{y}{f(t)}, && \forall (t,y,z)\in\mathcal{C}, \end{aligned} \tag{3.60}$$

$$\mathcal{L}A(t,y,z) = \frac{\partial}{\partial t}k(t,z) + h(t,z)\frac{\partial}{\partial z}k(t,z), \qquad \forall (t,y,z)\in\mathcal{C}^c, \tag{3.61}$$

$$\frac{\partial}{\partial y}A(t,y,z) = 0. \qquad \forall (t,y,z)\in\mathcal{C}^c. \tag{3.62}$$

Observe that since $Y^*(t)$ is linear in its initial value, $\forall t\in[0,T]$, we have for a constant λ that $\lambda Y^*(t)$ has the same dynamics as $Y^*(t)$, $\forall t\in[0,T]$. We now get by use of Itô's formula, (3.60)–(3.61), (3.22)–(3.24), and the fact that λ increases only at the boundary, that[8]

$$\begin{aligned}
&dA(t, \lambda(t, Z(t))Y^*(t), Z(t)) \\
&= [dA(t, \lambda Y^*(t), Z(t))]_{|\lambda=\lambda(t,Z(t))} + Y^*(t)\frac{\partial}{\partial y}A(t, \lambda(t, Z(t))Y^*(t), Z(t))d\lambda(t, Z(t)) \\
&= \lambda(t, Z(t))\frac{\partial}{\partial y}A(t, \lambda(t, Z(t))Y^*(t), Z(t))\sigma\theta^*(t)dW^Q(t) \\
&\quad + \Big[(r + \mu(t))A(t, \lambda(t, Z(t))Y^*(t), Z(t)) \\
&\quad + \ell(t) - \lambda(t, Z(t))c^*(t) - \mu(t)\lambda(t, Z(t))D(t)^*\Big] 1_{(\lambda(t,Z(t))Y^*(t)>b(t,Z(t)))}dt \\
&\quad + \left[\frac{\partial}{\partial t}k(t, Z(t)) + h(t, Z(t))\frac{\partial}{\partial z}k(t, Z(t))\right] 1_{(\lambda(t,Z(t))Y^*(t)\leq b(t,Z(t)))}dt \\
&\quad + Y^*(t)\frac{\partial}{\partial y}A(t, \lambda(t, Z(t))Y^*(t), Z(t))1_{\big(\lambda(t,Z(t))Y^*(t)=b(t,Z(t))\big)}d\lambda(t, Z(t)).
\end{aligned}$$

Since by (3.62) $\frac{\partial}{\partial y}A(t, \lambda(t)Y^*(t), Z(t)) = 0$ on the set $\{(t, \omega) : \lambda(t, Z(t))Y^*(t) = b(t, Z(t))\}$ this reduces to

$$\begin{aligned}
&dA(t, \lambda(t, Z(t))Y^*(t), Z(t)) \\
&= \lambda(t, Z(t))\frac{\partial}{\partial y}A(t, \lambda(t, Z(t))Y^*(t), Z(t))\sigma\theta^*(t)dW^Q(t) \\
&\quad + [(r + \mu(t))A(t, \lambda(t, Z(t))Y^*(t), Z(t)) + \ell(t) - \lambda(t, Z(t))c^*(t) - \mu(t)\lambda(t, Z(t))D(t)^*]\, dt \\
&\quad + \Bigg[\frac{\partial}{\partial t}k(t, Z(t)) + h(t, Z(t))\frac{\partial}{\partial z}k(t, Z(t)) \\
&\quad - [(r + \mu(t))k(t, Z(t)) + \ell(t) - \lambda(t, Z(t))c^*(t) - \mu(t)\lambda(t, Z(t))D(t)^*]\Bigg] 1_{(\lambda(t,Z(t))Y^*(t)\leq b(t,Z(t)))}dt.
\end{aligned}$$

Finally, since $\{(t, \omega) : \lambda(t, Z(t))Y^*(t) \leq b(t, Z(t))\} = \left\{(t, \omega) : \lambda(t, Z(t)) = \frac{b(t,Z(t))}{Y^*(t)}\right\}$ has a zero $dt \otimes dP$-measure we conclude that

$$\begin{aligned}
&dA(t, \lambda(t, Z(t))Y^*(t), Z(t)) \\
&= [(r + \mu(t))A(t, \lambda(t, Z(t))Y^*(t), Z(t)) + \ell(t) - \lambda(t, Z(t))c^*(t) - \mu(t)\lambda(t, Z(t))D(t)^*]\, dt \\
&\quad + \lambda(t, Z(t))\frac{\partial}{\partial y}A(t, \lambda(t, Z(t))Y^*(t), Z(t))\sigma\theta^*(t)dW^Q(t),
\end{aligned}$$

i.e. by (3.14) the strategy is admissible. □

[8] $\frac{\partial}{\partial y}$ now means differentiating w.r.t. the second variable.

4. Inconsistent investment and consumption problems

Abstract: In a traditional Black-Scholes market we develop a verification theorem for a general class of investment and consumption problems where the standard dynamic programming principle does not hold. The theorem is an extension of the standard Hamilton-Jacobi-Bellman equation in the form of a system of non-linear differential equations. We derive the optimal investment and consumption strategy for a mean-variance investor without pre-commitment endowed with labor income. In the case of constant risk aversion it turns out that the optimal amount of money to invest in stocks is independent of wealth. The optimal consumption strategy is given as a deterministic bang-bang strategy. In order to have a more realistic model we allow the risk aversion to be time and state dependent. Of special interest is the case were the risk aversion is inversely proportional to present wealth plus the financial value of future labor income net of consumption. Using the verification theorem we give a detailed analysis of this problem. It turns out that the optimal amount of money to invest in stocks is given by a linear function of wealth plus the financial value of future labor income net of consumption. The optimal consumption strategy is again given as a deterministic bang-bang strategy. We also calculate, for a general time and state dependent risk aversion function, the optimal investment and consumption strategy for a mean-standard deviation investor without pre-commitment. In that case, it turns out that it is optimal to take no risk at all.

Keywords: Investment costs, risk aversion, stochastic control, indifferent compensation measure, certainty equivalents.

4.1 Introduction

The dynamic asset allocation problem for a portfolio investor searching to maximize the mean-variance objective

$$E_{t,x}\left[X^{\pi}(T)\right] - \frac{\gamma}{2} Var_{t,x}\left[X^{\pi}(T)\right], \tag{4.1}$$

for a constant γ, has in recent years been subject to numerous studies. The problem is non-standard in the sense that it cannot be formalized as a standard stochastic control problem,

$$\sup_{u} E_{t,x}\left[\int_t^T C(s, X^u(s), u(s))ds + \phi(X^u(T))\right], \tag{4.2}$$

for some functions C and ϕ. Therefore, the traditional dynamic programming approach does not apply directly. This is due to the lack of the iterated expectation property, and, consequently, we refer to such problems as *time inconsistent*. For every time inconsistent control problem we can fix an initial point and then solve the problem. The corresponding optimal control will at a later fixed point in time then not be optimal. We refer to this solution as the optimal *pre-commitment* control (for the mean-variance case see Korn (1997a) and Zhou and Li (2000)). The first to solve the problem (4.1) *without pre-commitment* were Basak and Chabakauri (2010). They present the problem in a quite general incomplete Wiener driven framework and by applying a so called total variance formula they obtain an extension of the classical Hamilton-Jacobi-Bellman equation for solving the problem. Björk and Murgoci (2009) extend the class of standard solvable problems (4.2) to the class of objectives

$$E_{t,x}\left[\int_t^T C(x,s,X^u(s),u(s))ds + \phi(x,X^u(T))\right] + G\big(x, E_{t,x}\left[X^u(T)\right]\big), \tag{4.3}$$

for some function G. They work in a general Markovian financial market having the results of Basak and Chabakauri (2010) as a special case. In (4.3) time inconsistency enters at two points: First, the present state x appears in C, ϕ and G, and second, the function G is allowed to be non-linear in the conditional expectation. Another work having Basak and Chabakauri (2010) as a special case is Kryger and Steffensen (2010). They analyze, in a classic Black-Scholes market, the class of problems given by the objectives

$$f\left(t,x,E_{t,x}\left[\phi_1(X^\pi(T))\right],\ldots,E\left[\phi_n(X^\pi(T))\right]\right), \tag{4.4}$$

where f is allowed to be a non-affine function of the expectation of the ϕ functions. One special example of interest only contained in (4.4) is the dynamic asset allocation problem for a portfolio investor with mean-standard deviation criteria. Kryger and Steffensen (2010) show that the optimal strategy derived for a mean-standard deviation investor without pre-commitment is to take no risk at all. The latest contribution to the literature treating mean-variance optimization problems without pre-commitment comes from Björk et al. (2012). They argue that the somehow unsatisfactory solution to (4.1), saying that the optimal amount to invest in stocks is constant, is due to the fact that the risk aversion parameter γ is constant. They solve the problem (4.1) for a general risk aversion function $\gamma(x)$ depending on present wealth and obtain for the special case $\gamma(x) = \gamma/x$ that the corresponding optimal amount invested in stocks is linear in wealth.

Working with inconsistent stochastic optimization problems without pre-commitment it might not be totally clear what we mean by an optimal control. This is well discussed in both Björk and Murgoci (2009) and Björk et al. (2012). They argue that the right thing to do is to study time inconsistency within a game theoretic framework and then look for a subgame perfect Nash equilibrium point for this game. This approach is first described in Strotz (1955), and the first to give a precise definition of the game theoretic equilibrium concept in continuous time were Ekeland and Lazrak (2006), and, Ekeland and Pirvu (2008). Conceptually, we attack the problems in the same manner.

Björk and Murgoci (2009) also take consumption into account. However, their preferences over consumption do not contribute to the inconsistency in the sense that the consumption term is just added as in a standard stochastic control problem, see (4.3) and (4.2). In this paper we introduce a new class of optimization problems. In these problems inconsistency also arises from taking a non-linear function of the expected (utility of) consumption. In addition we also allow for a capital injection in the form of a deterministic labor income. This leads to some mathematical difficulties but we manage to establish a verification theorem containing a Bellman-type set of differential equations for determination of the optimal strategies. Two concrete examples of economic interest covered by our approach are the mean-variance and the mean-standard deviation problems without pre-commitment including consumption and labor income. Those cases are analyzed in details in Section 4.3–4.5. One should recognize that we consider the problems for a general risk aversion function. More specific, we allow the risk aversion function to be both time and state dependent.

The structure of the paper is as follows: In Section 4.2 we present our formal model and the problems of interest. We discuss the concept of inconsistency, admissible strategies, and what we mean by an optimal strategy. Finally a verification theorem characterizing the solution to our class of problems is provided, leaving the proof to the Appendix. In Section 4.3 we derive the optimal consumption and investment strategy for a mean-variance investor without pre-commitment and constant risk aversion. In Section 4.4 we derive the optimal consumption and investment strategy for a mean-variance investor without pre-commitment and risk aversion inversely proportional to present wealth plus the financial value of future labor income net of consumption. In Section 4.5 we show, for a general time and state dependent risk aversion function fulfilling some reasonable assumptions, that a mean-standard deviation investor without pre-commitment should optimally take no risk at all.

4.2 The basic framework

In this section we present the basic model and the problems of interest. We discuss what we mean by an inconsistent problem, admissible strategies, and a corresponding optimal strategy. To solve the problems one need a Bellman-type set of partial differential equations. These are presented at the end of this section in the verification theorem, Theorem 4.2.1. Proofs are outlined in the Appendix.

4.2.1 The economic model

The economic setup is a standard Black-Scholes model consisting of a bank account, B, with risk free short rate, r, and a stock, S, with dynamics given by

$$\begin{aligned} dB(t) &= rB(t)dt,\ B(0) = 1, \\ dS(t) &= \alpha S(t)dt + \sigma S(t)dW(t),\ S(0) = s_0 > 0. \end{aligned}$$

Here $\alpha, \sigma, r > 0$ are constants and it is assumed that $\alpha > r$. The process W is a standard Brownian motion on an abstract probability space $(\Omega, \mathcal{F}, P)$ equipped with the filtration $\mathbb{F}^W = \left(\mathcal{F}^W(t)\right)_{t\in[0,T]}$ given by the P-augmentation of the filtration $\sigma\{W(s); 0 \le s \le t\}$, $\forall t \in [0, T]$.

We consider an investor with time horizon $[0, T]$, $T > 0$, and wealth process $(X(t))_{t\in[0,T]}$. The investor is assumed to be endowed with a continuous deterministic labor income rate ℓ and an initial amount of money x_0. At time t the investor chooses a non-negative consumption rate $c(t)$ and places a proportion $\pi(t)$ of his wealth in the stock, and the remainder in the bank account. Denoting by $X^{c,\pi}(t)$ the investors wealth at time t given the consumption and investment strategy (c, π), from now on just called strategy, the dynamics of the investor's wealth become

$$\begin{aligned} dX^{c,\pi}(t) &= \left[(r + \pi(t)(\alpha - r))X^{c,\pi}(t) + \ell(t) - c(t)\right] dt + \pi(t)\sigma X^{c,\pi}(t)dW(t),\ t \in [0, T), \quad (4.5) \\ X(0) &= x_0 > 0. \end{aligned}$$

For later use, see Section 4.4, we present the equivalent martingale measure, $\widehat{P}$, which for the Black-Scholes market is well-known to be given by the unique Radon-Nikodym derivative

$$\frac{d\widehat{P}(t)}{dP(t)} = \exp\left(-\left(\frac{\alpha - r}{\sigma}\right)W(t) - \frac{1}{2}\left(\frac{\alpha - r}{\sigma}\right)^2 t\right),\ t \in [0, T]. \quad (4.6)$$

The process $W^{\widehat{P}}$ given by

$$W^{\widehat{P}}(t) = W(t) + \frac{\alpha - r}{\sigma}t,\ t \in [0, T],$$

is a standard Brownian motion under the martingale measure $\widehat{P}$.

4.2.2 The problems of interest

Before introducing the problems of interest we introduce two conditional expectations

$$y(t,x) = E\left[\int_t^T e^{-\rho(s-t)} c(s)ds + e^{-\rho(T-t)} X^{c,\pi}(T) \,\middle|\, X(t) = x\right],$$
$$z(t,x) = E\left[\left(\int_t^T e^{-\rho(s-t)} c(s)ds + e^{-\rho(T-t)} X^{c,\pi}(T)\right)^2 \,\middle|\, X(t) = x\right].$$

Here ρ is a constant discounting rate, possibly different from the interest rate r. Loosely speaking, the class of stochastic problems we consider is, for any $(t,x) \in [0,T) \times \mathbb{R}$, to maximize

$$f(t,x,y^{c,\pi}(t,x), z^{c,\pi}(t,x)),\ (c,\pi) \in \mathcal{A}, \tag{4.7}$$

where $f \in \mathcal{C}^{1,2,2,2}$ and $\mathcal{A}$ is the class of admissible strategies to be defined in Theorem 4.2.1. The class of problems given by (4.7) contains two examples (among others) of economic interest, which we analyze in Section 4.3–4.5:

- Mean-variance without pre-commitment:

$$f(t,x,y,z) = y - \frac{\psi(t,x)}{2}\left(z - y^2\right),\ (t,x) \in [0,T) \times \mathbb{R}, \tag{4.8}$$

  where $\psi \in \mathcal{C}^{1,2}$ is a function which is allowed to depend on time and wealth. The problem is non-standard because of the non-linearity in y and because of the presence of t and x in the function ψ[1]. For a pure portfolio investor (c and ℓ set to zero) variants of (4.8) have been treated: For ψ constant, the problem is treated, in an incomplete Wiener driven framework, by Basak and Chabakauri (2010). Further, the problem is studied as a special (the simplest) case by Björk and Murgoci (2009). The case $\psi(t,x) = \gamma/x$, for a constant γ, is investigated by Björk et al. (2012).

- Mean-standard deviation without pre-commitment:

$$f(t,x,y,z) = y - \psi(t,x)\left(z - y^2\right)^{\frac{1}{2}},\ (t,x) \in [0,T) \times \mathbb{R}, \tag{4.9}$$

  where $\psi \in \mathcal{C}^{1,2}$ is a function allowed to depend on time and wealth. In addition to the arguments for the mean-variance problem this problem is non-standard due to the non-linearity in z. For a pure portfolio investor (c and ℓ set to zero) the problem has been treated in Kryger and Steffensen (2010) for the case ψ constant.

To the authors knowledge mean-variance and mean-standard deviation without pre-commitment including consumption and terminal wealth have not before been analyzed.

4.2.3 Inconsistency and the concept of an optimal strategy

The stochastic problems presented in (4.8) and (4.9) are called *time inconsistent* in the sense that the Bellman Optimality Principle does not hold: Suppose that we find the optimal strategy (c^*, π^*) for the time-0 problem $\mathcal{P}_{0,x_0}$ and suppose that we use this strategy on the time interval $[0,t]$. Then at time t the strategy (c^*, π^*) will not be optimal for the time-t problem $\mathcal{P}_{t,X^{c^*,\pi^*}(t)}$. This is because the law of iterated expectations does not apply for a given strategy. If the investors preference really is to pre-commit at time 0, he should of course simply solve the problem $\mathcal{P}_{0,x_0}$ and follow the corresponding optimal pre-commitment strategy. Here optimal is

[1] The appearance of z and thereby terms like "the conditional expectations of cumulated consumption to the power of 2" also makes the problem non-standard.

interpreted as optimal from the point of view of time zero. For some investors, this might be meaningful.

On the other hand it could easily be argued that most investors assign the same weight to all points in time, i.e. they do not look for an optimal strategy from the point of view of (say) time zero. Or put it another way, it seems reasonable that the investor assign no particular importance to a single point in time. Therefore, Björk and Murgoci (2009) and Björk et al. (2012) attack the problems in a game theoretic framework. That is, our preferences change in a temporally inconsistent way as time goes by and we can thus think about the problem as a game where the players are the future incarnations of ourselves. More specific, at every point in time t we have a player (and incarnation of ourselves) which we denote $\mathcal{P}_t$. Player $\mathcal{P}_t$ chooses the strategy $(c(t), \pi(t))$ at time t. The reward to $\mathcal{P}_t$ of course depends on the choice made by $\mathcal{P}_t$, but also on the choices made by the players $\mathcal{P}_s$ for all $s \in (t, T]$. We can now loosely define a subgame perfect Nash equilibrium strategy as a strategy (c^*, π^*) for which the following holds (for all players):

- If $\mathcal{P}_t$ knows that all players coming after him will use the strategy (c^*, π^*), then it is optimal for $\mathcal{P}_t$ also to use (c^*, π^*).

In a discrete time setup the concept of a subgame perfect Nash equilibrium strategy is very intuitive and it seems natural to look for subgame perfect Nash equilibrium strategies. However, it turns out to be a lot more complicated to define an equilibrium strategy in continuous time. The problem is that a single point in time has Lebesque measure zero, i.e. the individual player $\mathcal{P}_t$ does not influence the outcome of the game. Ekeland and Lazrak (2006) and Ekeland and Pirvu (2008) gave a precise definition of the game theoretic equilibrium concept in continuous time:

Definition 4.2.1. *Consider a strategy (c^*, π^*), choose a fixed reel point (c, π), a fixed real number $h > 0$, and an arbitrary initial point (t, x). Define the strategy $(\widetilde{c}_h, \widetilde{\pi}_h)$ by*

$$(\widetilde{c}_h(s), \widetilde{\pi}_h(s)) = \begin{cases} (c, \pi), & \text{for } t \le s < t+h, \\ (c^*(s), \pi^*(s)), & \text{for } t+h \le s < T. \end{cases} \tag{4.10}$$

If

$$\liminf_{h \to 0} \frac{f\left(t, x, y^{c^*,\pi^*}(t,x), z^{c^*,\pi^*}(t,x)\right) - f\left(t, x, y^{\widetilde{c}_h,\widetilde{\pi}_h}(t,x), z^{\widetilde{c}_h,\widetilde{\pi}_h}(t,x)\right)}{h} \ge 0,$$

*for all $(c, \pi) \in \mathbb{R}_+ \times \mathbb{R}$, we say that (c^*, π^*) is an **equilibrium** strategy.*

Looking for equilibrium strategies as defined in Definition 4.2.1 Björk et al. (2012) solve inconsistent control problems in the form (4.3). Consequently, they refer to the function defining the expected value of using the equilibrium strategy as the equilibrium value function.

We also choose to look for equilibrium strategies and refer, as Björk et al. (2012), to the strategies by the term *optimal*. However, in contrast to Björk et al. (2012) we choose to refer to the corresponding value function as the *optimal* value function. Denoting the optimal value function by V we write

$$V(t, x) = f\left(t, x, y^{c^*,\pi^*}, z^{c^*,\pi^*}\right), \tag{4.11}$$

for a strategy (c^*, π^*) fulfilling the equilibrium criteria in Definition 4.2.1. Doing so, our problem is to look for the optimal value function and the corresponding optimal strategies for objectives in the form given by (4.7).

Remark 4.2.1. It is all about how we look at the problem. Assume we are standing at time t and consider, for example, the mean-variance problem given by

$$f(t,x,y^{c,\pi}(t,x),z^{c,\pi}(t,x)), \tag{4.12}$$

where $f(t,x,y,z) = y - \frac{\gamma}{2}\left(z - y^2\right)$. To write the pre-commitment version of this problem (that is the investor pre-commit to his time-t preferences) one should define

$$\widetilde{y} = E\left[\int_t^T e^{-\rho(s-t)}c(s)ds + e^{-\rho(T-t)}X^{c,\pi}(T) \,\middle|\, X(t) = x\right],$$

and then write the problem as

$$\widetilde{f}(t,x,y^{c,\pi}(t,x),z^{c,\pi}(t,x),\widetilde{y}), \tag{4.13}$$

where $\widetilde{f}(t,x,y,z,\widetilde{y}) = y - \frac{\gamma}{2}\left(z + (\widetilde{y})^2 - 2y\widetilde{y}\right)$. At this point the reader might be confused since, at time t, we obviously have

$$f(t,x,y^{c,\pi}(t,x),z^{c,\pi}(t,x)) = \widetilde{f}(t,x,y^{c,\pi}(t,x),z^{c,\pi}(t,x),\widetilde{y}),$$

i.e. from the starting point of view the two problems look identical. However, at time $s \in (t,T)$, we have

$$f(s,X^{c,\pi}(s),y(s,X^{c,\pi}(s)),z(s,X^{c,\pi}(s))) \neq \widetilde{f}(s,X^{c,\pi}(s),y(s,X^{c,\pi}(s)),z(s,X^{c,\pi}(s)),\widetilde{y}).$$

The point is that by (4.13) we have made it clear that $\widetilde{y}$, opposed to x, y and z, is not a dynamic variable, i.e. the investor pre-commits to the time-t target when evaluating the variance term. To solve the pre-commitment problem (4.13) the trick is to write the problem as

$$\sup_{(c,\pi)\in\mathcal{A},\ \widetilde{y}=K} \widetilde{f}(t,x,y^{c,\pi}(t,x),z^{c,\pi}(t,x),K).$$

This can be solved in two steps: First solve the problem for a general K (this is a standard control problem), thereby obtaining an optimal strategy, $(c^*(K),\pi^*(K))$, as a function of K. Then insert $(c^*(K),\pi^*(K))$ in $\widetilde{y}$ and determine the optimal K^* as the solution to the nonlinear equation $\widetilde{y} = K^*$. For references see Korn (1997a) and Zhou and Li (2000).

4.2.4 The main result

In this subsection we present an extension of the standard Hamilton-Jacobi-Bellman equation for characterization of the optimal value function and the corresponding optimal strategy. The power of the verification theorem, as in the classic HJB framework, is that it transforms the stochastic problem into a system of deterministic differential equations and a deterministic pointwise infimum problem.

Theorem 4.2.1. *Let $f : [0,T] \times \mathbb{R}^3 \to \mathbb{R}$ be a function in $\mathcal{C}^{1,2,2,2}$. Define the set of admissible strategies, $\mathcal{A}$, as those, (c,π), for which the partial differential equations (4.63)–(4.64) and (4.70)–(4.71) have solutions, and for which the stochastic integrals in (4.69), (4.75) and (4.86) are martingales. Let, $\forall (t,x) \in [0,T) \times \mathbb{R}$, the optimal value function $V(t,x)$ be defined by (4.11), and define*

$$y^{c,\pi}(t,x) = E\left[\int_t^T e^{-\rho(s-t)}c(s)ds + e^{-\rho(T-t)}X^{c,\pi}(T) \,\middle|\, X(t)=x\right], \tag{4.14}$$

$$z^{c,\pi}(t,x) = E\left[\left(\int_t^T e^{-\rho(s-t)}c(s)ds + e^{-\rho(T-t)}X^{c,\pi}(T)\right)^2 \,\middle|\, X(t)=x\right]. \tag{4.15}$$

If there exist three functions F, $F^{(1)}$ and $F^{(2)}$ such that, $\forall (t,x) \in (0,T) \times \mathbb{R}$, we have

$$\begin{aligned}
F_t &= \inf_{(c,\pi)\in\mathcal{A}} \Big\{ -[(r+\pi(\alpha-r))x + \ell - c](F_x - Q) - \frac{1}{2}\pi^2\sigma^2 x^2 (F_{xx} - U) + J \Big\}, && (4.16)\\
F(T,x) &= f\left(T,x,x,x^2\right),\\
F_t^{(1)} &= -[(r+\pi^*(\alpha-r))x + \ell - c^*]F_x^{(1)} - \frac{1}{2}\pi^{*2}\sigma^2x^2F_{xx}^{(1)} - c^* + \rho F^{(1)}, && (4.17)\\
F^{(1)}(T,x) &= x,\\
F_t^{(2)} &= -[(r+\pi^*(\alpha-r))x + \ell - c^*]F_x^{(2)} - \frac{1}{2}\pi^{*2}\sigma^2x^2F_{xx}^{(2)} - 2c^*F^{(1)} + 2\rho F^{(2)}, && (4.18)\\
F^{(2)}(T,x) &= x^2,
\end{aligned}$$

where

$$Q = f_x, \tag{4.19}$$

$$U = f_{xx} + \left(F_x^{(1)}\right)^2 f_{yy} + \left(F_x^{(2)}\right)^2 f_{zz} + 2F_x^{(1)}F_x^{(2)}f_{yz} + 2F_x^{(1)}f_{xy} + 2F_x^{(2)}f_{xz}, \tag{4.20}$$

$$J = \left(\rho F^{(1)} - c\right) f_y + 2\left(\rho F^{(2)} - cF^{(1)}\right) f_z + f_t, \tag{4.21}$$

and

$$(\pi^*, c^*) = \arg \inf_{(c,\pi)\in\mathcal{A}} \left\{ -[(r+\pi(\alpha-r))x + \ell - c](F_x - Q) - \frac{1}{2}\pi^2\sigma^2x^2(F_{xx} - U) + J \right\},$$

then

$$V(t,x) = F(t,x),\ y^{c^*,\pi^*}(t,x) = F^{(1)}(t,x),\ z^{c^*,\pi^*}(t,x) = F^{(2)}(t,x),$$

and the optimal strategy is given by (c^*,π^*).

Proof. See Appendix 4.6.1. □

4.3 Mean-variance with constant risk aversion

The simplest case of mean-variance optimization without pre-commitment including consumption and terminal wealth is obtained by assuming constant risk aversion. Using Theorem 4.2.1 we are able to derive the optimal strategy. It turns out that the optimal investment strategy corresponds to a constant amount of money invested in stocks and that the optimal consumption strategy becomes a deterministic bang-bang-strategy. The solution and the problem are discussed below.

4.3.1 Presenting and solving the problem

Consider the problem of finding the optimal strategy for the objective given by

$$E_{0,x_0}\left[\int_0^T e^{-\rho s}c(s)ds + e^{-\rho T}X^{c,\pi}(T)\right] - \frac{\gamma}{2}Var_{0,x_0}\left[\int_0^T e^{-\rho s}c(s)ds + e^{-\rho T}X^{c,\pi}(T)\right], \quad (4.22)$$

where $\gamma > 0$ is a constant defining the investor's risk aversion, and where we restrict the admissible strategies to $(c,\pi) \in \mathcal{A}\bigcap(\mathcal{D}\times\mathbb{R})$, where $\mathcal{D}(s) := [c_{\min}(s), c_{\max}(s)]$, $s \in [0,T]$, is a finite interval. The corresponding function f is given by

$$f(t,x,y,z) = y - \frac{\gamma}{2}\left(z - y^2\right).$$

The system of partial differential equations we want to solve in order to obtain the optimal value function and the optimal strategy is given by (4.16) and (4.17)[2]. A candidate for the optimal strategy in terms of the value function is found by differentiating with respect to c and π inside the curly brackets in (4.16). Thereby

$$c^*(t,x) = \begin{cases} c_{\max}(t), & \text{if } F_x(x,t) - Q(x,t) < 1, \\ \text{non-defined}, & \text{if } F_x(x,t) - Q(x,t) = 1, \\ c_{\min}(t), & \text{if } F_x(x,t) - Q(x,t) > 1, \end{cases} \quad (4.23)$$

$$\pi^*(t,x)x = -\frac{\alpha - r}{\sigma^2}\frac{F_x - Q}{F_{xx} - U}, \quad (4.24)$$

(provided $U > F_{xx}$). Clearly

$$\begin{aligned} &\psi_x = \psi_{xx} = \psi_t = 0, \\ &f_y = 1 + \gamma y, f_{yy} = \gamma, f_z = -\frac{\gamma}{2}, \\ &f_t = f_x = f_{xx} = f_{xz} = f_{xy} = f_{zz} = f_{yz} = 0. \end{aligned}$$

Inserting this into (4.19)–(4.21) gives

$$Q = 0, \quad (4.25)$$

$$U = \gamma\left(F_x^{(1)}\right)^2, \quad (4.26)$$

$$J = \rho F^{(1)} - c - \gamma\rho\left(F^{(2)} - \left(F^{(1)}\right)^2\right). \quad (4.27)$$

We now search for solutions in the form

$$\begin{aligned} F(t,x) &= A(t)x + B(t), \\ F^{(1)}(t,x) &= a(t)x + b(t), \end{aligned}$$

[2] Since $F_t = F_t^{(1)} - \frac{\gamma}{2}\left(F_t^{(2)} + 2F^{(1)}F_t^{(1)}\right)$ we only need to solve two of the three differential equations given by (4.16)–(4.18). We choose to solve (4.16) and (4.17).

where A, B, a and b are deterministic functions of time. In order to calculate J we need to derive $F^{(2)}$ from our guess. The forms of F and $F^{(1)}$ determine the form of $F^{(2)}$. We have that

$$F^{(2)}(t,x) = \frac{2}{\gamma}[a(t)x + b(t) - A(t)x - B(t)] + [a(t)x + b(t)]^2.$$

The partial derivatives of interest are

$$\begin{aligned} F_t &= A'(t)x + B'(t),\ F_x = A(t),\ F_{xx} = 0, \\ F_t^{(1)} &= a'(t)x + b'(t),\ F_x^{(1)} = a(t),\ F_{xx}^{(1)} = 0. \end{aligned}$$

Inserting this into (4.25)–(4.27) gives

$$Q(t,x) = 0, \tag{4.28}$$
$$U(t,x) = \gamma a(t)^2, \tag{4.29}$$
$$J(t,x) = \rho[a(t)x + b(t)] - c(t) - 2\rho[a(t)x + b(t) - A(t)x - B(t)]. \tag{4.30}$$

Plugging the relevant derivatives, (4.28) and (4.29) into (4.23) and (4.24) we can now write the candidate for the optimal strategy in terms of the deterministic functions A and a. We get

$$c^*(t,x) = \begin{cases} c_{\max}(t), & \text{if } A(t) < 1, \\ \text{non-defined}, & \text{if } A(t) = 1, \\ c_{\min}(t), & \text{if } A(t) > 1, \end{cases} \tag{4.31}$$
$$\pi^*(t,x)x = \frac{1}{\gamma}\frac{\alpha - r}{\sigma^2}\frac{A(t)}{a(t)^2}, \tag{4.32}$$

(provided $\gamma a^2 > 0$). Inserting (4.28)–(4.32) and the relevant derivatives into the differential equations (4.16) and (4.17) and including the terminal conditions gives

$$\begin{aligned} A_t x + B_t &= -rxA - \frac{1}{2\gamma}\frac{(\alpha - r)^2}{\sigma^2}\frac{A^2}{a^2} - \ell A + c^*(A-1) \\ &\quad + \rho(ax + b) - 2\rho(ax + b - Ax - B), \\ A(T) &= 1, \\ B(T) &= 0, \\ a_t x + b_t &= -rxa - \frac{1}{\gamma}\frac{(\alpha - r)^2}{\sigma^2}\frac{A}{a} - \ell a + c^*(a-1) + \rho(ax + b), \\ a(T) &= 1, \\ b(T) &= 0. \end{aligned}$$

We obtain the solutions

$$A(t) = a(t) = e^{(r-\rho)(T-t)}$$

and

$$B(t) = e^{2\rho t}\int_t^T \left[\frac{1}{2\gamma}\frac{(\alpha - r)^2}{\sigma^2} + \ell(s)e^{(r-\rho)(T-s)} - c^*(s)\left(e^{(r-\rho)(T-s)} - 1\right) + \rho b(s)\right] e^{-2\rho s} ds, \tag{4.33}$$

$$b(t) = e^{\rho t}\int_t^T \left[\frac{1}{\gamma}\frac{(\alpha - r)^2}{\sigma^2} + \ell(s)e^{(r-\rho)(T-s)} - c^*(s)\left(e^{(r-\rho)(T-s)} - 1\right)\right] e^{-\rho s} ds, \tag{4.34}$$

as well as the relation

$$b(t)-B(t)=\frac{1}{2\gamma}\frac{(\alpha-r)^2}{\sigma^2}\int_t^T e^{-2\rho(s-t)}ds=\begin{cases}\frac{1}{4\gamma\rho}\frac{(\alpha-r)^2}{\sigma^2}\left(1-e^{-2\rho(T-t)}\right), & \text{if } \rho>0,\\ \frac{1}{2\gamma}\frac{(\alpha-r)^2}{\sigma^2}(T-t), & \text{if } \rho=0.\end{cases} \tag{4.35}$$

The optimal strategy now follows directly from plugging the solutions for A and a into (4.31) and (4.32). We summarize the results as follows:

Proposition 4.3.1. *For the mean-variance problem given by (4.22) we have the following results.*

- *The optimal strategy is given by*

$$c^*(t,x)=\begin{cases}c_{\max}(t), & \text{if } r<\rho,\\ \text{non-defined}, & \text{if } r=\rho,\\ c_{\min}(t), & \text{if } r>\rho,\end{cases}$$

$$\pi^*(t,x)x=\frac{1}{\gamma}\frac{(\alpha-r)}{\sigma^2}e^{-(r-\rho)(T-t)}.$$

- *The optimal value function is given by*

$$V(t,x)=e^{(r-\rho)(T-t)}x+B(t).$$

- *The conditional expected value of the discounted cumulated optimal consumption plus the discounted optimal terminal wealth is given by*

$$E_{t,x}\left[\int_t^T e^{-\rho(s-t)}c^*(s)ds+e^{-\rho(T-t)}X^{c^*,\pi^*}(T)\right]=e^{(r-\rho)(T-t)}x+b(t).$$

- *The conditional variance of the discounted cumulated optimal consumption plus the discounted optimal terminal wealth is given by*[3]

$$Var_{t,x}\left[\int_t^T e^{-\rho(s-t)}c^*(s)ds+e^{-\rho(T-t)}X^{c^*,\pi^*}(T)\right]=\frac{2}{\gamma}(b(t)-B(t)).$$

Here B, b and $b-B$ are given by (4.33)–(4.35).

