Dynamic Asset Allocation

Claus Munk

Until August 2012:
Aarhus University, e-mail: cmunk@econ.au.dk

From August 2012:
Copenhagen Business School, e-mail: cm.fi@cbs.dk

this version: July 3, 2012

The document contains graphs in color, use color printer for best results.
## Contents

Preface v

1 Introduction to asset allocation 1
   1.1 Introduction ......................................................... 1
   1.2 Investor classes and motives for investments .......................... 1
   1.3 Typical investment advice ............................................. 3
   1.4 How do individuals allocate their wealth? ............................ 3
   1.5 An overview of the theory of optimal investments .................... 3
   1.6 The future of investment management and services .......................... 3
   1.7 Outline of the rest ....................................................... 3
   1.8 Notation ................................................................. 3

2 Preferences 5
   2.1 Introduction ............................................................. 5
   2.2 Consumption plans and preference relations ............................ 6
   2.3 Utility indices ............................................................ 9
   2.4 Expected utility representation of preferences ......................... 10
   2.5 Risk aversion ............................................................ 16
   2.6 Utility functions in models and in reality ............................... 20
   2.7 Preferences for multi-date consumption plans ........................... 26
   2.8 Exercises .............................................................. 33

3 One-period models 37
   3.1 Introduction ............................................................. 37
   3.2 The general one-period model ......................................... 37
   3.3 Mean-variance analysis ............................................... 43
   3.4 A numerical example ................................................... 49
   3.5 Mean-variance analysis with constraints ............................... 49
   3.6 Estimation .............................................................. 49
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7</td>
<td>Critique of the one-period framework</td>
<td>49</td>
</tr>
<tr>
<td>3.8</td>
<td>Exercises</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>Discrete-time multi-period models</td>
<td>51</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>A multi-period, discrete-time framework for asset allocation</td>
<td>51</td>
</tr>
<tr>
<td>4.3</td>
<td>Dynamic programming in discrete-time models</td>
<td>54</td>
</tr>
<tr>
<td>5</td>
<td>Introduction to continuous-time modelling</td>
<td>59</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>59</td>
</tr>
<tr>
<td>5.2</td>
<td>The basic continuous-time setting</td>
<td>59</td>
</tr>
<tr>
<td>5.3</td>
<td>Dynamic programming in continuous-time models</td>
<td>62</td>
</tr>
<tr>
<td>5.4</td>
<td>Loss from suboptimal strategies</td>
<td>66</td>
</tr>
<tr>
<td>5.5</td>
<td>Exercises</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>Asset allocation with constant investment opportunities</td>
<td>69</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>69</td>
</tr>
<tr>
<td>6.2</td>
<td>General utility function</td>
<td>70</td>
</tr>
<tr>
<td>6.3</td>
<td>CRRA utility function</td>
<td>72</td>
</tr>
<tr>
<td>6.4</td>
<td>Logarithmic utility</td>
<td>75</td>
</tr>
<tr>
<td>6.5</td>
<td>Discussion of the optimal investment strategy for CRRA utility</td>
<td>76</td>
</tr>
<tr>
<td>6.6</td>
<td>The life-cycle</td>
<td>78</td>
</tr>
<tr>
<td>6.7</td>
<td>Loss due to suboptimal investments</td>
<td>80</td>
</tr>
<tr>
<td>6.8</td>
<td>Infrequent rebalancing of the portfolio</td>
<td>81</td>
</tr>
<tr>
<td>6.9</td>
<td>Exercises</td>
<td>84</td>
</tr>
<tr>
<td>7</td>
<td>Stochastic investment opportunities: the general case</td>
<td>85</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>85</td>
</tr>
<tr>
<td>7.2</td>
<td>General utility functions</td>
<td>86</td>
</tr>
<tr>
<td>7.3</td>
<td>CRRA utility</td>
<td>93</td>
</tr>
<tr>
<td>7.4</td>
<td>Logarithmic utility</td>
<td>105</td>
</tr>
<tr>
<td>7.5</td>
<td>How costly are deviations from the optimal investment strategy?</td>
<td>105</td>
</tr>
<tr>
<td>7.6</td>
<td>Exercises</td>
<td>107</td>
</tr>
<tr>
<td>8</td>
<td>The martingale approach</td>
<td>111</td>
</tr>
<tr>
<td>8.1</td>
<td>The martingale approach in complete markets</td>
<td>111</td>
</tr>
<tr>
<td>8.2</td>
<td>Complete markets and constant investment opportunities</td>
<td>115</td>
</tr>
<tr>
<td>8.3</td>
<td>Complete markets and stochastic investment opportunities</td>
<td>119</td>
</tr>
<tr>
<td>8.4</td>
<td>The martingale approach with portfolio constraints</td>
<td>120</td>
</tr>
<tr>
<td>8.5</td>
<td>Exercises</td>
<td>127</td>
</tr>
<tr>
<td>9</td>
<td>Numerical methods for solving dynamic asset allocation problems</td>
<td>129</td>
</tr>
</tbody>
</table>
Contents

10 Asset allocation with stochastic interest rates
10.1 Introduction .................................................. 131
10.2 One-factor Vasicek interest rate dynamics .................. 132
10.3 One-factor CIR dynamics ................................... 135
10.4 A numerical example ....................................... 137
10.5 Two-factor Vasicek model .................................. 143
10.6 Other studies with stochastic interest rates ................ 146
10.7 Exercises .................................................... 149

11 Asset allocation with stochastic market prices of risk
11.1 Introduction .................................................. 153
11.2 Mean reversion in stock returns ........................... 153
11.3 Stochastic volatility ........................................ 160
11.4 More ......................................................... 164
11.5 Exercises .................................................... 164

12 Inflation risk and asset allocation with no risk-free asset
12.1 Introduction .................................................. 167
12.2 Real and nominal price dynamics .......................... 167
12.3 Constant investment opportunities ......................... 169
12.4 General stochastic investment opportunities ............... 172
12.5 Hedging real interest rate risk without real bonds ....... 172

13 Labor income
13.1 Introduction .................................................. 179
13.2 A motivating example ....................................... 179
13.3 Exogenous income in a complete market .................. 181
13.4 Exogenous income in incomplete markets ................. 189
13.5 Endogenous labor supply and income ...................... 191
13.6 More ......................................................... 194

14 Consumption and portfolio choice with housing

15 Other variations of the problem...
15.1 Multiple and/or durable consumption goods .............. 197
15.2 Uncertain time of death; insurance ....................... 197

16 International asset allocation

17 Non-standard assumptions on investors
17.1 Preferences with habit formation .......................... 201
17.2 Recursive utility ............................................ 203
17.3 Model/parameter uncertainty, incomplete information, learning ........................................ 210
17.4 Ambiguity aversion ........................................ 210
17.5 Other objective functions .................................. 210
17.6 Consumption and portfolio choice for non-price takers ........ 210
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.7</td>
<td>Non-utility based portfolio choice</td>
<td>210</td>
</tr>
<tr>
<td>17.8</td>
<td>Allowing for bankruptcy</td>
<td>211</td>
</tr>
<tr>
<td>18</td>
<td>Trading and information imperfections</td>
<td>213</td>
</tr>
<tr>
<td>18.1</td>
<td>Trading constraints</td>
<td>213</td>
</tr>
<tr>
<td>18.2</td>
<td>Transaction costs</td>
<td>213</td>
</tr>
<tr>
<td>A</td>
<td>Results on the lognormal distribution</td>
<td>219</td>
</tr>
<tr>
<td>B</td>
<td>Stochastic processes and stochastic calculus</td>
<td>223</td>
</tr>
<tr>
<td>B.1</td>
<td>Introduction</td>
<td>223</td>
</tr>
<tr>
<td>B.2</td>
<td>What is a stochastic process?</td>
<td>224</td>
</tr>
<tr>
<td>B.3</td>
<td>Brownian motions</td>
<td>231</td>
</tr>
<tr>
<td>B.4</td>
<td>Diffusion processes</td>
<td>234</td>
</tr>
<tr>
<td>B.5</td>
<td>Itô processes</td>
<td>237</td>
</tr>
<tr>
<td>B.6</td>
<td>Stochastic integrals</td>
<td>237</td>
</tr>
<tr>
<td>B.7</td>
<td>Itô’s Lemma</td>
<td>241</td>
</tr>
<tr>
<td>B.8</td>
<td>Important diffusion processes</td>
<td>242</td>
</tr>
<tr>
<td>B.9</td>
<td>Multi-dimensional processes</td>
<td>249</td>
</tr>
<tr>
<td>B.10</td>
<td>Change of probability measure</td>
<td>255</td>
</tr>
<tr>
<td>B.11</td>
<td>Exercises</td>
<td>258</td>
</tr>
<tr>
<td>C</td>
<td>Solutions to Ordinary Differential Equations</td>
<td>261</td>
</tr>
<tr>
<td>References</td>
<td>263</td>
<td></td>
</tr>
</tbody>
</table>
INCOMPLETE!
Preliminary and incomplete lecture notes intended for use at an advanced master’s level or an introductory Ph.D. level. I appreciate comments and corrections from Kenneth Brandborg, Jens Henrik Eggert Christensen, Heine Jepsen, Thomas Larsen, Jakob Nielsen, Nicolai Nielsen, Kenneth Winther Pedersen, Carsten Sørensen, and in particular Linda Sandris Larsen. Additional comments and suggestions are very welcome!

Claus Munk
Internet homepage: sites.google.com/site/munkfinance
CHAPTER 1

Introduction to asset allocation

1.1 Introduction

Financial markets offer opportunities to move money between different points in time and different states of the world. Investors must decide how much to invest in the financial markets and how to allocate that amount between the many, many available financial securities. Investors can change their investments as time passes and they will typically want to do so for example when they obtain new information about the prospective returns on the financial securities. Hence, they must figure out how to manage their portfolio over time. In other words, they must determine an investment strategy or an asset allocation strategy. The term asset allocation is sometimes used for the allocation of investments to major asset classes, e.g., stocks, bonds, and cash. In later chapters we will often focus on this decision, but we will use the term asset allocation interchangeably with the terms optimal investment or portfolio management.

It is intuitively clear that in order to determine the optimal investment strategy for an investor, we must make some assumptions about the objectives of the investor and about the possible returns on the financial markets. Different investors will have different motives for investments and hence different objectives. In Section 1.2 we will discuss the motives and objectives of different types of investors. We will focus on the asset allocation decisions of individual investors or households. Individuals invest in the financial markets to finance future consumption of which they obtain some felicity or utility. We discuss how to model the preferences of individuals in Chapter 2.

1.2 Investor classes and motives for investments

We can split the investors into individual investors (households; sometimes called retail investors) and institutional investors (includes both financial intermediaries – such as pension funds, insurance companies, mutual funds, and commercial banks – and manufacturing companies producing goods or services). Different investors have different objectives. Manufacturing companies probably invest mostly in short-term bonds and deposits in order to manage their liquidity needs and avoid the
deadweight costs of raising small amounts of capital very frequently. They will rarely set up long-term strategies for investments in the financial markets and their financial investments constitute a very small part of the total investments.

Individuals can use their money either for consumption or savings. Here we use the term savings synonymously with financial investments so that it includes both deposits in banks and investments in stocks, bonds, and possibly other securities. Traditionally most individuals have saved in form of bank deposits and maybe government bonds, but in recent years there has been an increasing interest of individuals for investing in the stock market. Individuals typically save when they are young by consuming less than the labor income they earn, primarily in order to accumulate wealth they can use for consumption when they retire. Other motives for saving is to be able to finance large future expenditures (e.g., purchase of real estate, support of children during their education, expensive celebrations or vacations) or simply to build up a buffer for “hard times” due to unemployment, disability, etc. We assume that the objective of an individual investor is to maximize the utility of consumption throughout the life-time of the investor. We will discuss utility functions in Chapter 2.

A large part of the savings of individuals are indirect through pension funds and mutual funds. These funds are the major investors in today’s markets. Some of these funds are non-profit funds that are owned by the investors in the fund. The objective of such funds should represent the objectives of the fund investors.

Let us look at pension funds. One could imagine a pension fund that determines the optimal portfolio of each of the fund investors and aggregates over all investors to find the portfolio of the fund. Each fund investor is then allocated the returns on his optimal portfolio, probably net of some servicing fee. The purpose of forming the fund is then simply to save transaction costs. A practical implementation of this is to let each investor allocate his funds among some pre-selected portfolios, for example a portfolio mimicking the overall stock market index, various portfolios of stocks in different industries, one or more portfolios of government bonds (e.g., one in short-term and one in long-term bonds), portfolios of corporate bonds and mortgage-backed bonds, portfolios of foreign stocks and bonds, and maybe also portfolios of derivative securities and even non-financial portfolios of metals and real estate. Some pension funds operate in this way and there seems to be a tendency for more and more pension funds to allow investor discretion with regards to the way the deposits are invested.

However, in many pension funds some hired fund managers decide on the investment strategy. Often all the deposits of different fund members are pooled together and then invested according to a portfolio chosen by the fund managers (probably following some general guidelines set up by the board of the fund). Once in a while the rate of return of the portfolio is determined and the deposit of each investor is increased according to this rate of return less some servicing fee. In many cases the returns on the portfolio of the fund are distributed to the fund members using more complicated schemes. Rate of return guarantees, bonus accounts,.... The salary of the manager of a fund is often linked to the return on the portfolio he chooses and some benchmark portfolio(s). A rational manager will choose a portfolio that maximizes his utility and that portfolio choice may be far from the optimal portfolio of the fund members....

Mutual funds...

This lecture note will focus on the decision problem of an individual investor and aims to analyze
and answer the following questions:

- What are the utility maximizing dynamic consumption and investment strategies of an individual?
- What is the relation between optimal consumption and optimal investment?
- How are financial investments optimally allocated to different asset classes, e.g., stocks and bonds?
- How are financial investments optimally allocated to single securities within each asset class?
- How does the optimal consumption and investment strategies depend on, e.g., risk aversion, time horizon, initial wealth, labor income, and asset price dynamics?
- Are the recommendations of investment advisors consistent with the theory of optimal investments?

1.3 Typical investment advice

TO COME... References: Quinn (1997), Siegel (2002)


1.4 How do individuals allocate their wealth?

TO COME...


Christiansen, Joensen, and Rangvid (2008): differences due to education
Yang (2009): house owners vs. non-owners

1.5 An overview of the theory of optimal investments

TO COME...

1.6 The future of investment management and services

TO COME... References: Bodie (2003), Merton (2003)

1.7 Outline of the rest

1.8 Notation

Since we are going to deal simultaneously with many financial assets, it will often be mathematically convenient to use vectors and matrices. All vectors are considered column vectors. The
superscript $\top$ on a vector or a matrix indicates that the vector or matrix is transposed. We will use the notation $\mathbf{1}$ for a vector where all elements are equal to 1; the dimension of the vector will be clear from the context. We will use the notation $\mathbf{e}_i$ for a vector $(0, \ldots, 0, 1, 0, \ldots, 0)^\top$ where the 1 is entry number $i$. Note that for two vectors $\mathbf{x} = (x_1, \ldots, x_d)^\top$ and $\mathbf{y} = (y_1, \ldots, y_d)^\top$ we have $\mathbf{x}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{x} = \sum_{i=1}^d x_i y_i$. In particular, $\mathbf{x}^\top \mathbf{1} = \sum_{i=1}^d x_i$ and $\mathbf{e}_i^\top \mathbf{x} = x_i$. We also define $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x} = \sum_{i=1}^d x_i^2$.

If $\mathbf{x} = (x_1, \ldots, x_n)$ and $f$ is a real-valued function of $\mathbf{x}$, then the (first-order) derivative of $f$ with respect to $\mathbf{x}$ is the vector

$$f'(\mathbf{x}) \equiv \mathbf{f}_x(\mathbf{x}) = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)^\top.$$ 

This is also called the gradient of $f$. The second-order derivative of $f$ is the $n \times n$ Hessian matrix

$$f''(\mathbf{x}) \equiv \mathbf{f}_{xx}(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$ 

If $\mathbf{x}$ and $\mathbf{a}$ are $n$-dimensional vectors, then

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{a}^\top \mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^\top \mathbf{a}) = \mathbf{a}.$$ 

If $\mathbf{x}$ is an $n$-dimensional vector and $\mathbf{A}$ is a symmetric [i.e., $\mathbf{A} = \mathbf{A}^\top$] $n \times n$ matrix, then

$$\frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^\top \mathbf{A} \mathbf{x}) = 2\mathbf{A} \mathbf{x}.$$ 

If $\mathbf{A}$ is non-singular, then $(\mathbf{A} \mathbf{A}^\top)^{-1} = (\mathbf{A}^\top)^{-1} \mathbf{A}^{-1}$. 

\[ \text{Chapter 1. Introduction to asset allocation} \]
2.1 Introduction

In order to say anything concrete about the optimal investments of individuals we have to formalize the decision problem faced by individuals. We assume that individuals have preferences for consumption and must choose between different consumption plans, i.e., plans for how much to consume at different points in time and in different states of the world. The financial market allows individuals to reallocate consumption over time and over states and hence obtain a consumption plan different from their endowment.

Although an individual will typically obtain utility from consumption at many different dates (or in many different periods), we will first address the simpler case with consumption at only one future point in time. In such a setting a “consumption plan” is simply a random variable representing the consumption at that date. Even in one-period models individuals should be allowed to consume both at the beginning of the period and at the end of the period, but we will first ignore the influence of current consumption on the well-being of the individual. We do that both since current consumption is certain and we want to focus on how preferences for uncertain consumption can be represented, but also to simplify the notation and analysis somewhat. Since we have in mind a one-period economy, we basically have to model preferences for end-of-period consumption.

Sections 2.2–2.4 discuss how to represent individual preferences in a tractable way. We will demonstrate that under some fundamental assumptions (“axioms”) on individual behavior, the preferences can be modeled by a utility index which to each consumption plan assigns a real number with higher numbers to the more preferred plans. Under an additional axiom we can represent the preferences in terms of expected utility, which is even simpler to work with and used in most models of financial economics. Section 2.5 defines and discusses the important concept of risk aversion. Section 2.6 introduces the utility functions that are typically applied in models of financial economics and provides a short discussion of which utility functions and levels of risk aversions that seem to be reasonable for representing the decisions of individuals. In Section 2.7
we discuss extensions to preferences for consumption at more than one point in time.

There is a large literature on how to model the preferences of individuals for uncertain outcomes and the presentation here is by no means exhaustive. The literature dates back at least to the Swiss mathematician Daniel Bernoulli in 1738 (see English translation in Bernoulli (1954)), but was put on a firm formal setting by von Neumann and Morgenstern (1944). For some recent textbook presentations on a similar level as the one given here, see Huang and Litzenberger (1988, Ch. 1), Kreps (1990, Ch. 3), Gollier (2001, Chs. 1-3), and Danthine and Donaldson (2002, Ch. 2).

2.2 Consumption plans and preference relations

It seems fair to assume that whenever the individual compares two different consumption plans, she will be able either to say that she prefers one of them to the other or to say that she is indifferent between the two consumption plans. Moreover, she should make such pairwise comparisons in a consistent way. For example, if she prefers plan 1 to plan 2 and plan 2 to plan 3, she should prefer plan 1 to plan 3. If these properties hold, we can formally represent the preferences of the individual by a so-called preference relation. A preference relation itself is not very tractable so we are looking for simpler ways of representing preferences. First, we will find conditions under which it makes sense to represent preferences by a so-called utility index which attaches a real number to each consumption plan. If and only if plan 1 has a higher utility index than plan 2, the individual prefers plan 1 to plan 3. If these properties hold, we can formally represent the preferences of the individual by a so-called preference relation. A preference relation itself is not very tractable so we are looking for simpler ways of representing preferences. First, we will find conditions under which it makes sense to represent preferences by a so-called utility index which attaches a real number to each consumption plan. If and only if plan 1 has a higher utility index than plan 2, the individual prefers plan 1 to plan 2. Attaching numbers to each possible consumption plan is also not easy so we look for an even simpler representation. We show that under an additional condition we can represent preferences in an even simpler way in terms of the expected value of a utility function. A utility function is a function defined on the set of possible levels of consumption. Since consumption is random it then makes sense to talk about the expected utility of a consumption plan. The individual will prefer consumption plan 1 to plan 2 if and only if the expected utility from consumption plan 1 is higher than the expected utility from consumption plan 2. This representation of preferences turns out to be very tractable and is applied in the vast majority of asset pricing models.

Our main analysis is formulated under some simplifying assumptions that are not necessarily appropriate. At the end of this section we will briefly discuss how to generalize the analysis and also discuss the appropriateness of the axioms on individual behavior that need to be imposed in order to obtain the expected utility representation.

We assume that there is uncertainty about how the variables affecting the well-being of an individual (e.g., asset returns) turn out. We model the uncertainty by a probability space \((\Omega, \mathcal{F}, \mathbb{P})\). In most of the chapter we will assume that the state space is finite, \(\Omega = \{1, 2, \ldots, S\}\), so that there are \(S\) possible states of which exactly one will be realized. For simplicity, think of this as a model of one-period economy with \(S\) possible states at the end of the period. The set \(\mathcal{F}\) of events that can be assigned a probability is the collection of all subsets of \(\Omega\). The probability measure \(\mathbb{P}\) is defined by the individual state probabilities \(p_\omega = \mathbb{P}(\omega), \omega = 1, 2, \ldots, S\). We assume that all \(p_\omega > 0\) and, of course, we have that \(p_1 + \ldots + p_S = 1\). We take the state probabilities as exogenously given and known to the individuals.

