

# VU Research Portal

## Optimal Financial Decision Making under Loss Averse Preferences

Siegmann, A.H.

2003

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Siegmann, A. H. (2003). *Optimal Financial Decision Making under Loss Averse Preferences*. Tinbergen Instituut.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

Optimal Financial Decision Making  
under Loss Averse Preferences

*Ph.D. Thesis*

Arjen Siegmann

July 18, 2002



*vrije* Universiteit *amsterdam*



# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Loss aversion . . . . .	2
1.2	Applied finance . . . . .	4
1.3	Overview of the thesis . . . . .	5
<b>2</b>	<b>Solution to the Multi-stage Mean-Shortfall model</b>	<b>9</b>
2.1	Introduction . . . . .	9
2.2	Model setup . . . . .	11
2.3	Results . . . . .	13
2.3.1	Solution to the multistage model . . . . .	13
2.3.2	Tuning the model . . . . .	17
2.3.3	Time-diversification . . . . .	19
2.4	The multi-asset case . . . . .	20
2.5	Conclusions and discussion . . . . .	23
2.A	Proofs . . . . .	26
<b>3</b>	<b>Extensions to the Mean-Shortfall Model</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	The behavioral value function . . . . .	36
3.3	Quadratic shortfall . . . . .	45
3.4	Extra kink . . . . .	49
3.5	Restrictions on asset allocations . . . . .	51
3.6	Conclusions . . . . .	54
3.A	Proofs . . . . .	55
<b>4</b>	<b>Optimal Pension Fund Investment Management</b>	<b>59</b>
4.1	Introduction . . . . .	59

4.2	Pension funding as a mean-shortfall optimization . . . . .	61
4.3	Results . . . . .	65
4.3.1	Base case . . . . .	65
4.3.2	Uncertainty in inflation . . . . .	69
4.3.3	Alternative interest rates . . . . .	72
4.4	Downside deviation . . . . .	73
4.5	Empirical evidence . . . . .	75
4.6	Discussion and conclusion . . . . .	79
<b>5</b>	<b>Household Savings under Loss Aversion</b>	<b>81</b>
5.1	Introduction . . . . .	81
5.2	The model . . . . .	82
5.3	Results . . . . .	84
5.4	Habit formation . . . . .	88
5.5	Kahneman & Tversky preferences . . . . .	92
5.6	Conclusion . . . . .	94
5.A	Proofs . . . . .	95
<b>6</b>	<b>Explaining Hedge Fund Returns by Loss Aversion</b>	<b>99</b>
6.1	Introduction . . . . .	99
6.2	Model . . . . .	103
6.3	Implications for hedge funds . . . . .	108
6.4	Conclusions . . . . .	111
6.A	Proofs . . . . .	113
<b>7</b>	<b>Conclusion</b>	<b>119</b>

---

# Introduction

---

This thesis studies investment decisions. The theory of financial decision making considers investors who allocate capital in an environment that is uncertain by definition. In mainstream finance, people's attitude toward uncertainty has been modeled in terms of risk aversion. Put informally, risk aversion means that if several investments have the same expected return/pay-off, the one with the smallest variation in the outcome is preferred. To model risk aversion for optimal decision making, one usually models the objective for an investor as a concave utility function in terms of total wealth. For such functions the main results are known since long, see the overview in Merton (1990) and Ingersoll (1987). For three different appearances of risk aversion, the fraction of initial capital that is invested in a risky asset is either decreasing, constant, or increasing in the initial wealth of the investor. Since these results were established, utility functions that show hyperbolic risk aversion have dominated the finance literature. Such functions are mathematically convenient.

There are, however, serious problems with these classical utility functions. First, there is little empirical evidence that people have preferences that can be modeled realistically with such functions. Although the idea of risk aversion is appealing and not necessarily untrue, the popular choice of utility function seems more motivated from a mathematical point than from the point of realism. Second, from the area of behavioral psychology there has emerged a concrete view on how people evaluate outcomes of financial decision. This is not compatible with the traditional way of modeling preferences

in economics. Consistently, people seem to value outcomes not in terms of their total financial wealth, but rather in terms of *changes* with respect to their current situation. Coupled with an aversion against negative changes, such attitudes can be characterized with the term *loss aversion*. The basic psychology book of Gleitman et al. (2000) has the following definition:

**“Loss Aversion:** A widespread pattern, evident in many aspects of decision making, in which people seem particularly sensitive to losses and eager to avoid them.”

This thesis is about financial decision making when investors are loss-averse. We provide new results and insights on the implied relation between wealth and risk-taking, applied to different areas of finance.

## 1.1 Loss aversion

Although it is a well-known and widespread pattern, research analyzing the effect of loss aversion in all aspects of finance and financial decision making has only just started. With respect to financial decision making, loss aversion is actively studied in two particular areas of finance. The first area is that of behavioral finance, which concentrates on behavioral phenomena in experimental and empirical (real-life) settings of financial decision-making. The second is the area of Asset/Liability Management and risk management, where loss aversion is used implicitly through downside risk measures.

### Behavioral finance

Behavioral finance is rooted in experimental work regarding the estimation of preferences by performing laboratory experiments. From these experiments, many of the assumptions in the classical framework on utility functions, expected utility maximization, and rationality were very soon found to be invalid.<sup>1</sup> An important article is that of Kahneman and Tversky (1979). Based on insights from behavioral psychology they propose a decision framework called ‘prospect theory’ in which

- value is assigned to gains and losses rather than to absolute wealth levels, and
- probabilities are replaced by decision weights.

---

<sup>1</sup>See Rabin (2002) for an extensive critique on classical economic theory and its failure to incorporate established behavioral insights.

Although both elements of prospect theory were not new at the time, Kahneman and Tversky were the first to integrate them in a framework for decision making under risk. Besides necessary extensions, Tversky and Kahneman (1992) estimate parameter values for a value- and weighting function. Especially at the end of the 1980s much research has been published on how people actually behave when faced with a decision situation involving an uncertain outcome. It is important to note that the effects modeled by Kahneman and Tversky are *systematic*. Given their systematic nature, the results need to be taken into account by any economist modeling risk-taking behavior.

In this thesis we do not consider the part of prospect theory that replaces probabilities by subjective decision weights. We represent loss aversion in its simplest form: if an outcome is above the reference point, its subjective value equals the monetary value of the outcome. Outcomes below the reference point are multiplied by a constant penalty that represents the degree of loss aversion. In a setting with uncertainty, our objective function can be called mean-shortfall, as it maximizes average (mean) wealth, penalizing expected shortfall below a reference point.

The empirical side of behavioral finance studies real-life economic behavior, applying behavioral insights to improve understanding of empirical phenomena. However, contrary to our approach of focusing on one particular preference structure, most empirical research considers a number of behavioral insights at the same time. For example, when evidence for loss aversion in buy and sell behavior of stocks is studied in Shefrin and Statman (1985), the existence of mental accounts is considered simultaneously. In the same way, other research integrates loss aversion with overconfidence, or with over- and under-reaction to information. See for example Shleifer (2000) for an extensive study of behavioral finance in relation to the efficient market hypothesis. Hirshleifer (2001) presents a survey of investor psychology in relation to security prices. There are many behavioral phenomena to consider and to integrate with loss aversion. However, the result is that it remains unclear until now what the sole effect of loss-averse preferences is on decision making.

## **Rationality**

Most behavioral phenomena can be considered cognitive biases. Cognitive biases are by definition irrational. Other than reading a book on the subject, e.g., Kahneman et al. (1982), most of us do not consciously know that we possess them. As such, integrating (irrational) biases with loss aversion sug-

gests that loss aversion is also an irrational bias. This is clearly not the case. In general, people actively express loss-averse attitudes. This suggests that most of us are rationally aware that our aversion to losses is stronger than our liking of gains. This is also evidenced by the use of downside risk measures in normative areas of finance, such as Asset/Liability Management and risk management, see Section 1.2. An additional problem of combining loss aversion with cognitive biases, is that it becomes difficult, if not impossible, to pinpoint the effect of loss aversion on decision making.

The way this thesis stands out in relation to other research in behavioral finance, is that we (i) focus solely on loss aversion, (ii) consider it a rational preference, and (iii) focus on the structure of the optimal decisions, rather than on an axiomatic approach to characterize loss aversion. This way, we isolate the effect of loss-averse preferences on financial decision making, and focus on the economics of the resulting decisions.

## 1.2 Applied finance

Whether founded in theory or not, in the area of applied finance there is already an asymmetric attitude with respect to gains and losses. The importance of downside risk measures has been recognized and used quite extensively. See also the work of Artzner et al. (1999), who derive axiomatic results on the coherence of certain (downside) risk measures. Two applied areas where the downside risk measures are most visible are those of banking regulation and Asset/Liability Management.

### **Banking regulation**

Since the early 90s, banks report their risks in a measure called ‘Value-at-Risk’ (VaR). VaR measures the location of a specified quantile of the loss distribution. Hence it explicitly associates risk with the downside of the financial position of a bank, in contrast with traditional measures that used the overall variability, e.g., the variance, of the return. See the study of Campbell (2001) who studies the optimal allocation and performance of portfolios when VaR is the risk measure.

### **Asset/Liability Management**

In applied models for Asset/Liability Management, the risk measure is very important, as it can influence the outcome of an ALM model considerably. In Cariño et al. (1994) an ALM model is described that was implemented for the Japanese insurance company Yasuda-Kasai. When considering the risk

measure to use, they do away with the traditional risk measure, variance of total return. Arguing that ‘it is unclear that excessive returns are undesirable’, they choose to measure the risk for the firm as the expected amount by which goals are not achieved. Besides the Value-at-Risk measure mentioned above, a specific risk measure that is considered in research, as well as in applied ALM models, is expected shortfall. It measures the expected amount by which a specified target is not met. Alternatively, downside deviation is used as a shortfall measure that penalizes quadratic shortfall. This way, larger shortfall gets a proportionally larger penalty.

Most research in ALM has evolved toward solving complex optimization models with a focus on the solution method, e.g., Consigli and Dempster (1998). Seeing ALM as a technical challenge makes it an attractive research subject with the discipline of operations research. Mathematical tools can be used to come up with optimal investment and funding policies to increase revenue or decrease costs for financial institutions. However, the problem with such an approach is that the relation between the risk measure and the resulting optimal policies remains unclear. At best, a specialized consultant working with ALM models on a daily basis will develop some intuition for the sensitivity of the outcomes to alternative parameter values. No fundamental insights in financial decision making are obtained, however.

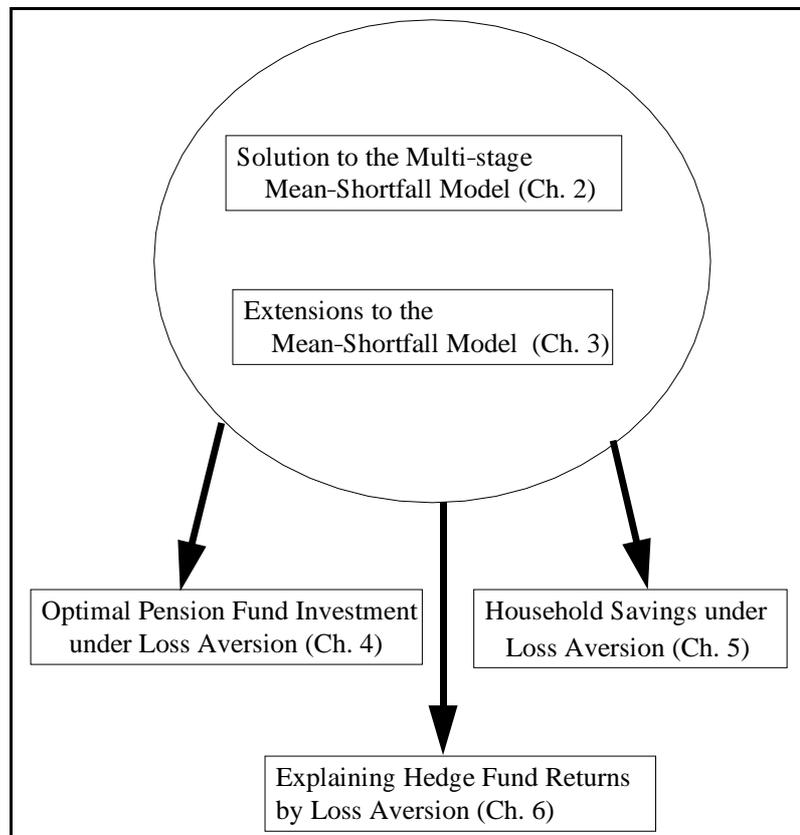
The relevance of our work for the area of Asset/Liability Management is that we (i) consider a parametric solution to models that are basic versions of advanced ALM models with a downside risk-measure, (ii) analyze the mechanisms behind the resulting patterns of risk-taking (instead of focusing on computational aspects), and (iii) obtain results on the robustness for different specifications of downside risk and other model formulations.

It is worth mentioning here the papers by Basak and Shapiro (2001), and Berkelaar and Kouwenberg (2000b). They are close in spirit to the perspective taken in this monograph. Both Basak and Shapiro and Berkelaar and Kouwenberg focus on optimal policies under loss-averse preferences. However, to retain tractability, they use continuous-time modeling, combined with the assumption of complete markets. Our approach, by contrast, considers incomplete markets, and is set in discrete time. To retain tractability, we focus on simpler loss-averse objective functions.

### 1.3 Overview of the thesis

The contents of the thesis are best introduced using the diagram in Figure 1.1.

In Chapter 2 the main results for the mean-shortfall model are derived.



**Figure 1.1: Overview of the thesis**

Optimizing over one risky and one risk-free asset, we find the optimal solution in terms of initial wealth. We give intuition for the typical risk-taking behavior that is observed.

We consider the model in Chapter 2 to be a basic representation of models used in ALM. Also, it is the most basic representation of loss aversion as used in behavioral finance. However, the exact formulation of loss aversion and downside risk measures varies in the literature. Chapter 3 considers alternative specifications of loss aversion to test the robustness of the result from Chapter 2. Therefore, Chapters 2 and 3 can be considered the basic building blocks of the thesis. The Chapters 4, 5, and 6 then provide applications of the results. All these chapters share the basic formulation of the objective function of Chapter 2.

In Chapter 4 we use loss aversion to analyze the outcomes of an ALM model for a stylized pension fund. As the financial position of pension funds is influenced by many sources of uncertainty, this chapter contains a sensitivity analysis of the results to the specification of uncertainty in the mean-shortfall

model. Given that pension funds actively perform ALM studies to assist in decision making, the chapter contains empirical data on actual investment behavior. We use these data to look for evidence of loss-averse preferences.

Chapter 5 applies the mean-shortfall objective to represent household preferences for consumption. There is already a large body of literature on household savings, but very little on the effect of loss aversion. This chapter shows savings behavior in a two-period model when households have loss-averse preferences. As commonly analyzed in the literature, we add the possibility of habit formation to the model, i.e., the second period benchmark is affected by first-period consumption. The relevance of loss aversion seems even more pronounced in the case of habit formation: the existence of a habit implies, almost by definition, that people do not wish to realize consumption below a given habitual level.

Chapter 6 analyzes financial decision making for hedge funds when loss-averse preferences are assumed. Hedge funds are a particular class of mutual funds, in which only qualified investors can invest. Consequently, they have a large degree of freedom in the investment strategies they can follow, being exempt from the regulation that holds for most other mutual funds. The case of hedge funds is interesting, as they are found to generate non-linear pay-offs relative to underlying financial indices. This calls for an extension of the mean-shortfall model with an asset that has a non-linear relation with the risky asset. We find that the resulting model retains analytical tractability. We derive a set of optimal pay-off patterns. The results match empirical patterns found in the literature. We discuss the optimal financial decisions in terms of specific dynamic investment strategies that hedge funds can follow. Chapter 7 concludes.



---

# Solution to the Multi-stage Mean-Shortfall model

---

## 2.1 Introduction

Over the past decade, we have witnessed a growing literature on financial planning models. Such models can assist financial institutions like pension funds, insurance companies and banks in their Asset/Liability Management (ALM), as illustrated in the book by Ziemba and Mulvey (1998). The key component in these models concerns the trade-off between risk and return. It is therefore of paramount importance which risk measure is put into the model. Traditionally, the variance or standard deviation has been the prominent measure of risk. Its main (and perhaps only) advantage is its computational simplicity. As argued by Sortino and Van der Meer (1991) however, the variance is an inadequate measure of risk in many practical circumstances. The main criticism to the use of the variance is its symmetric nature, whereas risk is typically perceived as an asymmetric phenomenon. Asymmetric or downside risk measures are generally more difficult to work with, both analytically and computationally. Given the current state of computer technology and the increased use of derivative assets in investment, however, their use in both theoretical and empirical financial planning models has increased rapidly.

Downside-risk measures are currently used extensively in the area of Asset/Liability Management (ALM). Recent research typically incorporates the down-side risk measure in a multi-stage stochastic programming (MSP) ap-

proach, as in Consigli and Dempster (1998), Mulvey and Thorlacius (1998), Cariño et al. (1994), Boender (1997), and Dert (1998). The main advantage of the MSP approach relative to the more traditional static mean-variance oriented approach is that the explicit dynamic nature of financial decisions can better be taken into account. For example, a decision now may be followed by recourse actions in the future. Moreover, different preferences and (dynamic) constraints can be modeled directly. As a result, the MSP approach generally produces significant improvements over static mean-variance based decisions. These improvements can be exploited when the MSP model is implemented in practice, as in Cariño et al. (1994). The MSP-models used for ALM, however, also have two interrelated drawbacks.

First, it is often impossible to solve an MSP analytically. As an alternative, most people seek for computational solutions built around discretizations of the original MSP, the so-called scenario approach. This gives rise to a deterministic program as an approximation to the MSP model. See for example Hiller and Eckstein (1993) who study a financial planning model from a numerical rather than an analytical or practical perspective. The computational effort needed to solve the deterministic programming problem increases rapidly in the number of scenarios used. Second, the numerical solution typically consists of optimal decisions at every point in time and in every possible state of nature (as represented in the scenario tree). Even for simple MSP models and realistically sized scenario structures, this gives an unwieldy set of numbers that lacks transparency and clear-cut economic interpretation. Moreover, the numerical solution is only optimal for the specific scenario tree, parameter values and initial state variables for which the problem is solved. See also Dupačová et al. (1998) who check the robustness of the optimal solution to a stochastic program with respect to out-of-sample scenarios. The two drawbacks are also acknowledged in Consigli and Dempster (1998), who note on the complexity of the solution to an MSP that “The solution to these very large and complex problems needs to be followed by a detailed computer-based analysis of the results in order to supply conveniently represented information to the decision maker.” In other words, for these models to be implemented and used by management or decision makers, there is a need for a method to summarize optimal decisions in terms of decision rules, linking optimal decisions in each state of nature to observed quantities like the assets to liabilities ratio, the state of the economy, etc.

Despite extensive research in the area of Stochastic Programming, analytic solutions to multistage stochastic programs are rare. The main contribution of this chapter is that we analytically characterize the solution to a basic multistage financial planning model. It shares the main character-

istics of more elaborate applied financial planning models that are widely used in the financial industry, see Ziemba and Mulvey (1998). Besides its immediate interpretation as a general financial planning model, the model that is introduced in this chapter will also serve as a basis for the rest of this thesis. By solving the basic model, we hope to provide insight into the key mechanisms driving the results of similar full-scale models used in empirical work. The interaction between the formulation of the MSP model and the resulting decisions is generally badly understood. By solving our basic model analytically, we can present the solution as decision rules in feedback-form, which have an economic interpretation.

It is this interpretation that extends to other chapters. What we learn in this chapter on decision taking under uncertainty will prove to be very useful in analyzing behavior that follows from assuming loss averse preferences for pension funds (Chapter 4), households (Chapter 5), and hedge funds (Chapter 6).

The remainder of this chapter is set up as follows. In Section 2.2 the basic multi-stage model is introduced. Section 2.3 gives the solution to the model and the corresponding optimal decision rules. It also gives a number of interesting consequences that follow from the solution. Section 2.4 considers the case of having multiple risky assets, as is the case in a portfolio optimization. Section 2.5 ends with a discussion of the found results and concludes. The Appendix gathers the proofs.

## 2.2 Model setup

Let  $W_t$  denote the wealth of an investor at time  $t = 0, \dots, T$ . At each time  $t$  there are two investment opportunities: a risk-free asset with certain return  $r_f$  in each period, and a risky asset with uncertain return  $u_t$  over period  $t$ . Restricting the investment categories to only two assets may appear restrictive at first sight. However, as shown by for example Merton (1990), if two-fund separation holds in an economy, all efficient asset allocations are completely spanned by investments in the risk-free asset and the market portfolio only. In this case, our  $u_t$  would represent the return on the market portfolio, i.e. a portfolio that holds all available securities in proportion to their market values. See Cass and Stiglitz (1970) for a general discussion of separation theorems. In Section 2.4 we extend the current model to more assets, and show that separation holds indeed. The certain return  $r_f$  can also be replaced by a time-dependent  $r_{f,t}$ , representing for example a term-structure of interest rates. This does not affect the outcome of the model, however, so we drop the subscript  $t$  for ease of exposition. Note further

that it is central to this chapter that we leave the exact distribution of  $u_t$  unspecified. This way, possible interpretations of  $u$  include returns on stock investments, derivatives, etc.

With the two investment categories given, wealth evolves as

$$W_{t+1} = W_t \cdot r_f + X_t \cdot (u_{t+1} - r_f), \quad t = 0, \dots, T - 1, \quad (2.1)$$

where  $X_t$  is the amount invested in the risky asset at the beginning of period  $t$ , which is allowed to be state-dependent. We assume that the  $u_{t+1}$ s are independent, though not necessarily identically distributed with absolute continuous distribution function  $G_{t+1}(\cdot)$  on  $[0, \infty)$ . We assume  $\mathbb{E}_t[u_{t+1}] > r_f$ , with  $\mathbb{E}_t(\cdot) = \int \cdot dG_{t+1}$ , i.e., the expected return on the risky asset always exceeds the risk-free return. The investor's objective function is given by

$$\max_{X_0, \dots, X_{T-1}} \mathbb{E}_0[W_T] - \lambda \cdot \mathbb{E}_0[(W_T^B - W_T)^+], \quad (2.2)$$

with  $W_T^B$  a benchmark level of wealth at the horizon,  $\lambda > 0$  a risk-aversion parameter, and  $(y)^+$  denoting the maximum of 0 and  $y$ . The investor thus trades off expected wealth or return against risk. As such, (2.2) contains the key ingredients of typical financial planning models. Risk in (2.2) is measured as the expected shortfall of final wealth with respect to the benchmark  $W_T^B$ . As risk is only associated with the down-side of the distribution of terminal wealth, the second term in (2.2) is called a downside-risk measure. Note that an equivalent optimization problem is obtained when we add a term  $-W^B$  to (2.2). Hence, the trade-off can be expressed in expected surplus versus expected shortfall.

If  $W^B = W_0$ , terminal wealth is measured against initial wealth and risk is interpreted in terms of expected losses. Alternative interpretations of  $W^B$  include pension liabilities and index or benchmark returns. The latter are relevant in a context of relative performance evaluation. Model (2.2) also serves as a typical example of a financial planning model in Birge and Louveaux (1997). There the setting is given by parents who wish to provide for a child's college education  $T$  years from now. In that case  $W^B$  represents the tuition goal and  $\lambda$  the cost of borrowing if the goal is not met.

Hiller and Eckstein (1993) use the same objective function as (2.2) in a stochastic dedication model for fixed-income portfolios. (2.2) also resembles the objective function of Cariño et al. (1994), albeit that they use a piecewise linear penalty function in terms of the expected loss, where we have a linear one. Cariño et al. (1994) also argue that in the context of banks and insurance companies risk measures as in (2.2) can easily be justified: these companies are faced with specific additional funding costs if reserves fall below critical threshold levels. The risk measure in (2.2) in terms of returns is

used by Rockafellar and Uryasev (2000) in a static stochastic portfolio optimization. The model we solve compares to the two-asset case of a model by Birge and Louveaux (1997). They use it to illustrate the use of scenarios in transforming the original MSP in a deterministic LP problem. We stick to the original MSP without resorting to its deterministic implementation. Note that optimal parametric solutions have been derived for models formulated in continuous-time like in Merton (1969) and Sethi (1998) for the consumption/investment problem and Ingersoll (1987) for the general portfolio optimization problem. See also the continuous-time model of Basak and Shapiro (2001), who optimize expected utility of terminal wealth under a constraint on downside-risk. By contrast, our model is set in discrete time, for which few analytic results are available.

The risk aversion parameter  $\lambda$  in (2.2) determines the trade-off between risk and return. Setting  $\lambda = 0$  implies risk neutrality, while increasing  $\lambda$  induces loss aversion. For the moment, we assume  $\lambda$  is determined directly by the decision maker. In Subsection 2.3.2 we prove that  $\lambda$  can also be set indirectly by fixing the shortfall probability.

## 2.3 Results

In this section we obtain three main results. First, using Dynamic Programming and exploiting the special functional form of the objective function, we obtain the complete solution to model (2.2) up to a set of  $2 * T$  parameters that depend on the model parameters and the probability distributions of the risky returns  $u_{t+1}$ . These parameters can be computed efficiently for any given set of distributions  $G_{t+1}, t = 0 \dots, T - 1$ . Second, we prove that there is a unique shortfall probability associated with any value of  $\lambda$  that gives a finite solution. This allows the investor to pick  $\lambda$  by specifying preferences with respect to shortfall probabilities only. Third, we prove that the sensitivity of the decision rules to changes in (relative) wealth increases when the horizon extends. This provides further evidence for the time-diversification controversy in finance.

### 2.3.1 Solution to the multistage model

Define  $W_t^B = W_T^B / r_f^{T-t}$  as the risk-free discounted value of benchmark wealth at time  $t$ , and  $S_t = W_t - W_t^B$  as the surplus. The next theorem has the main result.

**Theorem 2.3.1** *If there is a bounded solution to the optimization problem in (2.1) and (2.2), then it is given by*

$$X_t^* = \frac{r_f}{r_f - \bar{u}_{t+1}^*} \cdot S_t, \quad t = 0, \dots, T-1, \quad (2.3)$$

where  $\bar{u}_{t+1}^*$  is smaller than  $r_f$  for positive surplus and larger than  $r_f$  for negative surplus, so  $X_t^* \geq 0$ . Specifically,  $\bar{u}_{t+1}^*$  is one of two possible  $\bar{u}$ s that solve

$$\mathbb{E}_t[u_{t+1} - r_f] = \lambda_t \cdot \mathbb{E}_t[(r_f - u_{t+1}) \cdot I_{\{u_{t+1} \leq \bar{u}\}}]. \quad (2.4)$$

The two solutions to (2.4) are labeled  $\bar{u}_{t+1}^+ < r_f$  and  $\bar{u}_{t+1}^- > r_f$ . Further,  $\lambda_t$  is defined by

$$\lambda_t = \frac{\lambda \cdot (p_{t+1}^- - p_{t+1}^+)}{1 + \lambda \cdot p_{t+1}^+}, \quad (2.5)$$

where for  $k = t+1, \dots, T-1$  the  $p_k^+$  and  $p_k^-$  are defined recursively by

$$p_k^- = p_{k+1}^- \cdot G_{k+1}(\bar{u}_{k+1}^-) + p_{k+1}^+ (1 - G_{k+1}(\bar{u}_{k+1}^-)), \quad (2.6)$$

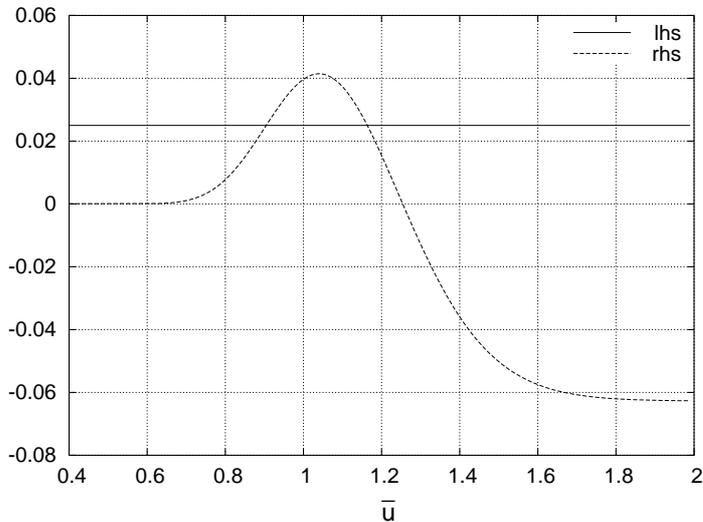
$$p_k^+ = p_{k+1}^- \cdot G_{k+1}(\bar{u}_{k+1}^+) + p_{k+1}^+ (1 - G_{k+1}(\bar{u}_{k+1}^+)), \quad (2.7)$$

and  $p_T^- \equiv 1$ ,  $p_T^+ \equiv 0$ .

**Proof:** See appendix.

The key result of the theorem is that in equation (2.3) we have the exact shape of the decision rule for  $X_t^*$ . Moreover, equation (2.4) can be easily solved numerically for any distribution function  $G_{t+1}(\cdot)$ . From the definition of  $\lambda_t$  in (2.5) it follows that its value is determined by  $\lambda$  and future values of  $\bar{u}_k^*$ ,  $k > t+1$ . Consequently, the model can be solved by sequentially solving equation (2.4) for  $t = T-1, \dots, 0$ , giving the parameters  $\bar{u}_{t+1}^*$  that determine the slope of the decision rule for  $X_t^*$ .

An important property of the optimal solution is that (2.3) and (2.4) represent the solution to a static model ( $T = 1$ ), with risk-aversion parameter  $\lambda$  equal to  $\lambda_t$ . The multi-stage problem is therefore almost myopic: (i) it can be solved by solving a sequence of static problems, but (ii) the static problems are linked through the time-varying risk aversion parameter  $\lambda_t$ . This can be exploited to solve the dynamic program numerically without resorting to scenarios and deterministic equivalents of the original MSP.



**Figure 2.1: Finding the optimal  $\bar{u}$**

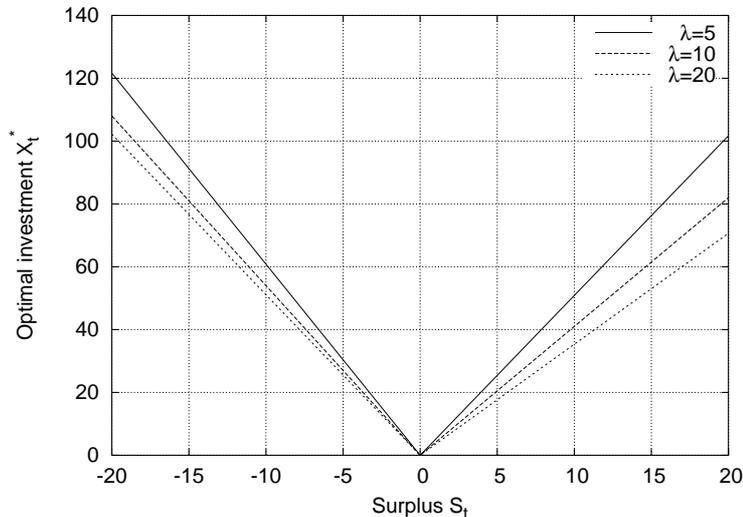
In this figure, the solid line ‘lhs’ represents the left-hand side of equation (2.4), which is a constant expression. The dashed line ‘rhs’ represents the right-hand side, which is a function of  $\bar{u}$ . The points of intersection are the  $\bar{u}$ s that solve (2.4). The risk-free rate  $r_f$  is 1.04, and the uncertain return  $u$  is distributed lognormal(0.085, 0.16), representing a typical stock return with a mean return of 10% and standard deviation of 17%.

With respect to the solution to (2.4), note that the right-hand side of (2.4) is unimodal in  $\bar{u}$ . The maximum is reached at  $\bar{u} = r_f$ . Consequently, if

$$-\lambda_t \cdot \mathbb{E}_t [(u_{t+1} - r_f) \cdot I_{\{u_{t+1} \leq r_f\}}] < \mathbb{E}_t [u_{t+1} - r_f], \quad t = 0, \dots, T-1, \quad (2.8)$$

(2.4) has two distinct solutions: one smaller and one larger than  $r_f$ , and (2.8) is a sufficient condition for having a bounded solution. Under (2.8), solving equation (2.4) can be visualized as in Figure 2.1. Solving (2.4) and finding the optimal decision rules (2.3) boils down to evaluating the conditional first moments of  $G_{t+1}(\cdot)$ . This can be done very efficiently. If condition (2.8) is not met, either  $\bar{u}_{t+1}^* = r_f$  or (2.4) has no solution. In both cases, the optimal decision  $X_t^*$  in (2.3) is unbounded.

It is shown in the appendix that under (2.8), the two  $\bar{u}$ s that solve (2.4) indeed result in the optimal  $X_t^*$ . As  $X_t^* > 0$  at the optimum, it follows from (2.3) that  $\bar{u} < r_f$  and  $\bar{u} > r_f$  correspond to the cases of a positive and a negative time  $t$  surplus  $W_t - W_t^B$ , respectively. We therefore label the solutions as  $\bar{u}_t^+$  and  $\bar{u}_t^-$  for  $W_t > W_t^B$  and  $W_t < W_t^B$ , respectively, implying piecewise linearity of the optimal decision rule in (2.3).



**Figure 2.2: Optimal investment in the risky asset**

For a one-period model, this figure shows the optimal investment in the risky asset,  $X_0^*$ , as a function of initial surplus,  $S_0$ .  $r_f = 1.04$ ,  $u$  has a lognormal distribution, with  $\mu = 0.085$ , and  $\sigma = 0.16$ , representing a typical stock return with a mean of 10% and standard deviation of 17%.

It is important to note that the piecewise linearity of the optimal decision rule for  $X_t$  in Theorem 2.3.1 holds for *any* set of absolute-continuous distribution functions  $G_{t+1}(\cdot)$ ,  $t = 0, \dots, T - 1$ .  $G_{t+1}(\cdot)$  only enters the optimal decision rule through the values of  $\bar{u}_{t+1}^+$  and  $\bar{u}_{t+1}^-$ , which follow easily from equation (2.4). Theorem 2.3.1 thus not only provides the solution to the dynamic investment problem for standard assets like a stock or bond index. Even if  $G_{t+1}$  is the return distribution of a complicated derivative instrument, piecewise linearity of  $X_t^*$  still holds<sup>1</sup>.

The third consequence of Theorem 2.3.1 follows from the typical shape of the decision rule for  $X_t^*$ , which we have plotted in Figure 2.2 for a relevant set of parameter values and different values for  $\lambda$ . We can observe that the amount invested in the risky asset is decreasing in the surplus when the surplus is negative, and increasing in case it is positive. This typical behavior of the optimal decision rule implies that in an unfavorable situation (negative surplus) the decision maker takes more risk as the surplus *decreases*. In the

<sup>1</sup>A logical extension of distribution functions is one that considers discrete probability distributions. In this case, absolute continuity is not retained, but the proof of Theorem 2.3.1 will still hold, with some adjustments. The outcome is predictable, as we already use discretizations of the probability distributions to plot the optimal solutions for specific parameter values.

favorable situation of a positive surplus, the decision maker takes more risk when the surplus *increases*. The fact that the slope of the decision rule is different for the two regimes is due to the different risk taking behavior implied in the objective function. In a situation of a negative surplus the decision maker must take more risk: the higher return on the portfolio is needed to make the probability of recovering from the unfavorable situation strictly positive. In the situation of a positive surplus risky investments allow one to profit from the higher expected return, subject to having an acceptable trade-off with downside-risk. Clearly, if the surplus is positive and wealth increases, more money can be invested in the risky asset at the cost of only a marginal increase in downside-risk. This induces a positive relation between the surplus and the investment in the risky asset. In the above line of reasoning, the effect of  $\lambda$  is clear, as visualized in Figure 2.2. Higher values of  $\lambda$  decrease risk taking, for comparable wealth levels.