4.3.2 Discussion of the solution and the problem

As mentioned in Björk et al. (2012) one can argue that from an economic point of view the optimal investment strategy does **not** make sense. From the expression of the optimal investment strategy we see that the optimal amount of money to invest in stocks is independent of wealth, i.e. for a given γ a rich investor and a poor investor optimally invest the same amount of money in stocks. For a one-period model the optimal investment strategy is reasonable since we would expect the richer investor to have a lower value of γ. However, for a multi-period model the strategy seems to be economically unreasonable. If the investor chooses γ such that it reflects his risk aversion corresponding to his initial wealth, then at a later point in time t, due to the progression of wealth, this γ (likely) no longer reflect his risk aversion corresponding to his present wealth $X^{c^*,\pi^*}(t)$. Obviously, the investor should choose his γ in a more sophisticated

[3] We have that

$$Var_{t,x}\left[\int_t^T e^{-\rho(s-t)}c^*(s)ds+e^{-\rho(T-t)}X^{c^*,\pi^*}(T)\right]=F^{(2)}(t,x)-\left(F^{(1)}(t,x)\right)^2=\frac{2}{\gamma}(b(t)-B(t)).$$

way. One approach is to let the risk aversion depend on present time and wealth. This case is analyzed in Section 4.4. The authors think that it is important, even though some might see it as equivalent concepts, to emphasize that we twist the objective function because we realize that the problem, and not the solution, is inappropriate.

We can interpret wealth as a pension saving account and labor income net of consumption as pension contributions. The constraint that the consumption rate c only is allowed to take values between a deterministic upper and lower boundary, $c_{\min}$ respectively $c_{\max}$, has a natural interpretation. If for example we have $c_{\min}(t) = k_1\ell(t)$ and $c_{\max}(t) = k_2\ell(t)$, for constants $0 < k_1 < k_2 < 1$, we have that the investor is forced to spend no less than a minimal fraction $1-k_2$ and no more than a maximal fraction $1-k_1$ of his labor income on pension contributions. This corresponds to a compulsory pension scheme and a subsistence level, respectively.

From the expression of the optimal consumption strategy we have that the investor optimally consumes the minimum (maximum) allowed if he is patient (impatient). That is, if he has a time preference parameter ρ smaller (greater) than the risk free interest rate r. If we take the optimal investment strategy as given this result is easy to understand: First of all the deterministic consumption strategy minimizes the variance term in (4.22). If the investor chooses to consume the minimum allowed he saves as much as possible. These savings earn, according to the optimal investment strategy, the risk-free interest rate. This consumption strategy is only optimal if these savings including interest are large enough for the investor to be willing to wait for them, i.e. if $r > \rho$. Reversely, if $r < \rho$ the investor is to impatient to wait for the savings including interest and chooses to consume the maximum allowed instead. However, it is important to emphasize that initial we could not have foreseen this trivial optimal consumption strategy since we search for c^* and π^* simultaneously.

Finally, one should notice that the optimal strategy does not guarantee that wealth stays non-negative (or above any other given lower boundary). This is also the case for the optimal strategy derived in Basak and Chabakauri (2010) and Björk and Murgoci (2009). However, the optimal strategy is perfectly reasonable. In the definition of the problem (4.22) we do not exclude strategies for which the corresponding wealth has positive probability of becoming negative. Proposition 4.3.1 simply tells us that, as a consequence hereof, it is optimal to continue to take risky investment decisions and consume even though this may punish the total utility obtained over the interval $[0, T]$ in form of a negative wealth at time T.

4.4 Mean-variance with time and state dependent risk aversion

By introducing a time and state dependent risk aversion function the mean-variance problem without pre-commitment including consumption and terminal wealth becomes much more complicated. For the case $c = \ell = 0$ this is analyzed, for state (but not time) dependent risk aversion, by Björk et al. (2012). We consider the special case of time and state dependent risk aversion where the investor's risk aversion is hyperbolic in present wealth plus the financial value of future labor income net of consumption. One can argue that the investor, hereby, can influence his own risk aversion by choosing his consumption rate in a certain way. This is in our opinion however perfectly reasonable. All it says is that if the investor knows that he in the future is going to save money by consuming less then he should act as if he already had more money and adapt his risk aversion to that situation. Solving the mean-variance problem with time and state dependent risk aversion we only allow the investor to look for strategies for which the corresponding wealth plus the financial value of future labor income net of consumption stays positive over the entire interval. This conforms with the well-known and often required constraint that wealth plus human capital must stay positive at all times. Consequently, terminal wealth becomes positive.

As in Section 4.3 we restrict consumption by a time dependent upper and lower boundary.

It turns out that the optimal investment strategy becomes linear in the investor's wealth plus financial value of future labor income net of consumption. Furthermore, as in the case of constant risk aversion, the optimal consumption rate becomes a deterministic bang-bang strategy. The deterministic function determining when it is optimal to consume the maximum or minimum allowed is given by a system of non-linear differential equations for which we have no explicit solution. Moreover, the constraint that wealth plus the financial value of future labor income net of consumption must stay positive at all times may become binding. That is, in order to be able to finance his consumption stream, the investor optimally consumes the minimum allowed from the point in time where the constraint (may) becomes active and onwards. Opposed to the case with constant risk aversion we find that it is not always optimal to either consume the maximum or minimum allowed at all times. For some investors we find that it is optimal to first, for a period of time, to consume the maximum allowed and then, for the remainder of the period, to consume the minimum allowed. The results are analyzed in details below.

4.4.1 Presenting and solving the problem

Define the time-t financial value of future labor income net of consumption by

$$K^{(c)}(t,x) := E_{t,x}^{\widehat{P}}\left[\int_t^T e^{-r(s-t)}\Big(\ell(s) - c\left(s, X^{c,\pi}(s)\right)\Big)ds\right], \tag{4.36}$$

and define, for a finite time dependent interval $\mathcal{D} \in \mathbb{R}$, the set of strategies

$$\mathcal{B}(t) := \Big\{(c,\pi) \;\Big|\; c(s) \in \mathcal{D}(s) := [c_{\min}(s), c_{\max}(s)],\ X^{c,\pi}(s) + K^{(c)}(s, X^{c,\pi}(s)) \geq 0, \forall s \in [t,T]\Big\}.$$

In (4.36) the expectation is derived under the martingale measure $\widehat{P}$ defined by (4.6). The condition $x + K^{(c)} \geq 0$ ensures that the strategies in $\mathcal{B}$ fulfill the natural requirement that, at any time, the investor should be able to finance his own consumption stream. That is, the investor has to be sure that, at any time, consumption can be financed by present wealth, capital gains and labor income. To ensure that the set of strategies $\mathcal{B}$ is non-empty we must assume that, initially, $c_{\min}$ fulfils the condition $x + K^{(c_{\min})} \geq 0$. Note that depending on the size of $c_{\max}$ and the size of labor income, consumption is always either restricted directly by the upper bound $c_{\max}$ *or* indirectly by the more technical constraint $x + K^{(c)} \geq 0$. A lower bound $c_{\min}$ seems natural since the investor is expected to have a subsistence level. At least consumption should naturally be restricted to stay non-negative.

We now consider the problem of finding the optimal strategy for the objective given by

$$\begin{aligned} &E_{0,x_0}\left[\int_0^T e^{-\rho s}c(s)ds + e^{-\rho T}X^{c,\pi}(T)\right] \\ &- \frac{\gamma}{2\left(x_0 + K^{(c)}(0,x_0)\right)} Var_{0,x_0}\left[\int_0^T e^{-\rho s}c(s)ds + e^{-\rho T}X^{c,\pi}(T)\right], \end{aligned} \tag{4.37}$$

where we restrict the class of admissible strategies to $(c,\pi) \in \mathcal{A}\bigcap\mathcal{B}$. Note that hereby the risk aversion factor on the conditional variance is a function of the consumption stream. The corresponding function f is given by

$$f(t,x,y,z) = y - \frac{\psi(t,x)}{2}\left(z - y^2\right),$$

where $\psi(t,x) = \frac{\gamma}{x + K^{(c)}(t,x)}$. The system of partial differential equations we want to solve in order to obtain the optimal value function and the optimal strategy is given by (4.17) and (4.18)[4].

[4]Since $F_t = F_t^{(1)} - \frac{\frac{\partial}{\partial t}\psi(t,x)}{2}\left(F^{(2)} - \left(F^{(1)}\right)^2\right) - \frac{\psi(t,x)}{2}F_t^{(2)} + \psi(t,x)F^{(1)}F_t^{(1)}$ we only need to solve two of the three differential equations given by (4.16)–(4.18). We choose to solve (4.17) and (4.18).

Clearly,

$$
\begin{aligned}
&f_t = -\frac{\psi_t}{2}\left(z-y^2\right), f_x = -\frac{\psi_x}{2}\left(z-y^2\right), f_{xx} = -\frac{\psi_{xx}}{2}\left(z-y^2\right),\\
&f_y = 1+\psi y, f_{yy} = \psi, f_z = -\frac{\psi}{2}, f_{xz} = -\frac{\psi_x}{2}, f_{xy} = \psi_x y,\\
&f_{zz} = f_{yz} = 0.
\end{aligned}
$$

Combining with the formulas (4.19)–(4.21) we obtain, after some rearrangement of terms, the following expressions which turn out to be useful

$$
\begin{aligned}
F_x - Q &= F_x^{(1)} + \psi F^{(1)} F_x^{(1)} - \frac{\psi}{2} F_x^{(2)}, &(4.38)\\
F_{xx} - U &= F_{xx}^{(1)} - \frac{\psi}{2} F_{xx}^{(2)} + \psi F^{(1)} F_{xx}^{(1)}, &(4.39)\\
J &= \rho F^{(1)} - c - \left(\psi\rho + \frac{\psi_t}{2}\right)\left(F^{(2)} - \left(F^{(1)}\right)^2\right). &(4.40)
\end{aligned}
$$

By characterizing (4.36) as the solution to a (Feynman-Kač) PDE we get that

$$
\psi_t = -\frac{\gamma}{\left(x+K^{(c)}\right)^2}\left(rK^{(c)} - \ell + c - (rx+\ell-c)K_x^{(c)} - \frac{1}{2}\pi^2\sigma^2x^2K_{xx}^{(c)}\right).
$$

Note that the coefficient of $K_x^{(c)}$ does not include the excess return $(\alpha - r)$ since $K^{(c)}$ is defined as a conditional expectation under the martingale measure $\widehat{P}$. Insert this in (4.40) to obtain

$$
\begin{aligned}
J =\ & \rho F^{(1)} - c - \frac{\gamma\rho}{x+K^{(c)}}\left(F^{(2)} - \left(F^{(1)}\right)^2\right) + \frac{\gamma}{2\left(x+K^{(c)}\right)^2}\Bigg(rK^{(c)} - \ell + c\\
&- (rx+\ell-c)K_x^{(c)} - \frac{1}{2}\pi^2\sigma^2x^2K_{xx}^{(c)}\Bigg)\left(F^{(2)} - \left(F^{(1)}\right)^2\right). \qquad (4.41)
\end{aligned}
$$

We are now ready to derive a candidate for the optimal strategy. To do this we consider the two cases where the constraint $x + K^{(c)} \geq 0$ is non-binding and binding, respectively. That is, we consider, for an arbitrary point (t,x), the two cases $x+K^{(c_{\min})}(t,x) > 0$ and $x+K^{(c_{\min})}(t,x) = 0$, respectively.

The non-binding case ($x + K^{(c_{\min})}(t,x) > 0$)

A candidate for the optimal strategy is found by differentiating with respect to c and π inside the curly brackets in (4.16). We get

$$
c^*(t,x) = \begin{cases} c_{\max}(t), & \text{if } C(t,x) < 1,\\ \text{non-defined}, & \text{if } C(t,x) = 1,\\ c_{\min}(t), & \text{if } C(t,x) > 1, \end{cases} \qquad (4.42)
$$

where

$$
C(t,x) := F_x(x,t) - Q(x,t) + \frac{\gamma}{2(x+K^{(c^*)}(t))^2}\left(1+K_x^{(c^*)}(t,x)\right)\left(F^{(2)}(t,x) - \left(F^{(1)}(t,x)\right)^2\right),
$$

and

$$
\pi^*(t,x)x = -\frac{\alpha - r}{\sigma^2}\frac{F_x - Q}{F_{xx} - U}, \qquad (4.43)
$$

(provided $U > F_{xx}$). The optimal consumption strategy is a bang-bang strategy and, for the moment, it appears to be stochastic and dependent on wealth. However, we now search for a solution to the problem in the set of solutions such that the optimal consumption strategy becomes deterministic. If we find such a solution this is of course no restriction. Thus, we search for solutions where $F^{(1)}$ and $F^{(2)}$ are in a form designed exactly such that this is the case. We propose that

$$\begin{aligned}
F^{(1)}(t,x) &= a(t)\left(x + K^{(c^*)}(t)\right) + b(t), \\
F^{(2)}(t,x) &= f(t)\left(x + K^{(c^*)}(t)\right)^2 + g(t)\left(x + K^{(c^*)}(t)\right) + h(t),
\end{aligned}$$

where a, b, f, g and h are deterministic functions of time, where the candidate for the optimal consumption strategy, c, is assumed to be independent of wealth, and where[5]

$$a(t)b(t) = \frac{g(t)}{2}, \tag{4.44}$$

$$h(t) = b(t)^2. \tag{4.45}$$

In this case we get that

$$K^{(c^*)}(t) = \int_t^T e^{-r(s-t)}(\ell(s) - c^*(s))ds.$$

Due to (4.44) and (4.45) we get that the variance term in the value function in (4.37) has the form

$$\begin{aligned}
& F^{(2)} - \left(F^{(1)}\right)^2 \\
= \;& \left(f(t) - a(t)^2\right)\left(x + K^{(c^*)}(t)\right)^2 + 2\left(\frac{g(t)}{2} - a(t)b(t)\right)\left(x + K^{(c^*)}(t)\right) + h(t) - b(t)^2 \\
= \;& \left(f(t) - a(t)^2\right)\left(x + K^{(c^*)}(t)\right)^2. \qquad (4.46)
\end{aligned}$$

The form of F is completely determined by $F^{(1)}$ and $F^{(2)}$. Using (4.46) we get that

$$\begin{aligned}
F(t,x) &= F^{(1)} - \frac{\gamma}{2\left(x + K^{(c^*)}(t)\right)}\left\{F^{(2)} - \left(F^{(1)}\right)^2\right\} \\
&= a(t)\left(x + K^{(c^*)}(t)\right) + b(t) - \frac{\gamma}{2}\left(f(t) - a(t)^2\right)\left(x + K^{(c^*)}(t)\right).
\end{aligned}$$

The partial derivatives of interest become

$$\begin{aligned}
F_t^{(1)} &= a'(t)\left(x + K^{(c^*)}(t)\right) + a(t)\left(rK^{(c^*)}(t) - \ell(t) + c(t)\right) + b'(t), \\
F_x^{(1)} &= a(t),\; F_{xx}^{(1)} = 0, \\
F_t^{(2)} &= f'(t)\left(x + K^{(c^*)}(t)\right)^2 + 2f(t)\left(rK^{(c^*)}(t) - \ell(t) + c(t)\right)\left(x + K^{(c^*)}(t)\right) \\
&\quad + g'(t)\left(x + K^{(c^*)}(t)\right) + g(t)\left(rK^{(c^*)}(t) - \ell(t) + c(t)\right) + h'(t), \\
F_x^{(2)} &= 2f(t)\left(x + K^{(c^*)}(t)\right) + g(t),\; F_{xx}^{(2)} = 2f(t).
\end{aligned}$$

[5]The assumptions given in (4.44) and (4.45) turn out to be consistent with the assumption that the candidate for the optimal consumption strategy is deterministic.

Inserting this in (4.38) gives

$$\begin{aligned}
F_x(t,x) - Q(t,x) &= a(t) + \frac{\gamma}{x+K^{(c^*)}(t)}\left[a(t)^2\left(x+K^{(c^*)}(t)\right) + a(t)b(t)\right] \\
&\quad - \frac{\gamma}{2\left(x+K^{(c^*)}(t)\right)}\left[2f(t)\left(x+K^{(c^*)}(t)\right) + g(t)\right] \\
&= a(t) + \gamma\left(a(t)^2 - f(t)\right) + \frac{\gamma}{x+K^{(c^*)}(t)}\left(a(t)b(t) - \frac{g(t)}{2}\right).
\end{aligned}$$

By assumption (4.44) this reduces to

$$F_x(t,x) - Q(t,x) = a(t) + \gamma\left(a(t)^2 - f(t)\right). \tag{4.47}$$

Inserting the partial derivatives in (4.39) gives

$$F_{xx}(t,x) - U(t,x) = -\frac{\gamma f(t)}{x+K^{(c^*)}(t)}. \tag{4.48}$$

Now, insert $K_x^{(c^*)} = 0$, (4.46) and (4.47) in (4.42) to obtain

$$c^*(t) = \begin{cases} c_{\max}(t), & \text{if } a(t) + \frac{\gamma}{2}\left(a(t)^2 - f(t)\right) < 1, \\ \text{non-defined}, & \text{if } a(t) + \frac{\gamma}{2}\left(a(t)^2 - f(t)\right) = 1, \\ c_{\min}(t), & \text{if } a(t) + \frac{\gamma}{2}\left(a(t)^2 - f(t)\right) > 1, \end{cases} \tag{4.49}$$

and insert (4.47) and (4.48) in (4.43) to obtain

$$\pi^*(t,x)x = \frac{\alpha - r}{\sigma^2\gamma f(t)}\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right]\left(x+K^{(c^*)}(t)\right), \tag{4.50}$$

provided that[6]

$$\frac{\gamma f(t)}{x+K^{(c^*)}(t)} > 0. \tag{4.51}$$

Now plug in (4.49), (4.50) and the relevant derivatives into (4.17) and include the terminal conditions to obtain

$$\boxed{\begin{aligned}
a_t &= -\left\{(r-\rho) + \frac{(\alpha-r)^2}{\sigma^2\gamma f}\left[a + \gamma(a^2 - f)\right]\right\}a, \\
a(T) &= 1, \\
b_t &= -c^* + \rho b, \\
b(T) &= 0.
\end{aligned}} \tag{4.52}$$

[6]The candidate of the optimal investment strategy (4.43) was derived under the condition that $U > F_{xx}$. By use of (4.48) we can write this condition as (4.51).

Now, insert (4.49), (4.50) and the relevant derivatives into (4.18) and include the terminal conditions to obtain

$$
\boxed{
\begin{aligned}
f_t &= -\left\{2\left((r-\rho)+\frac{(\alpha-r)^2}{\sigma^2\gamma f}\left[a+\gamma(a^2-f)\right]\right)+\frac{(\alpha-r)^2}{\sigma^2\gamma^2 f^2}\left[a+\gamma(a^2-f)\right]^2\right\}f,\\
g_t &= -\left(r+\frac{(\alpha-r)^2}{\sigma^2\gamma f}\left[a+\gamma(a^2-f)\right]\right)g-2c^*a+2\rho g,\\
h_t &= -2c^*b+2\rho h,\\
f(T) &= 1,\\
g(T) &= 0,\\
h(T) &= 0.
\end{aligned}
}
\tag{4.53}
$$

From (4.52) and (4.53) we immediately obtain the solutions

$$
\begin{aligned}
b(t) &= \int_t^T e^{-\rho(s-t)}c^*(s)ds,\\
h(t) &= \left(\int_t^T e^{-\rho(s-t)}c^*(s)ds\right)^2,
\end{aligned}
$$

i.e. assumption (4.45) is indeed fulfilled. Finally, we have three things left to verify!

- We need to show that assumption (4.44) is fulfilled.
- The candidate of the optimal investment strategy (4.50) was derived under the condition (4.51). This assumption has to be verified.
- The non-linear system of partial differential equations given by (4.52) and (4.53) does not satisfy the usual Lipschitz and growth conditions. Global existence and uniqueness are therefore not guaranteed. We need to show that the system of partial differential equations in fact has a unique solution.

This is all done in Appendix 4.6.2.

The binding case ($x + K^{(c_{\min})}(t,x) = 0$)

Whenever the constraint is active the investor is forced to consume at the minimum rate allowed. By the definition of $K^{(c)}$ given by (4.36) the only investment strategy which can finance this consumption stream, while keeping $x + K^{(c)} \geq 0$, is $\pi = 0$. We get that the only strategy, and thereby the optimal strategy, in $\mathcal{A}\bigcap\mathcal{B}$ is $c^*(t) = c_{\min}(t)$ and $\pi^*(t,x) = 0$. Obviously, once the restriction becomes active it becomes binding for the remaining time of the period. That is, if $x + K^{(c)}(t,x) = 0$ we get $X^{c,\pi}(s) + K^{(c)}(s, X^{c,\pi}(s)) = 0$ for all $s \in [t,T]$.

Collecting the results from the two cases we summarize as follows

Proposition 4.4.1. *For the mean-variance problem given by (4.37) we have the following results.*

- *The optimal strategy is given by*

$$c^*(t) = \begin{cases} \widehat{c}(t), & \text{if } t \in [0, t^*), \\ c_{\min}(t), & \text{if } t \in [t^*, T], \end{cases}$$

$$\pi^*(t,x)x = \frac{\alpha - r}{\sigma^2 \gamma f(t)} \left[a(t) + \gamma \left(a(t)^2 - f(t) \right) \right] \left(x + K^{(c^*)}(t) \right),$$

where

$$\widehat{c}(t) = \begin{cases} c_{\max}(t), & \text{if } a(t) + \frac{\gamma}{2} \left(a(t)^2 - f(t) \right) < 1, \\ \text{non-defined}, & \text{if } a(t) + \frac{\gamma}{2} \left(a(t)^2 - f(t) \right) = 1, \\ c_{\min}(t), & \text{if } a(t) + \frac{\gamma}{2} \left(a(t)^2 - f(t) \right) > 1, \end{cases}$$

and t^* is given by

$$\sup \left\{ t^* \in [t, T] \;\middle|\; X^{c,\pi}(t) + \int_t^{t^*} e^{-r(s-t)} (\ell(s) - \widehat{c}(s)) ds + \int_{t^*}^T e^{-r(s-t)} (\ell(s) - c_{\min}(s)) ds \geq 0 \right\}. \tag{4.54}$$

- *The optimal value function is given by*

$$V(t,x) = a(t) \left(x + K^{(c^*)}(t) \right) + b(t) - \frac{\gamma}{2} \left(f(t) - a(t)^2 \right) \left(x + K^{(c^*)}(t) \right).$$

- *The conditional expected value of the discounted cumulated optimal consumption plus the discounted optimal terminal wealth is given by*

$$E_{t,x} \left[\int_t^T e^{-\rho(s-t)} c^*(s) ds + e^{-\rho(T-t)} X^{c^*,\pi^*}(T) \right] = a(t) \left(x + K^{(c^*)}(t) \right) + b(t).$$

- *The conditional variance of the discounted cumulated optimal consumption plus the discounted optimal terminal wealth is given by*

$$Var_{t,x} \left[\int_t^T e^{-\rho(s-t)} c^*(s) ds + e^{-\rho(T-t)} X^{c^*,\pi^*}(T) \right] = \left(f(t) - a(t)^2 \right) \left(x + K^{(c^*)}(t) \right)^2.$$

Here,

$$K^{(c^*)}(t) = \int_t^T e^{-r(s-t)} (\ell(s) - c^*(s)) ds,$$

and a, b, f, g and h are given by the non-linear system of partial differential equations given by (4.52) and (4.53).

Remark 4.4.1. We get that (4.54) can be written as

$$\sup \left\{ t^* \in [0,T] \;\middle|\; x_0 + \int_0^{t^*} e^{-rs}(\ell(s) - \widehat{c}(s))ds + \int_{t^*}^T e^{-rs}(\ell(s) - c_{\min}(s))ds \geq 0 \right\}.$$

That is, t^* can be found initially at time $t = 0$ and will not change at a later point in time. To make this clear consider, at time $t = 0$, the two situations:

1. The cases $t^* < T$ and $t^* = T$ with $x_0 + K^{(c^*)}(0, x_0) = 0$: As already noted, in those cases, the constraint $x + K^{(c)} \geq 0$ becomes binding in the sense that the only investment strategy which can finance the consumption stream c^*, while keeping $x + K^{(c)} \geq 0$, is $\pi^*(t) = 0$, $\forall t \in [0,T]$. Obviously, we get $X^{c^*,\pi^*}(t) + K^{(c^*)}(t, X^{c^*,\pi^*}(t)) = 0$, $\forall t \in [0,T]$, i.e. t^* does not change at a later point in time.

2. The case $t^* = T$ with $x_0 + K^{(c^*)}(0, x_0) > 0$: We get due to π^* being linear in $x + K^{(c^*)}$ that $X^{c^*,\pi^*}(t) + K^{(c^*)}(t, X^{c^*,\pi^*}(t)) > 0$, $\forall t \in [0,T]$[7], i.e. $t^* = T$ at all future time points.

We emphasize that it is the binding nature of the constraint together with the dynamic investment strategy that makes t^*, and consequently the optimal consumption strategy, deterministic. Once again, the optimal consumption strategy $c^*(t),\ t \in [0,T]$, is completely known at time $t = 0$.

Remark 4.4.2. For big enough values of γ we may have (if $c_{\max}$ is big relative to x_0) that $x_0 + K^{(c^*)}(0, x_0) = 0$ (see Figure 4.2). One may argue that the optimal value function is then not well-defined. The concern is about the risk aversion function ψ being non-defined (division by zero). Naturally, the strategy (c^*, π^*) defined by Proposition 4.4.1 is given as the limit of the series of strategies $(c_n^*, \pi^*)_{n=1,2,\ldots}$ where the expression in (4.54) is strictly positive but tends to zero. From Appendix 4.6.2 formula (4.89) we have that $x + K^{(c_n^*)}$ follows a geometric Brownian motion. More precise, $X^{c_n^*,\pi^*}(T) = (x + K^{(c_n^*)}(t,x)) \exp(\ldots)$ where the stochastic exponential term depends on the dynamics of W. We conclude that the value function is well-defined since

$$\begin{aligned}
& E_{0,x_0}\left[\int_0^T e^{-\rho s} c_n^*(s)ds + e^{-\rho T} X^{c_n^*,\pi^*}(T)\right] \\
& - \frac{\gamma}{2\left(x_0 + K^{(c_n^*)}(0,x_0)\right)} Var_{0,x_0}\left[\int_0^T e^{-\rho s} c_n^*(s)ds + e^{-\rho T} X^{c_n^*,\pi^*}(T)\right] \\
= & \int_0^T e^{-\rho s} c_n^*(s)ds + \underbrace{\left(x_0 + K^{(c_n^*)}(0,x_0)\right)}_{\searrow 0} \exp(\ldots) - \underbrace{\frac{\gamma\left(x_0 + K^{(c_n^*)}(0,x_0)\right)^2}{2\left(x_0 + K^{(c_n^*)}(0,x_0)\right)}}_{\searrow 0} \underbrace{Var_{0,x_0}\left[\exp(\ldots)\right]}_{\searrow 0} \\
\longrightarrow & \int_0^T e^{-\rho s} c^*(s)ds,
\end{aligned}$$

which coincides with Proposition 4.4.1.

[7]See Appendix 4.6.2 formula (4.89) where we, for the non-binding case, show that $x + K^{(c^*)}$ is a geometric Brownian motion.

4.4.2 Discussion of the solution

The investment strategy

From the expression of the optimal investment strategy we see that the optimal amount of money to invest in stocks is proportional to wealth plus the financial value of future labor income net of consumption. From an economic point of view this seems to be a fairly reasonable investment strategy. First of all, a rich investor should invest more in stocks than a poor investor. Second, if we know that a large amount of money will be injected continuously into the savings, then we should also invest a large amount of money in stocks.

Two important questions are:

- How does the optimal proportion of wealth to invest in stocks develop as time goes by?
- How do shifts in parameter values (γ in particular) influence the optimal proportion of wealth to invest in stocks?

In order to answer those questions define

$$\widetilde{\pi}^*(t) = \frac{\alpha - r}{\sigma^2 \gamma f(t)} \left[a(t) + \gamma \left(a(t)^2 - f(t)\right)\right].$$

Then

$$\pi^*(t,x)x = \widetilde{\pi}^*(t)\left(x + K^{(c^*)}(t)\right).$$

In Appendix 4.6.2 we show that we have the following integral equation for $\widetilde{\pi}^*$:

$$\widetilde{\pi}^*(t) = \frac{\alpha - r}{\sigma^2 \gamma} \left\{ e^{-\int_t^T \left[(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)+\sigma^2\widetilde{\pi}^*(s)^2\right] ds} + \gamma e^{-\int_t^T \sigma^2 \widetilde{\pi}^*(s)^2 ds} - \gamma \right\}.$$

We recognize this integral equation from Björk et al. (2012) who derive this for the case without consumption and labor income. If $r > \rho$ we conclude that $\widetilde{\pi}^*$ is increasing in time. On the other hand we have, for the non-binding case, that $(X^{c^*,\pi^*}(t) + K^{(c^*)}(t))/X^{c^*,\pi^*}(t)$ is expected to decrease with time. How fast $\widetilde{\pi}^*$ increases and how fast $(X^{c^*,\pi^*}(t) + K^{(c^*)}(t))/X^{c^*,\pi^*}(t)$ decreases depends in a complex way on the value of γ[8]. From the optimization problem (4.37) we can argue that (since a smaller value of γ corresponds to giving the variance term a smaller weight) we expect a smaller value of γ to imply a more aggressive investment strategy in general, i.e. a larger value of $\widetilde{\pi}$. Due to the complexity of the system of partial differential equations given in (4.52) and (4.53) it seems difficult to prove this. For a average scenario it is in Figure 4.1 illustrated that a smaller γ indeed implies a more aggressive investment strategy. For small values of γ we also note that we should indeed expect the optimal proportion of wealth to invest in stocks to increase over time. However, as seen in Figure 4.1 ($\gamma = 8$), we also have that for γ large enough the optimal proportion of wealth to invest in stocks seems to be approximately constant.

The consumption strategy

As already mentioned the upper and lower boundary for the consumption rate has a natural interpretation when we think of wealth as a pension saving account (see Subsection 4.3.2).

Let us consider the non-binding case ($x_0 + K^{(c^*)}(0, x_0) > 0$). From the expression of the optimal consumption strategy we see that it is optimal either to consume the maximum or minimum allowed dependent on whenever the deterministic expression $a(t) + \frac{\gamma}{2}\left(a(t)^2 - f(t)\right)$ is smaller or larger than 1. Since the functions a and f are given by the non-linear system of

[8] $K^{(c^*)}$ depends on γ since obviously the optimal consumption strategy does so.

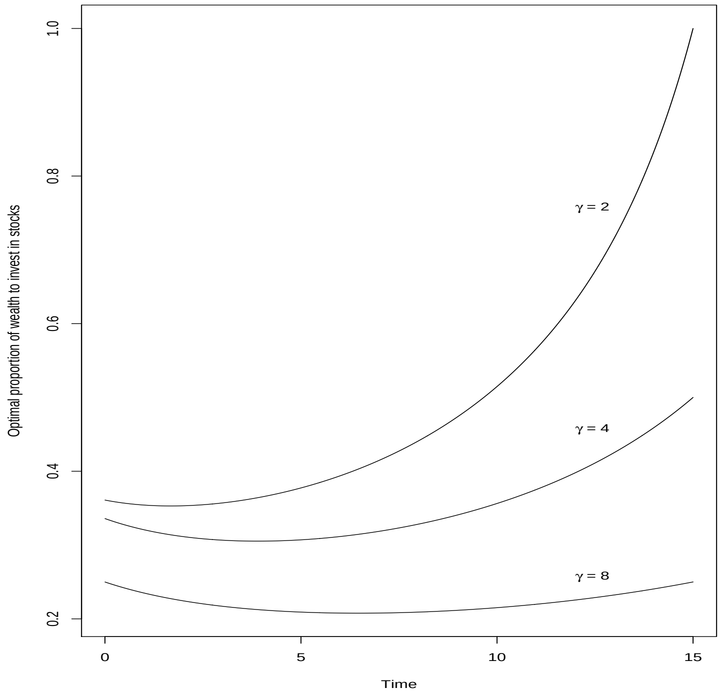

Figure 4.1: Parameter values are $T = 15,\ r = 0.04,\ \alpha = 0.12,\ \sigma = 0.2$ and $\rho = 0.02$. The initial wealth is $X_0 = 1000000$ DKK, labor income is during the hole time period 30000 DKK/month and the minimal consumption allowed (which turns out to equal the optimal consumption for all choices of γ) is during the hole time period 21000 DKK/month.

differential equations (4.52) and (4.53) it is difficult to analyze the optimal consumption strategy. However, some insight can be obtained by the following calculations:

$$\begin{aligned}
&\frac{\partial}{\partial t}\left(a(t) + \frac{\gamma}{2}(a(t)^2 - f(t))\right) \\
= \;& a'(t) + \gamma a(t)a'(t) - \frac{\gamma}{2} f'(t) \\
= \;& -\left\{(r-\rho) + \frac{(\alpha - r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]\right\} a(t) \\
& -\left\{(r-\rho) + \frac{(\alpha - r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]\right\} \gamma a(t)^2 \\
& +\left\{(r-\rho) + \frac{(\alpha - r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]\right\} \gamma f(t) \\
& +\frac{(\alpha - r)^2}{\sigma^2\gamma^2 f(t)^2}\left[a(t) + \gamma(a(t)^2 - f(t))\right]^2 \frac{\gamma}{2} f(t) \\
= \;& -\left\{(r-\rho) + \frac{(\alpha - r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]\right\}\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right] \\
& +\frac{1}{2}\frac{(\alpha - r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]^2 \\
= \;& -(r-\rho)\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right] - \frac{1}{2}\frac{(\alpha - r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]^2. \qquad (4.55)
\end{aligned}$$

Since for all $t \in [0,T]$ we have $a(t) + \gamma\left(a(t)^2 - f(t)\right) > 0$ [9], and $f(t) > 0$ [10], we can for the case $r \geq \rho$ conclude that $\frac{\partial}{\partial t}\left(a(t) + \frac{\gamma}{2}(a(t)^2 - f(t))\right) < 0,\ \forall t \in [0,T]$. Since we also have the terminal condition $\left(a(T) + \frac{\gamma}{2}(a(T)^2 - f(T))\right) = 1$ we can now make the following statement:

$$r \geq \rho \Longrightarrow \left(a(t) + \frac{\gamma}{2}(a(t)^2 - f(t))\right) > 1 \Longrightarrow c^*(t) = c_{\min}(t),\ \forall t \in [0,T). \tag{4.56}$$

If $r < \rho$ the expression (4.55) consists of a positive term minus a positive term, and we can therefore not make any conclusions about the optimal consumption behavior. Figure 4.2 shows how the optimal consumption strategy looks in the case of $r < \rho$. We see that for γ small enough it is optimal to consume the minimum allowed during the whole time period, and for γ large enough it is optimal to consume the maximum allowed during the whole time period. For certain values of γ we see that it is optimal in the beginning to consume the maximum allowed and then later to consume the minimal allowed.

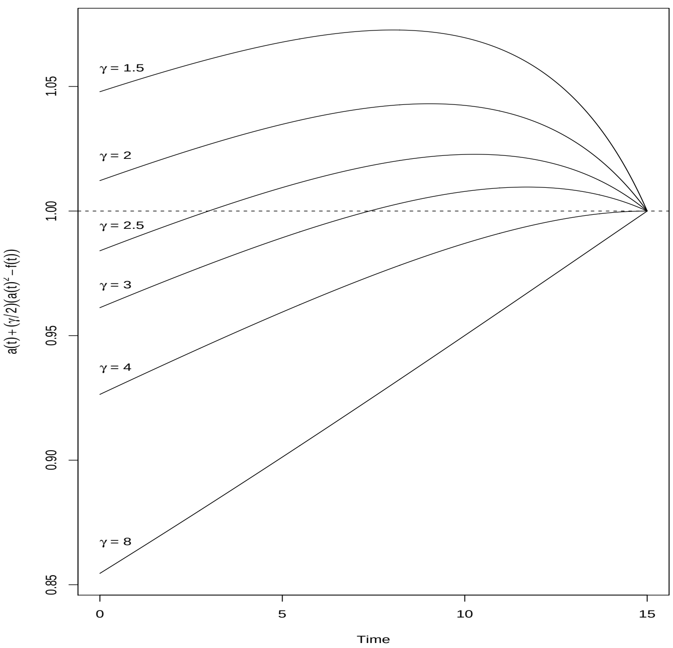

Figure 4.2: Parameter values are $T = 15$, $r = 0.04$, $\alpha = 0.12$, $\sigma = 0.2$ and $\rho = 0.06$. Points falling below $y = 1$ corresponds to an optimal consumption rate equal to the maximum allowed and points falling above $y = 1$ corresponds to an optimal consumption rate equal to the minimum allowed.