Individuals care about their consumption. It seems reasonable to assume that when an individual chooses between two different actions (e.g., portfolio choices), she only cares about the consumption
2.2 Consumption plans and preference relations

<table>
<thead>
<tr>
<th>state ω</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>state prob. $p_ω$</td>
<td>0.2</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>cons. plan 1, $c^{(1)}$</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>cons. plan 2, $c^{(2)}$</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>cons. plan 3, $c^{(3)}$</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>cons. plan 4, $c^{(4)}$</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2.1: The possible state-contingent consumption plans in the example.

plans generated by these choices. For example, she will be indifferent between two choices that generate exactly the same consumption plans, i.e., the same consumption levels in all states. In order to simplify the following analysis, we will assume a bit more, namely that the individual only cares about the probability distribution of consumption generated by each portfolio. This is effectively an assumption of state-independent preferences.

We can represent a consumption plan by a random variable $c$ on $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that there is only one consumption good and since consumption should be non-negative, $c$ is valued in $\mathbb{R}_+ = [0, \infty)$. As long as we are assuming a finite state space $\Omega = \{1, 2, \ldots, S\}$ we can equivalently represent the consumption plan by a vector $(c_1, \ldots, c_S)$, where $c_ω \in [0, \infty)$ denotes the consumption level if state $ω$ is realized, i.e., $c_ω \equiv c(ω)$. Let $\mathcal{C}$ denote the set of consumption plans that the individual has to choose among. Let $Z \subseteq \mathbb{R}_+$ denote the set of all the possible levels of the consumption plans that are considered, i.e., no matter which of these consumption plans we take, its value will be in $Z$ no matter which state is realized. Each consumption plan $c \in \mathcal{C}$ is associated with a probability distribution $\pi_c$, which is the function $\pi_c: Z \rightarrow [0, 1]$, given by

$$\pi_c(z) = \sum_{ω \in \Omega: c_ω = z} p_ω,$$

i.e., the sum of the probabilities of those states in which the consumption level equals $z$.

As an example consider an economy with three possible states and four possible state-contingent consumption plans as illustrated in Table 2.1. These four consumption plans may be the product of four different portfolio choices. The set of possible end-of-period consumption levels is $Z = \{1, 2, 3, 4, 5\}$. Each consumption plan generates a probability distribution on the set $Z$. The probability distributions corresponding to these consumption plans are as shown in Table 2.2. We see that although the consumption plans $c^{(3)}$ and $c^{(4)}$ are different they generate identical probability distributions. By assumption individuals will be indifferent between these two consumption plans.

Given these assumptions the individual will effectively choose between probability distributions on the set of possible consumption levels $Z$. We assume for simplicity that $Z$ is a finite set, but the results can be generalized to the case of infinite $Z$ at the cost of further mathematical complexity. We denote by $\mathcal{P}(Z)$ the set of all probability distributions on $Z$ that are generated by consumption plans in $\mathcal{C}$. A probability distribution $\pi$ on the finite set $Z$ is simply a function $\pi: Z \rightarrow [0, 1]$ with the properties that $\sum_{z \in Z} \pi(z) = 1$ and $\pi(A \cup B) = \pi(A) + \pi(B)$ whenever $A \cap B = \emptyset$.

We assume that the preferences of the individual can be represented by a preference relation $\succeq$ on $\mathcal{P}(Z)$, which is a binary relation satisfying the following two conditions:
Table 2.2: The probability distributions corresponding to the state-contingent consumption plans shown in Table 2.1.

<table>
<thead>
<tr>
<th>cons. level $z$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>cons. plan 1, $\pi_c(1)$</td>
<td>0</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cons. plan 2, $\pi_c(2)$</td>
<td>0.3</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>cons. plan 3, $\pi_c(3)$</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>cons. plan 4, $\pi_c(4)$</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
</tbody>
</table>

(i) if $\pi_1 \succeq \pi_2$ and $\pi_2 \succeq \pi_3$, then $\pi_1 \succeq \pi_3$ [transitivity]

(ii) $\forall \pi_1, \pi_2 \in \mathcal{P}(Z)$ : either $\pi_1 \succeq \pi_2$ or $\pi_2 \succeq \pi_1$ [completeness]

Here, $\pi_1 \succeq \pi_2$ is to be read as “$\pi_1$ is preferred to $\pi_2$”. We write $\pi_1 \not\succeq \pi_2$ if $\pi_1$ is not preferred to $\pi_2$. If both $\pi_1 \succeq \pi_2$ and $\pi_2 \succeq \pi_1$, we write $\pi_1 \sim \pi_2$ and say that the individual is indifferent between $\pi_1$ and $\pi_2$. If $\pi_1 \succeq \pi_2$, but $\pi_2 \not\succeq \pi_1$, we say that $\pi_1$ is strictly preferred to $\pi_2$ and write $\pi_1 \succ \pi_2$. If both $\pi_1 \succ \pi_2$ and $\pi_2 \succ \pi_1$, we write $\pi_1 \sim \pi_2$ and say that the individual is indifferent between $\pi_1$ and $\pi_2$.

Note that if $\pi_1, \pi_2 \in \mathcal{P}(Z)$ and $\alpha \in [0, 1]$, then $\alpha \pi_1 + (1 - \alpha)\pi_2 \in \mathcal{P}(Z)$. The mixed distribution $\alpha \pi_1 + (1 - \alpha)\pi_2$ assigns the probability $\alpha \pi_1(z) + (1 - \alpha)\pi_2(z)$ to the consumption level $z$. When can think of the mixed distribution $\alpha \pi_1 + (1 - \alpha)\pi_2$ as the outcome of a two-stage “gamble.” The first stage is to flip a coin which with probability $\alpha$ shows head and with probability $1 - \alpha$ shows tails. If head comes out, the second stage is the “consumption gamble” corresponding to the probability distribution $\pi_1$. If tails is the outcome of the first stage, the second stage is the consumption gamble corresponding to $\pi_2$. When we assume that preferences are represented by a preference relation on the set $\mathcal{P}(Z)$ of probability distributions, we have implicitly assumed that the individual evaluates the two-stage gamble (or any multi-stage gamble) by the combined probability distribution, i.e., the ultimate consequences of the gamble. This is sometimes referred to as consequentialism.

Let $z$ be some element of $Z$, i.e., some possible consumption level. By $1_z$ we will denote the probability distribution that assigns a probability of one to $z$ and a zero probability to all other elements in $Z$. Since we have assumed that the set $Z$ of possible consumption levels only has a finite number of elements, it must have a maximum element, say $z^*$, and a minimum element, say $z'$. Since the elements represent consumption levels, it is certainly natural that individuals prefer higher elements than lower. We will therefore assume that the probability distribution $1_{z^*}$ is preferred to any other probability distribution. Conversely, any probability distribution is preferred to the probability distribution $1_{z'}$. We assume that $1_{z^*}$ is strictly preferred to $1_{z'}$ so that the individual is not indifferent between all probability distributions. For any $\pi \in \mathcal{P}(Z)$ we thus have that,

$$1_{z^*} \succ \pi \succ 1_{z'} \quad \text{or} \quad 1_{z^*} \sim \pi \succ 1_{z'} \quad \text{or} \quad 1_{z^*} \succ \pi \sim 1_{z'}.$$
2.3 Utility indices

A utility index for a given preference relation $\succeq$ is a function $U : \mathcal{P}(Z) \to \mathbb{R}$ that to each probability distribution over consumption levels attaches a real-valued number such that

$$\pi_1 \succeq \pi_2 \iff U(\pi_1) \geq U(\pi_2).$$

Note that a utility index is only unique up to a strictly increasing transformation. If $U$ is a utility index and $f : \mathbb{R} \to \mathbb{R}$ is any strictly increasing function, then the composite function $V = f \circ U$, defined by $V(\pi) = f(U(\pi))$, is also a utility index for the same preference relation.

We will show below that a utility index exists under the following two axiomatic assumptions on the preference relation $\succeq$:

Axiom 2.1 (Monotonicity). Suppose that $\pi_1, \pi_2 \in \mathcal{P}(Z)$ with $\pi_1 \succ \pi_2$ and let $a, b \in [0, 1]$. The preference relation $\succeq$ has the property that

$$a > b \iff a\pi_1 + (1 - a)\pi_2 \succ b\pi_1 + (1 - b)\pi_2.$$

This is certainly a very natural assumption on preferences. If you consider a weighted average of two probability distributions, you will prefer a high weight on the best of the two distributions.

Axiom 2.2 (Archimedean). The preference relation $\succeq$ has the property that for any three probability distributions $\pi_1, \pi_2, \pi_3 \in \mathcal{P}(Z)$ with $\pi_1 \succ \pi_2 \succ \pi_3$, numbers $a, b \in (0, 1)$ exist such that

$$a\pi_1 + (1 - a)\pi_3 \succ \pi_2 \succ b\pi_1 + (1 - b)\pi_3.$$

The axiom basically says that no matter how good a probability distribution $\pi_1$ is, it is so that for any $\pi_2 \succ \pi_3$ we can find some mixed distribution of $\pi_1$ and $\pi_3$ to which $\pi_2$ is preferred. We just have to put a sufficiently low weight on $\pi_1$ in the mixed distribution. Similarly, no matter how bad a probability distribution $\pi_3$ is, it is so that for any $\pi_1 \succ \pi_2$ we can find some mixed distribution of $\pi_1$ and $\pi_3$ that is preferred to $\pi_2$. We just have to put a sufficiently low weight on $\pi_3$ in the mixed distribution.

We shall say that a preference relation has the continuity property if for any three probability distributions $\pi_1, \pi_2, \pi_3 \in \mathcal{P}(Z)$ with $\pi_1 \succ \pi_2 \succ \pi_3$, a unique number $\alpha \in (0, 1)$ exists such that

$$\pi_2 \sim \alpha \pi_1 + (1 - \alpha)\pi_3.$$ 

We can easily extend this to the case where either $\pi_1 \sim \pi_2$ or $\pi_2 \sim \pi_3$. For $\pi_1 \sim \pi_2 \succ \pi_3$, $\pi_2 \sim 1\pi_1 + (1 - 1)\pi_3$ corresponding to $\alpha = 1$. For $\pi_1 \succ \pi_2 \sim \pi_3$, $\pi_2 \sim 0\pi_1 + (1 - 0)\pi_3$ corresponding to $\alpha = 0$. In words the continuity property means that for any three probability distributions there is a unique combination of the best and the worst distribution so that the individual is indifferent between the third “middle” distribution and this combination of the other two. This appears to be closely related to the Archimedean Axiom and, in fact, the next lemma shows that the Monotonicity Axiom and the Archimedean Axiom imply continuity of preferences.

Lemma 2.1. Let $\succeq$ be a preference relation satisfying the Monotonicity Axiom and the Archimedean Axiom. Then it has the continuity property.

Proof. Given $\pi_1 \succ \pi_2 \succ \pi_3$. Define the number $\alpha$ by

$$\alpha = \sup\{k \in [0, 1] \mid \pi_2 \succ k\pi_1 + (1 - k)\pi_3\}.$$
By the Monotonicity Axiom we have that \( \pi_2 \succ k\pi_1 + (1 - k)\pi_3 \) for all \( k < \alpha \) and that \( k\pi_1 + (1 - k)\pi_3 \succeq \pi_2 \) for all \( k > \alpha \). We want to show that \( \pi_2 \sim \alpha\pi_1 + (1 - \alpha)\pi_3 \). Note that by the Archimedean Axiom, there is some \( k > 0 \) such that \( \pi_2 \succ k\pi_1 + (1 - k)\pi_3 \) and some \( k < 1 \) such that \( k\pi_1 + (1 - k)\pi_3 \succeq \pi_2 \). Consequently, \( \alpha \) is in the open interval \((0,1)\).

Suppose that \( \pi_2 \succ \alpha\pi_1 + (1 - \alpha)\pi_3 \). Then according to the Archimedean Axiom we can find a number \( b \in (0,1) \) such that \( \pi_2 \succ b\pi_1 + (1 - b)\{\alpha\pi_1 + (1 - \alpha)\pi_3\} \). The mixed distribution on the right-hand side has a total weight of \( k = b + (1 - b)\alpha = \alpha + (1 - \alpha)b > \alpha \) on \( \pi_1 \). Hence we have found some \( k > \alpha \) for which \( \pi_2 \succ k\pi_1 + (1 - k)\pi_3 \). This contradicts the definition of \( \alpha \). Consequently, we must have that \( \pi_2 \not\succ \alpha\pi_1 + (1 - \alpha)\pi_3 \).

Now suppose that \( \alpha\pi_1 + (1 - \alpha)\pi_3 \succeq \pi_2 \). Then we know from the Archimedean Axiom that a number \( a \in (0,1) \) exists such that \( a\{\alpha\pi_1 + (1 - \alpha)\pi_3\} \succeq (1 - a)\pi_3 \succ \pi_2 \). The mixed distribution on the left-hand side has a total weight of \( a\alpha < \alpha \) on \( \pi_1 \). Hence we have found some \( k < \alpha \) for which \( k\pi_1 + (1 - k)\pi_3 \succeq \pi_2 \). This contradicts the definition of \( \alpha \). We can therefore also conclude that \( \alpha\pi_1 + (1 - \alpha)\pi_3 \not\succeq \pi_2 \). In sum, we have \( \pi_2 \sim \alpha\pi_1 + (1 - \alpha)\pi_3 \).

The next result states that a preference relation which satisfies the Monotonicity Axiom and has the continuity property can always be represented by a utility index. In particular this is true when \( \succ \) satisfies the Monotonicity Axiom and the Archimedean Axiom.

**Theorem 2.1.** Let \( \succeq \) be a preference relation which satisfies the Monotonicity Axiom and has the continuity property. Then it can be represented by a utility index \( U \), i.e., a function \( U: \mathcal{P}(Z) \to \mathbb{R} \) with the property that

\[
\pi_1 \succeq \pi_2 \iff U(\pi_1) \geq U(\pi_2).
\]

**Proof.** Recall that we have assumed a best probability distribution \( 1_{z^*} \) and a worst probability distribution \( 1_{z^i} \) in the sense that

\[
1_{z^*} \succ \pi \succ 1_{z^i} \quad \text{or} \quad 1_{z^*} \sim \pi \succ 1_{z^i} \quad \text{or} \quad 1_{z^*} \succ \pi \sim 1_{z^i}
\]

for any \( \pi \in \mathcal{P}(Z) \). For any \( \pi \in \mathcal{P}(Z) \) we know from the continuity property that a unique number \( \alpha_\pi \in [0,1] \) exists such that

\[
\pi \sim \alpha_\pi 1_{z^*} + (1 - \alpha_\pi) 1_{z^i}.
\]

If \( 1_{z^*} \sim \pi \sim 1_{z^i} \), \( \alpha_\pi = 1 \). If \( 1_{z^*} \succ \pi \succ 1_{z^i} \), \( \alpha_\pi = 0 \). If \( 1_{z^*} \succ \pi \sim 1_{z^i} \), \( \alpha_\pi \in (0,1) \).

We define the function \( U: \mathcal{P}(Z) \to \mathbb{R} \) by \( U(\pi) = \alpha_\pi \). By the Monotonicity Axiom we know that \( U(\pi_1) \geq U(\pi_2) \) if and only if

\[
U(\pi_1) 1_{z^*} + (1 - U(\pi_1)) 1_{z^i} \succeq U(\pi_2) 1_{z^*} + (1 - U(\pi_2)) 1_{z^i},
\]

and hence if and only if \( \pi_1 \succeq \pi_2 \). It follows that \( U \) is a utility index.

### 2.4 Expected utility representation of preferences

Utility indices are functions of probability distributions on the set of possible consumption levels. With many states of the world and many assets to trade in, the set of such probability distributions will be very, very large. This will significantly complicate the analysis of optimal choice using utility indices to represent preferences. To simplify the analysis financial economists
traditionally put more structure on the preferences so that they can be represented in terms of expected utility.

We say that a preference relation $\succeq$ on $\mathcal{P}(Z)$ has an expected utility representation if there exists a function $u : Z \rightarrow \mathbb{R}$ such that

$$\pi_1 \succeq \pi_2 \iff \sum_{z \in Z} \pi_1(z)u(z) \geq \sum_{z \in Z} \pi_2(z)u(z). \quad (2.1)$$

Here $\sum_{z \in Z} \pi(z)u(z)$ is the expected utility of end-of-period consumption given the consumption probability distribution $\pi$, so (2.1) says that $E[u(c_1)] \geq E[u(c_2)]$, where $c_i$ is the random variable representing end-of-period consumption with associated consumption probability distribution $\pi_i$.

The function $u$ is called a von Neumann-Morgenstern utility function or simply a utility function. Note that $u$ is defined on the set $Z$ of consumption levels, which in general has a simpler structure than the set of probability distributions on $Z$. Given a utility function $u$, we can obviously define a utility index by $\mathcal{U}(\pi) = \sum_{z \in Z} \pi(z)u(z)$.

### 2.4.1 Conditions for expected utility

When can we use an expected utility representation of a preference relation? The next lemma is a first step.

**Lemma 2.2.** A preference relation $\succeq$ has an expected utility representation if and only if it can be represented by a linear utility index $\mathcal{U}$ in the sense that

$$\mathcal{U}(a\pi_1 + (1-a)\pi_2) = a\mathcal{U}(\pi_1) + (1-a)\mathcal{U}(\pi_2)$$

for any $\pi_1, \pi_2 \in \mathcal{P}(Z)$ and any $a \in [0,1]$.

**Proof.** Suppose that $\succeq$ has an expected utility representation with utility function $u$. Define $\mathcal{U} : \mathcal{P}(Z) \rightarrow \mathbb{R}$ by $\mathcal{U}(\pi) = \sum_{z \in Z} \pi(z)u(z)$. Then clearly $\mathcal{U}$ is a utility index representing $\succeq$ and $\mathcal{U}$ is linear since

$$\mathcal{U}(a\pi_1 + (1-a)\pi_2) = \sum_{z \in Z} (a\pi_1(z) + (1-a)\pi_2(z))u(z)$$

$$= a\sum_{z \in Z} \pi_1(z)u(z) + (1-a)\sum_{z \in Z} \pi_2(z)u(z)$$

$$= a\mathcal{U}(\pi_1) + (1-a)\mathcal{U}(\pi_2).$$

Conversely, suppose that $\mathcal{U}$ is a linear utility index representing $\succeq$. Define a function $u : Z \rightarrow \mathbb{R}$ by $u(z) = \mathcal{U}(1_z)$. For any $\pi \in \mathcal{P}(Z)$ we have

$$\pi \sim \sum_{z \in Z} \pi(z)1_z.$$

Therefore,

$$\mathcal{U}(\pi) = \mathcal{U} \left( \sum_{z \in Z} \pi(z)1_z \right) = \sum_{z \in Z} \pi(z)\mathcal{U}(1_z) = \sum_{z \in Z} \pi(z)u(z).$$

Since $\mathcal{U}$ is a utility index, we have $\pi_1 \succeq \pi_2 \iff \mathcal{U}(\pi_1) \geq \mathcal{U}(\pi_2)$, which the computation above shows is equivalent to $\sum_{z \in Z} \pi_1(z)u(z) \geq \sum_{z \in Z} \pi_2(z)u(z)$. Consequently, $u$ gives an expected utility representation of $\succeq$. \qed
Table 2.3: The probability distributions used in the illustration of the Substitution Axiom.

<table>
<thead>
<tr>
<th>z</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>π₁</td>
<td>0</td>
<td>0.2</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>π₂</td>
<td>0</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>π₃</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>π₄</td>
<td>0.5</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
</tr>
<tr>
<td>π₅</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

The question then is under what assumptions the preference relation $\succeq$ can be represented by a linear utility index. As shown by von Neumann and Morgenstern (1944) we need an additional axiom, the so-called Substitution Axiom.

**Axiom 2.3 (Substitution).** For all $\pi_1, \pi_2, \pi_3 \in \mathcal{P}(Z)$ and all $a \in [0, 1]$, we have

$$\pi_1 \succ \pi_2 \iff a\pi_1 + (1-a)\pi_3 \succ a\pi_2 + (1-a)\pi_3$$

and

$$\pi_1 \sim \pi_2 \iff a\pi_1 + (1-a)\pi_3 \sim a\pi_2 + (1-a)\pi_3.$$  

The Substitution Axiom is sometimes called the Independence Axiom or the Axiom of the Irrelevance of the Common Alternative. Basically, it says that when the individual is to compare two probability distributions, she needs only consider the parts of the two distributions which are different from each other. As an example, suppose the possible consumption levels are $Z = \{1, 2, 3, 4\}$ and consider the probability distributions on $Z$ given in Table 2.3. Suppose you want to compare the distributions $\pi_4$ and $\pi_5$. They only differ in the probabilities they associate with consumption levels 2, 3, and 4 so it should only be necessary to focus on these parts. More formally observe that

$$\pi_4 \sim 0.5\pi_1 + 0.5\pi_3 \quad \text{and} \quad \pi_5 \sim 0.5\pi_2 + 0.5\pi_3.$$  

$\pi_1$ is the conditional distribution of $\pi_4$ given that the consumption level is different from 1 and $\pi_2$ is the conditional distribution of $\pi_5$ given that the consumption level is different from 1. The Substitution Axiom then says that

$$\pi_4 \succ \pi_5 \iff \pi_1 \succ \pi_2.$$  

The next lemma shows that the Substitution Axiom is more restrictive than the Monotonicity Axiom.