### 2.3.2 Tuning the model

The next corollary shows that the multi-stage model can be tuned appropriately by relating the model parameter  $\lambda$  to a specified shortfall probability.

**Corollary 2.3.1** *If the optimal solution is bounded and initial surplus is positive, there is a unique  $\lambda^+$  such that under the optimal policies  $X_t^*, t = 0, \dots, T-1$ ,  $\Pr[W_T < W^B]$  is fixed. For a negative initial surplus, there is a unique  $\lambda^-$  associated with the probability  $\Pr[W_T > W^B]$ .*

**Proof:** See appendix.

Corollary 2.3.1 implies that the decision maker does not have to pick  $\lambda$  in the multi-stage model directly. As the probabilities of a positive and a negative final surplus are uniquely determined by a choice of  $\lambda$ , the manager only has to specify her probabilistic preferences with respect to shortfall. A value for  $\lambda$  then follows immediately. Of course the choices for the probabilities are limited by condition (2.8), which ensures that the optimal decisions  $X_t^*$  are finite.

Tuning the model by setting a shortfall or recovery probability rather than a value for  $\lambda$ , enhances the practical interpretability of the model. The financial industry widely uses the concept of Value-at-Risk (VaR), which is intimately related to the shortfall probabilities and benchmark wealth levels, see Jorion (2000). In a static context, Rockafellar and Uryasev (2000) propose a technique for stochastic portfolio optimization in which Mean Shortfall (or CVaR) is optimized and VaR is calculated simultaneously. As such Corollary 2.3.1 presents the same result for the multi-stage optimization problem.

Note that  $\lambda$  can not be interpreted as the Lagrange multiplier of an optimization problem that maximizes terminal wealth with a *restriction* on expected shortfall. In our model the probability of shortfall is fixed for a given  $\lambda$ , but not the expected shortfall itself. If expected shortfall appears in a restriction, the value of  $\lambda$  as a Lagrange multiplier would change with the surplus.

In a static context, it is obvious that expected shortfall is directly linked to a quantile of the return distribution. However, in a dynamic context we have not seen this relationship in the literature before. If applicable in more elaborate MSP models, it can be a valuable tool in fine-tuning empirical financial planning models.

**Corollary 2.3.2** *If  $\lambda \rightarrow \infty$ , the optimal solution to model (2.2) is*

$$X_t^* = \begin{cases} S_t & \text{if } S_t \geq 0, \\ \frac{r_f}{r_f - \bar{u}_{t+1}^*} S_t & \text{if } S_t < 0, \end{cases} \quad (2.9)$$

where  $\bar{u}_{t+1}^*$  is the  $\bar{u}$  that solves

$$\int_0^{\bar{u}} (r_f - u_t) dG_t = 0. \quad (2.10)$$

**Proof:** See appendix.

Corollary 2.3.2 shows that for positive surpluses, the optimal investment in the risky asset equals the surplus. In the case of a positive surplus, it is clear that increasing the shortfall penalty to infinity drives the allocation to generate zero expected shortfall. As only the surplus is invested in the risky asset, losing the complete investment in the risky asset still leaves exactly the next-period benchmark wealth  $W_{t+1}^B$ .

If the surplus is negative, the solution to equation (2.10) induces a positive investment in the risky asset, which only depends on the distribution of  $u_t$  and  $r_f$ . Compared to the case of a positive surplus, there is no possibility to rule out shortfall. Hence, the positive excess return on the risky asset leads to a positive risky investment. Moreover, if the  $u_t$ s are identically distributed, the allocation in the risky asset is the same at all time periods. Note that compared to only investing in the risk-free asset, the optimal allocation in the risky asset decreases the expected shortfall. In Chapter 3 we will derive results for the limiting portfolio when downside deviation is used as the risk measure, and also when an extra kink is added to the bilinear objective of this chapter.

### 2.3.3 Time-diversification

The following corollaries give important properties of the  $\lambda_t$ s and the optimal  $X_t^*$ , respectively.

**Corollary 2.3.3** *If the optimization problem (2.2) has a bounded solution, for  $T$  and  $\lambda$  fixed, the values of  $\lambda_t$  are increasing in  $t$ .*

**Proof:** See appendix.

**Corollary 2.3.4** *If the  $u_t$ s are identically distributed, the absolute slope  $r_f/|\bar{u}_{t+1}^* - r_f|$  of the optimal decision rules in (2.3) is monotonically increasing in the time to maturity  $T - t$ .*

**Proof:** See appendix.

As the  $\lambda_t$ s represent the investor's risk aversion, Corollary 2.3.3 states that the investor in the multi-stage setting becomes less risk averse the longer the time to maturity  $T - t$ . As a direct result, Corollary 2.3.4 concludes that the optimal policies are more sensitive to changes in surplus if the planning horizon is further away. Moreover, the increasing slopes of the decision rules imply riskier initial asset allocations for longer planning periods and given (non-zero) initial surplus.

Corollaries 2.3.3 and 2.3.4 have implications in the area of time diversification theory. Proponents of time diversification argue that investors hold riskier asset allocations whenever their investment horizon is further away. This is generally illustrated in the context of stocks and bonds. The intuition is that in the long run the increased uncertainty due to a more risky portfolio is more than compensated by the higher expected return. The uncertainty as measured by the return standard deviation increases with the square root of the horizon. By contrast, the expected return increases linearly. As a result, the shortfall probability of stocks decreases in the distance to the horizon, and for long horizons stocks almost certainly outperform bonds. Samuelson (1994), Kritzman and Rich (1998) and Merton and Samuelson (1974), however, argue that the spread of the distribution of terminal wealth also widens with the time horizon. Consequently, whereas the probability of a loss decreases, the potential magnitude of the loss increases accordingly if riskier investments are held. This is used as an argument to prove that it could well be that an investor chooses a less aggressive portfolio if the investment horizon increases. Corollary 2.3.4 proves, however, that if wealth is measured against a benchmark level and the magnitude of a loss is taken into account,

dynamically optimizing risk averse investors hold riskier asset allocations at longer horizons.

## 2.4 The multi-asset case

The optimization model in Section 2.2 includes only two assets. We argued that this is not a serious restriction of the model, as the return on the risky asset can be interpreted as the return on a market portfolio. In this section we give the solution to the model when there is more than 1 risky asset. This changes the focus of the analysis from one that looks only at risk-taking to one that considers the solution to the model in the broadest sense, namely as a portfolio selection problem.

We analyze a one-period optimization with two risky assets and one risk-free asset. For the situation of  $n$  risky assets, the interested reader is referred to the appendix.

Define  $u_A$  and  $u_B$  as the return on the risky assets A and B, respectively. The amounts invested in these assets are represented by  $X_A$  and  $X_B$ . The amount invested in the risk-free asset is therefore  $W_0 - X_A - X_B$ . The one-period optimization problem is now given by

$$\max_{X_A, X_B} \mathbb{E}_0[W_1] - \lambda \cdot \mathbb{E}_0 [(W^B - W_1)^+], \quad (2.11)$$

$$\text{s.t. } W_1 = W_0 \cdot r_f + X_A \cdot (u_A - r_f) + X_B \cdot (u_B - r_f), \quad (2.12)$$

where  $W^B$  represents the benchmark at time 1, and  $\lambda$  is again the loss aversion parameter. Defining  $S_0$  as the initial surplus  $W_0 - W^B/r_f$ , the main result is given in the next theorem.

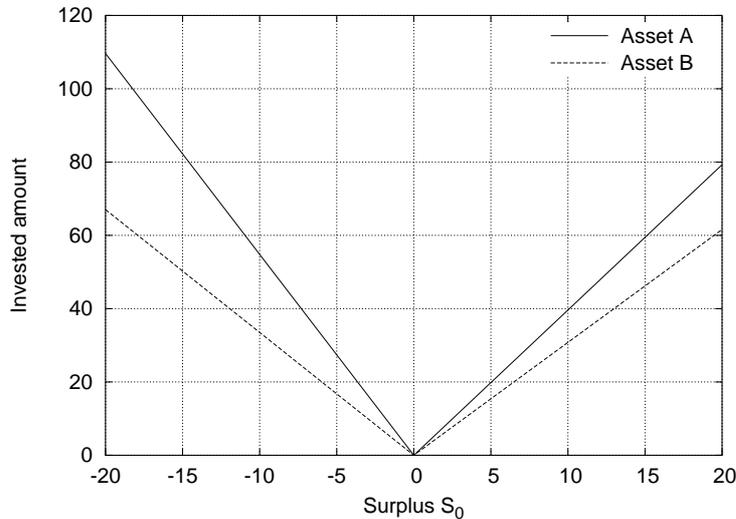
**Theorem 2.4.1** *If there is a bounded solution, the optimal solution to the optimization problem in (2.11) is given by*

$$X_A^* = \frac{r_f}{r_f - k_A^*} S_0, \quad (2.13)$$

$$X_B^* = \frac{r_f}{r_f - k_B^*} S_0, \quad (2.14)$$

where  $k_A^*$ , and  $k_B^*$  are equal to  $k_A^+ < r_f$ ,  $k_B^+ < r_f$ , and  $k_A^- > r_f$ ,  $k_B^- > r_f$  for positive and negative surplus, respectively.

Theorem 2.4.1 shows that for two risky assets, the expressions for the optimal amounts  $X_A^*$  and  $X_B^*$  are similar to those for the setting with one risky asset. There is a piecewise linear relation between the amount invested



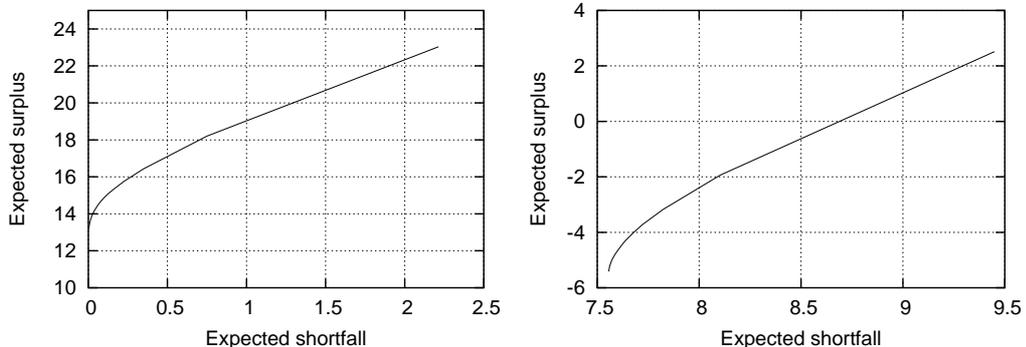
**Figure 2.3: Optimal investment with two risky assets**

Asset A is distributed lognormal with  $\mu = 0.085$ ,  $\sigma = 0.16$ , corresponding to a mean return of 10% and standard deviation of 17%. The return on asset B is uncorrelated with asset A, log-normally distributed with  $\mu = 0.10$ ,  $\sigma = 0.3$ , corresponding to a mean return of 16% and standard deviation of 33%.  $\lambda = 5$  and  $r_f = 1.04$ .

in the risky asset and the surplus. A consequence of this result it that the ratio of the investments in A and B, i.e., the portfolio composition, is a constant for either sign of the surplus. Figure 2.3 shows the allocation in two risky assets that have different risk-return properties.

From Figure 2.3 we can see that for positive and negative surpluses, different portfolio compositions are optimal, i.e., the optimal risk-return profile differs depending on the starting situation. We could not have observed this from Figure 2.2, where the investment in one risky asset reaches comparable levels for negative and positive surpluses. Allowing for nonlinear pay-offs, Chapter 6 further explores the optimality of pay-off patterns and finds more specific differences in investment allocation for initial wealth above and below the benchmark.

Elaborating further on the difference between a positive and negative initial surplus, there is a serious consequence of the results in terms of an efficient frontier as used in the CAPM. There, an efficient frontier is constructed based on the variance and mean of portfolio returns. Our results, however, imply that based on expected return and expected shortfall, the



**Figure 2.4: Risk-return frontiers**

This figure shows the efficient frontiers for average shortfall versus average surplus for two initial surplus levels. The left panel has initial surplus of +10%, the right panel of -10%. Parameter values are  $r_f = 1.04$ , and  $u_A \sim \log N(0.085, 0.16)$ ,  $u_B \sim \log N(0.10, 0.30)$ .

efficient frontier differs for a positive and negative surplus. We illustrate this for the model with two risky assets in Figure 2.4.

The left panel in Figure 2.4 shows the risk-return frontier in case of a positive surplus. The *efficient* frontier starts at zero shortfall and positive surplus and then shows expected surplus as a function of expected shortfall. The specific position on the line depends on the value of  $\lambda$ . The right panel in Figure 2.4 shows the frontier for a negative initial surplus. There, the frontier does not hit the zero-shortfall axis, i.e., there is no allocation that gives zero shortfall risk. Efficient allocations are located at the minimum-risk point and the points to the upper-right of that point. The 100% risk-free investment would be located to the lower right of the frontier, and is thus inefficient.

It is remarkable that the effect of having two frontiers instead of one has gone largely unnoticed in previous research on downside risk measures, for example in the work of Brouwer (1997), who has thoroughly analyzed efficient frontiers for portfolio optimization under downside risk measures. See also Huisman et al. (1999), who consider efficient frontiers in a Value-at-Risk framework. Only in the behavioral portfolio theory of Shefrin and Statman (2000), which is similar to a Value-at-Risk framework, also two efficient frontiers are found, based on the different ‘aspiration levels’ of investors.

A possible explanation of overlooking the existence of two efficient frontiers can be that most researchers assume an investor with a ‘positive surplus’, i.e., the starting situation is favorable. However, when an explicit down-side risk measure is used, it seems reasonable to consider the possibility of an investor actually reaching or being in a situation of negative surplus. In

that case, our results suggest that the investor should change his portfolio composition.

A final consequence of the Theorem 2.4.1 is that it sheds light on the interpretation of the risky return we gave in the model of Section 2.2. There we interpreted  $u_t$  as the return on *the* market portfolio, while we just derived that in a portfolio context there is not one, but there are two optimal asset mixes.<sup>2</sup> Although we do not prove it formally, it can be shown that the results in Section 2.2 continue to hold, except for the fact that the  $\bar{u}_t$ s need a more subtle interpretation. Such a solution would simultaneously characterize the solution to a  $T$ -period model with  $n$  risky assets.

## 2.5 Conclusions and discussion

In this chapter we analytically characterized the solution to a multi-stage financial planning model involving a trade-off between return and downside-risk. Our model shares the same basic characteristics of more elaborate empirical models like that of Cariño et al. (1994). Downside-risk was used because of its widespread popularity in the financial industry and its use in related academic work. Risk was measured with respect to a benchmark wealth level, which can be interpreted as a liability level in case of Asset/Liability Management (ALM) problems. We have derived the analytic characterization of the optimal solutions in feedback form, i.e., as decision rules. This form has a clear-cut economic interpretation, which constitutes a valuable addition to the existing literature where solutions are often derived numerically rather than analytically.

One of the salient findings of this chapter is the V-shape in terms of the investment in the risky asset as a function of the surplus. This result does not hinge on the probability distribution of the return on the risky asset. The distribution only determines the relative steepness of the decision rule for positive and negative values of the surplus.

In contrast to many traditional continuous-time analyses involving utility functions, our model results in additional risk taking behavior both in situations of under-funding and over-funding. In case of under-funding (i.e., a negative surplus), taking risk is the only way to make the probability of a positive final surplus positive. In case of over-funding (i.e., a positive surplus), making risky investments increases expected return without unduly affecting downside-risk measures.

---

<sup>2</sup>Notice that our results suggest three-fund separation for the static mean-shortfall model, opposed to two-fund separation in a mean-variance world.

For all practical purposes, it is important for the implementation of any mathematical model that model parameters can be chosen in an intuitive manner. For the basic financial planning model we presented in this chapter, we find that the choice for the parameter that reflects downside-risk aversion is directly connected to the probability of shortfall at the horizon. This shows that in this case a decision maker can make the trade-off between return and downside-risk explicit by choosing an appropriate shortfall probability. Using a more popular concept in the area of finance, this is equivalent to specifying a Value-at-Risk quantile for the wealth distribution at the horizon. Note however that the resulting optimal investment strategy still accounts for both the event and the extent of shortfall. As such, it does not suffer from the drawbacks of a purely VaR-based objective function, which only incorporates the event or probability of shortfall.

We have proved that the solution can be represented as the solution to a sequence of one-period models where the typical shape of the decision rules is constant over time. Its steepness, however, decreases because the appropriate risk aversion parameter increases over time. This leads directly to the paradigm of time diversification: investors hold riskier asset allocations if the planning period is longer. The increased riskiness of the portfolio is offset by the larger set of possible future recourse actions.

On the computational side, our analytic results also give rise to some interesting directions for future research. As the piece-wise linear structure of the optimal decision rules does not depend on the initial level of surplus nor on the precise form of the return distribution, optimal solutions to dynamic versions of our model can be found by an easy decomposition of the dynamic model into repeated one-period models. Specifically, more complex financial planning models could be solved through a decision-rule based approach, optimizing over the basis and slopes of a V shape. Such a decomposition allows for a considerable reduction in computational burden without affecting the optimality of the solution found. A useful line of future research would be to investigate to what extent similar properties hold and can be exploited in more elaborate empirical MSP models of this type.

Finally, we have motivated the choice for expected shortfall as the risk measure based on the observation that it is widely used in Asset/Liability Management studies. We have also explained that it is similar to the value function found empirically by Kahneman and Tversky (1979). Besides the fact that we can derive closed-form decision rules, the fact that the solution is not (yet) observed in practice makes it even more interesting. If the objective function is used in ALM, then the effect found is relevant for institutions and regulators alike, and is something to be reckoned with. However, there are many more possibilities of objective functions that can be called loss averse,

and so it remains to be seen how restrictive the bilinear formulation of the objective is, and whether the result carry over to other loss averse objectives. We explore the consequences of using other objective functions in Chapter 3.

# Appendix

## 2.A Proofs

### Proof of Theorem 2.3.1:

We solve the optimization problem as given by equations (2.1) and (2.2) by dynamic programming. Start with the following sequence of value functions:

$$V_t(W_t) = \max_{X_t} \mathbb{E}_t[V_{t+1}(W_{t+1}r_f + X_t(u_{t+1} - r_f))], \quad (2A.1)$$

$$t = 0, \dots, T,$$

$$V_T(W_T) = W_T - \lambda \cdot [W_T^B - W_T]^+ \quad (2A.2)$$

where  $\mathbb{E}_t$  is defined as the conditional expectation given  $u_{t-1}, u_{t-2}, \dots, u_1$ . Given  $W_t$ ,  $V_t(W_t)$  is the expected value of the objective function when all decisions from time  $t$  on are taken optimally. Clearly, solving the model is equivalent to finding  $V_0(W_0)$ .

The first order condition with respect to  $X_{T-1}$  is obtained by differentiating  $V_T(W_T)$  to  $X_{T-1}$  and is given by

$$\mathbb{E}_{T-1}[u_T - r_f] + \lambda \cdot \int_0^{\bar{u}_T} (u_T - r_f) dG_T(u_T) = 0, \quad (2A.3)$$

where  $\bar{u}_T$  is defined as

$$\bar{u}_T = r_f + \frac{W_T^B - W_T \cdot r_f}{X_{T-1}}. \quad (2A.4)$$

As shown in the main text, if condition (2.8) holds, (2A.3) solves for two distinct values of  $\bar{u}_T$ . To proof that the resulting  $X_{T-1}^*$  is optimal, we consider the second-order condition for the problem at  $T - 1$ :

$$\frac{\partial^2 V_T(W_T)}{\partial X_{T-1}^2} = -\lambda \cdot \frac{(W_T^B - W_{T-1}r_f)^2}{X_{T-1}^3} \cdot g_T(\bar{u}(X_{T-1})) < 0, \quad (2A.5)$$

where  $g_t(\cdot)$  is defined as the density function of  $u_t$ , for  $t = 1, \dots, T$ .

As  $\lambda > 0$ , it follows from (2A.5) that an optimal  $X_{T-1}^*$  should be positive at the optimum. Therefore, only a  $\bar{u}_T < r_f$  can be optimal if  $W_T^B - W_{T-1}r_f < 0$ . Otherwise,  $X_{T-1}^*$  would be negative. Also, only a  $\bar{u}_T > r_f$  can be optimal if  $W_T^B - W_{T-1}r_f > 0$ . As these two cases correspond to a positive and a negative surplus, respectively, we define  $\bar{u}_T^+$  and  $\bar{u}_T^-$  as the corresponding values of  $\bar{u}$ .

As  $X_{t-1}$  influences the objective function only through the effect on  $W_t$ , i.e. there are no direct costs associated with a choice of  $X_{t-1}$ , the first order condition with respect to  $X_{t-1}$ ,  $t < T$  is

$$\mathbb{E}_{t-1} \left[ \frac{\partial V_t(W_t)}{\partial X_{t-1}} \right] = \mathbb{E}_{t-1} \left[ \frac{\partial V_t(W_t)}{\partial W_t} \cdot \frac{\partial W_t}{\partial X_{t-1}} \right] = 0. \quad (2A.6)$$

Let  $W_{t+1}^* = r_f W_t + X_t^* \cdot (u_{t+1} - r_f)$ , with  $X_t^*$  the optimal decision at time  $t$ . We have

$$\begin{aligned} \frac{\partial V_t(W_t)}{\partial W_t} &= \mathbb{E}_t \left[ \frac{\partial V_{t+1}(W_{t+1})}{\partial W_{t+1}} \cdot \frac{\partial W_{t+1}^*}{\partial W_t} \right] = r_f \cdot \mathbb{E}_t \left[ \frac{\partial V_{t+1}(W_{t+1})}{\partial W_{t+1}} \right] \\ &= r_f^2 \cdot \mathbb{E}_t \left[ \frac{\partial V_{t+2}(W_{t+2})}{\partial W_{t+2}} \right] = \dots = r_f^{T-t} \mathbb{E}_t [1 + \lambda \cdot I_{\{S_T < 0\}}], \end{aligned} \quad (2A.7)$$

where  $I_A$  is the indicator function of the event  $A$ . In this case it is equal to 1 if  $S_T < 0$  and 0 otherwise.

Using (2A.7) and the fact that  $\partial W_t / \partial X_{t-1} = u_t - r_f$ , the first order condition in (2A.6) becomes

$$\mathbb{E}_{t-1} \left[ \frac{\partial V_t(W_t)}{\partial X_{t-1}} \right] = \mathbb{E}_{t-1} [r_f^{T-t} \cdot (u_t - r_f)] + \lambda \cdot \mathbb{E}_{t-1} [r_f^{T-t} \cdot (u_t - r_f) \cdot I_{\{S_T < 0\}}] = 0. \quad (2A.8)$$

Dividing by  $r_f^{T-t}$  and partitioning based on the sign of  $S_t$  gives

$$\mathbb{E}_{t-1} [u_t - r_f] + \lambda \cdot \mathbb{E}_{t-1} [(u_t - r_f) \cdot I_{\{S_T < 0\} \cap \{S_t < 0\}}] + \lambda \cdot \mathbb{E}_{t-1} [(u_t - r_f) \cdot I_{\{S_T < 0\} \cap \{S_t > 0\}}] = 0. \quad (2A.9)$$

Note that for given  $W_{t-1}$ ,  $X_{t-1}^*$  is fixed. As the  $u_t$ s are independent,  $S_t$  only varies monotonically with the realization of  $u_t$ . By absolute continuity of  $G_t(\cdot)$ , we do not have to consider the case  $S_t = 0$ , so we can define a  $\bar{u}_t$  such that  $S_t > 0$  for  $u_t > \bar{u}_t$  and  $S_t < 0$  for  $u_t < \bar{u}_t$ . This reduces (2A.9) to

$$\begin{aligned} \mathbb{E}_{t-1} [u_t - r_f] + \lambda \cdot \int_0^{\bar{u}_t} (u_t - r_f) \Pr(S_T < 0 | S_t) dG_t(u_t) \\ + \lambda \cdot \int_{\bar{u}_t}^{\infty} (u_t - r_f) \Pr(S_T < 0 | S_t) dG_t(u_t) = 0. \end{aligned} \quad (2A.10)$$

Assume

$$\Pr(S_T < 0 | S_t) = \begin{cases} p_t^+ & \text{if } S_t > 0, \\ p_t^- & \text{if } S_t < 0. \end{cases} \quad (2A.11)$$

The above assumption states that the probability of ending up with a negative terminal surplus only depends on the *sign* of the time  $t$  surplus  $S_t$  and not on its value. This clearly holds for  $t = T$  with  $p_T^+ = 0$  and  $p_T^- = 1$ . (2A.11) implies that equation (2A.10) can be rewritten as

$$\mathbb{E}_{t-1}[u_t - r_f] + \lambda \cdot p_t^- \cdot \int_0^{\bar{u}_t} (u_t - r_f) dG_t(u_t) + \lambda \cdot p_t^+ \cdot \int_{\bar{u}_t}^{\infty} (u_t - r_f) dG_t(u_t) = 0, \quad (2A.12)$$

which simplifies to

$$(1 + \lambda \cdot p_t^+) \cdot \mathbb{E}_{t-1}[u_t - r_f] + \lambda \cdot (p_t^- - p_t^+) \cdot \int_0^{\bar{u}_t} (u_t - r_f) dG_t = 0, \quad (2A.13)$$

or, dividing by  $(1 + \lambda \cdot p_t^+)$

$$\mathbb{E}_{t-1}[u_t - r_f] + \frac{\lambda \cdot (p_t^- - p_t^+)}{(1 + \lambda \cdot p_t^+)} \cdot \int_0^{\bar{u}_t} (u_t - r_f) dG_t = 0. \quad (2A.14)$$

Defining  $\lambda_{t-1}$  as

$$\lambda_{t-1} = \frac{\lambda \cdot (p_t^- - p_t^+)}{1 + \lambda \cdot p_t^+}, \quad (2A.15)$$

we can write (2A.14) as

$$\mathbb{E}_{t-1}[u_t - r_f] + \lambda_{t-1} \cdot \int_0^{\bar{u}_t} (u_t - r_f) dG_t = 0. \quad (2A.16)$$

As equation (2A.16) is the time- $t$  version of the first order condition for the problem at  $T-1$ , we can use the exact same reasoning as for the equation in (2A.3). So, if condition (2.8) holds this equation solves for two distinct  $\bar{u}_t$ , corresponding to a positive and a negative surplus, respectively. Moreover, the corresponding optimal  $X_{t-1}^*$  is such that the next period's surplus,  $S_t$ , is positive for a realization of the risky return  $u_t > \bar{u}_t$  and negative for  $u_t < \bar{u}_t$ . Therefore,

$$\Pr(S_T < 0 | S_{t-1}) = \Pr(S_T < 0 | S_t < 0) \cdot \Pr(S_t < 0 | S_{t-1}) \quad (2A.17)$$

$$\begin{aligned} &+ \Pr(S_T < 0 | S_t > 0) \cdot \Pr(S_t > 0 | S_{t-1}) \\ &= p_t^- \cdot \int_0^{\bar{u}_t} dG_t(u_t) + p_t^+ \cdot \int_{\bar{u}_t}^{\infty} dG_t(u_t), \end{aligned} \quad (2A.18)$$

where  $\bar{u}_t = \bar{u}_t^+$  for  $S_{t-1} > 0$  and  $\bar{u}_t = \bar{u}_t^-$  for  $S_{t-1} < 0$ . As both  $\bar{u}_t^+$  and  $\bar{u}_t^-$  are constant, (2A.18) clearly shows that  $\Pr(S_T < 0 | S_{t-1})$  also satisfies (2A.11). The proof now follows by induction.  $\blacksquare$

**Proof of Corollary 2.3.1:**

Define  $G_t^-$  and  $G_t^+$  by  $G_t(\bar{u}_t^-)$  and  $G_t(\bar{u}_t^+)$ , respectively. Given a value of  $\lambda$ ,  $\bar{u}_T^+$  and  $\bar{u}_T^-$  are uniquely determined for the decision problem at time  $T - 1$ . In turn,  $\bar{u}_T^+$  and  $\bar{u}_T^-$  determine  $G_T^+$  and  $G_T^-$  which determine  $p_{T-1}^+$  and  $p_{T-1}^-$ . Repeating this procedure for all previous subproblems finally gives a *unique*  $p_0^+$  or  $p_0^-$ , depending on whether  $S_0 > 0$  or  $S_0 < 0$ , respectively. This implies that if a certain  $p_0^+$  or  $p_0^-$  is chosen, there will be either an unbounded solution, or a bounded solution with a unique  $\lambda$  associated with it. ■

**Proof of Corollary 2.3.2:**

Consider the proof of Theorem 2.3.1. If  $\lambda \rightarrow \infty$ , the first order condition at  $T - 1$  simplifies to

$$\int_0^{\bar{u}} (r_f - u_T) dG_T = 0. \quad (2A.19)$$

This equation has two solutions. For positive surplus,  $\bar{u}^+ = 0$ , implying  $X_{T-1}^* = S_{T-1}$ . In the proof of Theorem 2.3.1, this puts  $p_T^+$  to zero. For negative surplus, there is a  $\bar{u}^- > r_f$  that solves (2A.19), see Figure 2.1.  $p_{T-1}^-$  is positive. Following the proof of Theorem 2.3.1, we arrive at a first order condition for  $X_t$  given by

$$\mathbb{E}_{t-1}[u_t - r_f] + \lambda \cdot p_t^- \cdot \int_0^{\bar{u}_t} (u_t - r_f) dG_t = 0, \quad (2A.20)$$

where  $p_t^-$  is recursively defined as  $p_{t+1}^- \cdot G_t(\bar{u}_t^-)$ , with  $p_T^- = 1$ . The result follows by dividing by  $\lambda p_t^-$  and letting  $\lambda \rightarrow \infty$ . ■

**Proof of Corollary 2.3.3:**

The risk aversion parameter at time  $t$ ,  $\lambda_t$  is defined in (2A.15). From (2A.18) we have that  $p_t^+$  and  $p_t^-$  are defined recursively as

$$p_t^- = p_{t+1}^- \cdot G_t(\bar{u}_t^-) + p_{t+1}^+ (1 - G_t(\bar{u}_t^-)), \quad (2A.21)$$

$$p_t^+ = p_{t+1}^- \cdot G_t(\bar{u}_t^+) + p_{t+1}^+ (1 - G_t(\bar{u}_t^+)), \quad (2A.22)$$

and  $p_T^- = 1$ ,  $p_T^+ = 0$ .

As  $\bar{u}_t^- > \bar{u}_t^+$ , we have  $1 > G_t^- > G_t^+ > 0$  for any  $t$ . Using (2A.21) and (2A.22), we can write

$$p_t^- - p_t^+ = (p_{t+1}^- - p_{t+1}^+) \cdot (G_{t+1}^- - G_{t+1}^+), \quad i = 0, \dots, n - 1. \quad (2A.23)$$

As  $p_T^- > p_T^+$  and  $(G_{t+1}^- - G_{t+1}^+) < 1$  for all  $t$ , this implies that

$$p_t^- - p_t^+ < p_{t+1}^- - p_{t+1}^+, \quad t = 0, \dots, T - 1, \quad (2A.24)$$

i.e.,  $\{p_t^- - p_t^+\}_0^T$  is a strictly positive and increasing sequence in  $t$ . Moreover, as  $p_t^- > p_t^+$ , it follows from (2A.22) that  $p_{t+1}^+ < p_t^+$ . Using this and (2A.24), it is easy to see that the numerator in the definition of  $\lambda_t$  (2A.15) is increasing in  $t$ , while the denominator is decreasing in  $t$ . Hence  $\lambda_t$  is increasing in  $t$ , and thus decreasing in the time to maturity  $T - t$ . ■

**Proof of Corollary 2.3.4:**

With identically distributed returns  $u_t$ ,  $t = 1, \dots, T$ , the solution to (2.4) depends only on the value of  $\lambda_t$ . As can be observed from Figure 2.1, a larger value of  $\lambda_t$  corresponds to a larger distance between the  $\bar{u}_{t+1}$ s and  $r_f$ . This implies that the distance  $|r_f - \bar{u}_{t+1}^*|$  is increasing in  $t$ . Hence, the absolute slope  $r_f/|r_f - \bar{u}_t^*|$  in the decision rule for  $X_t^*$  is decreasing in  $t$ . ■

**Proof of Theorem 2.4.1**

The optimization problem is given by

$$\max_{X_A, X_B} \mathbb{E}_0[W_1] - \lambda \cdot \mathbb{E}_0 [(W^B - W_1)^+], \quad (2A.25)$$

$$\text{s.t. } W_1 = W_0 \cdot r_f + X_A \cdot (u_A - r_f) + X_B \cdot (u_B - r_f), \quad (2A.26)$$

which can be written as

$$\max \mathbb{E}[W_1] - \lambda \cdot \int_{SF} (W^B - W_1) dG(u_A, u_B), \quad (2A.27)$$

with  $SF$  defined as the set

$$SF = \{(u_A, u_B) \in \mathbb{R}^2 | S_0 r_f + X_A(u_A - r_f) + X_B(u_B - r_f) < 0\}, \quad (2A.28)$$

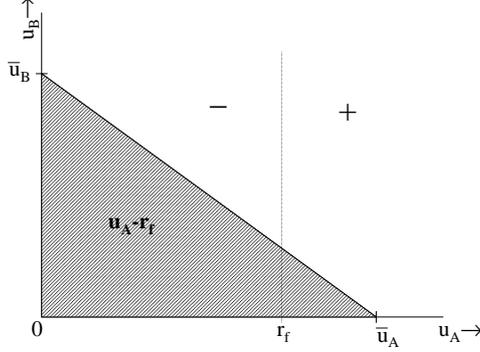
where  $S_0 = W_0 - W^B/r_f$ . For given  $S_0$ ,  $X_A$  and  $X_B$  the set  $SF$  defines the area in  $(u_A, u_B)$ -space where there is shortfall, i.e.,  $W_1 < W^B$ . Given the expression in (2A.27), the first order conditions for  $X_A$  and  $X_B$  are given by

$$X_A : \quad \mathbb{E}[u_A - r_f] + \lambda \cdot \int_{SF} (u_A - r_f) dG(u_A, u_B) = 0, \quad (2A.29)$$

$$X_B : \quad \mathbb{E}[u_B - r_f] + \lambda \cdot \int_{SF} (u_B - r_f) dG(u_A, u_B) = 0. \quad (2A.30)$$

These expressions are easily derived, as by definition shortfall is zero along the border of the area of integration  $SF$ . To visualize the area  $SF$ , define

$$\bar{u}_A = \frac{W^B - W_0 r_f + X_A r_f + X_B r_f}{X_A}, \quad (2A.31)$$



**Figure 2.5:** The area of integration,  $SF$ , as a function of  $\bar{u}_A$  and  $\bar{u}_B$ .

and

$$\bar{u}_B = \frac{W^B - W_0 r_f + X_B r_f + X_A r_f}{X_B}. \quad (2A.32)$$

The area of integration,  $SF$ , can now be represented as a triangular area where the points of the triangle are at  $(0, 0)$ ,  $(\bar{u}_A, 0)$ , and  $(0, \bar{u}_B)$ , as shown in Figure 2.5.

From the the definitions of  $\bar{u}_A$  and  $\bar{u}_B$  in (2A.31), and (2A.32), respectively, we can infer that the pairs  $(\bar{u}_A^*, \bar{u}_B^*)$  that solve the first-order conditions (2A.29) and (2A.30), directly determine the optimal  $X_A^*$  and  $X_B^*$ . Hence, solving (2A.25) boils down to finding the optimal  $\bar{u}_A^*$  and  $\bar{u}_B^*$ .