[9] The optimal investment strategy is given by

$$\pi^*(t,x)x = \frac{\alpha - r}{\sigma^2 \gamma f(t)}\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right]\left(x + K^{(c^*)}(t)\right).$$

From (4.89) we have (for the non-binding case) that $X^{c^*,\pi^*}(t) + K^{(c^*)(t)} > 0,\ t \in [0,T]$, and it follows that the optimal amount of money to invest in stocks is strictly positive over the interval $[0,T]$ iff $a(t) + \gamma\left(a(t)^2 - f(t)\right) > 0,\ \forall t \in [0,T]$. Since for a given strategy π_1 the strategy defined by $\pi_2(t,x)x = \max(0, \pi_1(t,x)x)$ results in a larger mean and a smaller variance of the terminal wealth, compared to the strategy π_1, we conclude that it will never be optimal to invest a negative amount of money in stocks. We therefore conclude that we must have $a(t) + \gamma\left(a(t)^2 - f(t)\right) \geq 0,\ \forall t \in [0,T]$. By the terminal condition $\left(a(T) + \frac{\gamma}{2}(a(T)^2 - f(T))\right) = 1$ and (4.55) we have that the slope is negative at $t = T$, and we therefore get the strict inequality $a(t) + \gamma\left(a(t)^2 - f(t)\right) > 0,\ \forall t \in [0,T]$.

[10] By (4.94) this is clearly the case.

In order to get a better understanding of the optimal consumption strategy we make the following observation. If we take the optimal investment strategy as given we can try to comment on the result given by (4.56) and Figure 4.2. First of all, in contrast to the case with constant risk aversion, the deterministic consumption strategy does influence the variance term. If we, during an infinitesimal time interval $[t, t+dt]$, choose to consume the minimum allowed we save the amount of money $(c_{\max} - c_{\min})dt$ in addition to $(\ell - c_{\max})\, dt$. The fraction $\widetilde{\pi}^*$ of these money is invested in stocks (which do contribute to the variance term) and the rest is invested in the bank account (which do not contribute to the variance term).

For the case $r \geq \rho$ the expected investment return on the saved amount of money is larger than ρ, and the mean term in (4.37) can therefore, in terms of consumption, be maximized by choosing the minimum consumption rate allowed. We also have that a minimum consumption rate minimizes the variance weight term $\gamma/2(x + K^{(c)}(t))$, since a minimum consumption rate maximizes $K^{(c)}$. On the other hand, choosing to consume the minimum allowed also maximizes the variance term in (4.37) through the investment. From (4.56) we can conclude that maximizing the mean term and minimizing the variance weight term (in terms of consumption) makes it more than up for a larger variance term (in terms of consumption through the investment strategy).

The chain of reasoning seems to stay true for the case $r < \rho$. In this case the expected return on the saved money is larger than ρ if and only if $\widetilde{\pi}^*$ is big enough. From Figure 4.1 we have that this is the case for a small enough value of γ. Correspondingly, Figure 4.2 shows that it is optimal to consume the minimum (maximum) allowed for a small (large) enough value of γ. The reason that it is optimal for an investor with a given value of γ (not too large neither too small) first to consume the maximum allowed and after some time the minimum allowed is due to the fact that $\widetilde{\pi}^*$ is increasing in time.

4.5 Mean-standard deviation without pre-commitment

In this section we study and solve the mean-standard deviation problem without pre-commitment including consumption and terminal wealth. We consider a general risk aversion function ψ, which we assume fulfills some reasonable constraints, and show that the optimal investment strategy becomes the same ($\pi^* = 0$) for all variants of ψ. Finally, we give an interpretation of the optimal strategy which also helps us understand the fundamental difference between pre-commitment and without pre-commitment.

4.5.1 Presenting and solving the problem

Consider the problem of finding the optimal strategy for the objective given by

$$E_{0,x_0}\left[\int_0^T e^{-\rho s} c(s) ds + e^{-\rho T} X^{c,\pi}(T)\right] - \psi(0,x_0)\left(Var_{0,x_0}\left[\int_0^T e^{-\rho s} c(s) ds + e^{-\rho T} X^{c,\pi}(T)\right]\right)^{\frac{1}{2}}, \tag{4.57}$$

where $\psi \in \mathcal{C}^{1,2}$, $\psi > 0$, is a risk aversion function with ψ_x, ψ_{xx} and ψ_t finite $\forall (t,x) \in [0,T] \times \mathbb{R}$, and where we again restrict the admissible strategies to $(c,\pi) \in \mathcal{A} \bigcap (\mathcal{D} \times \mathbb{R})$, where $\mathcal{D}(s) = [c_{\min}(s), c_{\max}(s)]$, $s \in [0,T]$, is a finite time dependent interval. The corresponding function f is given by

$$f(t,x,y,z) = y - \psi(t,x)\left(z - y^2\right)^{\frac{1}{2}}.$$

The system of partial differential equations we want to solve in order to obtain the optimal value function and the optimal strategy is given by (4.16) and (4.17)[11]. Clearly,

$$
\begin{aligned}
f_t &= -\psi_t \left(z-y^2\right)^{\frac{1}{2}},\ f_x = -\psi_x \left(z-y^2\right)^{\frac{1}{2}},\ f_{xx} = -\psi_{xx} \left(z-y^2\right)^{\frac{1}{2}}, \\
f_y &= 1+\psi \left(z-y^2\right)^{\frac{-1}{2}} y,\ f_{yy} = \psi \left(z-y^2\right)^{\frac{-3}{2}} y^2 + \psi \left(z-y^2\right)^{\frac{-1}{2}}, \\
f_z &= -\frac{\psi}{2} \left(z-y^2\right)^{\frac{-1}{2}},\ f_{zz} = \frac{\psi}{4} \left(z-y^2\right)^{\frac{-3}{2}}, \\
f_{xz} &= -\frac{\psi_x}{2} \left(z-y^2\right)^{\frac{-1}{2}},\ f_{xy} = \psi_x \left(z-y^2\right)^{\frac{-1}{2}} y,\ f_{yz} = \frac{-\psi}{2} \left(z-y^2\right)^{\frac{-3}{2}} y.
\end{aligned}
$$

Inserting this in (4.19)–(4.21) we obtain the following expression

$$
Q = -\psi_x \left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{\frac{1}{2}}, \tag{4.58}
$$

$$
\begin{aligned}
U = \frac{\psi}{4} &\left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{-\frac{3}{2}} \left(F^{(2)}_x - 2F^{(1)}F^{(1)}_x\right)^2 + \psi \left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{-\frac{1}{2}} \left(F^{(1)}_x\right)^2 \\
&- \psi_{xx} \left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{\frac{1}{2}} - \psi_x \left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{-\frac{1}{2}} \left(F^{(2)}_x - 2F^{(1)}F^{(1)}_x\right),
\end{aligned} \tag{4.59}
$$

$$
J = \rho F^{(1)} - c - \psi\rho \left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{\frac{1}{2}} - \psi_t \left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{\frac{1}{2}}. \tag{4.60}
$$

We are now ready to derive a candidate for the optimal strategy. Again this is done by differentiating with respect to c and π inside the curly brackets in (4.16). We get

$$
c^*(t,x) = \begin{cases} c_{\max}(t), & \text{if } C(t,x) < 0, \\ \text{non-defined}, & \text{if } C(t,x) = 0, \\ c_{\min}(t), & \text{if } C(t,x) > 0, \end{cases} \tag{4.61}
$$

where

$$
C(t,x) := F_x(x,t) - Q(x,t) - 1 - \left(\frac{\partial}{\partial c}\psi_t\right) \left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{\frac{1}{2}},
$$

and

$$
\pi^*(t,x)x = -\frac{\alpha - r}{\sigma^2} \frac{F_x - Q}{F_{xx} - U}, \tag{4.62}
$$

(provided $U > F_{xx}$). We now search for solutions in the form,

$$
\begin{aligned}
F(t,x) &= A(t)x + B(t), \\
F^{(1)}(t,x) &= a(t)x + b(t),
\end{aligned}
$$

where A, B, a, and b are deterministic functions of time. The derivatives of interest are

$$
\begin{aligned}
F_t = A'(t)x + B'(t),\ F_x = A(t),\ F_{xx} = 0, \\
F^{(1)}_t = a'(t)x + b'(t),\ F^{(1)}_x = a(t),\ F^{(1)}_{xx} = 0.
\end{aligned}
$$

[11]Since $F_t = F^{(1)}_t - \frac{\partial}{\partial t}\psi(t,x)\left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{\frac{1}{2}} - \frac{\psi(t,x)}{2}\left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{-\frac{1}{2}}\left(F^{(2)}_t - 2F^{(1)}F^{(1)}_t\right)$ we only need to solve two of the three differential equations given by (4.16)–(4.18). We choose to solve (4.16) and (4.17).

Inserting this into the system of partial differential equations given by (4.16) and (4.17) we find

$$\begin{aligned}A_t x + B_t &= -rx(A-Q) - \frac{1}{2}\pi^*(\alpha - r)x(A-Q) - \ell(A-Q) + c^*(A-Q) + J,\\ a_t x + b_t &= -rxa - \pi^*(\alpha-r)xa - \ell a + c^*(a-1) + \rho(ax+b).\end{aligned}$$

Quite surprisingly, this system of differential equations is solved by $\pi^* = 0$ via the parametrization $F^{(2)} = \left(F^{(1)}\right)^2$ [12]. Note that for this solution, actually U is infinite. However since $\pi^* U$ is finite, the solution is admissible. With $\pi^* = 0$ the system of partial differential equations reduces to

$$\begin{aligned}A_t x + B_t &= -rxA + \rho x a - \ell A + c^*(A-1) + \rho b,\\ A(T) &= 1,\\ B(T) &= 0,\\ a_t x + b_t &= -(r-\rho)xa - \ell a + c^*(a-1) + \rho b,\\ a(T) &= 1,\\ b(T) &= 0,\end{aligned}$$

where we have added the terminal conditions. We obtain the solutions

$$A(t) = a(t) = e^{(r-\rho)(T-t)},$$

and

$$B(t) = b(t) = e^{\rho t}\int_t^T \left[\ell(s)e^{(r-\rho)(T-s)} - c^*(s)\left(e^{(r-\rho)(T-s)} - 1\right)\right] e^{-\rho s} ds.$$

The optimal strategy now follows directly by plugging in the solutions for A and a together with the partial derivatives and the relation $F^{(2)} = \left(F^{(1)}\right)^2$ into (4.61) and (4.62). We summarize the results as follows.

[12] $F^{(2)} = \left(F^{(1)}\right)^2$ implies that $F_x^{(2)} - 2F^{(1)}F_x^{(1)} = 0$. Inserting this into (4.58) and (4.59) gives

$$\begin{aligned}Q &= 0,\\ U &= \psi(t,x)\left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{-\frac{1}{2}}\left(F_x^{(1)}\right)^2.\end{aligned}$$

Formula (4.62) now reduces to

$$\pi^*(t,x)x = -\frac{\alpha - r}{\sigma^2}\frac{A(t)\left(F^{(2)} - \left(F^{(1)}\right)^2\right)^{\frac{1}{2}}}{\psi(t,x)\left(F_x^{(1)}\right)^2}.$$

It is then clear that

$$\left\{\left(F^{(2)} = \left(F^{(1)}\right)^2\right) \Rightarrow (\pi^* = 0)\right\} \Leftrightarrow \left\{F_x^{(1)} \neq 0 \text{ and } A(t) \text{ bounded from above}\right\}.$$

Below we show that $F_x^{(1)} = a(t) = A(t) = e^{r(T-t)} > 0$, and clearly we get $\pi^* = 0$.

Proposition 4.5.1. *For the mean-standard deviation problem given by (4.57) we have the following results.*

- *The optimal strategy is given by*

$$c^*(t,x) = \begin{cases} c_{\max}(t), & \text{if } r < \rho, \\ \text{non-defined}, & \text{if } r = \rho, \\ c_{\min}(t), & \text{if } r > \rho, \end{cases}$$
$$\pi^*(t,x) = 0.$$

- *The optimal value function is given by*

$$V(t,x) = e^{(r-\rho)(T-t)}x + B(t).$$

- *The conditional expected value of the discounted cumulated optimal consumption plus the discounted optimal terminal wealth is given by*

$$E_{t,x}\left[\int_t^T e^{-\rho(s-t)}c^*(s)ds + e^{-\rho(T-t)}X^{c^*,\pi^*}(T)\right] = e^{(r-\rho)(T-t)}x + b(t).$$

- *The conditional variance of the discounted cumulated optimal consumption plus the discounted optimal terminal wealth is given by*

$$Var_{t,x}\left[\int_t^T e^{-\rho(s-t)}c^*(s)ds + e^{-\rho(T-t)}X^{c^*,\pi^*}(T)\right] = 0.$$

Here

$$B(t) = b(t) = e^{\rho t}\int_t^T \left[\ell(s)e^{(r-\rho)(T-s)} - c^*(s)\left(e^{(r-\rho)(T-s)} - 1\right)\right]e^{-\rho s}ds.$$

4.5.2 Discussion of the solution

The optimal consumption strategy coincides with the one found in Chapter 4.3. We refer to Subsection 4.3.2 for an interpretation.

The following intuitive explanation of the somehow surprising optimal investment strategy ($\pi^* = 0$) actually provides us with some helpful insight to understand the nature of inconsistent stochastic optimization problems. In *discrete* time, the definition of a subgame Nash equilibrium strategy is: Consider n players and split the time interval $[T_0, T_n]$ in n equally long intervals. Let player m, $1 \le m \le n$, decide on the strategy (c^*_{m-1}, π^*_{m-1}) used over the interval $[T_{m-1}, T_m)$.

- The equilibrium control (c^*_{n-1}, π^*_{n-1}) is obtained by letting player n optimize the value function at time T_{n-1}.
- The equilibrium control (c^*_{n-2}, π^*_{n-2}) is obtained by letting player $n-1$ optimize the value function at time T_{n-2} given the knowledge that player number n will use the strategy (c^*_{n-1}, π^*_{n-1}).
- Proceed recursively by induction.

Now, consider a single period pure portfolio optimization problem with time horizon Δt given by

$$V(0,x) = \sup_{\pi}\left\{E_{0,x}\left[X^{\pi}(\Delta t)\right] - \psi(0,x)\left(Var_{0,x}\left[X^{\pi}(\Delta t)\right]\right)^{\frac{1}{2}}\right\},$$

where $X^{c,\pi}(\Delta t) = r\Delta t(1-\pi)x + R\pi x$ and R is a random variable such that $E[R] = \alpha\Delta t$ and $Var[R] = \sigma^2\Delta t$. Then we have

$$
\begin{aligned}
&E_{0,x}\left[X^{\pi}(\Delta t)\right] - \psi(0,x)\left(Var_{0,x}\left[X^{\pi}(\Delta t)\right]\right)^{\frac{1}{2}} \\
&= r\Delta t x + (\alpha - r)\pi x\Delta t - \psi(0,x)\left(\sigma^2\Delta t\pi^2 x^2\right)^{\frac{1}{2}} \\
&= r\Delta t x + (\alpha - r)\pi x\Delta t - \psi(0,x)\sigma\sqrt{\Delta t}|\pi|x.
\end{aligned}
$$

We directly obtain the optimal strategy

$$
\pi^* = \begin{cases} 0, & \text{if } (\alpha - r)\Delta t < \psi(0,x)\sigma\sqrt{\Delta t}, \\ \text{non-defined}(\mathbb{R}_+), & \text{if } (\alpha - r)\Delta t = \psi(0,x)\sigma\sqrt{\Delta t}, \\ \infty, & \text{if } (\alpha - r)\Delta t > \psi(0,x)\sigma\sqrt{\Delta t}. \end{cases}
$$

Clearly, for any risk aversion function $\psi(0,x)$ there exist a small enough Δt such that $(\alpha - r)\Delta t < \psi(0,x)\sigma\sqrt{\Delta t}$, i.e. such that $\pi^* = 0$ [13]. Likewise, for a multi period problem we conclude that for any risk aversion function $\psi(t,x)$, fulfilling our reasonable assumptions, we get the Nash equilibrium strategy $\pi^* = 0$ whenever the discretization of the interval of optimization is fine enough. To obtain this note, by the argumentation above, that for a fine enough discretization of the interval player n (the last player in the game) chooses to take no risk at all. Consequently, player number $n-1$ faces, since there is no randomness after time T_{n-1}, also a single period problem and by the same argumentation player number $n-1$ also chooses to take no risk at all. Proceeding recursively we obtain $\pi^* = 0$ for all players.

For any risk aversion function $\psi(t,x)$, fulfilling our reasonable assumptions, we now conclude that over an infinitesimal time interval, dt, standard deviation is of the order $\sqrt{dt}$, which means that the punishment is so hard that any risk taking is unattractive.

Acknowledgments

The authors wish to thank the anonymous referee for an extremely detailed and valuable correspondence and Esben Masotti Kryger for valuable discussions and comments.

4.6 Appendix

4.6.1

Proof of Theorem 4.2.1. Consider an arbitrary admissible strategy (c,π).

1: First we argue that if there exist a function Y such that, $\forall (t,x) \in (0,T) \times \mathbb{R}$, we have

$$
Y_t = -[(r + \pi(\alpha - r))x + \ell - c]Y_x - \frac{1}{2}\pi^2\sigma^2 x^2 Y_{xx} - c + \rho Y, \tag{4.63}
$$

$$
Y(T,x) = x, \tag{4.64}
$$

then

$$
Y(t,x) = y^{c,\pi}(t,x), \tag{4.65}
$$

where $y^{c,\pi}$ is given by (4.14).

[13] This is also indicated by Zhou and Li (2000) who get the same solution for a sufficiently short time horizon.

To show this define

$$\widetilde{Y}(t,x) = e^{-\rho t} Y(t,x). \tag{4.66}$$

From (4.63) and (4.64) we get, $\forall (t,x) \in (0,T) \times \mathbb{R}$, that

$$\widetilde{Y}_t = -[(r + \pi(\alpha - r))x + \ell - c]\widetilde{Y}_x - \frac{1}{2}\pi^2\sigma^2 x^2 \widetilde{Y}_{xx} - e^{-\rho t}c, \tag{4.67}$$

$$\widetilde{Y}(T,x) = e^{-\rho T}x. \tag{4.68}$$

Using Itô's formula

$$\begin{aligned}
& \widetilde{Y}(t, X^{c,\pi}(t)) \\
= & -\int_t^T d\widetilde{Y}(s, X^{c,\pi}(s)) + \widetilde{Y}(T, X^{c,\pi}(T)) \\
= & -\int_t^T \Big(\widetilde{Y}_s(s, X^{c,\pi}(s)) + [(r + \pi(s)(\alpha - r))X^{c,\pi}(s) + \ell(s) - c(s)]\, \widetilde{Y}_x(s, X^{c,\pi}(s)) \\
& + \frac{1}{2}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 \widetilde{Y}_{xx}(s, X^{c,\pi}(s))\Big) ds \\
& - \int_t^T \pi(s)\sigma X^{c,\pi}(s)\widetilde{Y}_x(s, X^{c,\pi}(s))dW(s) + \widetilde{Y}(T, X^{c,\pi}(T)).
\end{aligned}$$

Plug in (4.67) and (4.68) to obtain

$$\widetilde{Y}(t, X^{c,\pi}(t)) = \int_t^T e^{-\rho s}c(s)ds + e^{-\rho T}X^{c,\pi}(T) - \int_t^T \pi(s)\sigma X^{c,\pi}(s)\widetilde{Y}_x(s, X^{c,\pi}(s))dW(s). \tag{4.69}$$

Since (c,π) is an admissible strategy taking the conditional expectation given $X(t) = x$ on both sides of the equality leaves us with

$$\widetilde{Y}(t,x) = E\left[\int_t^T e^{-\rho s}c(s)ds + e^{-\rho T}X^{c,\pi}(T) \,\middle|\, X(t) = x\right].$$

It is now clear that

$$Y(t,x) = e^{\rho t}\widetilde{Y}(t,x) = y^{c,\pi}(t,x).$$

2: Second we argue that if there exist a function Z such that, $\forall (t,x) \in (0,T) \times \mathbb{R}$, we have

$$Z_t = -[(r + \pi(\alpha - r))x + \ell - c]Z_x - \frac{1}{2}\pi^2\sigma^2 x^2 Z_{xx} - 2cY + 2\rho Z, \tag{4.70}$$

$$Z(T,x) = x^2, \tag{4.71}$$

then

$$Z(t,x) = z^{c,\pi}(t,x), \tag{4.72}$$

where $z^{c,\pi}$ is given by (4.15).

To show this define $\widetilde{Z}(t,x) = e^{-2\rho t}Z(t,x)$. From (4.70) and (4.71) we get, $\forall (t,x) \in (0,T) \times \mathbb{R}$, that

$$\widetilde{Z}_t = -[(r + \pi(\alpha - r))x + \ell - c]\widetilde{Z}_x - \frac{1}{2}\pi^2\sigma^2 x^2 \widetilde{Z}_{xx} - 2e^{-\rho t}c\widetilde{Y}, \tag{4.73}$$

$$\widetilde{Z}(T,x) = \left(e^{-\rho T}x\right)^2, \tag{4.74}$$

where $\widetilde{Y}(t,x)$ is given in (4.66). Using Itô's formula

$$
\begin{aligned}
&\widetilde{Z}(t, X^{c,\pi}(t)) \\
&= -\int_t^T d\widetilde{Z}(s, X^{c,\pi}(s)) + \widetilde{Z}(T, X^{c,\pi}(T)) \\
&= -\int_t^T \Big(\widetilde{Z}_s(s, X^{c,\pi}(s)) + [(r+\pi(s)(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)]\,\widetilde{Z}_x(s, X^{c,\pi}(s)) \\
&+ \frac{1}{2}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 \widetilde{Z}_{xx}(s, X^{c,\pi}(s))\Big)ds \\
&- \int_t^T \pi(s)\sigma X^{c,\pi}(s)\widetilde{Z}_x(s, X^{c,\pi}(s))dW(s) + \widetilde{Z}(T, X^{c,\pi}(T)).
\end{aligned}
$$

Plug in (4.73) and (4.74) and thereafter (4.69) to obtain

$$
\begin{aligned}
&\widetilde{Z}(t, X^{c,\pi}(t)) \\
&= 2\int_t^T e^{-\rho s}c(s)\widetilde{Y}(s, X^{c,\pi}(s))ds + \left(e^{-\rho T}X^{c,\pi}(T)\right)^2 \qquad (4.75)\\
&- \int_t^T \pi(s)\sigma X^{c,\pi}(s)\widetilde{Z}_x(s, X^{c,\pi}(s))dW(s) \\
&= 2\int_t^T e^{-\rho s}c(s)\int_s^T e^{-\rho y}c(y)dyds + 2\int_t^T e^{-\rho s}c(s)ds\; e^{-\rho T}X^{c,\pi}(T) \\
&- 2\int_t^T e^{-\rho s}c(s)\int_s^T \pi(y)\sigma X^{c,\pi}(y)\widetilde{Y}_x(y, X^{c,\pi}(y))dW(y)ds \\
&+ \left(e^{-\rho T}X^{c,\pi}(T)\right)^2 - \int_t^T \pi(s)\sigma X^{c,\pi}(s)\widetilde{Z}_x(s, X^{c,\pi}(s))dW(s). \qquad (4.76)
\end{aligned}
$$

Since (c,π) is an admissible strategy taking the conditional expectation given $X(t)=x$ on both sides of the equality leaves us with

$$
\begin{aligned}
\widetilde{Z}(t,x) = E\Big[&2\int_t^T e^{-\rho s}c(s)\int_s^T e^{-\rho y}c(y)dyds \\
&+ 2\int_t^T e^{-\rho s}c(s)ds\; e^{-\rho T}X^{c,\pi}(T) + \left(e^{-\rho T}X^{c,\pi}(T)\right)^2 \;\Big|\; X(t)=x\Big].
\end{aligned}
$$

Provided that $2\int_t^T e^{-\rho s}c(s)\int_s^T e^{-\rho y}c(y)dyds = \left(\int_t^T e^{-\rho s}c(s)ds\right)^2$ we now have that

$$
\widetilde{Z}(t,x) = E\left[\left(\int_t^T e^{-\rho s}c(s)ds + e^{-\rho T}X^{c,\pi}(T)\right)^2 \;\middle|\; X(t)=x\right].
$$

This is however easily realized since

$$
\frac{\partial}{\partial t}\left(2\int_t^T e^{-\rho s}c(s)\int_s^T e^{-\rho y}c(y)dyds\right) = -2e^{-\rho t}c(t)\int_t^T e^{-\rho y}c(y)dy = \frac{\partial}{\partial t}\left(\int_t^T e^{-\rho s}c(s)ds\right)^2
$$

$$
\left(2\int_t^T e^{-\rho s}c(s)\int_s^T e^{-\rho y}c(y)dyds\right)_{\big|t=T} = 0 = \left(\int_t^T e^{-\rho s}c(s)ds\right)^2_{\big|t=T}.
$$

It is now clear that

$$Z(t,x) = e^{2\rho t}\widetilde{Z}(t,x) = z^{c,\pi}(t,x).$$

3: At last we argue that if there exist a function F such that, $\forall(t,x) \in (0,T)\times\mathbb{R}$, we have

$$F_t = \inf_{(c,\pi)\in\mathcal{A}} \left\{ -[(r+\pi(\alpha-r))x + \ell - c](F_x - f_x) - \frac{1}{2}\pi^2\sigma^2x^2(F_{xx} - U) + J \right\}, \tag{4.77}$$

$$F(T,x) = f\left(T,x,x,x^2\right), \tag{4.78}$$

where

$$U = f_{xx} + \left(F_x^{(1)}\right)^2 f_{yy} + \left(F_x^{(2)}\right)^2 f_{zz} + 2F_x^{(1)}F_x^{(2)}f_{yz} + 2F_x^{(1)}f_{xy} + 2F_x^{(2)}f_{xz}, \tag{4.79}$$

$$J = \left(\rho F^{(1)} - c\right) f_y + 2\left(\rho F^{(2)} - cF^{(1)}\right) f_z + f_t, \tag{4.80}$$

with $F^{(1)}$ and $F^{(2)}$ fulfilling (4.17) and (4.18), respectively, then

$$F(t,x) = V(t,x),$$

where V is the optimal value function defined by (4.11).

First step is to derive an expression for

$$f(t, X^{c,\pi}(t), y^{c,\pi}(t, X^{c,\pi}(t)), z^{c,\pi}(t, X^{c,\pi}(t))).$$

By (4.65) and (4.72) this equals

$$f(t, X^{c,\pi}(t), Y(t, X^{c,\pi}(t)), Z(t, X^{c,\pi}(t))).$$

Using Itô's formula (we have assumed $f \in \mathcal{C}^{1,2,2,2}$) we get that

$$\begin{aligned}
& f(t, X^{c,\pi}(t), y^{c,\pi}(t, X^{c,\pi}(t)), z^{c,\pi}(t, X^{c,\pi}(t))) \\
= \; & -\int_t^T df(s, X^{c,\pi}(s), Y(s, X^{c,\pi}(s)), Z(s, X^{c,\pi}(s))) \\
& + f(T, X^{c,\pi}(T), Y(T, X^{c,\pi}(T)), Z(T, X^{c,\pi}(T))) \\
= \; & -\int_t^T \Big\{ (f_s + f_yY_s + f_zZ_s)ds + (f_x + f_yY_x + f_zZ_x)dX^{c,\pi}(s) \\
& + \frac{1}{2}\Big[f_yY_{xx} + f_zZ_{xx} + f_{xx} + f_{yy}(Y_x)^2 + f_{zz}(Z_x)^2 \\
& + 2f_{xy}Y_x + 2f_{xz}Z_x + 2f_{yz}Y_xZ_x \Big] \pi(s)^2\sigma^2 X^{c,\pi}(s)^2 ds \Big\} \\
& + f(T, X^{c,\pi}(T), Y(T, X^{c,\pi}(T)), Z(T, X^{c,\pi}(T))),
\end{aligned} \tag{4.81}$$

where we have skipped some arguments under the integral. Inserting (4.63), (4.70) and the dynamics of X given by (4.5) we have that

$$
\begin{aligned}
& f(t, X^{c,\pi}(t), y^{c,\pi}(t, X^{c,\pi}(t)), z^{c,\pi}(t, X^{c,\pi}(t))) \\
= \quad & -\int_t^T \Bigg\{ \Bigg[f_s + f_y \Bigg(-\big[(r+\pi(s)(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big] Y_x \\
& -\frac{1}{2}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 Y_{xx} - c(s) + \rho Y \Bigg) + f_z \Bigg(-\big[(r+\pi(s)(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big] Z_x \\
& -\frac{1}{2}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 Z_{xx} - 2c(s)Y + 2\rho Z \Bigg) \Bigg] ds \\
& +(f_x + f_y Y_x + f_z Z_x)\Big(\big[(r+\pi(s)(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big] ds + \pi(s)\sigma X^{c,\pi}(s) dW(s) \Big) \\
& +\frac{1}{2}\Big[f_y Y_{xx} + f_z Z_{xx} + f_{xx} + f_{yy}\left(Y_x\right)^2 + f_{zz}\left(Z_x\right)^2 \\
& +2f_{xy}Y_x + 2f_{xz}Z_x + 2f_{yz}Y_x Z_x \Big] \pi(s)^2\sigma^2 X^{c,\pi}(s)^2 ds \Bigg\} \\
& +f(T, X^{c,\pi}(T), Y(T, X^{c,\pi}(T)), Z(T, X^{c,\pi}(T))).
\end{aligned}
$$

After reduction this is

$$
\begin{aligned}
& f(t, X^{c,\pi}(t), y^{c,\pi}(t, X^{c,\pi}(t)), z^{c,\pi}(t, X^{c,\pi}(t))) \\
= \quad & -\int_t^T \Bigg\{ \Bigg(f_s + f_y\big(- c(s) + \rho Y \big) + f_z\big(- 2c(s)Y + 2\rho Z \big) \Bigg) ds \\
& +f_x\big[(r+\pi(s)(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big] ds + (f_x + f_y Y_x + f_z Z_x)\pi(s)\sigma X^{c,\pi}(s) dW(s) \\
& +\frac{1}{2}\Big[f_{xx} + f_{yy}\left(Y_x\right)^2 + f_{zz}\left(Z_x\right)^2 + 2f_{xy}Y_x + 2f_{xz}Z_x + 2f_{yz}Y_x Z_x \Big] \pi(s)^2\sigma^2 X^{c,\pi}(s)^2 ds \Bigg\} \\
& +f(T, X^{c,\pi}(T), Y(T, X^{c,\pi}(T)), Z(T, X^{c,\pi}(T))).
\end{aligned}
$$

Define for an arbitrary admissible strategy (c, π) the quantities corresponding to (4.79) and (4.80) by

$$
\widetilde{U} = f_{xx} + \left(Y_x\right)^2 f_{yy} + \left(Z_x\right)^2 f_{zz} + 2Y_x Z_x f_{yz} + 2Y_x f_{xy} + 2Z_x f_{xz}, \tag{4.82}
$$

$$
\widetilde{J} = \left(\rho Y - c\right) f_y + 2\left(\rho Z - cY\right) f_z + f_t. \tag{4.83}
$$

We now get

$$
\begin{aligned}
& f(t, X^{c,\pi}(t), y^{c,\pi}(t, X^{c,\pi}(t)), z^{c,\pi}(t, X^{c,\pi}(t))) \\
= \quad & -\int_t^T \Big\{ \widetilde{J}(s) + f_x\big[(r+\pi(s)(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big] ds \\
& +(f_x + f_y Y_x + f_z Z_x)\pi(s)\sigma X^{c,\pi}(s) dW(s) + \frac{1}{2}\widetilde{U}(s)\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 ds \Big\} \\
& +f(T, X^{c,\pi}(T), Y(T, X^{c,\pi}(T)), Z(T, X^{c,\pi}(T))).
\end{aligned} \tag{4.84}
$$

Next step is, by use of (4.84), to show that for any admissible strategy (c, π) we have

$$
\begin{aligned}
& f(t, x, y^{c,\pi}(t, x), z^{c,\pi}(t, x)) \\
\leq\ & F(t, x) + \int_t^T \left(J(s) - \widetilde{J}(s) + \frac{1}{2}\sigma^2\pi(s)^2 X^{c,\pi}(s)^2 (U(s) - \widetilde{U}(s)) \right) ds.
\end{aligned} \tag{4.85}
$$

By use of Itô's formula we get that

$$\begin{aligned}
F(t,X^{c,\pi}(t)) &= -\int_t^T dF(s,X^{c,\pi}(s)) + F(T,X^{c,\pi}(T)) \\
&= -\int_t^T \Big(F_s ds + F_x dX^{c,\pi}(s) + \frac{1}{2}F_{xx}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 ds\Big) \\
&\quad + F(T,X^{c,\pi}(T)).
\end{aligned}$$

Since F solves the pseudo Hamilton-Jacobi-Bellman equation (4.77) we see that for the arbitrary strategy (c,π) we have, $\forall (t,x) \in (0,T)\times\mathbb{R}$, that

$$F_t \le -\big[(r+\pi(\alpha-r))x + \ell - c\big](F_x - f_x) - \frac{1}{2}\pi^2\sigma^2 x^2(F_{xx} - U) + J.$$

Inserting this with $x = X^{c,\pi}(s)$, inserting the dynamics of X given by (4.5), and inserting the terminal conditions (4.78), (4.64) and (4.71) we get that

$$\begin{aligned}
F(t,X^{c,\pi}(t)) \ge -\int_t^T \Big\{ \Big(&- \big[(r+\pi(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big](F_x - f_x) \\
&- \frac{1}{2}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2(F_{xx} - U(s)) + J(s)\Big) ds \\
&+ F_x\Big(\big[(r+\pi(s)(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big]ds + \pi(s)\sigma X^{c,\pi}(s)dW(s)\Big) \\
&+ \frac{1}{2}F_{xx}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 ds \Big\} + f(T,X^{c,\pi}(T),Y(T,X^{c,\pi}(T)),Z(T,X^{c,\pi}(T))).
\end{aligned}$$

After reduction this is

$$\begin{aligned}
F(t,X^{c,\pi}(t)) \ge -\int_t^T &\Big\{ \Big(f_x\big[(r+\pi(\alpha-r))X^{c,\pi}(s) + \ell(s) - c(s)\big] \\
&+ \frac{1}{2}\pi(s)^2\sigma^2 X^{c,\pi}(s)^2 U(s) + J(s)\Big) ds + F_x\pi(s)\sigma X^{c,\pi}(s)dW(s) \Big\} \\
&+ f(T,X^{c,\pi}(T),Y(T,X^{c,\pi}(T)),Z(T,X^{c,\pi}(T))).
\end{aligned}$$

Insert (4.84) to obtain

$$\begin{aligned}
F(t,X^{c,\pi}(t)) \ge &\int_t^T \big(f_x + f_y Y_x + f_z Z_x - F_x\big)\pi(s)\sigma X^{c,\pi}(s)dW(s) \\
&- \int_t^T \Big(J(s) - \widetilde{J}(s) + \frac{1}{2}\sigma^2\pi(s)^2 X^{c,\pi}(s)^2(U(s) - \widetilde{U}(s))\Big) ds \\
&+ f(t,X^{c,\pi}(t),y^{c,\pi}(t,X^{c,\pi}(t)),z^{c,\pi}(t,X^{c,\pi}(t))).
\end{aligned} \tag{4.86}$$

Since (c,π) is an arbitrarily chosen admissible strategy, taking the conditional expectation given $X(t) = x$ of the inequality gives (4.85).