**Lemma 2.3.** If a preference relation $\succeq$ satisfies the Substitution Axiom, it will also satisfy the Monotonicity Axiom.

**Proof.** Given $\pi_1, \pi_2 \in \mathcal{P}(Z)$ with $\pi_1 \succ \pi_2$ and numbers $a, b \in [0, 1]$. We have to show that

$$a > b \iff a\pi_1 + (1-a)\pi_2 \succ b\pi_1 + (1-b)\pi_2.$$  

Note that if $a = 0$, we cannot have $a > b$, and if $a\pi_1 + (1-a)\pi_2 \succ b\pi_1 + (1-b)\pi_2$ we cannot have $a = 0$. We can therefore safely assume that $a > 0$.  

\[\]
First assume that \( a > b \). Observe that it follows from the Substitution Axiom that
\[
a \pi_1 + (1 - a) \pi_2 > a \pi_2 + (1 - a) \pi_2
\]
and hence that \( a \pi_1 + (1 - a) \pi_2 > \pi_2 \). Also from the Substitution Axiom we have that for any \( \pi_3 > \pi_2 \), we have
\[
\pi_3 \sim \left(1 - \frac{b}{a}\right) \pi_3 + \frac{b}{a} \pi_3 \succ \left(1 - \frac{b}{a}\right) \pi_2 + \frac{b}{a} \pi_3.
\]
Due to our observation above, we can use this with \( \pi_3 = a \pi_1 + (1 - a) \pi_2 \). Then we get
\[
a \pi_1 + (1 - a) \pi_2 \succ \frac{b}{a} \{a \pi_1 + (1 - a) \pi_2\} + \left(1 - \frac{b}{a}\right) \pi_2
\]
\[
\sim b \pi_1 + (1 - b) \pi_2,
\]
as was to be shown.

Conversely, assuming that
\[
a \pi_1 + (1 - a) \pi_2 \succ b \pi_1 + (1 - b) \pi_2,
\]
we must argue that \( a > b \). The above inequality cannot be true if \( a = b \) since the two combined distributions are then identical. If \( b \) was greater than \( a \), we could follow the steps above with \( a \) and \( b \) swapped and end up concluding that \( b \pi_1 + (1 - b) \pi_2 \succ a \pi_1 + (1 - a) \pi_2 \), which would contradict our assumption. Hence, we cannot have neither \( a = b \) nor \( a < b \) but must have \( a > b \).

Next we state the main result:

**Theorem 2.2.** Assume that \( Z \) is finite and that \( \succeq \) is a preference relation on \( \mathcal{P}(Z) \). Then \( \succeq \) can be represented by a linear utility index if and only if \( \succeq \) satisfies the Archimedean Axiom and the Substitution Axiom.

**Proof.** First suppose the preference relation \( \succeq \) satisfies the Archimedean Axiom and the Substitution Axiom. Define a utility index \( U : \mathcal{P}(Z) \rightarrow \mathbb{R} \) exactly as in the proof of Theorem 2.1, i.e.,
\[
U(\pi) = \alpha_\pi, \quad \text{where } \alpha_\pi \in [0, 1] \text{ is the unique number such that }
\]
\[
\pi \sim \alpha_\pi \mathbf{1}_{z^*} + (1 - \alpha_\pi) \mathbf{1}_{x^*}.
\]
We want to show that, as a consequence of the Substitution Axiom, \( U \) is indeed linear. For that purpose, pick any two probability distributions \( \pi_1, \pi_2 \in \mathcal{P}(Z) \) and any number \( a \in [0, 1] \). We want to show that \( U(a \pi_1 + (1 - a) \pi_2) = aU(\pi_1) + (1 - a)U(\pi_2) \). We can do that by showing that
\[
a \pi_1 + (1 - a) \pi_2 \sim (aU(\pi_1) + (1 - a)U(\pi_2)) \mathbf{1}_{z^*} + (1 - \{aU(\pi_1) + (1 - a)U(\pi_2)\}) \mathbf{1}_{x^*}.
\]
This follows from the Substitution Axiom:
\[
a \pi_1 + (1 - a) \pi_2 \sim a\{U(\pi_1) \mathbf{1}_{z^*} + (1 - U(\pi_1)) \mathbf{1}_{x^*}\} + (1 - a)\{U(\pi_2) \mathbf{1}_{z^*} + (1 - U(\pi_2)) \mathbf{1}_{x^*}\}
\]
\[
\sim (aU(\pi_1) + (1 - a)U(\pi_2)) \mathbf{1}_{z^*} + (1 - \{aU(\pi_1) + (1 - a)U(\pi_2)\}) \mathbf{1}_{x^*}.
\]

Now let us show the converse, i.e., if \( \succeq \) can be represented by a linear utility index \( U \), then it must satisfy the Archimedean Axiom and the Substitution Axiom. In order to show the Archimedean
Axiom, we pick \( \pi_1 \succ \pi_2 \succ \pi_3 \), which means that \( U(\pi_1) > U(\pi_2) > U(\pi_3) \), and must find numbers \( a, b \in (0,1) \) such that
\[
a \pi_1 + (1 - a) \pi_3 > \pi_2 > b \pi_1 + (1 - b) \pi_3,
\]
i.e., that
\[
U(a \pi_1 + (1 - a) \pi_3) > U(\pi_2) > U(b \pi_1 + (1 - b) \pi_3).
\]
Define the number \( a \) by
\[
a = 1 - \frac{1}{2} \frac{U(\pi_1) - U(\pi_2)}{U(\pi_1) - U(\pi_3)}.
\]
Then \( a \in (0, 1) \) and by linearity of \( U \) we get
\[
U(a \pi_1 + (1 - a) \pi_3) = aU(\pi_1) + (1 - a)U(\pi_3)
= U(\pi_1) + (1 - a)(U(\pi_3) - U(\pi_1))
= U(\pi_1) - \frac{1}{2}(U(\pi_1) - U(\pi_2))
= \frac{1}{2}(U(\pi_1) + U(\pi_2))
> U(\pi_2).
\]

Similarly for \( b \).

In order to show the Substitution Axiom, we take \( \pi_1, \pi_2, \pi_3 \in \mathcal{P}(Z) \) and any number \( a \in (0, 1] \).
We must show that \( \pi_1 \succ \pi_2 \) if and only if \( a \pi_1 + (1 - a) \pi_3 > a \pi_2 + (1 - a) \pi_3 \), i.e.,
\[
U(\pi_1) > U(\pi_2) \iff U(a \pi_1 + (1 - a) \pi_3) > U(a \pi_2 + (1 - a) \pi_3).
\]
This follows immediately by linearity of \( U \):
\[
U(a \pi_1 + (1 - a) \pi_3) = aU(\pi_1) + U((1 - a) \pi_3)
> aU(\pi_2) + U((1 - a) \pi_3)
= U(a \pi_2 + (1 - a) \pi_3)
\]
with the inequality holding if and only if \( U(\pi_1) > U(\pi_2) \). Similarly, we can show that \( \pi_1 \sim \pi_2 \) if and only if \( a \pi_1 + (1 - a) \pi_3 \sim a \pi_2 + (1 - a) \pi_3 \).

The next theorem shows which utility functions that represent the same preference relation. The proof is left for the reader as Exercise 2.1.

**Theorem 2.3.** A utility function for a given preference relation is only determined up to a strictly increasing affine transformation, i.e., if \( u \) is a utility function for \( \succeq \), then \( v \) will be so if and only if there exist constants \( a > 0 \) and \( b \) such that \( v(z) = au(z) + b \) for all \( z \in Z \).

If one utility function is an affine function of another, we will say that they are equivalent. Note that an easy consequence of this theorem is that it does not really matter whether the utility is positive or negative. At first, you might find negative utility strange but we can always add a sufficiently large positive constant without affecting the ranking of different consumption plans.

Suppose \( U \) is a utility index with an associated utility function \( u \). If \( f \) is any strictly increasing transformation, then \( V = f \circ U \) is also a utility index for the same preferences, but \( f \circ u \) is only the utility function for \( V \) if \( f \) is affine.
The expected utility associated with a probability distribution \(\pi\) on \(\mathbb{Z}\) is \(\sum_{z \in \mathbb{Z}} \pi(z)u(z)\). Recall that the probability distributions we consider correspond to consumption plans. Given a consumption plan, i.e., a random variable \(c\), the associated probability distribution is defined by the probabilities 

\[ \pi(z) = \mathbb{P}\{\omega \in \Omega|c(\omega) = z\} = \sum_{\omega \in \Omega: c(\omega) = z} p_{\omega}. \]

The expected utility associated with the consumption plan \(c\) is therefore 

\[ E[u(c)] = \sum_{\omega \in \Omega} p_{\omega} u(c(\omega)) = \sum_{z \in \mathbb{Z}} \sum_{\omega \in \Omega: c(\omega) = z} p_{\omega} u(z) = \sum_{z \in \mathbb{Z}} \pi(z)u(z). \]

Of course, if \(c\) is a risk-free consumption plan in the sense that a \(z\) exists such that \(c(\omega) = z\) for all \(\omega\), then the expected utility is \(E[u(c)] = u(z)\). With a slight abuse of notation we will just write this as \(u(c)\).

2.4.2 Some technical issues

**Infinite \(\mathbb{Z}\).** What if \(\mathbb{Z}\) is infinite, e.g., \(\mathbb{Z} = \mathbb{R}_+ = [0, \infty)\)? It can be shown that in this case a preference relation has an expected utility representation if the Archimedean Axiom, the Substitution Axiom, an additional axiom (“the sure thing principle”), and “some technical conditions” are satisfied. Fishburn (1970) gives the details.

Expected utility in this case: \(E[u(c)] = \int_{\mathbb{Z}} u(z)\pi(z) \, dz\), where \(\pi\) is a probability density function derived from the consumption plan \(c\).

**Boundedness of expected utility.** Suppose \(u\) is unbounded from above and \(\mathbb{R}_+ \subseteq \mathbb{Z}\). Then there exists \((z_n)_{n=1}^\infty \subseteq \mathbb{Z}\) with \(z_n \to \infty\) and \(u(z_n) \geq 2^n\). Expected utility of consumption plan \(\pi_1\) with \(\pi_1(z_n) = 1/2^n\): 

\[ \sum_{n=1}^{\infty} u(z_n)\pi_1(z_n) \geq \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = \infty. \]

If \(\pi_2, \pi_3\) are such that \(\pi_1 \succ \pi_2 \succ \pi_3\), then the expected utility of \(\pi_2\) and \(\pi_3\) must be finite. But for no \(b \in (0, 1)\) do we have

\[ \pi_2 \succ b\pi_1 + (1 - b)\pi_3 \quad [\text{expected utility} = \infty]. \]

- no problem if \(\mathbb{Z}\) is finite
- no problem if \(\mathbb{R}_+ \subseteq \mathbb{Z}\), \(u\) is concave, and consumption plans have finite expectations:

\[ u \text{ concave} \Rightarrow u \text{ is differentiable in some point } b \text{ and} \]

\[ u(z) \leq u(b) + u'(b)(z - b), \quad \forall z \in \mathbb{Z}. \]

If the consumption plan \(c\) has finite expectations, then

\[ E[u(c)] \leq E[u(b) + u'(b)(c - b)] = u(b) + u'(b) (E[c] - b) < \infty. \]
Subjective probability. We have taken the probabilities of the states of nature as exogenously given, i.e., as objective probabilities. However, in real life individuals often have to form their own probabilities about many events, i.e., they form subjective probabilities. Although the analysis is a bit more complicated, Savage (1954) and Anscombe and Aumann (1963) show that the results we developed above carry over to the case of subjective probabilities. For an introduction to this analysis, see Kreps (1990, Ch. 3).

### 2.4.3 Are the axioms reasonable?

The validity of the Substitution Axiom, which is necessary for obtaining the expected utility representation, has been intensively discussed in the literature. Some researchers have conducted experiments in which the decisions made by the participating individuals conflict with the Substitution Axiom.

The most famous challenge is the so-called Allais Paradox named after Allais (1953). Here is one example of the paradox. Suppose \( Z = \{0, 1, 5\} \). Consider the consumption plans in Table 2.4. The Substitution Axiom implies that \( \pi_1 \succ \pi_2 \Rightarrow \pi_4 \succ \pi_3 \). This can be seen from the following:

\[
0.11(\$1) + 0.89(\$1) \sim \pi_1 \succ \pi_2 \sim 0.11 \left( \frac{1}{11}(\$0) + \frac{10}{11}(\$5) \right) + 0.89(\$1) \Rightarrow
\]

\[
\pi_2 \sim 0.11 \left( \frac{1}{11}(\$0) + \frac{10}{11}(\$5) \right) + 0.89(\$0) \sim 0.9(\$0) + 0.1(\$5) \pi_3 \sim
\]

Nevertheless individuals preferring \( \pi_1 \) to \( \pi_2 \) often choose \( \pi_3 \) over \( \pi_4 \). Apparently people tend to over-weight small probability events, e.g., \( \$0 \) in \( \pi_2 \).

Other “problems”:

- the “framing” of possible choices, i.e., the way you get the alternatives presented, seem to affect decisions
- models assume individuals have unlimited rationality

### 2.5 Risk aversion

In this section we focus on the attitudes towards risk reflected by the preferences of an individual. We assume that the preferences can be represented by a utility function \( u \) and that \( u \) is strictly increasing so that the individual is “greedy,” i.e., prefers high consumption to low consumption. We assume that the utility function is defined on some interval \( Z \) of \( \mathbb{R} \), e.g., \( Z = \mathbb{R}_+ \equiv [0, \infty) \).

<table>
<thead>
<tr>
<th>( z )</th>
<th>( 0 )</th>
<th>( 1 )</th>
<th>( 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_1 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( \pi_2 )</td>
<td>0.01</td>
<td>0.89</td>
<td>0.1</td>
</tr>
<tr>
<td>( \pi_3 )</td>
<td>0.9</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>( \pi_4 )</td>
<td>0.89</td>
<td>0.11</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2.4: The probability distributions used in the illustration of the Allais Paradox.
2.5 Risk aversion

2.5.1 Risk attitudes

Fix a consumption level $c \in \mathbb{Z}$. Consider a random variable $\varepsilon$ with $E[\varepsilon] = 0$. We can think of $c + \varepsilon$ as a random variable representing a consumption plan with consumption $c + \varepsilon(\omega)$ if state $\omega$ is realized. Note that $E[c + \varepsilon] = c$. Such a random variable $\varepsilon$ is called a fair gamble or a zero-mean risk.

An individual is said to be (strictly) risk-averse if she for all $c \in \mathbb{Z}$ and all fair gambles $\varepsilon$ (strictly) prefers the sure consumption level $c$ to $c + \varepsilon$. In other words, a risk-averse individual rejects all fair gambles. Similarly, an individual is said to be (strictly) risk-loving if she for all $c \in \mathbb{Z}$ (strictly) prefers $c + \varepsilon$ to $c$, and said to be risk-neutral if she for all $c \in \mathbb{Z}$ is indifferent between accepting any fair gamble or not. Of course, individuals may be neither risk-averse, risk-neutral, or risk-loving, for example if they reject fair gambles around some values of $c$ and accept fair gambles around other values of $c$. Individuals may be locally risk-averse, locally risk-neutral, and locally risk-loving. Since it is generally believed that individuals are risk-averse, we focus on preferences exhibiting that feature.

We can think of any consumption plan $c$ as the sum of its expected value $E[c]$ and a fair gamble $\varepsilon = c - E[c]$. It follows that an individual is risk-averse if she prefers the sure consumption $E[c]$ to the random consumption $c$, i.e., if $u(E[c]) \geq E[u(c)]$. By Jensen’s Inequality, this is true exactly when $u$ is a concave function and the strict inequality holds if $u$ is strictly concave and $c$ is a non-degenerate random variable, i.e., it does not have the same value in all states. Recall that $u : \mathbb{Z} \to \mathbb{R}$ concave means that for all $z_1, z_2 \in \mathbb{Z}$ and all $a \in (0, 1)$ we have

$$u(az_1 + (1-a)z_2) \geq au(z_1) + (1-a)u(z_2).$$

If the strict inequality holds in all cases, the function is said to be strictly concave. By the above argument, we have the following theorem:

**Theorem 2.4.** An individual with a utility function $u$ is (strictly) risk-averse if and only if $u$ is (strictly) concave.

Similarly, an individual is (strictly) risk-loving if and only if the utility function is (strictly) convex. An individual is risk-neutral if and only if the utility function is affine.

2.5.2 Quantitative measures of risk aversion

We will focus on utility functions that are continuous and twice differentiable on the interior of $\mathbb{Z}$. By our assumption of greedy individuals, we then have $u' > 0$, and the concavity of the utility function for risk-averse investors is then equivalent to $u'' \leq 0$.

The certainty equivalent of the random consumption plan $c$ is defined as the $c^* \in \mathbb{Z}$ such that

$$u(c^*) = E[u(c)],$$

i.e., the individual is just as satisfied getting the consumption level $c^*$ for sure as getting the random consumption $c$. With $\mathbb{Z} \subseteq \mathbb{R}$, $c^*$ uniquely exists due to our assumptions that $u$ is continuous and strictly increasing. From the definition of the certainty equivalent it is clear that an individual will rank consumption plans according to their certainty equivalents.
For a risk-averse individual we have the certainty equivalent \( c^* \) of a consumption plan is smaller than the expected consumption level \( E[c] \). The **risk premium** associated with the consumption plan \( c \) is defined as \( \lambda(c) = E[c] - c^* \) so that

\[
E[u(c)] = u(c^*) = u(E[c] - \lambda(c)).
\]

The risk premium is the consumption the individual is willing to give up in order to eliminate the uncertainty.

The degree of risk aversion is associated with \( u'' \), but a good measure of risk aversion should be invariant to strictly positive, affine transformations. This is satisfied by the Arrow-Pratt measures of risk aversion defined as follows. The **Absolute Risk Aversion** is given by

\[
\text{ARA}(c) = -\frac{u''(c)}{u'(c)}.
\]

The **Relative Risk Aversion** is given by

\[
\text{RRA}(c) = -\frac{cu''(c)}{u'(c)} = c \text{ARA}(c).
\]

We can link the Arrow-Pratt measures to the risk premium in the following way. Let \( \bar{c} \in Z \) denote some fixed consumption level and let \( \varepsilon \) be a fair gamble. The resulting consumption plan is then \( c = \bar{c} + \varepsilon \). Denote the corresponding risk premium by \( \lambda(\bar{c}, \varepsilon) \) so that

\[
E[u(\bar{c} + \varepsilon)] = u(c^*) = u(\bar{c} - \lambda(\bar{c}, \varepsilon)). \tag{2.2}
\]

We can approximate the left-hand side of (2.2) by

\[
E[u(\bar{c} + \varepsilon)] \approx E \left[ u(\bar{c}) + \varepsilon u'(\bar{c}) + \frac{1}{2} \varepsilon^2 u''(\bar{c}) \right] = u(\bar{c}) + \frac{1}{2} \text{Var}[\varepsilon] u''(\bar{c}),
\]

using \( E[\varepsilon] = 0 \) and \( \text{Var}[\varepsilon] = E[\varepsilon^2] - E[\varepsilon]^2 = E[\varepsilon^2] \), and we can approximate the right-hand side of (2.2) by

\[
u(\bar{c} - \lambda(\bar{c}, \varepsilon)) \approx u(\bar{c}) - \lambda(\bar{c}, \varepsilon) u'(\bar{c}).
\]

Hence we can write the risk premium as

\[
\lambda(\bar{c}, \varepsilon) \approx -\frac{1}{2} \frac{u''(\bar{c})}{u'(\bar{c})} = \frac{1}{2} \text{Var}[\varepsilon] \text{ARA}(\bar{c}).
\]

Of course, the approximation is more accurate for “small” gambles. Thus the risk premium for a small fair gamble around \( \bar{c} \) is roughly proportional to the absolute risk aversion at \( \bar{c} \). We see that the absolute risk aversion \( \text{ARA}(\bar{c}) \) is constant if and only if \( \lambda(\bar{c}, \varepsilon) \) is independent of \( \bar{c} \).