From the definition of  $\bar{u}_A$  and  $\bar{u}_B$ , write

$$\bar{u}_B - \frac{\bar{u}_B}{\bar{u}_A} r_f = -\frac{S_0 r_f}{X_B} + r_f. \quad (2A.33)$$

Assuming that  $X_A$  and  $X_B$  are both nonnegative at the optimum<sup>3</sup>, we find that

$$\bar{u}_B - \frac{\bar{u}_B}{\bar{u}_A} r_f = \begin{cases} < r_f & \text{if } S_0 > 0, \\ > r_f & \text{if } S_0 < 0. \end{cases} \quad (2A.34)$$

---

<sup>3</sup>We could have derived this from the second-order condition, but leave it to an assumption here. It is also quite obvious, given that the expected return on assets  $A$  and  $B$  is strictly larger than  $r_f$

Now observe that the expression  $\bar{u}_B - \bar{u}_B/\bar{u}_A r_f$  is exactly the point in Figure 2.5 on the y-axis where the line from  $(0, \bar{u}_A)$  to  $(0, \bar{u}_B)$  intersects with the vertical line at  $u_A = r_f$ . That means that the above definition can be visualized as asserting that for a positive surplus, the point  $(r_f, r_f)$  lies outside the grey triangle in Figure 2.5, while for a negative surplus  $(r_f, r_f)$  lies on the inside. This separates the possible solutions  $\bar{u}_A^*$  and  $\bar{u}_B^*$  to the first order condition, based on the sign of the surplus.

Assume that there are pairs  $(\bar{u}_A^+, \bar{u}_B^+)$ , and  $(\bar{u}_A^-, \bar{u}_B^-)$  that solve the system given by equations (2A.29) and (2A.30) for positive and negative surplus, respectively. In that case, we can gather terms involving  $X_A$  and  $X_B$  in the objective function (2A.27), which are multiplied by 0. What is left is the value of the objective function given by

$$V^*(W_0) = W_0 r_f + \lambda \cdot (W_0 r_f - W^B) \cdot \int_{SF} dG(u_A, u_B). \quad (2A.35)$$

For positive surplus, the expression for  $V^*$  in (2A.35) is maximal for the  $(\bar{u}_A^+, \bar{u}_B^+)$  that give the highest probability mass over the area of integration. In the case of a negative surplus,  $V^*$  is maximal for the area with smallest probability mass. A direct consequence is that the optimal solution is characterized by two pairs  $(\bar{u}_A^+, \bar{u}_B^+)$  and  $(\bar{u}_A^-, \bar{u}_B^-)$  for a positive and negative surplus, respectively, that give maximal objective value. Pairs  $(\bar{u}_A, \bar{u}_B)$  that solve the first order conditions do not have to be unique. For each sign of the surplus only the pair that gives the highest objective value determines the optimal solution.

The expressions for  $X_A^*$  and  $X_B^*$  follow from the definitions of  $\bar{u}_A$  in (2A.31) and  $\bar{u}_B$  in (2A.32). The  $k_A^*$  and  $k_B^*$  in Theorem 2.4.1 are given by

$$k_A^* = \bar{u}_A^* - \frac{\bar{u}_A^*}{\bar{u}_B^*} r_f, \quad (2A.36)$$

$$k_B^* = \bar{u}_B^* - \frac{\bar{u}_B^*}{\bar{u}_A^*} r_f, \quad (2A.37)$$

which completes the proof. ■

### ***n* risky assets**

For more than two risky assets, the proof follows along the same lines as that of Theorem 2.4.1. First order condition for risky asset  $i$  is given by

$$\mathbb{E}[u_i - r_f] + \lambda \cdot \int_{SF} (u_i - r_f) dG_i = 0, \quad i = 1, \dots, n, \quad (2A.38)$$

where  $SF$  is the set  $\{(u_1, \dots, u_n) \in \mathbb{R}^n | S_0 r_f + \sum_i^n (X_i (u_i - r_f)) < 0\}$ .

Define

$$\bar{u}_i = \frac{X_1 r_f + \dots + X_n r_f - S_0 r_f}{X_i}, \quad i = 1, \dots, n. \quad (2A.39)$$

Now, the area of integration  $SF$  is  $\mathbb{R}_+^n$ , bounded by the hyperplane through the points  $(\bar{u}_1 e_1, \dots, \bar{u}_n e_n)$ , where  $e_i$  is a vector of length  $n$  with zeros everywhere, and a 1 at position  $i$ .

We can deduce that

$$\bar{u}_i - \frac{\sum_{j \neq i} X_j^* r_f}{X_i^*} = -\frac{S_0}{X_i^*} r_f + r_f \begin{cases} \leq r_f & \text{if } S_0 \geq 0, \\ > r_f & \text{if } S_0 < 0, \end{cases} \quad (2A.40)$$

assuming the  $X_j^*$ s are all positive, compare equation (2A.34). From the definition of the  $\bar{u}_j$ s we have that at the optimum

$$X_j^* = X_i^* \cdot \bar{u}_i / \bar{u}_j, \quad (2A.41)$$

which we can use to rewrite (2A.40) as

$$\bar{u}_i^* \cdot \left( 1 - \sum_{j \neq i} \frac{r_f}{\bar{u}_j^*} \right) \begin{cases} \leq r_f & \text{if } S_0 \geq 0, \\ > r_f & \text{if } S_0 < 0. \end{cases} \quad (2A.42)$$

When expression on the left-hand side of (2A.42) is equal to  $r_f$ , the hyperplane that limits the set  $SF$  goes through the point with coordinate  $(r_f, r_f, \dots, r_f)$ . Hence, the set of solutions can be split in solutions that are valid for positive and those that are valid for negative surplus. The  $\bar{u}_i^*$ s that give the highest objective value for each sign of the surplus solve the optimization problem with  $n$  risky assets.



---

# Extensions to the Mean-Shortfall Model

---

## 3.1 Introduction

The previous chapter has presented the optimal solution to the mean-shortfall model. The optimal decision(rule) took the form of a V-shape, with risky investments increasing in the extent of being above or below a pre-specified benchmark wealth level. However, in practice and theory, other specifications of loss averse preferences are used. Therefore, the aim of this chapter is to see whether the V-shape found in Chapter 2 is only the result of our specific choice of the bilinear objective, or whether it results from a larger class of loss averse objectives.

To test the robustness of the results, we generalize the model in 4 directions relevant empirically. First, we explore the consequences of taking the value function estimated by Kahneman and Tversky (1979) as the objective. This objective function is closely related to the mean-shortfall objective of Chapter 2, and is referred to in the literature as ‘the behavioral value function’. The most important difference with the bilinear objective is that the marginal penalty on a shortfall is decreasing in the extent of the loss. Section 3.2 explores the properties of this objective function, and its results when used in an optimization. We discuss the appropriateness of its use in finance and use our results to shed new light on some well-known papers in behavioral finance.

Second, Section 3.3 establishes results on the optimal investment strat-

egy as a function of wealth if the marginal penalty on losses increases in the magnitude of a loss. Such risk-measures, e.g. downside deviation, are predominantly used in ALM studies. For financial institutions, the effect of increased risk-taking when wealth decreases, is seen by many as undesirable. In the same section we also want to increase understanding of the outcomes that describe behavior in unfavorable circumstances.

Third, in Section 3.4 we add an extra ‘kink’ to the bilinear objective function. The additional kink reflects an increased aversion to shortfall with respect to a risk-free investment on top of the aversion to shortfall with respect to the pre-specified benchmark.

Fourth, in Section 3.5 we address the issue of restrictions on the asset allocation. The original model can lead to allocations in one asset category of far more than 100%. This is not practically feasible for most economic agents. It can also create a positive probability of negative wealth. The consequences of limiting the allowable investment fractions are explored in Section 3.5. Section 3.6 concludes.

Note that this chapter focuses on generalizations of Chapter 2 in terms of the objective function and allowable investment strategies. An alternative direction of generalization of Chapter 2 concerns the specification of uncertainty in the model. This is left to Chapter 4, where extra uncertainty is introduced in a framework that addresses the economic relevance of the loss averse model for pension funding.

## 3.2 The behavioral value function

The previous chapter already presented much of the evidence that is found in the literature on loss aversion. Specifically, we saw that modeling utility in terms of gains and losses already dates back to Markowitz (1952). Recent literature mostly refers to the work of Kahneman and Tversky (1979). First, they introduced a transformation function that transforms objective probabilities into subjective probabilities, bringing about a deviation from expected utility. Second, they estimated specific shapes for the utility function below and above a reference point. In experiments they found that subjects were risk loving in losses, and risk averse in gains. That is, with respect to gains people rather prefer a sure gain to an uncertain gamble with the same expected pay-off. For losses, people prefer the gamble with uncertain pay-off to the sure loss. The typical function that has these properties is generally referred to as the ‘behavioral value function’, a term we will use as well. Considering the transformation of probabilities is beyond the scope of this chapter. We assume in the following that the optimizer knows and

uses the objective probabilities.

The possible existence of a behavioral function has serious implications for situations in which people make decisions under risk. When a substantial loss is suffered, people can have the tendency to increase risk-taking, even when the expected return is negative. The latter is documented in for example Thaler and Johnson (1990), who find that when decision makers have prior losses, outcomes which offer the opportunity to “break even” are especially attractive. Going further, Shefrin and Statman (1985) pose that professional investors are also risk-loving with respect to losses. Investors have a “disposition to ride losers too long”, i.e., holding on to stocks with decreasing market value too long. The explanation in their paper is based on the ‘S’ shape of the value function to be introduced further below in (3.1). If this explanation is correct, and the S shape is indeed a distinguishing feature compared to the bilinear approximation we considered in the previous chapter, there is all the more reason to investigate what optimal investment policies with these preferences look like in the framework of Chapter 2. At the end of this section we examine whether the S shaped value function results in an aversion to the realization of a losses claimed by Shefrin and Statman.

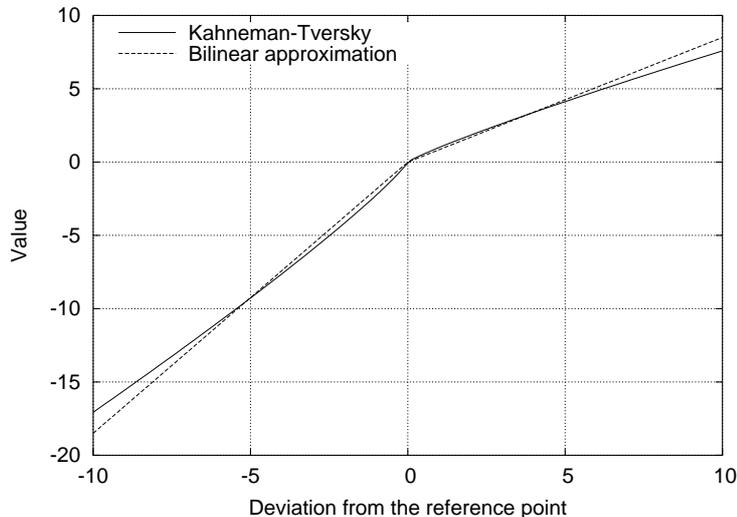
We start our analysis of the consequences of using the behavioral value function by considering the value function as estimated by Tversky and Kahneman (1992). With  $x$  the deviation from the reference point, a value function  $v(\cdot)$  assigns a subjective valuation to any outcome  $x$  according to

$$v(x) = \begin{cases} x^\alpha & \text{if } x \geq 0, \\ -\gamma \cdot (-x)^\beta & \text{if } x < 0, \end{cases} \quad (3.1)$$

where  $\alpha = \beta = 0.88$ , and  $\gamma = 2.25$ . The bilinear formulation of the objective function in Chapter 2 can be considered a special case of this formulation by setting  $\alpha = \beta = 1$ . In Figure 3.1 a plot of the value function (3.1) against a fit of the piecewise-linear approximation, (2.2) of Chapter 2, is shown.

A first observation from Figure 3.1 is that the fit of the approximation on the value function (3.1) is very good. Any bilinear function can be written in the form of a mean-shortfall model around the reference point. The observed fit does not depend on the choice of scale, a consequence of the basic power functions of which the value function is constructed.

A second observation concerns the shape of the value function. From the formula in (3.1), and given that  $\alpha$  and  $\beta$  are both smaller than 1, it follows that the function has an S-shape. In Figure 3.1 this is not clear to see for gains, but it is visible for negative deviations from the reference point. The convexity in the domain of losses implies risk-loving behavior in losses, while



**Figure 3.1: The Kahneman-Tversky value function**

The solid line plots the Kahneman-Tversky value function as given in (3.1). The dashed line plots a bilinear approximation that has a slope of 0.85 to the right of the reference point, and of -1.85 on the left-hand side.

the concavity on the side of gains represents risk-averse behavior in gains.

To see the impact in terms of optimal investment decisions, we analyze a one-period investment model with the same setup as in Chapter 2, but with a different objective function. Using the definition of the value function  $v(\cdot)$  in (3.1), the optimization problem is given by

$$\max_{X_0} \mathbb{E}[v(W_0 r_f + X_0(u - r_f) - W^B)], \quad (3.2)$$

where, as in Chapter 2,  $W_0$  is the initial wealth,  $r_f$  the risk-free (gross) rate of return,  $X_0$  the amount of stock investment,  $u$  the uncertain return on stocks and  $W^B$  the benchmark level of wealth. This problem is again the most basic representation of optimal investment under uncertainty.

The solution to problem (3.2) is analyzed for a stock return distribution  $G(\cdot)$  with lognormal distribution that has an expected return of 10% and standard deviation of 17%, representing plausible parameter values for historical stock returns. The risk-free rate is 1.04, the benchmark  $W^B$  is 104. This way, an initial wealth level of 100 is needed to reach the benchmark  $W^B$  with certainty by investing only in the risk-free asset. We solve (3.2) by discretizing the return  $u$  and optimizing numerically.

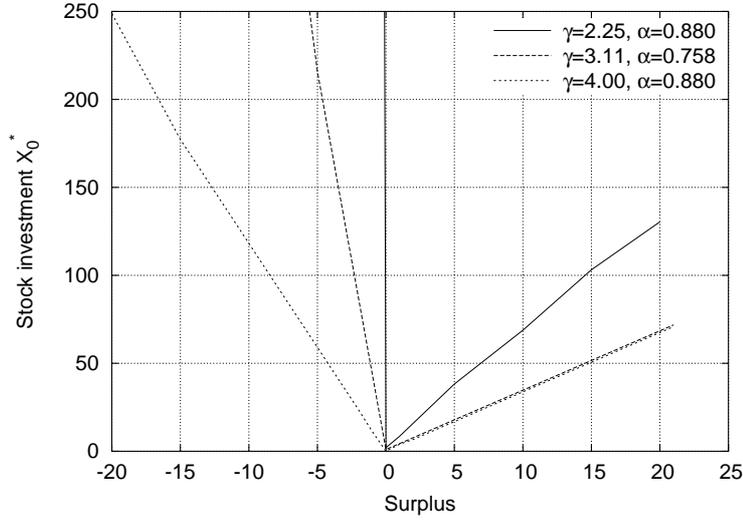
The first result of solving (3.2) is that for nonpositive surpluses the optimal solution is unbounded, i.e., the optimal stock investment is infinite. So either the distributional assumptions for  $G(\cdot)$  and  $r_f$  are not realistic, or the setup of the investment problem in (3.2) is not comparable with the setting in which the value function is estimated. The first possibility can be ruled out, as the assumption on the risk-free rate and stock-return are not extraordinary, see Siegel (1992) for an overview of averages and standard deviations for stock returns, measured over different time periods. Rather, the Kahneman-Tversky value function with parameter values estimated in Tversky and Kahneman (1992) is probably not an adequate reflection of aggregate preferences of stock market participants. Note that for positive surpluses the optimal solution is bounded, which will be presented shortly.

Having established that the value function as estimated by Tversky and Kahneman (1992) is not appropriate for the basic investment model, it is too early to discard the general S-shaped value function completely, as it is possible that the estimated parameter values are just not suitable for the investment context. To explore the investment behavior induced by an S-shaped value function in non-degenerate cases, the results of Shumway (1997) are useful. There,  $\gamma$  and  $\alpha$  in (3.1) are estimated using stock return data. Shumway finds estimates  $\hat{\gamma} = 3.11$ , and  $\hat{\alpha} = 0.758$ . These parameter values give bounded solutions for a relevant range of surpluses, but lead to unrealistically high risky investments for slight underfunding. Our optimization framework illustrates that parameter values estimated from aggregate data may lead to unrealistic behavior at the level of individual optimizing agents. As the stock investments for negative surpluses are extremely high, we add a third set of parameter values, namely  $\gamma = 4$ , and  $\alpha = 0.88$ . The optimal stock investment as a function of the surplus for all three parameter combinations is displayed in Figure 3.2.

Looking at the results in Figure 3.2, we see that the parameter values as estimated by Shumway (1997) lead to extremely high stock investment for negative surpluses. This is a worrying result, as the parameter values were not estimated in a laboratory experiment, but were estimated using actual stock return data. The outcomes suggest that realistic risk preferences cannot be inferred from stock price behavior as done by Shumway, or that the specification of the behavioral value function is flawed.

The decision rule for the parameter set with  $\gamma = 4$  clearly shows a V-shape. We already found that the bilinear objective function is a good approximation to the behavioral value function. Figure 3.2 shows that result to carry over to the evaluation of the shape of the decision rule.

An additional check to see if the left- and right-hand sides of the V in the figure are truly straight lines is to consider the probability of underfunding,



**Figure 3.2: Optimal solution for a behavioral value function**

This figure shows the optimal level of stock investment  $X_0^*$  using Kahneman and Tversky's value function with different parameter values for  $\gamma$  and  $\alpha$ . The parameter values  $\gamma = 2.25$ ,  $\alpha = 0.88$  (solid line), are estimated in Tversky and Kahneman (1992). We use these parameters only for positive surplus, giving unbounded solutions for surplus  $\leq 0$ . Parameter values  $\gamma = 3.11$ ,  $\alpha = 0.758$  are from Shumway (1997). The horizontal axis gives the value of the surplus  $S_0 = W_0 - W^B/r_f$ . The solution is computed numerically with 100 discretizations for the stock return distribution, lognormally distributed with parameters  $\mu = 0.085$ ,  $\sigma = 0.16$ .

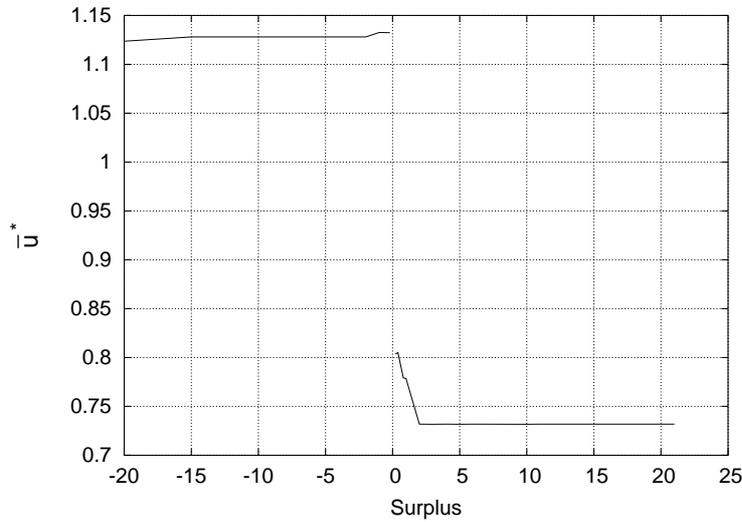
or, equivalently, the return  $\bar{u}$  that gives final wealth of exactly  $W^B$ . By definition

$$\begin{aligned} \bar{u} &= \frac{W^B - W_0 r_f}{X_0} + r_f \\ &= r_f - r_f \cdot S_0 / X_0, \end{aligned} \tag{3.3}$$

see definition (2A.4) in Chapter 2. From this definition it is immediately clear that whenever  $X_0^*$  is linear in the surplus  $W_0 - W^B/r_f$ , the associated  $\bar{u}$  is a constant. To check the linearity of the V-shape, Figure 3.3 shows the values of  $\bar{u}^*$  as a function of  $W_0$ .

We see that the values of  $\bar{u}^*$  are almost constant on either side of the discounted benchmark, with a very small area around zero surplus where the  $\bar{u}^*$ s approach 1. For absolute levels of the surplus larger than 2, the lines are horizontal straight lines.

Note that the observed similarity of the shape of the decision rules with those of the previous chapter does not mean that the optimal fraction in stocks is the same as for the bilinear objective. The actual solution depends



**Figure 3.3: Values of  $\bar{u}^*$**

This figure shows  $\bar{u}^* = r_f - r_f \cdot S_0/X_0^*$ , against different values of initial surplus,  $S_0$ . There is a discontinuity at zero, where  $\bar{u}$  is undefined.

on the specific parameter values chosen.

### Ride losers too long, sell winners too early

Given the results derived so far, we now consider whether the S shape of the behavioral value function results in the behavior suggested by Shefrin and Statman (1985). As we want to discuss their claim in detail, consider the following citation quote in which they motivate their claim:

*“To see how the disposition to sell winners and ride losers emerges in prospect theory, consider an investor who purchased a stock one month ago for \$50 and who finds that the stock is now selling at \$40. The investor must now decide whether to realize the loss or hold the stock for one more period. To simplify the discussion, assume that there are no taxes or transaction costs. In addition, suppose that one of two equiprobable outcomes will emerge during the coming period: either the stock will increase in price by \$10 or decrease in price by \$10. According to prospect theory, our investor frames his choice as a choice between the following two lotteries:*

A. Sell the stock now, thereby realizing what had been a \$10

*“paper loss.”*

*B. Hold the stock for one more period, given 50-50 odds between losing an additional \$10 or “breaking even.”*

*Since the choice between these lotteries is associated with the convex portion of the S-shaped value function, prospect theory implies that B will be selected over A. That is, the investor will ride his losing stock. An analogous argument demonstrates why prospect theory gives rise to a disposition to realize gains.”*

The importance of the quoted citation is that the reasoning is very straightforward, and appeals to intuition. Having a utility function that is convex in some area induces excessive risk-taking. Too much concavity in gains induces profit-taking behavior. Two problems arise, however, after having read the above claim. First, it is unclear to what benchmark case the terms ‘too long’ and ‘too short’ refer. It implies that the authors have an idea of normative optimality, which is not met by the investment strategy following from the behavioral value function. Second, it is not known under what conditions the claimed effect will or will not arise. We leave the first point for what it is, interpreting the disposition as ‘riding losers long, selling winners early’. This way the problem of having to devise a proper benchmark case is neutralized. We concentrate on modeling buy and sell behavior in an investment model with a behavioral value function. As a side note, observe that Shefrin and Statman’s use of the term prospect theory only considers the use of the value function, and not the transformation of probabilities, so their claim should hold under expected utility maximization.

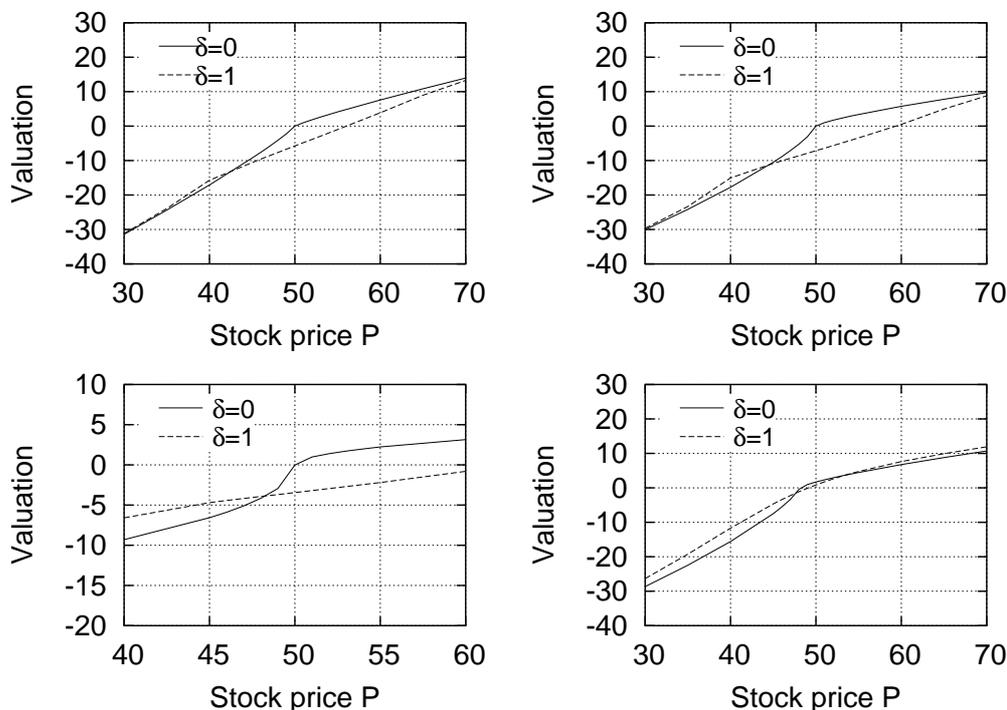
To analyze Shefrin and Statman’s claim, we solve an optimization model that matches the cited investment problem. Define again  $v(\cdot)$  as the Kahneman-Tversky value function defined in (3.1). The optimization problem is given by

$$\max_{\delta} \quad v(S), \tag{3.4}$$

$$\text{s.t. } S = \delta \cdot P_1 \cdot u + (1 - \delta) \cdot P_1 r_f - P_0, \tag{3.5}$$

$$\delta \in \{0, 1\}, \tag{3.6}$$

where  $P_0$  represents the initial stock price of \$50,  $P_1$  is the current price of the stock, which is \$40 in the citation quote above. The decision variable  $\delta$  represents the decision to either hold or sell the stock. If  $\delta = 1$ , the investor holds the stock and receives  $P_1 \cdot u$  in the next period. If  $\delta = 0$ , the proceeds of selling the stock are put in the risk-free asset, realizing proceeds of  $P_1 \cdot r_f$ . Consequently, the variable  $S$  represents the surplus relative to the benchmark at the end of the next period.



**Figure 3.4: Comparison of holding and selling the stock**

For holding ( $\delta = 1$ ) and selling ( $\delta = 0$ ) the stock, these figures give the prospective value as a function of the current stock price  $P$ .  $P = 40$  is the point of interest in the example of Shefrin and Statman (1985). The top left panel gives the result for the parameters that match Shefrin and Statman's setup, i.e., the behavioral value function given in (3.1) with  $\gamma = 2.25$  and  $\alpha = 0.88$ ,  $u$  is either 1.25 or 0.74 with probability 0.5, and  $r_f = 1$ . The top right panel has  $\gamma = 3.11$ ,  $\alpha = 0.758$ . Lower left panel has  $\gamma = 3.11$ ,  $\alpha = 0.5$ . Lower right panel has  $\gamma = 3.11$ ,  $\alpha = 0.758$ ,  $r_f = 1.04$ , and  $u \sim \log N(0.085, 0.16)$ .

Figure 3.4 shows the value of the objective for the two possible decisions for  $\delta$  as a function of the current price  $P_1$  of the stock. Because we want to analyze the sensitivity of the outcome for the choice of parameters, Figure 3.4 plots the results for different sets of parameter values.

The situation  $P_1 = 40$  in the top left panel gives the results for the exact situation presented by Shefrin and Statman (1985). In this panel the stock return is +25% or -25%, both with probability 1/2. For  $P_1$  equal to \$40, this matches the specification of uncertainty that is considered by Shefrin and Statman. Also, the parameter values in the value function  $v(\cdot)$  are those estimated by Tversky and Kahneman (1992). In the panel, we see that for a small region below and for a larger region above a stock price of 50, the line for  $\delta = 0$  lies above that of  $\delta = 1$ , i.e., it is optimal to sell the stock. For  $P_1 < 43$ , holding the stock is just slightly better than

selling it. Hence, we must conclude that the claim that the disposition effect is a straightforward consequence of the estimated behavioral value function by Tversky and Kahneman (1992), is incorrect. There is a large range of stock prices for which selling is optimal, but it is not confined to ranges above the initial price of 50. The difference in objective for the range in which holding the stock is optimal, appears too small to be of any significance in an actual decision situation.

Again, it is possible that the parameter values  $\gamma = 2.25$  and  $\alpha = 0.88$  are not suitable for an investment situation. Therefore, the upper right panel shows the solutions when only the parameter values are changed to those estimated by Shumway (1997). The pattern of the solution remains the same.

It is possible that the ex-ante intuitive appeal of Shefrin and Statman's claim lies in the way they visualize the S-shape, suggesting it to be more curved than is justified for realistic parameter values. Therefore, the lower left panel has parameter values  $\gamma = 3.11$ ,  $\alpha = 0.5$ . These values give the value function a distinct S-shape, as can be observed from the solid line for  $\delta = 0$  in the figure. For this strongly curved objective function the panel reveals the behavior that is worded in the citation quote above. For stock prices below 48, holding the stock is clearly preferred. Above  $P_1 = 50$ , the valuation for selling the stock clearly exceeds that of holding on to it.

To complete the analysis, the lower right panel gives the results for a stock return distribution that is more realistic than the distribution that was used for the other panels. In contrast with the upper panels, here the largest difference between holding and selling is in the range of losses. Below 50, holding on to the stock gives significantly higher value. For prices above 50, however, the difference in valuation between the two outcomes is only marginal. Concluding, the results from our straightforward investment model does not give firm support nor completely discards the claim from Shefrin and Statman. It shows that the outcome crucially depends on the parameter values and specification of uncertainty that one wants to consider. Added to the V-shaped optimal decision rules that were found at the beginning of this section, we have shown that for more realistic distributions of stock returns the behavior that they assume does not really exist.

We conclude our discussion with two problems associated with the actual application of the value function in the setting of financial planning, namely representativeness and normative desirability. With regard to the first, we doubt whether the laboratory setting in which Tversky and Kahneman (1992) estimated the value function is representative for a typical investment situation. In the laboratory experiments, 25 graduate students from Berkeley and Stanford university “with no special training in decision theory” were

selected. Each student was paid \$25 for participation. Many decision makers in investment, however, are professionals that have been trained and/or are experienced to make investment decisions given the uncertainty of financial markets. Also, the amounts at stake in investment are generally a large multiple of \$25. Moreover, it is difficult to imagine students participating in an experiment in which it is possible to actually suffer a substantial monetary loss. However, this possibility is a reality in many investment situations.

Another issue is whether the decisions people take are the result of rational thinking, or that in some situations people actually lack “self-control”, a term used by Shefrin and Statman (1985). See also Hirshleifer (2001), who discusses the role of self-control and social interactions in financial decisions. Thaler and Shefrin (1981) go as far as modeling the lack of self-control as part of an agency problem, boiling down to an intra-personal conflict between a rational part and an emotional part of one’s self. Specifically with regard to the value function (3.1), it is difficult to imagine this value function being selected as an objective function in an optimization framework for a financial planner. The behavioral value function only seems useful for empirical research in finance, and not for the normative area of financial planning. In that situation, a decision maker would be forced to make a normative statement on suitable preferences, and it seems more likely that he would select an objective function that has an increasing marginal penalty on losses. This idea is confirmed by the widespread use of downside deviation as risk measure in applied Asset/Liability Management studies. The next section will consider a downside risk measure that puts an increasing marginal penalty on losses.

### 3.3 Quadratic shortfall

The analysis in Chapter 2 involved a downside risk measure that was linear in shortfall. Subsection 3.2 explored the consequences of replacing the linear penalty with a convex penalty in losses. In this section, we analyze a downside risk-measure that is concave in losses, i.e., it puts an increasing marginal penalty on losses. Such risk measures overcome the conceptual problems on representativeness and normative desirability we raised in the previous section. For that reason, they are heavily used in Asset/Liability Management and other areas of investment.

Using a risk measure that is concave in losses is recommended in for example Sortino and Van der Meer (1991) and Harlow (1991). Specifically, they argue that downside deviation, or semideviation, as it is called by Markowitz (1959), is the risk measure to be used in investment. It is also in Boender

(1997) and in the practice of Asset/Liability Management, see Ziemba and Mulvey (1998).

Our aim is to determine if and how the V-shaped asset allocation policies persist if the risk measure becomes quadratic shortfall. To do so, we analyze a one-period investment model as before, but now with an objective function that penalizes squared shortfall below a benchmark level of wealth. The formulation of the objective is

$$\max_{X_0} \mathbb{E}[W_1] - \nu \cdot \mathbb{E}[(\{W^B - W_1\}^+)^2]. \quad (3.7)$$

The transition of  $W_1$  is again given by

$$W_1 = W_0 \cdot r_f + X_0 \cdot (u_1 - r_f), \quad (3.8)$$

where  $W^B$  represents benchmark wealth,  $W_1$  final wealth,  $X_0^*$  the investment in the risky asset,  $u$  the return on the risky asset, and  $r_f$  the risk-free rate. The parameter  $\nu$  represents loss aversion with respect to the level  $W^B$ , and has a similar interpretation as  $\lambda$  in the mean-shortfall model of Chapter 2.

Figure 3.5 shows the objective in (3.7) as a function of final wealth  $W_1$ , together with the mean-shortfall objective from Chapter 2. In the figure, we can see that the fit is not as good as for the Kahneman-Tversky objective in Figure 3.1. The quadratic shortfall objective has a distinctly different (concave) shape on the loss side.

Define  $S_0 = W_0 - W^B/r_f$  as the level of initial surplus. The main results for the model with squared shortfall objective are in the following theorem.

**Theorem 3.3.1** *For  $\nu > 0$ , the optimal solution  $X_0^*$  to problem (3.7) has the following properties:*

- (i) *The optimal stock investment  $X_0^*$  for zero surplus ( $S_0 = 0$ ) is positive, decreasing in  $\nu$ , and given by*

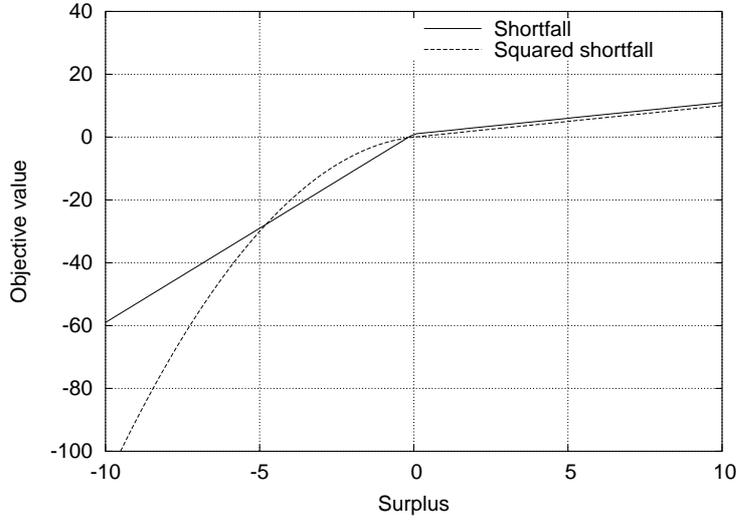
$$X_0^* = \frac{\mathbb{E}[u - r_f]}{2\nu \cdot \mathbb{E}[(r_f - u)^+]^2}. \quad (3.9)$$

- (ii) *The decision rule for  $X_0^*$  as a function of  $S_0$  has one minimum for a non-positive surplus. For any  $\nu$ , the location of the minimum is characterized by*

$$X_0^{min} = \frac{r_f}{r_f - \bar{u}^m} S_0^{min}, \quad (3.10)$$

where  $\bar{u}^m$  is the  $\bar{u} > r_f$  that solves

$$\int_0^{\bar{u}} (u - r_f) dG = 0. \quad (3.11)$$



**Figure 3.5: Squared shortfall vs. linear shortfall**

This figure plots the mean-shortfall objective and mean-squared shortfall objective functions as a function of the final surplus  $S_1$ .  $\lambda = 5, \nu = 1$

For surplus lower than  $S_0^{min}$ ,  $X_0^*$  is decreasing in  $S_0$ . Above  $S_0^{min}$ ,  $X_0^*$  is increasing in  $S_0$ .