Consider now the specific strategy (c^*,π^*) fulfilling the infimum in (4.77). By (4.65) and (4.72) it follows that

$$\begin{aligned}
F^{(1)}(t,x) &= y^{c^*,\pi^*}(t,x), \\
F^{(2)}(t,x) &= z^{c^*,\pi^*}(t,x).
\end{aligned}$$

Final step is to show that the Nash equilibrium criteria given by Definition 4.2.1 is fulfilled.

Going through the same calculations as above, and using that for the specific strategy (c^*, π^*) we have, $\forall (t,x) \in (0,T) \times \mathbb{R}$, that

$$F_t = -[(r + \pi^*(\alpha - r))x + \ell - c^*](F_x - f_x) - \frac{1}{2}(\pi^*)^2\sigma^2 x^2 (F_{xx} - U) + J,$$

we get

$$\begin{aligned} F\left(t, X^{c^*,\pi^*}(t)\right) = \int_t^T & \left\{(f_x + f_y Y_x + f_z Z_x - F_x)\, \pi^*(s)\sigma X^{c^*,\pi^*}(s) dW(s)\right\} \\ & + f\left(t, X^{c^*,\pi^*}(t), Y\left(t, X^{c^*,\pi^*}(t)\right), Z\left(t, X^{c^*,\pi^*}(t)\right)\right). \end{aligned} \tag{4.87}$$

Since (c^*, π^*) is an admissible strategy, taking the conditional expectation given $X(t) = x$ on both sides of the inequality we obtain

$$F(t,x) = f\left(t, x, y^{c^*,\pi^*}(t,x), y^{c^*,\pi^*}(t,x)\right). \tag{4.88}$$

Now let $(\widetilde{c}_h, \widetilde{\pi}_h)$ be defined by (4.10). To make it explicit that the expressions in (4.82) and (4.83) depend on h, when using the strategy $(\widetilde{c}_h, \widetilde{\pi}_h)$, we write $\widetilde{U}_h$ and $\widetilde{J}_h$, respectively. By (4.85) and (4.88) we get that

$$\begin{aligned} & \liminf_{h \to 0} \frac{f\left(t, x, y^{c^*,\pi^*}(t,x), z^{c^*,\pi^*}(t,x)\right) - f\left(t, x, y^{\widetilde{c}_h,\widetilde{\pi}_h}(t,x), z^{\widetilde{c}_h,\widetilde{\pi}_h}(t,x)\right)}{h} \\ & = \liminf_{h \to 0} \frac{\int_t^T \left(\widetilde{J}_h(s) - J(s) + \frac{1}{2}\sigma^2 \widetilde{\pi}_h(s)^2 X^{\widetilde{c}_h,\widetilde{\pi}_h}(s)^2 (\widetilde{U}_h(s) - U(s))\right) ds}{h} \\ & = \liminf_{h \to 0} \frac{\int_t^{t+h} \left(\widetilde{J}_h(s) - J(s) + \frac{1}{2}\sigma^2 \pi(s)^2 X^{c,\pi}(s)^2 (\widetilde{U}_h(s) - U(s))\right) ds}{h} \\ & = \widetilde{J}_0(t) - J(t) + \frac{1}{2}\sigma^2 \pi(t)^2 X^{c,\pi}(t)^2 (\widetilde{U}_0(t) - U(t)) \\ & = 0, \end{aligned}$$

where we have used that $(\widetilde{c}_h, \widetilde{\pi}_h)$ coincides with (c^*, π^*) on $[t+h, T]$, and with (c, π) on [t,t+h). We conclude that $F(t,x) = V(t,x)$, and that (c^*, π^*) is the corresponding optimal strategy.

□

4.6.2

First. We want to show that the assumption made in (4.44) given as

$$a(t)b(t) = \frac{g(t)}{2}$$

is fulfilled. This is easily seen to be the case since $a(T)b(T) = \frac{g(T)}{2} = 0$, and by use of (4.52) and (4.53)

$$\begin{aligned}
&\frac{\partial}{\partial t}(a(t)b(t)) = a'(t)b + ab'(t) \\
= \quad & -\left\{(r-\rho) + \frac{(\alpha-r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]\right\} a(t)b(t) - c^*(t)a(t) + \rho a(t)b(t),
\end{aligned}$$

and

$$\frac{g'(t)}{2} = -\left\{(r-\rho) + \frac{(\alpha-r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma(a(t)^2 - f(t))\right]\right\}\frac{g(t)}{2} - c^*(t)a(t) + \rho\frac{g(t)}{2}.$$

□

Second. We want to show that the condition (4.51) for the non-binding case given as

$$\frac{\gamma f(t)}{x + K^{(c^*)}(t)} > 0, \ \forall (t,x) \in [0,T) \times \mathbb{R},$$

required for the investment strategy (4.50) to be optimal, is fulfilled.

In order to verify this condition notice that by use of the candidate for the optimal strategy given by (4.49) and (4.50) we get that

$$\begin{aligned}
& d\left(X^{c^*,\pi^*}(t) + K^{(c^*)}(t)\right) \\
= \quad & \left[rX^{c^*,\pi^*}(t) + \frac{(\alpha-r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right]\left(X^{c^*,\pi^*}(t) + K^{(c^*)}(t)\right) + \ell(t) - c^*(t)\right] dt \\
& + \frac{\alpha - r}{\sigma\gamma f(t)}\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right]\left(X^{c^*,\pi^*}(t) + K^{(c^*)}(t)\right) dW(t) + \left(rK^{(c^*)}(t) - \ell(t) + c^*(t)\right) dt \\
= \quad & \left(r + \frac{(\alpha-r)^2}{\sigma^2\gamma f(t)}\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right]\right)\left(X^{c^*,\pi^*}(t) + K^{(c^*)}(t)\right) dt \\
& + \frac{\alpha - r}{\sigma\gamma f(t)}\left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right]\left(X^{c^*,\pi^*}(t) + K^{(c^*)}(t)\right) dW(t).
\end{aligned}$$

We see that $(X^{c^*,\pi^*}(t) + K^{(c^*)}(t))$ takes the form of a geometric Brownian motion. We get the solution

$$\begin{aligned}
& X^{c^*,\pi^*}(t) + K^{(c^*)}(t) \\
= & \left(x_0 + K^{(c^*)}(0,x_0)\right)\exp\left\{\int_0^t \left(r + \frac{(\alpha-r)^2}{\sigma^2\gamma f(s)}\left[a(s) + \gamma\left(a(s)^2 - f(s)\right)\right]\right.\right. \\
& \left.\left. - \frac{1}{2}\frac{(\alpha-r)^2}{\sigma^2\gamma^2 f(s)^2}\left[a(s) + \gamma\left(a(s)^2 - f(s)\right)\right]^2\right) ds + \int_0^t \frac{\alpha-r}{\sigma\gamma f(s)}\left[a(s) + \gamma\left(a(s)^2 - f(s)\right)\right] dW(s)\right\}.
\end{aligned} \tag{4.89}$$

From (4.94) below it follows that f is a strictly positive function over the time interval $[0,T)$. Finally, since for the non-binding case $x_0 + K^{(c^*)}(0,x_0) > 0$ we conclude that the condition is fulfilled. □

Third. We want to show that the highly non-linear system of partial differential equations given by (4.52) and (4.53) has a unique global solution. In order to show this take conditional expectation in (4.89) for $t=T$ to get that (remember that $K^{(c^*)}(T)=0$)

$$E_{t,x}\left[X^{c^*,\pi^*}(T)\right] = \left(x+K^{(c^*)}(t)\right)e^{\int_t^T[r+(\alpha-r)\widetilde{\pi}^*(s)]ds} \tag{4.90}$$

$$E_{t,x}\left[\left(X^{c^*,\pi^*}(T)\right)^2\right] = \left(x+K^{(c^*)}(t)\right)^2 e^{2\int_t^T\left[r+(\alpha-r)\widetilde{\pi}^*(s)+\frac{1}{2}\sigma^2\widetilde{\pi}^*(s)^2\right]ds}, \tag{4.91}$$

where[14]

$$\widetilde{\pi}^*(t) = \frac{\alpha-r}{\sigma^2\gamma f(t)}\left[a(t)+\gamma\left(a(t)^2-f(t)\right)\right]. \tag{4.92}$$

We now get that

$$\begin{aligned}
& a(t)\left(x+K^{(c^*)}(t)\right)+b(t) \\
= \;& E_{t,x}\left[\int_t^T e^{-\rho(s-t)}c^*(s)ds + e^{-\rho(T-t)}X^{c^*,\pi^*}(T)\right] \\
= \;& \left(x+K^{(c^*)}(t)\right)e^{\int_t^T[(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)]ds} + \int_t^T e^{-\rho(s-t)}c^*(s)ds
\end{aligned}$$

and

$$\begin{aligned}
& f(t)\left(x+K^{(c^*)}(t)\right)^2 + g(t)\left(x+K^{(c^*)}(t)\right)+h(t) \\
= \;& E_{t,x}\left[\left(\int_t^T e^{-\rho(s-t)}c^*(s)ds + e^{-\rho(T-t)}X^{c^*,\pi^*}(T)\right)^2\right] \\
= \;& e^{-2\rho(T-t)}E_{t,x}\left[\left(X^{c^*,\pi^*}(T)\right)^2\right] + 2\left(\int_t^T e^{-\rho(s-t)}c^*(s)ds\right)e^{-\rho(T-t)}E_{t,x}\left[X^{c^*,\pi^*}(T)\right] \\
& +\left(\int_t^T e^{-\rho(s-t)}c^*(s)ds\right)^2 \\
= \;& \left(x+K^{(c^*)}(t)\right)^2 e^{2\int_t^T\left[(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)+\frac{1}{2}\sigma^2\widetilde{\pi}^*(s)^2\right]ds} \\
& +2\left(\int_t^T e^{-\rho(s-t)}c^*(s)ds\right)\left(x+K^{(c^*)}(t)\right)e^{\int_t^T[(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)]ds} \\
& +\left(\int_t^T e^{-\rho(s-t)}c^*(s)ds\right)^2.
\end{aligned}$$

[14]Notice that $\widetilde{\pi}$ defines the optimal proposition of $X^{c^*,\pi^*}+K^{(c^*)}$ to invest in stocks S.

Collecting terms we obtain

$$a(t) = e^{\int_t^T [(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)]ds}, \tag{4.93}$$

$$b(t) = \int_t^T e^{-\rho(s-t)} c^*(s) ds,$$

$$f(t) = e^{2\int_t^T \left[(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)+\frac{1}{2}\sigma^2\widetilde{\pi}^*(s)^2\right]ds}, \tag{4.94}$$

$$g(t) = 2e^{\int_t^T [(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)]ds} \int_t^T e^{-\rho(s-t)} c^*(s) ds,$$

$$h(t) = \left(\int_t^T e^{-\rho(s-t)} c^*(s) ds\right)^2.$$

Now insert (4.93) and (4.94) into (4.92) to get the following integral equation for $\widetilde{\pi}^*$

$$\begin{aligned} \widetilde{\pi}^*(t) &= \frac{\alpha - r}{\sigma^2 \gamma f(t)} \left[a(t) + \gamma\left(a(t)^2 - f(t)\right)\right] \\ &= \frac{\alpha - r}{\sigma^2\gamma}\left\{e^{-\int_t^T \left[(r-\rho)+(\alpha-r)\widetilde{\pi}^*(s)+\sigma^2\widetilde{\pi}^*(s)^2\right]ds} + \gamma e^{-\int_t^T \sigma^2\widetilde{\pi}^*(s)^2 ds} - \gamma\right\}. \end{aligned} \tag{4.95}$$

The key question now is whether the integral equation (4.95) has a unique global solution. Fortunately this is the case. The technical work is done in Björk et al. (2012). They show that the algorithm given by

$$\begin{aligned} \widetilde{\pi}_0(t) &:= 1, \\ \widetilde{\pi}_{n+1}(t) &= \frac{\alpha - r}{\sigma^2\gamma}\left[e^{-\int_t^T [(r-\rho)+(\alpha-r)\widetilde{\pi}_n(s)+\sigma^2\widetilde{\pi}_n^2(s)]ds} + \gamma e^{-\int_t^T \sigma^2\widetilde{\pi}_n^2(s)ds} - \gamma\right], \end{aligned} \tag{4.96}$$

converges in $\mathcal{C}[0,T]$, and that the full sequence $\{\widetilde{\pi}_n\}$ converges to the solution $\widetilde{\pi}^*$. □

5. Why you should care about investment costs: A risk-adjusted utility approach

Abstract: Under the assumption of zero correlation between cost ratios and expected investment returns we analyze the impact of proportional investment costs. We consider a constant relative risk aversion investor optimizing expected utility from terminal wealth and identify, in addition to the direct effect due to the additional costs incurred, an indirect effect due to a less risky stock position induced by investment costs. By use of an indifferent compensation measure, defined as the minimum relative increase in the initial wealth the investor demands in compensation to accept incurring investment costs of a certain size, we quantify the impact of investment costs. We obtain, for realistic parameters for risk aversion and the financial market, the surprising results that the indirect effect is between half (myopic investor) and the same (long term investor) size as the direct effect, and that the investment decision seems to be of very little importance compared to the size of the investment costs.

Keywords: Investment costs, risk aversion, stochastic control, indifferent compensation measure, certainty equivalents.

5.1 Introduction

Most investors seem primarily to focus on the ability of excellent stock picking, when deciding which fund should manage their savings. At the same time costs charged by funds seem to differ by a great deal. The average U.S. equity mutual fund charges around 1.3–1.5 percent, but cost ratios range from as low as 0.2 (index funds) to as high as 2 percent. In general, costs can vary substantially across comparable funds, and larger funds and fund complexes charge lower costs (see e.g. Khorana et al. (2008)). Clearly, the argument for charging high costs is excellent stock picking. The expensive fund managers are likely to claim that the additional return they are expected to generate (compared to any cheaper fund manager) more than compensates for the extra costs. However, the vast majority of the large literature analyzing whether excellent stock picking skills exist rejects this hypothesis. This is e.g. demonstrated by Gil-Bazo and Ruiz-Verdú (2009) who consider a data set including all open-end U.S. mutual funds that are active in the 1961 to 2005 period. They consider a series of robustness checks consisting of checking for the impact of funds with extreme cost ratios and extreme risk-adjusted performance; the impact of small funds; exclusive focussing on funds for which annual operating costs account for 100 percent of all costs or focussing only on funds with loads; splitting time into subperiods; splitting mutual funds into categories. In all cases the conclusion stays the same: The hypothesis of a unit slope relation between risk-adjusted before-fee performance and cost ratios falls at any conventional significance level. In fact, ironically, the estimated slope becomes negative. See also Carhart (1997) who, using the same data set, concludes that higher costs depress investment performance while increasing fund companies' profitability. Also Fama and French (2010) report that only very few funds produce benchmark-adjusted expected returns sufficient to cover their costs. One of the reasons that some funds are more expensive is due to the more actively managed investments. Huang et al. (2013) report, using a sample of 2979 U.S. equity funds over the period between 1980 and 2009, that the top and bottom decile of funds on average change their annualized volatility by more than six percentage points. They also find, by use of a holding-based measure of risk shifting, that funds which alter risk perform worse than funds that keep stable risk levels over time, suggesting that risk shifting either is an indication of inferior ability or is motivated by agency issues. Summing up, it seems hard to prove that good performance is anything but a random phenomena.

Consequently, the impact of investment costs is naturally analyzed under the assumption of a zero correlation between the cost ratio and the expected investment return. The literature, e.g. the references above, seems only to focus on the loss in rate of return. However, the loss in rate of return simply induced by paying higher investment costs might not describe the actual loss suffered by the investor. A more sophisticated approach would be to take into account the risk aversion of the investor when evaluating the impact of investment costs, thereby also introducing a change in the investment strategy induced by investment costs. Introducing proportional investment costs and by use of utility functions, this is the approach taken in our paper. Two related papers, also taking the investor's risk aversion into account while considering proportional costs are Guillén et al. (2013) and Palczewski et al. (2013) (the latter analyzes the impact of transaction costs).

Guillén et al. (2013) consider a Value at Risk investor (VaR-investor) who invests in a Black-Scholes market concerned about a given α-percent quantile of the terminal wealth distribution. By introducing investment costs the investor is forced to invest less in the stock market in order to maintain the same α-percent quantile. Consequently, the loss in the geometric rate of return splits into two effects: (a) A direct effect due to the additional expense incurred and (b) an indirect effect due to a less risky stock position. Some of the capital the investor, prior to introducing investment costs, was willing to risk losing is now used to pay investment costs. Guillén et al. (2013) refer to the indirect effect as *loss of investment opportunities*. The main drawback of the VaR-approach is that no monetary quantification of how much the investor actually suffers from investment costs seems to be possible. Focussing at the geometric rate of return seems a bit ad hoc since, in the first place, when deciding upon the investment strategy, the VaR-investor had no particular preferences for a high median. Using the geometric rate of return to measure the impact of investment costs also restricts the parameter space since for very risk seeking investors, introducing investment costs actually *increases* the geometric rate of return.

In contrast, Palczewski et al. (2013) use utility functions, but focus instead on the impact of transaction costs. They optimize expected utility from investing in a market consisting of a risk free asset and a risky asset modeled by a diffusion model with state-dependent drift. The effect of costs can again be divided into a direct and an indirect effect. This time the indirect effect is due to less trading in the asset portfolio. By calculating the indifference price, defined as the amount of money the investor is willing to pay up front to avoid incurring transaction costs, they find that in general the loss in utility due to proportional transaction costs is about twice as large as the direct expenses incurred.

Similar findings are offered by our paper for the case of proportional investment costs. We focus on a constant relative risk aversion (CRRA) utility optimizer who hands over his savings to a fund investing in a frictionless Black-Scholes market while being charged proportional investment costs. As in Guillén et al. (2013) and Palczewski et al. (2013) we obtain a direct and an indirect effect due to a change in investment costs. In contrast to the VaR-approach of Guillén et al. (2013) the change in investment strategy and, consequently, the change in geometric rate of return induced by a change in investment costs becomes independent of the investment horizon. The change in geometric rate of return is the same for both short and long term investors. In order to quantify the financial impact of investment costs we calculate the indifferent compensation ratio (ICR), defined as the minimum relative increase in the initial wealth the investor demands in compensation to accept incurring investment costs of a certain size. For a CRRA utility optimizing investor the ICR is proved to be equal to the relative change in certainty equivalents. By comparing the ICR value to the fraction of initial wealth expected to be spent on investment costs, we find, similar to Palczewski et al. (2013), that the magnitude of the indirect effect equals the direct effect when considering a long-term investor (40 years horizon, i.e.

investing for retirement). For a short term (myopic) investor we find that the magnitude of the indirect effect is half the size of the direct effect. That is, the amount of money needed up front to be compensated for investment costs can be twice as big as the amount of money expected to be used on paying investment costs. We refer to the effect as the risk-adjusted impact of investment costs or, in the words of Jens Perch Nielsen, the more catchy phrase; *The double blow of investment costs*. Finally, we undertake a study of whether the investment strategy or the size of investment costs is of most importance. Specifically, we study an investor facing high investment costs and an optimal investment strategy (w.r.t. his risk aversion profile) and ask which suboptimal investment strategies the investor is willing to accept if he at the same time is offered lower investment costs. The conclusion is independent of the time horizon and very clear: The asset allocation is of very little importance compared to the size of investment costs.

The outline of the paper is as follows: In Section 5.2 we introduce the investor and the financial market to be considered, the wealth dynamics, and quantiles for the terminal wealth distribution together with the geometric rate of return. In Section 5.3 we analyze for a utility optimizing investor the change in investment strategy induced by a change in investment costs, and we compare the results with the VaR-approach by Guillén et al. (2013). Finally, in Section 5.4, we use the indifferent compensation measure to quantify the impact of investment costs, and to evaluate whether the investment strategy or the size of investment costs is of most importance.

5.2 The financial model

Consider an investor with time horizon $T > 0$ who has the possibility to invest in a Black-Scholes market given by a risky stock, S, and a risk-free bank account, B, with dynamics given by

$$dB(t) = rB(t)dt,\ B(0) = 1, \tag{5.1}$$

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t),\ S(0) = s_0 > 0. \tag{5.2}$$

We assume the risk-free rate, r, the expected return of the stock, μ, and the volatility of the stock, σ, to be constant with $\mu > r$. The risky part of the stock, W, is a standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ equipped with the filtration $\mathbb{F}^W = (\mathcal{F}^W(t))_{t\in[0,T]}$ given by the P-augmentation of the filtration $\sigma\{W(s);\, 0 \le s \le t\}, \forall t \in [0, T]$. In principle, a financial market model should contain more than one stock. However, it is well known from the Mutual Fund Theorem by Merton (1971) that for a HARA (Hyperbolic Absolute Risk Aversion) utility investor, as we shall consider, the optimal asset allocation will always be between an optimal portfolio of the stocks weighted by their Sharp ratios, and the risk-free bank account (at least for diffusion processes with deterministic coefficients).

Let X denote the wealth process of the investor and π the proportion of the wealth invested in risky stocks. Consequently, the proportion $1 - \pi$ of the wealth is invested in the risk-free bank account. Assume further that the investor hands over the asset management to a second party (e.g. a mutual fund or a pension company), and therefore is subject to investment costs. More precisely, assume that a constant fraction, ν, of the amount of money invested in the stock market is deducted from the wealth process. The dynamics of the wealth process becomes

$$\begin{aligned} dX(t) &= \left((1-\pi(t))\frac{dB(t)}{B(t)} + \pi(t)\frac{dS(t)}{S(t)} - \pi(t)\nu\right)X(t) \\ &= (r + \pi(t)(\mu - \nu - r))X(t)dt + \sigma\pi(t)X(t)dW(t), \qquad (5.3) \\ X(0) &= x_0. \qquad (5.4) \end{aligned}$$

Note that costs from investing in the risk-free asset are not explicitly indicated, but are indirectly present through a possibly lower risk-free rate of return. However, compared to costs emerging from investing in stocks, costs from investing in the risk-free bank asset are likely to be negligible.

5.2.1 Wealth percentiles for a constant investment strategy

One extremely popular investment strategy, which also turns out to be optimal for the popular CRRA (Constant Relative Risk Aversion) utility optimizing investor introduced by the pioneering work of Merton (1969), is to hold, at all times, a constant fraction of wealth in stocks. In this paper we consider only such constant strategies. Note that the word *constant* is rather misleading since the amount of money invested in stocks is dynamically reallocated to keep the fraction of wealth invested in stocks constant. Letting π denote a constant investment strategy, the wealth dynamics (5.3) takes the form of a geometric Brownian motion with solution

$$X(t) = x_0 \exp\left(\left(r + \pi(\mu - \nu - r) - \frac{1}{2}\pi^2\sigma^2\right) t + \pi\sigma W(t)\right). \tag{5.5}$$

Since $W(t) \stackrel{\mathcal{D}}{=} \sqrt{t}U$, where U is standard normal distributed, we get from (5.5) that the α-quantile, q_α, of the terminal wealth distribution is given by

$$q_\alpha(\nu, \pi) = x_0 \exp\left(\left(r + \pi(\mu - \nu - r) - \frac{1}{2}\pi^2\sigma^2\right) T + \pi\sigma\sqrt{T} d_\alpha\right), \tag{5.6}$$

where d_α is the α-quantile of the standard normal distribution. In particular, the median is given by

$$q_{50}(\nu, \pi) = x_0 \exp\left(\left(r + \pi(\mu - \nu - r) - \frac{1}{2}\pi^2\sigma^2\right) T\right). \tag{5.7}$$

From this we get that the median rate of return (geometric rate of return), from now on referred to as simply *the rate of return*, ρ, is given by

$$\rho(\nu, \pi) = r + \pi(\mu - \nu - r) - \frac{1}{2}\pi^2\sigma^2. \tag{5.8}$$

Note that the rate of return, in contrast to the expected arithmetic rate of return, possesses a maximum with respect to the risky stock allocation.

5.3 Investment costs' impact on the investment strategy

Consider a *naive investor* who for some reasons invests, without any objective in mind, a constant fraction, π, of his wealth in risky stocks. For the naive investor the gain/loss in terms of rate of return caused by a change in the investment costs from ν_1 to ν_2 is given by

$$\rho(\nu_2, \pi) - \rho(\nu_1, \pi) = \pi(\nu_1 - \nu_2), \tag{5.9}$$

where ρ is defined by (5.8). This simple calculation seems to be how investment costs are normally quantified.

However, a *sophisticated investor* should take his risk aversion into account and thereby adjust his investment strategy accordingly when investment costs change. Our main focus will be how a utility investor adjusts his investment strategy in response to changes in the investment costs and how to quantify the associated gain or loss. However, for comparison, we first consider a VaR-investor as done by Guillén et al. (2013). The main point is that the sophisticated investor experiences a greater change in the rate of return induced by a change in investment costs compared to the naive investor.

5.3.1 Base case parameter values

Throughout this paper we will, as a base case example for illustrating results, use the market parameters $\mu = 7\%$, $\sigma = 20\%$ and $r = 3\%$. When illustrating the impact of investment costs we consider the difference between paying investment costs at rate $\nu_1 = 1.4\%$ and $\nu_2 = 0.6\%$, respectively. Most papers focus on the impact of introducing investment costs. We focus on an individual who prefers to hand over the asset management to a fund. In that situation it seems unattainable to obtain zero investment costs. Therefore, we consider instead the "price" of investing through an expensive fund (ν_1) compared to a cheap fund (ν_2). The investment costs considered are consistent with those observed in the market (see references in Section 5.1).

5.3.2 The Value at Risk approach

This subsection briefly presents and explains the impact of investment costs for a VaR-investor, as presented in Guillén et al. (2013). A VaR(α)-investor is an investor being concerned with the value of the α-quantile of the terminal wealth distribution, i.e. with the level for which the realized terminal wealth risk to fall below with a probability of α-percent. Denote the lower level of wealth by $\widehat{X}(T)$. To be more precise, the VaR(α)-investor facing investment costs ν_1 chooses his investment strategy, $\widehat{\pi}_1$, such that

$$q_\alpha(\nu_1, \widehat{\pi}_1) = \widehat{X}(T), \tag{5.10}$$

where q_α is given by (5.6). The point is that if investment costs change from ν_1 to ν_2 this relation holds no more, i.e.

$$q_\alpha(\nu_2, \widehat{\pi}_1) \neq \widehat{X}(T). \tag{5.11}$$

Obviously, the VaR-(α)-investor should adjust his investment strategy such that the VaR-criteria (5.11) is still fulfilled. We therefore look for a change in the investment strategy, Δ, fulfilling the relation

$$q_\alpha(\nu_2, \widehat{\pi}_1 + \Delta) = q_\alpha(\nu_1, \widehat{\pi}_1), \tag{5.12}$$

i.e. being faced with investment costs ν_2 the investor should instead invest $\widehat{\pi}_2 = \widehat{\pi}_1 + \Delta$ of his wealth in risky stocks. We get the solution

$$\Delta = \frac{-b \pm \sqrt{D}}{2a}, \tag{5.13}$$

where $D = b^2 - 4ac$ and

$$\begin{aligned} a &= \frac{\sigma}{2}, & (5.14)\\ b &= -(\mu - \nu_2 - r) + \widehat{\pi}_1\sigma^2 - \frac{\sigma d_\alpha}{\sqrt{T}}, & (5.15)\\ c &= \widehat{\pi}_1(\nu_2 - \nu_1). & (5.16) \end{aligned}$$

Of the two solutions we focus on the one allowing for more stocks in the case of lower costs ($\nu_2 < \nu_1$) and fewer stocks in the case of higher costs ($\nu_2 > \nu_1$). Guillén et al. (2013) derive restrictions on the parameter space ensuring that the median increases and the α-quantile decreases when exposure in the risky asset increases. In contrast to (5.9) the change in rate of return becomes

$$\rho(\nu_2, \widehat{\pi}_2) - \rho(\nu_1, \widehat{\pi}_1) = \widehat{\pi}_1(\nu_1 - \nu_2) + \Delta(\mu - \nu_2 - r) - \frac{1}{2}\Delta^2\sigma^2 - \widehat{\pi}_1\Delta\sigma^2. \tag{5.17}$$

Note that since Δ and $\widehat{\pi}_1$ depend on the investment horizon so does the change in rate of return. Figure 5.1 illustrates for the base case example (see Subsection 5.3.1) the VaR-calibration concept for an investor investing 1 unit initially. First, when investment costs are $\nu_1 = 1.4\%$ the VaR(10%)-investor chooses to invest $\widehat{\pi}_1 = 60\%$ of his wealth in stocks corresponding to a 10-percent quantile for the terminal wealth distribution of $q_{10\%}(\nu_1, 60\%) = 1.75$ and a median of $q_{50\%}(\nu_1, 60\%) = 4.65$. Not changing the investment strategy, but now being charged

$\nu_2 = 0.6\%$ in investment costs, the 10-percent quantile becomes $q_{10\%}(\nu_2, 60\%) = 2.12$ and the median increases, due to paying less costs, to $q_{50\%}(\nu_1, 60\%) = 5.63$. However, due to gained investment opportunities the investor are able to increase the fraction of wealth invested in stocks to $\widehat{\pi}_2 = 74.4\%$ thereby obtaining the original 10-percent target quantile $q_{10\%}(\nu_1, 60\%) = 1.75$ and at the same time an even bigger median of size $q_{50\%}(\nu_1, 60\%) = 5.87$. In other words, Figure 5.1 illustrates the direct and the indirect effect of investment costs.

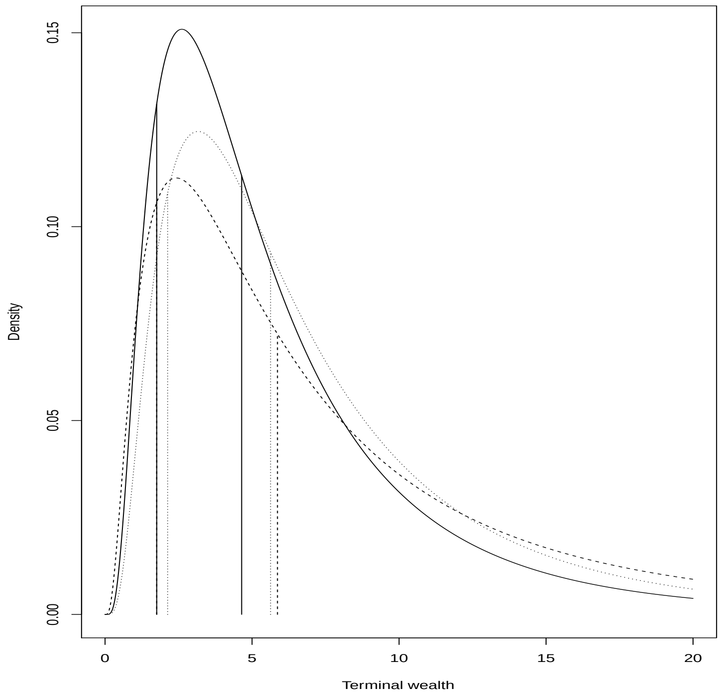

Figure 5.1: Illustration of the VaR-calibration: Terminal wealth distribution for an investor with time horizon $T = 40$ years investing $\widehat{\pi}_1 = 60\%$ (solid curve), $\pi_2 = 60\%$ (dotted curve) or $\widehat{\pi}_2 = 74.4\%$ (dashed curve) of his wealth in stocks while being charged $\nu_1 = 1.4\%$, $\nu_2 = 0.6\%$ or $\nu_2 = 0.6\%$ in investment costs, respectively. The 10-percent quantiles and the medians are indicated by vertical lines.

Guillén et al. (2013) focus on the loss in rate of return incurred by the VaR-investor from introducing investment costs. For some parameters, going from no costs to costs actually charged by funds around the world, they conclude that the loss in rate of return calculated by (5.17) is double the size of the naive loss calculated by (5.9), i.e. the indirect effect is as big as the direct effect. However, going from low investment costs around 0.6 percent to higher costs at around 1.4 percent, both within the interval of normal costs charged by real life funds, we get a much more modest additional loss in the rate of return; the loss is about $1/5$ higher than the loss from the naive calculation. To be concrete, for the case illustrated in Figure 5.1, the direct loss in the rate of return is 0.48 percent, and the indirect loss due to lost investment opportunities is 0.10 percent.

One question the VaR-approach is not capable of answering is how to evaluate the impact of investment cost taking the entire distribution of terminal wealth into account. E.g. how much is it "worth" being charged ν_2 percent instead of ν_1 percent? This is the reason we take on another approach and consider a utility optimizing investor instead.

5.3.3 The utility approach

Consider an investor measuring his attitude towards risk by a constant relative risk aversion (CRRA) function defined by

$$u(x) = \begin{cases} \frac{x^\gamma}{\gamma}, & \text{if } x > 0, \\ \lim_{x \searrow 0} \frac{x^\gamma}{\gamma}, & \text{if } x = 0, \\ -\infty, & \text{if } x < 0, \end{cases} \tag{5.18}$$

for some $\gamma \in (-\infty, 1)\backslash\{0\}$ ($\gamma = 0$ corresponds to the logarithmic case). A commonly used object for the investor is to try maximizing the expected utility from terminal wealth. Formally, the investor looks for the investment strategy π^* fulfilling

$$\sup_\pi E\left[u(X^\pi(T))\right] = E\left[u\left(X^{\pi^*}(T)\right)\right], \tag{5.19}$$

where we have indicated the π-dependence of the terminal wealth in the notation. The problem given by (5.19) is one of the most well-known optimization problems from the literature of financial stochastic control and is often referred to as *Merton's problem* after the pioneering work of Merton (1969). The solution to the problem can be obtained by use of dynamic programming, which turns the stochastic optimization problem into a deterministic optimization problem and a set of partial differential equations (Hamilton-Jacobi-Bellman equations), or by use of the modern, more direct, method called the "Martingale method" developed by Cvitanic et al. (1987) and Cox and Huang (1989), which builds upon modern mathematical finance theory. The solution turns out to depend on the market price of risk, the investor's risk aversion, and the investment costs, and is given by

$$\pi^*(\nu) = \frac{1}{1-\gamma}\frac{\mu - \nu - r}{\sigma^2}. \tag{5.20}$$

We see that the optimal investment strategy is to hold a constant fraction of wealth in risky assets. This is a feature of CRRA utility which, from a mathematical point of view, is rather appealing (see Subsection 5.2.1). The strategy itself dictates directly the change in the risky position induced by a change in investment costs. We get

$$\Delta = \pi^*(\nu_2) - \pi^*(\nu_1) = \frac{1}{1-\gamma}\frac{\nu_1 - \nu_2}{\sigma^2}. \tag{5.21}$$

As for the VaR-approach we see that, naturally, if the investment costs decrease (increase) more stocks should be bought (sold). However, opposed to the VaR-approach the change in the risky position is always unique (no short-selling solution). Even more appealing we get in contrast to the VaR-approach (see (5.17)) that the change in rate of return becomes independent of the investment horizon. We get, with $\pi_i^* = \pi^*(\nu_i)$, that

$$\rho(\nu_2, \pi_2^*) - \rho(\nu_1, \pi_1^*) = \pi_1^*(\nu_1 - \nu_2) + \Delta(\mu - \nu_2 - r) - \frac{1}{2}\Delta^2\sigma^2 - \pi_1^*\Delta\sigma^2. \tag{5.22}$$

Clearly, the rather large change in rate of return induced by lost/gained investment opportunities, due to a change in investment costs, was for the VaR-investor partly because a very long investment horizon was considered. Since many private investors mainly invest through being a member of a pension scheme, i.e. long time horizon, the conclusion still seems to be highly relevant. However, the fact that for the utility investor the change in rate of return induced by a change in investment costs does not depend on the investment horizon counters criticism claiming that the effect is negligible for short horizons.