Loosely speaking, the absolute risk aversion \( \text{ARA}(\bar{c}) \) measures the aversion to a fair gamble of a given dollar amount around \( \bar{c} \), such as a gamble where there is an equal probability of winning or losing 1000 dollars. Since we expect that a wealthy investor will be less averse to that gamble than a poor investor, the absolute risk aversion is expected to be a decreasing function of wealth. Note that

\[
\text{ARA}'(c) = -\frac{u''(c)u'(c) - u''(c)^2}{u'(c)^2} = \left( \frac{u''(c)}{u'(c)} \right)^2 - \frac{u'''(c)}{u'(c)} < 0 \implies u'''(c) > 0,
\]

that is, a positive third-order derivative of \( u \) is necessary for the utility function \( u \) to exhibit decreasing absolute risk aversion.
2.5 Risk aversion

Now consider a “multiplicative” fair gamble around \( \hat{c} \) in the sense that the resulting consumption plan is \( c = \hat{c} (1 + \varepsilon) = \hat{c} + \varepsilon \hat{c} \), where \( E[\varepsilon] = 0 \). The risk premium is then

\[
\lambda(\varepsilon, \hat{c} \varepsilon) \approx \frac{1}{2} \text{Var}[\varepsilon] \text{ARA}(\hat{c}) = \frac{1}{2} \hat{c}^2 \text{Var}[\varepsilon] \text{ARA}(\hat{c}) = \frac{1}{2} \hat{c} \text{Var}[\varepsilon] \text{RRA}(\hat{c})
\]

implying that

\[
\frac{\lambda(\varepsilon, \hat{c} \varepsilon)}{\varepsilon} \approx \frac{1}{2} \text{Var}[\varepsilon] \text{RRA}(\hat{c}). \tag{2.3}
\]

The fraction of consumption you require to engage in the multiplicative risk is thus (roughly) proportional to the relative risk aversion at \( \hat{c} \). Note that utility functions with constant or decreasing (or even modestly increasing) relative risk aversion will display decreasing absolute risk aversion.

Some authors use terms like risk tolerance and risk cautiousness. The **absolute risk tolerance** at \( c \) is simply the reciprocal of the absolute risk aversion, i.e.,

\[
\text{ART}(c) = \frac{1}{\text{ARA}(c)} = -\frac{u'(c)}{u''(c)}.
\]

Similarly, the relative risk tolerance is the reciprocal of the relative risk aversion. The **risk cautiousness** at \( c \) is defined as the rate of change in the absolute risk tolerance, i.e., \( \text{ART}^*(c) \).

2.5.3 Comparison of risk aversion between individuals

An individual with utility function \( u \) is said to be more risk-averse than an individual with utility function \( v \) if for any consumption plan \( c \) and any fixed \( \hat{c} \in Z \) with \( E[u(c)] \geq u(\hat{c}) \), we have \( E[v(c)] \geq v(\hat{c}) \). So the \( v \)-individual will accept all gambles that the \( u \)-individual will accept – and possibly some more. Pratt (1964) has shown the following theorem:

**Theorem 2.5.** Suppose \( u \) and \( v \) are twice continuously differentiable and strictly increasing. Then the following conditions are equivalent:

(a) \( u \) is more risk-averse than \( v \),

(b) \( \text{ARA}_u(c) \geq \text{ARA}_v(c) \) for all \( c \in Z \),

(c) a strictly increasing and concave function \( f \) exists such that \( u = f \circ v \).

**Proof.** First let us show (a) \(\Rightarrow\) (b): Suppose \( u \) is more risk-averse than \( v \), but that \( \text{ARA}_u(\hat{c}) < \text{ARA}_v(\hat{c}) \) for some \( \hat{c} \in Z \). Since \( \text{ARA}_u \) and \( \text{ARA}_v \) are continuous, we must then have that \( \text{ARA}_u(c) < \text{ARA}_v(c) \) for all \( c \) in an interval around \( \hat{c} \). Then we can surely find a small gamble around \( \hat{c} \), which the \( u \)-individual will accept, but the \( v \)-individual will reject. This contradicts the assumption in (a).

Next, we show (b) \(\Rightarrow\) (c): Since \( v \) is strictly increasing, it has an inverse \( v^{-1} \) and we can define a function \( f \) by \( f(x) = u(v^{-1}(x)) \). Then clearly \( f(v(c)) = u(c) \) so that \( u = f \circ v \). The first-order derivative of \( f \) is

\[
f'(x) = \frac{u'(v^{-1}(x))}{v'(v^{-1}(x))},
\]

which is positive since \( u \) and \( v \) are strictly increasing. Hence, \( f \) is strictly increasing. The second-order derivative is

\[
f''(x) = \frac{u''(v^{-1}(x)) - \{v''(v^{-1}(x))u'(v^{-1}(x))\}/v'(v^{-1}(x))}{v'(v^{-1}(x))^2} = \frac{u'(v^{-1}(x))}{v'(v^{-1}(x))^2} \left( \text{ARA}_v(v^{-1}(x)) - \text{ARA}_u(v^{-1}(x)) \right).
\]
From (b), it follows that $f''(x) < 0$, hence $f$ is concave.

Finally, we show that (c) $\Rightarrow$ (a): assume that for some consumption plan $c$ and some $\bar{c} \in Z$, we have $E[u(c)] \geq u(\bar{c})$ but $E[v(c)] < v(\bar{c})$. We want to arrive at a contradiction.

$$f(v(\bar{c})) = u(\bar{c}) \leq E[u(c)] = E[f(v(c))]$$

$$< f(E[v(c)])$$

$$< f(v(\bar{c})),$$

where we use the concavity of $f$ and Jensen’s Inequality to go from the first to the second line, and we use that $f$ is strictly increasing to go from the second to the third line. Now the contradiction is clear. 

### 2.6 Utility functions in models and in reality

#### 2.6.1 Frequently applied utility functions

**CRRA utility.** (Also known as power utility or isoelastic utility.) Utility functions $u(c)$ in this class are defined for $c \geq 0$:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}, \quad (2.4)$$

where $\gamma > 0$ and $\gamma \neq 1$. Since

$$u'(c) = c^{-\gamma} \quad \text{and} \quad u''(c) = -\gamma c^{-\gamma-1},$$

the absolute and relative risk aversions are given by

$$\text{ARA}(c) = \frac{u''(c)}{u'(c)} = \frac{\gamma}{c}, \quad \text{RRA}(c) = c \text{ARA}(c) = \gamma.$$  

The relative risk aversion is constant across consumption levels $c$, hence the name CRRA (Constant Relative Risk Aversion) utility. Note that $u'(0+) \equiv \lim_{c \to 0} u'(c) = \infty$ with the consequence that an optimal solution will have the property that consumption/wealth $c$ will be strictly above 0 with probability one. Hence, we can ignore the very appropriate non-negativity constraint on consumption since the constraint will never be binding. Furthermore, $u'(\infty) \equiv \lim_{c \to \infty} u'(c) = 0$.

Some authors assume a utility function of the form $u(c) = c^{1-\gamma}$, which only makes sense for $\gamma \in (0, 1)$. However, empirical studies indicate that most investors have a relative risk aversion above 1, cf. the discussion below. The absolute risk tolerance is linear in $c$:

$$\text{ART}(c) = \frac{1}{\text{ARA}(c)} = \frac{c}{\gamma}.$$  

Except for a constant, the utility function

$$u(c) \equiv \frac{c^{1-\gamma} - 1}{1-\gamma},$$

is identical to the utility function specified in (2.4). The two utility functions are therefore equivalent in the sense that they generate identical rankings of consumption plans and, in particular, identical optimal choices. The advantage in using the latter definition is that this function has a well-defined limit as $\gamma \to 1$. From l'Hôpital’s rule we have that

$$\lim_{\gamma \to 1} \frac{c^{1-\gamma} - 1}{1-\gamma} = \lim_{\gamma \to 1} \frac{-c^{1-\gamma} \ln c}{-1} = \ln c,$$
which is the important special case of logarithmic utility. When we consider CRRA utility, we will assume the simpler version (2.4), but we will use the fact that we can obtain the optimal strategies of a log-utility investor as the limit of the optimal strategies of the general CRRA investor as $\gamma \to 1$.

Some CRRA utility functions are illustrated in Figure 2.1.

**HARA utility.** (Also known as extended power utility.) The absolute risk aversion for CRRA utility is hyperbolic in $c$. More generally a utility function is said to be a HARA (Hyperbolic Absolute Risk Aversion) utility function if

$$\text{ARA}(c) = \frac{-u''(c)}{u'(c)} = \frac{1}{\alpha c + \beta}$$

for some constants $\alpha, \beta$ such that $\alpha c + \beta > 0$ for all relevant $c$. HARA utility functions are sometimes referred to as affine (or linear) risk tolerance utility functions since the absolute risk tolerance is

$$\text{ART}(c) = \frac{1}{\text{ARA}(c)} = \alpha c + \beta.$$ 

The risk cautiousness is $\text{ART}'(c) = \alpha$.

How do the HARA utility functions look like? First, let us take the case $\alpha = 0$, which implies that the absolute risk aversion is constant (so-called CARA utility) and $\beta$ must be positive.

$$\frac{d(\ln u'(c))}{dc} = \frac{u''(c)}{u'(c)} = -\frac{1}{\beta}$$

implies that

$$\ln u'(c) = -\frac{c}{\beta} + k_1 \Rightarrow u'(c) = e^{k_1} e^{-c/\beta}$$
for some constant $k_1$. Hence,
\[ u(c) = -\frac{1}{\beta} e^{k_1} e^{-c/\beta} + k_2 \]
for some other constant $k_2$. Applying the fact that increasing affine transformations do not change decisions, the basic representative of this class of utility functions is the negative exponential utility function
\[ u(c) = -e^{-ac}, \quad c \in \mathbb{R}, \]
where the parameter $a = 1/\beta$ is the absolute risk aversion. Constant absolute risk aversion is certainly not very reasonable. Nevertheless, the negative exponential utility function is sometimes used for computational purposes in connection with normally distributed returns, e.g., in one-period models.

Next, consider the case $\alpha \neq 0$. Applying the same procedure as above we find
\[ \frac{d(\ln u'(c))}{dc} = \frac{u''(c)}{u'(c)} = -\frac{1}{\alpha c + \beta} \Rightarrow \ln u'(c) = -\frac{1}{\alpha} \ln(\alpha c + \beta) + k_1 \]
so that
\[ u'(c) = e^{k_1} \exp \left\{ -\frac{1}{\alpha} \ln(\alpha c + \beta) \right\} = e^{k_1} (\alpha c + \beta)^{-1/\alpha}. \quad (2.5) \]
For $\alpha = 1$ this implies that
\[ u(c) = e^{k_1} \ln(c + \beta) + k_2. \]
The basic representative of such utility functions is the extended log utility function
\[ u(c) = \ln(c - \bar{c}), \quad c > \bar{c}, \]
where we have replaced $\beta$ by $-\bar{c}$. For $\alpha \neq 1$, Equation (2.5) implies that
\[ u(c) = \frac{1}{\alpha} e^{k_1} \frac{1}{1 - \frac{1}{\alpha}} (\alpha c + \beta)^{1-1/\alpha} + k_2. \]

For $\alpha < 0$, we can write the basic representative is
\[ u(c) = -\left(\bar{c} - c\right)^{1-\gamma}, \quad c < \bar{c}, \]
where $\gamma = 1/\alpha < 0$. We can think of $\bar{c}$ as a satiation level and call this subclass satiation HARA utility functions. The absolute risk aversion is
\[ \text{ARA}(c) = \frac{-\gamma}{\bar{c} - c}, \]
which is increasing in $c$, conflicting with intuition and empirical studies. Some older financial models used the quadratic utility function, which is the special case with $\gamma = -1$ so that $u(c) = -(\bar{c} - c)^2$. An equivalent utility function is $u(c) = c - ac^2$.

For $\alpha > 0$ (and $\alpha \neq 1$), the basic representative is
\[ u(c) = \frac{(c - \bar{c})^{1-\gamma}}{1 - \gamma}, \quad c > \bar{c}, \]
where $\gamma = 1/\alpha > 0$. The limit as $\gamma \to 1$ of the equivalent utility function $\frac{(c-\bar{c})^{1-\gamma}}{1 - \gamma}$ is equal to the extended log utility function $u(c) = \ln(c - \bar{c})$. We can think of $\bar{c}$ as a subsistence level of wealth or
consumption (which makes sense only if \( \bar{c} \geq 0 \)) and refer to this subclass as **subsistence HARA utility** functions. The absolute and relative risk aversions are

\[
\text{ARA}(c) = \frac{\gamma}{c - \bar{c}}, \quad \text{RRA}(c) = \frac{\gamma c}{c - \bar{c}} = \frac{\gamma}{1 - (\bar{c}/c)},
\]

which are both decreasing in \( c \). The relative risk aversion approaches \( \infty \) for \( c \to \bar{c} \) and decreases to the constant \( \gamma \) for \( c \to \infty \). Clearly, for \( \bar{c} = 0 \), we are back to the CRRA utility functions so that these also belong to the HARA family.

### Mean-variance preferences.

For some problems it is convenient to assume that the expected utility associated with an uncertain consumption plan only depends on the expected value and the variance of the consumption plan. This is certainly true if the consumption plan is a normally distributed random variable since its probability distribution is fully characterized by the mean and variance. However, it is generally not appropriate to use a normal distribution for consumption (or wealth or asset returns).

For a quadratic utility function, \( u(c) = c - ac^2 \), the expected utility is

\[
\]

which is indeed a function of the expected value and the variance of the consumption plan. Alas, the quadratic utility function is inappropriate for several reasons. Most importantly, it exhibits increasing absolute risk aversion.

For a general utility function the expected utility of a consumption plan will depend on all moments. This can be seen by the Taylor expansion of \( u(c) \) around the expected consumption, \( E[c] \):

\[
u(c) = u(E[c]) + u'(E[c])(c - E[c]) + \frac{1}{2} u''(E[c])(c - E[c])^2 + \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[c])(c - E[c])^n,
\]

where \( u^{(n)} \) is the \( n \)'th derivative of \( u \). Taking expectations, we get

\[
E[u(c)] = u(E[c]) + \frac{1}{2} u''(E[c]) \text{Var}[c] + \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[c]) E[(c - E[c])^n].
\]

Here \( E[(c - E[c])^n] \) is the central moment of order \( n \). The variance is the central moment of order 2. Obviously, a greedy investor (which just means that \( u \) is increasing) will prefer higher expected consumption to lower for fixed central moments of order 2 and higher. Moreover, a risk-averse investor (so that \( u'' < 0 \)) will prefer lower variance of consumption to higher for fixed expected consumption and fixed central moments of order 3 and higher. But when the central moments of order 3 and higher are not the same for all alternatives, we cannot just evaluate them on the basis of their expectation and variance. With quadratic utility, the derivatives of \( u \) of order 3 and higher are zero so there it works. In general, mean-variance preferences can only serve as an approximation of the true utility function.

### 2.6.2 What do we know about individuals’ risk aversion?

From our discussion of risk aversion and various utility functions we expect that individuals are risk averse and exhibit decreasing absolute risk aversion. But can this be supported by empirical
evidence? Do individuals have constant relative risk aversion? And what is a reasonable level of risk aversion for individuals?

You can get an idea of the risk attitudes of an individual by observing how they choose between risky alternatives. Some researchers have studied this by setting up “laboratory experiments” in which they present some risky alternatives to a group of individuals and simply see what they prefer. Some of these experiments suggest that expected utility theory is frequently violated, see e.g., Grether and Plott (1979). However, laboratory experiments are problematic for several reasons. You cannot be sure that individuals will make the same choice in what they know is an experiment as they would in real life. It is also hard to formulate alternatives that resemble the rather complex real-life decisions. It seems more fruitful to study actual data on how individuals have acted confronted with real-life decision problems under uncertainty. A number of studies do that.

Friend and Blume (1975) analyze data on household asset holdings. They conclude that the data is consistent with individuals having roughly constant relative risk aversion and that the coefficients of relative risk aversion are “on average well in excess of one and probably in excess of two” (quote from page 900 in their paper). Pindyck (1988) finds support of a relative risk aversion between 3 and 4 in a structural model of the reaction of stock prices to fundamental variables.

Other studies are based on insurance data. Using U.S. data on so-called property/liability insurance, Szpiro (1986) finds support of CRRA utility with a relative risk aversion coefficient between 1.2 and 1.8. Cicchetti and Dubin (1994) work with data from the U.S. on whether individuals purchased an insurance against the risk of trouble with their home telephone line. They conclude that the data is consistent with expected utility theory and that a subsistence HARA utility function performs better than log utility or negative exponential utility.

Ogaki and Zhang (2001) study data on individual food consumption from Pakistan and India and conclude that relative risk aversion is decreasing for poor individuals, which is consistent with a subsistence HARA utility function.

It is an empirical fact that even though consumption and wealth have increased tremendously over the years, the magnitude of real rates of return has not changed dramatically. As indicated by (2.3) relative risk premia are approximately proportional to the relative risk aversion. As discussed in, e.g., Munk (2012), basic asset pricing theory implies that relative risk premia on financial assets (in terms of expected real return in excess of the real risk-free return) will be proportional to the “average” relative risk aversion in the economy. If the “average” relative risk aversion was significantly decreasing (increasing) in the level of consumption or wealth, we should have seen decreasing (increasing) real returns on risky assets in the past. The data seems to be consistent with individuals having “on average” close to CRRA utility.

To get a feeling of what a given risk aversion really means, suppose you are confronted with two consumption plans. One plan is a sure consumption of \( \bar{c} \), the other plan gives you \( (1 - \alpha)\bar{c} \) with probability 0.5 and \( (1 + \alpha)\bar{c} \) with probability 0.5. If you have a CRRA utility function \( u(c) = c^{1 - \gamma} / (1 - \gamma) \), the certainty equivalent \( c^* \) of the risky plan is determined by

\[
\frac{1}{1 - \gamma} (c^*)^{1 - \gamma} = \frac{1}{2} \frac{1}{1 - \gamma} ((1 - \alpha)\bar{c})^{1 - \gamma} + \frac{1}{2} \frac{1}{1 - \gamma} ((1 + \alpha)\bar{c})^{1 - \gamma},
\]
\[
\gamma = \text{RRA} \quad \alpha = 1\% \quad \alpha = 10\% \quad \alpha = 50\%
\]

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00%</td>
<td>0.25%</td>
<td>6.70%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01%</td>
<td>0.50%</td>
<td>13.40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01%</td>
<td>1.00%</td>
<td>25.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02%</td>
<td>2.43%</td>
<td>40.72%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05%</td>
<td>4.42%</td>
<td>46.00%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10%</td>
<td>6.76%</td>
<td>48.14%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02%</td>
<td>8.72%</td>
<td>49.29%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05%</td>
<td>9.37%</td>
<td>49.65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5: Relative risk premia for a fair gamble of the fraction \( \alpha \) of your consumption.

which implies that

\[
c^* = \left( \frac{1}{2} \right)^{1/(1-\gamma)} \left[ (1-\alpha)^{1-\gamma} + (1+\alpha)^{1-\gamma} \right]^{1/(1-\gamma)} \bar{c}.
\]

The risk premium \( \lambda(\bar{c}, \alpha) \) is

\[
\lambda(\bar{c}, \alpha) = \bar{c} - c^* = \left( 1 - \left( \frac{1}{2} \right)^{1/(1-\gamma)} \left[ (1-\alpha)^{1-\gamma} + (1+\alpha)^{1-\gamma} \right]^{1/(1-\gamma)} \right) \bar{c}.
\]

Both the certainty equivalent and the risk premium are thus proportional to the consumption level \( \bar{c} \). The relative risk premium \( \lambda(\bar{c}, \alpha)/\bar{c} \) is simply one minus the relative certainty equivalent \( c^*/\bar{c} \). These equations assume \( \gamma \neq 1 \). In Exercise 2.5 you are asked to find the certainty equivalent and risk premium for log-utility corresponding to \( \gamma = 1 \).

Table 2.5 shows the relative risk premium for various values of the relative risk aversion coefficient \( \gamma \) and various values of \( \alpha \), the “size” of the risk. For example, an individual with \( \gamma = 5 \) is willing to sacrifice 2.43\% of the safe consumption in order to avoid a fair gamble of 10\% of that consumption level. Of course, even extremely risk averse individuals will not sacrifice more than they can loose but in some cases it is pretty close. Looking at these numbers, it is hard to believe in \( \gamma \)-values outside, say, \([1, 10] \). In Exercise 2.6 you are asked to compare the exact relative risk premia shown in the table with the approximate risk premia given by (2.3).