- (iii) For a positive surplus, and  $\nu \rightarrow \infty$ ,  $X_0^* = S_0$ .  
For a negative surplus, and  $\nu \rightarrow \infty$ ,  $X_0^*$  is given by

$$X_0^* = \frac{r_f}{r_f - \bar{u}^*} S_0, \quad (3.12)$$

where  $\bar{u}^* > r_f$  is the  $\bar{u}$  that solves

$$\int_0^{\bar{u}} (r_f - u) dG = \frac{1}{r_f - \bar{u}} \int_0^{\bar{u}} (u - r_f)^2 dG. \quad (3.13)$$

Moreover, for a negative surplus,  $X_0^*$  is smaller than for the mean-shortfall objective.

**Proof:** See appendix.

Theorem 3.3.1 gives specific properties of the solution, without specifying a parametric decision rule for the optimal  $X_0^*$  as a function of  $S_0$ . In the following, we discuss the results one by one, comparing them with the results

in Theorem 2.3.1 that gives the optimal solution to the mean-shortfall model. Remarkably, the existence of a bounded solution does not depend on a lower bound on  $\nu$ , due to the concavity in losses.

First, equation (3.9) shows that the optimal investment in the risky asset is non-zero at zero surplus. Looking at Figure 3.5, we see that this is caused by the marginal penalty on a loss being very small for losses close to zero. This is different from the mean-shortfall objective, where the marginal penalty is constant regardless of the distance to the benchmark. A large value of  $\nu$ , however, will drive  $X_0^*$  to zero.

Second, the implication of Theorem 3.3.1(ii) is that the optimal investment in the risky asset has a V-shaped<sup>1</sup> relation with the surplus  $S_0$ . For different values of  $\nu$ , the minima of the V-shapes lie on a straight line through the point with  $S_0 = X_0^* = 0$ . Equation (3.11) resembles the first-order condition for the mean-shortfall model in equation (2.4). Given that for the mean-shortfall model  $X_0 = 0$  for  $S_0 = 0$ , and given the non-zero minimum investment  $X_0^{min}$  for the quadratic shortfall case, the implication is that for each value of  $\nu$  and  $\lambda$  there is an area around  $S_0 = 0$ , where the investment in the risky asset is higher in the squared-shortfall model than in the linear shortfall model.

The limiting behavior for  $X_0^*$  when  $\nu \rightarrow \infty$  in (iii) shows that for large values of  $\nu$ , both sides of the V-shape become linear in the surplus. For positive surplus, the optimal  $X_0^*$  is just equal to the surplus, as the aversion to downside risk has become too large. For negative surplus, however, the solution converges to an allocation that gives minimum quadratic shortfall. Although smaller than in the linear shortfall case, the investment in the risky asset does not go to zero for negative surplus. Such behavior may be desired by regulators. Ex ante intuition might say that high loss aversion induces low-risk allocations. This is clearly not true if one starts in a situation of shortfall.

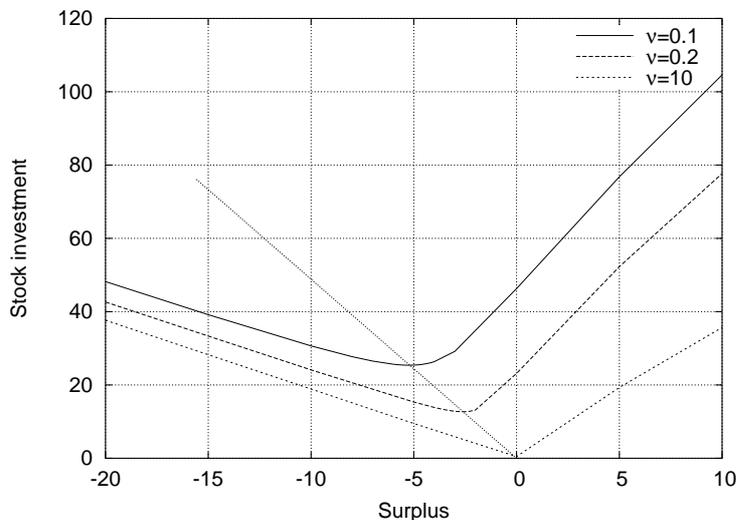
Figure 3.6 shows the numerical solutions to model (3.7) given three different values of  $\nu$ . It illustrates the results in the theorem, and also shows the typical shapes of the optimal decision rules, together with an idea of the sensitivity of the optimal solution to  $\nu$ .

With respect to the V-shapes in Figure 3.6, we see that the the right-hand side of the V is more sensitive to changes in  $\nu$  than the left-hand side. For  $\nu = 10$  we see a piecewise linear V-shape, the limiting behavior suggested by the theorem.

Compared with linear shortfall as a risk measure, squared shortfall in-

---

<sup>1</sup>Strictly speaking, the relation does not have to be V-shaped, but we will use the term for consistency, denoting a general U-shaped figure.



**Figure 3.6: Solution for the quadratic-loss objective**

This figure shows the optimal stock investment as a function of initial surplus for the quadratic-loss objective and different values of  $\nu$ . Benchmark is  $W^B = 104$ ,  $r_f = 1.04$ , and  $u \sim \text{logN}(0.085, 0.16)$ . The added line goes through the minima of the decision rules.

duces less extreme-risk allocations for surplus values not too close to 0. Even for very small values of loss aversion, the slopes of the optimal decision rules remain quite stable, while for linear shortfall the slopes of the V-shape are very sensitive to changes in loss aversion. Using squared shortfall as a risk measure ensures a smoother investment policy than with linear shortfall.

### 3.4 Extra kink

Until now, the objective functions all had one feature in common, having one benchmark or reference point that represented a target level of wealth. The drawback of limiting the formulation to such objectives is that it does not take into account possible aversion against future shortfall below current wealth. The existence of such aversion becomes evident in the popularity of financial products that give a guarantee on (part of) the invested amount. Therefore, in this section an extra benchmark is added to the mean-shortfall objective as introduced in Chapter 2. The added benchmark lies at the risk-free wealth level in the next period. Again formulated as a one-period model,

the objective becomes

$$\max_{X_0} \mathbb{E}[W_1] - \lambda_1 \cdot \mathbb{E}[(W^B - W_1)^+] - \lambda_2 \cdot \mathbb{E}[(W_0 r_f - W_1)^+], \quad (3.14)$$

where  $\lambda_2 \geq 0$  is a new penalty parameter, penalizing the shortfall below the risk-free wealth level  $W_0 r_f$ .  $\lambda_1$  represents loss aversion with respect to  $W^B$ , the benchmark level of wealth, which is independent of initial wealth  $W_0$ . The mean-shortfall objective of Chapter 2 is obtained from (3.14) by setting  $\lambda_2 = 0$ . For a positive  $\lambda_2$  the objective function can be visualized as a piece-wise linear increasing function in  $W_1$  with two kinks. The next theorem presents the solution.

**Theorem 3.4.1** *With  $S_0$  defined as  $W_0 - W^B/r_f$ , the solution to (3.14) is given by*

$$X_0^* = \frac{r_f}{r_f - \bar{u}} S_0, \quad (3.15)$$

where  $\bar{u}^*$  is one of the  $\bar{u}$ s that solve

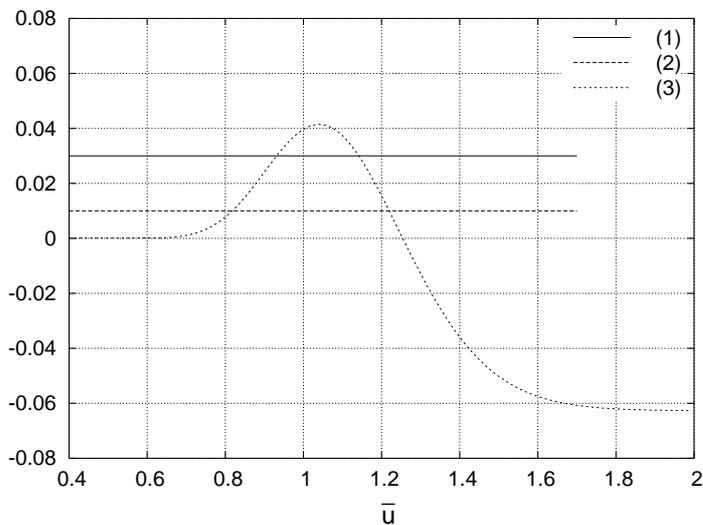
$$\frac{\mathbb{E}[u - r_f]}{\lambda_1} - \frac{\lambda_2}{\lambda_1} \int_0^{r_f} (r_f - u) dG = \int_0^{\bar{u}} (r_f - u) dG. \quad (3.16)$$

For  $S_0 < 0$ ,  $\bar{u}^* = \bar{u}^- > r_f$ , for positive  $S_0$ ,  $\bar{u}^* = \bar{u}^+ < r_f$ .

**Proof:** See appendix

Theorem 3.4.1 shows that the inclusion of an extra kink in the objective at  $W_0 r_f$ , retains analytical tractability. Compared with the first-order condition for the mean-shortfall model (2.4), the first order condition (3.16) differs only by a constant term, which is the second term on the left-hand side of (3.16).

We have visualized the solution to equation (3.16) in Figure 3.7. If  $\lambda_2 = 0$ , we are back in the situation of the mean-shortfall model, and the points of intersection with lines (1) and (3) in the figure give the optimal  $\bar{u}^*$ s. For  $\lambda_2 > 0$ , the dashed line (2) decreases with increasing  $\lambda_2$ , and the points of intersection give the  $\bar{u}$ s that solve (3.16). For  $\lambda_2$  sufficiently large, we can drive  $\bar{u}^+ < r_f$  equal to zero, implying that for a positive surplus the investment in the risky asset is just the level of the surplus. If  $\lambda_2$  increases even more, there is no  $\bar{u}^+ < r_f$  that solves (3.16). The risky investment stays unchanged at the surplus as a corner solution to (3.14). For  $\bar{u}^- > r_f$ , increasing  $\lambda_2$  drives  $\bar{u}^-$  to infinity. The limiting point for  $\lambda_2$  lies at the point where the left-hand side of (3.16) is equal to  $-\mathbb{E}[u - r_f]$ . In this case, with  $\bar{u}^- = \infty$ , the optimal investment in the risky asset is zero, just as for larger



**Figure 3.7: Solving the first-order condition**

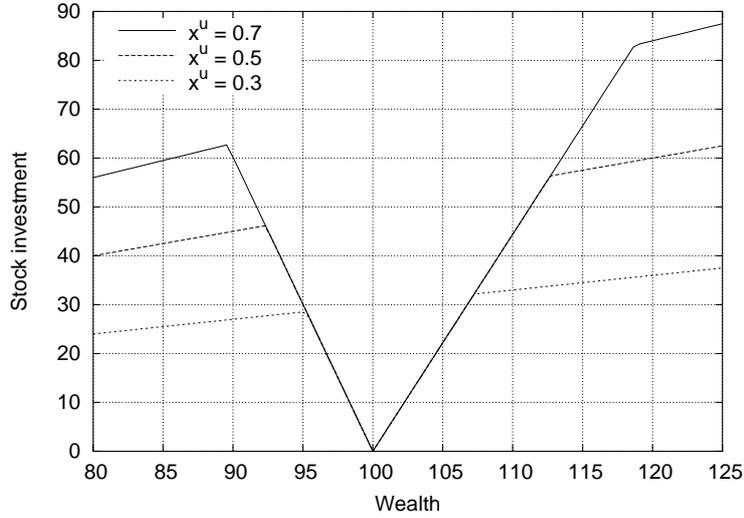
This figure visualizes the solution to equation (3.16). In the figure, line (1) is the first term on the left-hand side, line (2) is the sum of the terms on the left-hand side, the curved line (3) is equal to the right-hand side. Parameter values are  $r_f = 1.04$ ,  $u \sim \log N(0.085, 0.16)$ ,  $\lambda_1 = 2$ , and  $\lambda_2 = 1.33$ .

values of  $\lambda_2$ . Hence, by increasing the penalty parameter  $\lambda_2$  for the second benchmark  $W_0 r_f$ , the left side of the V can become as flat as desired. The intuition is clear: if the penalty on losing money with respect to the risk-free investment goes to infinity, the optimal action is to invest all wealth in the risk-free asset. This amounts to taking ones losses. This is different than in the linear and squared shortfall case, where  $\lambda \rightarrow \infty$  and  $\nu \rightarrow \infty$  induce a minimum-risk allocation that contains a non-zero investment in the risky asset.

### 3.5 Restrictions on asset allocations

There may exist institutional restrictions that limit stock investment to a certain maximum fraction of total assets. In most Asset/Liability Management models such restrictions are commonplace.

We study the consequences for the solutions to the mean-shortfall model from Chapter 2 of imposing restrictions on the asset mix. Recall that the



**Figure 3.8: Optimal one-period solution to the restricted model**

For a one-period model, this figure shows the optimal stock investment as a function of initial surplus under restriction (3.19).  $W^B = 104$ , the stock return  $u$  is distributed  $\log N(0.085, 0.16)$ ,  $r_f = 1.04$ ,  $\lambda = 5$ .

optimization model is given by

$$\max_{X_0, \dots, X_{T-1}} \mathbb{E}_0[W_T] - \lambda \cdot \mathbb{E}_0[(W_T^B - W_T)^+], \quad (3.17)$$

$$\text{s.t. } W_{t+1} = W_t \cdot r_f + X_t \cdot (u_{t+1} - r_f), \quad t = 0, \dots, T-1, \quad (3.18)$$

see Section 2.2.

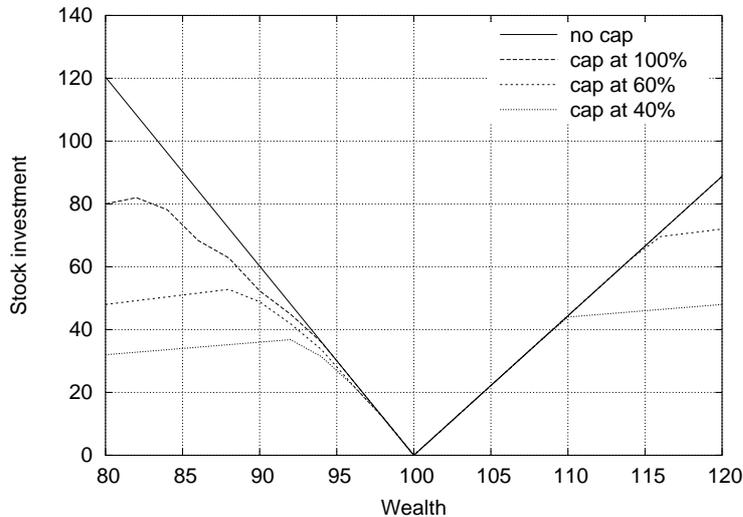
Define a cap on the fraction invested in the risky asset, denoted by  $x^u$ . In a one-period model, i.e.,  $T = 1$ , the restriction to have a maximum fraction of stocks is formulated as

$$X_0 \leq x^u \cdot W_0. \quad (3.19)$$

For such a simple one-period restriction, the consequences are straightforward. If the unrestricted optimal fraction  $X_0^*/W_0$  is strictly smaller than  $x^u$ ,  $X_0^*$  is also the optimal solution for the restricted model. If it is larger than  $x^u$ , the restriction in (3.19) becomes binding, and  $X_0^* = x^u \cdot W_0$ . For different values of  $x^u$ , optimal solutions to the restricted model are visualized in Figure 3.8.

However, if  $T > 1$  and the restriction on the asset mix holds for any time  $t$  during the planning period, i.e.,

$$X_t \leq x^u \cdot W_t, \quad t = 0, \dots, T-1, \quad (3.20)$$



**Figure 3.9: Optimal solution at  $t = 0$  for the permanent cap.**

For a two-period model, this figure presents the amount invested in the risky asset at time 0 for different permanent caps on the fraction invested in the risky asset. Presented is the solution to a model with  $u_1, u_2$  both distributed  $\log N(0.085, 0.16)$ ,  $W^B = 108.16$ ,  $r_f = 1.04$ , and  $\lambda = 10$ .

then the resulting  $X_t^*$  can not be determined in a straightforward manner for all  $t$ . At time  $T - 1$ , the solution is still the capped optimal solution at  $x^u \cdot W_{T-1}$ . At times  $t < T - 1$ , however, the optimal solution is not only driven by having a current cap on stock investment, but also by the prospect of limited decision freedom in future periods. The optimal  $X_t^*$ s in the unrestricted mean-shortfall model are only optimal given the future possible recourse actions. Hence, if the space of possible future actions is restricted, the current decision is bound to change.

To show the consequence of a persistent cap on the fraction invested in stock, Figure 3.9 shows the optimal stock investment in terms of amounts in the risky asset at time 0.

We observe that in the case of a cap in both periods, initial stock investment is strictly smaller than in the unrestricted model *for all* wealth levels. This is caused by the fact that losses in future periods cannot be followed by an optimal unrestricted investment in stocks.

## 3.6 Conclusions

This chapter has considered several extensions to the framework of Chapter 2. The results for the Kahneman-Tversky value function revealed both conceptual problems as well as problems with unbounded solutions. These results suggest that the Kahneman-Tversky value function combined with the parameter values estimated by Tversky and Kahneman (1992) or Shumway (1997) is not useful in a setting of financial planning. However, one can consider the value function to be of use in research that looks at empirical behavior in an investment situation. In that case, the resulting optimal stock investment for a somewhat larger loss aversion parameter was insightful. It showed that for bounded solutions a clear V-shape can be found in the decision rule.

Another downside risk measure was analyzed in the case of squared shortfall. The convex loss penalty has desirable properties for application in financial planning. At the outset, it seemed possible that the V-shapes that were found for the mean-shortfall model, and persisted for the Kahneman-Tversky case, would disappear. The results showed otherwise. Numerical solutions to the mean-quadratic shortfall model showed investment patterns with clear V-shapes again. Although analytical tractability of the complete solution was not possible, we were able to give additional properties of the solution.

In addition to loss aversion for wealth below a fixed benchmark, Section 3.4 added a benchmark level at current wealth to the shortfall risk-measure. With the extra loss aversion parameter, risk-taking at negative surpluses can be driven to zero. The V-shape disappears.

Adding a restriction on the allowed fraction invested in the risky asset first of all bounds the V-shape from above. Second, in a multi-period framework the slope of the decision rule becomes lower, as the space for future recourse actions is limited.

Concluding, this chapter has made the point that the mean-shortfall model of Chapter 2 can be seen as a benchmark case for a larger class of loss averse models. We were able to use the same analytical techniques to find properties of the optimal solution, and established direct links between the difference in the specification of the objective and the resulting decision rules. The V-shape of stock investments as a function of initial wealth or surplus appears a persistent phenomenon if agents are loss averse.

# Appendix

## 3.A Proofs

### Proof of Theorem 3.3.1:

The maximization problem to solve is

$$\max_{X_0} \mathbb{E}[W_1] - \nu \mathbb{E}[(\{W^B - W_1\}^+)^2], \quad (3A.1)$$

where  $W_1 = W_0 r_f + X_0 \cdot (u - r_f)$ . The first order condition to this problem can be written as

$$\mathbb{E}[u - r_f] - 2\nu r_f S_0 \int_0^{\bar{u}} (u - r_f) dG - 2\nu X_0 \int_0^{\bar{u}} (u - r_f)^2 dG = 0, \quad (3A.2)$$

where  $S_0 = W_0 - W^B/r_f$ , and  $\bar{u}$  is defined as

$$\bar{u} = -\frac{S_0 r_f}{X_0} + r_f, \quad (3A.3)$$

giving the return for which  $W_1 = W^B$ . The second order condition is given by

$$-2\nu \int_0^{\bar{u}} (u - r_f)^2 dG < 0, \quad (3A.4)$$

which is satisfied for  $\nu > 0$ .

We prove the four parts of the theorem separately.

(i) If  $S_0 = 0$ ,  $\bar{u} = r_f$ , so equation (3A.2) simplifies to

$$\mathbb{E}[u - r_f] - 2\nu X_0 \int_0^{r_f} (u - r_f)^2 dG = 0. \quad (3A.5)$$

The left-hand side of (3A.5) is a linear equation in  $X_0$ , with solution

$$X_0^* = \frac{\mathbb{E}[u - r_f]}{2\nu \mathbb{E}[(r_f - u)^+]^2}, \quad (3A.6)$$

which establishes that  $X_0^*$  at zero surplus is positive and decreasing in  $\nu$ . ■

- (ii) Given the first-order condition (3A.2) and using the implicit function theorem, the derivative of  $X_0^*$  with respect to  $S_0$  is given by

$$\frac{\partial X_0^*}{\partial S_0} = -\frac{\partial F/\partial S_0}{\partial F/\partial X_0}, \quad (3A.7)$$

where we have used  $F(\cdot)$  to denote the derivative of the objective function with respect to  $X_0$ , being the left-hand side of (3A.2). From the second-order condition in 3A.4 we know that the denominator of (3A.7) is negative. The numerator is given by

$$\frac{\partial F}{\partial S_0} = -2\nu r_f \int_0^{\bar{u}} (u - r_f) dG, \quad (3A.8)$$

which is zero for  $\bar{u} = 0$  and a  $\bar{u}^m > r_f$ . We can ignore the case  $\bar{u} = 0$ , as  $\bar{u} \downarrow 0$  is only a solution to the first order condition (3A.2) if  $\nu \rightarrow \infty$ , or  $|S_0| \rightarrow \infty$ . From the definition of  $\bar{u}$  in (3A.3) and the fact that  $\bar{u}^m > r_f$  it follows that the surplus at which  $X_0^* = X_0^{min}$  lies at a negative surplus. Combining (3A.2) with (3A.8), an alternative expression for  $X_0^{min}$  is given by

$$X_0^{min} = \frac{\mathbb{E}[u - r_f]}{2\nu \int_0^{\bar{u}^m} (u - r_f)^2 dG}. \quad (3A.9)$$

From the second-order derivative  $\partial^2 F/\partial S_0^2$ , which is positive, we find that  $X_0^{min}$  is the minimum investment in the risky asset.

With respect to the sign of  $X_0^*$ , note that from (i) we have that for  $X_0^*(S_0 = 0) > 0$ . Since we have that  $\partial X_0^*/\partial S_0$  is a continuous function of  $S_0$ , and  $X_0 = 0$  does not solve the first order condition (3A.2), we find that  $X_0^*$  is positive. ■

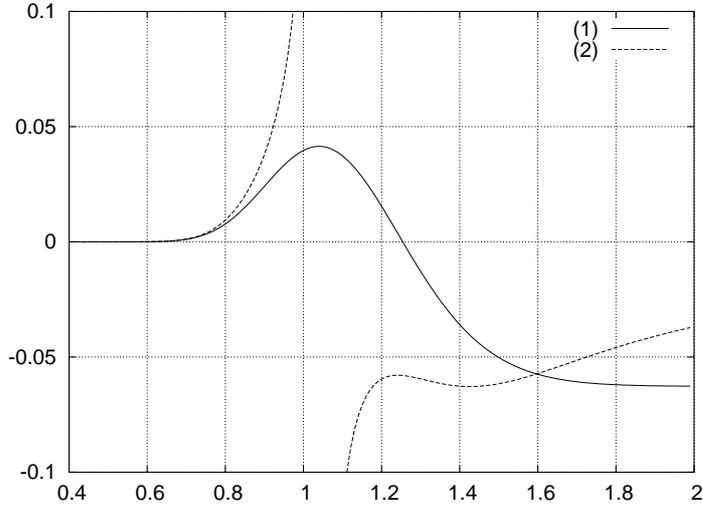
- (iii) Dividing the left-hand side expression of the first-order condition (3A.2) by  $2\nu S_0 r_f$  gives

$$\frac{\mathbb{E}[u - r_f]}{2r_f \nu S_0} - \int_0^{\bar{u}} (u - r_f) dG + \frac{1}{\bar{u} - r_f} \int_0^{\bar{u}} (u - r_f)^2 dG = 0. \quad (3A.10)$$

For  $S_0 \neq 0$  and  $\nu \rightarrow \infty$ , the expression in (3A.10) simplifies to

$$\int_0^{\bar{u}} (r_f - u) dG = \frac{1}{r_f - \bar{u}} \int_0^{\bar{u}} (u - r_f)^2 dG, \quad (3A.11)$$

which depends on  $r_f$  and  $G(\cdot)$  only. As (3A.11) does not include terms  $W_0$  or  $X_0$ , we find that in the limit,  $X_0^*$  is determined by the  $\bar{u}$ s that



**Figure 3.10: Solutions to equation (3A.11)**

This figure shows the left-hand(1) and right-hand(2) side of equation (3A.11), as a function of  $\bar{u}$ .  $r_f = 1.04$ ,  $u \sim \log N(0.085, 0.16)$ .

solve (3A.11). For given parameter values, Figure 3.10 shows the left- and right-hand side of equation (3A.11), as a function of  $\bar{u}$ . There is a clear point of intersection for  $\bar{u} > r_f$ . For  $\bar{u} < r_f$ , the left-hand side of (3A.11) is smaller than the right-hand side, so for  $\bar{u} < r_f$  it solves only for  $\bar{u} = 0$ , corresponding to  $X_0^* = S_0$ . ■

To compare  $X_0^*$  for a negative surplus and  $\nu \rightarrow \infty$  with the one in the mean-shortfall model for  $\lambda \rightarrow \infty$ , observe that in the mean-shortfall case, the limiting first-order condition is in equation (3A.11), but with right-hand side equal to zero. With  $\bar{u}^- > r_f$ , for a negative surplus the right-hand side of (3A.11) is negative, so the limiting  $\bar{u}^\infty$  is larger than in the mean-shortfall model. From the definition of  $\bar{u}$  it follows that the limiting  $X_0^*$  is smaller than in the mean-shortfall model. ■

### Proof of Theorem 3.4.1

We start by rewriting the optimization problem (3.14) as

$$\begin{aligned} \max_{X_0} W_0 r_f + X_0 \mathbb{E}[u - r_f] - \lambda_1 \cdot \int_0^{\bar{u}} W^B - W_0 r_f - X_0(u - r_f) dG \\ - \lambda_2 \int_0^{r_f} W^B - W_0 r_f - X_0(u - r_f) dG, \end{aligned} \quad (3A.12)$$

where  $\bar{u}$  is defined as

$$\bar{u} = \frac{W^B - W_0 r_f}{X_0} + r_f. \quad (3A.13)$$

Using Leibniz' rule, the first order condition to this problem is given by

$$\mathbb{E}[u - r_f] + \lambda_1 \int_0^{\bar{u}} u - r_f dG + \lambda_2 \int_0^{r_f} u - r_f dG = 0. \quad (3A.14)$$

The result follows by observing that this condition is the same as that for the model in Chapter 2, except for the added third term in (3A.14). This term, however, does not depend on  $S_0$  or  $X_0^*$ , so the rest of the proof is the same of that for the mean-shortfall model, see the proof of Theorem 2.3.1. ■

---

# Optimal Pension Fund Investment Management

---

## 4.1 Introduction

In this chapter we simultaneously provide a generalization and interpretation of the model in Chapter 2. We generalize the model to incorporate stochastic liabilities, interpreting the model in a setting of Asset-Liability Management for pension funds.

In recent years the strategic investment policy by Dutch pension funds has to be supported by a so-called Asset/Liability Management(ALM) study. In the context of pension funds these studies give insight into the expected cost and risk aspects of investment and contribution policies. This way, an ALM study serves as a foundation of contribution and investment policies toward employers, workers, pensioners and regulator. Even in cases where ALM studies are not compulsory, they are used heavily by financial institutions, such as banks, insurance companies and pension funds, see Ziemba and Mulvey (1998) for an extensive overview of the use of and research into ALM.

An important feature of any financial planning model such as used in ALM is the objective function. In the objective function, risk and return measures have to be specified to be able to make inferences on the desirability of policies under consideration. For a long time, the financial world has settled on the use of the mean and variance of investment returns to measure performance. The developments of the last decade, however, have lead to

using more explicit risk measures in ALM objectives. This is where the research of the previous chapters and the area of Asset/Liability Management coincide.

Dutch pension funds are the most likely candidates to apply a loss averse framework to. First, they actively use Asset/Liability Management in their financial planning process. In fact, doing an ALM study is mandatory for the motivation of investment policy in The Netherlands. For an overview of ALM and pension funds, see Boender and Vos (2000). Dert (1995) provides a characteristic example, modeling the ALM problem for pension funds as a stochastic programming model. Second, the risk measures that are used in most ALM studies are down-side risk measures. These risk-measures can be seen as representing loss averse preferences directly. The use of down-side risk measures in a defined benefit context is widespread. See for example Sortino and Van der Meer (1991), Harlow (1991), and Boender (1997).

In comparison with other studies of portfolio management for pension funds, Randall and Satchell (1997) is a good example of a traditional analysis. They compare the efficiency of different portfolios for pension fund assets with respect to the mean and variance of the return. This has the advantage of being able to use the standard toolkit of mean-variance analysis as developed by Markowitz (1952) in his CAPM. The disadvantage, however, is that it does not take into account that pension funds experience a serious downside when returns are (too) low. This is even acknowledged by the authors, who note that pension funds cannot afford “to lose a huge amount of money, even if they are frequently making small amounts of money. This asymmetry is not accounted for by a model which is defined over mean and variance only.” Their analysis also prohibits an inference on how the investment decision depends on the (initial) financial position of the fund.

A good illustration of the predominant technical approach to Asset/Liability Management in the literature is given by Zenios (1995). He considers the computational problems of solving stochastic programming formulations for Asset/Liability Management in a fixed-income environment. Although one of the models he considers includes penalty parameters for the downside and upside deviation of the portfolio return, most attention is paid to a multi-period model in which the utility of the investor is given by a traditional iso-elastic function, e.g., power utility. Clearly, the paper is not focused on the structure of the optimal decisions with respect to risk and return, but rather on the computational and model building aspects. See also Kusy and Ziemba (1986), Hiller and Eckstein (1993), Maranas et al. (1997) and Zenios et al. (1998).

The setup of this chapter is as follows. In Section 4.2 we interpret the mean-shortfall model of Chapter 2 in terms of a defined-benefit pension fund.

Section 4.3 derives implications for pension fund investment policy in a relevant setting, and considers the sensitivity of the results to the specification of uncertainty. The sensitivity analysis is repeated for a model with downside deviation as the risk measure in Section 4.4. In Section 4.5 we explore empirical evidence of loss averse preferences for pension funds by examining actual pension fund investment policies. Section 4.6 ends with a discussion and conclusions.

## 4.2 Pension funding as a mean-shortfall optimization

Recall the formulation of the mean-shortfall model that we analyzed in Chapter 2. As a  $T$  period optimization problem, it is given by

$$\max_{X_0, \dots, X_{t-1}} \quad \mathbb{E}[W_T] - \lambda \cdot \mathbb{E}[(W^B - W_T)^+], \quad (4.1)$$

$$\text{s.t.} \quad W_{t+1} = W_t r_f + X_t \cdot (u_t - r_f), \quad t = 0, \dots, T-1, \quad (4.2)$$

where  $W_t$  represents wealth at time  $t$ ,  $r_f$  the risk-free (gross) return,  $u_t$  the uncertain return over period  $t$ ,  $W^B$  the benchmark wealth at the horizon and  $\lambda$  the loss aversion parameter. Furthermore, we find it convenient to define  $W_0^B$  as the risk-free discounted benchmark, given by  $W^B / r_f^T$ .

### Formulation of the objective

By choosing linear shortfall as down-side risk measure and not, for example, quadratic shortfall, the relation with the results of Chapter 2 is most clearly visible. However, since ALM studies predominantly use downside deviation as a risk-measure, the sensitivity of the results is analyzed in Section 4.4. Note that the results in Chapter 3 have already shown that the general shape of the solution to the model in (4.1) and (4.2) is the same for linear and quadratic shortfall.

A severe limitation of the objective function in (4.1) is that it is one-dimensional, i.e., it is defined in terms of the single variable  $W_T$ . The real objective for a benefit-defined fund, however, consists of several competing objectives. Typical objectives are: minimal contributions, maximal indexation of pensions, and minimal risk with respect to funding. With respect to the latter, the key ratio that is reported as a measure of financial soundness of a fund is the actuarial funding ratio. It is simply the value of assets divided by the present value of the liabilities,  $W_0 / W_0^B$ . Another measure is the surplus, which can be defined as  $W_0 - W_0^B$ . Now, if the funding ratio

(or surplus) is high, the risk of underfunding is low, the contribution level can be lowered, and indexation can fully compensate for inflation. Hence, in the simple model of pension fund investment, we use an ALM objective that is only defined in terms of the surplus of the fund, which has a 1-1 relation with the funding ratio. If the surplus is high, the other objectives are met at the same time. If the surplus is low, or if there is a deficit, the associated consequences are carried over to the other objectives, e.g., for a fund with a low funding level, contributions are raised and indexation might be postponed. See Boender and Vos (2000) for an analysis of the mechanism of allocating financial risks over multiple objectives for a pension fund, which they call risk-budgeting.

Leibowitz et al. (1992) model shortfall-aversion of pension funds by incorporating a shortfall constraint on the ‘surplus return’. The asset returns are then analyzed in a mean-variance setting. Additionally, a “surplus return” is defined as the ratio of surplus growth divided by the initial value of liabilities. Their definition of this return implies that their subsequent mean-variance analysis is equivalent to having a constraint on the amount of surplus shortfall. Their setting can be compared with our objective function in (4.1) by interpreting  $\lambda$  as a Lagrange-multiplier, whose value depends on initial wealth  $W_0$ . The problem with such a setup is that there is not always a solution possible. Also, in our framework of two assets, any feasible solution would be trivial: invest the maximum amount in stocks, such that the shortfall constraint is binding. The usefulness of their approach is that they explicitly reckon with the duration of pension liabilities versus the duration of available fixed-income instruments.

### Interpretation of the parameters

The symbols  $r_f$  and  $u_t$  represent the returns on the investment categories available to the fund. In practice, pension funds will invest in more than only two categories, but we have seen in Chapter 2 that the two-asset case generalizes to that of more assets.  $W_t$  is a natural representation for the total value of assets of the fund.

Interpreting  $W^B$  is key to the present chapter. It is the reference level relative to which either a gain or a loss is measured. For a benefit-defined pension fund, a natural candidate for the reference level is the value of the liabilities that needs to be covered at time  $T$ . The liabilities of a pension fund consist of the pension rights built-up by the contributors. The value of these rights can be computed by discounting the future pension payments. In the Dutch system, future pension payments are discounted at an actuarial discount rate of 4%. As such the present value of future payments can be

computed, resulting in the liability value for the fund.  $W^B$  can be interpreted as the discounted value at time  $T$  of the guaranteed pension payments at times  $T+1$  and further. This is not all that can be said of the computation of the liabilities. Subection 4.3.3 discusses an alternative method, and explores the consequences in terms of our model.

There are several factors that influence the liabilities of a defined-benefit pension fund,  $W^B$  in our model. In the following, we consider only one factor, namely price inflation. This is the key variable of interest for a pension fund, as it influences the real value of pensions. In general, pension funds aim to cover the decrease in real value of nominal pension rights by applying indexation. That brings about a direct link between price inflation and the nominal level of the liabilities,  $W^B$ . Note that it is possible to consider other variables influencing the level  $W^B$ , such as the career developments of individual (active) members, and wage inflation.

The effect of price inflation on  $W^B$  is modeled by introducing a variable  $\pi$ , representing inflation over the period of optimization. Defining the *real* pension obligations at the end date as  $W^{B_{real}}$ , the nominal reference point  $W^B$  is determined as

$$W^B = (1 + \pi) \cdot W^{B_{real}}, \quad (4.3)$$

where  $1 + \pi$  is the gross inflation factor, assuming annual compounding. The setup in (4.3) makes it easy to study the effect of inflation on the optimal solution. It can also make clear the way that current regulation sees the valuation of liabilities. Under the current Dutch regulation, nominal liabilities are discounted at a rate of 4%. It is difficult to find a formal rationale of this method, but the opinion seems to be that 4% is a reasonable proxy for the long-term real rate of interest<sup>1</sup>. Using equation (4.3) and the definition of  $W_0^B$ , we can see that if the level of liabilities is multiplied by inflation, i.e., full indexation, the actuarial interest rate of 4% corresponds to a nominal risk-free return of  $(1 + \pi) \cdot 1.04$ .