Figure 5.2 illustrates for the base case example (see Subsection 5.3.1) the utility-calibration concept for an investor investing 1 unit initially. First, when investment costs are $\nu_1 = 1.4\%$ the investor optimally invests $\pi_1^* = 60\%$ of his wealth in stocks. However, due to gained investment opportunities the investor increases this fraction to $\pi_2^* = 78.5\%$ when investment costs decline to $\nu_2 = 0.6\%$. In contrast to the VaR-approach we see in Figure 5.2 that the 10-percent quantiles of the terminal wealth distributions are not equal for the two cases. The utility investor does not focus solely on the 10-percent quantile, but evaluates the entire distribution of terminal wealth when he decides how much of the saved investment costs to invest in stocks. In this respect he acts more sophisticated than the VaR-investor.

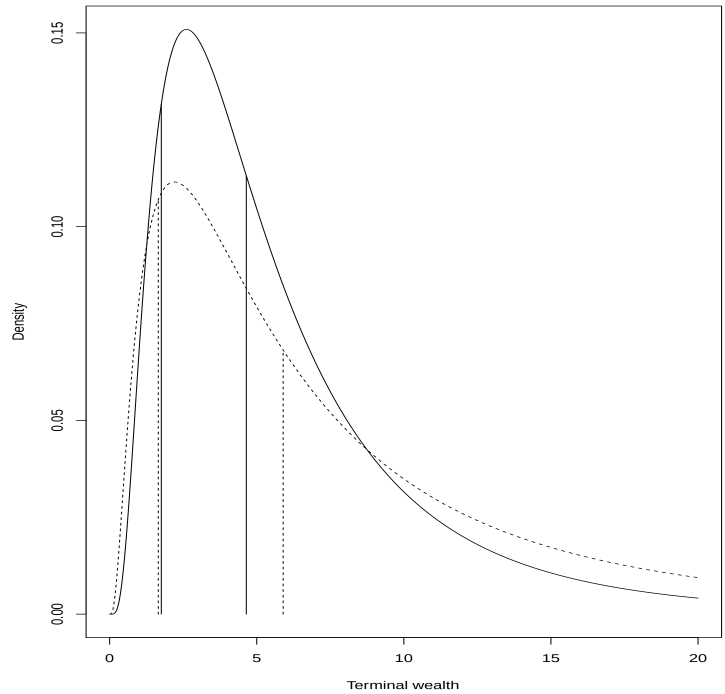

Figure 5.2: Illustration of the utility-calibration: Terminal wealth distribution for an investor with time horizon $T = 40$ years investing $\pi_1^* = 60\%$ (solid curve) or $\pi_2^* = 78.5\%$ (dashed curve) of his wealth in stocks while being charged $\nu_1 = 1.4\%$ or $\nu_2 = 0.6\%$ in investment costs, respectively. The 10-percent quantiles and the medians are indicated by vertical lines.

5.3.4 Comparing the VaR and utility approach

In Subsections 5.3.2 and 5.3.3 we considered a VaR-investor and a utility optimizing investor, respectively. The VaR-investor picked his investment strategy in order to obtain a target 10-percent quantile for the terminal wealth distribution equal to 1.75. The utility investor picked his investment strategy in order to maximize expected utility from terminal wealth while using the power utility function given by (5.18) with risk aversion given by $\gamma = -0.0833$. In order to be able to compare the two investors' reaction to changes in investment costs we have constructed the two examples such that when investment costs are $\nu_1 = 1.4\%$ both investors prefer to invest $\pi = 60\%$ of the wealth in stocks.

Changing investment costs from $\nu_1 = 1.4\%$ to $\nu_2 = 0.6\%$ we saw in Subsections 5.3.2 and 5.3.3 that the change in the investment strategy was $\Delta = 14.4\%$ for the VaR-investor and $\Delta = 18.5\%$ for the utility investor. From this example we conclude that the utility investor reacts more strongly compared to the VaR-investor when investment costs change. However, as illustrated by Figure 5.3, this is not a general rule of thumb. In fact, by changing the risk profiles such that both investors prefer a $\pi = 30\%$ position of wealth in stocks when investment costs are $\nu_1 = 1.4\%$, lowering the investment costs now makes the VaR-investor change his investment strategy more than the utility optimizing investor.

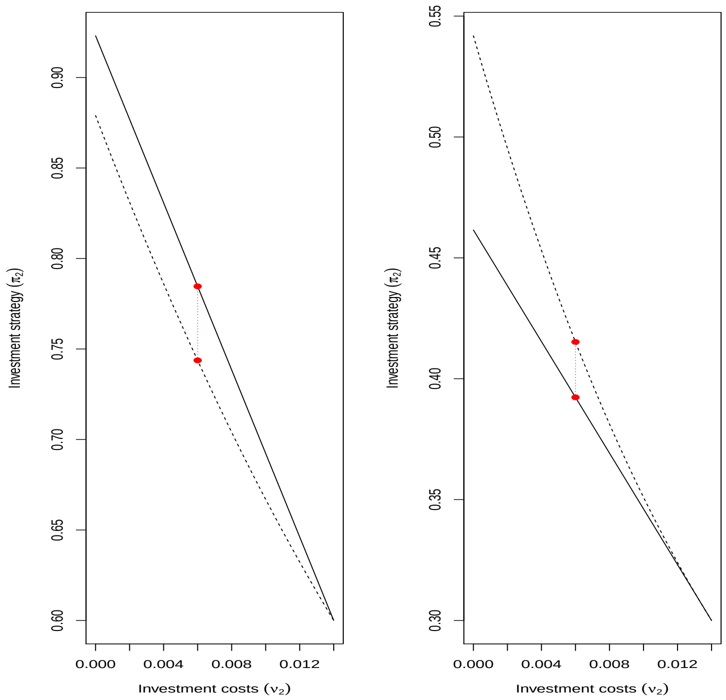

Figure 5.3: Illustrations of how the utility optimizing investor (solid curves) and the VaR-investor (dashed curves) change investment strategy when investment costs decrease from $\nu_1 = 1.4\%$ to $\nu_2 \in [0, \nu_1]$. The left plot illustrates the case where the investors prefers a $\pi = 60\%$ allocation in stocks, and the right plot the case of a preferable $\pi = 30\%$ allocation in stocks, when investment costs are $\nu_1 = 1.4\%$. The dots indicate the base case example from earlier subsections where investment costs decrease to $\nu_2 = 0.06\%$. The investment horizon is $T = 40$.

5.4 Quantification of the impact of investment costs

In this section we use the utility approach described in Subsection 5.3.3 to quantify the impact of investment costs. Inserting the optimal investment strategy (5.20) into (5.5) we are able to calculate the optimal expected utility given by (5.19). We get

$$
\begin{aligned}
&\sup_\pi E\left[u\left(X^{(\nu,\pi)}(T)\right)\right] \\
&= E\left[u\left(X^{(\nu,\pi^*)}(T)\right)\right] \\
&= \frac{1}{\gamma} x_0^\gamma \exp\left(\left(r + \pi^*(\mu - r - \nu) - \frac{1}{2}(\pi^*)^2\sigma^2\right) T\gamma + \frac{1}{2}(\pi^*)^2\sigma^2\gamma^2 T\right) \\
&= \frac{1}{\gamma} x_0^\gamma \exp\left\{\left(r + \frac{1}{2}\frac{\gamma}{1-\gamma}\left(\frac{\mu - r - \nu}{\sigma}\right)^2\right) T\right\}, \qquad (5.23)
\end{aligned}
$$

where x_0 denotes the initial wealth. Now, define the Indifferent Compensation Ratio as the minimum relative increase in the initial wealth the investor demands in compensation to accept incurring higher investment costs. Formally, the indifferent compensation ratio for two given levels of investment costs, $ICR_{(\nu_1,\nu_2)}$, is for $v_2 \leq v_1$ given by the relation

$$\begin{aligned} &\sup_\pi E\left[u\left(X^{(\nu_2,\pi)}(T)\right)\Big|X^{(\nu_2,\pi)}(0)=x_0\right] \\ &= \sup_\pi E\left[u\left(X^{(\nu_1,\pi)}(T)\right)\Big|X^{(\nu_1,\pi)}(0)=x_0\left(1+ICR_{(\nu_1,\nu_2)}\right)\right]. \end{aligned} \tag{5.24}$$

The certainty equivalent is the smallest amount of money the investor is willing to receive with certainty at the horizon in exchange for the possibility to invest in the stock market. In formula, the certainty equivalent for a given level of investment costs, $CEQ_{(\nu)}$, is defined by the relation

$$u\left(CEQ_{(\nu)}\right) = \sup_\pi E\left[u\left(X^{(\nu,\pi)}(T)\right)\right]. \tag{5.25}$$

Since by (5.5) the wealth process is linear in initial wealth we obtain

$$u\left(CEQ_{(\nu_2)}\right) = u\left(\left(1+ICR_{(\nu_1,\nu_2)}\right)CEQ_{(\nu_1)}\right). \tag{5.26}$$

From this we conclude that for power utility the two measures are equivalent in the sense that

$$\frac{CEQ_{(\nu_2)}-CEQ_{(\nu_1)}}{CEQ_{(\nu_1)}} = \frac{\left(1+ICR_{(\nu_1,\nu_2)}\right)CEQ_{(\nu_1)}-CEQ_{(\nu_1)}}{CEQ_{(\nu_1)}} = ICR_{(\nu_1,\nu_2)}, \tag{5.27}$$

i.e. the relative change in certainty equivalents equals the indifferent compensation ratio. This is another appealing feature of power utility. Note that the indifferent compensation ratio given by (5.24) can easily be calculated by (5.27) and (5.23).

We are now in a position to quantify the impact of investment costs using the indifferent compensation measure given by (5.24). Consider the base case example of Subsection 5.3.1 with one unit initially invested and investment costs $\nu_1 = 1.4\%$ and $\nu_2 = 0.6\%$, respectively. We get

$$ICR_{(\nu_1,\nu_2)} = \frac{CEQ_{(\nu_2)}-CEQ_{(\nu_1)}}{CEQ_{(\nu_1)}} = \frac{5.66-4.54}{4.54} = 0.248, \tag{5.28}$$

i.e. the investor profits a gain equal to 24.8 percent of the initial wealth when investment costs decrease from ν_1 to ν_2. In Subsection 5.3.2 we considered the change in rate of return induced by a change in investment costs and compared the naive change given by (5.9) with the change the VaR-investor experienced given by (5.17). For the utility optimizing investor it seems natural to compare the indifferent compensation ratio with the present value of expected accumulated investment costs over the horizon. In order to do so note that

$$\begin{aligned} d(e^{-rt}X(t)) &= -re^{-rt}X(t)dt + e^{-rt}dX(t) \\ &= e^{-rt}X(t)\pi(\mu-r-\nu)dt + e^{-rt}X(t)\pi\sigma dW(t), \end{aligned} \tag{5.29}$$

which implies

$$e^{-rT}X(T) = x_0 + \int_0^T e^{-rt}X(t)\pi(\mu-r-\nu)dt + \int_0^T e^{-rt}X(t)\pi\sigma dW(t). \tag{5.30}$$

Taking expectation, noting that the wealth process is strictly positive, and using (5.5), we get the following expression for the expected accumulated investment costs over the horizon.

$$\begin{aligned} adm(\nu,\pi) \equiv E\left[\int_0^T e^{-rt}X(t)\pi\nu dt\right] &= E[x_0 - e^{-rT}X(T) + \int_0^T e^{-rt}X(t)\pi(\mu-r)dt] \\ &= x_0\left(\frac{\nu}{\mu-r-\nu}\right)\left(e^{\pi(\mu-\nu-r)T}-1\right). \end{aligned} \tag{5.31}$$

From this we define the Relative present value of the Expected Change in accumulated investment costs induced by lowering investment costs from ν_1 to ν_2, $REC_{(\nu_1,\nu_2)}$, by

$$REC_{(\nu_1,\nu_2)} \equiv \frac{adm(\nu_1,\pi_1^*) - adm(\nu_2,\pi_2^*)}{x_0}. \tag{5.32}$$

Note that, as is the case for the indifferent compensation ratio, this value is independent of the size of the initial wealth. We interpret (5.32) as the direct effect and the difference between (5.27) and (5.32) as the indirect effect of investment costs. The left plot in Figure 5.4 compares the indifferent compensation ratio (5.27) with (5.32) for different investment costs. In other words, we compare the size of the compensation sum with the expected accumulated investment costs over the horizon, relative to the size of the initial wealth. For the base case example $(\nu_1^*,\nu_2^*) = (1.4\%, 0.6\%)$ with corresponding optimal investment strategies $(\pi_1,\pi_2) = (60\%, 78.4\%)$ we get

$$REC_{(\nu_1,\nu_2)} = 0.467 - 0.337 = 0.130, \tag{5.33}$$

i.e. 13 percent of the initial wealth is expected to be used to pay for the extra investment costs. In contrast, we got in (5.28) that the investor demands an increase of 24.8% of initial wealth in compensation to accept incurring the higher investment costs $\nu_1 = 1.4\%$ instead of the lower $\nu_2 = 0.6\%$. The great gap between these two values (the indirect effect) is due to the risk aversion of the utility investor. The "fear" that the actual amount of money spent on investment costs exceeds the expectation is overwhelming compared to the "benefit" that it is below. This is exactly what most analyses seem to forget. The fact that for reasonable values of investment costs the effect of higher costs seems to double is quiet surprising. As mentioned, one could refer to the phenomenon as *The double blow of investment costs*. Clearly, one wonders how much this double effect is due to the long horizon (T=40) considered by the base case example. The right plot in Figure 5.4 illustrates that the effect is substantial even for a short term investor. Still, for a myopic investor the indirect effect is half the size of the direct effect.

Finally we analyze the relative importance of the investment strategy and investment costs. Specifically, we consider a power utility optimizing investor who hands over his savings to a fund charging him high investment costs at a rate ν_1. In return, the fund offers a tailored investment strategy π_1^* fitted to meet the risk preferences of the client, i.e. the object given by (5.19) is optimized. The investor now becomes aware that another fund offers to manage his savings while only charging him investment costs at a lower rate ν_2. However, in order to offer this cheap product, the fund is organized as an investment collective meaning that all members follow the same investment strategy π_2. Obviously, if the common investment strategy π_2 exercised by the cheap fund happens to equal the tailored optimal investment strategy π_2^* it's a no-brainer — the investor should move to the cheap fund. However, how much can the collective investment strategy π_2 offered by the cheap fund differ from the optimal tailored investment strategy π_2^* while still being preferable for the investor to move to the cheap fund? In formula, we look for the investment strategies π_2 satisfying the relation

$$u\left(CEQ_{(\nu_1)}\right) = E\left[u\left(X^{(\nu_2,\pi_2)}(T)\right)\right]. \tag{5.34}$$

By use of (5.23) we get the solution

$$\pi_2 = \frac{-b \pm \sqrt{D}}{2a}, \tag{5.35}$$

where $D = b^2 - 4ac$ and

$$a = \frac{\sigma}{2}\gamma(1-\gamma), \tag{5.36}$$

$$b = -(\mu - \nu_2 - r), \tag{5.37}$$

$$c = \frac{1}{2}\frac{\gamma}{1-\gamma}\left(\frac{\mu - v_1 - r}{\sigma^2}\right)^2. \tag{5.38}$$

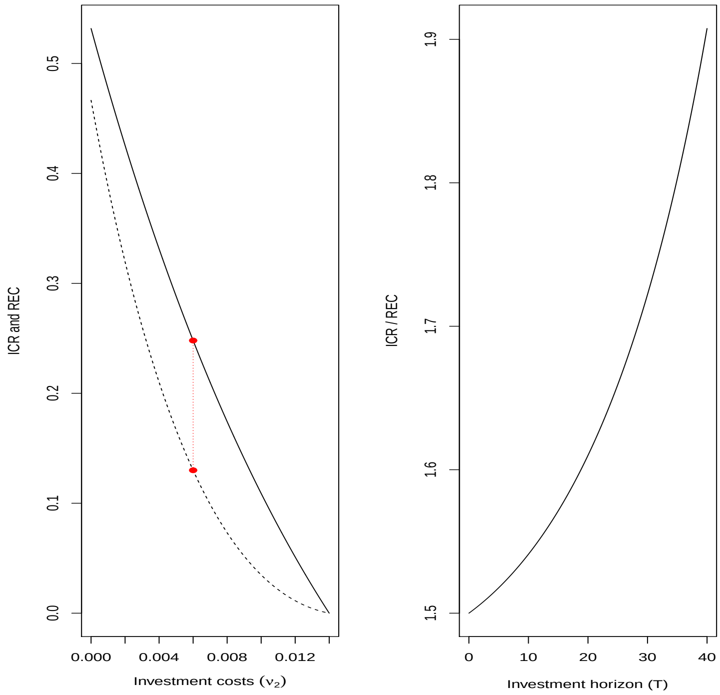

Figure 5.4: Illustrations of the indifferent compensation ratio (ICR) and the relative present value of the expected change in accumulated investment costs (REC). The left plot illustrates for a fixed horizon $T = 40$ the values when investment costs decreases from $\nu_1 = 1.4\%$ to $\nu_2 \in [0, \nu_1]$ (ICR solid curve, REC dashed curve). The dots indicate the base case example from earlier subsections where investment costs decrease to $\nu_2 = 0.06\%$. The right plot illustrates for a decrease in investment costs from $\nu_1 = 1.4\%$ to $\nu_2 = 0.6\%$ the proportion between the two measures (ICR/REC) for varying time horizons.

Note that the solution does not depend on the investment horizon.

The right-most curve in Figure 5.5 illustrates, for the parameters $(\nu_1, \pi_1^*) = (1.4\%, 60\%)$ and $\nu_2 \in [0, \nu_1]$, the cost-dependent value of π_2 for which the investor is indifferent between the two funds. From left to right the next 5 curves illustrates cost-dependent values of π_2 for which the investor is 10, 20, 30, 40 and 50 percent better off being a member of the cheaper fund. The line illustrates the cost-dependent optimal risk allocation π_2^*. The conclusion is surprisingly clear. If the investor is offered the lower cost rate $\nu_2 = 0.6\%$ instead of the higher cost rate $\nu_1 = 1.4\%$, he is better off almost no matter how much the investment strategy differs from his risk preferences. Any investment strategy $\pi_2 \in (0.28, 1.29)$ makes the cheaper fund preferable (remember $\pi_2^* = 78.5\%$ is optimal). As seen in Figure 5.5 this interval shrinks when considering scenarios where the investor is 10, 20, 30, 40 and 50 percent better off. Still, the range of investment strategies are incredible large. Once again, the conclusion, which is independent of the investment horizon, is very clear: The investor should be much more concerned with investment costs compared to being concerned with which investment strategy exactly meets his risk preferences.

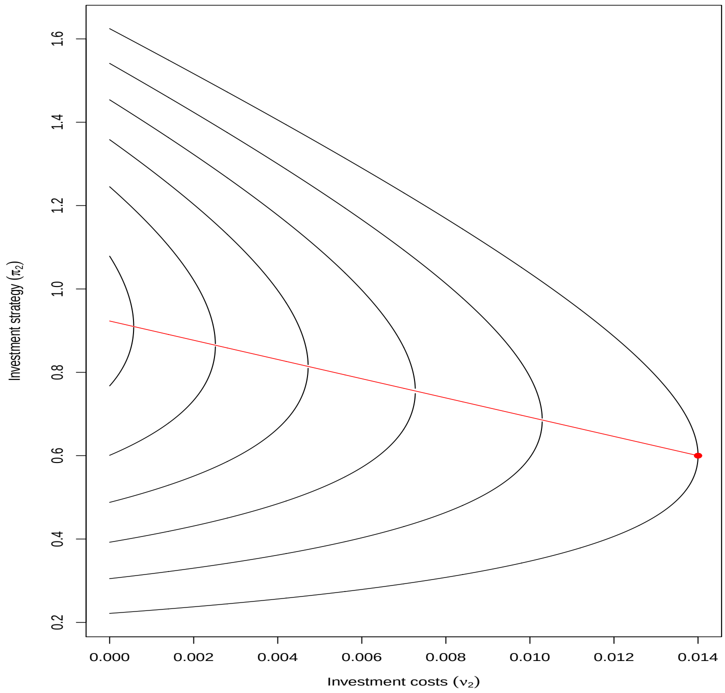

Figure 5.5: The curves illustrate for given investment costs $\nu_2 \in [0, \nu_1]$ the investment strategies π_2 which makes the investor (from right to left) 0, 10, 20, 30, 40 and 50 percent better off compared to being charged $\nu_1 = 1.4\%$ while using the optimal investment strategy $\pi_1^* = 60\%$. The line illustrates the cost-dependent optimal stock allocation.

6. Entrance times of random walks: With applications to pension fund modeling

Abstract: The purpose of the paper is twofold. First, we consider entrance times of random walks, i.e. the time of first entry to the negative axis. Partition sum formulas are given for entrance time probabilities, the n^{th} derivative of the generating function, and the n^{th} falling factorial entrance time moment. Similar formulas for the characteristic function of the position of the random walk both conditioned on entry and conditioned on no entry are also established. Second, we consider a model for a with-profits collective pension fund. The model has previously been studied by approximate methods, but we give here an essentially complete theoretical description of the model based on the entrance time results. We also conduct a mean-variance analysis for a fund in stationarity. To facilitate the analysis we devise a simple and effective exact simulation algorithm for sampling from the stationary distribution of a regenerative Markov chain.

Keywords: Random walks, entrance times, generating function, factorial moment, partition sum, stationarity, exact simulation, collective pension fund, with-profits contracts.

6.1 Introduction

This paper analyzes a "with-profits collective pension scheme"; this type of scheme and variants hereof are widespread, among other places, in the Nordic countries and the Netherlands. Members of the scheme are guaranteed a minimum benefit. The guarantees are a liability for the pension fund for which it must reserve an amount of money equal to the net present value of future guaranteed benefits (the *reserve*). In addition to the already guaranteed benefits members may receive *bonus* in the form of increased guarantees. Bonus is attributed periodically, e.g. annually, when the ratio of total assets to the reserve (the *funding ratio*) is sufficiently high. The phrase "with-profits" refers to this profit-sharing mechanism.

Assets in excess of the reserve are termed the *bonus potential*. The bonus potential allows the fund to invest in risky assets by absorbing adverse investment results. The scheme is "collective" in the sense that the bonus potential is considered common to all members. It is also collective in the sense that the investment strategy and bonus policy is the same for all members. Collective funds generally benefit from economy of scale in the form of low administration and investment costs. The flip side is the lack of an individual investment strategy.

We consider a model for a collective pension fund in which bonus is attributed when the funding ratio exceeds a given *bonus threshold*. The fund follows a CPPI (Constant Proportion Portfolio Insurance) investment strategy in order to stay solvent, i.e. to ensure that total assets exceed the reserve. The paper gives an essentially complete description of the fund dynamics including the time between bonuses, the (conditional) expected bonus percentage and the (conditional) expected funding ratio. The analysis is based on a detailed study of an embedded one-sided random walk obtained by a transformation of the funding ratio process sampled at the discrete set of time points where bonus can be attributed. We consider both a fund starting at the bonus threshold and a fund in stationarity. Furthermore, we use the results to perform a mean-variance analysis of a standardized benefit payout. The analysis is performed for a fund in stationarity representing the "average" member. To facilitate the analysis we also employ an *exact simulation algorithm*, which might be of independent interest.

The theoretical foundation for the analysis is a series of new results for entrance times of random walks derived in this paper. For a random walk started at the origin the entrance time is the time of entry into $(-\infty, 0]$. The main theoretical results are partition sum formulas for the n^{th} derivative of the entrance time generating function and for the n^{th} falling factorial entrance time moment. The latter result generalizes the well-known formula for the mean entrance time. We also give a partition sum formula for entrance time probabilities, and a similar formula for the position of the random walk conditioned on entrance at time n. These results are implicit in Spitzer (1956) and Asmussen (2003), but the proofs are new and simpler. Finally, we give a new partition sum formula for the position of the random walk conditioned on entrance taken place after time n.

The results allow explicit calculations of the entrance time probabilities and moments in terms of the marginal distribution of the random walk. The computational effort gradually becomes prohibitive, but the first 100, say, entrance time probabilities and moments are computationally feasible. We use the results to study the one-sided random walk embedded in the funding ratio process by utilizing the fact that a one-sided random walk and its associated (unrestricted) random walk are identical up to the time of first entry into $(-\infty, 0]$. However, the results are generally applicable and not limited to our pension fund application.

Optimizing utility from terminal wealth for an individual saving for retirement is treated by numerous papers. The foundation was laid by Richard (1975) and to mention a few, who among other results obtain optimal investment strategies, there is, Huang and Milevsky (2008) who allow for unspanned labor income; Huang et al. (2008) who separate the breadwinner income process from the family consumption process; Steffensen and Kraft (2008) who generalize to a multi-state Markov chain framework typically used by actuaries for modeling a series of life history events; Bruhn and Steffensen (2011) who generalize to a multi-person household, with focus on a married couple with economically and/or probabilistically dependent members; Kwak et al. (2011) who also consider a household but focus on generation issues; Kronborg and Steffensen (2013) who calculate the optimal investment strategy for a pension saver in the present of a minimum rate guarantee; and Gerber and Shiu (2000) who present a comprehensive discussion of terminal utility optimization in a pension saving context. There is also a vast literature on modern investment management, founded by Markowitz (1952) and Merton (1971), aimed at finding optimal investment strategies without the pension aspect.

In contrast to the literature cited above this paper takes the point of view of a pension fund where a group of people share a common investment strategy. Investment gains are shared through a bonus strategy by which collective funds above a threshold are transferred to individual guarantees. Based on results on utility optimization of durable goods by Hindi and Huang (1993) it can be shown that the optimal bonus strategy is to continuously attribute bonus whenever the funding ratio exceeds a certain barrier — thereby not allowing the funding ratio to exceed the barrier. In this paper, and in real life, the transfer is done periodically rather than continuously. References for continuous-time analysis of pension schemes taking both assumed (technical) returns, realized returns and bonus into account include Norberg (1999), Steffensen (2000), Norberg (2001), Steffensen (2004) and Nielsen (2006).

We assume the fund to follow a CPPI strategy. This strategy ensures that the fund remains funded, and it is "locally" optimal if we consider the periods between possible bonus attributions as local horizons. More precisely, Preisel et al. (2010) point out that CPPI is optimal in a finite horizon setting with HARA-utility and a subsistence level corresponding to a terminal funding ratio of one. CPPI strategies are treated, for the unrestricted case, by Cox and Huang (1989), and for the restricted case, by Teplá (2001). Using a CPPI strategy, and thereby avoiding insolvency, as done in this paper, stands in contrast to the literature on constructing contracts that are fair between owners and policyholders, see e.g. Briys and de Varenne (1997) and Grosen and Jørgensen (2000).

The current paper is related to Preisel et al. (2010), Kryger (2010) and Kryger (2011). We use the same underlying funding ratio dynamics, but the pension product and the terms by which members enter and leave the fund differ. In the cited papers members pay a fixed share (possibly zero) of contributions to the bonus potential on entry. This raises a number of issues regarding intergenerational fairness. In the present setup the share depends on the funding status of the fund in such a way that the contract is always financially fair. Methodologically, the cited papers use various analytical approximations while the current paper relies almost exclusively on exact results.

The main insight of Preisel et al. (2010) is that a given year's apparent success of a large bonus resulting from a high equity allocation can come at the even higher price of subsequent large losses trapping the company at a low funding ratio for a long period. They also derive approximations to the expected bonus and funding ratio in stationarity. Kryger (2010) finds optimal investment strategies for power utility and mean-variance criteria. For fixed values of the bonus threshold, he finds optimal investment strategies in the class of CPPI strategies for a fund in stationary. It is found that different investment strategies imply only modest differences in utility and, hence, that an investment collective can accommodate quite different attitudes towards risk. Finally, Kryger (2011) studies the impact of the pension design on efficiency and intergenerational fairness.

The rest of the paper is organized as follows. Section 6.2 presents the theoretical contributions on entrance times and moments of random walks. Section 6.3 describes the pension fund model, and Section 6.4 applies the random walk results to study bonus waiting times, size of bonus and the funding ratio. Results are given for a fund started at the bonus threshold and for a fund in stationarity. Section 6.5 contains a comprehensive numerical application including a mean-variance analysis. It also explains the exact simulation algorithm used in the analysis. Finally, the appendix contains proofs for the results of Section 6.2 and additional lemmas.

6.2 Random walks

In this section we present a series of results on entrance times and conditional characteristic functions of random walks. The results will be used in subsequent sections to provide a detailed description of the distribution of bonus times, size of bonus and funding ratio of the collective pension fund model under study. However, the results are generally applicable and can be applied in many other contexts as well.

The entrance time of a random walk is defined as the (first) time of entry into $(-\infty, 0]$ after time 0. The results to follow devise how a number of quantities related to entrance times can be computed as sums over partition sets. We present both new results and existing results with new and simpler proofs. The results fall in three parts.

First, we derive a closed-form formula for the entrance time probabilities of a random walk started at the origin (Theorem 6.2.3). This result is also implicit in the seminal paper by Spitzer (1956), but we give here a simpler self-contained proof. Second, we derive an expression for the n^{th} derivative of the generating function for the entrance time (Theorem 6.2.5), which we subsequently use to derive a formula for the factorial moments (Theorem 6.2.6). These results are new. Third, we derive formulas for the characteristic function of the position of the random walk conditioned on entrance at time n (Theorem 6.2.8) and on entrance after time n (Theorem 6.2.9). The first of these results is known, but the proof is new, while the second result is new. Most of the proofs rely on combinatorial arguments, some of which might be of independent interest, in particular Lemma 6.6.1.

6.2.1 Entrance times and partitions

Consider the random walk

$$S_0 = 0 \text{ and } S_n = S_{n-1} + X_n \text{ for } n \in \mathbb{N}, \tag{6.1}$$

where $X_1, X_2, \ldots$ are i.i.d. random variables. Following the notation and terminology of Asmussen (2003) we let τ_- denote the *entrance time* to $(-\infty, 0]$, also known as the first *(weak) descending ladder epoch*, defined by

$$\tau_- = \inf\{n \geq 1 : S_n \leq 0\}. \tag{6.2}$$

In this section our prime interest is the calculation of the entrance time probabilities

$$\tau_n = P(\tau_- = n) = P(S_1 > 0, \ldots, S_{n-1} > 0, S_n \leq 0), \tag{6.3}$$

i.e. the probability that the entry into $(-\infty, 0]$ occurs at the n^{th} step. To facilitate the study of $(\tau_n)_{n\in\mathbb{N}}$ we introduce its generating function, defined for $0 \leq s \leq 1$ by

$$\tau(s) = \sum_{n=1}^{\infty} \tau_n s^n. \tag{6.4}$$

Let $p_n = P(S_n \leq 0)$ for $n \geq 1$. The following surprising theorem, originally due to Andersen (1954)[1], expresses τ in terms of the (marginal) probabilities p_n. The original proof is complicated but a simple combinatorial proof now exists, see e.g. Theorem XII.7.1 of Feller (1971).

Theorem 6.2.1. (Sparre Andersen Theorem) *For* $0 \leq s < 1$

$$\log\left(\frac{1}{1-\tau(s)}\right) = \sum_{n=1}^{\infty} \frac{s^n}{n} p_n. \tag{6.5}$$

By Theorem 6.2.1 we can write the generating function as

$$\tau(s) = 1 - e^{H(s)}, \text{ where } H(s) = -\sum_{n=1}^{\infty} \frac{s^n}{n} p_n. \tag{6.6}$$

Now, since

$$\tau_n = \frac{\tau^{(n)}(0)}{n!}, \tag{6.7}$$

where $\tau^{(n)}$ denotes the n^{th} derivative of τ, the entrance time probabilities can in principle be calculated by repeated differentiation of expression (6.6). However, direct differentiation leads to an exponentially increasing number of terms (the number of terms almost triples on each iteration) so this approach is infeasible in practise for all but the smallest n. Fortunately, the number of different terms is substantially smaller. This observation gives rise to a summation formula which makes it computationally feasible to calculate τ_n for n up to at least 100. In order to state the result we define the set of integer partitions of a given order.

Definition 6.2.1. *Define for* $n \geq 1$ *the* partition set *of order* n *by*

$$\mathcal{D}_n = \left\{(\sigma_1, \ldots, \sigma_n)\middle|\sigma_1 \in \mathbb{N}_0, \ldots, \sigma_n \in \mathbb{N}_0, \sum_{i=1}^{n} i\sigma_i = n\right\}, \tag{6.8}$$

where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$.

[1]Erik Sparre Andersen (1919–2003) was a Danish mathematician and actuary. He played a prominent role in the design and operation of the Danish Labour Market Supplementary Pension Scheme (ATP) from its foundation in 1964 and in the years to follow. The results of this paper rest to a large extent on the work of our former colleague.

By definition $\mathcal{D}_0$ consists of the empty partition. Integer partitions occur in number theory and combinatorics, and the size of $\mathcal{D}_n$ as a function of n (the partition function) is a well-studied object. The following asymptotic expression is due to Hardy and Ramanujan (1918)

$$\#\mathcal{D}_n \approx \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{2n/3}}. \tag{6.9}$$

We note that the size of $\mathcal{D}_n$ increases sub-exponentially in n.

For later use, we define for $\sigma \in \mathcal{D}_n$ the sign of the partition and two combinatorial coefficients

$$\operatorname{sgn}(\sigma) = (-1)^{\sum_{i=1}^n \sigma_i}, \quad d_\sigma = \prod_{i=1}^n \sigma_i! i^{\sigma_i}, \quad c_\sigma = \frac{n!}{\prod_{i=1}^n \sigma_i!(i!)^{\sigma_i}}. \tag{6.10}$$

We also need the following fundamental identity often used in connection with generating functions, see e.g. Chapter 7 of Szpankowski (2001) for a proof.

Theorem 6.2.2. *Provided* $\sum_{n=1}^\infty a_n b^n$ *converges absolutely*

$$\exp\left(\sum_{n=1}^\infty a_n b^n\right) = 1 + \sum_{m=1}^\infty \sum_{\sigma\in\mathcal{D}_m} \frac{b^m}{\sigma_1!\sigma_2!\cdots\sigma_m!} \prod_{n=1}^m a_n^{\sigma_n}. \tag{6.11}$$

Combination of Theorem 6.2.1 and Theorem 6.2.2 yields a partition sum formula for the entrance time probabilities in terms of the probabilities p_n. This result can also be derived from Spitzer (1956), but the present proof is considerably simpler.

Theorem 6.2.3. *For* $n \geq 1$

$$\tau_n = -\sum_{\sigma\in\mathcal{D}_n} \frac{\operatorname{sgn}(\sigma)}{d_\sigma} \prod_{i=1}^n p_i^{\sigma_i}. \tag{6.12}$$

Proof. See Appendix 6.6.1. □

Provided p_n are available Theorem 6.2.3 makes it feasible to calculate entrance time probabilities for fairly large values of n. Table 6.2.1 shows the size of $\mathcal{D}_n$, i.e. the number of terms in the partition sum (6.12), and the number of terms when (6.7) is used directly.[2] Clearly, the computational gain is massive.

n	1	2	3	4	5	10	40	100	200
$\#\mathcal{D}_n$	1	2	3	5	7	42	37338	$1.9057 \cdot 10^8$	$3.9730 \cdot 10^{12}$
#Naive	1	2	4	10	26	$5.476 \cdot 10^3$	$1.1257 \cdot 10^{18}$	$4.7720 \cdot 10^{46}$	$2.4594 \cdot 10^{94}$

Table 6.1: Number of terms needed to calculate the entrance time probabilities by formula (6.12) and (6.7), respectively.

Theorem 6.2.3 also provides the following purely combinatoric result (which is used in the proof of Lemma 6.6.1). For $n \geq 2$

$$\sum_{\sigma\in\mathcal{D}_n} \frac{\operatorname{sgn}(\sigma)}{d_\sigma} = 0. \tag{6.13}$$

This follows from (6.12) by considering the degenerate case $X_i \equiv 0$ in which case $p_n = 1$ for all n, $\tau_1 = 1$ and $\tau_n = 0$ for $n \geq 2$.

The coefficients d_σ obey a number of other interesting relations, e.g. the following theorem which shows that $1/d_\sigma$ can be interpreted as a probability distribution on $\mathcal{D}_n$. The proof of the theorem also serves as an illustration of the combinatorial method used throughout.