2.6.3 Two-good utility functions and the elasticity of substitution

Consider an atemporal utility function \( f(c, z) \) of two consumption of two different goods at the same time. An indifference curve in the \((c, z)\)-space is characterized by \( f(c, z) = k \) for some constant \( k \). Changes in \( c \) and \( z \) along an indifference curve are linked by

\[
\frac{\partial f}{\partial c} dc + \frac{\partial f}{\partial z} dz = 0
\]

so that the slope of the indifference curve (also known as the marginal rate of substitution) is

\[
\frac{dz}{dc} = -\frac{\frac{\partial f}{\partial c}}{\frac{\partial f}{\partial z}}.
\]
Unless the indifference curve is linear, its slope will change along the curve. Indifference curves are generally assumed to be convex. The elasticity of substitution tells you by which percentage you need to change \( \frac{z}{c} \) in order to obtain a one percent change in the slope of the indifference curve. It is a measure of the curvature or convexity of the indifference curve. If the indifference curve is very curved, you only have to move a little along the curve before its slope has changed by one percent. Hence, the elasticity of substitution is low. If the indifference curve is almost linear, you have to move far away to change the slope by one percent. In that case the elasticity of substitution is very high. Formally, the elasticity of substitution is defined as

\[
\psi = - \frac{d \left( \frac{z}{c} \right) / \frac{z}{c}}{d \left( \frac{\partial f}{\partial z} \right) / \frac{\partial f}{\partial c} / \frac{\partial f}{\partial z},}
\]

which is equivalent to

\[
\psi = - \frac{d \ln \left( \frac{z}{c} \right)}{d \ln \left( \frac{\partial f}{\partial z} / \frac{\partial f}{\partial c} \right)}.
\]

Assume now that

\[
f(c, z) = (ac^\alpha + bz^\alpha)^{1/\alpha}, \tag{2.6}
\]

where \( \alpha < 1 \) and \( \alpha \neq 0 \). Then

\[
\frac{\partial f}{\partial c} = ac^{\alpha - 1} (ac^\alpha + bz^\alpha)^{\frac{1}{\alpha} - 1}, \quad \frac{\partial f}{\partial z} = bz^{\alpha - 1} (ac^\alpha + bz^\alpha)^{\frac{1}{\alpha} - 1},
\]

and thus

\[
\frac{\partial f}{\partial c} = \frac{b}{a} \left( \frac{z}{c} \right)^{\alpha - 1}.
\]

Computing the derivative with respect to \( \frac{z}{c} \), we get

\[
\frac{d \left( \frac{\partial f}{\partial c} / \frac{z}{c} \right)}{d \left( \frac{z}{c} \right)} = \frac{b}{a} (\alpha - 1) \left( \frac{z}{c} \right)^{\alpha - 2}
\]

and thus

\[
\psi = - \frac{b}{a} \left( \frac{z}{c} \right)^{\alpha - 1} \frac{1}{\frac{b}{a} (\alpha - 1) \left( \frac{z}{c} \right)^{\alpha - 2}} = - \frac{1}{\alpha - 1} = \frac{1}{1 - \alpha},
\]

which is independent of \((c, z)\). Therefore the utility function (2.6) is referred to as CES (Constant Elasticity of Substitution) utility.

For the Cobb-Douglas utility function

\[
f(c, z) = c^a z^{1-a}, \quad 0 < a < 1, \tag{2.7}
\]

the intertemporal elasticity of substitution equals 1. In fact, the Cobb-Douglas utility function (2.7) can be seen as the limit of the utility function (2.6) assuming \( b = 1 - a \) as \( \alpha \to 0 \).

### 2.7 Preferences for multi-date consumption plans

Above we implicitly considered preferences for consumption at one given future point in time. We need to generalize the ideas and results to settings with consumption at several dates. In one-period models individuals can consume both at time 0 (beginning-of-period) and at time 1 (end-of-period). In multi-period models individuals can consume either at each date in the discrete
time set $\mathcal{T} = \{0, 1, 2, \ldots, T\}$ or at each date in the continuous time set $\mathcal{T} = [0, T]$. In any case a
consumption plan is a stochastic process $c = (c_t)_{t \in \mathcal{T}}$ where each $c_t$ is a random variable representing
the state-dependent level of consumption at time $t$.

Consider the discrete-time case and, for each $t$, let $Z_t \subseteq \mathbb{R}$ denote the set of all possible consump-
tion levels at date $t$ and define $Z = Z_0 \times Z_1 \times \cdots \times Z_T \subseteq \mathbb{R}^{T+1}$, then any consumption plan $c$ can
again be represented by a probability distribution $\pi$ on the set $Z$. For finite $Z$, we can again apply
Theorem 2.1 so that under the relevant axioms, we can represent preferences by a utility index $U$, which to each consumption plan $(c_t)_{t \in \mathcal{T}} = (c_0, c_1, \ldots, c_T)$ attaches a real number $U((c_0, c_1, \ldots, c_T))$ with higher numbers to the more preferred consumption plans. If we further impose the Substitu-
tion Axiom, Theorem 2.2 ensures an expected utility representation, i.e., the existence of a utility function
$U : Z \to \mathbb{R}$ so that consumption plans are ranked according to their expected utility, i.e.,

$$U((c_0, c_1, \ldots, c_T)) = E[U(c_0, c_1, \ldots, c_T)] = \sum_{\omega \in \Omega} p_\omega U((c_0, c_1(\omega), \ldots, c_T(\omega))).$$

We can call $U$ a multi-date utility function since it depends on the consumption levels at all
dates. Again this result can be extended to the case of an infinite $Z$, e.g., $Z = \mathbb{R}^{T+1}$, but also
to continuous-time settings where $U$ will then be a function of the entire consumption process
$c = (c_t)_{t \in [0, T]}$.

### 2.7.1 Additively time-separable expected utility

Often time-additivity is assumed so that the utility the individual gets from consumption in
one period does not directly depend on what she consumed in earlier periods or what she plan to
consume in later periods. For the discrete-time case, this means that

$$U((c_0, c_1, \ldots, c_T)) = \sum_{t=0}^{T} u_t(c_t)$$

where each $u_t$ is a valid “single-date” utility function. Still, when the individual has to choose her
current consumption rate, she will take her prospects for future consumption into account. The
continuous-time analogue is

$$U((c_t)_{t \in [0, T]}) = \int_0^T u_t(c_t) \, dt.$$

In addition it is typically assumed that $u_t(c_t) = e^{-\delta t} u_t(c_t)$ for all $t$. This is to say that the direct
utility the individual gets from a given consumption level is basically the same for all dates, but
the individual prefers to consume any given number of goods sooner than later. This is modeled by
the subjective time preference rate $\delta$, which we assume to be constant over time and independent
of the consumption level. More impatient individuals have higher $\delta$’s. In sum, the life-time utility
is typically assumed to be given by

$$U((c_0, c_1, \ldots, c_T)) = \sum_{t=0}^{T} e^{-\delta t} u_t(c_t)$$
in discrete-time models and

$$U((c_t)_{t \in [0, T]}) = \int_0^T e^{-\delta t} u_t(c_t) \, dt.$$
in continuous-time models. In both cases, \( u \) is a “single-date” utility function such as those discussed in Section 2.6.\(^1\)

Time-additivity is mostly assumed for tractability. However, it is important to realize that the time-additive specification does not follow from the basic axioms of choice under uncertainty, but is in fact a strong assumption, which most economists agree is not very realistic. One problem is that time-additive preferences induce a close link between the reluctance to substitute consumption across different states of the economy (which is measured by risk aversion) and the willingness to substitute consumption over time (which can be measured by the so-called elasticity of intertemporal substitution). Solving intertemporal utility maximization problems of individuals with time-additive CRRA utility, it turns out that an individual with a high relative risk aversion will also choose a very smooth consumption process, i.e., she will have a low elasticity of intertemporal substitution. There is nothing in the basic theory of choice that links the risk aversion and the elasticity of intertemporal substitution together. For one thing, risk aversion makes sense even in an atemporal (i.e., one-date) setting where intertemporal substitution is meaningless and, conversely, intertemporal substitution makes sense in a multi-period setting without uncertainty in which risk aversion is meaningless. The close link between the two concepts in the multi-period model with uncertainty is an unfortunate consequence of the assumption of time-additive expected utility.

According to Browning (1991), non-additive preferences were already discussed in the 1890 book “Principles of Economics” by Alfred Marshall. See Browning’s paper for further references to the critique on intertemporally separable preferences. Let us consider some alternatives that are more general and still tractable.

### 2.7.2 Habit formation and state-dependent utility

The key idea of habit formation is to let the utility associated with the choice of consumption at a given date depend on past choices of consumption. In a discrete-time setting the utility index of a given consumption process \( c \) is now given as \( E[\sum_{t=0}^{T} e^{-\delta t} u(c_t, h_t)] \), where \( h_t \) is a measure of the standard of living or the habit level of consumption, e.g., a weighted average of past consumption rates such as

\[
    h_t = h_0 e^{-\beta t} + \alpha \sum_{s=1}^{t-1} e^{-\beta (t-s)} c_s,
\]

where \( h_0, \alpha, \) and \( \beta \) are non-negative constants. It is assumed that \( u \) is decreasing in \( h \) so that high past consumption generates a desire for high current consumption, i.e., preferences display intertemporal complementarity. In particular, models where \( u(c, h) \) is assumed to be of the power-linear form,

\[
    u(c, h) = \frac{1}{1-\gamma} (c - h)^{1-\gamma}, \quad \gamma > 0, c \geq h,
\]

\(^1\)Some utility functions are negative, including the frequently used power utility \( u(c) = c^{1-\gamma}/(1 - \gamma) \) with a constant relative risk aversion \( \gamma > 1 \). When \( \delta > 0 \), we will then have that \( e^{-\delta t} u(c) \) is in fact bigger (less negative) than \( u(c) \), which may seem to destroy the interpretation of \( \delta \) stated in the text. However, for the decisions made by the investor it is the marginal utilities that matter and, when \( \delta > 0 \) and \( u \) is increasing, \( e^{-\delta t} u'(c) \) will be smaller than \( u'(c) \) so that, other things equal, the individual will choose higher current than future consumption. Therefore, it is fair to interpret \( \delta \) as a time preference rate and expect it to be positive.
2.7 Preferences for multi-date consumption plans

Preferecns for multi-date consumption plans turn out to be computationally tractable. This is closely related to the subsistence HARA utility, but with habit formation the “subsistence level” \( h \) is endogenously determined by past consumption. The corresponding absolute and relative risk aversions are

\[
\text{ARA}(c, h) \equiv -\frac{u_{cc}(c, h)}{u_c(c, h)} = \frac{\gamma}{c - h}, \\
\text{RRA}(c, h) \equiv -\frac{u_{cc}(c, h)}{u_c(c, h)} = \frac{\gamma c}{c - h},
\]

(2.8)

where \( u_c \) and \( u_{cc} \) are the first- and second-order derivatives of \( u \) with respect to \( c \). In particular, the relative risk aversion is decreasing in \( c \). Note that the habit formation preferences are still consistent with expected utility.

A related line of extension of the basic preferences is to allow the preferences of an individual to depend on some external factors, i.e., factors that are not fully determined by choices made by the individual. One example that has received some attention is where the utility which some individual attaches to her consumption plan depends on the consumption plans of other individuals or maybe the aggregate consumption in the economy. This is often referred to as “keeping up with the Jones’es.” If you see your neighbors consume at high rates, you want to consume at a high rate too. Utility is state-dependent. Models of this type are sometimes said to have an external habit, whereas the habit formation discussed above is then referred to as internal habit.

If we denote the external factor by \( X_t \), a time-additive life-time expected utility representation is

\[
E[\sum_{t=0}^{T} e^{-\delta t} u(c_t, X_t)],
\]

and a tractable version is

\[
u(c, X) = \frac{1}{1-\gamma} (c - X)^{1-\gamma}
\]

very similar to the subsistence CRRA or the specific habit formation utility given above. In this case, however, “subsistence” level is determined by external factors. Another tractable specification is

\[
u(c, X) = \frac{1}{1-\gamma} (c/X)^{1-\gamma}.
\]

The empirical evidence of habit formation preferences is mixed. The time variation in risk aversion induced by habits as shown in (2.8) will generate variations in the Sharpe ratios of risky assets over the business cycle, which are not explained in simple models with CRRA preferences and appear to be present in the asset return data. Campbell and Cochrane (1999) construct a model with a representative individual having power-linear external habit preferences in which the equilibrium Sharpe ratio of the stock market varies counter-cyclically in line with empirical observations. However, a counter-cyclical variation in the relative risk aversion of a representative individual can also be obtained in a model where each individual has a constant relative risk aversion, but the relative risk aversions are different across individuals, as explained, e.g., by Chan and Kogan (2002). Various studies have investigated whether a data set of individual decisions on consumption, purchases, or investments are consistent with habit formation in preferences. To mention a few studies, Ravina (2007) reports strong support for habit formation, whereas Dynan (2000), Gomes and Michaelides (2003), and Brunnermeier and Nagel (2008) find no evidence of habit formation at the individual level.

2.7.3 Recursive utility

Another preference specification gaining popularity is the so-called recursive preferences or Epstein-Zin preferences, suggested and discussed by, e.g., Kreps and Porteus (1978), Epstein and Zin (1989, 1991), and Weil (1989). The original motivation of this representation of preferences is that it allows individuals to have preferences for the timing of resolution of uncertainty, which is not consistent with the standard multi-date expected utility theory and violates the set of behavioral axioms.
In a discrete-time framework Epstein and Zin (1989, 1991) assumed that life-time utility from time \( t \) on is captured by a utility index \( U_t \) (in this literature sometimes called the “felicity”) satisfying the recursive relation

\[ U_t = f(c_t, z_t), \]

where \( z_t = CE_t(U_{t+1}) \) is the certainty equivalent of \( U_{t+1} \) given information available at time \( t \) and \( f \) is an aggregator on the form

\[ f(c, z) = (ac^\alpha + bz^\alpha)^{1/\alpha}. \]

The aggregator is identical to the two-good CES utility specification (2.6) and, since \( z_t \) here refers to future consumption or utility, \( \psi = 1/(1-\alpha) \) is called the intertemporal elasticity of substitution.

An investor’s willingness to substitute risk between states is modeled through \( z_t \) as the certainty equivalent of a constant relative risk aversion utility function. Recall that the certainty equivalent for an atemporal utility function \( u \) is defined as

\[ CE = u^{-1}(E[u(x)]). \]

In particular for CRRA utility \( u(x) = x^{1-\gamma}/(1-\gamma) \) we obtain

\[ CE = (E[x^{1-\gamma}])^{\frac{1}{1-\gamma}}, \]

where \( \gamma > 0 \) is the relative risk aversion.

To sum up, Epstein-Zin preferences are specified recursively as

\[ U_t = \left( ac_t^{\alpha} + b \left( E_t[U_{t+1}^{1-\gamma}] \right)^{\frac{1}{1-\gamma}} \right)^{\frac{1}{1-\gamma}}. \]  

Using the fact that \( \alpha = 1 - \frac{1}{\psi} \), we can rewrite \( U_t \) as

\[ U_t = \left( ac_t^{\frac{1}{\psi}} + b \left( E_t[U_{t+1}^{1-\gamma}] \right)^{\frac{1}{1-\gamma}} \right)^{\frac{1}{1-\gamma}}. \]

Introducing \( \theta = (1-\gamma)/(1-\frac{1}{\psi}) \), we have

\[ U_t = \left( ac_t^{\frac{\theta}{\psi}} + b \left( E_t[U_{t+1}^{1-\gamma}] \right)^{\frac{1}{1-\gamma}} \right)^{\frac{\theta}{1-\gamma}}. \]

(2.10)

When the time horizon is finite, we need to specify the utility index \( U_T \) at the terminal date. If we allow for consumption at the terminal date and for a bequest motive, a specification like

\[ U_T = (ac_T^{\alpha} + \varepsilon a W_T^{\alpha})^{1/\alpha} \]

(2.11)

assumes a CES-type weighting of consumption and bequest in the terminal utility with the same CES-parameter \( \alpha \) as above. The parameter \( \varepsilon \geq 0 \) can be seen as a measure of the relative importance of bequest compared to consumption. Note that (2.11) involves no expectation as terminal wealth is known at time \( T \). Alternatively, we can think of \( cT_{t-1} \) as being the consumption over the final period and specify the terminal utility index as

\[ U_T = (\varepsilon a W_T^{\alpha})^{1/\alpha} = (\varepsilon a)^{1/\alpha} W_T. \]  

(2.12)
Bansal (2007) and other authors assume that \( a = 1 - b \), but the value of \( a \) is in fact unimportant as it does not affect optimal decisions and therefore no interpretation can be given to \( a \). At least this is true for an infinite time horizon and for a finite horizon when the terminal utility takes the form (2.11) or (2.12). In order to see this, first note that we can rewrite (2.9) as

\[
U_t = a^{1/\alpha} \left( c_t^{\alpha} + b a^{-1} \left( E_t \left[ U_{t+1}^{1-\gamma} \right] \right)^{\alpha/(\alpha-1)} \right) \]

which implies that

\[
a^{-1/\alpha} U_t = \left( c_t^{\alpha} + b \left( E_t \left[ \left( a^{-1/\alpha} U_{t+1} \right)^{1-\gamma} \right] \right)^{\alpha/(\alpha-1)} \right)^{1/\alpha},
\]

This suggests that the utility index \( \tilde{U} \) defined for any \( t \) by \( \tilde{U}_t = a^{-1/\alpha} U_t \) is equivalent to the utility index \( U \), since it is just a scaling, and it does not involve \( a \). With a finite time horizon and terminal utility given by (2.11), we see that \( \tilde{U}_T = a^{-1/\alpha} U_T = \left( c_T^\alpha + \varepsilon W_T^\alpha \right)^{1/\alpha} \), which also does not involve \( a \). Similarly when terminal utility is specified as in (2.12). Without loss of generality we can therefore let \( a = 1 \).

Time-additive power utility is the special case of recursive utility where \( \gamma = 1/\psi \). In order to see this, first note that with \( \gamma = 1/\psi \), we have \( \alpha = 1 - \gamma \) and \( \theta = 1 \) and thus

\[
U_t = \left( a c_t^{1-\gamma} + b E_t \left[ U_{t+1}^{1-\gamma} \right] \right)^{1/(1-\gamma)}
\]

or

\[
U_t^{1-\gamma} = a c_t^{1-\gamma} + b E_t \left[ U_{t+1}^{1-\gamma} \right].
\]

If we start unwinding the recursions, we get

\[
U_t^{1-\gamma} = a c_t^{1-\gamma} + b E_t \left[ a c_{t+1}^{1-\gamma} + b E_{t+1} \left[ U_{t+2}^{1-\gamma} \right] \right] = a E_t \left[ c_t^{1-\gamma} + b a^{1-\gamma} + b^2 E_t \left[ U_{t+2}^{1-\gamma} \right] \right].
\]

If we continue this way and the time horizon is infinite, we obtain

\[
U_t^{1-\gamma} = a \sum_{s=0}^{\infty} E_t b^s c_{t+s}^{1-\gamma},
\]

whereas with a finite time horizon and the terminal utility index (2.12), we obtain

\[
U_t^{1-\gamma} = a \left( \sum_{s=0}^{T-t} b^s E_t \left[ c_{t+s}^{1-\gamma} \right] + \varepsilon b^{T-t} E_t \left[ W_T^{1-\gamma} \right] \right).
\]

In any case, observe that

\[
V_t = \frac{1}{a(1-\gamma)} U_t^{1-\gamma}
\]

is an increasing function of \( U_t \) and will therefore represent the same preferences as \( U_t \). Moreover, \( V_t \) is clearly equivalent to time-additive expected utility. Note that \( b \) plays the role of the subjective discount factor which we often represent by \( e^{-\delta} \).
Chapter 2. Preferences

The Epstein-Zin preferences are characterized by three parameters: the relative risk aversion $\gamma$, the elasticity of intertemporal substitution $\psi$, and the subjective discount factor $b = e^{-\delta}$. Relative to the standard time-additive power utility, the Epstein-Zin specification allows the relative risk aversion (attitudes towards atemporal risks) to be disentangled from the elasticity of intertemporal substitution (attitudes towards shifts in consumption over time). Moreover, Epstein and Zin (1989) shows that when $\gamma > 1/\psi$, the individual will prefer early resolution of uncertainty. If $\gamma < 1/\psi$, late resolution of uncertainty is preferred. For the standard utility case $\gamma = 1/\psi$, the individual is indifferent about the timing of the resolution of uncertainty. Note that in the relevant case of $\gamma > 1$, the auxiliary parameter $\theta$ will be negative if and only if $\psi > 1$. Empirical studies disagree about reasonable values of $\psi$. Some studies find $\psi$ smaller than one (for example Campbell 1999), other studies find $\psi$ greater than one (for example Vissing-Jørgensen and Attanasio 2003).