## Planning horizon

We choose the horizon  $T$  equal to 15 years, intended to match the duration of the liabilities of an average pension fund. The choice for the static model excludes the possibility of policy changes between time 0 and  $T$ . Thus, the decision  $X_0^*$  gives an an initial asset mix that is optimal over the whole planning period. Strategies in ALM studies are often also static rather than dynamic. The static model is straightforward to analyze, and we have seen

---

<sup>1</sup>Note that insurance companies currently use 3%.

that its solution is representative for the solution to the multistage model, see Chapter 2.

In an attempt to explain the equity premium puzzle, Benartzi and Thaler (1995) postulate that investors and pension funds are loss averse and myopic, i.e.,  $T$  is equal to 1 year. They motivate the suggestion of myopia by pointing out that investors account for their gains and losses at the end of each year. The only evidence they give, however, is that the average asset mix from their model is comparable to the average mix found in their dataset of institutional investors. Besides the debatable evidence, at the outset myopia is unrealistic for pension funds. Pension fund managers frequently express their view that they are, and should be, long-term investors, and should not be pinned down on 1-year performance figures. This view is consistent with the practice of motivating investment policies by ALM studies that calculate risk and return over a long-term horizon.

### Shortfall risk

The objective function in (4.1) involves the maximization of wealth (or surplus), with a penalty on the expected amount of deficit below 100% actuarial funding. loss aversion with respect to deficits below  $W^B$  represents that, in a defined-benefit system, the liabilities need to be covered by available assets. The formulation in (4.1) allows for a deficit. The parameter  $\lambda$  represents the aversion to shortfall. In the context of pension funds, the extent of loss aversion is influenced by the fund's flexibility in a situation of underfunding. A fund with much flexibility with regard to contribution rises, additional capital injections by the sponsor, lowering of pension rights, and skipping of indexation, might be less loss averse than a fund that does not have such flexibility.

A problem with incorporating loss aversion in the objective of a pension fund, is that there is an obligation by law that the current value of all pension liabilities should always be covered by available assets. The Dutch law on pensions, the PSF (Pensioen- en Spaarfondsenwet), stipulates in article 9a sub 1 that "the assets of a pension fund should ... be sufficient to cover the pension liabilities." This constitutes the most serious financial check that is to be performed by the regulator, the PVK (pensions and insurance supervisory authority of The Netherlands). Note that by interpreting  $W^B$  as the level of pension liabilities, the situation  $W_T < W^B$  is against the law. However, the possibility of underfunding is a reality, and the present Chapter will provide insight in optimal behavior when a fund is in a situation of a deficit.

An additional obligation for most pension funds is the statutory commit-

ment to safeguard the *real* value of pensions, by means of providing indexation on accrued pension rights. Such indexation can usually be postponed if the financial position of the fund does not allow it, but if the position gets better in the future the fund first needs to apply the postponed indexation.

## 4.3 Results

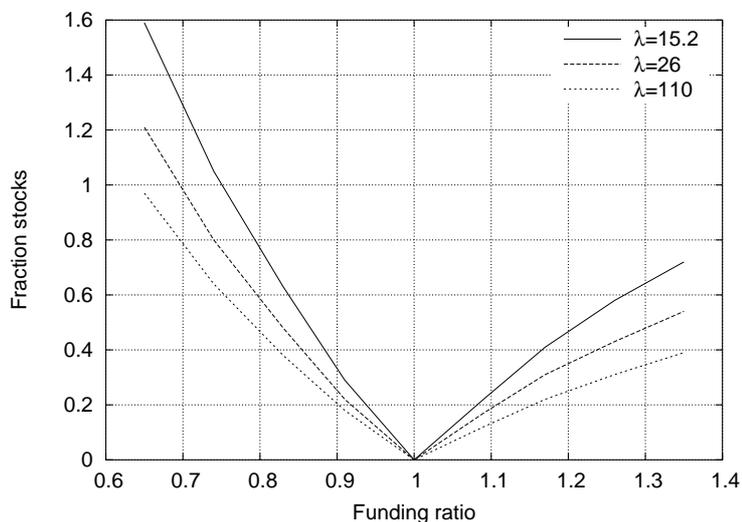
In Subsection 4.3.1, inflation  $\pi$  is assumed constant. We take  $r_f$  such that the actuarial discount rate of 4% is correct. Subsection 4.3.2 specifies a joint probability distribution of inflation and the stock return, while Section 4.3.3 considers the consequences of having a risk-free rate that is not in accordance with the actuarial assumption of a real rate of 4%.

### 4.3.1 Base case

In the most simple case, we assume that  $W^{B_{real}}$ , the real level of liabilities at time 1, is known with certainty at time  $t = 0$ . In other words, inflation  $\pi$  is a known constant. As the planning horizon is 15 years, a reasonable assumption for the expectation of inflation is 2% per year, the inflation target for the European Central Bank (ECB). That makes  $W^B = (1.02)^T \cdot W^{B_{real}}$ . Assuming that the current practice of using an actuarial discount rate of 4% for nominal liabilities is valid, we set  $r_f = (1 + 6.1\%)^T$ . To see the effect, observe that a real liability of 180 at time T has a present value of  $180 \cdot (1.02/1.061)^T = 180/1.04^T = 100$ . The only uncertainty in the model comes from the return on the risky asset  $u$ . As in Chapter 2, we take  $u$  to have a lognormal probability distribution with an average return of 10% and standard deviation of 17%, representing typical historical figures for stock returns.

To solve model (4.1), we discretize the distribution of  $u$  using 500 points, and find the optimal stock investment through numerical optimization with Gams. Doing this for different values of the initial wealth  $W_0$  results in pairs  $(W_0, X^*)$ . Figure 4.1 shows the resulting optimal fractions  $X^*/W_0$  as a function of  $W_0$  for different values of the loss aversion parameter  $\lambda$ . Table 4.1 lists the values of  $\lambda$  and the associated probabilities of shortfall.

From Figure 4.1 we can observe the typical V-shape that we have seen in earlier chapters. In the pension fund context this figure has an important consequence. Pension funds usually use an ALM study to support an investment policy that is aimed at holding a “strategic mix”. The figure shows, however, that the optimal initial asset mix for the 15 year period depends on the initial funding ratio.



**Figure 4.1: Optimal fraction invested in stocks for the base case**  
 This figure shows the optimal investment in stock as a fraction of initial wealth against the funding ratio for different values of the loss aversion parameter  $\lambda$ .  $T = 15$

**Table 4.1: The values of  $\lambda$  used to compute the optimal investment decisions in Figure 4.1**

The second column lists the probability of underfunding, given a positive initial surplus. The third column gives the probability that the final surplus is positive, given a negative initial surplus.

$\lambda$	P(underfunding)	P(overfunding)
15.2	10%	16.8%
26	5%	11.8%
110	1%	7.8%

In terms of the funding ratio of the pension fund, the interpretation of the figure is as follows. For a funding ratio of 100%, the future nominal pension payments can be fully guaranteed. In this case, the optimal allocation does not contain stocks. We call this point the minimum-risk portfolio.

At a funding ratio of more than 100%, the surplus asset value leads to an increasingly higher fraction in stocks, thus using the risk premium on stock investment to make the fund even wealthier (in expectation).

At funding ratios below 100%, stock investments increase in the extent of underfunding. When the pension liabilities cannot be guaranteed with certainty anymore, the fund has to take risk to prevent a further erosion of the pension claims. A higher percentage in stocks is the only way to reach the benchmark  $W^B$  with positive probability. In a real-world setting, an alternative to increasing  $X$  can be an increased contribution from the side of the sponsor, or a more frugal pension claim by lowering indexation. Both alternatives can be visualized as having the effect of shifting the financial position of the fund to the right on the horizontal axis of Figure 4.1, either by increasing the numerator or by decreasing the denominator of the funding ratio.

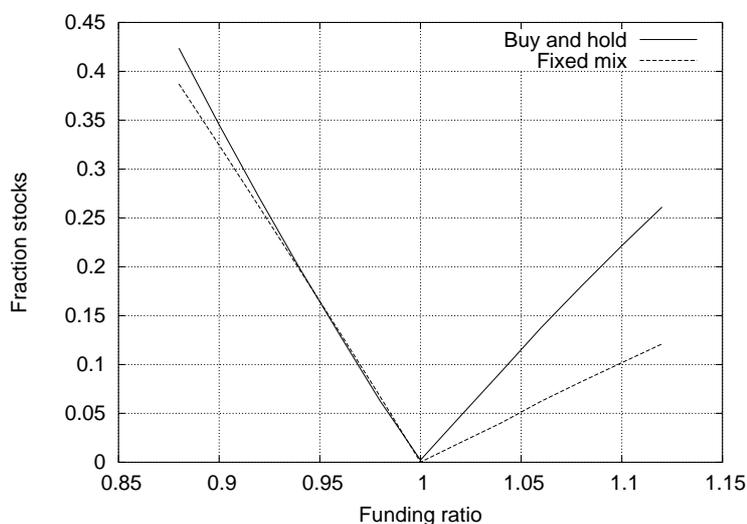
The loss aversion parameter  $\lambda$  has a decreasing effect on the steepness of the V-shape in Figure 4.1. A higher loss aversion leads to a safer asset mix, as the certainty of the risk-free return becomes more attractive than the possible extra return on stocks. For the rest of this chapter, we take  $\lambda = 110$ , representing a 1% probability of underfunding.

A marked difference between practice and the model in (4.1) is that in ALM studies one usually optimizes over a fixed mix, where we optimize over the initial mix only. The optimal policy that is plotted in Figure 4.1 can be labeled as a ‘buy-and-hold’-policy, i.e., after the initial mix is set, the fraction invested in stocks only changes with the returns that are made. To analyze the consequences of this assumption, we perform a hybrid simulation/optimization as in Boender (1997)<sup>2</sup> to find the optimal fixed-mix policy for the model in (4.1). The results are in Figure 4.2, which shows the optimal fraction of stocks for the buy-and-hold against the fixed-mix alternative.

We see that the V-shape is retained for the fixed-mix policy, although the initial allocation is different. For the fixed-mix strategy, the right side of the V-shape is lower than for the buy-and-hold strategy. To further explore

---

<sup>2</sup>The method is in fact a genetic algorithm. Starting from an initial sample of random asset mixes, stepwise improvements in the best mix are achieved through subsequent replication and selection steps.



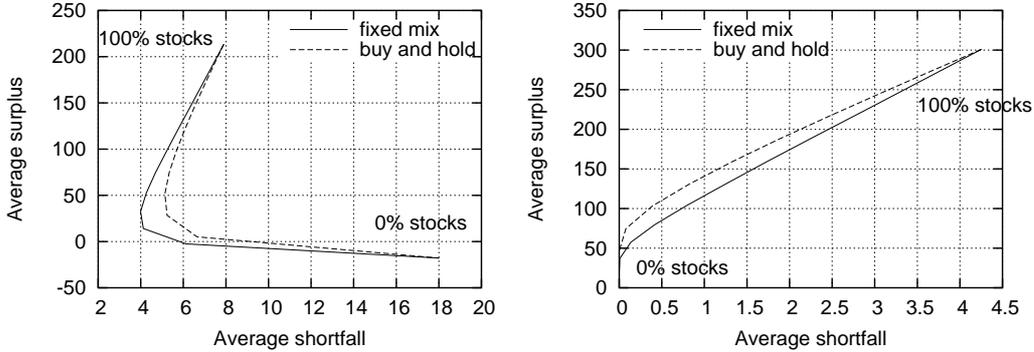
**Figure 4.2: Optimal investment for buy-and-hold versus fixed-mix**

This figure shows the optimal fraction invested in stocks for two different assumption on the intertemporal rebalancing. The buy-and-hold strategy implies a choice for an initial asset mix, and no rebalancing. The fixed-mix strategy implies that the fraction stocks is rebalanced to the same ratio at the beginning of each time period  $0 < t \leq T$  of the planning period.

the difference between the two strategies, consider the risk-return diagram in Figure 4.3.

To start with the left panel, it turns out that the objective value for a fund starting in underfunding is higher for the fixed-mix strategy than for the same asset mix under buy-and-hold. This is because in underfunding the buy-and-hold strategy induces the exact opposite behavior of what is dynamically optimal. From Chapter 2 we see that in under a dynamically optimal strategy, the fraction of stocks should increase if the surplus drops sufficiently low. Under buy-and-hold, however, if the surplus decreases due to a fall in stock prices, the fraction invested in stocks decreases as well.

In the right panel, the risk-return curve for the buy-and-hold strategy is entirely above the fixed mix one. This means that for the same average shortfall, a higher average surplus is achieved. It explains the difference in the optimal stock fraction for the right-hand side of the V in Figure 4.2. An explanation for the effect is that a buy-and-hold policy has an automatic downside protection: even if the initial investment in stocks becomes worthless, between time 0 and  $T$ , the investment in the risk-free asset is left



**Figure 4.3: Risk and return for buy-and-hold versus fixed-mix**

For a range of asset mixes, this figure shows the average surplus versus average shortfall, using 10000 scenarios for the stock return. The left panel has an initial funding ratio of 90%, the right panel of 110%.  $\lambda = 110$ ,  $u \sim \log N(0.085, 0.16)$ ,  $r_f = 1.04$ ,  $T = 15$ .

unchanged. For the fixed-mix, however, negative stock returns are followed by a rebalancing to the fixed-mix at the end of each period. With respect to the upside, under buy-and-hold the fraction in stocks grows proportionally with the stock return, so the upside potential is larger than under the fixed-mix policy. The relatively bad performance of a fixed-mix has been pointed out before in Perold and Sharpe (1988), who note on constant-mix strategies that “They have less downside protection than, and not as much upside as, buy-and-hold strategies”.

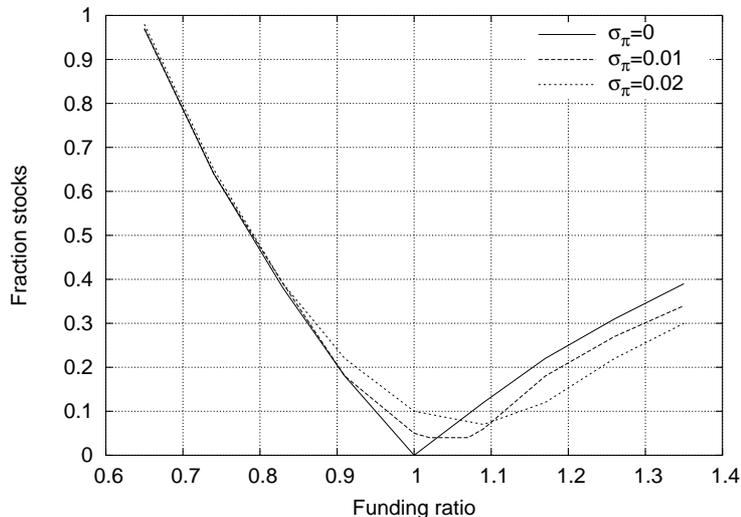
So far, we have assumed a deterministic value for  $W^B$ . In the next subsection we consider the robustness of the shape of the decision rules to the introduction of an uncertain inflation rate that influences the level  $W^B$ .

### 4.3.2 Uncertainty in inflation

With an uncertain inflation rate  $\pi$ , and the reference point  $W^B$  in the objective function (4.1) becomes a stochastic variable. Log-stock returns and inflation are modeled as bivariate normal. This way, the covariance matrix  $\Sigma$  of the logs looks as

$$\Sigma = \begin{bmatrix} \sigma_\pi^2 & \rho \cdot \sigma_\pi \sigma_u \\ \rho \cdot \sigma_\pi \sigma_u & \sigma_u^2 \end{bmatrix}, \quad (4.4)$$

where  $\rho$  is the correlation coefficient between inflation  $\pi$  and the log-stock return  $u$ . For the stock return we take a mean of 10% and standard deviation of 17%. These correspond to a normal distribution with  $\mu = 0.085$  and



**Figure 4.4: Optimal stock investment under inflation uncertainty.**

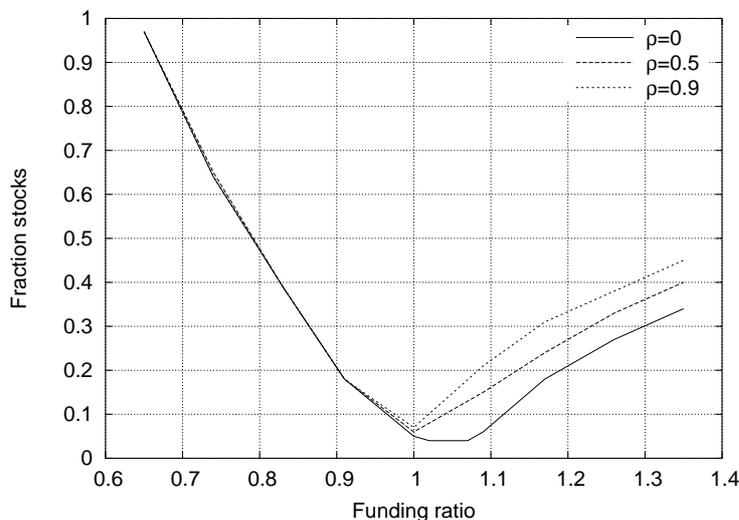
In this figure the optimal stock fractions to problem (4.1) are plotted for different values for  $\sigma_\pi$ , the standard deviation of the inflation.  $\lambda = 110$ ,  $T = 15$ ,  $\rho = 0$ .

$\sigma = 0.16$  for the log-stock return. Mean inflation is again equal to 2% and  $r_f = 1.061$ .

Note that if we model inflation as generated by a normal distribution, negative values will occur. This is not a problem in the model, it only shows that modeling inflation this way is a crude approximation of the real probability distribution. As the mean and variance of inflation are based on historical data, the analysis will also consider alternative values. With sufficient controversy over the size of the correlation between inflation and stock returns, the parameter  $\rho$  is also varied in the analysis, starting from a value of zero. To start with, results for varying values of the standard deviation  $\sigma_\pi$  of inflation are in Figure 4.4.

From Figure 4.4 we observe again the basic pattern from Figure 4.1. The percentage in stocks has a V-shaped relation with the funding ratio. The difference with Figure 4.4 is that the kink of the V does not lie at a zero stock-investment. The minimum-risk portfolio contains stocks. When inflation can only be hedged partially, the bottom line is that there is always uncertainty with respect to the level of the final pension payment.

Next, we consider the optimal solution to the loss averse model when the correlation between stocks and inflation is non-zero. In the long run stocks are considered a partial hedge for inflation risk. This is one of the motivations

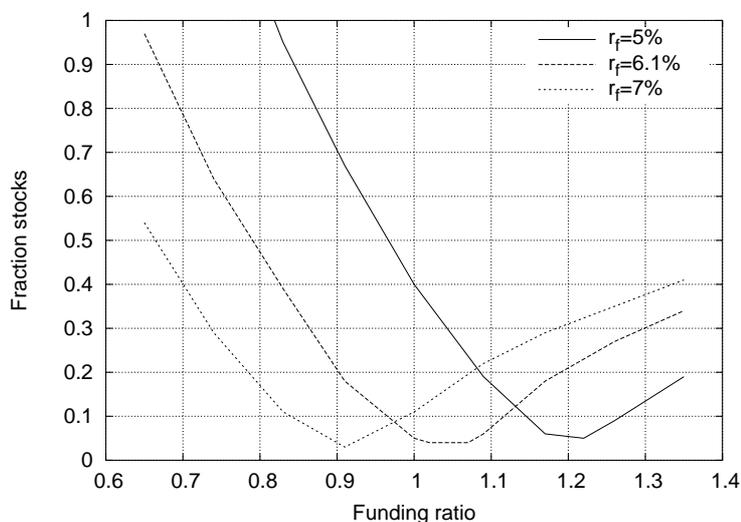


**Figure 4.5: Optimal stock investment for different correlation coefficients**

In this figure the optimal stock fractions to problem (4.1) are plotted for different values for  $\rho$ , the correlation coefficient between the log-stock return and inflation.  $\sigma_\pi = 0.01$ ,  $\lambda = 110$ ,  $T = 15$ .

for investing in stocks in case of indexed pension claims, see Leibowitz et al. (1994). Randall and Satchell (1997) find a positive correlation for UK data of 0.16 between equity and paid-out pensions. However, for Dutch data from 1956 to 1994, Dert (1995) finds an annual negative correlation coefficient of -0.24 between log-stock returns and price inflation.

Figure 4.5 shows the effect of changes in the correlation  $\rho$  between stock returns and inflation. For a given funding ratio, a higher correlation results in a larger stock investment. The effect on the investment policy is only moderate, however. With a yearly correlation of 90%, which is very high, the maximum difference in fraction stocks compared to no correlation is 15 percentage points. It is surprising that this effect is so limited. As mentioned before: the presence of correlation between stock returns and inflation motivates stock investment as an inflation hedge. Given that the true correlation is suggested to lie between 0.1 and 0.5, this motivation seems to be of limited important in practice. The positive risk-premium of stocks is more important than the effect of a positive correlation with inflation.



**Figure 4.6: Optimal stock fraction for different values of the risk-free rate**

In this figure the solutions to model (4.1) are shown when the risk-free rate  $r_f$  is varied. With expected inflation of 2%, the rate of 6.1% corresponds to a real rate of 4%.  $T = 1$ ,  $\rho = 0$ ,  $\sigma_\pi = 1\%$ .

### 4.3.3 Alternative interest rates

Until now, we assumed that the actuarial interest rate of 4% was exactly the real risk-free interest rate. In this section, a final plot of stock investment versus the funding ratio is shown, now in a situation where the real rate differs from 4%.

All earlier calculations assumed that the actuarial interest rate of 4% is a reasonable representation of the long-term average of the real interest rate. In this paragraph we extend the analysis to include the fact that the real interest rate can differ from 4 percent.

A real interest rate other than 4% has a direct effect on the market value of the liabilities. Valuation of the liabilities according to market value is finding more ground in the Dutch pension fund sector. Using a real interest rate lower than 4% gives lower funding ratio, using a higher interest rate gives a higher funding ratio. To illustrate the consequences of a real interest rate other than 4%, Figure 4.6 shows the solutions to model (4.1) for varying values of the risk-free interest rate  $r_f$ .

From the figure we can observe that the location of the minimum-risk portfolio, i.e., the bottom of the V-shape, is very sensitive to the value of the risk-free rate. We know from Chapter 2 that without uncertainty in  $W^B$

the minimum-risk portfolio lies at zero surplus, see also Figure 4.1 of this section. The wealth level that gives zero surplus equals the discounted value of the liabilities  $W^B$  against the risk-free rate  $r_f$ . A larger  $r_f$  decreases the discounted value, a smaller  $r_f$  increases it. Hence, as the funding ratio on the x-axis of Figure 4.6 is computed regardless of the true value of  $r_f$ , the shift of the V-shape is due to the shift in the risk-free discounted value of the liabilities.

Note that a low future interest rate is predicted by Chaveau and Loufir (1997), as cited in Bovenberg (2001). They have studied the effect of the ageing of the population on the interest rate and predict that the real interest rate will decline to a level of 3.25% in 2025.

## 4.4 Downside deviation

The previous section has evaluated the consequences of using different specifications of uncertainty in the standard mean-shortfall model. The mean-shortfall model was selected, because it is the most simple loss averse model with an analytical characterization of the solution in Chapter 2. In the practice of ALM for pension funds, however, downside deviation is predominantly used as the risk-measure. This section explores the sensitivity of the outcomes from Section 4.3 for taking this risk measure. The outcome without inflation uncertainty has been studied in Chapter 3, Section 3.3. It showed that for a fixed benchmark, the resulting decision rule has the same shape as in the mean-shortfall model.

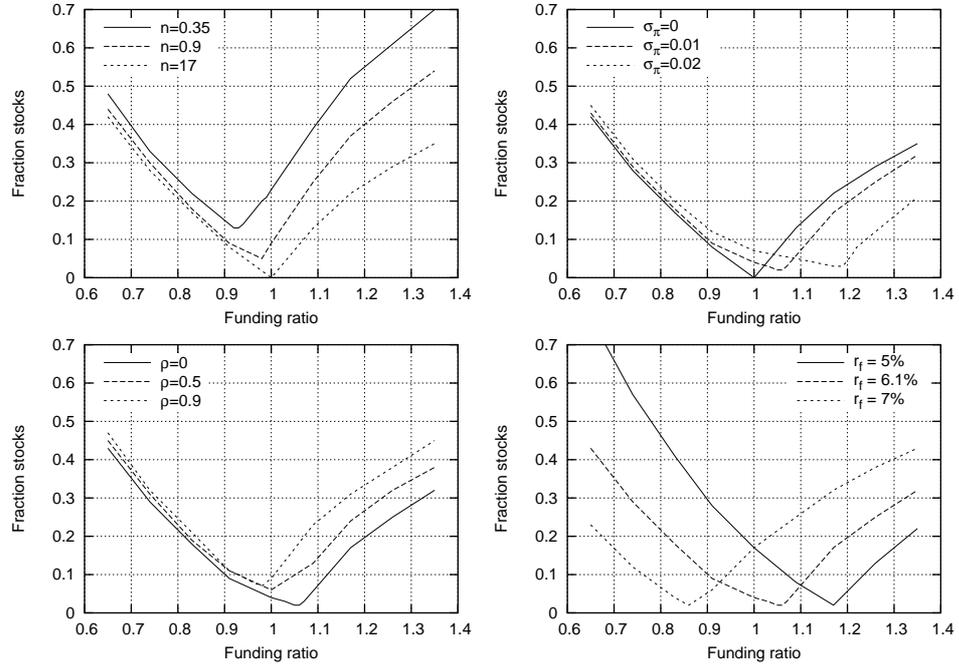
Using downside deviation as a risk-measure boils down to taking quadratic shortfall. The popularity of its use is due to the fact that it punishes large losses more than proportionally heavy than small losses. The formulation of the objective becomes,

$$\max_{X_0, \dots, X_{t-1}} \quad \mathbb{E}[W_T] - \nu \cdot \mathbb{E}[(W^B - W_T)^+ ]^2, \quad (4.5)$$

$$\text{s.t.} \quad W_{t+1} = W_t r_f + X_t \cdot (u_t - r_f), \quad t = 0, \dots, T - 1, \quad (4.6)$$

with the parameters defined as before, and  $T = 15$ . The difference with the objective function (4.1) is the power of 2 in the second term of (4.5). Also, in accordance with notation of Chapter 3, the loss aversion parameter is now represented by  $\nu$ , signaling that the risk-measure is different.

Figure 4.7 shows 4 panels in which the sensitivity analyses of the previous section are repeated for downside-deviation as the risk measure.



**Figure 4.7: Optimal policies for downside-deviation objective**

Top left panel shows the optimal stock investment fraction for different values of the loss aversion parameter  $\nu$ , inflation constant at 0.02,  $\rho = 0$ ,  $u \sim \log N(0.085, 0.16)$ ,  $r_f = 1.061$ . The values of  $\nu$  have been chosen to roughly compare with the 1%, 5%, and 10% probabilities of underfunding, as in Figure 4.1. For  $\nu = 8$ , the top right panel shows the solution for different values of the standard deviation of inflation, as in Figure 4.4. The lower left panel has different values for the correlation between inflation and log-stock returns,  $\rho$ , while  $\sigma_\pi$  is kept at 0.01, compare Figure 4.5. The lower right panel has different values for the riskfree rate  $r_f$ , while  $\rho = 0$ , and  $\sigma_\pi = 0.01$ .

Starting with the top left panel and comparing with Figure 4.1, we recognize the effect of the V-shape moving in the upper-left direction. Also, the left-hand side of the V stays at the same position, the right-hand side is much more sensitive to values of  $\nu$ . As the results in the figure do not include uncertainty for inflation, we can refer to Chapter 3 for a detailed explanation of the results.

The top-right, and lower-left panel show the effect of uncertainty in inflation on the outcomes. We find that  $\sigma_\pi$  and  $\rho$  have the same effect on the decision rule as in the mean-shortfall model. Increasing inflation uncertainty makes the minimum-risk point of the V-shape shift to the right. Increasing correlation between inflation and stock returns, the minimum stock allocation increases, as does the fraction of stocks for positive surpluses. As in Figure 4.5, the left-hand side of the V is not affected significantly by a posi-

**Table 4.2: Six large Dutch pension funds**

This table presents assets and funding ratios for six of the largest Dutch pension funds. Data is collected from annual reports and vvb data.

Fund	Asset value(bln €)	Funding ratio ultimo 2001
ABP	147.3	1.12
Shell	12.2	1.39
PGGM	49.1	1.12
Philips	14.7	1.28
BPMT	17.5	1.18
SPF	9.9	1.45

tive correlation.

## 4.5 Empirical evidence

The previous sections have shown the specific investment policies that are optimal in the loss aversion framework for a stylized pension fund. In this section, we seek empirical evidence that can shed light on the question whether pension funds are loss averse. We present historical data on the actuarial funding ratios and investment portfolios of Dutch pension funds. If loss aversion is a plausible assumption for the preferences of a pension fund, and if pension funds invest rationally, then the strategies that are followed should be comparable with the optimal V-shaped strategies found in this chapter.

In this section we provide data on the actual financial positions of Dutch pension funds. If pensions funds are loss averse, some of them might already be on the left-hand side of a V-shaped investment rule. Whether funds' investment policies actually show such V-shaped behavior is explored in more detail.

Table 4.2 shows the asset values and funding ratios of six large Dutch pension funds. ABP is the pension fund for all Dutch government personnel, and the largest in terms of asset value The Netherlands. The Shell pension fund is for employees of Royal Dutch Shell. PGGM is the fund for workers in the health sector. Philips is obviously for Philips workers. BPMT is the fund for people working in the steel and technical industries. Finally, SPF is the fund that covers the pension for workers at one of the companies associated with the Dutch railways. In total, these funds represent 58% of the total of pension assets in The Netherlands of €435 billion.

All six funds in Table 4.2 are overfunded, i.e., they have a funding ratio

**Table 4.3: Asset returns 1997-2001**

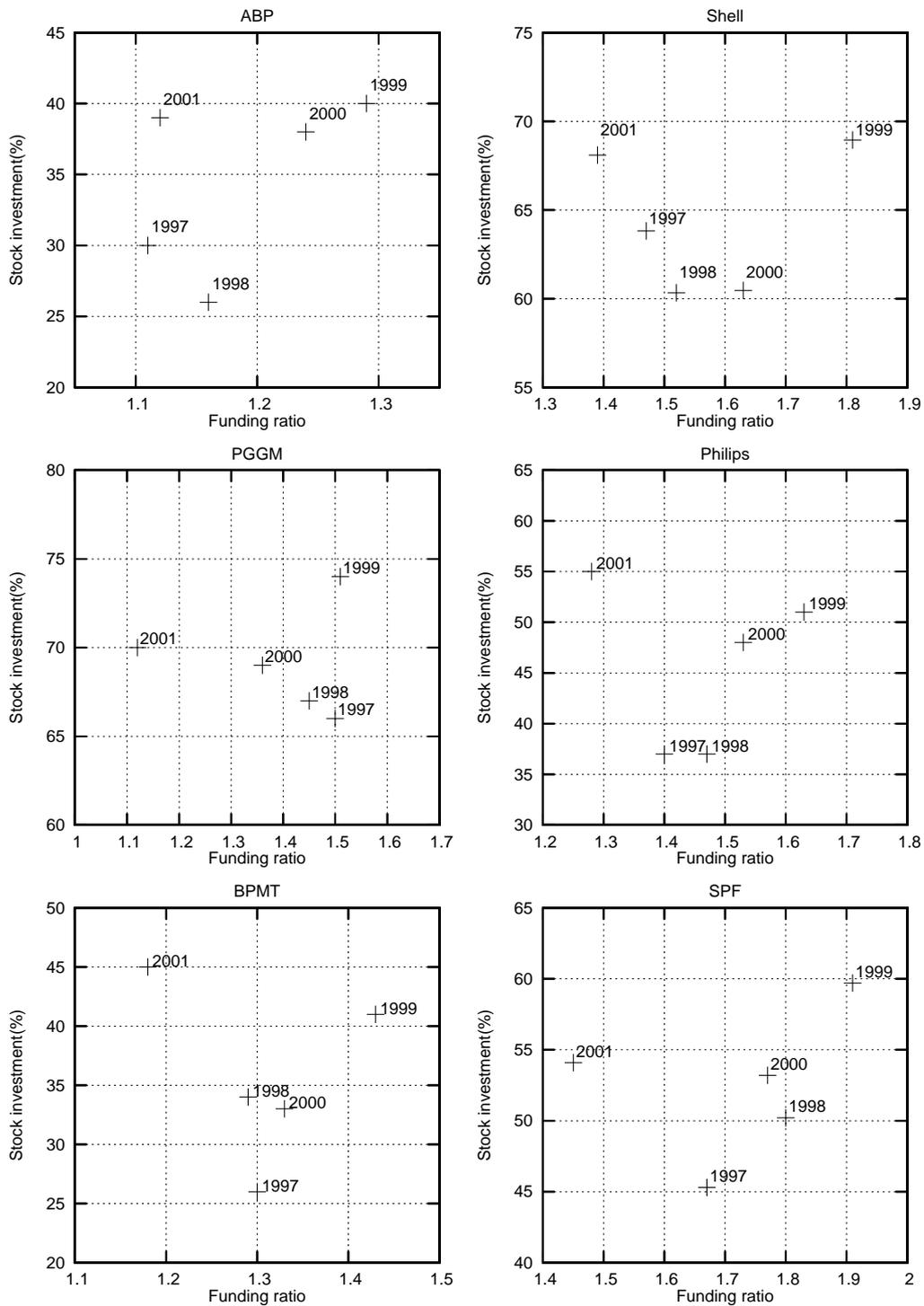
This table gives the yearly stock returns as measured by the MSCI world index, measured in local currency. Bonds is the Lehman Brothers Aggregate Bond Index return. Interest is the 1-Year Treasury Constant Maturity Rate

year	MSCI	Bonds	Interest
1997	20.8%	10.2%	5.6%
1998	19.2%	8.6%	5.1%
1999	26.3%	-2.0%	5.1%
2000	-10.8%	9.4%	6.1%
2001	-15.3%	10.4%	3.5%

larger than 100%. We saw in Subsection 4.3.3 that the kink of the V-shape does not necessarily have to lie at 100% actuarial funding. While there are worries about the diminishing buffers of pension funds against the 4% actuarial discount rate, the problems could be even more pronounced if the market value of liabilities is taken to calculate the funding ratio.

Having seen current funding ratios, we are interested in actual investment policies over the past 5 years. Under loss aversion and rationality, we would expect to see patterns of investment that show resemblance with some of the V- or U-shaped patterns that were optimal for our loss averse model. Figure 4.8 presents the scatter plots of the funding ratio versus fractional stock investment for the six pension funds listed in Table 4.2. Table 4.3 lists annual returns for representative stock and bond indices.

As can be seen in Table 4.3, from 1997 to 2000 we have witnessed an enormous surge in stock prices, followed by decline in 2000 and 2001. Bond return have shown the exact opposite patterns, while one-year interest rates have been fairly stable. For pension funds, the variability of liability growth is very small compared to that of stock returns. Therefore, the movement of the funding ratios in Figure 4.8 is for the most part determined by the investment returns. For all funds we see the same relative movement of the funding ratio: increasing from 1997 to 1999, decreasing from 1999 to 2001. The only exceptions are PGGM and BPMT, who show a slight decrease of the funding ratio over the years 1997-1998. Note that stock investments do not have to be the only risky investments that pension funds have. Several funds invest a significant part of their assets, i.e., more than 5%, in risky investments such as private equity or real-estate. However, the fractions invested in those categories were either too small or relatively stable over time, so they are not considered in the analysis.