[2]The number of terms obtained by differentiating the generating function n times without collecting terms.

Theorem 6.2.4. *For $n \geq 1$ and $1 \leq k \leq n$*

$$\sum_{\sigma \in \mathcal{D}_n} \frac{1}{d_\sigma} = \sum_{\sigma \in \mathcal{D}_n} \frac{k\sigma_k}{d_\sigma} = 1. \tag{6.14}$$

6.2.2 Entrance time moments

In principle the entrance time moments (and other characteristics) can be calculated from the entrance time probabilities, τ_n. However, since the calculation of τ_n becomes increasingly difficult it is both theoretically and practically important to have more direct means of calculating moments. In this section we present a formula for the falling factorial entrance time moments. This result generalizes the well-know formula for the mean entrance time[3]

$$E(\tau_-) = \exp\left(\sum_{k=1}^{\infty} \frac{1}{k} P(S_k > 0)\right). \tag{6.15}$$

The formula for the factorial moments rely on the following main result which gives a partition sum representation of the n^{th} derivative of the generating function. We denote by $f^{(n)}$ the n^{th} derivative of a function f.

Theorem 6.2.5. *For $n \geq 1$ and $0 \leq s < 1$*

$$\tau^{(n)}(s) = -e^{H(s)} \sum_{\sigma \in \mathcal{D}_n} c_\sigma H_\sigma(s), \tag{6.16}$$

where c_σ is given by (6.10) and

$$H_\sigma(s) = \prod_{i=1}^{n} \left(H^{(i)}(s)\right)^{\sigma_i}. \tag{6.17}$$

Proof. See Appendix 6.6.2. □

Note that Theorem 6.2.3 can be derived from Theorem 6.2.5 since $\tau_n = \tau^{(n)}(0)/n!$ and $H^{(i)}(0) = -(i-1)!p_i$. This constitutes an alternative proof of Theorem 6.2.3 which does not rely on Theorem 6.2.2.

For integers m and n we denote by $(m)_n$ the n^{th} falling factorial of m,

$$(m)_n = m(m-1)\cdots(m-n+1). \tag{6.18}$$

In particular, $(m)_1 = m$ and $(m)_n = 0$ for $n > m$. By monotone convergence the n^{th} factorial moment of τ_- is given by

$$E((\tau_-)_n) = \lim_{s \to 1-} \tau^{(n)}(s), \tag{6.19}$$

whether or not the limit is finite. Using Theorem 6.2.5 and (6.19) in combination with Lemmas 6.6.1 and 6.6.2 the following result for the n^{th} factorial moment can be derived.

Theorem 6.2.6. *If $n \geq 1$ and $\sum_{k=1}^{\infty} k^{n-2} P(S_k > 0) < \infty$ then*

$$E((\tau_-)_n) = \exp\left(\sum_{k=1}^{\infty} \frac{1}{k} P(S_k > 0)\right) n! \sum_{\sigma \in \mathcal{D}_{n-1}} \prod_{i=1}^{n-1} \left(\frac{h_i}{i!}\right)^{\sigma_i} \frac{1}{\sigma_i!} < \infty, \tag{6.20}$$

where, for $n = 1$, the last sum is 1 by definition, and for $1 \leq i \leq n-1$

$$h_i = \sum_{k=i}^{\infty} \frac{(k)_i}{k} P(S_k > 0). \tag{6.21}$$

---

[3]Expression (6.15) can be derived from Theorem 6.2.1 by a limit argument, see e.g. Theorem XII.7.3 of Feller (1971).

Proof. See Appendix 6.6.2. $\square$

By use of Theorem 6.2.6 we can calculate entrance time moments of any order. For the first three moments we get

$$E(\tau_-) = \exp\left(\sum_{k=1}^{\infty} \frac{q_k}{k}\right), \tag{6.22}$$

$$E\left(\tau_-^2\right) = E((\tau_-)_2) + E(\tau_-) = E(\tau_-)\left(2\sum_{k=1}^{\infty} q_k + 1\right), \tag{6.23}$$

$$\begin{aligned} E\left(\tau_-^3\right) &= E((\tau_-)_3) + 3E((\tau_-)_2) + E(\tau_-) \\ &= E(\tau_-)\left(3\sum_{k=2}^{\infty}(k-1)q_k + 3\sum_{k=1}^{\infty} q_k \left[\sum_{k=1}^{\infty} q_k + 2\right] + 1\right), \end{aligned} \tag{6.24}$$

where $q_k = 1 - p_k = P(S_k > 0)$. These formulas are all easy to evaluate to any desired degree of accuracy.

6.2.3 Conditional characteristic functions

In this section we present results characterizing the position of the random walk upon entrance to $(-\infty, 0]$ (the weak descending ladder height) and the position when entrance has not yet occurred. We will need the combined generating and characteristic function defined for $|s| < 1$ and $\zeta \in \mathbb{R}$ by

$$\chi(s, \zeta) = E\left(s^{\tau_-} e^{i\zeta S_{\tau_-}}\right). \tag{6.25}$$

For a random variable X and an event A we write $E(X; A)$ for $E(X 1_A)$, and $E(X|A)$ for $E(X 1_A)/P(A)$. From Theorem VII 4.1 of Asmussen (2003) we have the following generalization of Theorem 6.2.1

Theorem 6.2.7. *For* $|s| < 1$ *and* $\zeta \in \mathbb{R}$

$$\log\left(\frac{1}{1 - \chi(s, \zeta)}\right) = \sum_{n=1}^{\infty} \frac{s^n}{n} E\left(e^{i\zeta S_n}; S_n \le 0\right). \tag{6.26}$$

By combining Theorem 6.2.7 and Theorem 6.2.2 we obtain a partition sum formula for the characteristic function of the random walk given entrance to $(-\infty, 0]$ at time n. This result is similar to Theorem 6.2.3 for the entrance time probabilities.

Theorem 6.2.8. *For* $n \ge 1$ *and* $\zeta \in \mathbb{R}$

$$E\left(e^{i\zeta S_n} | \tau_- = n\right) = -\frac{1}{\tau_n} \sum_{\sigma \in \mathcal{D}_n} \frac{\operatorname{sgn}(\sigma)}{d_\sigma} \prod_{k=1}^{n} \left(E\left(e^{i\zeta S_k}; S_k \le 0\right)\right)^{\sigma_k}. \tag{6.27}$$

Proof. See Appendix 6.6.3. $\square$

It is also of interest to know the distribution of the random walk given that it has not yet entered $(-\infty, 0]$. It turns out that the characteristic function for this distribution can also be calculated as a partition sum. The result is established by subtracting the characteristic function of Theorem 6.2.8 up to time n from the unconditional characteristic function and using Lemma 6.6.1 to identify the resulting structure.

Theorem 6.2.9. *For $n \geq 1$ and $\zeta \in \mathbb{R}$*

$$E(e^{i\zeta S_n}|\tau_- > n) = \frac{1}{P(\tau_- > n)} \sum_{\sigma \in \mathcal{D}_n} \frac{1}{d_\sigma} \prod_{k=1}^{n} \left(E(e^{i\zeta S_k}; S_k > 0)\right)^{\sigma_k}. \tag{6.28}$$

Proof. See Appendix 6.6.3. □

Note that if $P(\tau_- < \infty) = 1$ then $P(\tau_- > n) = 1 - \sum_{i=1}^{n} \tau_i$ such that (6.28) can indeed be calculated.

6.3 Pension fund model

We consider a model for a collective pension fund with a with-profits pension product. Each contribution is split into a part giving a guaranteed payment and a part invested in a, possibly leveraged, investment portfolio. The product is with-profit in the sense that all guaranteed payments are increased, known as bonus, when the funding ratio exceeds a given threshold level. The investment strategy and bonus policy are common and all members receive the same bonus (percentage). In our model members enter and leave the fund on financially fair terms, although this is not necessarily strictly true in practice. Despite its simplicity the model resembles the traditional collective pension funds known from e.g. the Nordic countries and the Netherlands. The random walk results presented in Section 6.2 will be used to give an essentially complete description of the dynamics of the fund.

First, consider a frictionless Black-Scholes market consisting of a bank account, B, with risk free short rate, r, and a risky stock, S, with dynamics given by

$$dB(t) = rB(t)dt,\ B(0) = 1, \tag{6.29}$$

$$dZ(t) = (r + \mu)Z(t)dt + \sigma_Z Z(t)dW(t),\ Z(0) = z_0 > 0. \tag{6.30}$$

Here r, μ and σ_Z are strictly positive constants. The process W is a standard Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ equipped with the filtration $\mathbb{F}^W = (\mathcal{F}^W(t))_{t\geq 0}$ given by the P-augmentation of the filtration $(\sigma\{W(s);\ 0 \leq s \leq t\})_{t\geq 0}$.

The decision to attribute bonus or not is taken at a set of equidistant, discrete set of time points $0 = t_0 < t_1 < \ldots$. We assume for simplicity that contributions and benefits also fall at these times. The market value of the guaranteed benefits, the reserve, is denoted $R(t)$. Not taking mortality into account (paramount to assuming that mortality risk can be neglected by the law of large numbers) the evolution of the reserve between potential bonus times is given by

$$R(t) = R(t_i)e^{r(t-t_i)}, \quad \text{where } i = \max\{j \in \mathbb{N}_0 : t_j \leq t\}. \tag{6.31}$$

The total assets of the fund is denoted $A(t)$ and the funding ratio of the fund is defined as

$$F(t) = \frac{A(t)}{R(t)}. \tag{6.32}$$

The difference between total assets and the reserve, $A(t) - R(t)$, is called the *bonus potential* (or surplus). We assume that the fund attributes bonus according to a threshold bonus strategy such that at time t_i all guaranteed payments are increased by

$$r_i^B = \begin{cases} 0 & \text{if } F(t_i-) \leq \kappa, \\ \frac{F(t_i-)-\kappa}{\kappa} & \text{if } F(t_i-) > \kappa, \end{cases} \tag{6.33}$$

where κ is assumed to be strictly larger than 1. Note that immediately after a bonus attribution the funding ratio equals κ. Let $\tilde{F}_i$ denote the funding ratio at time t_i after (possible) bonus attribution, but before contributions and benefits have fallen. Thus $\tilde{F}_i = F(t_i-)/(1+r_i^B) = \min\{F(t_i-), \kappa\}$.

The pension product:

- Contributions do not affect the funding ratio, i.e. for contributions received at time t_i only the fraction $1/\tilde{F}_i$ is guaranteed (at rate r) and enters to the reserve while the remainder enters to the bonus potential.
- The initially guaranteed benefit is entitled to bonus from the time of contribution to the time of payment.
- Benefits do not affect the funding ratio, i.e. guaranteed benefits paid out at time t_i are increased by $\tilde{F}_i$ (terminal bonus).

Note that the fraction of contributions guaranteed at the risk-free rate[4] depends on the current funding ratio of the fund. Also note that for each contribution the member pays a price to enter the collective fund, but he also receives his share of the surplus for each benefit paid out.

Let c_i and b_i denote the contributions and benefits, respectively, at time t_i, and let c_i^G and b_i^G denote the part of contributions and benefits guaranteed. Total assets and the reserve at time t_i is then given by

$$A(t_i) = A(t_i-) + c_i - b_i, \tag{6.34}$$

$$R(t_i) = \left(1 + r_i^B\right) R(t_i-) + c_i^G - b_i^G, \tag{6.35}$$

where $c_i^G = c_i/\tilde{F}_i$ and $b_i^G = b_i/\tilde{F}_i$. Hence, by construction $F(t_i) = \tilde{F}_i$ irrespective of the size of contributions and benefits. It is not hard to show that the fact that pension savers enter and leave the pension fund without effecting the funding ratio makes the scheme financially fair. There is no redistribution of wealth between generations.

The simplest pension product to be considered is the case in which a contribution of one is made at time $t = 0$ for a benefit paid out in its entirety at $t = T$ (retirement). In this case the benefit at retirement becomes

$$O_T = \frac{F(T)}{F(0)} e^{rT} \prod_{i=1}^{T} \left(1 + r_i^B\right). \tag{6.36}$$

Different contribution and benefit profiles can also be analyzed within this model. The point to note is that the payoff depends solely on the funding ratio dynamics.

We assume that the pension fund has to stay funded at all times, i.e. its assets must not fall below the reserve or, equivalently, the funding ratio must not fall below one. In order to achieve this the fund pursues a CPPI (constant proportion portfolio insurance) strategy by which a constant fraction, C, of the bonus potential is invested in stocks. The remaining assets are invested in the risk-free asset. We allow for values of C greater than one, i.e. leverage of the bonus potential is possible. The dynamics of the assets between time t_i and t_{i+1} is given by

$$dA(t) = (r + \gamma(t)\mu)dt + \gamma(t)\sigma_Z dW(t), \tag{6.37}$$

where $\gamma(t) = C\frac{F(t)-1}{F(t)}$.

Let Δ denote the time between possible bonus attributions, $t_i = i\Delta$, and let $F_i = F(t_i)$. It follows from Preisel et al. (2010)) that the funding ratio process sampled at t_i evolves live a (discrete-time) Markov chain with dynamics

$$F_i = \min\left\{(F_{i-1} - 1)e^{\left(C\mu - \frac{1}{2}C^2\sigma_Z^2\right)\Delta + C\sigma_Z\sqrt{\Delta}U_i} + 1, \kappa\right\}, \tag{6.38}$$

[4] We disregard throughout mortality and the associated mortality bequest from the deceased to the survivors which acts as an increased rate of return in the case of life annuities.

where the U_i's are i.i.d. standard normal variables. In particular, if $F_0 > 1$ all subsequent F_i's are strictly larger than 1 (and at most κ).

The fund has to decide on an investment strategy, C, and a bonus policy, κ. A high bonus threshold implies that the fund can invest more freely but also that only a small fraction of the pension is guaranteed. This may or may not be in the interest of the members. Similarly, an aggressive investment strategy implies a higher probability of very high returns, but also a higher risk of very low returns (on the bonus potential).

The purpose of the rest of the paper is twofold. First, we characterize the impact of C and κ in terms of bonus time (time between bonus attributions), bonus size and funding ratio. The characterization is provided for funds started at κ and in stationarity (the long-run average). Second, we propose a criterion by which κ and C can be determined. The criterion is evaluated in stationarity to reflect the fact that the fund is collective and should be designed for the benefit of the average member.

6.4 Bonus and funding ratio

The time between consecutive bonuses, the size of the bonus and the funding ratio given no bonus has yet been attributed can all be analyzed by the random walk results of Section 6.2.

Consider the following transformation

$$Y_n = -\log\left(\frac{F_n - 1}{\kappa - 1}\right) \text{ for } n \in \mathbb{N}_0. \tag{6.39}$$

This transformation turns the funding ratio process (6.38) into the one-sided random walk

$$Y_n = (Y_{n-1} + X_n)^+, \tag{6.40}$$

where the X_n's are i.i.d. normally distributed with mean $-\left(C\mu - \frac{1}{2}C^2\sigma_Z^2\right)\Delta$ and variance $C^2\sigma_Z^2\Delta$. Note that $F_n = \kappa$ corresponds to $Y_n = 0$, while funding ratios close to one correspond to high values of Y.

Along with Y we also consider the (unrestricted) random walk of Section 6.2,

$$S_0 = 0 \text{ and } S_n = S_{n-1} + X_n \text{ for } n \in \mathbb{N}, \tag{6.41}$$

with the same X_n's as in (6.40). The distribution of S_n is given by

$$S_n \sim N\left(-n\left(C\mu - \frac{1}{2}C^2\sigma_Z^2\right)\Delta, nC^2\sigma_Z^2\Delta\right). \tag{6.42}$$

Thus, in the notation of Section 6.2 we have

$$p_n = P(S_n \leq 0) = \Phi\left(\sqrt{n\Delta}\frac{\mu - \frac{1}{2}C\sigma_Z^2}{\sigma_Z}\right), \tag{6.43}$$

where Φ denotes the cumulative distribution function (CDF) of a standard normal distribution.

6.4.1 Stationarity

The first question of interest is whether the fund admits a stationary distribution or not. In the stationary case the funding ratio distribution converges towards a non-degenerate distribution, otherwise it converges (in probability) towards one. The following result answers the question in terms of the aggressiveness of the investment strategy (the result was also by Preisel et al. (2010) albeit in a different parametrization).

Proposition 6.4.1. *The funding ratio process (6.38) admits a stationary distribution if and only if* $C < \frac{2\mu}{\sigma_Z^2}$.

Proof. By Proposition 11.5.3 of Meyn and Tweedie (2009) we have that the Y-process, and hence the F-process, admits a stationary distribution iff the mean of the increments X_n is strictly negative, i.e. iff $C < \frac{2\mu}{\sigma_Z^2}$. □

When it exists, we will denote the stationary funding ratio distribution by π. We note that the existence of a stationary distribution is independent of how often (Δ) and at which level (κ) bonus is allotted. A stationary distribution exists if and only if the median return on the bonus potential is positive. If the bonus potential is invested more aggressively than that it will eventually get lost (in the boundary case, $C = 2\mu/\sigma_Z^2$, bonus will in fact be attributed infinitely often, but the average time between each bonus is infinite!).

6.4.2 Bonus times

We refer to the time between consecutive bonus attributions as bonus times, or more precisely bonus waiting time. Formally, these are defined by

$$\mathcal{T}_1 = \inf\{n \geq 1 : F_n = \kappa\} = \inf\{n \geq 1 : Y_n = 0\}, \tag{6.44}$$

and, recursively, for $k \geq 2$

$$\mathcal{T}_k = \inf\{n \geq 1 : F_{\mathcal{T}_{k-1}+n} = \kappa\} = \inf\{n \geq 1 : Y_{\mathcal{T}_{k-1}+n} = 0\}. \tag{6.45}$$

Consider first the case where $F_0 = \kappa$. Then $Y_0 = 0$ and the first bonus time coincides with the entrance time of S to $(-\infty, 0]$, i.e. $\mathcal{T}_1 = \tau_-$, where τ_- is given by (6.2) of Section 6.2. Further, since the funding ratio is κ after a bonus attribution it follows by the Markov property that all subsequent bonus times are independent and distributed as $\mathcal{T}_1$.

Consider next the stationary case and assume that F_0 is distributed according to the stationary funding ratio distribution. Imagine that the fund has been operating since time minus infinity. The probability that at time 0 we are in a period (between two bonuses) of length k is then given by $k\tau_k / \sum_{n=1}^{\infty} n\tau_n$, i.e. the probability is proportional to the length of the period times the frequency by which it occurs. Further, given that we are in a period of length k the probability that we are n places ($n \leq k$) away from the end is $1/k$, since each position is equally likely. Summing over all possible k's we get that the probability that the next bonus occurs at time $n \geq 1$ is given by

$$P_\pi(\mathcal{T}_1 = n) = \sum_{k=n}^{\infty} \frac{1}{k} \left(\frac{k\tau_k}{\sum_{n=1}^{\infty} n\tau_n} \right) = \frac{\sum_{k=n}^{\infty} \tau_k}{E(\tau_-)} = \frac{1 - \sum_{k=1}^{n-1} \tau_k}{E(\tau_-)}, \tag{6.46}$$

where we use subscript π to denote that F_0 is drawn from the stationary distribution. When the first bonus (after time 0) has been attributed the funding ratio is κ. Hence, all subsequent bonus times are distributed as τ_-.

Using subscript κ to denote the case $F_0 = \kappa$ we thus have

Proposition 6.4.2. *For* $k \geq 1$ *and* $n \geq 1$

$$P_\kappa(\mathcal{T}_k = n) = \tau_n, \tag{6.47}$$

and, provided the stationary distribution exists,

$$P_\pi(\mathcal{T}_k = n) = \begin{cases} \left(1 - \sum_{k=1}^{n-1} \tau_k\right) / E(\tau_-) & \text{for } k = 1, \\ \tau_n & \text{for } k \geq 2, \end{cases} \tag{6.48}$$

where τ_n *is given by (6.12) of Theorem 6.2.3 and* $E(\tau_-)$ *is given by (6.15) of Section 6.2.2*

Note that in stationarity the probability of receiving bonus in any given year is

$$P_\pi(r_1^B > 0) = P_\pi(Y_1 = 0) = P_\pi(\mathcal{T}_1 = 1) = \frac{1}{E(\tau_-)}. \tag{6.49}$$

This relationship is also known as Kac's theorem.

In the stationary case the drift of S is negative and it is not hard to show that the criterion of Theorem 6.2.6 is satisfied for all n. Hence, the time between bonuses has factorial moments of all orders and these can be calculated by (6.20) of Theorem 6.2.6.[5]

6.4.3 Number of bonuses

The number of bonuses in a given period can be calculated from the bonus time distribution. Let $R_k = \mathcal{T}_1 + \ldots + \mathcal{T}_k$ denote the time of the k'th bonus, also known as the renewal epochs. For $F_0 = \kappa$ and $F_0 \sim \pi$ the distribution of R_k can be calculated by the recursion

$$P_*(R_1 = n) = P_*(\mathcal{T}_1 = n), \tag{6.50}$$

$$P_*(R_k = n) = \sum_{j=k-1}^{n-1} P_*(R_{k-1} = j)\tau_{n-j}. \tag{6.51}$$

where $P_*(\mathcal{T}_1 = n)$ is given by Proposition 6.4.2 (and $*$ is either κ or π).

Let N_n denote the number of bonuses from time 1 to time n. We have

$$P_*(N_n \le k) = P_*(R_{k+1} > n) = 1 - \sum_{j=k+1}^{n} P_*(R_{k+1} = j), \tag{6.52}$$

where $*$ is either κ or π.

6.4.4 Bonus percentage

The bonus percentage distribution can be derived from the (descending) ladder height distribution of the random walk. For $F_0 = \kappa$ the events $(\mathcal{T}_1 = n)$ and $(\tau_- = n)$ are identical and on this event the bonus percentage given by (6.33) becomes

$$r_n^B = \frac{\kappa - 1}{\kappa}\left(e^{-S_n} - 1\right). \tag{6.53}$$

The mean bonus percentage given the time of bonus can then be calculated by use of Theorem 6.2.8.

Proposition 6.4.3. *For $n \ge 1$*

$$E_\kappa\left(r_n^B \middle| \mathcal{T}_1 = n\right) = \frac{\kappa - 1}{\kappa}\left[E\left(e^{-S_n} \middle| \tau_- = n\right) - 1\right], \tag{6.54}$$

where

$$E\left(e^{-S_n} \middle| \tau_- = n\right) = -\frac{1}{\tau_n} \sum_{\sigma \in \mathcal{D}_n} \frac{\operatorname{sgn}(\sigma)}{d_\sigma} \prod_{k=1}^{n} \left(E\left(e^{-S_k}; S_k \le 0\right)\right)^{\sigma_k} \tag{6.55}$$

with

$$E\left(e^{-S_k}; S_k \le 0\right) = e^{kC\mu\Delta}\Phi\left(\sqrt{k\Delta}\frac{\mu + \frac{1}{2}C\sigma_Z^2}{\sigma_Z}\right). \tag{6.56}$$

Further, for $n \ge 1$ and provided the stationary distribution exists

$$E_\pi\left(r_n^B \middle| \mathcal{T}_1 = n\right) = \frac{1}{P_\pi(\mathcal{T}_1 = n)} \sum_{k=0}^{\infty} E_\kappa\left(r_{n+k}^B \middle| \mathcal{T}_1 = n + k\right) \frac{\tau_{k+n}}{E(\tau_-)}. \tag{6.57}$$

[5] In fact, it can be shown that Y is so-called geometrically ergodic which implies exponential moments of the return time to 0, i.e. the time between bonuses.

Proof. Formula (6.54) follows directly from (6.53). Theorem 6.2.8 identifies the conditional distribution of S_n given $\tau_- = n$ as a linear combination of conditional normal tail measures. Since the normal distribution has exponential moments of all orders we conclude (6.55) by dominated convergence. Expression (6.56) follows from (6.42) by standard calculations.

For the stationary case we consider a fund which has been run since time minus infinity. Let λ denote time since bonus was last attributed,

$$\lambda = \inf\{n \geq 0 : F_{-n} = \kappa\} = \inf\{n \geq 0 : Y_{-n} = 0\}. \tag{6.58}$$

By the argument of Section 6.4.2 leading to (6.46) and the Markov property we have

$$P_\pi\left(\mathcal{T}_1 = n, \lambda = k\right) = \frac{1}{n+k}\frac{(n+k)\tau_{k+n}}{E(\tau_-)} = \frac{\tau_{k+n}}{E(\tau_-)}, \tag{6.59}$$

$$E_\pi\left(r_n^B \middle| \mathcal{T}_1 = n, \lambda = k\right) = E_\kappa\left(r_{n+k}^B \middle| \mathcal{T}_1 = n+k\right). \tag{6.60}$$

By summing over all possible values of λ we obtain

$$\begin{aligned} E_\pi\left(r_n^B \middle| \mathcal{T}_1 = n\right) &= \frac{1}{P_\pi(\mathcal{T}_1 = n)} E_\pi\left(r_n^B; \mathcal{T}_1 = n\right) \\ &= \frac{1}{P_\pi(\mathcal{T}_1 = n)} \sum_{k=0}^{\infty} E_\pi\left(r_n^B; \mathcal{T}_1 = n, \lambda = k\right) \\ &= \frac{1}{P_\pi(\mathcal{T}_1 = n)} \sum_{k=0}^{\infty} E_\pi\left(r_n^B \middle| \mathcal{T}_1 = n, \lambda = k\right) P_\pi\left(\mathcal{T}_1 = n, \lambda = k\right). \end{aligned} \tag{6.61}$$

Finally, inserting (6.59) and (6.60) in (6.61) yields (6.57). □

Higher order polynomial moments of r^B can be expressed in terms of exponential moments of S_n by expanding (6.53). It is straightforward to extend Proposition 6.4.3 to cover this case also.

6.4.5 Funding ratio

We consider at last the funding ratio of the fund given that no bonus has yet been attributed. For $F_0 = \kappa$ the events $(\mathcal{T}_1 > n)$ and $(\tau_- > n)$ are identical and on this event

$$F_n = (\kappa - 1)e^{-S_n} + 1. \tag{6.62}$$

Polynomial funding ratio moments can be derived by use of Theorem 6.2.9. For the mean funding ratio we have the following result.

Proposition 6.4.4. *For $n \geq 1$*

$$E_\kappa\left(F_n \middle| \mathcal{T}_1 > n\right) = (\kappa - 1)E\left(e^{-S_n} \middle| \tau_- > n\right) + 1, \tag{6.63}$$

where

$$E(e^{-S_n}|\tau_- > n) = \frac{1}{P(\tau_- > n)} \sum_{\sigma \in \mathcal{D}_n} \frac{1}{d_\sigma} \prod_{k=1}^{n} \left(E(e^{-S_k}; S_k > 0)\right)^{\sigma_k} \tag{6.64}$$

with

$$E\left(e^{-S_k}; S_k > 0\right) = e^{kC\mu\Delta}\Phi\left(-\sqrt{k\Delta}\frac{\mu + \frac{1}{2}C\sigma_Z^2}{\sigma_Z}\right). \tag{6.65}$$

Further, for $n \geq 1$ and provided the stationary distribution exists

$$E_\pi\left(F_n \middle| \mathcal{T}_1 > n\right) = \frac{1}{P_\pi(\mathcal{T}_1 > n)} \sum_{k=0}^{\infty} E_\kappa\left(F_{n+k} \middle| \mathcal{T}_1 > n+k\right) P_\pi(\mathcal{T}_1 = n+k+1), \tag{6.66}$$

Proof. We will only prove (6.66). With λ as in (6.58) and by use of (6.59) we have

$$P_\pi(\mathcal{T}_1 > n, \lambda = k) = \sum_{i=n+1}^{\infty} P_\pi(\mathcal{T}_1 = i, \lambda = k) = \sum_{i=n+1}^{\infty} \frac{\tau_{i+k}}{E(\tau_-)} = P_\pi(\mathcal{T}_1 = n+k+1), \tag{6.67}$$

and, by the Markov property,

$$E_\pi\left(F_n \middle| \mathcal{T}_1 > n, \lambda = k\right) = E_\kappa\left(F_{n+k} \middle| \mathcal{T}_1 > n+k\right). \tag{6.68}$$

Then

$$\begin{aligned} E_\pi\left(F_n \middle| \mathcal{T}_1 > n\right) &= \frac{1}{P_\pi(\mathcal{T}_1 > n)} E_\pi\left(F_n; \mathcal{T}_1 > n\right) \\ &= \frac{1}{P_\pi(\mathcal{T}_1 > n)} \sum_{k=0}^{\infty} E_\pi\left(F_n; \mathcal{T}_1 > n, \lambda = k\right) \\ &= \frac{1}{P_\pi(\mathcal{T}_1 > n)} \sum_{k=0}^{\infty} E_\pi\left(F_n \middle| \mathcal{T}_1 > n, \lambda = k\right) P_\pi(\mathcal{T}_1 > n, \lambda = k). \end{aligned} \tag{6.69}$$

By inserting (6.67) and (6.68) into (6.69) we obtain (6.66). □

6.5 Numerical application

In the following we calculate key statistics for the pension fund model of Section 6.3, and we illustrate how these statistics are influenced by the bonus threshold (κ) and the investment strategy (C). Based on the results of Section 6.4 we calculate the bonus time distribution, bonus time moments, the number of bonuses, the expected bonus and the expected funding ratio given no bonus. Statistics are calculated for a fund at the bonus threshold and for a fund in stationarity.

In Section 6.5.2 we consider a pension saver paying one monetary unit to the pension fund and receiving 40 years later his terminal pension benefit as a lump sum. Closed form expressions for the pension benefit mean and variance are derived for a fund at the bonus threshold at the time of the contribution. Further, we present in Proposition 6.5.1 an exact simulation algorithm which allows the calculation of the pension benefit mean and variance in stationarity. This is used to find the investment strategy optimizing the expected payout in stationarity for given bonus threshold, i.e. the expected payout for the average saver.

Exact samples from the stationary distribution can also be obtained by the algorithm of Ensor and Glynn (2000). Their algorithm uses exponential tilting and requires exponential moments of the innovation distribution to generate independent, identically distributed samples. In contrast, our algorithm generates partly dependent, identically distributed samples with no distributional assumptions.

We assume throughout that bonus is attributed (possibly) once a year ($\Delta = 1$), and we use the following capital market parameters $r = 3\%$, $\mu = 4\%$, and $\sigma_Z = 15\%$.

6.5.1 Characterization

As a base case example we choose $C = 1.5$ and $\kappa = 1.5$. Thus when the fund is at the bonus threshold $2/3$ of contributions are guaranteed the risk-free rate r. For contributions committed to the fund at lower funding ratios, i.e. in periods between bonuses, a larger fraction is guaranteed.

Note that since C is larger than one the bonus potential is leveraged, i.e. the amount invested in stocks is larger than the bonus potential. At the bonus threshold the fraction of total assets invested in stocks is given by $(1-(1/\kappa))C$. For a base case fund at the bonus threshold a fraction of $1/2$ of total assets are invested in stocks.

Stationarity

Only investment strategies which give rise to a stationary funding ratio process are considered viable options for a collective pension fund. Otherwise the funding ratio will (essentially) converge to one implying that all assets are invested in the risk-free asset only or, equivalently, that all contributions are fully guaranteed. Since one of the purposes of entering a collective fund is to get access to the capital market in a cost-effective way, the latter situation defies the purpose of an investment collective.[6]

Proposition 6.4.1 provides an upper bound on the investment strategy, C, for a stationary distribution to exist. The bound depends on the capital market parameters only, and neither on the threshold (κ) nor the frequency of possible bonus attributions (Δ). For the capital market parameters stated above the fund admits a stationary distribution if and only if C is at most 3.56. Hence, with $C = 1.5$ the base case fund is stationary.

For higher values of σ_Z and/or smaller values of μ the upper bound is appreciably smaller. For the higher, but not unrealistic, volatility of $\sigma_Z = 20\%$ and with the same risk-premium of $\mu = 4\%$ the upper bound on C is 2. For a base case fund at the threshold the bonus potential constitutes one third of total assets. In this case, a bound of 2 implies that the fund can invest at most two thirds of its assets in stocks. Thus, the stationarity requirement can impose material constraints on the investment strategy. Figure 6.1 illustrates the upper bound on C for different market parameter sets.

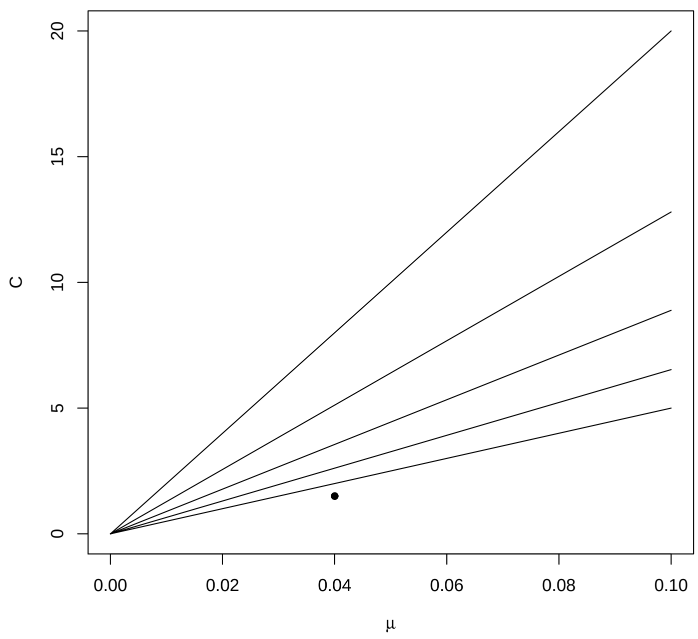

Figure 6.1: Upper bound on the fraction of the bonus potential invested in risky assets (C) for a stationary funding ratio process to exist. The bound is shown as a function of expected excess return of the risky asset (μ) for selected values of volatility (σ_Z). From highest to lowest the bounds correspond to $\sigma_Z = 10\%$, 12.5%, 15%, 17.5% and 20%. The dot indicates the base case values of $(\mu, C) = (4\%, 1.5)$.

[6]The other main purpose of a collective fund is the ability to provide lifelong benefit streams through "diversification" of the individual member's time of death. However, in the current paper we do not consider this aspect.

Bonus times

The first bonus time, $\mathcal{T}_1$, measures the time of the first bonus after time zero. For a fund initially at the bonus threshold the distribution of $\mathcal{T}_1$ can be identified as the distribution of τ_-, i.e. the entrance time to $(-\infty, 0]$ of the associated random walk. For a fund in stationarity, however, it typically takes longer before the first bonus is attributed since the funding ratio at time zero is often below the bonus threshold. Once the first bonus is attributed, the waiting time between all subsequent bonuses is distributed as τ_- regardless of the funding ratio at time zero.

It turns out, perhaps somewhat surprising, that for a fund starting out either at the bonus threshold or in stationary the distribution of $\mathcal{T}_1$ depends only on the investment strategy, and not on the bonus threshold.[7] The top left plot of Figure 6.2 shows the distribution of $\mathcal{T}_1$ in these two cases as given by Proposition 6.4.2. For a base case fund at the bonus threshold the probability of bonus first year is over fifty percent, while the same probability in stationarity is only about twenty percent. The large difference between these values implies that in stationarity the fund is typically "between bonuses".

The probability of bonus first year as a function of the investment strategy is shown in the bottom left plot of Figure 6.2. For a fund at the bonus threshold, the probability is over fifty percent for all considered investment strategies albeit decreasing in C. The stationary probability on the other hand tends to zero as C approaches the upper bound for stationarity of 3.56. Recall that the stationarity probability also has the interpretation as the long-term average, i.e. the frequency with which bonus will be attributed over long horizons (regardless of the initial funding status).