The continuous-time equivalent of recursive utility is called stochastic differential utility and studied by, e.g., Duffie and Epstein (1992). The utility index $U_t$ associated at time $t$ with a given consumption process $c$ over the remaining lifetime $[t, T]$ is recursively given by

$$U_t = E_t \left[ \int_t^T f(c_s, U_s) \, ds \right]$$

where we assume a zero utility of terminal wealth, $U_T = 0$. Here $f$ is a so-called normalized aggregator. A somewhat tractable version of $f$ is

$$f(c, U) = \begin{cases} 
\frac{\delta}{1-\psi} c^{1-\psi} \left[(1-\gamma)\ln c - \ln (1-\gamma)\right] - \delta \theta U, & \text{for } \psi \neq 1 \\
(1-\gamma)\delta \ln c - \delta U \ln (1-\gamma)U, & \text{for } \psi = 1 \\
\frac{\delta}{1-\psi} e^{-1-\psi} - \frac{\delta}{1-\psi} U - \delta \theta U, & \text{for } \gamma = 1, \psi \neq 1 \\
\delta \ln c - \delta U, & \text{for } \gamma = \psi = 1 
\end{cases}$$

(2.13)

where $\theta = (1-\gamma)/(1-\frac{1}{\psi})$. This can be seen as the continuous-time version of the discrete-time Epstein-Zin preferences in (2.10). Again, $\delta$ is a subjective time preference rate, $\gamma$ reflects the degree of risk aversion towards atemporal bets, and $\psi > 0$ reflects the intertemporal elasticity of substitution towards deterministic consumption plans. It is also possible to define a normalized aggregator for $\gamma = 1$ and for $0 < \gamma < 1$ but we focus on the empirically more reasonable case of $\gamma > 1$. As in the discrete-time framework, the special case where $\psi = 1/\gamma$ (so that $\theta = 1$) corresponds to the classic time-additive power utility utility specification. Let us confirm that for the case $\psi = 1/\gamma \neq 1$, where the first definition in (2.13) applies. In this case

$$U_t = E_t \left[ \int_t^T \left( \frac{\delta}{1-\gamma} c_s^{1-\gamma} - \delta U_s \right) \, ds \right] = E_t \left[ \int_t^T \frac{\delta}{1-\gamma} c_s^{1-\gamma} \, ds \right] - \delta E_t \left[ \int_t^T U_s \, ds \right].$$

This recursive relation is satisfied by

$$U_t = \delta E_t \left[ \int_t^T e^{-\delta(s-t)} \frac{1}{1-\gamma} c_s^{1-\gamma} \, ds \right],$$

(2.14)

With a finite time horizon and a bequest motive, there is really a fourth parameter, namely the relative weight of bequest and consumption, as represented by the constant $\varepsilon$ in (2.11) or (2.12).
2.8 Exercises

because then

\[ E_t \left[ \int_t^T \delta c_{s}^{1-\gamma} ds \right] = \delta E_t \left[ \int_t^T \left( \int_t^s e^{-\delta(s-t)} \frac{1}{1-\gamma} c_v^{1-\gamma} dv \right) ds \right] \]

\[ = \delta E_t \left[ \int_t^T \left( \int_t^s e^{-\delta(s-t)} ds \right) \frac{1}{1-\gamma} c_v^{1-\gamma} dv \right] \]

\[ = E_t \left[ \int_t^T \left( 1 - e^{-\delta(v-t)} \right) \frac{1}{1-\gamma} c_v^{1-\gamma} dv \right], \]

where the second equality follows by changing the order of integration, and consequently

\[ E_t \left[ \int_t^T \frac{\delta}{1-\gamma} c_{s}^{1-\gamma} ds \right] - \delta E_t \left[ \int_t^T \delta c_{s}^{1-\gamma} ds \right] = \delta E_t \left[ \int_t^T \left( 1 - e^{-\delta(s-t)} \right) \frac{1}{1-\gamma} c_v^{1-\gamma} ds \right] \]

\[ = \delta E_t \left[ \int_t^T e^{-\delta(s-t)} \frac{1}{1-\gamma} c_v^{1-\gamma} ds \right] = \delta \mathcal{U}_t. \]

The utility index in (2.14) is a positive multiple of—and therefore equivalent to—the traditional time-additive power utility specification.

Note that, in general, recursive preferences are not consistent with expected utility since \( \mathcal{U}_t \) depends non-linearly on the probabilities of future consumption levels.

2.7.4 Two-good, multi-period utility

For studying some problems it is useful or even necessary to distinguish between different consumption goods. Until now we have implicitly assumed a single consumption good which is perishable in the sense that it cannot be stored. However, individuals spend large amounts on durable goods such as houses and cars. These goods provide utility to the individual beyond the period of purchase and can potentially be resold at a later date so that it also acts as an investment. Another important good is leisure. Individuals have preferences both for consumption of physical goods and for leisure. A tractable two-good utility function is the Cobb-Douglas function:

\[ u(c_1, c_2) = \frac{1}{1-\gamma} \left( c_1^{\psi} c_2^{1-\psi} \right)^{1-\gamma}, \]

where \( \psi \in [0,1] \) determines the relative weighting of the two goods.

2.8 Exercises

Exercise 2.1. Give a proof of Theorem 2.3.

Exercise 2.2 (Adapted from Problem 3.3 in Kreps (1990)). Consider the following two probability distributions of consumption. \( \pi_1 \) gives 5, 15, and 30 (dollars) with probabilities \( \frac{1}{3}, \frac{5}{9}, \) and \( \frac{1}{9}, \) respectively. \( \pi_2 \) gives 10 and 20 with probabilities \( \frac{2}{3} \) and \( \frac{1}{3}, \) respectively.

(a) Show that we can think of \( \pi_1 \) as a two-step gamble, where the first gamble is identical to \( \pi_2. \) If the outcome of the first gamble is 10, then the second gamble gives you an additional 5 (total 15) with probability \( \frac{1}{2} \) and an additional \(-5 \) (total 5) also with probability \( \frac{1}{2}. \) If the
outcome of the first gamble is 20, then the second gamble gives you an additional 10 (total 30) with probability 1/3 and an additional −5 (total 15) with probability 2/3.

(b) Observe that the second gamble has mean zero and that π₁ is equal to π₂ plus mean-zero noise. Conclude that any risk-averse expected utility maximizer will prefer π₂ to π₁.

Exercise 2.3 ((Adapted from Chapter 3 in Kreps (1990)).) Imagine a greedy, risk-averse, expected utility maximizing consumer whose end-of-period income level is subject to some uncertainty. The income will be Y with probability \( \bar{p} \) and \( Y' < Y \) with probability \( 1 - \bar{p} \). Think of \( \Delta = Y - Y' \) as some loss the consumer might incur due an accident. An insurance company is willing to insure against this loss by paying \( \Delta \) to the consumer if she sustains the loss. In return, the company wants an upfront premium of \( \delta \). The consumer may choose partial coverage in the sense that if she pays a premium of \( a \delta \), she will receive \( a \Delta \) if she sustains the loss. Let \( u \) denote the von Neumann-Morgenstern utility function of the consumer. Assume for simplicity that the premium is paid at the end of the period.

(a) Show that the first order condition for the choice of \( a \) is

\[
\bar{p}u'(Y - a\delta) = (1 - \bar{p})(\Delta - \delta)u'(Y - (1 - a)\Delta - a\delta).
\]

(b) Show that if the insurance is actuarially fair in the sense that the expected payout \( (1 - \bar{p})\Delta \) equals the premium \( \delta \), then the consumer will purchase full insurance, i.e., \( a = 1 \) is optimal.

(c) Show that if the insurance is actuarially unfair, meaning \( (1 - \bar{p})\Delta < \delta \), then the consumer will purchase partial insurance, i.e., the optimal \( a \) is less than 1.

Exercise 2.4. Consider a one-period choice problem with four equally likely states of the world at the end of the period. The consumer maximizes expected utility of end-of-period wealth. The current wealth must be invested in a single financial asset today. The consumer has three assets to choose from. All three assets have a current price equal to the current wealth of the consumer. The assets have the following end-of-period values:

<table>
<thead>
<tr>
<th>state</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>probability</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>asset 1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>asset 2</td>
<td>81</td>
<td>100</td>
<td>100</td>
<td>144</td>
</tr>
<tr>
<td>asset 3</td>
<td>36</td>
<td>100</td>
<td>100</td>
<td>225</td>
</tr>
</tbody>
</table>

(a) What asset would a risk-neutral individual choose?

(b) What asset would a power utility investor, \( u(W) = \frac{1}{1-\gamma}W^{1-\gamma} \) choose if \( \gamma = 0.5 \)？ If \( \gamma = 2 \)？ If \( \gamma = 5 \)？

Now assume a power utility with \( \gamma = 0.5 \).

(c) Suppose the individual could obtain a perfect signal about the future state before she makes her asset choice. There are thus four possible signals, which we can represent by \( s_1 = \{1\} \), \( s_2 = \{2\} \), \( s_3 = \{3\} \), and \( s_4 = \{4\} \). What is the optimal asset choice for each signal? What is her expected utility before she receives the signal, assuming that the signals have equal probability?

(d) Now suppose that the individual can receive a less-than-perfect signal telling her whether the state is in \( s_1 = \{1, 4\} \) or in \( s_2 = \{2, 3\} \). The two possible signals are equally likely. What is the expected utility of the investor before she receives the signal?
2.8 Exercises

Exercise 2.5. Consider an individual with log utility, \( u(c) = \ln c \). What is her certainty equivalent and risk premium for the consumption plan which with probability 0.5 gives her \((1-\alpha)c\) and with probability 0.5 gives her \((1+\alpha)c\)? Confirm that your results are consistent with numbers for \( \gamma = 1 \) shown in Table 2.5.

Exercise 2.6. Use Equation (2.3) to compute approximate relative risk premia for the consumption gamble underlying Table 2.5 and compare with the exact numbers given in the table.

Exercise 2.7. Consider an atemporal setting in which an individual has a utility function \( u \) of consumption. His current consumption is \( c \). As always, the absolute risk aversion is \( \text{ARA}(c) = -u''(c)/u'(c) \) and the relative risk aversion is \( \text{RRA}(c) = -cu''(c)/u'(c) \).

Let \( \varepsilon \in [0,c] \) and consider an additive gamble where the individual will end up with a consumption of either \( c + \varepsilon \) or \( c - \varepsilon \). Define the additive indifference probability \( \pi(W, \varepsilon) \) for this gamble by

\[
\pi(W, \varepsilon) = \left( \frac{1}{2} + \pi(c, \varepsilon) \right) u(c + \varepsilon) + \left( \frac{1}{2} - \pi(c, \varepsilon) \right) u(c - \varepsilon). \tag{1}
\]

Assume that \( \pi(c, \varepsilon) \) is twice differentiable in \( \varepsilon \).

(a) Argue that \( \pi(c, \varepsilon) \geq 0 \) if the individual is risk-averse.

(b) Show that the absolute risk aversion is related to the additive indifference probability by the following relation

\[
\text{ARA}(c) = 4 \lim_{\varepsilon \to 0} \frac{\partial \pi(c, \varepsilon)}{\partial \varepsilon} \tag{2}
\]

and interpret this result. Hint: Differentiate twice with respect to \( \varepsilon \) in (1) and let \( \varepsilon \to 0 \).

Now consider a multiplicative gamble where the individual will end up with a consumption of either \( (1 + \varepsilon)c \) or \( (1 - \varepsilon)c \), where \( \varepsilon \in [0,1] \). Define the multiplicative indifference probability \( \Pi(W, \varepsilon) \) for this gamble by

\[
u(c) = \left( \frac{1}{2} + \Pi(c, \varepsilon) \right) u((1 + \varepsilon)c) + \left( \frac{1}{2} - \Pi(c, \varepsilon) \right) u((1 - \varepsilon)c). \tag{3}
\]

Assume that \( \Pi(c, \varepsilon) \) is twice differentiable in \( \varepsilon \).

(c) Derive a relation between the relative risk aversion \( \text{RRA}(c) \) and \( \lim_{\varepsilon \to 0} \frac{\partial \Pi(c, \varepsilon)}{\partial \varepsilon} \) and interpret the result.
One-period models

3.1 Introduction

TO COME...

3.2 The general one-period model

Given $d$ risky assets with (stochastic) rates of return $R = (R_1, \ldots, R_d)^\top$ and a risk-free asset with a (certain) rate of return $r$ over the period of interest. Consider an investor having an initial wealth $W_0$ and no income from non-financial sources. If the investor invests amounts $\theta = (\theta_1, \ldots, \theta_d)^\top$ in the risky assets and the remainder $\theta_0 = W_0 - \theta^\top 1$ in the risk-free asset, he will end up with wealth

$$W = W_0 + \theta^\top R + \theta_0 r = (1 + r)W_0 + \theta^\top (R - r1)$$

at the end of the period. Letting $\pi_i = \theta_i/W_0$ denote the fraction of wealth invested in the $i$'th asset, we can rewrite the terminal wealth as

$$W = W_0 [1 + r + \pi^\top (R - r1)],$$

where $\pi = (\pi_1, \ldots, \pi_d)^\top$.

We assume that preferences can be represented by expected utility of end-of-period consumption or wealth so the decision problem is to choose $\theta$ or, equivalently, $\pi$ to maximize $E[u(W)]$, where $u$ is a utility function. We will assume throughout the chapter that $u$ is increasing and concave and is sufficiently smooth for all the relevant derivatives to exist. Note that we ignore any consumption decision at the beginning of the planning period, i.e., we assume that the consumption decision has already been taken independently of the investment decision.

The first-order condition for the problem

$$\sup_{\theta \in \mathbb{R}^d} E[u((1 + r)W_0 + \theta^\top (R - r1))]$$

is

$$E[u'((1 + r)W_0 + \theta^\top (R - r1))(R - r1)] = 0. \quad (3.1)$$
The second-order condition for a maximum will be satisfied since we will assume that \( u \) is concave. Hence, the first-order condition alone will characterize the optimal investment.

Without further assumptions, Arrow (1971), Pratt (1964), and others have shown a number of interesting results on the optimal portfolio choice. We will state only a few and refer to Merton (1992, Ch. 2) for further properties of the general solution to this utility maximization problem.

### 3.2.1 One risky asset

First we will specialize to the case with a single risky asset so that the first-order condition simplifies to

\[
\mathbb{E}\left[u'((1 + r)W_0 + \theta(R - r))(R - r)\right] = 0. \tag{3.2}
\]

Assuming a single risky asset may seem very restrictive, but we will later see that under some conditions, all individuals will optimally combine the risk-free asset and a single portfolio of the available risky asset. In the results below, the only risky asset can thus be interpreted as that portfolio.

The first result concerns the sign of the optimal investment in the risky asset:

**Theorem 3.1.** Assume a single risky asset and a strictly increasing and concave utility function \( u \). The optimal risky investment \( \theta \) is positive/zero/negative if and only if the excess expected return \( \mathbb{E}[R] - r \) is positive/zero/negative.

**Proof.** Define \( f(\theta) = \mathbb{E}[u'( (1 + r)W_0 + \theta(R - r)) (R - r)] \). The first-order condition (3.2) for \( \theta \) is \( f(\theta) = 0 \). Note that \( f'(\theta) = \mathbb{E}[u''((1 + r)W_0 + \theta(R - r))(R - r)^2] \), which is negative since \( u'' < 0 \). Hence, \( f(\theta) \) is decreasing in \( \theta \). Also note that \( f(0) = \mathbb{E}[u'((1 + r)W_0)(R - r)] = u'((1 + r)W_0)(\mathbb{E}[R] - r) \). Since \( u' > 0 \), we have \( f(0) > 0 \) if and only if \( \mathbb{E}[R] > r \). For \( \mathbb{E}[R] > r \), the equation \( f(\theta) = 0 \) is therefore satisfied for a \( \theta > 0 \). \( \square \)

The next result describes how the optimal investment in the risky asset varies with initial wealth:

**Theorem 3.2.** Assume a single risky asset with \( \mathbb{E}[R] > r \) and assume a strictly increasing and concave utility function \( u \). The optimal risky investment \( \theta = \theta(W_0) \) has the following properties:

(i) If ARA(\( \cdot \)) is uniformly decreasing (respectively increasing; constant), then \( \theta \) is increasing (respectively decreasing; constant) in \( W_0 \). 

(ii) If RRA(\( \cdot \)) is uniformly decreasing (respectively increasing; constant), then \( \pi = \theta/W_0 \) is increasing (respectively decreasing; constant) in \( W_0 \).

**Proof.** (i) Suppose that ARA is decreasing; the other cases can be handled similarly. By the assumption \( \mathbb{E}[R] > r \) and Theorem 3.1, we have \( \theta > 0 \). For states in which the realized return on the risky asset exceeds the risk-free return, we will therefore have that end-of-period wealth satisfies \( W > (1 + r)W_0 \). With decreasing ARA, this implies that ARA(\( W \)) \( \leq \) ARA((1+r)W_0) or, equivalently,

\[
u''(W) \geq -\text{ARA}((1 + r)W_0) u'(W).
\]

Multiplying by \( R - r > 0 \) gives

\[
u''(W)(R - r) \geq -\text{ARA}((1 + r)W_0) u'(W)(R - r). \tag{3.3}
\]
3.2 The general one-period model

For states in which the realized return on the risky asset is smaller than the risk-free return, we obtain

\[ u''(W) \leq -\text{ARA} \left( (1 + r)W_0 \right) u'(W), \]

and multiplying by \( R - r < 0 \), we have to reverse the inequality, so that we again obtain (3.3), which is therefore true for all realized returns. Taking expectations, we have

\[ E[u''(W)(R - r)] \geq -\text{ARA} \left( (1 + r)W_0 \right) E[u'(W)(R - r)] = 0, \]

(3.4)
due to the first-order condition (3.2).

Now, differentiating the first-order condition with respect to \( W_0 \) gives

\[ E \left[ u''(W)(R - r) \left( 1 + r \frac{\partial \theta}{\partial W_0}(R - r) \right) \right] = 0, \]

which implies that

\[ \frac{\partial \theta}{\partial W_0} = \frac{E[u''(W)(R - r)]}{-E[u''(W)(R - r)^2]}. \]

(3.5)
The denominator is strictly positive since \( u'' < 0 \) and the numerator is positive due to (3.4). Hence \( \frac{\partial \theta}{\partial W_0} \geq 0 \).

(ii) Rewrite the first-order condition as

\[ E \left[ u'(1 + r)W_0 + W_0 \left( \frac{\theta}{W_0} \right) (R - r) \right] = 0, \]

Then the proof of the result is similar to the proof of (i) with the relative risk aversion replacing the absolute risk aversion. The details are left for the reader (see Exercise 3.1).

The following results provide insights about how the optimal investments depend on returns. Differentiating the first-order condition (3.2) with respect to the risk-free rate \( r \), we get

\[ E \left[ u''(W) \left( W_0 - \theta + \frac{\partial \theta}{\partial r}(R - r) \right)(R - r) - u'(W) \right] = 0, \]

which implies that

\[ \frac{\partial \theta}{\partial r} = \frac{E[u'(W)]}{E[u''(W)(R - r)^2]} - (W_0 - \theta) \frac{E[u''(W)(R - r)]}{E[u''(W)(R - r)^2]}. \]

(3.6)
Applying (3.5), we arrive at

\[ \frac{\partial \theta}{\partial r} = \frac{E[u'(W)]}{E[u''(W)(R - r)^2]} + \frac{W_0 - \theta}{1 + r} \frac{\partial \theta}{\partial W_0}. \]

The first term on the right-hand side can be interpreted as the substitution effect and is strictly negative. If the risk-free rate increases, the risk-free asset is more attractive, and the individual will invest more in the risk-free asset and less in the risky asset. The second term on the right-hand side is the income effect. Note that \( W_0 - \theta \) is the investment in the risk-free asset. Assuming this is positive, an increase in the risk-free rate will make the individual wealthier. For a unit increase in the risk-free rate, the end-of-period wealth will increase by exactly \( W_0 - \theta \), and the present value of that is \((W_0 - \theta)/(1 + r)\). This increase in present wealth is multiplied by the derivative \( \frac{\partial \theta}{\partial W_0} \) to get the impact on the optimal risky investment. The income effect can be positive or negative.
If the income effect is negative, then the sum of the substitution and the income effects is clearly negative so that \( \frac{\partial \theta}{\partial r} < 0 \). This will be the case if \( \theta \leq W \) and \( \frac{\partial \theta}{\partial W} > 0 \). The latter condition is satisfied when the absolute risk aversion is increasing in wealth, cf. Theorem 3.2, but this is an unrealistic assumption on preferences. A more interesting result is the following:

**Theorem 3.3.** Assume a single risky asset with limited liability so that the return satisfies \( R \geq -1 \). Assume a strictly increasing and concave utility function \( u \) so that the relative risk aversion \( \text{RRA}(W) \leq 1 \) for all \( W \). Then the optimal risky investment is strictly decreasing in the risk-free rate.

*Proof.* First note that we can rewrite (3.6) as

\[
\frac{\partial \theta}{\partial r} = \frac{E \left[ u'(W) - (W_0 - \theta)u''(W)(R - r) \right]}{E \left[ u'(W)(R - r)^2 \right]}
\]

\[
= \frac{E \left[ u'(W)(1 + \text{ARA}(W)(W_0 - \theta)(R - r)) \right]}{E \left[ u'(W)(R - r)^2 \right]}
\]

\[
= \frac{E \left[ u'(W)(1 - \text{RRA}(W) + \text{ARA}(W)W_0(1 + R)) \right]}{E \left[ u'(W)(R - r)^2 \right]}
\]

The denominator is negative. Under the assumptions of the theorem, the numerator is surely non-negative. Hence \( \frac{\partial \theta}{\partial r} \leq 0 \).

Under the assumptions of the theorem, the income effect is positive but it is dominated by the negative substitution effect. Note however that the relative risk aversion is generally believed to exceed 1. From the proof, we can see that if the relative risk aversion is “sufficiently higher” than 1, we will typically end up with the opposite conclusion, i.e., \( \frac{\partial \theta}{\partial r} \geq 0 \).