**Figure 4.8: Stock investment in relation to reported funding ratios**

This figure shows funding ratios and fractions invested in stocks for six large Dutch pension funds. Data is from annual reports of the respective funds, Vereniging van Bedrijfstakpensioenfondsen (vfb). Where necessary, liabilities have been computed through (linear) extrapolation.

Ultimo 2001, the funds differ considerably in the fraction of stocks they have. ABP and BPMT have around 40% stocks, with funding ratios of 1.1 and 1.2, respectively. Philips and SPF have 55%, with funding ratios of 1.3 and 1.45. PGGM and Shell have a high 70% stock investment, and funding ratios of 1.1 and 1.4. If we look at the three pairs of funds with comparable stock investments, PGGM stands out, having one of the lowest funding ratios, with the highest stock investment. It also has the most stable fraction of stocks over time, so it might be the result of taking a long-term calculated risk, rather than a double-or-nothing policy.

Turning to the investment patterns present in Figure 4.8, we see a diffuse behavior for the first period 1997-1998. Stock returns were very high in this year, but pension funds do not show a shared opinion on whether to increase or decrease the fraction of stocks. In the top panels, ABP and Shell had an increasing funding ratio, and decreasing stock fraction. The middle two panels show funding ratios and stock fractions with only little variation. The lower two panels show an increasing stock fraction with either a stable or increasing funding ratio. Clearly, investment policies differed to a large extent during the year 1997.

From 1998 onward, however, identical patterns emerge. At the end of 1999 the fraction of stocks has increased with the funding ratio, and has decreased again at the end of 2000. From 2000 to 2001, the funding ratios decrease further, but now the stock investments increase. The extent of the stock increase in the last year is only small for ABP, PGGM, and SPF, but given the large decreases in funding ratios (see the large negative returns on stocks), it represents a significant investment decision for all funds. The pattern can partly be explained by funds following a fixed-mix strategy. For four of the six panels, namely ABP, PGGM, Philips and SPF, one could imagine that the observed stock investments are the result of a fixed-mix strategy, where rebalancing occurs at the beginning of each year. For SPF, ABP, and Philips, though, we need to add the assumption that the strategic asset mix has changed after 1997 or 1998. For two of the six funds, namely Shell and BPMT, it is difficult to see how a fixed-mix policy would lead to the observed patterns. Over the period 1998-2001, both funds show a series of consecutive increasing and decreasing stock fractions of more than 7%-points.

At the end of Subsection 4.3.1, we have shown that buy-and-hold and fixed-mix strategies have different risk-return properties, the attractiveness of the strategies depending on whether the initial position has a positive or negative surplus. This result might explain a pattern that is seen in for example the panel for SPF. From 1997 to 2000, the points are more or less on a straight line, implying a buy-and-hold policy. Such a policy was preferred

in a situation of positive surplus in our model. Then, from 2000 to 2001, the asset mix remains fixed. This might be compared with the better risk-return profile of the buy-and-hold policy when the initial situation is one of underfunding.

Clearly, the preliminary analysis of Figure 4.8 can at best be only suggestive about the validity of loss averse preferences for pension funds. The main puzzle so far is to explain the increase or non-decrease in the fraction of stocks in the portfolio from 2000 to 2001, despite two consecutive large negative stock returns. So far, it appears that loss aversion might help in at least explaining part of this puzzle.

## 4.6 Discussion and conclusion

This chapter has given an interpretation of the loss averse model introduced in Chapter 2, where loss aversion was interpreted as an aversion to (actuarial) funding ratios below 100%. The sensitivity of the optimal investment strategy has been explored for changes in the economic assumptions on inflation, the correlation between stock returns and inflation, and the level of the real interest rate. Results for a different risk-measure, namely downside-deviation were also obtained, closely resembling those of the mean-shortfall model.

Section 4.5 considered patterns of portfolio adjustment versus funding ratios for six large Dutch pension funds. It did not give clear evidence that the pattern of portfolio adjustment by pension funds is in accordance with the optimal dynamic strategy under loss aversion. Further research using a larger and richer dataset should be done to come up with better representations of the investment strategies pension funds are following.



---

# Household Savings under Loss Aversion

---

## 5.1 Introduction

For at least half a decade, the life-cycle framework has been the standard way for economists to model household behavior with respect to the allocation of time, effort, and money. As explained in Browning and Crossley (2001), the attractive properties of the life-cycle framework is that it propagates the idea that agents make sequential decisions to achieve a coherent goal using currently available information as best they can. A direct consequence of this idea is that the economic behavior of households should be aimed at keeping the marginal value of money constant over time, representing the optimum in a dynamic optimization. Within the framework of the life-cycle hypothesis, the theory has led to a large theoretical and empirical literature on consumption. However, in an overview of theories and facts on household saving, Browning and Lusardi (1996) note that compared with the large body of literature on consumption, there is a “relative ignorance that still surrounds the reasons why households save”. Especially the existence and level of precautionary savings is not completely understood. The precautionary savings motive represents the idea that some households save more because they face more income uncertainty than others. Carroll (1998) notes that the range of results that give a relation between savings and uncertainty is disturbingly large. A major obstacle in estimating such relationships is that theory provides no analytical result on how to specify

uncertainty.

In this chapter we use the mean-shortfall objective from Chapter 2 to represent household preferences with respect to consumption. That way, consumption risk is represented as the explicit measure of expected shortfall below a reference level. In analyzing the results we specifically look at the relation between savings and wealth. As the model includes a benchmark level of consumption, we come quite naturally to a notion of precautionary savings. Precaution is seen as a motive to explain consumption patterns in empirical studies, as in Caballero (1990), Normandin (1994), and Normandin (1997). Another issue is the way in which uncertainty influences savings. Aizenman (1998) shows that disappointment aversion induces a positive relation between uncertainty and the size of a buffer stock. Modeling a household as a loss averse agent in Section 5.2, we derive the explicit decision rules for consumption and savings behavior in Section 5.3. There, we also trace the relation between uncertainty and savings. Section 5.4 introduces habit formation and reveals how it affects the optimal consumption behavior. Section 5.5 shows the optimal consumption rules when the value function from Kahneman-Tversky is taken as a utility function. Section 5.6 concludes.

## 5.2 The model

We consider an agent with current wealth  $w$  who lives two periods and has to decide on his current saving, which is denoted as the variable  $s$ . His first and second period budget constraints are

$$c_1 = w - s \tag{5.1}$$

$$c_2 = s \cdot r, \tag{5.2}$$

where  $c_1, c_2$  denote consumption over period  $i$ ,  $i = 1, 2$ , and  $r$  is the real return on investment. We assume  $c_1 > 0$  and that the return  $r$  has an arbitrary absolute-continuous probability distribution function  $G(\cdot)$  with support  $(0, \infty)$ . It represents the gross return on investment, which can be produced by a simple savings account, a portfolio of stocks, etc. Having an arbitrary probability distribution  $G(\cdot)$  contrasts with Aizenman (1998) and Bowman et al. (1999), who also examine loss aversion but restrict uncertainty in income to a binomial type of probability distribution. If in the following the terms negative and positive return are used, they indicate that we are talking about  $r$  in terms of the *net* return, i.e.,  $r - 1$ , which can take on values in  $(-1, \infty)$ .

We assume that the agent or household has a per-period loss averse utility

function

$$u(c_i, b_i) = c_i - \lambda \cdot (b_i - c_i)^+, \quad (5.3)$$

where  $\lambda > 0$  is a loss aversion parameter,  $(y)^+$  is used to denote the maximum of 0 and  $y$ . The parameter  $b_i$  is the benchmark level of consumption, representing the level of consumption below which a loss is suffered, be it physical, financial or mental. In Section 5.3 we assume  $b_i$  is equal to the fixed level  $b$ . Section 5.4 deals with the case of  $b_2$  being a function of  $b_1$  and consumption in the first period,  $c_1$ .

The utility function in (5.3) incorporates the main features of Kahneman and Tversky's value function, as it measures utility relative to a reference point, and treats positive and negative deviations from the reference point asymmetrically. The penalty parameter  $\lambda$  represents a measure for the costs of not achieving a desired consumption level. The utility function in (5.3) has been used before by Benartzi and Thaler (1995) in an attempt to explain the equity premium puzzle. It is also the piecewise-linear equivalent of the one used by Aizenman (1998), where  $\lambda$  is the disappointment aversion rate and  $b$  the certainty equivalent consumption. Bowman et al. (1999) derive axiomatic results for loss averse utility functions of which (5.3) is a special case. Note that the two-period model can still be considered to be within the standard life-cycle framework. Having two periods can be seen as modeling life-time consumption in which smoothing happens at low frequencies, i.e., across two specific stages of the life-cycle. This observation puts the remark by Browning and Crossley (2001) in context, who note that many psychological or behavioral explanations rule out the life-cycle framework. Given the behavioral roots of the utility function above, we must conclude that they probably refer to work, e.g. Thaler (1994), in which no formal optimization of consumption and savings is considered. Clearly, that is not the case here.

With (5.3) as the instantaneous utility function, the agent determines savings through the following maximization problem

$$\max_{c_1} c_1 - \lambda \cdot (b - c_1)^+ + \mathbb{E}(c_2) - \lambda \cdot \mathbb{E}(b - c_2)^+, \quad (5.4)$$

subject to budget constraints (5.1) and (5.2). Note that we abstract from time-discounting. In the above model, it would be appropriate to include a parameter  $\rho$  for the last two terms, representing a discount with respect to future consumption. However, this only increases the number of parameters in the model, without changing the results. If appropriate, in the following we will add a note to show how the results would change if time-discounting was included.

## 5.3 Results

The main results follow from the following theorem.

**Theorem 5.3.1** *Under the condition*

$$\lambda > \max\left(0, \frac{1 - \mathbb{E}[r]}{\mathbb{E}[r]}\right), \quad (5.5)$$

the optimal solution  $0 \leq c_1^* \leq w$  to (5.4) is given by

$$c_1^* = \begin{cases} \min(b, w - b/r_p^*) & \text{if } \mathbb{E}[r] \geq 1, \\ \max(b, w - b/r_n^*) & \text{if } \mathbb{E}[r] < 1. \end{cases} \quad (5.6)$$

where  $r_p^*$ ,  $r_n^*$  are the two  $\bar{r}$ s that solve

$$\int_0^{\bar{r}} r dG = \frac{1 - \mathbb{E}[r]}{\lambda} + 1_{\{\mathbb{E}[r] > 1\}}, \quad (5.7)$$

and  $I_{\{A\}}$  is the indicator function with respect to the event  $A$ .  $r_p^*$  solves (5.7) for a positive and  $r_n^*$  for a negative (net) expected return.

The wealth levels at which consumption is equal to the benchmark are given by  $w_p^*$  and  $w_n^*$ , corresponding to a positive and negative expected return economy, respectively. They are defined as

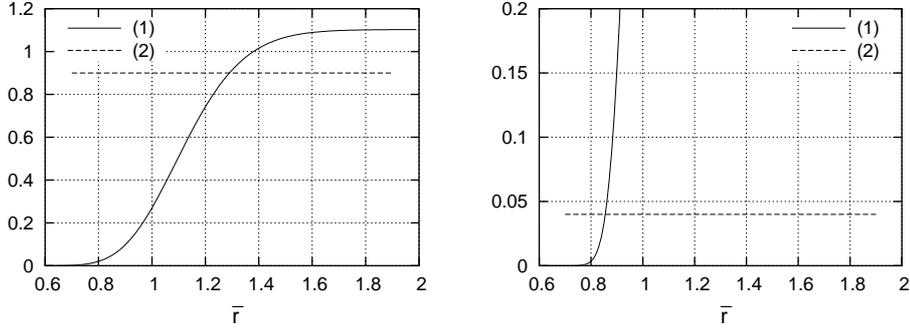
$$w_p^* = b \cdot (1 + 1/r_p^*), \quad (5.8)$$

$$w_n^* = b \cdot (1 + 1/r_n^*). \quad (5.9)$$

**Proof:** See appendix.

Equation (5.6) in Theorem 5.3.1 gives parametric rules for the optimal consumption strategy, which differs for positive and negative expected real returns. For two different expected returns, Figure 5.1 graphically shows the solutions to equation (5.7). Note that although a positive expected return on investment is the most common assumption, one can easily imagine a situation in which the expected positive (nominal) return on investments is smaller than the combined effect of inflation and time-discounting. These situations have occurred in economies with hyperinflation for example. Note that a distinction in consumption/savings behavior based on the sign of the expected return on savings has not been documented before in the literature.

To facilitate the interpretation of Theorem 5.3.1, we plot the optimal consumption as a function of wealth in Figure 5.2, for a negative and a positive expected return, respectively. The two levels  $w_p^*$  and  $w_n^*$  give the



**Figure 5.1: Solving equation (5.7)**

In the figures, the line (1) represents the left-hand side of equation (5.7) as a function of  $\bar{r}$ . The line (2) represents the right-hand side of the equation. The point of intersection gives the  $\bar{r}$  that solves (5.7).

In the left panel,  $\lambda = 1$ , and  $r \sim \log N(0.085, 0.16)$ , so that  $\mathbb{E}[r] = 1.10$  and the line (2) lies at 0.9. The right panel has  $\lambda = 1$ ,  $r \sim \log N(-0.04, 0.07)$ , so that  $\mathbb{E}[r] = 0.96$ , and the line (2) lies at 0.04

point of the kink in the consumption rule for positive and negative expected return, respectively.

In the case of a positive expected return, optimal consumption is the result of a trade-off between first and (expected) second period utility. There is an upper bound at level  $b$ , which is reached for wealth equal to or larger than  $w_p^*$ . Any additional wealth above  $w_p^*$  is fully put into savings.

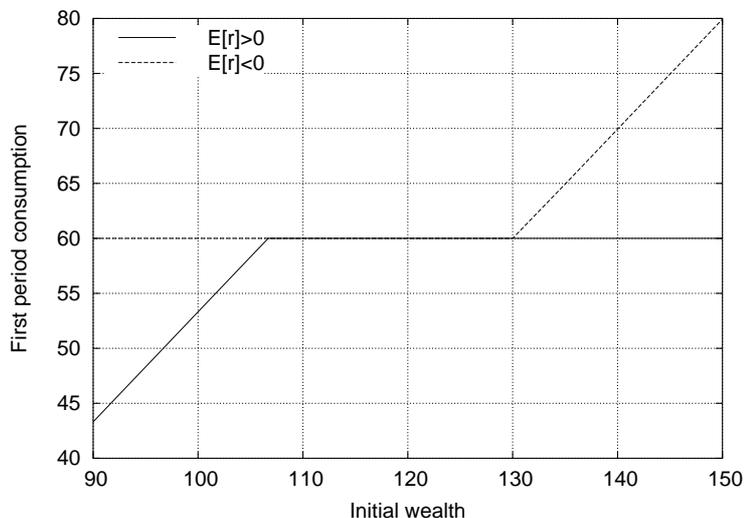
A negative expected return leads to having a lower bound on first period consumption at level  $b$ . This level is reached for wealth below  $w_p^*$ . Above  $w_n^*$ , optimal consumption results from a trade-off between current and future utility of consumption. Note that in the figure  $w_p^* < w_n^*$ . This need not be the case in general, as  $w_p^*$  and  $w_n^*$  represent wealth levels in different economies.

Given Theorem 5.3.1, we can completely specify a parametric relation for optimal savings. This contrasts with other studies like Bowman et al. (1999), Aizenman (1998) and Carroll (1998), that have derived axiomatic or numerical results for the consumption/savings decision of loss averse agents under uncertainty. In our model the explicit savings rules are given by

$$s^* = \begin{cases} \min(w - b, b/r_n^*) & \text{if } \mathbb{E}[r] \geq 1, \\ \max(w - b, b/r_p^*) & \text{if } \mathbb{E}[r] < 1. \end{cases} \quad (5.10)$$

Expression (5.10) leads to a number of interesting results.

First, observe that the derivative of (5.10) with respect to wealth  $w$  is either 0 or 1, implying that savings is nondecreasing in wealth. This can be

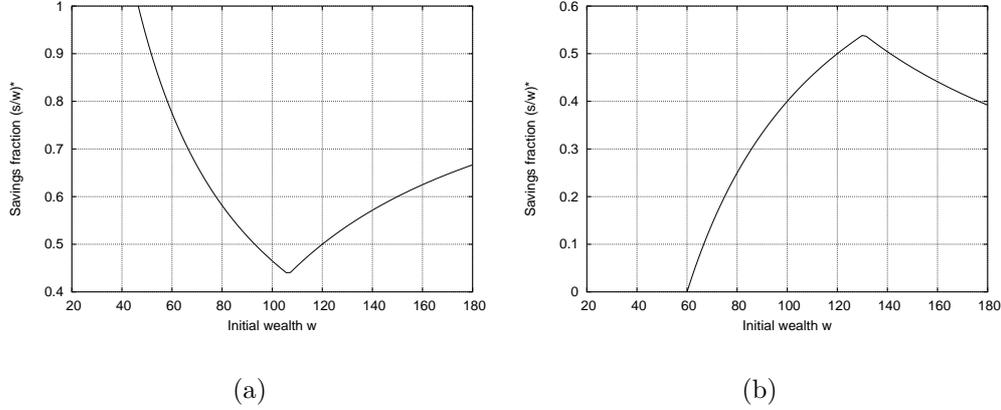


**Figure 5.2: Optimal period 1 consumption as a function of initial wealth**

For the positive expected return,  $r$  has a lognormal distribution with  $\mu = 0.085$ , and  $\sigma = 0.16$ . For the negative expected return,  $G$  is lognormal with  $\mu = -0.04$ , and  $\sigma = 0.07$ . Other parameter values are  $\lambda = 1$  and  $b = 60$ .

explained by the loss aversion with respect to period 2 consumption: even if the return on savings is negative, the punishment on period 2 shortfall outweighs the expected negative return on saving.

Second, (5.10) reveals the relation between the return distribution  $G(\cdot)$  and the amount of savings. The dependence on the return distribution only enters through the values of  $r_p$  and  $r_n$ . Carroll (1998) shows that in other theoretical models of consumption and savings behavior there can be multiple measures of income uncertainty. This implies that for many of these models it is not trivial to give the relation between savings and uncertainty. In the model presented here, however, the per-period utility function in (5.3) gives an explicit and unequivocal measure of the relevant uncertainty, namely downside-risk. It is the lower expectation of  $r$  below the return level  $r_p$  or  $r_n$ , depending on whether the expected savings return is positive or negative. Moreover, from Theorem 5.3.1 it follows that for a given amount of savings, the riskiness of the investment return influences saving unambiguously. To see this, observe from equation (5.7) that if the expectation of  $r$  below  $\bar{r}$  increases,  $\bar{r}$  itself should decrease to satisfy the equation. As  $r_p^*$  and  $r_n^*$  have a nonnegative effect on optimal savings through equation (5.10) it follows that an increase in the riskiness of period 2 wealth has a nondecreasing



**Figure 5.3: Optimal savings fraction**

This figure shows the optimal fraction of wealth saved,  $s_0^*$ , as a function of wealth at time 0,  $w$ . For both graphs  $\lambda = 1$ ,  $b = 60$ . In (a)  $r$  is distributed lognormal with  $\mu = 0.085$ ,  $\sigma = 0.16$ , so  $\mathbb{E}[r] > 1$ . In (b)  $G$  is also lognormal,  $\mu = -0.04$ ,  $\sigma = 0.07$ , so  $\mathbb{E}[r] < 1$ .

effect on savings. This means that precautionary saving, i.e. saving to build up a buffer against future risk, is a motive for a loss averse agent in our model, regardless of the expected return on saving. It corroborates the result by Aizenman (1998), who finds that for disappointment averse (i.e. loss averse) developing countries, buffer stocks are efficient, with size increasing in the volatility of second period income. It also confirms a result by Bowman et al. (1999) for the case of general loss averse utility functions. Restricting income uncertainty to a binary symmetric distribution, they find that an increase in uncertainty of income increases savings.

Finally, we derive the optimal *fraction* of wealth saved. This quantity is relevant because most empirical studies on savings behavior consider this measure as a dependent variable. Otherwise, savings behavior is not comparable between households with different wealth levels. Dividing left- and right-hand sides of equation (5.10) by  $w$ , the fraction of wealth saved is clearly restricted to the unit interval and given by

$$(s/w)^* = \begin{cases} \max\{1 - b/w, \frac{b}{w \cdot r_p^*}\} & \mathbb{E}[r] \geq 1, \\ \min\{1 - b/w, \frac{b}{w \cdot r_n^*}\} & \mathbb{E}[r] < 1. \end{cases} \quad (5.11)$$

Figure 5.3 shows a graph of the savings fraction as a function of initial wealth.

Browning and Lusardi (1996) note on the motives for saving by households that “there is a widespread feeling that the wealthy have different mo-

tives to save from the less wealthy.” This suggestion is explicitly confirmed and illustrated through (5.11), as illustrated in Figure 5.3. In discussing the difference in savings between the poor and the wealthy, we treat economies with a positive and a negative real return separately.

To start with a positive return-economy, the kink in the savings-rule in Figure 5.3(a) lies at  $w_p^*$ . Define rich households to have wealth above  $w_p^*$  and poor household to have wealth below  $w_p^*$ . Figure 5.3(a) then shows that for poor households a higher initial wealth results in a smaller fraction of wealth saved. For wealthy households the opposite effect holds: an increase in wealth results in increased relative savings. From the list of motives given by Browning and Lusardi (1996), the behavior of the poor households could be well explained by the motive of precautionary savings. As wealth increases, less money needs to be saved to reach the period 2 benchmark  $b$ . For wealthy households, the savings pattern is explained by the intertemporal-substitution motive. As wealth increases and expected return is positive, more money is saved to enjoy a higher expected period 2 consumption.

For a negative return-economy Figure 5.3(b) shows a completely different situation. The fraction of wealth saved is increasing in wealth for poor households and decreasing in wealth for rich households. The contrast with the positive expected return economy can be explained through the fact that under a negative expected return, the marginal effect of increased savings on period 2 utility is lower. Therefore, once wealth is so high that the interior optimum exceeds the threshold, any extra wealth is consumed in period 1. Hence the fraction saved decreases when wealth exceeds the (higher) threshold.

Finally, we can take the freedom to view the comparative statics of Figure 5.3 as the decision rules for household saving with wealth as input, as done in Bowman et al. (1999). In a setting with loss aversion, Bowman et al. (1999) focus on the effects of shocks to income on consumption. They find evidence that gains and losses in wealth have an asymmetric impact on household consumption and saving behavior. In the current model we have made this effect explicit in the optimal savings rules of Theorem 5.3.1, illustrated in Figure 5.3. The fraction saved either increases or decreases in wealth, depending on which side of the threshold the agent’s wealth is.

## 5.4 Habit formation

The benchmark level  $b$  in the formulation of the model in Section 5.2 is fixed. However, the formulation of the objective function and state equation, allow for a natural extension in which the benchmark in the second period

is partly determined by consumption in the first period. Such a mechanism is called habit formation. Habit formation is used in several contexts in economics where consumption preferences are modeled. Constantinides (1990) uses habit formation in an attempt to resolve the equity premium puzzle. Recent papers by Seckin(2000a,b) explore the effect of habit formation on precautionary savings in a classical model of consumption and savings. Habit formation is also in the loss averse model of the previously cited study of Bowman et al. (1999).

Duesenberry (1949) discusses habit formation, showing that it implies that consumption decisions are not independent of each other. This is another way of saying that with habit formation, the per-period utility function is not time-separable, complicating the analysis of standard consumption models, see the discussion in Browning and Lusardi (1996). In the current section we show that including habit formation in the model of Section 5.2 retains tractability, while increasing insight in the effects of loss aversion on consumption and savings.

Extending model 5.4 to include habit formation, we introduce two benchmark levels,  $b_1$  and  $b_2$ , for period 1 and 2, respectively. Habit persistence is then modeled by having  $b_2$  depend on period 1 consumption, using the same specification as in Bowman et al. (1999):

$$b_2 = (1 - \alpha) \cdot b_1 + \alpha \cdot c_1, \quad (5.12)$$

where  $\alpha$  denotes the degree of habit persistence.<sup>1</sup>  $\alpha = 0$  implies a fixed benchmark for both periods, as in the previous sections.  $\alpha = 1$  implies that the period 2 benchmark is equal to period 1 consumption.

**Theorem 5.4.1** *Consider the cases of a positive(I) and negative(II) expected return separately.*

I. If  $\mathbb{E}[r] > 0$ , define

$$\bar{w}_p = b_1 \cdot (1 + 1/r_p^*), \quad (5.13)$$

where  $r_p^*$  is the  $\bar{r}$  that solves

$$\lambda - \mathbb{E}[r] - \lambda \int_0^{\bar{r}} (\alpha + r) dG = 0. \quad (5.14)$$

---

<sup>1</sup>Note that it is possible to use a separate  $\alpha$  for upward and downward adjustments to  $b_2$ , say  $\alpha^u$  and  $\alpha^d$ . This has a trivial effect on the optimal solution however, and does not structurally change the results.

The optimal consumption rule is given by

$$c_1^* = \begin{cases} w - \frac{\alpha w + (1-\alpha)b_1}{r_p^* + \alpha} & \text{if } w \leq \bar{w}_p, \\ b_1 & \text{if } w > \bar{w}_p. \end{cases} \quad (5.15)$$

II. If  $\mathbb{E}[r] < 0$ , define

$$\bar{w}_L = b_1 \cdot (1 + 1/\bar{r}_n^L), \quad (5.16)$$

$$\bar{w}_H = b_1 \cdot (1 + 1/\bar{r}_n^H), \quad (5.17)$$

where  $\bar{r}_n^L$  is the  $\bar{r}$  that solves

$$\lambda - \mathbb{E}[r] - \lambda \int_0^{\bar{r}} (\alpha + r) dG = 0, \quad (5.18)$$

and  $\bar{r}_n^H$  solves

$$-\mathbb{E}[r] - \lambda \int_0^{\bar{r}} (\alpha + r) dG = 0. \quad (5.19)$$

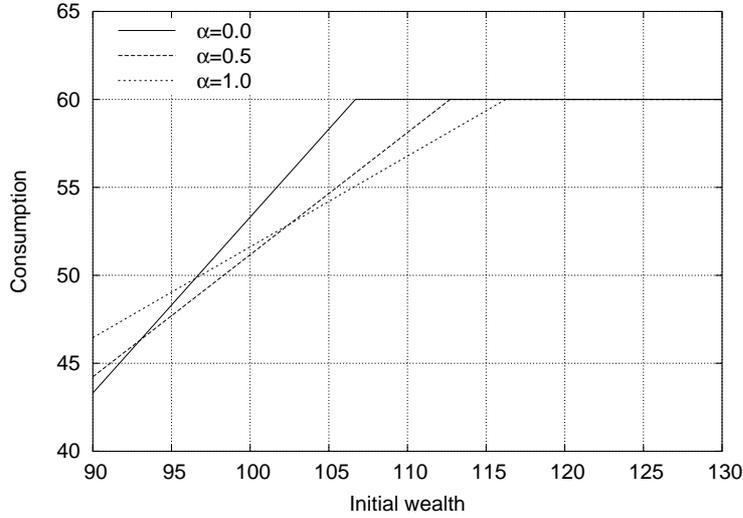
From equations (5.18) and (5.19) it follows that  $\bar{r}_n^L > \bar{r}_n^H$ , so  $\bar{w}_L < \bar{w}_H$ . Optimal consumption is given by

$$c_1^* = \begin{cases} w - \frac{\alpha \cdot w + (1-\alpha)b_1}{\bar{r}_n^L + \alpha} & \text{if } w \leq \bar{w}_L, \\ b_1 & \text{if } \bar{w}_L < w < \bar{w}_H, \\ w - \frac{\alpha \cdot w + (1-\alpha)b_1}{\bar{r}_n^H + \alpha} & \text{if } w \geq \bar{w}_H. \end{cases} \quad (5.20)$$

**Proof: See appendix.**

To analyze the consequences of Theorem 5.4.1, let us start with the case of positive expected return, case I in the theorem. From equation (5.15), we see that the optimal consumption decision looks similar to that of the model without habit formation, see equation (5.6) in Theorem 5.3.1. The difference lies in the influence of the degree of habit persistence  $\alpha$ , that is influencing the threshold wealth and the slope of the decision rule. To see the effect of  $\alpha$ , Figure 5.4 shows the optimal consumption as a function of initial wealth for different values of  $\alpha$ .

For  $\alpha = 0$ , i.e., no habit persistence, the optimal consumption rule is the same as in the previous section, compare Figure 5.2. Observe that  $\bar{w}_p$  is the wealth level at the kink. When  $\alpha$  increases, we see that the threshold wealth  $\bar{w}_p$  increases and the slope of the optimal consumption rule decreases.



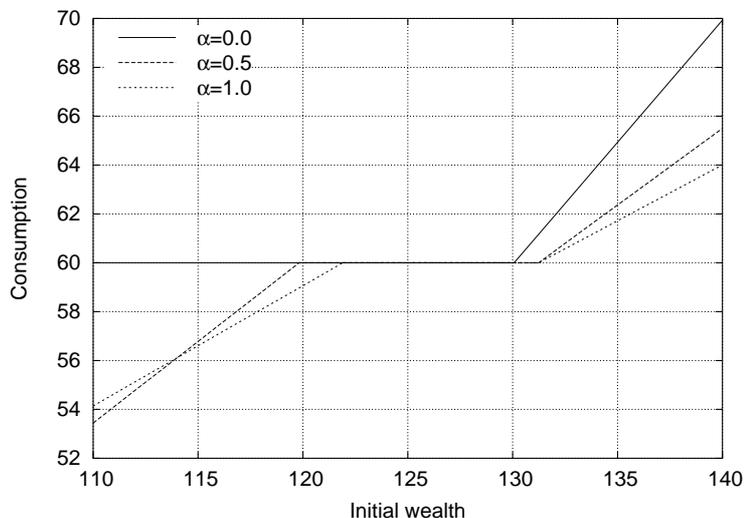
**Figure 5.4: Consumption with habit formation and positive expected return**

For a positive expected return,  $\lambda = 1$ ,  $b_1 = 60$ , this figure shows the optimal consumption as a function of initial wealth for varying degrees of habit persistence.  $\alpha = 0$  corresponds to no habit persistence,  $\alpha = 1$  implies  $b_2 = c_1$ . Parameter values are  $\mu = 0.085$ ,  $\sigma = 0.16$ ,  $\lambda = 1$ .

Both these observations follow more or less directly from Theorem 5.4.1. From equation (5.14) it is clear that  $\alpha$  has a decreasing effect on the upper integral limit  $\bar{r}_p$ , which has itself a decreasing effect on  $\bar{w}_p$ . From the optimal consumption rule in (5.15) we see that the effect of  $\alpha$  on the slope of the decision rule is given by  $-\alpha/(\bar{r}^* + \alpha)$ , which is decreasing in  $\alpha$ . Comparing habit-formers ( $\alpha = 1$ ) to agents without habit formation ( $\alpha = 0$ ) shows that for high wealth levels a lower consumption is accepted for the benefit of a lower second period reference point. For low wealth, consumption may be higher, as the second period reference point will be lower anyhow.

For the case of a negative expected return, the optimal consumption rule is visualized in Figure 5.5.

Let us first concentrate on the left side of Figure 5.5. When there is habit persistence ( $\alpha > 0$ ), we see that below the threshold  $\bar{w}_L$ , first period consumption is not kept at the benchmark level. Again, it is the combination of increased savings and a lower period 2 benchmark level, that makes this an optimal strategy. We see that for  $\alpha = 0$ , the increase of savings alone does not justify a consumption below the benchmark, as the return on savings is negative. The change in slope of the consumption decision shows the same



**Figure 5.5: Consumption with habit formation and negative expected return**

For a negative expected return,  $\lambda = 1$ ,  $b_1 = 60$ , this figure shows the optimal consumption as a function of initial wealth for varying degrees of habit persistence.  $\alpha = 0$  corresponds to no habit persistence,  $\alpha = 1$  implies  $b_2 = c_1$ . Parameter values are  $\mu = -0.04$ ,  $\sigma = 0.07$ ,  $\lambda = 1$ .

behavior as for the positive return case.

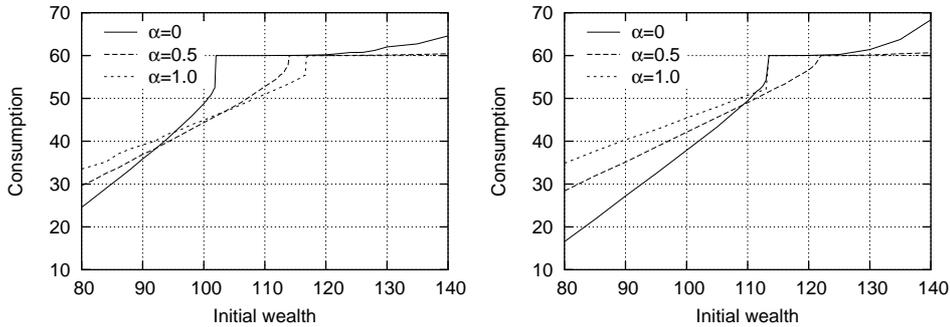
For larger wealth levels, Figure 5.5 shows that habit persistence leads to a decrease in consumption relative to the case  $\alpha = 0$ . That is because an increase in consumption does not only decrease period 2 expected consumption, but also increases the period 2 benchmark and thus decreases utility further.

## 5.5 Kahneman & Tversky preferences

The Kahneman-Tversky value function as estimated in Tversky and Kahneman (1992) is given by

$$v(c) = \begin{cases} x^{0.89} & \text{if } x \geq 0, \\ -2.25 \cdot (-x)^{0.89} & \text{if } x < 0, \end{cases} \quad (5.21)$$

where the variable  $x$  represents the deviation from the reference point. See also Chapter 3. In the context of consumption and savings, we simply take  $x = c - b$  and solve the optimization problem with (5.21) as instantaneous



**Figure 5.6: Optimal solution for the Kahneman-Tversky objective function**

The left panel shows the results for a positive expected return,  $r \sim \log N(0.085, 0.16)$ , the right panel for a negative expected return  $r \sim \log N(-0.04, 0.16)$ .  $\alpha = 0$  corresponds to no habit persistence,  $\alpha = 1$  to full habit persistence. Parameter values are  $b_1 = 60$ .

utility function. The results are in Figure 5.6 for both positive and negative expected return.

Comparing Figure 5.6 with Figures 5.4 and 5.5, we see a large degree of similarity. There are two differences, however. First, for  $\alpha = 0$  in the positive-return case, for Kahneman-Tversky utility the first-period consumption increases beyond the benchmark, while for bilinear utility the benchmark is never exceeded. This is a consequence of the convexity of the behavioral value function in the domain of losses. The convexity is not large enough to increase consumption significantly above the benchmark when there is habit persistence though. A second difference is in the shape of the lines in Figure 5.6. They resemble the shape of the value function itself: a small area below the reference point has a very high slope, the marginal utility of consumption goes to infinity when consumption approaches the benchmark level. In the figure, we can observe that when consumption is close to the benchmark level of 60, the optimal decision becomes extremely sensitive to the initial wealth level. This is very different from the bilinear objective, where consumption was a bilinear function of wealth.

For the negative expected return-case, the consumption patterns for Kahneman - Tversky utility stay roughly the same. A difference with Figure 5.5 emerges for habit persistence in that consumption above the benchmark does not take off, while for the bilinear objective, consumption rises linearly with wealth. The difference is caused by the concavity in gains for the Kahneman-Tversky objective: the marginal value of first period consumption above the benchmark is decreasing in the level. Together with an increased second-

period reference point and high marginal utility around the reference point, increasing first period consumption is not attractive in the case of Kahneman-Tversky utility.

## 5.6 Conclusion

In this paper we have incorporated a bilinear objective function from Chapter 2 into a standard consumption-savings framework. The formulation makes it possible to derive explicit decision rules for optimal saving. From these rules we concluded that the amount of saving is nondecreasing in wealth and in risk, the latter being a shortfall measure on the left tail of the return distribution. Furthermore, a remarkable result is that policies in terms of the fraction saved differ remarkable between high-wealth and low-wealth households and for different types of economies, i.e., expected returns.