Moments of the time between bonuses can be calculated by Theorem 6.2.6. The mean and standard deviation for various values of C are shown in Table 6.2; as mentioned above the distribution and hence the moments depend only on the investment strategy. In the base case the mean is five years, but with a standard deviation of almost fourteen years. Thus there is considerable variability in the length of the periods between bonuses. For larger values of C the mean and, in particular, the standard deviation increase. A fund with $C = 3$ is still stationary but there will be decades, and even centuries, with no bonus. However, by Proposition 6.4.1 we have that the funding ratio admits a stationary distribution iff the median return on the bonus potential is positive (for a base case fund that is C<3.56). In other words, the stationarity requirement imposes a median time between bonuses[8] of 1. We conclude that the distribution of the waiting times between bonuses is heavily right skewed.

C	1	1.5	2	2.5	3
$E(\tau_-)$	4.12	5.02	6.49	9.35	17.39
$SD(\tau_-)$	9.87	13.73	20.93	37.55	98.60

Table 6.2: The mean and standard deviation of time between bonuses for different investment strategies C. In all cases the median is 1.

Number of bonuses

The number of bonuses from time 1 to time n is denoted N_n. The distribution of N_n can be calculated from the distributions of $\mathcal{T}_1$ and τ_- as described in Section 6.4.3. For a fund starting out either at the bonus threshold or in stationary the distribution of N_n depends only on the investment strategy, and not on the bonus threshold.

The top right plot of Figure 6.2 shows the distribution of N_{40} for a base case fund. For a base case fund starting out at the bonus threshold the distribution of N_{40} is unimodal and the

[7] However, for a fund starting out at a given funding ratio (below the bonus threshold) the distribution of $\mathcal{T}_1$ does of course depend on the bonus threshold also.

[8] Defined as $\min\{n \in \mathbb{N} : \sum_{i=1}^{n} \tau_i \geq 0.5\}$

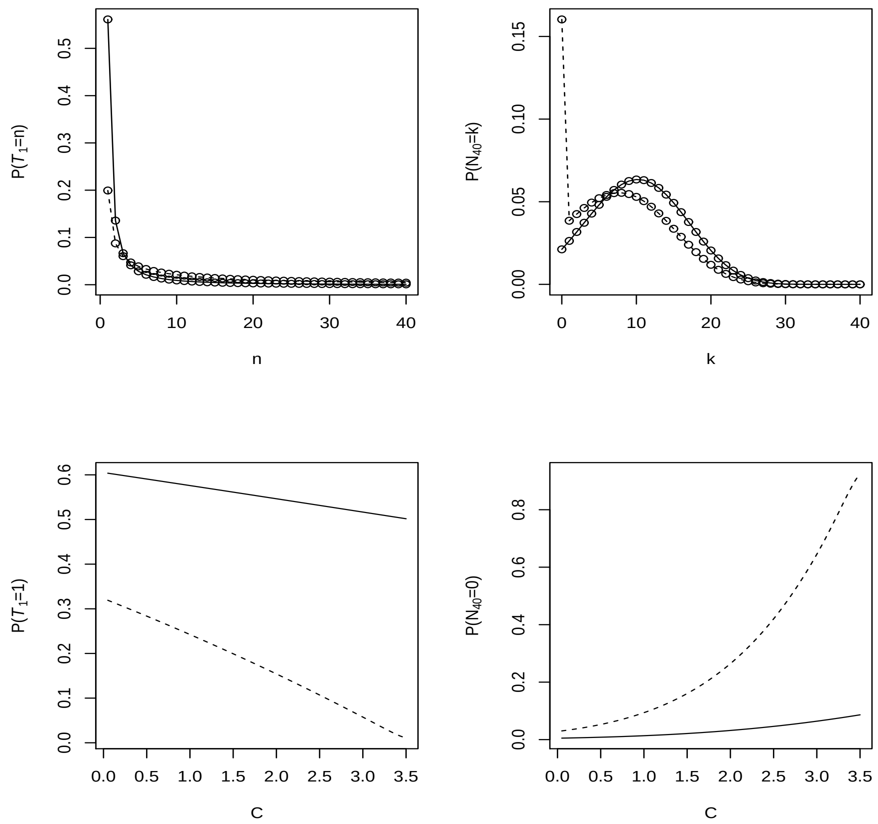

Figure 6.2: Solid curves correspond to $F_0 = \kappa$ and dashed curves to the stationary case, $F_0 \sim \pi$. For the upper plots $C = 1.5$. Upper left: Distribution of first bonus. Lower left: Probability of bonus first year as a function of the investment strategy C. Upper right: Distribution of number of bonuses in 40 years. Lower right: Probability of no bonuses in 40 years as a function of the investment strategy C.

number of bonuses will most likely be around 10. In the stationarity case, however, there is an additional peak at zero. There is thus a rather large probability of about fifteen percent of no bonus at all in forty years, corresponding to the events where the fund is initially at a (very) low funding level. If bonus is attributed at least once the fund evolves like a fund started at the bonus threshold for the remaining period. The part of the stationary distribution of N_{40} at one and above therefore looks like a scaled and shifted version of the "threshold" distribution.

The probability of no bonuses in forty years as function of the investment strategy is depicted in the bottom right plot of Figure 6.2. For increasing C, the probability is modestly increasing for a fund at the bonus threshold, while the stationary probability increases to one.

Bonus percentage

In contrast to the time of bonus, the bonus percentage (conditioned on bonus being given) depends on both the investment strategy and the bonus threshold. It also depends on the time since last bonus.

The bonus percentage is determined by the funding ratio distribution the year prior; the higher the funding ratio the higher the conditional expected bonus.[9] If a fund starting at the bonus threshold does not give bonus in the first year the funding ratio will be strictly below the bonus threshold. This implies that bonus in the second year (if given) is smaller than bonus in the first year (if given). This argument is hard to continue formally, but it seems at least intuitively reasonable that the expected bonus will be decreasing in the time since last bonus.

The bonus percentage conditioned on the value of $\mathcal{T}_1$ can be calculated by Proposition 6.4.3. It is shown in the top left plot of Figure 6.3 for the base case fund. We see that the conditional expected bonus quickly drops by one to two percentage points depending on how the fund is started, and then levels off to just below 5.5%.

The middle left plot of Figure 6.3 shows the expected value of the first bonus,

$$E(r^B_{\mathcal{T}_1}) = \sum_{n=1}^{\infty} E(r^B_n|\mathcal{T}_1 = n)P(\mathcal{T}_1 = n), \tag{6.70}$$

for the base case fund. The expected bonus (when given) is increasing in C, both in stationarity and for a fund starting at the bonus threshold. In stationarity, however, the frequency with which bonus is attributed decreases with C, cf. lower left plot of Figure 6.2. The average bonus in stationarity is therefore a trade-off between many, small bonuses and few, large bonuses. The average bonus in stationarity,

$$E_\pi(r^B) = E_\pi(r^B_1|\mathcal{T}_1 = 1)P_\pi(\mathcal{T}_1 = 1), \tag{6.71}$$

is shown in the lower left plot of Figure 6.3. It is seen that the long-term average bonus is maximized for C just below 2.

It follows from Proposition 6.4.3 that the average bonus for funds with the same C but different bonus thresholds are linearly related. Specifically, the average bonus in stationarity is related by

$$E^{(C,\kappa_2)}_\pi(r^B) = \frac{\kappa_2 - 1}{\kappa_2}\frac{\kappa_1}{\kappa_1 - 1}E^{(C,\kappa_1)}_\pi(r^B). \tag{6.72}$$

This implies that plots for different thresholds are scaled versions of each other. In particular, the average bonus in stationarity is maximized for the same C. It also follows that the average bonus for a fund with $\kappa = 3$ is twice as high as the average bonus for a fund with $\kappa = 1.5$ (base case value). Of course, the guaranteed part to which bonus is applied is correspondingly smaller.

Funding ratio

The pension payoff depends on the funding ratio when contributions are committed, the bonuses up to the time of payout, and the funding ratio at payout. Thus to evaluate the payoff we need to consider both the funding ratio at the time money enters the fund, and the funding ratio at the time money leaves the fund.

The expected funding ratio as a function of the time since last bonus can be calculated by Proposition 6.4.4. This is shown for the base case fund in the upper right plot of Figure 6.3. The expected funding ratio is decreasing in the time since last bonus. Intuitively, this seems reasonable since absence of bonus indicates that the fund is experiencing poor investment results. It is perhaps surprising, however, that the (expected) funding ratio seems to level off, at around 120%. Thus beyond a certain point the funding ratio does not deteriorate any further. Limiting distributions of Markov chains conditioned on non-absorption (or in our case no bonus) are known as Yaglom limits. We conjecture that the fund possesses a Yaglom limit both in stationarity and when started at the threshold, i.e. that the funding ratio distribution conditioned on no bonus converges to a non-degenerate distribution. However, establishing existence, let alone identifying, Yaglom limits is non-trivial and a formal study is outside the scope of this paper.

[9] This in fact is not obvious, since we condition on bonus being given. It can nevertheless be shown as a consequence of stochastic ordering and log-concavity of the normal distribution.

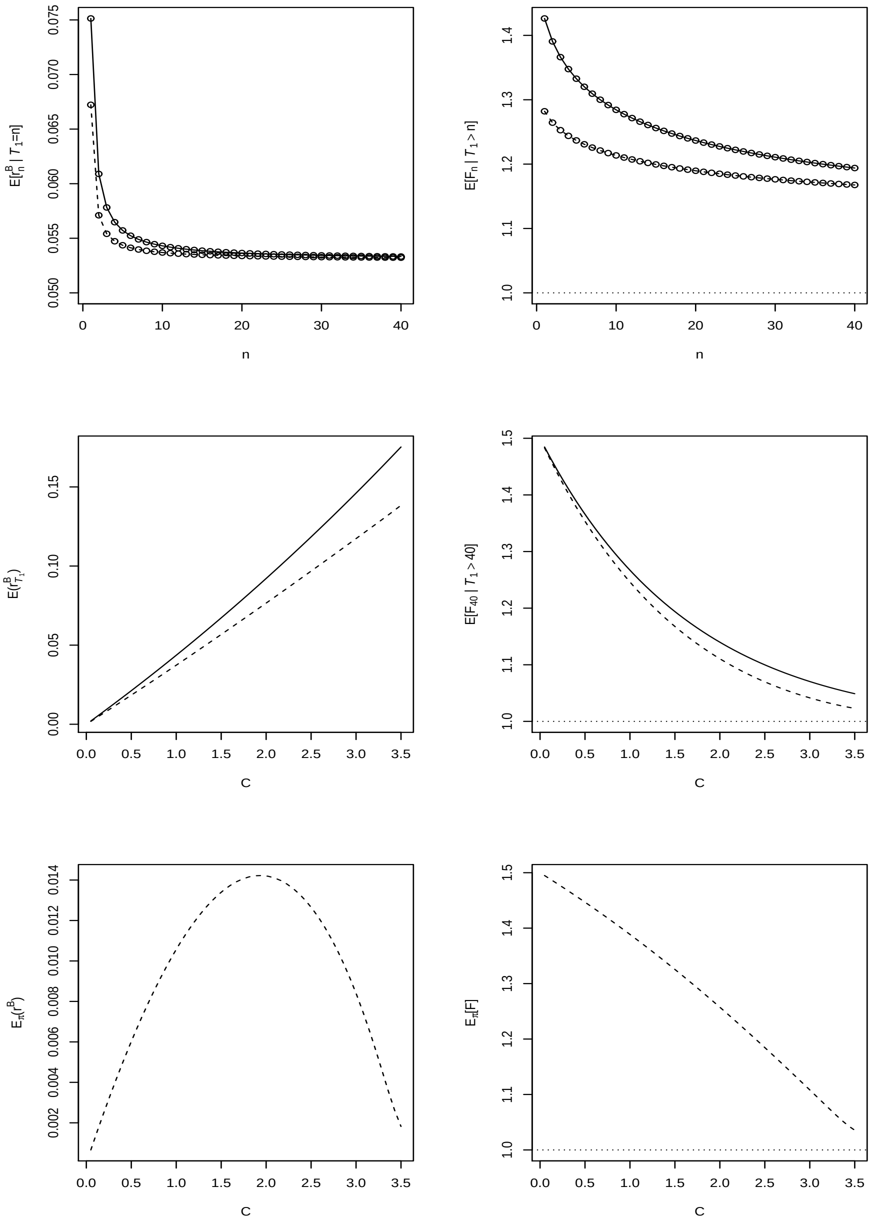

Figure 6.3: Solid curves correspond to $F_0 = \kappa$ and dashed curves to the stationary case, $F_0 \sim \pi$. Bonus threshold $\kappa = 1.5$ in all plots. Upper left: Conditional expected bonus percentage ($C = 1.5$). Middle left: Expected first bonus. Lower left: Expected bonus in stationarity. Upper right: Conditional expected funding ratio given no bonus ($C = 1.5$). Middle right: Conditional expected funding ratio given no bonus in 40 years. Lower right: Expected funding ratio in stationarity.

The interested reader is referred to the specialist literature on quasi-stationarity, e.g. Tweedie (1974); Jacka and Roberts (1995); Lasserre and Pearce (2001).[10]

[10]The partial result that the funding ratio given no bonus does not converge to one can be obtained without too

We know that high equity exposures lead to high, but infrequent, bonuses. It also leads to low expected funding ratios. The expected funding ratio in stationarity,

$$E_\pi(F) = \kappa P_\pi(\mathcal{T}_1 = 1) + E(F_1|\mathcal{T}_1 = 1)P_\pi(\mathcal{T}_1 > 1), \tag{6.73}$$

is shown in the lower right plot of Figure 6.3, while the expected funding ratio after 40 years with no bonus is shown in the middle right plot. For low values of C both the unconditional and conditional expected funding ratio is close to the maximum of 1.5, while as C approaches the upper limit for stationarity the (expected) funding ratio tends to one.

We finally note that it follows from Proposition 6.4.4 that the expected funding ratio for funds with the same C but different thresholds are related by an affine transformation. Specifically, the expected funding ratio in stationarity is related by

$$E_\pi^{(C,\kappa_2)}(F) = \frac{\kappa_2 - 1}{\kappa_1 - 1}\left(E_\pi^{(C,\kappa_1)}(F) - 1\right) + 1. \tag{6.74}$$

Higher bonus thresholds thereby lead to higher expected funding ratios and hence lower guarantees.

6.5.2 Pension benefits

The rationale behind guaranteeing a fixed return on a part of the contributions is that it ensures a certain minimum benefit. However, guarantees reduce the risk capacity for risky assets impairing expected returns. Expected returns can be increased by leverage of the bonus potential, but this in turn increases the variability. The (minimum) fraction guaranteed and the expected return/variability are controlled by κ and C, respectively.

In this section we calculate the mean and variance of the payout considered in Section 6.3 (repeated here for ease of reference),

$$O_T = \frac{F_T}{F_0} e^{rT} \prod_{i=1}^{T} \left(1 + r_i^B\right). \tag{6.75}$$

This is used to calculate the optimal C for a mean-variance criterion for given κ. We take the perspective of the "average" member and we therefore perform the optimization for a fund in stationarity. The analysis proceeds in two steps. First, we calculate the mean and variance of O_T for a fund starting at the threshold. Second, based on these results we apply an exact simulation algorithm to find the stationary mean and variance. The algorithm is considerably simpler than existing algorithms and might be of independent interest.

Fund at bonus threshold

The first and second order moment of O_T (and thereby the variance) can be calculated by a so-called last-exit decomposition, see e.g. Meyn and Tweedie (2009) p. 178. Let $\mathcal{U}$ denote the last time bonus was given before and including time T. Note that $\mathcal{U}$ is not a stopping time, and

much effort. Loosely speaking, it follows since the bonus waiting time has exponential moments, cf. footnote 5, and since bonus can only be attributed when the funding ratio is "close" to κ the year prior. Consequently, the probability that the funding ratio is above a certain level can be bounded away from zero at least along a subsequence. Essentially the same conditions (exponential moments of time to absorbtion and increasing absorbtion times from states far away) are used in Ferrari et al. (1995) to show the existence of a quasi-stationary distribution for a continuous-time Markov chain on a discrete state space. Also note Martinez et al. (1998) which studies quasi-stationarity of a Brownian motion conditioned to stay positive. In our setup this can be seen as the limiting case where bonus is attributed continuously ($\Delta \approx 0$).

that the decomposition is not a consequence of the Markov property. We have

$$\begin{aligned}E_\kappa(O_T) =& E_\kappa(O_T;\mathcal{T}_1 > T) + \sum_{j=1}^{T} E_\kappa(O_T;\mathcal{U}=j)\\ =& \frac{e^{rT}}{\kappa}e_\kappa(T) + \frac{e^{rT}}{\kappa}\sum_{j=1}^{T}\left[\sum_{\sigma\in\mathcal{D}_j}\widetilde{c}_\sigma\prod_{i=1}^{j}\left(\tau_i E_\kappa\left(1+r_i^B\middle|\mathcal{T}_1=i\right)\right)^{\sigma_i}\right]e_\kappa(T-j), \qquad (6.76)\end{aligned}$$

where for $\sigma \in \mathcal{D}_j$ and $n = 0,\dots,T$,

$$\tilde{c}_\sigma = \frac{(\sum_{i=1}^{j}\sigma_i)!}{\prod_{i=1}^{j}\sigma_i!}, \quad e_\kappa(n) = P_\kappa(\mathcal{T}_1 > n)E_\kappa(F_n\big|\mathcal{T}_1 > n). \qquad (6.77)$$

The expression for $E_\kappa(O_T;\mathcal{U} = j)$ follows by considering the different ways bonus can be attributed such that the last bonus falls at time j. The different patterns of time between bonuses are given by the permutations in $\mathcal{D}_j$. For each pattern the probability of it occurring and the associated expected bonus can be calculated by the Markov property, and this has to be multiplied by the number of ways the "bonus waiting periods" can be arranged, given by $\tilde{c}_\sigma$. Finally, we multiply by the expected funding ratio given that no bonuses are given for the remaining period, given by $e_\kappa(T-j)$. The quantities appearing in expressions (6.76) and (6.77) can be calculated by Propositions 6.4.2–6.4.4.

For the second order moment we similarly find

$$\begin{aligned}E_\kappa(O_T^2) =& E_\kappa(O_T^2;\mathcal{T}_1 > T) + \sum_{j=1}^{T} E_\kappa(O_T^2;\mathcal{U}=j)\\ =& \frac{e^{2rT}}{\kappa^2}s_\kappa(T) + \frac{e^{2rT}}{\kappa^2}\sum_{j=1}^{T}\left[\sum_{\sigma\in\mathcal{D}_j}\widetilde{c}_\sigma\prod_{i=1}^{j}\left(\tau_i E_\kappa\left(\left(1+r_i^B\right)^2\middle|\mathcal{T}_1=i\right)\right)^{\sigma_i}\right]s_\kappa(T-j), \quad (6.78)\end{aligned}$$

where $\tilde{c}_\sigma$ is given by (6.77) and for $n = 0,\dots,T$,

$$s_\kappa(n) = P_\kappa(\mathcal{T}_1 > n)E_\kappa(F_n^2\big|\mathcal{T}_1 > n). \qquad (6.79)$$

In order to calculate $s_\kappa(k)$ note that on the event $(\mathcal{T}_1 > n)$

$$F_n^2 = \left((\kappa-1)e^{-S_n}+1\right)^2 = (\kappa-1)^2e^{-2S_n} + 2(\kappa-1)e^{-S_n} + 1. \qquad (6.80)$$

Hence, we need to calculate conditional expectations of e^{-S_n} and e^{-2S_n}. The first of these is given by (6.64) of Proposition 6.4.4. The latter can be calculated by the same formula upon replacing the term $E(e^{-S_n}; S_n > 0)$ with $E(e^{-2S_n}; S_n > 0)$. To evaluate (6.78) we also need to calculate the second order moment of the bonus percentage. Similarly to F_n^2 the term $(1 + r_i^B)^2$ can be expanded and expressed in terms of e^{-S_i} and e^{-2S_i}. The appropriate conditional expectations of the latter quantities can be calculated by formula (6.55) of Proposition 6.4.3. We omit the details.

Fund in stationarity

We are interested in calculating the stationary mean and variance of O_T. From Propositions 6.4.2–6.4.4 we know the time and size of first bonus, moments of the initial funding ratio and moments of the terminal funding ratio conditioned on no bonuses. Unfortunately, we need the joint distribution of these quantities and this is not available in an analytically tractable form. Instead we will apply a simulation algorithm based on samples from the joint stationary distribution of $(F_0, \mathcal{T}_1, r^B_{\mathcal{T}_1}, F_{\mathcal{T}_1\wedge T})$. The idea is to split the period in two, the time up to first bonus (if it occurs before time T) and the time after first bonus. Moments of O_T can be obtained by combining samples for the period up to first bonus with analytic results for the period after first bonus.

Samples from the joint distribution of $(F_0, \mathcal{T}_1, r^B_{\mathcal{T}_1}, F_{\mathcal{T}_1 \wedge T})$ can be obtained in several ways. Perhaps the most obvious is to simulate F_0 from the stationarity funding ratio distribution. Given F_0 we can then simulate the evolution of the fund until first bonus and record the time and size of the bonus. Ensor and Glynn (2000) give an algorithm which can be used to obtain exact samples of F_0, and Preisel et al. (2010) give an alternative algorithm by which F_0 can be sampled to any desired level of accuracy. Both algorithms rely on the fact that the invariant distribution of a one-sided random walk (the Y-chain of Section 6.4) equals the distribution of the maximum of the associated unrestricted random walk (the S-chain of Section 6.4).

We employ a different idea based on the argument presented in Section 6.4.2. In stationarity the probability that at time 0 we are in a period between two bonuses of length k is proportional to $k\tau_k$. Further, given we are in a period of length k at time 0 it is equally likely that we are in any of the k positions. This observation gives rise to a very simple algorithm for simulating in stationarity: For a fund started *at the threshold* simulate the path up to first bonus, say, at time k. This happens with probability τ_k. Now, use this path to generate k samples by shifting it n places to the left for $n = 0, \ldots, k-1$. Repeat the algorithm to obtain more samples.

The algorithm generates partly dependent samples from the stationary distribution. However, when used to estimate expectations with respect to the stationary distribution the dependence does not pose a problem. For evaluating moments of O_T in stationarity we propose the following method. The method combines the (exact) samples with the analytic results obtained previously to obtain a consistent estimate of $E_\pi(O_T)$. Estimates of second, and higher, order moments of O_T are obtained by suitable modifications of the G_T-functional.

Proposition 6.5.1. *Let N be given. Starting at the bonus threshold simulate N paths until first bonus. Denote the funding ratio paths by $(F_0^{(i)}, \ldots, F^{(i)}_{\mathcal{T}_1^{(i)}-1})$ and the first bonus by $r^{(i)}$ for $i = 1, \ldots, N$. Let $M = \sum_{i=1}^N \mathcal{T}_1^{(i)}$. A consistent estimate of $E_\pi(O_T)$ can be obtained by*

$$\hat{E}_\pi(O_T) = \frac{1}{M} \sum_{i=1}^{N} \sum_{k=0}^{\mathcal{T}_1^{(i)}-1} G_T\left(F_k^{(i)}, \mathcal{T}_1^{(i)} - k, r^{(i)}, F^{(i)}_{\mathcal{T}_1^{(i)} \wedge (T+k)}\right), \tag{6.81}$$

where

$$G_T(F_0, \mathcal{T}_1, r^B, F_{\mathcal{T}_1 \wedge T}) = \frac{F_{\mathcal{T}_1 \wedge T}}{F_0} \times \begin{cases} e^{r\mathcal{T}_1}(1 + r^B) E_\kappa(O_{T-\mathcal{T}_1}) & \text{for } \mathcal{T}_1 \le T, \\ e^{rT} & \text{for } \mathcal{T}_1 > T, \end{cases} \tag{6.82}$$

and $O_0 = 0$ by convention.

Note that the two cases in G_T correspond to whether or not the first bonus (in the shifted path) occurs before or after time T. Also note that bonus occurring before or at time T for the k-shifted path is equivalent to $\mathcal{T}_1^{(i)} \ge T + k$, and in this case the last argument of G_T equals κ.

Although Proposition 6.5.1 is presented as a method for estimating a specific quantity the same method can be used to estimate any stationary expectation. The estimator can also be made unbiased by replacing M by its expectation, $NE_\kappa(\mathcal{T}_1)$. As a simple example, stationary probabilities for F_0 can be estimated unbiasedly by

$$\hat{P}_\pi(F_0 \in A) = \frac{1}{NE_\kappa(\mathcal{T}_1)} \sum_{i=1}^{N} \sum_{k=0}^{\mathcal{T}_1^{(i)}-1} 1_A\left(F_k^{(i)}\right) \tag{6.83}$$

for any event A.[11]

[11] Remark: Theorem 10.4.9 of Meyn and Tweedie (2009) together with Kac's theorem yields the representation

Mean-variance analysis

We are now in a position to do a mean-variance analysis of the payout O_T with respect to the strategy parameters κ and C. We assume a horizon of $T = 40$ years. This corresponds approximately to the average time between when contributions are made and benefits are paid out in a pension fund with life-long memberships.

We first consider a fund starting at the bonus threshold. Table 6.3 states the mean, the minimum (guarantee) and the standard deviation of the payout for different sets of (κ, C). In all cases the mean equals 6. For the smallest value of κ the payout is guaranteed to be at least 2.656. To reach an expected payout of 6 the bonus potential has to be leveraged substantially, and this in turn leads to a large standard deviation. As κ increases the guarantee decreases and an expected payout of 6 can be achieved by investing the bonus potential less aggressively. This reduces the standard deviation at the price of a larger downside risk (lower guarantee). In the limit as κ tends to infinity there is no guarantee and an expected payout of 6 can be achieved with a standard deviation of 2.17 by investing 37 percent of total assets in stocks.[12]

κ	1.25	1.5	2	3	5	10
C	2.705	1.259	0.782	0.570	0.468	0.413
mean	6	6	6	6	6	6
guarantee	2.656	2.213	1.660	1.107	0.664	0.332
std. dev.	3.662	2.603	2.356	2.256	2.214	2.191

Table 6.3: Mean, guarantee and standard deviation of O_{40} for a pension fund starting at the bonus threshold. All sets of strategies (κ, C) imply a mean payout of 6.

Consider next a pension fund which has fixed the bonus threshold at κ. This implies that at least $1/\kappa$ of contributions is guaranteed the risk free rate. The value of the threshold might be stipulated by regulation to ensure a certain minimum pension, or it might be decided by the board of the pension fund based on social economic considerations. In either case, the fund needs to determine an investment strategy C. One (common) way to balance the desire for a high payout against unwanted variability is by use of a mean-variance optimization criterion. Being a collective pension fund we want to optimize the fund for the benefit of the average member, i.e. in stationarity. Hence, we consider the following stationary mean-variance problem for fixed κ

$$\sup_C \{E_\pi(O_T) - \gamma \mathrm{Var}_\pi(O_T)\}. \tag{6.84}$$

In Figure 6.4 the optimization problem with $\gamma = 0.07468$ and $T = 40$ is illustrated for $\kappa = 1.5$ and $\kappa = 3$. The value of γ is chosen such that 60 percent in stocks is optimal for a mean-variance investor with no guarantee and a constant proportion of total assets in stocks. We see from Figure 6.4 that neither the mean nor the standard deviation is monotone in C. The expected payout is decreasing for C sufficiently large because very aggressive strategies lead to low funding ratios in stationarity. The fund with low guarantees ($\kappa = 3$) has an optimal C of about 80 percent, while the fund with high guarantees ($\kappa = 1.5$) has an optimal C of about 150 percent. The equity exposure, at the threshold, as a fraction of total assets is about 50 percent for both cases.

The mean-variance criterion is normally applied in situations where more risk (higher C) leads to higher expected return and higher variability. It might be argued that the mean-variance criterion is considered out of necessity to avoid degenerate solutions in these situations.

result $P_\pi(F_0 \in A) = E_\kappa(\sum_{k=0}^{\mathcal{T}_1 - 1} 1_A(F_k))/E_\kappa(\mathcal{T}_1)$. This result can also be obtained from (6.83) by taking expectation on both sides. Conversely, it follows from the representation that the right-hand side of (6.83) is an unbiased estimator of $P_\pi(F_0 \in A)$ as claimed.

[12] For the optimal mean-variance investment strategy see Korn (1997a) and Zhou and Li (2000).

In stationarity, however, the expected payout as a function of C is unimodal. We might therefore alternatively define the optimal investment strategy C^* as the one maximizing the expected payout in stationarity. The optimal investment strategy C^* thus defined is illustrated in Table 6.4 and Figure 6.4 upper left for different values of κ. We see that C^* is considerably higher than the one obtained from the mean-variance criterion (for given κ). However, the associated standard deviation is also considerably higher implying than the payout will typically deviate substantially from its expectation.

κ	1	1.25	1.5	2	3	5	10
C^*	N/A	2.143	2.313	2.473	2.700	2.850	2.951
maximal mean	3.320	4.923	6.886	11.73	23.66	48.50	93.61
std. dev.	0	2.213	6.649	26.13	151.5	826.9	3540

Table 6.4: The optimal investment strategy, C^*, maximizing the expectation of O_{40} in stationarity for different values of κ. The corresponding mean and standard deviation of O_{40} are also shown.

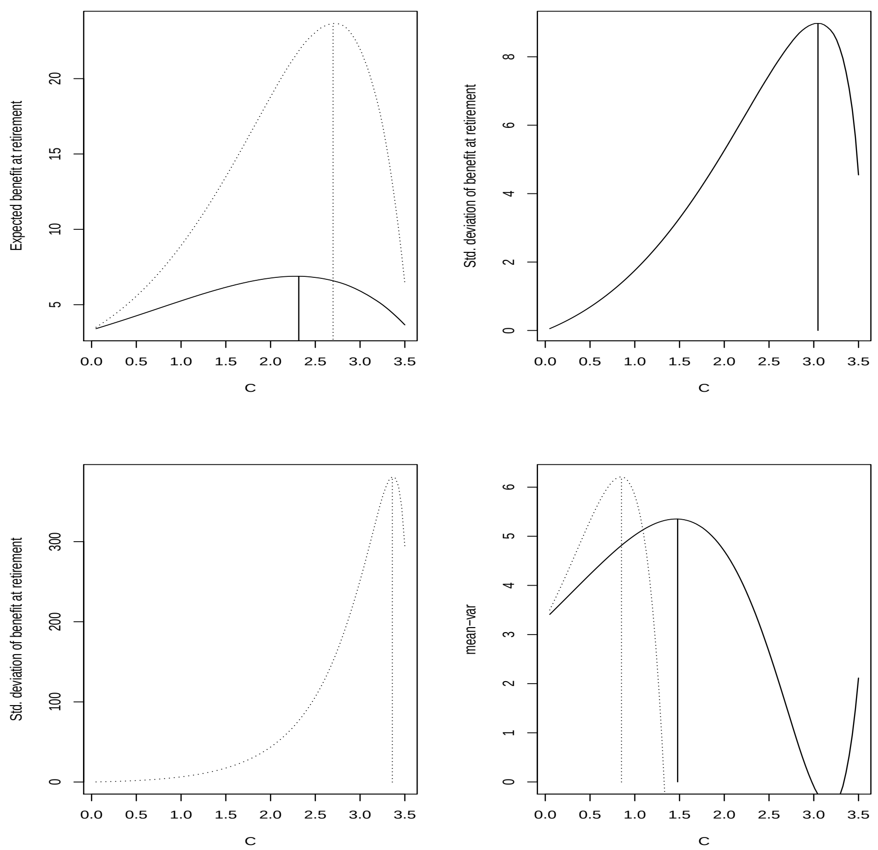

Figure 6.4: Stationary mean and standard deviation of O_{40} as a function of the investment strategy C. Solid curves correspond to $\kappa = 1.5$ and dotted curves to $\kappa = 3$. Upper left: Expected pension payout. Upper right and lower left: Standard deviation of the pension payout (note the different scales). Lower right: Illustration of the mean-variance optimization problem (6.84).

6.6 Appendix (proofs)

6.6.1 Entrance times and partitions

Proof of Theorem 6.2.3. Rearranging (6.5) we get for $0 \le s < 1$

$$1 - \exp\left(\sum_{n=1}^{\infty}(-1)\frac{p_n}{n}s^n\right) = \sum_{n=1}^{\infty}\tau_n s^n. \tag{6.85}$$

From Theorem 6.2.2 we get that

$$\begin{aligned}\exp\left(\sum_{n=1}^{\infty}(-1)\frac{p_n}{n}s^n\right) &= 1 + \sum_{m=1}^{\infty}\sum_{\sigma\in\mathcal{D}_m}\frac{s^m}{\sigma_1!\sigma_2!\cdots\sigma_m!}\prod_{n=1}^{m}\left((-1)\frac{p_n}{n}\right)^{\sigma_n}\\ &= 1 + \sum_{m=1}^{\infty}\sum_{\sigma\in\mathcal{D}_m}\frac{\operatorname{sgn}(\sigma)}{d_\sigma}s^m\prod_{n=1}^{m}p_n^{\sigma_n}. \end{aligned}\tag{6.86}$$

Insert this in (6.85) to obtain

$$-\sum_{m=1}^{\infty}\sum_{\sigma\in\mathcal{D}_m}\frac{\operatorname{sgn}(\sigma)}{d_\sigma}s^m\prod_{n=1}^{m}p_n^{\sigma_n} = \sum_{n=1}^{\infty}\tau_n s^n. \tag{6.87}$$

Finally, by inspection of the left and right hand side of the above equality we obtain (6.12). □

Proof of Theorem 6.2.4. We need to prove two identities

$$\sum_{\sigma\in\mathcal{D}_n}\frac{k\sigma_k}{d_\sigma} = 1, \tag{6.88}$$

and

$$\sum_{\sigma\in\mathcal{D}_n}\frac{1}{d_\sigma} = 1. \tag{6.89}$$

We first note that (6.89) follows from (6.88) by summing over k. Using that $\sum_{k=1}^{n}k\sigma_k = n$ for $\sigma\in\mathcal{D}_n$ we get

$$n = \sum_{k=1}^{n}\sum_{\sigma\in\mathcal{D}_n}\frac{k\sigma_k}{d_\sigma} = \sum_{\sigma\in\mathcal{D}_n}\frac{\sum_{k=1}^{n}k\sigma_k}{d_\sigma} = \sum_{\sigma\in\mathcal{D}_n}\frac{n}{d_\sigma}, \tag{6.90}$$

and dividing through by n yields (6.89).

We next prove (6.88) by induction in n. For $n = 1$ the relation is trivially satisfied. Assume that (6.88) holds for $n-1$ and all $1 \le k \le n-1$; then (6.89) also holds for $n-1$ as just shown. To prove that (6.88) holds for n and $1 \le k \le n$ there are three cases to consider:

For $k = n$ the only term in the sum different from zero occurs for $\sigma = (0,\dots,0,1)$, and hence

$$\sum_{\sigma\in\mathcal{D}_n}\frac{n\sigma_n}{\prod_{i=1}^{n}\sigma_i!i^{\sigma_i}} = \frac{n}{1!n} = 1. \tag{6.91}$$

For $1 < k < n$ we use that there is a one-to-one mapping between permutations in $\mathcal{D}_n$ with $\sigma_k > 0$ and permutations in $\mathcal{D}_{n-1}$ with $\pi_{k-1} > 0$ defined by $\pi = (\sigma_1,\dots,\sigma_{k-2},\sigma_{k-1}+1,\sigma_k -$

$1, \sigma_{k+1}, \ldots, \sigma_{k-1})$. Noting that $\sigma_k > 0$ with $k < n$ implies $\sigma_n = 0$ we have

$$\begin{aligned}
&\sum_{\sigma \in \mathcal{D}_n} \frac{k\sigma_k}{\prod_{i=1}^{n} \sigma_i! i^{\sigma_i}} \\
&= \sum_{\sigma \in \mathcal{D}_n : \sigma_k > 0} \frac{k\sigma_k}{\prod_{i=1}^{n-1} \sigma_i! i^{\sigma_i}} \\
&= \sum_{\pi \in \mathcal{D}_{n-1} : \pi_{k-1} > 0} \frac{k(\pi_k + 1)}{\prod_{i=1}^{n-1} \pi_i! i^{\pi_i}} \frac{\pi_{k-1}!(k-1)^{\pi_{k-1}} \pi_k! k^{\pi_k}}{(\pi_{k-1} - 1)!(k-1)^{\pi_{k-1}-1}(\pi_k + 1)! k^{\pi_k + 1}} \\
&= \sum_{\pi \in \mathcal{D}_{n-1} : \pi_{k-1} > 0} \frac{(k-1)\pi_{k-1}}{\prod_{i=1}^{n-1} \pi_i! i^{\pi_i}} \\
&= 1,
\end{aligned} \tag{6.92}$$

where the last equality follows by the induction hypothesis.