How does the optimal investment depend on the expected return on the risky asset? Decompose the risky return as \( R = \mu + \varepsilon \), where \( \mu = E[R] \) so that \( \varepsilon \) is the unexpected return. The first-order condition can then be rewritten as

\[
E \left[ u' \left( (1 + r)W_0 + \theta(\mu + \varepsilon - r) \right) (\mu + \varepsilon - r) \right] = 0.
\]

If we differentiate with respect to \( \mu \) and use (3.5), we find

\[
\frac{\partial \theta}{\partial \mu} = \frac{E[u'(W)]}{-E[u''(W)(R - r)^2]} + \theta \frac{\partial \theta}{\partial W} \frac{\partial \theta}{\partial \mu}.
\]

The first term on the right-hand side (the substitution effect) is positive for \( u \) increasing and concave. The second term on the right-hand side (the income effect) will be positive if \( \theta \geq 0 \) and \( \frac{\partial \theta}{\partial W} \geq 0 \), which is true if \( \mu \geq r \) and the absolute risk aversion is decreasing in wealth, as we expect it to be. We summarize the conclusion as follows:

**Theorem 3.4.** Assume a single risky asset with \( E[R] \geq r \). Assume that the utility function is strictly increasing and concave and exhibits a decreasing absolute risk aversion, \( \text{ARA}'(W) \leq 0 \). Then the optimal risky investment is increasing in the expected return on the risky asset.

### 3.2.2 Multiple risky assets

Now we return to the case with multiple risky assets. First we state a very intuitive result for general utility functions.
Theorem 3.5. An individual with strictly increasing and concave $u$ will undertake risky investments if and only if $E[R_j] > r$ for some $j \in \{1, \ldots, d\}$.

Proof. Define $f(\theta) = E[u'(1+r)W_0 + \theta^T(R-r1)(R-r1)]$. As in the proof of Theorem 3.1, it can be shown that $f$ is decreasing in each $\theta_j$. If, and only if, the optimal portfolio has $\theta_j \leq 0$ for all $j = 1, \ldots, d$, then
\[ E[u'(1+r)W_0] (R_j - r) \leq 0, \quad \forall j = 1, \ldots, d, \]
or, equivalently,
\[ u'(1+r)W_0) E[R_j - r] \leq 0, \quad \forall j = 1, \ldots, d. \]
Since $u'(\cdot) > 0$, this condition holds exactly when $E[R_j] \leq r$ for all $j = 1, \ldots, d$. 

The optimal portfolio will contain a positive position in some risky asset $i$ as long as at least one of the risky assets, say asset $j$, have an expected return exceeding the risk-free rate. But, with multiple risky assets, you cannot be sure that $i = j$, that will depend on the correlation between the risky assets.

For the special case of HARA utility where the absolute risk aversion is of the form
\[ \text{ARA}(z) = -\frac{u''(z)}{u'(z)} = \frac{1}{\alpha z + \beta} \]
we can say more about the optimal investments. Recall from Section 2.6 that, ignoring unimportant constants, marginal utility is given either by
\[ u'(z) = (\alpha z + \beta)^{-1/\alpha} \quad (3.7) \]
or by
\[ u'(z) = ae^{-az} \quad (3.8) \]
where $a = 1/\beta$ and the parameter $\alpha$ in the absolute risk aversion is zero.

Theorem 3.6. For an investor with HARA utility, the amount optimally invested in each risky asset is affine in wealth, i.e.,
\[ \theta^*(W_0) = (\alpha(1+r)W_0 + \beta)k \quad (3.9) \]
for some vector $k = (k_1, \ldots, k_d)^T$ independent of wealth and of the parameter $\beta$.

Note that the amount optimally invested in the risk-free asset is then also affine in wealth since
\[ \theta^*_0(W_0) = W_0 - (\theta^*(W_0))^T 1 = (1 - \alpha(1+r)k^T 1) W_0 - \beta k^T 1. \]
We give a proof of the theorem for the case (3.7) and leave the case with negative exponential utility for the reader as Exercise 3.2.

Proof. With marginal utility given by (3.7), the first-order condition (3.1) becomes
\[ E \left[ (\alpha(1+r)W_0 + \beta + \alpha \theta^*(R-r1))^T (R-r1) \right]^{-1/\alpha} (R-r1) = 0. \quad (3.10) \]
Fix some initial wealth $\tilde{W}_0$. Then the corresponding optimal portfolio $\theta^*(\tilde{W}_0)$ satisfies
\[ E \left[ (\alpha(1+r)\tilde{W}_0 + \beta + \alpha \left( \theta^*(\tilde{W}_0) \right)^T (R-r1) \right]^{-1/\alpha} (R-r1) = 0. \]
If we divide through by \( (\alpha(1 + r)\hat{W}_0 + \beta)^{-1/\alpha} \), we get
\[
E \left[ 1 + \frac{\alpha}{\alpha(1 + r)\hat{W}_0 + \beta} \left( \theta^*(\hat{W}_0) \right)^\top (R - r\mathbf{1}) \right]^{-1/\alpha} (R - r\mathbf{1}) = 0. \tag{3.11}
\]

Next, we multiply through by \( (\alpha(1 + r)\hat{W}_0 + \beta)^{-1/\alpha} \) and arrive at
\[
E \left[ (\alpha(1 + r)W_0 + \beta + \alpha(1 + r)\hat{W}_0 + \beta) \left( \theta^*(\hat{W}_0) \right)^\top (R - r\mathbf{1}) \right]^{-1/\alpha} (R - r\mathbf{1}) = 0.
\]

Comparing this with (3.10), we see that the optimal portfolio with initial wealth \( W_0 \) is
\[
\theta^*(W_0) = \frac{\alpha(1 + r)W_0 + \beta}{\alpha(1 + r)\hat{W}_0 + \beta} \theta^*(\hat{W}_0)
\]
so that (3.9) is satisfied with \( k = \theta^*(\hat{W}_0)/[\alpha(1 + r)\hat{W}_0 + \beta] \). If we substitute \( \theta^*(\hat{W}_0) = k[\alpha(1 + r)\hat{W}_0 + \beta] \) into (3.11), we get that the vector \( k \) satisfies
\[
E \left[ (1 + \alpha k^\top (R - r\mathbf{1}))^{-1/\alpha} (R - r\mathbf{1}) \right] = 0
\]
so that it cannot depend on \( \beta \).

\[\square\]

### 3.2.3 Examples with explicit solutions

For the special case of quadratic utility,
\[
u(z) = -(\bar{z} - z)^2, \quad u'(z) = 2(\bar{z} - z),
\]
the first-order condition is
\[
E [(\bar{z} - (1 + r)W_0 - \theta^\top (R - r\mathbf{1})) (R - r\mathbf{1})] = 0,
\]
which implies that
\[
(\bar{z} - (1 + r)W_0) (E [R] - r\mathbf{1}) - E [(R - r\mathbf{1}) (R - r\mathbf{1})^\top] \theta = 0.
\]
We then get the explicit solution
\[
\theta = (\bar{z} - (1 + r)W_0) (E [(R - r\mathbf{1}) (R - r\mathbf{1})^\top])^{-1} (E [R] - r\mathbf{1}),
\]
which is (3.9) with \( \alpha = -1, \beta = \bar{z}, \) and \( k = (E [(R - r\mathbf{1}) (R - r\mathbf{1})^\top])^{-1} (E [R] - r\mathbf{1}). \)

Under the assumption that the returns on the risky assets are normally distributed, we can also derive an explicit expression for the optimal portfolio for the special case of negative exponential utility, \( u(W) = -e^{-aW} \). If \( R \sim \mathcal{N}(\mu, \Sigma) \) where \( \mu \) is a \( d \)-dimensional vector of the expected rates of return and \( \Sigma \) is the \( d \times d \) variance-covariance matrix of these rates of return, then the end-of-period wealth for any given portfolio \( \theta \) is also normally distributed, \( W \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2) \), with mean and variance given by
\[
\mu_\theta = W_0(1 + r) + \theta^\top (\mu - r\mathbf{1}), \quad \sigma_\theta^2 = \theta^\top \Sigma \theta.
\]
Therefore,
\[
E[u(W)] = -E[e^{-aW}] = -e^{-a\mu_\theta + \frac{1}{2}a^2\sigma_\theta^2}.
\]
The function $x \mapsto -e^{-ax}$ is an increasing function so the portfolio $\theta$ that maximizes expected utility will also maximize

$$\mu_0 - \frac{a}{2} \sigma_0^2 = W_0(1 + r) + \theta^\top (\mu - r 1) - \frac{a}{2} \theta^\top \Sigma \theta.$$ 

This is achieved by the portfolio

$$\theta^* = \frac{1}{a} \Sigma^{-1} (\mu - r 1),$$

which is independent of wealth. This is consistent with Theorem 3.6 since $\alpha = 0$ for negative exponential utility. With normally distributed returns and constant absolute risk aversion, the amount optimally invested in each risky asset is independent of wealth.

### 3.3 Mean-variance analysis

Mean-variance analysis was introduced by Markowitz (1952, 1959). Mean-variance analysis assumes that the portfolio choice of investors will depend only on the mean and variance of their end-of-period wealth and hence on the mean and variances of the portfolios investors can form. A portfolio is said to be mean-variance efficient if it has the lowest return variance for a given expected return. The mean-variance efficient portfolios can thus be found by solving constrained optimization problems. We will follow Merton (1972) and use the Lagrangian optimization technique to solve for the efficient portfolios. For an alternative characterization see Hansen and Richard (1987) or Cochrane (2005, Ch. 5). Before we go into the derivations of optimal portfolios, let us discuss the theoretical foundation of mean-variance analysis.

#### 3.3.1 Theoretical foundation

In general an individual’s utility of wealth will depend on all moments of wealth. This can be seen by the Taylor expansion of $u(W)$ around the expected wealth, $E[W]$:

$$u(W) = u(E[W]) + u'(E[W])(W - E[W]) + \frac{1}{2} u''(E[W])(W - E[W])^2 + \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[W])(W - E[W])^n,$$

where $u^{(n)}$ is the $n’$th derivative of $u$. Taking expectations, we get

$$E[u(W)] = u(E[W]) + \frac{1}{2} u''(E[W]) \text{Var}(W) + \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[W]) E[(W - E[W])^n].$$

Here $E[(W - E[W])^n]$ is the central moment of order $n$. The variance is the central moment of order 2. Obviously, a greedy investor (which just means that $u$ is increasing) will prefer higher expected wealth to lower for fixed central moments of order 2 and higher. Moreover, a risk averse investor (so that $u'' < 0$) will prefer lower variance of wealth to higher for fixed expected wealth and fixed central moments of order 3 and higher. But when the central moments of order 3 and higher are not the same for all alternatives, we cannot just evaluate them on the basis of their expectation and variance. Of course, with quadratic utility, the derivatives of $u$ of order 3 and higher are zero, so the higher order moments of wealth are irrelevant. However, quadratic utility is a very unrealistic model of investor preferences.

Mean-variance analysis is valid if the returns on the risky assets are multivariate normally distributed, $\mathbf{R} \sim N(\mu, \Sigma)$. Here, $\mu$ is a vector of the expected rates of return on the risky assets,
and $\Sigma = (\Sigma_{ij})$ is the variance-covariance matrix of these rates of return, so that $\Sigma_{ij}$ denotes the covariance between the returns on asset $i$ and asset $j$. Given that the returns on all individual assets are normally distributed, the return on any portfolio—being a weighted average of the returns on the assets in the portfolio—will also be normally distributed. A portfolio characterized by the portfolio weights $\pi = (\pi_1, \ldots, \pi_d)^T$ on the risky assets and the weight $\pi_0 = 1 - \pi^T 1$ on the risk-free asset has a return of

$$R^\pi \equiv \pi_0 r + \pi^T R = r + \pi^T (R - r 1) = r + \sum_{i=1}^d \pi_i (R_i - r),$$

which is normally distributed with mean and variance given by

$$\mu(\pi) \equiv \mathbb{E}[R^\pi] = \pi_0 r + \pi^T \mu = r + \pi^T (\mu - r 1) = r + \sum_{i=1}^d \pi_i (\mu_i - r),$$

$$\sigma^2(\pi) \equiv \text{Var}[R^\pi] = \pi^T \Sigma \pi = \sum_{i=1}^d \sum_{j=1}^d \pi_i \pi_j \Sigma_{ij}.$$  

Consequently, the end-of-period wealth of each investor will also be normally distributed for any portfolio choice. All higher-order moments of wealth can be written in terms of mean and variance so that expected utility depends only on expected wealth and the variance of wealth.

An obvious short-coming of the assumption of normally distributed returns is the possibility of rates of returns smaller than -100%, which is inconsistent with limited liability of securities. It also allows for negative end-of-period wealth and hence negative consumption with positive probability, which is clearly unreasonable. An alternative which at first looks promising is to assume that the end-of-period prices of individual assets are lognormally distributed, ruling out negative prices and rates of return below 100%. The lognormal distribution is also fully described by its first two moments. Unfortunately, such an assumption is not tractable in a one-period setting since neither the value nor the return on a portfolio will then be lognormally distributed (the lognormal distribution is not stable under addition).

### 3.3.2 Mean-variance analysis with only risky assets

Assume that the variance-covariance matrix $\Sigma$ is non-singular, which is the case if none of the assets are redundant, i.e., no asset has a return which is a linear combination of the returns of other assets. The inverse of $\Sigma$ is denoted by $\Sigma^{-1}$. A portfolio is said to be mean-variance efficient if it has the minimum return variance among all the portfolios with the same mean return. Given the normality assumption on returns, greedy and risk averse investors will only choose among the mean-variance efficient portfolios. Assuming that there are no portfolio constraints, we can find a mean-variance efficient portfolio with expected return $\bar{\mu}$ by solving the quadratic minimization problem

$$\min_{\pi} \frac{1}{2} \pi^T \Sigma \pi$$

s.t. $\pi^T \mu = \bar{\mu},$

$$\pi^T 1 = 1.$$ 

The $\frac{1}{2}$ in the objective will be notationally convenient when we solve the problem. Clearly, the portfolio that minimizes half the variance will also minimize the variance.
We solve the problem by the Lagrange technique. Letting $\alpha$ and $\beta$ denote the Lagrange multipliers of the two constraints, the Lagrangian is

$$L = \frac{1}{2} \pi^\top \Sigma \pi + \alpha (\bar{\mu} - \pi^\top \mu) + \beta (1 - \pi^\top 1).$$

The first-order condition with respect to $\pi$ is

$$\frac{\partial L}{\partial \pi} = \Sigma \pi - \alpha \mu - \beta 1 = 0,$$

which implies that

$$\pi = \alpha \Sigma^{-1} \mu + \beta \Sigma^{-1} 1. \quad (3.12)$$

The first-order conditions with respect to the multipliers simply give the two constraints to the minimization problem. Substituting the expression (3.12) for $\pi$ into the two constraints, we obtain the equations

$$\alpha \mu^\top \Sigma^{-1} \mu + \beta 1^\top \Sigma^{-1} \mu = \bar{\mu},$$
$$\alpha \mu^\top \Sigma^{-1} 1 + \beta 1^\top \Sigma^{-1} 1 = 1.$$

Defining

$$A = \mu^\top \Sigma^{-1} \mu, \quad B = \mu^\top \Sigma^{-1} 1 = 1^\top \Sigma^{-1} \mu, \quad C = 1^\top \Sigma^{-1} 1, \quad D = AC - B^2, \quad (3.13)$$

we can write the solution to these two equations in $\alpha$ and $\beta$ as

$$\alpha = \frac{C \bar{\mu} - B}{D}, \quad \beta = \frac{A - B \bar{\mu}}{D}.$$

Substituting this into (3.12) we obtain

$$\pi = \pi(\bar{\mu}) = \frac{C \bar{\mu} - B}{D} \Sigma^{-1} \mu + \frac{A - B \bar{\mu}}{D} \Sigma^{-1} 1. \quad (3.14)$$

Some tedious calculations show that the variance of the return on this portfolio is equal to

$$\sigma^2(\bar{\mu}) \equiv \pi^\top(\bar{\mu}) \Sigma \pi(\bar{\mu}) = \frac{C \bar{\mu}^2 - 2B \bar{\mu} + A}{D}. \quad (3.15)$$

This is to be shown in Exercise 3.3. We see that the combinations of variance and mean form a parabola in a (mean, variance)-diagram.

Traditionally the portfolios are depicted in a (standard deviation, mean)-diagram. The above relation can also be written as

$$\frac{\sigma^2(\bar{\mu})}{1/C} = \frac{(\bar{\mu} - B/C)^2}{D/C^2} = 1,$$

from which it follows that the optimal combinations of standard deviation and mean form a hyperbola in the (standard deviation, mean)-diagram. This hyperbola is called the mean-variance frontier of risky assets. The mean-variance efficient portfolios are sometimes called frontier portfolios.

Before we proceed let us clarify a point in the derivation above. We have assumed that $D$ is non-zero. In fact, $D > 0$. To see this is true, first recall the following definition. A symmetric $d \times d$ matrix $\Sigma$ is said to be positive definite if $\pi^\top \Sigma \pi > 0$ for any non-zero $d$-vector $\pi$. Since in our case $\pi^\top \Sigma \pi$ equals the variance of the portfolio $\pi$ and all portfolios of risky assets will have a return with positive variance, the variance-covariance matrix $\Sigma$ is indeed a positive definite matrix.
A result in linear algebra says that the inverse $\Sigma^{-1}$ is then also positive definite, i.e., $x^T \Sigma^{-1} x > 0$ for any non-zero $d$-vector $x$. In particular we have $A > 0$ and $C > 0$. Also

$$AD = A(AC - B^2) = (B\mu - A1)^T \Sigma^{-1} (B\mu - A1) > 0$$

and since $A > 0$ we must have $D > 0$.

The minimum-variance portfolio is the portfolio that has the minimum variance among all portfolios. We can find this directly by solving the constrained minimization problem

$$\min_{\pi} \frac{1}{2} \pi^T \Sigma \pi$$

s.t. $\pi^T 1 = 1$

where there is no constraint on the mean portfolio return. Alternatively, we can minimize the variance $\sigma^2(\bar{\mu})$ in (3.15) over all $\bar{\mu}$. Taking the latter route, we find that the minimum variance is obtained when the mean return is $\bar{\mu}_{\text{min}} = B/C$ and the minimum variance is given by $\sigma^2_{\text{min}} = \sigma^2(\bar{\mu}_{\text{min}}) = 1/C$. From (3.14) we get that the minimum-variance portfolio is

$$\pi_{\text{min}} = \frac{1}{C} \Sigma^{-1} 1 = \frac{1}{1^T \Sigma^{-1} 1} \Sigma^{-1} 1.$$ (3.16)

It can be shown that the portfolio

$$\pi_{\text{slope}} = \frac{1}{B} \Sigma^{-1} \mu = \frac{1}{1^T \Sigma^{-1} \mu} \Sigma^{-1} \mu$$ (3.17)

is the portfolio that maximizes the slope of a straight line between the origin and a point on the mean-variance frontier in the $(\sigma, \mu)$-diagram. (This follows as a special case of the tangency portfolio derived in the following subsection.) Let us call $\pi_{\text{slope}}$ the maximum slope portfolio. This portfolio has mean $A/B$ and variance $A/B^2$. From (3.14) we see that any mean-variance optimal portfolio can be written as a linear combination of the maximum slope portfolio and the minimum-variance portfolio:

$$\pi(\bar{\mu}) = \frac{(C\bar{\mu} - B)B}{D} \pi_{\text{slope}} + \frac{(A - B\bar{\mu})C}{D} \pi_{\text{min}}.$$  

Note that the two multipliers of the portfolios sum to one. This is a two-fund separation result. If the investors can only form portfolios of the $d$ risky assets with normally distributed returns, any greedy and risk-averse investor will choose a combination of two special portfolios or funds, namely the maximum slope portfolio and the minimum-variance portfolio. These two portfolios are said to generate the mean-variance frontier of risky assets. In fact, it can be shown that any other two frontier portfolios generate the entire frontier.

Figure 3.1 shows an example of the mean-variance frontier generated from 10 individual assets.

### 3.3.3 Mean-variance analysis with both risky assets and a risk-free asset

A risk-free asset corresponds to a point $(0, r)$ in the $(\text{standard deviation}, \text{mean})$-diagram. The investors can combine any portfolio of risky assets with an investment in the risk-free asset. The (standard deviation, mean)-pairs that can be obtained by such a combination form a straight line between the point $(0, r)$ and the point corresponding to the portfolio of risky assets. Suppose for example that we invest a fraction $\alpha \leq 1$ of wealth in the risk-free asset and the fraction $1 - \alpha \geq 0$ in
a given portfolio of risky assets with some expected rate of return $\hat{\mu}$ and some standard deviation $\hat{\sigma}$. Then the mean and standard deviation of the combined portfolio are

$$\mu(\alpha) = \alpha r + (1 - \alpha)\hat{\mu}, \quad \sigma(\alpha) = (1 - \alpha)\hat{\sigma}.$$ 

Consequently,

$$\mu(\alpha) = \alpha r + \frac{\hat{\mu}}{\hat{\sigma}} \sigma(\alpha)$$

so that the set of points $\{(\sigma(\alpha), \mu(\alpha)) \mid \alpha \leq 1\}$ will form a straight line.\(^1\)

Other things equal, greedy and risk-averse investors want high expected return and low standard deviation so they will move as far to the “north-west” as possible in the diagram. Therefore they will pick a point somewhere on the upward-sloping line that is tangent to the mean-variance frontier of risky assets and goes through the point $(0, r)$. The point where this line is tangent to the frontier of risky assets corresponds to a portfolio which we refer to as the **tangency portfolio**. This is a portfolio of risky assets only. It is the portfolio that maximizes the Sharpe ratio over all risky portfolios. The Sharpe ratio of a portfolio is the ratio $(\mu(\pi) - r) / \sigma(\pi)$ between the excess expected return of a portfolio and the standard deviation of the return.