The inclusion of habit formation showed that it can drastically change consumption and savings behavior. For positive as well as negative expected return, it increases the wealth level at which the first-period benchmark is consumed. Consumption below the benchmark decreases first-period utility, but increases second-period utility through the combined effect of increased savings and a lower reference point. The figures of optimal consumption have shown that the average slope of the consumption function as a function of wealth becomes smaller, i.e., consumption changes take place more gradually.

Finally, in accordance with the results of Section 3.2 of Chapter 3, we find that the results found do not change significantly when the Kahneman-Tversky value function is taken as instantaneous utility function. This implies that the results for the model analyzed in this chapter can be relevant for empirical work on consumption, investment, and savings.

# Appendix

## 5.A Proofs

### Proof of Theorem 5.3.1:

We start with a reformulation of the optimization problem in (5.4):

$$\max_{c_1} c_1 - \lambda \cdot (b - c_1)^+ + \mathbb{E}[s \cdot r] - \lambda \cdot \mathbb{E}[b - s \cdot r]^+, \quad (5A.1)$$

$$\text{s.t. } s = w - c_1. \quad (5A.2)$$

Denote the value of the objective function (5A.1) for a fixed value of  $c_1$  by  $v(c_1)$ . The first order condition for an interior optimum is given by

$$\mathbb{E} \left[ \frac{\partial v}{\partial c_1} \right] = 0, \quad (5A.3)$$

where

$$\mathbb{E} \left[ \frac{\partial v}{\partial c_1} \right] = \begin{cases} 1 + \lambda - \mathbb{E}[r] - \lambda \cdot \int_0^{\bar{r}} r dG & \text{if } c_1 \leq b, \\ 1 - \mathbb{E}[r] - \lambda \cdot \int_0^{\bar{r}} r dG & \text{if } c_1 > b, \end{cases} \quad (5A.4)$$

and  $\bar{r}$  is defined as

$$\bar{r} = \frac{b}{w - c_1}. \quad (5A.5)$$

$\bar{r}$  is the threshold return that gives  $b - s \cdot r = 0$ .

The second order condition to problem (5A.1) is given by

$$\mathbb{E} \left[ \frac{\partial^2 v}{\partial c_1^2} \right] = -\lambda \cdot \frac{b^2}{(w - c_1)^3} \cdot g(\bar{r}) < 0, \quad (5A.6)$$

where  $g(\cdot)$  is the density function of the return  $r$ .  $g(\cdot)$  is nonnegative by definition, so the second order condition ensures that any  $c_1 < w$  that satisfies (5A.4) is optimal.

As  $G(\cdot)$  has support on  $(0, \infty)$ , the derivative for  $c_1 \leq b$  as given in (5A.4) is strictly decreasing in  $\bar{r}$ . For  $\mathbb{E}[r] \geq 1$  it follows from (5A.4) that  $c_1^* \leq b$ , as for  $c_1 > b$  the derivative with respect to  $c_1$  is negative for all  $\bar{r}$ . Hence, the appropriate derivative in the case  $\mathbb{E}[r] \geq 1$  is the first line in (5A.4). This includes the case  $\mathbb{E}[r] > 1 + \lambda$  when  $c_1^* = 0$ .

If  $\mathbb{E}[r] < 1$ , it follows from (5A.4) that  $c_1^* > b$ , as for  $c_1 \leq b$  the derivative with respect to  $c_1$  is strictly positive for all  $\bar{r}$ . Hence, the appropriate

derivative is the second line in equation (5A.4). The derivative is zero for a  $\bar{r} < \infty$ , if  $\lambda > (1 - \mathbb{E}[r])/\mathbb{E}[r]$ , which is ensured by condition (5.5).

The solution is now determined by the  $r_p^*$  and  $r_n^*$  that follow from putting the appropriate derivative in (5A.4) to zero and solving for  $\bar{r}$ . As there is no other term including  $c_1$  in (5A.4), the optimal  $c_1^*$  follows directly from the calculated bounds on  $c_1^*$  and the definition of  $\bar{r}$ . It is given by

$$c_1^* = \begin{cases} \min(b, w - b/r_p^*) & \text{if } \mathbb{E}[r] \geq 1, \\ \max(b, w - b/r_n^*) & \text{if } \mathbb{E}[r] < 1. \end{cases} \quad (5A.7)$$

The levels  $w_p^*$  and  $w_n^*$  follow from putting  $c_1^* = b$  and solving for  $w$ . ■

**Proof of Theorem 5.4.1:**

The proof starts along the same lines of the proof of Theorem 5.3.1. The optimization with habit formation is given as

$$\max_{c_1} c_1 - \lambda \cdot (b_1 - c_1)^+ + \mathbb{E}[s \cdot r] - \lambda \cdot \mathbb{E}[b_2 - s \cdot r]^+, \quad (5A.8)$$

$$\text{s.t. } s = w - c_1, \quad (5A.9)$$

$$\text{and } b_2 = (1 - \alpha) \cdot b_1 + \alpha \cdot c_1. \quad (5A.10)$$

Denote the value of the objective function (5A.8) for a fixed value of  $c_1$  by  $v(c_1)$ . The first order condition for an interior optimum is given by

$$\mathbb{E} \left[ \frac{\partial v}{\partial c_1} \right] = 0, \quad (5A.11)$$

where

$$\mathbb{E} \left[ \frac{\partial v}{\partial c_1} \right] = \begin{cases} 1 + \lambda - \mathbb{E}[r] - \lambda \cdot \int_0^{\bar{r}} (\alpha + r) dG & \text{if } c_1 \leq b, \\ 1 - \mathbb{E}[r] - \lambda \cdot \int_0^{\bar{r}} (\alpha + r) dG & \text{if } c_1 > b, \end{cases} \quad (5A.12)$$

and  $\bar{r}$  is defined as

$$\bar{r} = \frac{(1 - \alpha) \cdot b_1 + \alpha c_1}{w - c_1} - 1. \quad (5A.13)$$

The second order condition to problem (5A.8) is given by

$$\mathbb{E} \left[ \frac{\partial^2 v}{\partial c_1^2} \right] = -\lambda \frac{(w - c_1) \cdot \alpha + (1 - \alpha) \cdot b_1 + \alpha \cdot c_1}{(w - c_1)^2} \cdot (\alpha + \bar{r}) \cdot g(\bar{r}), \quad (5A.14)$$

where  $g(\cdot) > 0$  is the density function of the return  $r$ . For any  $c_1 < w$ , expression (5A.14) is negative, ensuring that any  $c_1 < w$  that satisfies (5A.12) is an interior optimal solution.

For  $\mathbb{E}[r] \geq 1$  it follows from (5A.12) that  $c_1^* \leq b_1$ , as for  $c_1 > b_1$  the derivative with respect to  $c_1$  is negative for all  $\bar{r}$ . Hence, the appropriate derivative when there is an interior solution is the first line in (5A.12). If there is no interior solution, either  $c_1^* = 0$  or  $c_1^* = b_1$ . If for a given  $\alpha$ ,  $\bar{r}^\alpha$  solves (5A.12), then the wealth level  $\bar{w}_\alpha^p$  at which  $c_1^* = b_1$  can be derived from the definition of  $\bar{r}$  in (5A.13). It is given by

$$\bar{w}_\alpha^p = b_1 \cdot (1 + 1/\bar{r}^\alpha), \quad (5A.15)$$

which is decreasing in  $\bar{r}^\alpha$ . As  $\alpha > 0$ , it follows from (5A.12) that  $\bar{r}^\alpha$  is decreasing in  $\alpha$ , so  $\bar{w}_\alpha^p$  is increasing in  $\alpha$ . This completes the proof of part I.

Unlike the case without habit formation, if  $\mathbb{E}[r] < 1$  the derivative for  $c_1 < b_1$  is not positive for all  $\bar{r}$ . This can only be true for all  $\alpha$  when we would demand  $(\lambda + 1)(\mathbb{E}[r]) < 1$ , which is the exact opposite of the (reasonable) assumption we made in Theorem 5.3.1. We have to conclude that both lines in (5A.12) can constitute a local optimum, which need to be analyzed both.

Having the two areas for a possible optimum, we analyze the wealth levels for both areas at which  $c_1$  is exactly equal to  $b_1$ . Define

$$\bar{w}_L = b_1 \cdot (1 + 1/\bar{r}_n^L), \quad (5A.16)$$

$$\bar{w}_H = b_1 \cdot (1 + 1/\bar{r}_n^H), \quad (5A.17)$$

where  $\bar{r}_n^L$  and  $\bar{r}_n^H$  are the  $\bar{r}$  that solve (5A.12) for consumption lower and higher than the benchmark  $b$ , respectively. It follows directly from the first order conditions that  $\bar{r}_n^L > \bar{r}_n^H$ , so  $\bar{w}_H > \bar{w}_L$ . The latter implies that for wealth below  $\bar{w}_L$ , the optimal consumption follows from the definition of  $\bar{r}$  and the value of  $\bar{r}^*$  that solves (5A.12) for  $c_1 \leq b_1$ . For wealth above  $\bar{w}_H$ , the  $\bar{r}^*$  for  $c_1 > b_1$  is the relevant one. If wealth lies between the two thresholds, optimal consumption is equal to  $b_1$ . ■



# Explaining Hedge Fund Returns by Loss Aversion

---

## 6.1 Introduction

There is a growing interest in hedge fund performance among investors, academics, and regulators alike. Investors and academics are intrigued by the unconventional performance characteristics of these funds. Regulators on the other hand are concerned with the market impact of hedge funds' reported speculative activities during major market events.

Most hedge funds not being formally regulated, are not limited in the type of assets they can hold. Moreover, they face less restrictions on short sales than standard mutual funds and can be highly leveraged and concentrated in specific sectors, countries and/or asset categories. The fund's management compensation is based on the fund's financial performance, something that is less common for conventional mutual funds. According to Edwards (1999), mutual fund managers are generally compensated a flat-fee structure of assets under management. A typical hedge fund, by contrast, charges a 1% fixed fee and 20% of profits.

Given the dramatic increase in the number of hedge funds over the past decade and their large degree of freedom in investment behavior, there has been much recent attention devoted to measuring the performance and return characteristics of hedge funds, see for example Amin and Kat (2001), Fung and Hsieh (1997), Fung and Hsieh (2001), Agarwal and Naik (2000b), and Mitchell and Pulvino (2001). Popular opinion has it that hedge funds,

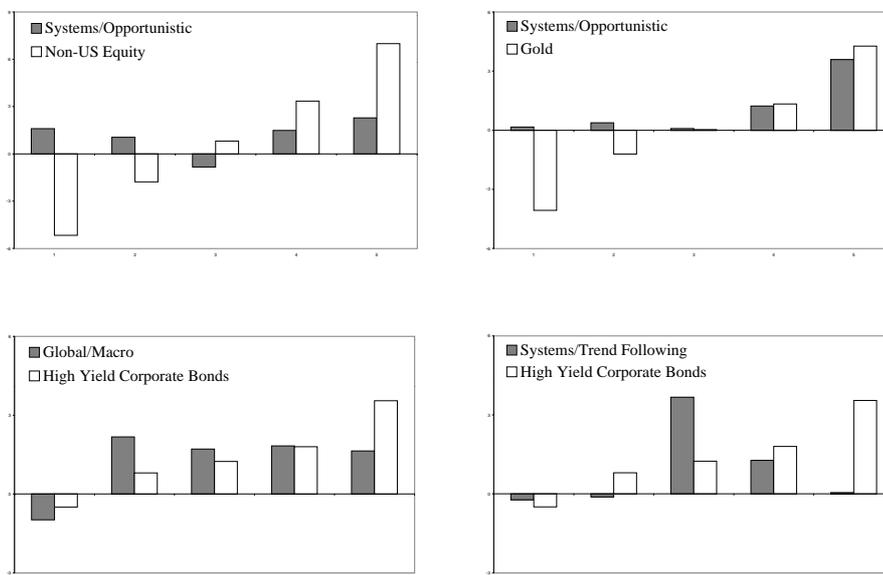
through their large freedom and degree of specialization, deliver exceptionally high returns and are market-neutral. That is, it is believed that hedge fund returns are not correlated with (stock)market returns. If this is true, this is an attractive feature for many institutional investors. The papers cited above concentrate on hedge funds' investment strategies, explaining hedge fund returns empirically. They focus on market efficiency and value-added created by hedge funds and their managers. They find that hedge funds do not outperform stock market returns in terms of average return and volatility. However, the standard method of measuring performance does not work for hedge funds.

A simple approach to performance measurement that works well for standard mutual funds is the asset class factor model of Sharpe (1992). Sharpe models the return  $R_i$  on a fund  $i$  as

$$R_i = w_{i1}F_1 + w_{i2}F_2 + \dots + w_{in}F_n + e_i, \quad (6.1)$$

where  $R_i$  is the vector of returns on fund  $i$ ,  $F_j$  is the value of systematic factor  $j$ , and  $e_i$  presents the non-systematic factors in the return for fund  $i$ . The  $w_{ij}$ s represent the sensitivities of the returns  $R_i$  to factor  $F_j$ . Sharpe selects 12 major asset classes for the factors  $F_1, \dots, F_{12}$  and finds a correlation coefficient of about 90% for a sample of US mutual funds. This implies that the return of mutual funds can be well approximated by a linear combination of returns on standard asset classes. This systematic part of the return is referred to as 'style' by Sharpe. The other part, the return variation that cannot be explained, is called 'selection'. The excess return  $e_i$  should be attributed to the skills of management in selecting the individual securities, hence the name 'selection'. The basic result of Sharpe that standard asset classes can explain most of the variation in mutual funds' returns have been confirmed and extended by various other authors, for example Brown and Goetzmann (1997) and De Roon et al. (2000).

In contrast to the style regressions of type (6.1) for mutual funds, Fung and Hsieh (1997) find that for hedge funds the straightforward approach of Sharpe does not provide a good fit. These results are confirmed by Agarwal and Naik (2000a), who find low correlations with different indices. Also, Brealey and Kaplanis (2001) find evidence for changing factor loadings over time. Fung and Hsieh argue that the lack of fit of (6.1) for hedge funds is due to the extensive use of dynamic trading strategies. Mutual funds generally follow a relatively stable investment strategy, resulting in  $w_{ij}$ 's between zero and one with modest time variation. By contrast, according to Fung and Hsieh hedge funds have weights  $w_{ij}$  between -10 and 10. In addition, the managers' opportunism may cause the  $w_{ij}$ 's to change quickly over time and



**Figure 6.1: Empirical nonlinear relations between systematic hedge fund return factors and standard asset class returns.**

The figures are based on data in Fung and Hsieh (1997), Table 1. We use their systematic risk factors for hedge fund returns corresponding to a specific qualitative descriptor. Given a factor return  $F_t$  and a traditional asset class return  $R_t$ , the pairs  $(F_t, R_t)$  are sorted on  $R_t$  and grouped into quintiles. For each quintile, the average return on  $F_t$  and  $R_t$  is plotted in the graphs.

across market conditions. This helps to explain why traditional Sharpe style regressions (6.1) fail dramatically for hedge funds.

The inappropriateness of Sharpe's model for hedge funds, leads Fung and Hsieh (1997) to consider dynamic strategies. As there is an infinite number of possible dynamic trading strategies, factor analysis is used to determine the dominant styles in hedge funds. They find that adding the style factors to Sharpe's model explains a significantly larger part of the return variation. Moreover, they identify distinctly nonlinear relations between their new factors and traditional asset class returns as used in a typical Sharpe style regression. Figure 6.1 presents four of the most prevalent examples of nonlinear relations as found by Fung and Hsieh.

The hedge fund industry already has its qualitative descriptors for certain types of hedge funds. These can be used to qualify the style factors. Comparable sets of qualifiers are used throughout the hedge fund literature, see for example Osterberg and Thomson (1999), Edwards and Caglayan (2001)

and Brown and Goetzmann (1997). In Figure 6.1 the qualifier for each style factor is from Fung and Hsieh and comes from the fund that has the highest correlation with that style factor. “Systems/Trend Following” refers to traders who use technical trading rules and are mostly trend followers. “Systems/Opportunistic” refers to technically driven traders who also take occasional bets on market events relying on rule-based models. “Global/Macro” refers to managers who primarily trade in the most liquid markets in the world, such as currencies and government bonds, typically taking bets on macroeconomic events such as changes in interest rate policies and currency devaluations. These strategies rely mostly on their assessments of economic fundamentals.

Fung and Hsieh (1997) argue that Figure 6.1 shows option-like pay-off patterns in hedge fund returns. For example, the upper-left panel may be identified as a long straddle, whereas the upper-right and lower-left panels roughly correspond to long call and short put positions, respectively. Fung and Hsieh (2001) then proceed by including returns on buy and hold strategies of lookback-straddles. A similar approach is followed by Agarwal and Naik (2000b), who further show that the descriptive model also has significant out-of-sample forecasting power. Both these papers increase our understanding of the kind of strategies that hedge funds are likely to follow. There is an important question, however, that current research has not yet addressed. If hedge fund returns are so highly nonlinear and strategies are very different from standard buy-and-hold or constant-fraction portfolio strategies as derived in for example Merton (1990), what preferences drive the investment managers of these funds? Even deeper, why are these funds there at all?

Some may argue that the observed patterns of hedge fund returns are merely a statistical artifact and that neither investors in hedge funds, nor hedge fund managers themselves, really know what type of pay-off pattern they generate. These arguments may be supported by the renowned secrecy surrounding hedge fund strategies and investment policies. In this chapter we refrain from yielding directly to the argument of irrational investors. Instead, we offer a framework of rational, loss averse investors that optimally choose payoff patterns that are remarkably similar to those presented in Figure 6.1.

Loss aversion originates from the area of behavioral finance. As a perspective for explaining observed financial markets’ or institutions’ behavior it is rapidly finding its way into the finance literature. Originating with the work of Kahneman and Tversky (1979), we have seen its application to studying the equity premium puzzle, see Benartzi and Thaler (1995); rationalizing the momentum effect, see Shefrin and Statman (1985); explaining the behavior of asset prices in equilibrium, in particular boom and bust patterns, see Barberis et al. (2001) and Berkelaar and Kouwenberg (2000a); and explaining

individual stock returns, see Barberis and Huang (2001). Building on the same framework and model specification for loss aversion as these earlier papers, we provide a rational explanation for observed hedge fund payoff patterns.

Besides evidence from behavioral finance there is another *ex ante* reason to suspect that loss aversion is an important phenomenon for hedge funds. This is the existence and structure of incentive fees. According to Edwards (1999), hedge funds pay managers large incentive fees as a fraction of the return achieved. Brown et al. (1999) mention common fees of 1% of funds under management and 20% of the profits, the same numbers as found by Liang (1999) in a large sample of hedge funds. Together with high incentive fees, however, investors usually require hedge fund managers to put a substantial amount of their own wealth in the fund. This requirement is obviously rooted in the preference of investors, who do not wish management to adopt a “recklessly risky” strategy. In conjunction with the use of high-water marks, managers can suffer a substantial personal loss if returns end up below a certain threshold level. As noted in Carpenter (2000), the most typical benchmark for hedge fund managers is a constant like the discounted value of current funds under management, or a benchmark representing a safe return, like a Treasury yield. We interpret the incentive scheme that is described above as (i) stimulating the manager to maximize expected fund wealth (or return) on the one hand, and (ii) making him loss averse to avoid moral hazard problems, i.e. excessive risk taking. The contribution of this chapter is that we combine these two effects into a model where the trade-off is between maximizing wealth and minimizing expected shortfall below a fixed target level. The present chapter can also be read as a generalization of a static version of the model in Chapter 2, as we include linear and non-linear investment products in the asset allocation decision.

The next section presents the model in its relation to other theoretical work in this area. From this model, we derive that the typical graphs of Figure 6.1 are exactly the four patterns that are optimal for a rational investor with mean-shortfall preferences. In Section 3 we highlight the model’s main implications. Section 4 concludes.

## 6.2 Model

In this section we introduce a simple model in which the hedge fund investment decision is represented through an investor with a loss averse objective function. We consider an investor with initial wealth  $W_0$  who optimizes expected utility defined over terminal wealth level  $W_1$ . We choose wealth

instead of return for ease of exposition. It is clear that there is an equivalent formulation in terms of possible return. The objective function we propose to model investor preferences is the mean-shortfall objective from Chapter 2, given by

$$\max \mathbb{E}[W_1] - \lambda \cdot \mathbb{E}[(W^B - W_1)^+], \quad (6.2)$$

$\lambda$  is the loss aversion parameter. The investor thus faces a trade-off between expected wealth on the one hand, and expected shortfall below the benchmark wealth level  $W^B$  on the other hand. If  $W^B = W_0$ , the investor weighs losses differently from gains. Instead of imposing a constraint on downside risk in the optimization problem, as in Basak and Shapiro (2001), we have incorporated the constraint in the objective function. This allows us to describe additional empirical features in hedge fund returns to the methodology followed by Basak and Shapiro. Barberis et al. (2001) use the expected shortfall measure in (6.2) as a risk measure to shed light on the behavior of firm-level stock returns in an asset-pricing framework. In their set up,  $W^B$  represents the *historical benchmark wealth level*, which may represent an average of recent portfolio wealth or the wealth at the end of a year. Berkelaar and Kouwenberg (2000b) use the more general specification of Kahneman and Tversky (1979) to solve a similar problem in continuous time.

Note that (6.2) is also relevant empirically. Sharpe (1998) explains how (6.2) is used by Morningstar to construct its ‘risk-adjusted rating’ for mutual funds. As these ratings in turn profoundly influence the flow of money to a mutual fund, see Guercio and Tkac (2001), (6.2) de facto reflects actual preferences of at least part of the investment industry. Benartzi and Thaler (1995) also use (6.2) as an approximation to the behavioral value function to explain the equity premium puzzle.

To model the investment opportunities available to a hedge fund, we assume that the investor can select 3 assets, namely a risk-free asset, a linear risky asset, and an option on the risky asset. We label the risky asset as stock in the rest of this chapter and normalize its initial price to 1. It should be kept in mind, however, that our results are not limited to stock investments. Alternative interpretations of the risky asset comprise stock indices, bonds or interest rates, and currencies. The stock has an uncertain pay-off  $u$  with distribution function  $G(u)$ . We assume that  $G(\cdot)$  is defined on  $(0, \infty)$ , is twice continuously differentiable and satisfies  $\mathbb{E}[u - r_f] > 0$ , i.e., there is a positive equity premium. The option is modeled as a European call option on the stock with strike price  $x$ . Its current price is denoted by  $c$ . To avoid making a particular choice for the option’s pricing model, we set the planning period equal to the option’s time to maturity. The option’s pay-off,  $R_c$ , is

now completely determined by the stock return as  $(u-x)^+$ . We assume there is a positive risk premium for the option as well, i.e.,  $\mathbb{E}[(u-x)^+/c] > r_f$ . Concentrating on hedge funds, we do not introduce any constraints on the positions the investor can take in any of these assets. We obtain

$$W_1 = W_0 r_f + X_0 \cdot (u - r_f) + X_1 \cdot (R_c - c \cdot r_f), \quad (6.3)$$

where  $r_f$  is the pay-off on the risk-free asset,  $X_0$  is the number of shares, and  $X_1$  the number of call options. The investor now maximizes (6.2) over  $\{X_0, X_1\}$ . Note that the evolution of wealth as described here compares with equation (2.1) of the model in Chapter 2. The only difference is the possibility of investing in a call option on the risky asset, and the fact that the present model is only a single-period model.

For ease of exposition we focus on the one-period model. The optimization problem introduced above is static. This may seem inappropriate for hedge funds, which are known to follow highly dynamic investment strategies. As argued in the introduction, however, there is ample empirical evidence that static models with non-linear instruments like options can explain a large part of the variation in hedge fund returns, both in-sample and out-of-sample, see Fung and Hsieh (1997) and Agarwal and Naik (2000b). Also, Chapter 2 of this thesis shows that the solution to the multi-period version of (6.2) gives decision rules at each period that have the same shape as the solution to the one-period model. However, another objection to the present set-up might be that we only include a single option. The main advantage of focusing on one option only, is that we are able to highlight the main features of the present model without introducing unnecessary complications. Even in this simple set-up, the model can describe several different pay-off patterns observed empirically. Moreover, Glosten and Jagannathan (1994) find that including more than one option in their contingent claims analysis does not give a better fit to their dataset of mutual fund returns. Similar results have been established by Agarwal and Naik (2000b), who show that for most hedge fund returns adding one option related factor in the Sharpe style regressions suffices to capture most of the non-linearity.

The following theorem gives our main result.

**Theorem 6.2.1** *The optimal investment strategy for a finite solution to problem (6.2) is given by one of the following:*

$$I: \quad X_0^* = 0, \text{ and } X_1^* = S_0/c, \quad (6.4)$$

$$II: \quad X_0^* = -S_0/p, \text{ and } X_1^* = -X_0^*, \quad (6.5)$$

$$III: \quad \begin{pmatrix} X_0 + X_1 \\ -X_0 \end{pmatrix} = \frac{1}{A} \cdot \begin{pmatrix} x - \bar{u}_1 \\ \bar{u}_2 - x \end{pmatrix} \cdot S_0 \cdot r_f, \quad (6.6)$$

where  $\bar{u}_1 < x < \bar{u}_2$ , and  $A > 0$ .  $S_0$  represents time 0 surplus, defined as  $W_0 - W^B/r_f$ . Through put-call parity,  $p$  is the price of a put option with strike price  $x$  as  $p = x/r_f + c - 1$ .

**Proof:** See appendix.

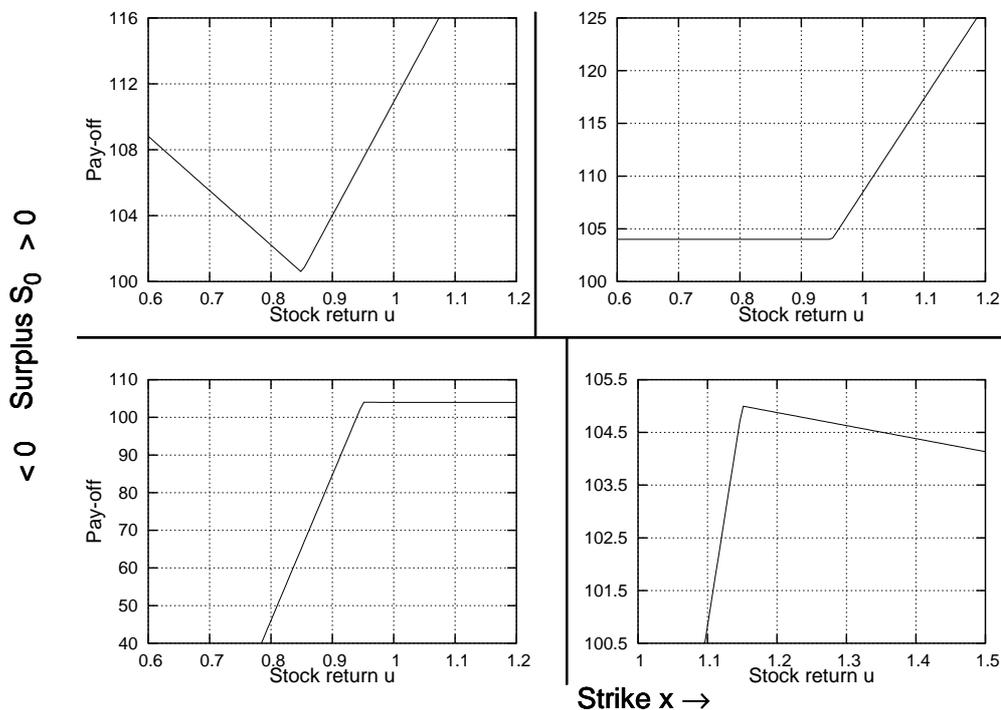
Theorem 6.2.1 states that if (6.2) has a finite solution, then the optimal investment strategy takes one out of three possible forms. A finite solution is ensured by a sufficiently high loss aversion parameter  $\lambda$  in (6.2). Unbounded solutions are less interesting in this setting, as they are not observed in practice.

Strategy I corresponds to a long position in the risk-free asset and the call option. The amount invested in the call option is exactly equal to the time 0 surplus. Strategy II corresponds to a short put position: an amount equal to the net shortfall is earned by selling puts. Strategy III is a condensed representation of either a long or short straddle<sup>1</sup> position. As  $A$ ,  $x - \bar{u}_1$ , and  $\bar{u}_2 - x$  are all assumed positive under III, the sign of  $X_0 + X_1$  and  $-X_0$  are completely determined by the sign of the surplus  $S_0$ . If the surplus is positive, we obtain a long straddle position. There is a short position in stocks,  $-X_0 > 0$ , which is offset by the long call position for sufficiently high stock prices,  $X_0 + X_1 > 0$ . Similarly, if the surplus is negative, we obtain a short straddle pay-off pattern. The appendix shows how  $\bar{u}_1$  and  $\bar{u}_2$  are derived from the model parameters and defines  $A$  as a function of  $\bar{u}_1$ ,  $\bar{u}_2$ , and  $r_f$  only.

Theorem 6.2.1 shows that the optimal pay-offs from model (6.2) are exactly those found by Fung and Hsieh (1997) in Figure 6.1. To make this point even stronger, we conduct a simple numerical experiment. Using a lognormal  $G(\cdot)$ , we compute the optimal solution to (6.2) for different strike prices and surplus levels. Computing the optimum is straightforward. It can either be done by discretizing  $G(\cdot)$ , or by solving the first order conditions. The latter are derived in the appendix in order to prove Theorem 6.2.1 and are very easy to solve numerically for any  $G(\cdot)$ . Figure 6.2 presents the results. For a positive surplus, we obtain a long straddle or a long call strategy, depending on whether the strike price is low or high, respectively. For negative surplus levels, the short put and short straddle are optimal for low and high strike prices, respectively. The similarity between Figure 6.2 and Figure 6.1 is striking: the present simple set-up provides a unified framework that can explain a large portion of all the different pay-off patterns observed empirically. Strike prices and surplus levels determine which of the pay-offs

---

<sup>1</sup>We use the term straddle to denote a portfolio of long put and call positions, where the number of puts and calls are not necessarily equal.



**Figure 6.2:** Characteristics of the optimal pay-offs as a function of the strike price  $x$  and the surplus  $W_0 - W^B/r_f$ .

The figure displays optimal pay-offs as a function of the risky return  $u$  for four different combinations of initial surplus  $S_0$  and strike price of the option. Though the precise form and steepness of these four pay-offs may vary if other combinations of  $S_0$  and strike are used, they are representative (in terms of positive/negative slope to the left/right of the strike) for the area in which they are plotted. These areas are bounded by the bold lines in the figure. The horizontal line separates positive from negative surplus. The two vertical lines separate ‘high’ from ‘low’ strike prices (for positive and negative surplus, respectively). The upper left panel has a strike of 0.85, the upper right and lower left panel of 0.95, the right panel of 1.15. The stock return is distributed  $\text{lognormal}(0.085, 0.16)$ , the call is priced using Black-Scholes.

is optimal in a particular setting.

It can be seen from Theorem 6.2.1 that higher absolute surplus levels  $|S_0|$  lead to more ‘aggressive’ investment policies, i.e., larger investments in the risky asset. For example, for increasingly large and positive values of the surplus, the number of long straddles or long puts increases as well, resulting in a steeper pay-off over the non-flat segments of the pay-off pattern. A similar result holds if the surplus becomes increasingly negative.

## 6.3 Implications for hedge funds

In Section 6.2 we found the optimal investment policies for a loss averse investor in a static setting. As mentioned, these results can be linked to pay-off patterns generated by dynamic investment strategies following the ideas of Fung and Hsieh (1997). Their paper, however, is mainly empirical. In this section we provide additional arguments linking the results of Theorem 6.2.1 and Figure 6.2 to dynamic investment strategies that are actually used by hedge funds.

The first dynamic strategy is that of a market timer as described by Merton (1981). Merton shows that the return pattern of such a strategy resembles a straddle on the traded asset. This idea is pursued further by Fung and Hsieh (2001), who find that using the return to a synthetic lookback straddle captures most of the variation in returns of Trend Following hedge funds. The upper-left panel in Figure 6.2 therefore clearly corresponds to at least one dynamic investment strategy.

Another popular dynamic investment strategy is portfolio insurance, see for example Leland (1980). The basic idea is to reduce the proportion of stock if prices fall, and to increase it if prices rise. This mimics a delta hedge strategy of a call option, which is precisely the pay-off pattern given in the upper-right panel of Figure 6.2. Portfolio insurance strategies are especially attractive for financial institutions facing short-term restrictions on their asset value, like certain pension funds, as argued by Shefrin and Statman (1985), Benninga and Blume (1985), Brennan and Solanki (1981).

Convergence bets as a dynamic investment strategy typically generate a pay-off pattern resembling a short put position, see the lower-left panel in Figure 6.2. A good example of this type of strategy is merger arbitrage as documented by Mitchell and Pulvino (2001). By taking a long position in the stock of the target in a merger or takeover, and a corresponding short position in the stock of the acquirer in case payment is in stock rather than cash, Mitchell and Pulvino (2001) show that positive returns are possible in bull markets. These returns are largely uncorrelated with the market, i.e., they have a beta equal to zero. In bear markets, mergers are more likely to fail, such that the merger arbitrage strategy results in potentially large losses there. These losses usually correlate positively with the market. The correlation pattern documented by Mitchell and Pulvino is precisely that of a short put. It is not surprising, therefore, that including a put-option return in the Sharpe style regressions for merger arbitrage returns significantly increases the explanatory power. Mitchell and Pulvino also show that hedge funds specialized in merger arbitrage generate pay-offs very similar to short

puts, albeit that they have a slightly positive beta in bull markets. Different interpretations of the short put pattern are also possible, for example, convergence bets in credit markets. In that case, the short put may be seen as a direct reflection of the credit spread, see Merton (1974). By taking off-setting positions in corporate and government bonds, or in bonds of governments with different credit ratings, one can lock in the spread in most cases. In case of default (of the long position), however, the strategy results in a large loss. Strategies of this type are known to have been implemented by, for example, LTCM. Jorion (2000) notes on the strategies followed by LTCM, that “Another view is that these strategies are actually designed to take a big loss once in a while ,like a short position in an option.”

The short straddle position in the lower-right panel is more difficult to link up with well-known dynamic investment strategies, and we will conjecture on some of the possible explanations for this further below. Thus far, we have established a link between the optimal pay-off patterns emerging from our model and the pay-offs on dynamic investment strategies that are documented, either empirically or theoretically. In the remainder of this section, we further explore the validity and implications of our model for hedge funds.

It is clear from Figure 6.2 that the shape of the optimal pay-off crucially depends on two variables, namely the (sign of the) surplus and the location of the strike price. We start with discussing the former. Figure 6.2 clearly shows that there is a remarkable difference between strategies for a positive and negative surplus, respectively. The interpretation of the surplus in the context of hedge funds is not straightforward and crucially depends on the interpretation of the benchmark  $W_B$ . As discussed in the introduction and documented in Brown et al. (1999) and Brown et al. (1997), most hedge fund managers get their bonuses if the fund earns a return above high-water marks. The mark can be given by a treasury bill return, a fixed return, or a return based on average market performance. If, for example, the high-water mark is fixed or is linked to a treasury yield, a sufficiently high intermediate return over part of the measurement period brings the fund into a situation with a positive ‘surplus’. So with respect to the managers of the fund, we can interpret the surplus in terms of the return that is necessary to attain the high-water mark. Analogously, a negative intermediate return may jeopardize the manager’s future fee and bring him in a situation of a negative surplus. Alternatively, Barberis et al. (2001) and Barberis and Huang (2001) use the idea of a surplus together with mental accounting practices adopted by loss averse investors. In their set-up, surplus represents the difference between the current price of a stock or fund and its historical benchmark. The historical benchmark may represent an average of recent stock prices,

or some specific historical stock price, such as the price at the end of the year. The difference between the realized stock price and the benchmark, if positive, is the investor's personal measure of how much 'he is up' on his investment and conversely, if negative, how much 'he is down'. If this way of mental accounting is relevant for modeling investor preferences, then our results suggest that the behavior of hedge funds is in line with the preferences of those who invest in it. This fits in with the incentive schemes as discussed before, such that these schemes can be seen as the proper instrument for attaining alignment.