For $k = 1$ the mapping $\pi = (\sigma_1 - 1, \sigma_2, \ldots, \sigma_{n-1})$ is one-to-one between permutations in $\mathcal{D}_n$ with $\sigma_1 > 0$ and *all* permutations in $\mathcal{D}_{n-1}$. We then have

$$\sum_{\sigma \in \mathcal{D}_n} \frac{\sigma_1}{\prod_{i=1}^{n} \sigma_i! i^{\sigma_i}} = \sum_{\pi \in \mathcal{D}_{n-1}} \frac{\pi_1 + 1}{\prod_{i=1}^{n-1} \pi_i! i^{\pi_i}} \frac{\pi_1!}{(\pi_1 + 1)!} = \sum_{\pi \in \mathcal{D}_{n-1}} \frac{1}{\prod_{i=1}^{n-1} \pi_i! i^{\pi_i}} = 1, \tag{6.93}$$

where the last equality follows from (6.89) which holds for $n-1$ by the induction hypothesis. □

6.6.2 Entrance time moments

Proof of Theorem 6.2.5. Let $G_1 = H'$ and define recursively

$$G_n = H' G_{n-1} + G'_{n-1} \quad (n \geq 2). \tag{6.94}$$

We first establish the relation

$$\tau^{(n)} = -e^H G_n \quad (n \geq 1). \tag{6.95}$$

Clearly, the relation holds for $n = 1$. Assuming (6.95) holds for $n - 1$ we have by (6.94)

$$\begin{aligned}
\tau^{(n)} &= \left(-e^H G_{n-1}\right)' \\
&= -e^H \left(H' G_{n-1} + G'_{n-1}\right) \\
&= -e^H G_n,
\end{aligned} \tag{6.96}$$

and hence (6.95) holds for all n by induction.

Having established (6.95) we now need to prove

$$G_n = \sum_{\sigma \in \mathcal{D}_n} c_\sigma H_\sigma \quad (n \geq 1). \tag{6.97}$$

As before we will prove this by induction. For $n = 1$ the equation reads $G_1 = H'$ which is true by definition. Assume (6.97) holds for $n - 1$. By (6.94) and the induction hypothesis we have

$$G_n = \sum_{\sigma \in \mathcal{D}_{n-1}} c_\sigma H' H_\sigma + \sum_{\sigma \in \mathcal{D}_{n-1}} c_\sigma (H_\sigma)', \tag{6.98}$$

where for $\sigma = (\sigma_1, \ldots, \sigma_{n-1}) \in \mathcal{D}_{n-1}$

$$
\begin{aligned}
(H_\sigma)' &= \sigma_1 H_1^{\sigma_1 - 1} H_2^{\sigma_2} \prod_{i=2}^{n-1} H_i^{\sigma_i} + H_1^{\sigma_1} \left(\prod_{i=2}^{n-1} H_i^{\sigma_i} \right)' \\
&= \ldots \\
&= \sum_{k=1}^{n-1} \sigma_k H_k^{\sigma_k - 1} H_{k+1} \prod_{i \neq k} H_i^{\sigma_i}.
\end{aligned} \tag{6.99}
$$

Thus

$$
G_n = \sum_{\sigma \in \mathcal{D}_{n-1}} c_\sigma H' H_\sigma + \sum_{\sigma \in \mathcal{D}_{n-1}} \sum_{k=1}^{n-1} c_\sigma \sigma_k H_k^{\sigma_k - 1} H_{k+1} \prod_{i \neq k} H_i^{\sigma_i}. \tag{6.100}
$$

Let $\sigma^* = (\sigma_1^*, \ldots, \sigma_n^*) \in \mathcal{D}_n$ be fixed, but arbitrary. We will identify the terms in (6.100) which contain the factor H^{σ^*} and show that the coefficients add up to c_{σ^*}. There are three cases (note that in the first two cases the condition implies that $\sigma_n^* = 0$):

1. $\sigma_1^* > 0$: Then $H' H_\sigma = H^{\sigma^*}$ with $\sigma = (\sigma_1^* - 1, \sigma_2^*, \ldots, \sigma_{n-1}^*) \in \mathcal{D}_{n-1}$. The first sum in (6.100) contains a term with this factor and coefficient c_σ.
2. $\sigma_j^* > 0$ for some $2 \leq j \leq n-1$: Then $H_k^{\sigma_k - 1} H_{k+1} \prod_{i \neq k} H_i^{\sigma_i} = H^{\sigma^*}$ with $\sigma = (\sigma_1^*, \ldots, \sigma_{j-2}^*, \sigma_{j-1}^* + 1, \sigma_j^* - 1, \sigma_{j+1}^*, \ldots, \sigma_{n-1}^*) \in \mathcal{D}_{n-1}$ and $k = j - 1$. The second sum in (6.100) contains a term with this factor and coefficient $c_\sigma(\sigma_{j-1}^* + 1)$.
3. $\sigma_n^* > 0$: In this case $\sigma^* = (0, \ldots, 0, 1)$ and hence $H^{\sigma^*} = H_n = H_k^{\sigma_k - 1} H_{k+1} \prod_{i \neq k} H_i^{\sigma_i}$ for $\sigma = (0, \ldots, 0, 1) \in \mathcal{D}_{n-1}$ and $k = n - 1$. This term is included in the second sum of (6.100) with coefficient c_σ.

Hence, for each positive component of σ^* there exists a term in (6.100) containing the factor H^{σ^*}. Conversely, every term in (6.100) corresponds to a $\sigma^* \in \mathcal{D}_n$ and is of the form covered in one of the three cases above. Hence, we have

$$
G_n = \sum_{\sigma^* \in \mathcal{D}_n} H^{\sigma^*} \sum_{1 \leq j \leq n : \sigma_j^* > 0} (\sigma_{j-1}^* + 1) c_{\sigma^{(j)}} \tag{6.101}
$$

where $\sigma^{(j)} = (\sigma_1^*, \ldots, \sigma_{j-2}^*, \sigma_{j-1}^* + 1, \sigma_j^* - 1, \sigma_{j+1}^*, \ldots, \sigma_{n-1}^*) \in \mathcal{D}_{n-1}$ and we, for notational convenience, define $\sigma_0^* \equiv 0$. Using (6.10) we can write the inner sum in (6.101) as

$$
\begin{aligned}
&\sum_{1 \leq j \leq n : \sigma_j^* > 0} (\sigma_{j-1}^* + 1) c_{\sigma^{(j)}} \\
&= \sum_{1 \leq j \leq n : \sigma_j^* > 0} (\sigma_{j-1}^* + 1) \frac{(n-1)!}{\prod_{i=1}^{n-1} \sigma_i^{(j)}! (i!)^{\sigma_i^{(j)}}} \\
&= \frac{1}{\prod_{i=1}^{n} \sigma_i^*! (i!)^{\sigma_i^*}} \sum_{1 \leq j \leq n : \sigma_j^* > 0} (\sigma_{j-1}^* + 1)(n-1)! \frac{\prod_{i=1}^{n} \sigma_i^*! (i!)^{\sigma_i^*}}{\prod_{i=1}^{n-1} \sigma_i^{(j)}! (i!)^{\sigma_i^{(j)}}} \\
&= \frac{1}{\prod_{i=1}^{n} \sigma_i^*! (i!)^{\sigma_i^*}} \sum_{1 \leq j \leq n : \sigma_j^* > 0} (\sigma_{j-1}^* + 1)(n-1)! \frac{\sigma_j^* j!}{(\sigma_{j-1}^* + 1)(j-1)!} \\
&= \frac{1}{\prod_{i=1}^{n} \sigma_i^*! (i!)^{\sigma_i^*}} (n-1)! \sum_{1 \leq j \leq n : \sigma_j^* > 0} \sigma_j^* j \\
&= \frac{n!}{\prod_{i=1}^{n} \sigma_i^*! (i!)^{\sigma_i^*}} \\
&= c_{\sigma^*},
\end{aligned} \tag{6.102}
$$

where the penultimate equality uses that $\sum_{1\le j\le n:\sigma_j^*>0}\sigma_j^* j = n$ since $\sigma^*\in\mathcal{D}_n$. This shows that (6.97) holds and we are finished. $\square$

Before proving Theorem 6.2.6 we need to state the following two lemmas:

Lemma 6.6.1. *For $n\ge 1$, $(a_1,\dots,a_n)\in\mathbb{C}^n$ and $b\in\mathbb{C}$*

$$\sum_{\sigma\in\mathcal{D}_n}\prod_{i=1}^{n}\left(a_i-\frac{b^i}{i}\right)^{\sigma_i}\frac{1}{\sigma_i!}=\sum_{\sigma\in\mathcal{D}_n}\prod_{i=1}^{n}\frac{a_i^{\sigma_i}}{\sigma_i!}-b\sum_{\sigma\in\mathcal{D}_{n-1}}\prod_{i=1}^{n-1}\frac{a_i^{\sigma_i}}{\sigma_i!},\tag{6.103}$$

where the last sum is 1 by definition for $n=1$.

Proof. We first note that the left-hand side of (6.103) can be written

$$\sum_{\sigma\in\mathcal{D}_n}\prod_{i=1}^{n}\frac{a_i^{\sigma_i}}{\sigma_i!}-b\sum_{\sigma\in\mathcal{D}_n}\sigma_1\frac{a_1^{\sigma_1-1}}{\sigma_1!}\prod_{i=2}^{n}\frac{a_i^{\sigma_i}}{\sigma_i!}+\dots,\tag{6.104}$$

where $\dots$ denotes higher order terms of b. Since there is only a contribution to the first order term if $\sigma_1>0$, and since $\sigma\in\mathcal{D}_n$ with $\sigma_1>0$ if and only if $(\sigma_1-1,\sigma_2,\dots,\sigma_{n-1})\in\mathcal{D}_{n-1}$, we have that the first order term is given by

$$b\sum_{\sigma\in\mathcal{D}_n:\sigma_1>0}\sigma_1\frac{a_1^{\sigma_1-1}}{\sigma_1!}\prod_{i=2}^{n}\frac{a_i^{\sigma_i}}{\sigma_i!}=b\sum_{\sigma\in\mathcal{D}_{n-1}}\prod_{i=1}^{n-1}\frac{a_i^{\sigma_i}}{\sigma_i!}.\tag{6.105}$$

Hence we are finished if we show that all higher order terms in (6.104) vanish.

For $m\ge 2$, the m'th order terms are of the form

$$\prod_{i=1}^{n}a_i^{\eta_i}\left(\frac{-b^i}{i}\right)^{\nu_i}\frac{1}{(\eta_i+\nu_i)!}\binom{\eta_i+\nu_i}{\nu_i}=\frac{\operatorname{sgn}(\nu)}{d_\nu}b^m\prod_{i=1}^{n}\frac{a_i^{\eta_i}}{\eta_i!},\tag{6.106}$$

where $\nu\in\mathcal{D}_m$ and $\eta\in\mathcal{D}_{n-m}$, with the convention that $\nu_j=0$ for $j>m$ and, similarly, $\eta_j=0$ for $j>n-m$ (for $m=n$, $\eta_j=0$ for all j). The higher order terms in (6.104) can then be written

$$\sum_{m=2}^{n}b^m\sum_{\eta\in\mathcal{D}_{n-m}}\prod_{i=1}^{n}\frac{a_i^{\eta_i}}{\eta_i!}\sum_{\nu\in\mathcal{D}_m}\frac{\operatorname{sgn}(\nu)}{d_\nu}=0,\tag{6.107}$$

since the inner sum is zero for $m\ge 2$ by equation (6.13). $\square$

We also need the following combination of Kronecker's lemma, see e.g. Theorem 2.5.5 of Durrett (2010), and Frobenius' theorem, see e.g. Chapter 7 of Duren (2012).

Lemma 6.6.2. *If $\sum_{k=1}^{\infty}a_k$ converges to a finite limit then*

$$f(s)=(1-s)\sum_{k=1}^{\infty}s^k k a_k\tag{6.108}$$

converges for $|s|<1$ and $f(s)\to 0$ as $s\to 1-$.

Proof. Let $b_k=k$ and $x_k=ka_k$. Then $b_k\uparrow\infty$ and, by assumption, $\sum_{k=1}^{\infty}x_k/b_k=\sum_{k=1}^{\infty}a_k$ converges. By Kronecker's lemma this implies that

$$\frac{1}{b_n}\sum_{k=1}^{n}x_k=\frac{1}{n}\sum_{k=1}^{n}ka_k\to 0 \text{ as } n\to\infty.\tag{6.109}$$

Next, let $a_0 = 0$ and $c_k = ka_k - (k-1)a_{k-1}$ for $k \geq 1$. Summation by parts yields

$$f(s) = (1-s)\sum_{k=1}^{\infty} s^k k a_k = \sum_{k=1}^{\infty} s^k (ka_k - (k-1)a_{k-1}) = \sum_{k=1}^{\infty} s^k c_k, \tag{6.110}$$

where the series $\sum_{k=1}^{\infty} c_k$ is Cesàro summable to zero by (6.109), i.e. the average of the partial sums, $\sum_{m=1}^{k} c_m = ka_k$, converges to zero. By Frobenius' theorem we conclude that $f(s)$ converges for $|s| < 1$ and $f(s) \to 0$ as $s \to 1-$. □

With those two lemmas available we are now ready to prove Theorem 6.2.6.

Proof of Theorem 6.2.6. By Theorem 6.2.5 we have for $|s| < 1$

$$\tau^{(n)}(s) = -e^{H(s)} \sum_{\sigma \in \mathcal{D}_n} c_\sigma H_\sigma(s), \tag{6.111}$$

where $H(s) = -\sum_{k=1}^{\infty} \frac{s^k}{k} P(S_k \leq 0)$ and

$$H_\sigma(s) = \prod_{i=1}^{n} \left(-\sum_{k=i}^{\infty} \frac{(k)_i}{k} s^{k-i} P(S_k \leq 0) \right)^{\sigma_i}. \tag{6.112}$$

Using that $\log(1-s) = -\sum_{k=1}^{\infty} s^k/k$, and hence $(i-1)!/(1-s)^i = \sum_{k=i}^{\infty} s^{k-i}(k)_i/k$, we get

$$\begin{aligned} e^{H(s)} &= \exp\left(\sum_{k=1}^{\infty} \frac{s^k}{k}(1 - P(S_k \leq 0)) - \sum_{k=1}^{\infty} \frac{s^k}{k} \right) \\ &= \exp\left(\sum_{k=1}^{\infty} \frac{s^k}{k} P(S_k > 0) \right)(1-s), \end{aligned} \tag{6.113}$$

and

$$\begin{aligned} c_\sigma H_\sigma(s) &= n! \prod_{i=1}^{n} \left(\sum_{k=i}^{\infty} \frac{(k)_i}{k} s^{k-i}(1 - P(S_k \leq 0)) - \sum_{k=i}^{\infty} \frac{(k)_i}{k} s^{k-i} \right)^{\sigma_i} \frac{1}{\sigma_i!(i!)^{\sigma_i}} \\ &= n! \prod_{i=1}^{n} \left(\sum_{k=i}^{\infty} \frac{(k)_i}{k} s^{k-i} P(S_k > 0) - \frac{(i-1)!}{(1-s)^i} \right)^{\sigma_i} \frac{1}{\sigma_i!(i!)^{\sigma_i}} \\ &= n! \prod_{i=1}^{n} \left(a_i - \frac{b^i}{i} \right)^{\sigma_i} \frac{1}{\sigma_i!}, \end{aligned} \tag{6.114}$$

where $a_i = \sum_{k=i}^{\infty} \frac{(k)_i}{k} s^{k-i} P(S_k > 0)/i!$ and $b = 1/(1-s)$. By Lemma 6.6.1 we then have

$$\sum_{\sigma \in \mathcal{D}_n} c_\sigma H_\sigma(s)/n! = \sum_{\sigma \in \mathcal{D}_n} \prod_{i=1}^{n} \frac{a_i^{\sigma_i}}{\sigma_i!} - b \sum_{\sigma \in \mathcal{D}_{n-1}} \prod_{i=1}^{n-1} \frac{a_i^{\sigma_i}}{\sigma_i!}. \tag{6.115}$$

Theorem 6.2.5 in combination with (6.113) and (6.115) yields

$$\begin{aligned} \tau^{(n)}(s) = &\exp\left(\sum_{k=1}^{\infty} \frac{s^k}{k} P(S_k > 0) \right) \sum_{\sigma \in \mathcal{D}_{n-1}} n c_\sigma \prod_{i=1}^{n-1} \left(\sum_{k=i}^{\infty} \frac{(k)_i}{k} s^{k-i} P(S_k > 0) \right)^{\sigma_i} \\ &- \exp\left(\sum_{k=1}^{\infty} \frac{s^k}{k} P(S_k > 0) \right) \sum_{\sigma \in \mathcal{D}_n} (1-s) c_\sigma \prod_{i=1}^{n} \left(\sum_{k=i}^{\infty} \frac{(k)_i}{k} s^{k-i} P(S_k > 0) \right)^{\sigma_i}. \end{aligned} \tag{6.116}$$

By assumption $\sum_{k=1}^{\infty} k^{n-2} P(S_k > 0) < \infty$, and by dominated convergence, all series of order at most $n-2$ thereby converge to finite limits as s tends to $1-$. In combination with Lemma 6.6.2 we also have

$$\lim_{s \to 1-} (1-s) \sum_{k=n}^{\infty} \frac{(k)_n}{k} s^{k-n} P(S_k > 0) = 0, \tag{6.117}$$

and we conclude

$$E((\tau_-)_n) = \lim_{s \to 1-} \tau^{(n)}(s) = \exp\left(\sum_{k=1}^{\infty} \frac{1}{k} P(S_k > 0)\right) n! \sum_{\sigma \in \mathcal{D}_{n-1}} \prod_{i=1}^{n-1} \left(\frac{h_i}{i!}\right)^{\sigma_i} \frac{1}{\sigma_i!} < \infty. \tag{6.118}$$

□

6.6.3 Conditional characteristic functions

Proof of Theorem 6.2.8. Define the (partial) characteristic functions

$$\gamma_n(\zeta) = E\left(e^{i\zeta S_n}; \tau_- = n\right), \tag{6.119}$$

in terms of which χ can be written as

$$\chi(s, \zeta) = \sum_{n=1}^{\infty} s^n \gamma_n(\zeta). \tag{6.120}$$

Rearranging (6.26) we get

$$1 - \exp\left(-\sum_{n=1}^{\infty} \frac{s^n}{n} g_n\right) = \sum_{n=1}^{\infty} s^n \gamma_n(\zeta), \tag{6.121}$$

where we have defined $g_n = E\left(e^{i\zeta S_n}; S_n \le 0\right)$. From Theorem 6.2.2 we get that

$$\begin{aligned} \exp\left(\sum_{n=1}^{\infty} (-1) \frac{g_n}{n} s^n\right) &= 1 + \sum_{m=1}^{\infty} \sum_{\sigma \in \mathcal{D}_m} \frac{s^m}{\sigma_1! \sigma_2! \cdots \sigma_m!} \prod_{n=1}^{m} \left((-1)\frac{g_n}{n}\right)^{\sigma_n} \\ &= 1 + \sum_{m=1}^{\infty} \sum_{\sigma \in \mathcal{D}_m} \frac{\operatorname{sgn}(\sigma)}{d_\sigma} s^m \prod_{n=1}^{m} g_n^{\sigma_n}. \end{aligned} \tag{6.122}$$

Insert this in (6.121) to obtain

$$-\sum_{m=1}^{\infty} \sum_{\sigma \in \mathcal{D}_m} \frac{\operatorname{sgn}(\sigma)}{d_\sigma} s^m \prod_{n=1}^{m} g_n^{\sigma_n} = \sum_{n=1}^{\infty} s^n \gamma_n(\zeta). \tag{6.123}$$

By inspection of the left and right hand side of the equality above we obtain

$$\gamma_n(\zeta) = -\sum_{\sigma \in \mathcal{D}_n} \frac{\operatorname{sgn}(\sigma)}{d_\sigma} \prod_{k=1}^{n} \left(E\left(e^{i\zeta S_k}; S_k \le 0\right)\right)^{\sigma_k}. \tag{6.124}$$

Finally, divide by τ_n to obtain (6.27). □

Proof of Theorem 6.2.9. For fixed ζ and n we have the following first-entrance decomposition

$$\begin{aligned} E(e^{i\zeta S_n}) &= \sum_{k=1}^{n} E(e^{i\zeta S_n}; \tau_- = k) + E(e^{i\zeta S_n}; \tau_- > n) \\ &= \sum_{k=1}^{n} E(e^{i\zeta S_k}; \tau_- = k) E(e^{i\zeta S_{n-k}}) + E(e^{i\zeta S_n}; \tau_- > n), \end{aligned} \tag{6.125}$$

where the second equality follows from the Markov property and the random walk structure. For ease of notation we let $e_k = E(e^{i\zeta S_k})$. In this notation, the decomposition above can be written

$$E(e^{i\zeta S_n};\tau_- > n) = e_n - \sum_{k=1}^{n} E(e^{i\zeta S_k};\tau_- = k)e_{n-k}. \tag{6.126}$$

Multiplying both sides of (6.126) by e_1 yields the relation

$$\begin{aligned} E(e^{i\zeta S_n};\tau_- > n)e_1 &= e_{n+1} - \sum_{k=1}^{n} E(e^{i\zeta S_k};\tau_- = k)e_{n+1-k} \\ &= E(e^{i\zeta S_{n+1}};\tau_- > n+1) + E(e^{i\zeta S_{n+1}};\tau_- = n+1), \end{aligned} \tag{6.127}$$

where we have used that for all k, $e_k e_1 = e_{k+1}$.

We are now ready to prove (6.28) by induction in n. Since the events $(\tau_- > 1)$ and $(S_1 > 0)$ are identical equation (6.28) holds for $n = 1$. Next, assume that (6.28) holds for n. Let $e_k^+ = E(e^{i\zeta S_k}; S_k > 0)$ and note that $E(e^{i\zeta S_k}; S_k \leq 0) = e_k - e_k^+$. From Theorem 6.2.8 we have

$$E(e^{i\zeta S_{n+1}};\tau_- = n+1) = -\sum_{\sigma\in\mathcal{D}_{n+1}} \prod_{k=1}^{n+1} \left(\frac{e_k^+ - e_k}{k}\right)^{\sigma_k} \frac{1}{\sigma_k!}. \tag{6.128}$$

Using that $e_k = e_1^k$ we get from (6.127), the induction hypothesis and (6.128) that

$$\begin{aligned} E(e^{i\zeta S_{n+1}};\tau_- > n+1) &= e_1 \sum_{\sigma\in\mathcal{D}_n} \prod_{k=1}^{n} \left(\frac{e_k^+}{k}\right)^{\sigma_k} \frac{1}{\sigma_k!} + \sum_{\sigma\in\mathcal{D}_{n+1}} \prod_{k=1}^{n+1} \left(\frac{e_k^+ - e_1^k}{k}\right)^{\sigma_k} \frac{1}{\sigma_k!} \\ &= \sum_{\sigma\in\mathcal{D}_{n+1}} \prod_{k=1}^{n+1} \left(\frac{e_k^+}{k}\right)^{\sigma_k} \frac{1}{\sigma_k!}, \end{aligned} \tag{6.129}$$

where the second equality uses Lemma 6.6.1 with $a_k = e_k^+/k$ and $b = e_1$. This shows that (6.28) holds for $n+1$ and we are finished. □

Bibliography

E. S. Andersen. On the fluctuations of sums of random variables II. *Mathematica Scandinavica 2*, pages 195–223, 1954.

S. Asmussen. *Applied Probability and Queues, second edition.* Springer-Verlag, New York, 2003.

S. Basak and G. Chabakauri. Dynamic mean-variance asset allocation. *The Review of Financial Studies 23(8)*, pages 2970–3016, 2010.

T. Björk and A. Murgoci. A general theory of Markovian time inconsistent stochastic control problems. *Working paper, Stockholm School of Economics*, 2009.

T. Björk, A. Murgoci, and X. Zhou. Mean variance portfolio optimization with state dependent risk aversion. *Mathematical Finance*, pages 1–24, 2012.

E. Briys and F. de Varenne. On the risk of insurance liabilities: debunking some common pitfalls. *Journal of Risk and Insurance 64(4)*, pages 673–694, 1997.

K. Bruhn and M. Steffensen. Household consumption, investment and life insurance. *Insurance Mathematics and Economics 48(3)*, pages 315–325, 2011.

M. M. Carhart. On persistence in mutual fund performance. *Journal of Finance 52*, pages 57–82, 1997.

J. C. Cox and C.-F. Huang. Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of Economic Theory 49(1)*, pages 33–83, 1989.

J. C. Cox and C.-F. Huang. A variational problem arising in financial economics. *Journal of Mathematical Economics 20(5)*, pages 465–487, 1991.

J. Cvitanic and I. Karatzas. Convex duality in constrained portfolio optimization. *Annals of Applied Probability 2(4)*, pages 767–818, 1992.

J. Cvitanic, J. P. Lehoczky, and I. Karatzas. Optimal portfolio and consumption decisions for a 'small investor' on a finite horizon. *SIAM Journal on Control and Optimization 25(6)*, pages 1557–1586, 1987.

T. M. Døskeland and H. A. Nordahl. Optimal pension insurance design. *Journal of Banking & Finance 32(3)*, pages 382–392, 2008.

D. Duffie and T. Zariphopoulou. Optimal investment with undiversifiable income risk. *Mathematical Finance 3(2)*, pages 135–148, 1993.

D. Duffie, W. Fleming, H. Soner, and T. Zariphopoulou. Hedging in incomplete markets with HARA utility. *Journal of Economic Dynamics and Control 21(4–5)*, pages 753–782, 1997.

P. Duren. *Invitation to Classical Analysis.* American Mathematical Society, 2012.

R. Durrett. *Probability: Theory and Examples, fourth edition.* Cambridge University Press, Cambridge, 2010.

P. H. Dybvig and H. Liu. Lifetime consumption and investment: Retirement and constrained borrowing. *Journal of Economic Theory 145(3)*, pages 885–907, 2010.

I. Ekeland and A. Lazrak. Being serious about non-commitment: subgame perfect equilibrium in continuous time. *Priprint, University of British Columbia*, 2006.

I. Ekeland and T. A. Pirvu. Investment and consumption without commitment. *Mathematics and Financial Economics 2(1)*, pages 57–86, 2008.

N. El-Karoui and M. Jeanblanc. Optimization of consumption with labor income. *Finance and Stochastic 2(4)*, pages 409–440, 1998.

N. El-Karoui and I. Karatzas. *The optimal stopping problem for a general American put-option. In Mathematics Finance (M. Davis, D. Duffie, W. Fleming. and S. Shreve, eds.)* ***IMA*** *Vol. 65, 77–88.* Springer-Verlag, New York, 1995.

N. El-Karoui, M. Jeanblanc, and V. Lacoste. Optimal portfolio management with American capital guarantee. *Journal of Economic Dynamics and Control 29(3)*, pages 449–468, 2005.

K. B. Ensor and P. W. Glynn. Simulating the maximum of a random walk. *Journal of Statistical Planning and Inference 85*, pages 127–135, 2000.

E. F. Fama and K. R. French. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance 5*, pages 1915–1947, 2010.

E. Farhia and S. Panageas. Saving and investing for early retirement: A theoretical analysis. *Journal of Financial Economics 83(1)*, pages 87–121, 2007.

W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II, second edition.* Wiley, New York, 1971.

P. A. Ferrari, H. Kesten, S. Martinez, and P. Picco. Existence of quasi-stationary distributions. A renewal dynamical approach. *Annals of Probability 23(2)*, pages 501–521, 1995.

W. H. Fleming and R. W. Richel. *Deterministic and Stochastic Optimal Control.* Springer-Verlag, New York, Berlin, Heidelberg, 1975.

H. U. Gerber and W. W. Shiu. Investing for retirement: Optimal capital growth and dynamic asset allocation. *North American Actuarial Journal 4*, pages 42–58, 2000.

J. Gil-Bazo and P. Ruiz-Verdú. The relation between price and performance in the mutual fund industry. *Journal of Finance 64(5)*, pages 2153–2183, 2009.

A. Grosen and P. L. Jørgensen. Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies. *Insurance: Mathematics and Economics 26(1)*, pages 37–57, 2000.

M. Guillén, S. F. Jarner, J. P. Nielsen, and A. M. Pérez-Marín. Risk-adjusted impact of administration costs in the distribution of terminal wealth for passive investment, mutual and hedge funds. *Working Paper*, 2013.

G. H. Hardy and S. Ramanujan. Asymtotic formulæ in combinatory analysis. *Proceedings of the London Mathematical Society 2.1*, pages 75–115, 1918.

H. He and H. F. Pagès. Labor income, borrowing constraints, and equilibrium asset prices. *Economic Theory 3(4)*, pages 663–696, 1993.

A. Hindi and C. Huang. Optimal consumption and portfolio rules with durability and local substitution. *Econometrica 6(1)*, pages 85–121, 1993.

H. Huang and M. A. Milevsky. Portfolio choice and mortality-contingent claims: The general HARA case. *The Jounal of Banking and Finance 32(11)*, pages 2444–2452, 2008.

H. Huang, M. A. Milevsky, and J. Wang. Portfolio choice and life insurance: The CRRA case. *The Jounal of Risk and Insurance 75(4)*, pages 847–872, 2008.

J. Huang, C. Sialm, and H. Zhang. Risk shifting and mutual fund performance. *Review of Financial Studies 24(8)*, pages 2575–2616, 2013.

S. D. Jacka and G. O. Roberts. Weak convergence of conditioned processes on a countable state space. *Journal of Applied Probability 32(4)*, pages 902–916, 1995.

B. A. Jensen and C. Sørensen. Paying for minimum interest rate guarantees: Who should compensate who? *European Financial Management 7(2)*, pages 183–211, 2001.

I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, Berlin, Heidelberg, 1991.

I. Karatzas and S. E. Shreve. *Methods of Mathematical Finance*. Springer-Verlag, New York, Berlin, Heidelberg, 1998.

A. Khorana, H. Seraes, and P. Tufano. Mutual fund fees around the world. *Review of Financial Studies 22(3)*, pages 1279–1310, 2008.

H. Koo. Consumption and portfolio selection with labor income: A continuous time approach. *Mathematical Finance 8(1)*, pages 49–65, 1998.

R. Korn. Some applications of L^2-hedging with a non-negative wealth process. *Applied Mathematical Finance 4(1)*, pages 65–79, 1997a.

R. Korn. *Optimal Portfolios*. World Scientific, 1997b.

H. Kraft and M. Steffensen. A dynamic programming approach to constrained portfolios. *Working paper*, 2012.

M. T. Kronborg. Optimal consumption and investment with labor income and Eurpean/American capital guarantee. *Working paper*, 2011.

M. T. Kronborg and M. Steffensen. Optimal consumption, investment and life insurance strategies with surrender option guarantee. *Scandinavian Actuarial Journal*, page Available at: http://dx.doi.org/10.1080/03461238.2013.775964, 2013.

E. M. Kryger. Pension fund design under long-term fairness constraints. *The Geneva Risk and insurance Review 35*, pages 130–159, 2010.

E. M. Kryger. Fairness versus efficiency of pension schemes. *European Actuarial Journal 1(1)*, pages 85–100, 2011.

E. M. Kryger and M. Steffensen. Some solvable portfolio problems with quadratic and collective objectives. *Working paper: Available at SSRN: http://ssrn.com/abstract=1577265*, 2010.

M. Kwak, Y. H. Shin, and U. J. Choi. Optimal investment and consumption decision of a family with life insurance. *Insurance: Mathematics and Economics 48(2)*, pages 176–188, 2011.

P. Lakner and L. M. Nygren. Portfolio optimization with downside constraints. *Mathematical Finance 16(2)*, pages 283–299, 2006.

J. B. Lasserre and C. E. M. Pearce. On the existence of a quasistationary measure for a Markov chain. *Annals of Probability 29(1)*, pages 437–446, 2001.

H. Markowitz. Portfolio selection. *Journal of Finance 7(1)*, pages 77–91, 1952.

S. Martinez, P. Picco, and J. S. Martin. Domain of attraction of quasi-stationary distributions for the Brownian motion with drift. *Advances in Applied Probability 30(2)*, pages 385–408, 1998.

R. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics 51(3)*, pages 247–257, 1969.

R. Merton. Optimum consumption and portfolio rules in a continuous-time model case. *Journal of Economic Theory 3*, pages 373–413, 1971.

S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability, 2nd edition*. London: Springer-Verlag, 2009.

M. A. Milevsky and V. R. Young. Annuitization and asset allocation. *Jounal of economic dynamics and control 31(9)*, pages 3138–3177, 2007.

C. Munk. Optimal consumption/investment policies with undiversifiable income risk and liquidity constraints. *Journal of Economic Dynamics and Control 24(9)*, pages 1315–1343, 2000.

P. H. Nielsen. Utility maximization and risk minimization in life and pension insurance. *Finanse and Stochastics 10*, pages 75–97, 2006.

P. H. Nielsen and M. Steffensen. Optimal investment and life insurance strategies under minimum and maximum constraints. *Insurance Mathematics and Economics 43(1)*, pages 15–28, 2008.

R. Norberg. A theory of bonus in life insurance. *Finance and Stochastics 3*, pages 373–390, 1999.

R. Norberg. On bonus and bonus prognoses in life insurance. *Scandinavian Actuarial Journal 22*, pages 126–147, 2001.

J. Palczewski, R. Poulsen, K. R. Schenck-Hoppé, and H. Wang. Dynamic portfolio optimization with transaction costs and state-dependent drift. *Working paper*, page http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2346304, 2013.

S. R. Pliska and J. Ye. Optimal life insurance purchase and consumption/investment under uncertain lifetime. *Journal of Banking and Finance 31(5)*, pages 1307–1319, 2007.

M. Preisel, S. F. Jarner, and R. Eliasen. Investing for retirement through a with-profits pension scheme: a client's perspective. *Scandinavian Actuarial Journal 1*, pages 15–35, 2010.

S. Richard. Optimal consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model. *Journal of Finance Economics 2(2)*, pages 187–203, 1975.

F. Spitzer. A combinatorial lemma and its application to probability theory. *Transactions of the American Mathematical Society 82*, pages 323–339, 1956.

M. Steffensen. Contigent claims analysis in life and pension insurance. *Proceedings 10th AFIR Colloquium*, pages 587–603, 2000.

M. Steffensen. On Merton's problem for life insurers. *ASTIN Bulletin 34(1)*, pages 5–25, 2004.

M. Steffensen and H. Kraft. Optimal consumption and insurance: A continuous-time markov chain approach. *ASTIN Bulletin 38(1)*, pages 231–257, 2008.

R. Strotz. Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies 23(3)*, pages 165–180, 1955.

W. Szpankowski. *Average Case Analysis of Algorithms on Sequences.* Wiley-Interscience, 2001.

L. Teplá. Optimal investment with minimum performance constraints. *Journal of Economic Dynamics and Control 25(10)*, pages 1629–1645, 2001.

R. L. Tweedie. Quasi-stationary distributions for Markov chains on a general state space. *Journal of Applied Probability 11(4)*, pages 726–741, 1974.

M. E. Yaari. Uncertain lifetime, life insurance and the theory of the consumer. *Review of Economic Studies 32(2)*, pages 137–150, 1965.

X. Y. Zhou and D. Li. Continuous-time mean-variance portfolio selection: A stochastic LQ framwork. *Applied Mathematical Optimization 42(1)*, pages 19–33, 2000.