One of the results also mentioned in the previous section was that larger absolute values of the surplus  $|S_0|$  result in steeper pay-off patterns. In particular, if the surplus becomes increasingly negative, our model predicts that a more aggressive investment strategy is adopted. Intuition for this can be found in Brown et al. (1997). They discuss the consequences of high-water mark thresholds used by hedge fund managers and note: 'If a fund has a negative return, the manager is out of the money and presumably has an incentive to increase risk'. Theorem 6.2.1 formalizes this intuition.

Given the above interpretation of surplus, our model predicts that funds with high water marks are more likely to follow convergence strategies. Vice versa, if our proposed rational framework has empirical content, we would expect higher high-water marks for firms focusing on convergence bets. Some anecdotal empirical evidence supporting our claim is available for LTCM. Jorion (2000) explains that the core strategy of LTCM was a relative-value or convergence-arbitrage trade on credit spreads, while Edwards (1999) states that LTCM had one of the highest incentive fees in the industry.

The second important variable driving influencing the shape of the optimal pay-off pattern is the strike price  $x$  of the option. The strike price links to the dynamic strategy followed by the fund. In our framework, a fund focusing on a particular strategy is tantamount to the fund picking its strike price for the option. Again, our model gives rise to several implications that are corroborated by the empirical literature. For example, consider the lower-left panel in Figure 6.2. As mentioned earlier, the short put pay-off pattern corresponds to a trading strategy focusing on convergence bets. Following our model, such a strategy is only optimal if the strike is sufficiently low, i.e., if the put is sufficiently far out of the money. This is supported by the findings of Mitchell and Pulvino (2001). As mentioned earlier, they capture much of the variation in merger arbitrage returns by including a put option. It is remarkable to note that the optimal strike price of their option is at a return level of -4%. This resembles an out of the money put along the lines predicted by our present simple framework. Another example is given by the upper-left panel in Figure 6.2. Originally suggested by Merton

(1981), Fung and Hsieh (2001) use the long straddle pay-off to describe the returns on market timers, i.e. those traders that buy when they believe the market goes up and sell short when they believe it going down. Our model implies that strike prices must not be too high in order for the long straddle to be optimal. Fung and Hsieh use at-the-money straddles in their empirical work. Our results suggest that out-of-the-money straddles would provide an even better fit in their regression model.

The above interpretation of the strike price as the chosen dynamic strategy of the fund raises the question what type of strategy will be chosen by investors if they can freely choose among different hedge fund styles. Extending the model (6.2) to optimize over  $\{X_0, X_1\}$  and  $x$  simultaneously leads to the optimality of the lower-left and upper-right panels in Figure 6.2, depending on whether the surplus is negative or positive, respectively. This has two implications. First, it may in part explain why the lower-right panel does not match any of the well-known dynamic investment styles used in the industry. It is simply not optimal to follow this strategy if the strategy choice is at the discretion of the hedge fund manager or the investor. The argument is strengthened by noting that the lower two panels describe a situation with a negative surplus. In such a setting, an efficient choice of the asset mix including the strike  $x$  is even more important than in a situation of a positive surplus. This may explain why we can observe the long straddle strategy (trend followers) empirically, while this is harder for the short straddle. Our second implication of the optimal choice of the strike  $x$ , however, is that the (long) straddle type pay-off patterns (upper-left panel in Figure 6.2) may not survive in the long run, either. As investors become more aware of the funds' properties and potential alternatives, our results suggests that they will shift out of trend following funds into long call strategies.

## 6.4 Conclusions

In this chapter we discussed a very simple financial optimization model based on loss aversion for which exactly four different pay-off patterns can be optimal. These patterns closely match the various patterns in hedge fund returns observed in empirical work. We provided several reasons to suspect that loss aversion is an important phenomenon in hedge funds. The most explicit piece of evidence given in the literature relates to incentive schemes for hedge fund managers: the limited liability for managers is reduced by requiring them to put a substantial amount of their own money at stake. In that sense, loss aversion is a result of the common 'put your money where your mouth is'-policy that hedge funds use to attract potential investors and signal their

commitment.

Our model provides a unified, rational framework for explaining patterns in hedge fund returns. Using traditional utility functions that do not have the loss aversion property, capturing the wide variety in hedge fund return patterns into one unifying framework is much more difficult. With respect to robustness, numerical computations reveal very similar patterns for alternative specifications for the loss averse objective function. In particular, using the original specification of Kahneman and Tversky (1979) our findings do not change significantly.

Interestingly, our model also gives rise to some new predictions pertaining to the relation between the incentive schemes and adopted dynamic strategies. Partial evidence supporting some of the predictions is available from empirical work in the literature. More detailed data sets, however, are needed to substantiate the empirical validity of some other implications of our theoretical model. Of course, in practice hedge fund managers may adopt a number of trading strategies and/or change strategies over time. It is clear that such complications fall outside the scope of our current simple modeling framework. We believe, however, that the model and its results provide a useful step in a further understanding of hedge funds.

# Appendix

## 6.A Proofs

### Proof of Theorem 6.2.1

We start by restating the optimization problem in (6.2) as

$$\max_{X_0, X_1} V(X_0, X_1), \quad (6A.1)$$

$$\text{with } V(X_0, X_1) = \mathbb{E}[W_1] - \lambda \cdot \mathbb{E}[(W^B - W_1)^+], \quad (6A.2)$$

subject to

$$W_1 = W_0 r_f + X_0 \cdot (u - r_f) + X_1 \cdot (R_c - c \cdot r_f). \quad (6A.3)$$

Define  $p = x/r_f + c - 1$ , the price of the put option corresponding to price of the call following from put-call parity, see e.g. Hull (1997). Define  $R_{c,x,G(\cdot)}$  as the expected return on the call option with strike  $x$  on an asset with return  $u \sim G(\cdot)$ , given by  $\mathbb{E}[(u-x)^+/c]$ . For ease of notation, we drop the subscripts  $x$  and  $G(\cdot)$ . Likewise, we denote the expected return on the put option with  $R_p$ . To ensure a finite optimal solution we need the following assumptions.

A:  $\lambda r_f G(x) > R_c - r_f$ ,

B:  $\lambda \int_0^x (x-u)^+/p - r_f dG > -(R_p - r_f)$ ,

C:  $R_c$  is increasing in  $x$ .

The motivation for these two assumptions will follow from the proofs below. In short, assumptions A and B put a lower bound on the loss aversion parameter  $\lambda$  to ensure that the trade-off between risk and return leads to a finite solution. Assumption C is tested empirically in Coval and Shumway (2001), who find that the expected return of S&P index option returns increases with the strike price.

There are four possible pay-off patterns resulting from the combination of a risk-free asset, a stock, and a call option on the stock, namely decreasing-decreasing(I), increasing-increasing(II), decreasing-increasing(III), increasing-decreasing(IV), where for example case (I) refers to a setting where the pay-off increases in  $u$  both before and after the strike price  $x$ .

**pattern I (decreasing-decreasing)**

Conditions for case I are  $X_0 \leq 0$  and  $X_0 + X_1 \leq 0$ . The first order conditions in this case are given by

$$\frac{\partial V}{\partial X_0} = 0 \Rightarrow 0 = \mathbb{E}[u - r_f] + \lambda \int_{\bar{u}}^{\infty} u - r_f dG, \quad (6A.4)$$

and

$$\frac{\partial V}{\partial X_1} = 0 \Rightarrow 0 = \mathbb{E}[(u - x)^+ - c \cdot r_f] + \lambda \int_{\bar{u}}^{\infty} (u - x)^+ - c \cdot r_f dG, \quad (6A.5)$$

where  $\bar{u}$  is a constant depending on  $(X_0, X_1)$ . Using  $\mathbb{E}[u] > r_f$  and  $\mathbb{E}[(u - x)^+] > c \cdot r_f$ , we obtain that the right-hand sides of both (6A.4) and (6A.5) are positive for any value of  $\bar{u}$ , such that there is no interior optimum. The solution in situation I is, therefore, to set  $X_0^* = X_1^* = 0$ .

**pattern II (increasing, increasing)**

Conditions for case II are  $X_0 > 0$  and  $X_0 + X_1 > 0$ . First order conditions are given by the system

$$\begin{cases} \mathbb{E}[u - r_f] + \lambda \int_0^{\bar{u}} u - r_f dG = 0, \\ \mathbb{E}[(u - x)^+ - c \cdot r_f] + \lambda \int_0^{\bar{u}} (u - x)^+ - c \cdot r_f dG = 0, \end{cases} \quad (6A.6)$$

where  $\bar{u}$  is again a constant depending on  $(X_0, X_1)$ . Each equation in (6A.6) has either zero or two solutions. The zero-solution case for the first order condition corresponds to unbounded solutions for the original optimization problem (6A.1), since the left-hand side in (6A.6) must then necessarily be positive. We have abstracted from unbounded solutions however. Since the integrands in the two equations of (6A.6) are different, the two equations will not be satisfied for the same value of  $\bar{u}$ . Hence, the optimum is attained at the extremals. In this case, for II the extremals are defined by two sets of parameter values, given by

$$X_0 = 0, \quad X_1 > 0, \quad (6A.7)$$

or

$$X_0 > 0, \quad X_0 + X_1 = 0. \quad (6A.8)$$

Starting with the former, investing only in the call option implies an optimization problem with the following first order condition for an interior optimum:

$$\mathbb{E}[(u - x)^+ - c \cdot r_f] + \lambda \int_0^{\bar{u}} (u - x)^+ - c \cdot r_f dG = 0, \quad (6A.9)$$

where  $\bar{u} = (W^B - W_0 r_f)/X_1 + x + c \cdot r_f$ . By definition,  $\bar{u} \geq x$ . Under assumption A, we find that the FOC is never fulfilled, i.e. the derivative with respect to  $X_1$  is negative. Without an interior optimum, the optimal solution is given by

$$X_1^* = \begin{cases} (W_0 - W^B/r_f)/c & \text{if } W_0 > W^B/r_f, \\ 0 & \text{if } W_0 \leq W^B/r_f. \end{cases} \quad (6A.10)$$

We call this the long call strategy.

Now for the second case of extremals in situation I:

Define  $X_2 = X_0 + X_1$ , and  $p = (x + c \cdot r_f - r_f)/r_f$ . Condition for an interior optimum is

$$\mathbb{E}[(x - u)^+ - p \cdot r_f] + \lambda \int_0^{\bar{u}} (x - u)^+ - p \cdot r_f dG = 0, \quad (6A.11)$$

where  $\bar{u} = (W_0 r_f - W^B)/X_2 + x - p \cdot r_f$ . By definition,  $\bar{u} \leq x$ . Under assumption B, the FOC has no solution, i.e. the derivative with respect to  $X_2$  is positive.

Without an interior optimum, the optimal solution is given by

$$X_2^* = \begin{cases} (W^B/r_f - W_0)/p & \text{if } W_0 < W^B/r_f, \\ 0 & \text{if } W_0 \geq W^B/r_f. \end{cases} \quad (6A.12)$$

We call this the short put strategy.

For  $W_0 \geq W^B/r_f$  the long call strategy has a higher objective value than the short put. This is seen from the objective values, which are  $W_0 r_f + (W_0 - W^B/r_f)(R_c - r_f)$  for the long call versus the short put value of  $W_0 r_f$ .

For  $W_0 < W^B/r_f$  the short put strategy has higher objective value than the long call. This is seen from the objective values, which are  $W_0 r_f + \lambda \cdot (W_0 r_f - W^B)$  for the long call versus the short put value that is larger than  $W_0 r_f + \lambda \cdot (W_0 r_f - W^B) \cdot G(x)$ . The last inequality follows from Assumption B.

**pattern III (decreasing-increasing: straddle)**

Situation III is characterized by  $X_0 < 0$ ,  $X_0 + X_1 > 0$ .

The two values for which  $W_1 = W^B$  are given by  $\bar{u}_1 < x$  and  $\bar{u}_2 > x$ . They are defined as

$$\bar{u}_1 = \frac{W^B - W_0 \cdot r_f - X_0 \cdot r_f - X_1 \cdot c \cdot r_f}{X_0} \quad (6A.13)$$

$$\bar{u}_2 = \frac{W^B - W_0 \cdot r_f - X_0 \cdot r_f - X_1 \cdot x - X_1 \cdot c \cdot r_f}{X_0 + X_1} \quad (6A.14)$$

The first order conditions are given by

$$\frac{\partial V}{\partial X_0} = \mathbb{E}[u - r_f] + \lambda \cdot \int_{\bar{u}_1}^{\bar{u}_2} (u - r_f) dG = 0, \quad (6A.15)$$

$$\frac{\partial V}{\partial X_1} = \mathbb{E}[(u - x)^+ - c \cdot r_f] + \lambda \cdot \int_{\bar{u}_1}^{\bar{u}_2} ((u - x)^+ - c \cdot r_f) dG = 0 \quad (6A.16)$$

$$\text{s.t.} \quad \bar{u}_1 < x < \bar{u}_2. \quad (6A.17)$$

It can be checked that under the current assumptions the Hessian is negative definite. If the FOC is fulfilled for a feasible  $(X_0, X_1)$ , it constitutes a local optimum. Note that if the FOC is satisfied, the value of the objective function can be written as

$$W_0 r_f + \lambda \cdot r_f \cdot (W_0 - W^B / r_f) \cdot (G(\bar{u}_2^*) - G(\bar{u}_1^*)) \leq W_0 r_f + \lambda \cdot S_0. \quad (6A.18)$$

The function value of the optimum in situation II for positive surplus is given by the value of the long call strategy as

$$W_0 r_f + (W_0 - W^B / r_f) \cdot \mathbb{E}[(u - x)^+ / c - r_f] = W_0 r_f + S_0 \cdot [R_c - r_f]. \quad (6A.19)$$

Using assumption C, which says that  $R_c - r_f$  is increasing in  $x$ , we find that for  $x \rightarrow 0$  and  $S_0 > 0$  the straddle pay-off is better than the long call pay-off of case II.

**pattern IV (increasing, decreasing: short straddle)**

Situation IV is similar to case III and characterized by  $X_0 > 0$ ,  $X_0 + X_1 < 0$ .  $\bar{u}_1$  and  $\bar{u}_2$  are the same as in situation III.

The first order conditions are given by

$$\frac{\partial V}{\partial X_0} = \mathbb{E}[u - r_f] + \lambda \cdot \int_0^{\bar{u}_1} (u - r_f) dG + \lambda \cdot \int_{\bar{u}_2}^{\infty} (u - r_f) dG, \quad (6A.20)$$

$$\begin{aligned} \frac{\partial V}{\partial X_1} &= \mathbb{E}[(u - x)^+ - c \cdot r_f] \\ &+ \lambda \cdot \int_0^{\bar{u}_1} ((u - x)^+ - c \cdot r_f) dG + \lambda \cdot \int_{\bar{u}_2}^{\infty} ((u - x)^+ - c \cdot r_f) dG \end{aligned} \quad (6A.21)$$

For the moment, it is left as an exercise to the reader to verify that the Hessian is negative definite. If the FOC is fulfilled for a feasible  $(X_0, X_1)$ , it constitutes a local optimum. Note that if the FOC is satisfied, the value of the objective function can be written as

$$W_0 r_f + \lambda \cdot (W_0 r_f - W^B) \cdot (1 - (G(\bar{u}_2^*) - G(\bar{u}_1^*))). \quad (6A.22)$$

The function value of the optimum in situation II for negative surplus is given by the short put strategy as

$$W_0 r_f + X_2 \cdot p \cdot \left( \mathbb{E}[R_p - r_f] + \lambda \cdot \int_0^x (x - u)^+ - p \cdot r_f dG \right) + \lambda \cdot (W_0 r_f - W^B) \cdot G(x), \quad (6A.23)$$

which is, according to Assumption B, larger than

$$W_0 r_f + \lambda \cdot (W_0 r_f - W^B) \cdot G(x). \quad (6A.24)$$

This implies that there is a strike  $y$  such that for strikes  $x < y$ , the short put strategy of pattern II has a higher objective value than the current short-straddle pattern IV.

Having found the optimal pay-offs for each pair  $(x, S_0)$ , we end with defining  $A$  in Theorem 6.2.1. From the definition of  $\bar{u}_1$  and  $\bar{u}_2$  in (6A.13) and (6A.14), we can write  $X_0$  and  $X_1$  as a function of  $\bar{u}_1$  and  $\bar{u}_2$  in the following way:

$$\begin{pmatrix} X_0 + X_1 \\ -X_0 \end{pmatrix} = \frac{1}{A} \cdot \begin{pmatrix} x - \bar{u}_1 \\ \bar{u}_2 - x \end{pmatrix} \cdot S \cdot r_f, \quad (6A.25)$$

where  $A = -(\bar{u}_2 - x)(x - \bar{u}_1) + cr_f(x - \bar{u}_1) + (\bar{u}_2 - x)pr_f$ , and  $pr_f = x + cr_f - r_f$ . As the long straddle pay-off is only optimal for positive surplus, we find  $A > 0$ . This concludes the proof.  $\blacksquare$



---

## Conclusion

---

In the introduction we gave a definition of loss aversion from Gleitman et al. (2000). However, we did not quote the full citation, which reads

**“Loss Aversion:** A widespread pattern, evident in many aspects of decision making, in which people seem particularly sensitive to losses and eager to avoid them. In many cases, this manifests itself as an increased willingness to take risks in hopes of reducing the loss.”

The second sentence of the quoted definition is the kind of behavior that we have formally derived in Chapter 2 to be optimal under the bilinear specification of loss averse preferences. To characterize the investment behavior, we introduced the notion of a surplus as the difference between initial wealth and the risk-free discounted level of the benchmark wealth. For a negative surplus, the investment in the risky asset is increasing when the surplus becomes even more negative. For positive surpluses, the risky investment increases with the surplus. Hence, the amount of risky investments as a function of initial wealth takes a V-shaped form. The V-shaped patterns also represent the solution to the multi-stage problem, where the slopes of the V’s are decreasing in (i) the degree of loss aversion, and (ii) the time to the horizon. The first effect is fairly trivial. However, the fraction invested in the risky asset does not go to zero for infinite loss aversion. We concluded that investing in the risky asset decreases the expected shortfall, regardless of the (linear) degree of loss aversion. The second effect has specific implications in the area of time-diversification. It confirms the popular idea that

stocks are more attractive than bonds when the planning period is longer and preferences are loss averse.

Chapter 3 showed that the characteristic asset allocation rules of Chapter 2 are robust to changes in the specification of loss aversion. We considered two alternatives to the linear penalty on losses, a convex and a concave one. The first was the Kahneman-Tversky value function, which plays a central role in the behavioral finance literature. The second was an objective with a quadratic penalty on shortfall. This specification is used in applied Asset/Liability Management studies. Proponents argue that it punishes large losses more severely, thus ensuring a safer investment policy. We saw that this is indeed the case for most (though not all) surplus values, although the V-shaped investment rules are still optimal. Moreover, despite the concavity in losses, the fraction in the risky asset does not go to zero if the loss aversion parameter goes to infinity. The results for both alternate specifications of loss aversion suggest that in many practical and empirical situations, it suffices to use the bilinear formulation of loss aversion.

A final consideration concerned the inclusion of restrictions on the composition of the asset mix, which have a dampening effect on the V-shaped rules. This effect becomes larger for models with a longer horizon.

The results in Chapters 2 and 3 are promising: for all loss averse settings we considered, the typical asset allocation rules have a persistent shape. In the rest of the thesis, we set out to put the mean-shortfall model into perspective. The first application to consider was in the area of Asset/Liability Management for pension funds. We saw that the parameters of the mean-shortfall model have a natural interpretation in terms of pension funding. This is not surprising, since the types of models that are typically used in ALM were the initial source of inspiration to consider the mean-shortfall model and its solution. We considered extensions to the framework of the mean-shortfall optimization, making the benchmark level stochastic. This increased insight in how major economic variables impact the optimal investment strategies for pension funds under loss averse preferences. We did not find conclusive empirical evidence of pension funds being actually loss averters. However, using data on funding ratios and stock allocations, we were able to point out specific patterns that cannot be explained by ‘standard’ strategies alone. We suggested that loss aversion may provide a possible explanation for part of the empirical puzzle.

A natural area of application for optimal financial decision making is the area of household consumption and investment. In Chapter 5 we interpreted the mean-shortfall objective as representing the per-period utility of consumption. We derived optimal investment decision rules, and analyzed them in the context of precautionary savings. Using the parametric solution,

and expected shortfall as the natural risk measure, it was straightforward to derive a relation between uncertainty and savings/investment. A very natural extension was the inclusion of habit formation. As the model has an explicit reference point, habit formation was a natural mechanism to consider, transforming the fixed benchmark level of the earlier chapters into an endogenously determined reference point.

A final application of the mean-shortfall model was in Chapter 6. It involved assumptions on the incentives of hedge fund managers, as well as the investors. We modeled possible dynamic strategies of hedge funds by extending the framework to include a call-option on the risky asset. The optimal pay-off patterns that emerged as part of the optimal solution matched empirical patterns of pay-offs from hedge fund return factors. Until now, no other explanation in terms of incentives or preferences was given to explain the existence of typical nonlinear pay-off patterns for hedge funds.

Concluding, this thesis can be seen as giving an in-depth view on the effects of loss averse preferences on financial decision making. The mean-shortfall model was solved and interpreted, giving both analytical and numerical results on optimal risk taking. We put the mean-shortfall model in the perspective by considering three different economic settings of financial decision making. In each setting there are numerous interesting directions for future research, one of the most interesting being under what specific conditions the found V-shaped behavior occurs empirically.



# BIBLIOGRAPHY

- Agarwal, V. and N. Y. Naik (2000a, Spring). On taking the "alternative" route: The risks, rewards, and performance persistence of hedge funds. *Journal of Alternative Investments* 2(4), 6–23.
- Agarwal, V. and N. Y. Naik (2000b, July). Performance evaluation of hedge funds with option-based and buy-and-hold strategies. London Business School, Working paper No. 300, July 2000.
- Aizenman, J. (1998). Buffer stocks and precautionary savings with loss aversion. *Journal of International Money and Finance* 17, 931–947.
- Amin, G. S. and H. M. Kat (2001). Hege fund performance 1990-2000. do the 'money machines' really add value? Working paper, ISMA Centre, University of Reading.
- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9, 203–228.
- Barberis, N. and M. Huang (2001). Mental accounting, loss aversion, and individual stock returns. *The Journal of Finance* 56(4), 1247–1292.
- Barberis, N., M. Huang, and T. Santos (2001). Prospect theory and asset prices. *Quarterly Journal of Economics* 66(1), 1 – 53.
- Basak, S. and A. Shapiro (2001). Value-at-risk based risk management: Optimal policies and asset prices. *Review of Financial Studies* 14(2), 371–405.
- Benartzi, S. and R. H. Thaler (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics* 110, 73–92.
- Benninga, S. and M. Blume (1985). On the optimality of portfolio insurance. *Journal of Finance* 40(5), 1341–1352.

- Berkelaar, A. and R. Kouwenberg (2000a). From boom til bust: How loss aversion affects asset prices. Econometric Institute Report EI-2000-21/A, Erasmus University Rotterdam.
- Berkelaar, A. and R. Kouwenberg (2000b). Optimal portfolio choice under loss aversion. Econometric Institute Report EI 2000-08/A.
- Birge, J. R. and F. Louveaux (1997). *Introduction to Stochastic Programming*. Springer-Verlag New York, Inc.
- Boender, C. and M. Vos (2000). Risk-return budgeting at pension plans. In R. Layard-Lieschling (Ed.), *Risk Budgeting: A Cutting-Edge Guide to Enhancing Fund Management*, pp. 80–88. The Institutional Investor Inc., New York.
- Boender, C. G. E. (1997). A hybrid simulation/optimisation scenario model for asset/liability management. *European Journal of Operational Research* 99, 126–135.
- Bovenberg, A. (2001). Hoe houden we de pensioenpolder droog? uitdagingen voor pensioenfondsen op de drempel van de 21e eeuw. OCFEB Research Memorandum 0106.
- Bowman, D., D. Minehart, and M. Rabin (1999). Loss aversion in a consumption-savings model. *Journal of Economic Behavior & Organization* 38, 155–178.
- Brealey, R. A. and E. Kaplanis (2001). Changes in the factor exposures of hedge funds. London Business School, IFA Working Paper 320, January 2001.
- Brennan, M. and R. Solanki (1981). Optimal portfolio insurance. *Journal of Financial and Quantitative Analysis* 16, 279–300.
- Brouwer, F. (1997). *Applications of the Mean-Downside Risk Investment Model*. Ph. D. thesis, Vrije Universiteit Amsterdam.
- Brown, S. J. and W. N. Goetzmann (1997). Mutual fund styles. *Journal of Financial Economics* 43, 373–399.
- Brown, S. J., W. N. Goetzmann, and R. G. Ibbotson (1999). Offshore hedge funds: Survival and performance, 1989-95. *Journal of Business* 72(1), 91–117.
- Brown, S. J., W. N. Goetzmann, and J. Park (1997). Conditions for survival: Changing risk and the performance of hedge fund managers and CTAs. Yale University, unpublished.
- Browning, M. and T. F. Crossley (2001). The life-cycle model of consumption and saving. *Journal of Economic Perspectives* 15(3), 3–22.

- Browning, M. and A. Lusardi (1996, December). Household saving: Micro theories and micro facts. *Journal of Economic Literature* 34, 1797 – 1855.
- Caballero, R. J. (1990). Consumption puzzles and precautionary savings. *Journal of monetary economics* 25, 113–136.
- Campbell, R. (2001). *Rethinking Risk in International Financial Markets*. Ph. D. thesis, Erasmus University Rotterdam.
- Cariño, D., T. Kent, D. Myers, C. Stacy, M. Sylvanus, A. Turner, K. Watanabe, and W. Ziemba (1994, January-February). The russell-yasuda kasai model: An asset/liability model for a japanese insurance company using multistage stochastic programming. *Interfaces* 24, 29–49.
- Carpenter, J. N. (2000, October). Does option compensation increase managerial risk appetite? *The journal of finance* 55(5), 2311–2331.
- Carroll, C. D. (1998). How important is precautionary saving? *Review of Economics and Statistics* 80(3), 410–419.
- Cass, D. and J. Stiglitz (1970, June). The structure of investor preferences and asset returns, and separability in portfolio allocation: A contribution to the pure theory of mutual funds. *Journal of Economic Theory* 2, 122–160.
- Chaveau, T. and R. Loufir (1997). The future of public pensions in the seven major economies. In D. Broer and J. Lassila (Eds.), *Pension Policies and Public Debt in Dynamic CGE Models*. Heidelberg: Phycia-Verlag.
- Consigli, G. and M. A. Dempster (1998). The CALM stochastic programming model for dynamic asset-liability management. *Annals of Operations Research*.
- Constantinides, G. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy* 98, 519–543.
- Coval, J. D. and T. Shumway (2001, June). Expected option returns. *The journal of Finance* 54(3), 983–1009.
- De Roon, F. A., T. E. Nijman, and J. R. Ter Horst (2000, March). Evaluating style analysis. Erasmus University Rotterdam, ERIM Report Series ERS-2000-11-F&A.
- Dert, C. L. (1995). *Asset Liability Management for Pension Funds: A multistage chance constrained programming approach*. Ph. D. thesis, Erasmus University Rotterdam.

- Dert, C. L. (1998). A dynamic model for asset liability management for defined benefit pension funds. In W. T. Ziemba and J. M. Mulvey (Eds.), *Worldwide Asset Liability Modeling*. Cambridge University Press.
- Duesenberry, J. S. (1949). *Income, Saving and the Theory of Consumer Behavior*. Harvard University Press.
- Dupačová, J., M. Bertocchi, and V. Moriggia (1998). Postoptimality for scenario based financial planning models with an application to bond portfolio management. In W. T. Ziemba and J. M. Mulvey (Eds.), *Worldwide Asset Liability Modeling*, Chapter 12, pp. 263–285. Cambridge University Press.
- Edwards, F. R. (1999, Spring). Hedge funds and the collapse of long-term capital management. *Journal of Economic Perspectives* 13(2), 189–210.
- Edwards, F. R. and M. O. Caglayan (2001). Hedge fund performance and manager skill. *The journal of futures markets* 21(11), 1003–1028.
- Fung, W. and D. A. Hsieh (1997). Empirical characteristics of dynamic trading strategies: The case of hedge funds. *The Review of Financial Studies* 10(2), 275–302.
- Fung, W. and D. A. Hsieh (2001, Summer). The risk in hedge fund strategies: Theory and evidence from trend followers. *The review of financial studies* 14(2), 313–341.
- Gleitman, H., A. J. Fridlund, and D. Reisberg (2000). *Basic Psychology* (5th ed.). W.W. Norton & Co.
- Glosten, L. and R. Jagannathan (1994). A contingent claim approach to performance evaluation. *Journal of Empirical Finance* 1, 133–160.
- Guercio, D. D. and P. A. Tkac (2001). Star power: The effect of morningstar ratings on mutual fund flows. Federal Reserve Bank of Atlanta, Working Paper 2001-15, August 2001.
- Harlow, W. (1991, September-October). Asset allocation in a downside-risk framework. *Financial Analysts Journal*, 28–40.
- Hiller, R. S. and J. Eckstein (1993, November). Stochastic dedication: Designing fixed income portfolios using massively parallel benders decomposition. *Management Science* 39(11), 1422–1438.
- Hirshleifer, D. (2001). Investor psychology and asset pricing. *The Journal of Finance* 56(4), 1533–1597.

- Huisman, R., K. G. Koedijk, and R. A. Pownall (1999). Asset allocation in a value-at-risk framework. Rotterdam Institute of Financial Management, working paper no. 99-3.
- Hull, J. C. (1997). *Options, Futures, and Other Derivatives*. Prentice-Hall Inc.
- Ingersoll, J. E. (1987). *Theory of Financial Decision Making*. Rowman & Littlefield Publishers, Inc.
- Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk* (2nd ed.). McGraw-Hill publishing company.
- Kahneman, D., P. Slovic, and A. Tversky (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge University Press.
- Kahneman, D. and A. Tversky (1979, March). Prospect theory: An analysis of decision under risk. *Econometrica* 47, 263–291.
- Kritzman, M. and D. Rich (1998, Summer). Beware of dogma: The truth about time diversification. *The Journal of Portfolio Management*, 66–77.
- Kusy, M. I. and W. Ziemba (1986). A bank asset and liability management model. *Operations Research* 34(3), 356–376.
- Leibowitz, M. L., S. Kogelman, and L. N. Bader (1994). Funding ratio return. *Journal of Portfolio Management* 21(1), 39–47.
- Leibowitz, M. L., S. Kogeman, and L. N. Bader (1992, Winter). Asset performance and surplus control: A dual-shortfall approach. *Journal of Portfolio Management*, 28–37.
- Leland (1980). Who should buy portfolio insurance? *The Journal of Finance* 35(2), 581–594.
- Liang, B. (1999). On the performance of hedge funds. Weatherhead school of management, Case Western Reserve university, working paper March 1999.
- Maranas, C. D., I. P. Androulakis, C. A. Floudas, A. J. Berger, and J. M. Mulvey (1997). Solving long-term financial planning problems via global optimization. *Journal of Economics dynamics and control* 21, 1405–1425.
- Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance* 7(1), 77–91.
- Markowitz, H. M. (1959). *Portfolio Selection*. John Wiley & Sons, New York.

- Merton, R. C. (1969, August). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics* 51, 247–257.
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance* 29, 449–470.
- Merton, R. C. (1981). On market timing and investment performance. i. an equilibrium theory of value for market forecasts. *Journal of Business* 54(3), 363–406.
- Merton, R. C. (1990). *Continuous-Time Finance*. Padstow, UK: Basil Blackwell Inc.
- Merton, R. C. and P. A. Samuelson (1974). Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods. *Journal of Financial Economics* 1, 67–94.
- Mitchell, M. and T. Pulvino (2001, December). Characteristics of risk and return in risk arbitrage. *Journal of Finance* 56(6), 2135 – 2175.
- Mulvey, J. M. and A. E. Thorlacius (1998). The towers perrin global capital market scenario generation system. In J. M. Mulvey and W. T. Ziemba (Eds.), *Worldwide Asset and Liability Modeling*, pp. 286–312. Cambridge University Press.
- Normandin, M. (1994, April). Precautionary saving: An explanation for excess sensitivity of consumption. *Journal of Business and Economic Statistics* 12(2), 205–219.
- Normandin, M. (1997). Precautionary saving and the deaton paradox. *Applied Economics Letters* 4, 187–190.
- Osterberg, W. P. and J. B. Thomson (1999, May 1). The truth about hedge funds. Federal Reserve Bank of Cleveland, Economic Commentary Series.
- Perold, A. F. and W. F. Sharpe (1988). Dynamic strategies for asset allocation. *Financial Analysts Journal*, 16–27. January-February 1988.
- Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review* 46, 657–685.
- Randall, J. and S. Satchell (1997). An analysis of the hedging approach to modelling pension fund liabilities. Pensions Institute Discussion Paper PI-9714, Birkbeck College, University of London.
- Rockafellar, R. and S. Uryasev (2000). Optimization of conditional value-at-risk. *The Journal of Risk* 2(3), 21–41.

- Samuelson, P. A. (1994, Fall). The long-term case for equities, and how it can be oversold. *The Journal of Portfolio Management*, 15–24.
- Seckin, A. (2000a). Consumption with habit formation. Cirano working paper 2000s-39, University of Montréal.
- Seckin, A. (2000b). Habit formation: A kind of prudence? Cirano working paper 2000s-42, University of Montréal.
- Sethi, S. P. (1998). Optimal consumption-investment decisions allowing for bankruptcy: A survey. In W. T. Ziemba and J. M. Mulvey (Eds.), *Worldwide Asset and Liability Modeling*. Cambridge University Press.
- Sharpe, W. F. (1992). Asset allocation: Management style and performance measurement. *Journal of Portfolio Management* 18, 7–19.
- Sharpe, W. F. (1998, July/August). Morningstar’s risk-adjusted ratings. *Financial Analysts Journal* 54(4), 21–33.
- Shefrin, H. and M. Statman (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance* 40(3), 777–790.
- Shefrin, H. and M. Statman (2000). Behavioral portfolio theory. *Journal of Financial and Quantitative Analysis* 35(2), 127–151.
- Shleifer, A. (2000). *Inefficient Markets*. Oxford University Press.
- Shumway, T. (1997). Explaining returns with loss aversion. Working Paper. University of Michigan Business School.
- Siegel, J. J. (1992, January-February). The equity premium: Stock and bond returns since 1802. *Financial Analysts Journal*, 28–38.
- Sortino, F. and R. Van der Meer (1991, Summer). Downside risk: Capturing what’s at stake in investment situations. *The Journal of Portfolio Management*, 27–31.
- Thaler, R. and H. Shefrin (1981). An economic theory of self-control. *Journal of Political Economy* 89, 392–410.
- Thaler, R. H. (1994). Psychology and savings policies. *American Economic Review* 84(2), 186–192.
- Thaler, R. H. and E. J. Johnson (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science* 36(6), 643–660.
- Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5, 297–323.

- Zenios, S. A. (1995). Asset/Liability management under uncertainty for fixed-income securities. *Annals of Operations Research* 59, 77–97.
- Zenios, S. A., M. R. Holmer, R. McKendall, and C. Vassiadou-Zeniou (1998). Dynamic models for fixed-income portfolio management under uncertainty. *Journal of Economic Dynamics and Control*. 22, 1517–1541.
- Ziemia, W. and J. Mulvey (Eds.) (1998). *Worldwide Asset and Liability Modeling*. Cambridge University Press.