To determine the tangency portfolio we consider the problem

$$\max_{\pi} \frac{\pi^T \mu - r}{(\pi^T \Sigma \pi)^{1/2}}$$

s.t. $\pi^T 1 = 1$.

\(^1\)For $\alpha > 1$, the standard deviation of the combined portfolio is $\sigma(\alpha) = -(1 - \alpha)\hat{\sigma}$ so that we get $\mu(\alpha) = \alpha r - (\hat{\mu}/\hat{\sigma})\sigma(\alpha)$. 

**Figure 3.1: The mean-variance frontier.** The curve shows the mean-variance frontier generated from the 10 individual assets corresponding to the red x’s.
Applying the constraint, the objective function can be rewritten as

$$f(\pi) = \frac{\pi^\top (\mu - r 1)}{(\pi^\top \Sigma \pi)^{1/2}} = \pi^\top (\mu - r 1) \left(\frac{\pi^\top \Sigma \pi}{\pi^\top \Sigma \pi}\right)^{-1/2}.$$  

The derivative is

$$\frac{\partial f}{\partial \pi} = (\mu - r 1) \left(\frac{\pi^\top \Sigma \pi}{\pi^\top \Sigma \pi}\right)^{-1/2} - \left(\frac{\pi^\top \Sigma \pi}{\pi^\top \Sigma \pi}\right)^{-3/2} \pi^\top (\mu - r 1) \Sigma \pi$$

and \(\frac{\partial f}{\partial \pi} = 0\) implies that

$$\frac{\pi^\top (\mu - r 1)}{\pi^\top \Sigma \pi} = \Sigma^{-1} (\mu - r 1),$$

which we want to solve for \(\pi\). Note that the equation has a vector on each side. If two vectors are identical, they will also be identical after a division by the sum of the elements of the vector. The sum of the elements of the vector on the left-hand side of (3.18) is

$$1^\top \left(\frac{\pi^\top (\mu - r 1)}{\pi^\top \Sigma \pi}\right) = \frac{\pi^\top (\mu - r 1)}{\pi^\top \Sigma \pi} 1^\top \pi = \frac{\pi^\top (\mu - r 1)}{\pi^\top \Sigma \pi},$$

where the last equality is due to the constraint. The sum of the elements of the vector on the right-hand side of (3.18) is simply \(1^\top \Sigma^{-1} (\mu - r 1)\). Dividing each side of (3.18) with the sum of the elements we obtain the tangency portfolio

$$\pi_{\text{tan}} = \frac{\Sigma^{-1} (\mu - r 1)}{1^\top \Sigma^{-1} (\mu - r 1)}.$$  

(3.19)

The expectation and standard deviation of the rate of return on the tangency portfolio are given by

$$\mu_{\text{tan}} = \mu^\top \pi_{\text{tan}} = \frac{\mu^\top \Sigma^{-1} (\mu - r 1)}{1^\top \Sigma^{-1} (\mu - r 1)},$$

$$\sigma_{\text{tan}} = \left(\pi_{\text{tan}}^\top \Sigma \pi_{\text{tan}}\right)^{1/2} = \frac{((\mu - r 1)^\top \Sigma^{-1} (\mu - r 1))^{1/2}}{1^\top \Sigma^{-1} (\mu - r 1)}.$$  

The maximum Sharpe ratio, i.e., the slope of the line, is thus

$$\frac{\mu_{\text{tan}} - r}{\sigma_{\text{tan}}} = \frac{\mu^\top \Sigma^{-1} (\mu - r 1) - r}{1^\top \Sigma^{-1} (\mu - r 1)} = \frac{\mu^\top \Sigma^{-1} (\mu - r 1) - r [1^\top \Sigma^{-1} (\mu - r 1)]}{((\mu - r 1)^\top \Sigma^{-1} (\mu - r 1))^{1/2}}$$

$$= \frac{(\mu - r 1)^\top \Sigma^{-1} (\mu - r 1)}{((\mu - r 1)^\top \Sigma^{-1} (\mu - r 1))^{1/2}} = \left((\mu - r 1)^\top \Sigma^{-1} (\mu - r 1)\right)^{1/2}.$$  

The upward-sloping straight line between the points \((0, r)\) and \((\sigma_{\text{tan}}, \mu_{\text{tan}})\) constitutes the mean-variance frontier of all assets. Again we have two-fund separation since all investors will combine just two funds, where one fund is simply the risk-free asset and the other is the tangency portfolio. This result is the basis for the famous Capital Asset Pricing Model (CAPM) developed by Sharpe (1964), Lintner (1965), and Mossin (1966). Note that also in this setting all investors will hold different risky assets in the same proportion to each other, i.e., for any \(i, j \in \{1, \ldots, d\}\) the ratio \(\pi_i/\pi_j\) is the same for all investors.

Exactly which combination of the two generating portfolios that a particular investor prefers is in general difficult to determine. For the unrealistic case of negative exponential utility (CARA)
the optimal combination can be determined in closed form as shown in Section 3.2. For other utility functions numerical optimization is necessary. In this regard the only advantage of the mean-variance framework is the two fund separation result since that allows us to look for a single portfolio weight (the fraction of wealth invested in the tangency portfolio) rather than portfolio weights of all risky assets. The numerical optimization is thus simpler assuming the mean-variance set-up.

Note that due to the assumption of normally distributed returns, the terminal wealth of the investor can go anywhere from $-\infty$ to $+\infty$ as long as some non-zero amount is invested in some risky asset. For utility functions with infinite marginal utility at a level higher than $-\infty$, the utility-maximizing decision will be to invest the entire wealth in the risk-free asset. This is for example the case for CRRA utility. The assumptions of the mean-variance analysis thus rule out its applications for reasonable utility functions!

### 3.4 A numerical example

TO COME...

### 3.5 Mean-variance analysis with constraints

TO COME...


Alexander, Baptista, and Yan (2007): Value-at-risk type constraints

### 3.6 Estimation

Mean-variance optimization is quite sensitive to the magnitudes of the inputs, i.e., expected returns, variances, and covariances. Chopra and Ziemba (1993) show that it is particularly important to obtain precise estimates of the expected returns. On the other hand, the expected returns are very hard to estimate precisely from historical returns, cf., e.g., Merton (1980).

For more on estimation and model uncertainty and how that affects optimal portfolio choice, see Garlappi, Uppal, and Wang (2007) and the references therein...

### 3.7 Critique of the one-period framework

- Investors typically get utility from consumption at many points in time and not simply the wealth level at one particular date.

- Even in the case where the investor only obtains utility from wealth at one date, she has the opportunity to change her portfolio over time, which she would normally do as new information arises (e.g., when stock prices and interest rates change) or simply because time passes. Investors live in a dynamic model and will take decisions dynamically. Of course, the existence of transaction costs is a reason for not changing the portfolio too frequently, but if we are really worried about transaction costs we should explicitly model that imperfection; the analysis of such models is quite difficult, however.
Consumption and investment decisions are generally not to be separated from each other. Investments are meant to generate future consumption!

The normality (or similar sufficient distributional) assumption employed in the mean-variance analysis is not reasonable, neither from a theoretical nor an empirical point of view. For example, the normal distribution allocates a strictly positive probability to a return below -100%, which cannot happen for investments in securities with limited liability.

3.8 Exercises

Exercise 3.1. Provide the details of the proof of part (ii) in Theorem 3.2.

Exercise 3.2. Give a proof of Theorem 3.6 for the case of negative exponential utility where marginal utility is given by (3.8).

Exercise 3.3. Show Equation (3.15).

Exercise 3.4. Let $R^\pi$ denote the return on a portfolio located on the mean-variance efficient frontier for risky assets only and suppose that $\pi$ is different from the minimum-variance portfolio. Show that there is a portfolio $z(\pi)$ also located on the mean-variance efficient frontier for risky assets only, which has the property that $\text{Cov}[R^\pi, Z^z(\pi)] = 0$. Show that $E[R^{z(\pi)}] = (A - B E[R^\pi])/(B - C E[R^\pi])$, where $A$, $B$, and $C$ are the constants defined in (3.13). \textit{Hint:} First show that the covariance between the return on the efficient portfolio with mean $m_1$ and the return on the efficient portfolio with mean $m_2$ is equal to $(Cm_1m_2 - B[m_1 + m_2] + A)/D$.

Exercise 3.5. Let $R_{\text{min}}$ denote the return on the minimum-variance portfolio of risky assets. Let $R$ be the return on any risky asset or portfolio of risky assets, efficient or not. Show that $\text{Cov}[R, R_{\text{min}}] = \text{Var}[R_{\text{min}}]$. \textit{Hint:} Consider a portfolio consisting of a fraction $a$ in this risky asset and a fraction $(1-a)$ in the minimum-variance portfolio. Compute the variance of the return on this portfolio and realize that the variance has to be minimized for $a = 0$.

Exercise 3.6. Let $R_1$ denote the return on a mean-variance efficient portfolio of risky assets and let $R_2$ denote another, not necessarily efficient, portfolio of risky assets with $E[R_2] = E[R_1]$. Show that $\text{Cov}[R_1, R_2] = \text{Var}[R_1]$ and conclude that $R_1$ and $R_2$ are positively correlated.
CHAPTER 4

Discrete-time multi-period models

4.1 Introduction

To study dynamic consumption and investment decisions, several papers have looked at multi-period, discrete-time models where the investor has the opportunity to consume and rebalance her portfolio at a number of fixed dates. Certainly this is a valuable extension of the single-period setting, but it is still a limitation that the investor can only change her decisions at pre-specified points in time and not react to new information arriving between these points in time. A continuous-time model seems more reasonable. Furthermore, the results on optimal consumption and investment strategies are typically clearer in continuous-time models than in discrete-time models, and the necessary mathematical computations are much more elegant in a continuous-time framework. Therefore, we will not give much attention to multi-period, discrete-time models. However, some aspects of the set-up of continuous-time models may be easier to understand if we start by looking at a discrete-time model and then take the limit as the period length goes to zero. The basic references for the discrete-time models are Samuelson (1969), Hakansson (1970), Fama (1970, 1976), and Ingersoll (1987, Ch. 11).

4.2 A multi-period, discrete-time framework for asset allocation

We consider an individual living over the time interval \([0, T]\) and assume that the individual can revise consumption and investment decisions at time points \(t_n = n\Delta t\), cf. the time line below. The terminal date \(T\) is assumed to be a multiple of the decision frequency, \(T = N\Delta t\). We define the set \(\mathcal{T} = \{t_0, t_1, \ldots, t_{N-1}\}\) of time points, where decisions are made. At the terminal date \(T\) no decisions are made.
We will assume that at any time $t \in \mathcal{T}$, the individual can invest in $d + 1$ assets. Asset 0 is an asset with a known return $r_t \Delta t$ over the next period, i.e., over the interval $[t, t + \Delta t]$, so that $r_t$ is the annualized short-term risk-free rate at time $t$. The returns on this asset in later periods are not necessarily known yet, but at least the asset is risk-free over the next period. The value at time $t$ of a dollar invested at time 0 and subsequently rolled over at the risk-free rate is denoted by $P^0_t$. We will refer to this investment as a “unit bank account.” The other assets $1, 2, \ldots, d$ are risky assets, i.e., assets with unknown returns even over the next period. For any $t \in \mathcal{T}$ and $t = T$, we denote by $P_t = (P^1_t, \ldots, P^d_t)^\top$ the vector of prices of the $d$ risky assets at time $t$. We assume for notational simplicity that the assets do not pay intermediate dividends so that returns are given only by percentage price changes. Let $R^{i}_{t+\Delta t} = (P^{i}_{t+\Delta t} - P^i_t)/P^i_t$ denote the return on risky asset $i$ over the interval $[t, t + \Delta t]$ and let $R_t = (R^1_{t+\Delta t}, \ldots, R^d_{t+\Delta t})^\top$ denote the vector of returns on all the risky assets over the same interval.

At any time $t \in \mathcal{T}$ the investor chooses a portfolio which is held unchanged until time $t + \Delta t$ and a consumption rate $c_t$ such that the total consumption in the interval $[t, t + \Delta t]$ is $c_t \cdot \Delta t$. (We assume that there is a single consumption good so that $c_t$ is one-dimensional.) This is subtracted from her wealth at time $t$. Of course, the portfolio and consumption chosen at time $t$ for the interval $[t, t + \Delta t]$ can only be based on the information known at time $t$. We assume that there is no consumption or investment beyond time $T$, which we can think of as the time of death (assumed to be known in advance!).

For the purposes of deriving the budget constraint we will first represent the portfolio by the number of units of each asset held. For any $t \in \mathcal{T}$, we let $M^i_t$ denote the number of units of asset $i = 0, 1, \ldots, d$ held in the period $[t, t + \Delta t)$. We will allow for the case where the agent earns income from other sources than his financial investments. We let $y_t$ be the rate of income earned in the period $[t, t + \Delta t)$ such that the entire income in this period is $y_t \cdot \Delta t$. We assume that the agent receives this amount at time $t$. Note that we do not model the labor supply decision resulting in this income, but take $y_t$ as exogenously given.

The agent enters date $t \in \mathcal{T}$ with a wealth of

$$W_t = \sum_{i=0}^{d} M^i_{t-\Delta t} P^i_t.$$

This is the value of her portfolio chosen in the previous period. She then receives income $y_t \cdot \Delta t$ and simultaneously has to choose the consumption rate $c_t$ and the new portfolio represented by $M^0_t, M^1_t, \ldots, M^d_t$. The budget restriction on these choices is that

$$(y_t - c_t) \Delta t = \sum_{i=0}^{d} \left[ M^i_t - M^i_{t-\Delta t} \right] P^i_t.$$
i.e., that income net of consumption equals the extra amount invested in the financial market. We then get that

\[ W_{t+\Delta t} - W_t = \sum_{i=0}^{d} M_i^t P_{i,t+\Delta t} - \sum_{i=0}^{d} M_i^{t-\Delta t} P_i^t \]

\[ = \sum_{i=0}^{d} M_i^t (P_{i,t+\Delta t} - P_i^t) + \sum_{i=0}^{d} (M_i^t - M_i^{t-\Delta t}) P_i^t \]

\[ = \sum_{i=0}^{d} M_i^t (P_{i,t+\Delta t} - P_i^t) + (y_t - c_t) \Delta t. \]

Let \( \theta_i^t = M_i^t P_i^t \) denote the amount invested in asset \( i \) at time \( t \in \mathcal{T} \) and let \( \theta_t = (\theta_1^t, \ldots, \theta_d^t)^\top \). Then the change in wealth can be rewritten as

\[ W_{t+\Delta t} - W_t = \theta_0^t r_t \Delta t + \theta_t^\top R_{t+\Delta t} + (y_t - c_t) \Delta t. \] (4.1)

We can also represent the portfolio by the fractions of wealth invested in the different assets. After receiving income and consuming at time \( t \), the funds invested will be \( W_t + (y_t - c_t) \Delta t \). Assuming this is non-zero, we can define the portfolio weight of asset \( i \) at time \( t \) as

\[ \pi_i^t = \frac{\theta_i^t}{W_t + (y_t - c_t) \Delta t}, \quad i = 0, 1, \ldots, d. \]

The vector of portfolio weights in the risky assets is denoted by \( \pi_t = (\pi_1^t, \ldots, \pi_d^t)^\top \). By construction the portfolio weight of the bank account is given by \( \pi_0^t = 1 - \pi_1^t \mathbf{1} = 1 - \sum_{i=1}^{d} \pi_i^t \). The end-of-period wealth can then be restated as

\[ W_{t+\Delta t} = (W_t + y_t \Delta t - c_t \Delta t) R_{t+\Delta t}^W, \] (4.2)

where

\[ R_{t+\Delta t}^W = 1 + r_t \Delta t + \pi_t^\top (R_{t+\Delta t} - r_t \mathbf{1}). \] (4.3)

Note that the only random variable (seen from time \( t \)) on the right-hand side of these wealth expressions is the return vector \( R_{t+\Delta t} \). Let us decompose the return into an expected and an unexpected part,

\[ R_{t+\Delta t} = \mu_t \Delta t + \varepsilon_{t+\Delta t} \sqrt{\Delta t}. \] (4.4)

Here \( \mu_t \) is the vector of expected rates of return per year, \( \varepsilon_{t+\Delta t} \) is a vector of independent stochastic shocks all with mean zero and variance one, and \( \varepsilon_{t+\Delta t} \) is a matrix determining how the returns are affected by these shocks. The values of \( \mu_t \) and \( \varepsilon_{t+\Delta t} \) are known at time \( t \). The realization of the shock vector \( \varepsilon_{t+\Delta t} \) will be known at time \( t + \Delta t \), just before the consumption and portfolio decisions at that date are taken. It follows that, seen at time \( t \), the variance-covariance matrix of \( R_{t+\Delta t} \) is given by \( \Sigma_{\varepsilon_{t+\Delta t}} \). The elements in \( \Sigma_{\varepsilon_{t+\Delta t}} = \Sigma_{\varepsilon_{t+\Delta t}} \) are hence annualized variances and covariances.

The wealth dynamics (4.1) can now be rewritten as

\[ W_{t+\Delta t} - W_t = [\theta_0^t r_t + \theta_t^\top \mu_t + y_t - c_t] \Delta t + \theta_t^\top \varepsilon_{t+\Delta t} \sqrt{\Delta t}. \] (4.5)

At time 0 the investor must choose the entire consumption rate process \( c = (c_t)_{t \in \mathcal{T}} \) and the entire portfolio process represented by \( \pi = (\pi_t)_{t \in \mathcal{T}} \) or \( \theta = (\theta_t)_{t \in \mathcal{T}} \). In other words, she must choose the current values \( c_0 \) and \( \pi_0 \) and for each future date \( t_n \) (with \( n = 1, \ldots, N - 1 \)) she must
choose a consumption rate \( c_n(\omega) \) and a portfolio \( \pi_n(\omega) \) for each possible state of the world \( \omega \) at day \( t_n \).

We assume that the life-time utility of consumption and terminal wealth is given by

\[
U(c_0, c_1, \ldots, c_{N-1}, W_T) = \sum_{n=0}^{N-1} e^{-\delta t_n} u(c_{t_n}) \Delta t + e^{-\delta T} \bar{u}(W_T)
\]
as discussed in Section 2.7. The maximal obtainable expected life-time utility seen from time 0 is therefore

\[
J_0 = \sup_{(c_n, \pi_n)} E \left[ \sum_{n=0}^{N-1} e^{-\delta t_n} u(c_{t_n}) \Delta t + e^{-\delta T} \bar{u}(W_T) \right],
\]
where the supremum is taken over all budget-feasible consumption and investment strategies. Similarly, for each \( t = i \Delta t \in \mathcal{T} \), we define

\[
J_t = \sup_{(c_n, \pi_n)} E_t \left[ \sum_{n=i}^{N-1} e^{-\delta (t_n-t)} u(c_{t_n}) \Delta t + e^{-\delta (T-t)} \bar{u}(W_T) \right],
\]
where the subscript on the expectations operator denotes that the expectation is taken conditional on the information known to the agent at time \( t = t_i \). \( J \) is often called the indirect or derived utility of wealth process or function, since it measures the highest attainable expected life-time utility the investor can derive from her current wealth in the current state of the world. Note that \( J_T = \bar{u}(W_T) \).

### 4.3 Dynamic programming in discrete-time models

In the definition of indirect utility in (4.6) the maximization is over both the current and all future consumption rates and portfolios. This is clearly a complicated maximization problem. We will now show that we can alternatively perform a sequence of simpler maximization problems. This result is based on the following manipulations, where \( t = t_i = i \Delta t \) as before:

\[
J_t = \sup_{(c_n, \pi_n)} E_t \left[ \sum_{n=i}^{N-1} e^{-\delta (t_n-t)} u(c_{t_n}) \Delta t + e^{-\delta (T-t)} \bar{u}(W_T) \right]
\]

\[
= \sup_{(c_n, \pi_n)} E_t \left[ u(c_t) \Delta t + \sum_{n=i}^{N-1} e^{-\delta (t_n-t)} u(c_{t_n}) \Delta t + e^{-\delta (T-t)} \bar{u}(W_T) \right]
\]

\[
= \sup_{(c_n, \pi_n)} E_t \left[ u(c_t) \Delta t + e^{-\delta \Delta t} E_{t+\Delta t} \left[ \sum_{n=i+1}^{N-1} e^{-\delta (t_n-[t+\Delta t])} u(c_{t_n}) \Delta t + e^{-\delta (T-[t+\Delta t])} \bar{u}(W_T) \right] \right]
\]

\[
= \sup_{c_t, \pi_t} E_t \left[ u(c_t) \Delta t + e^{-\delta \Delta t} \sup_{(c_n, \pi_n)} E_{t+\Delta t} \left[ \sum_{n=i+1}^{N-1} e^{-\delta (t_n-[t+\Delta t])} u(c_{t_n}) \Delta t + e^{-\delta (T-[t+\Delta t])} \bar{u}(W_T) \right] \right]
\]

Here, the first equality is simply due to the definition of indirect utility, the second equality comes from separating out the first term of the sum, the third equality is valid according to the law of iterated expectations, the fourth equality comes from separating out the discount term \( e^{-\delta \Delta t} \), and the final equality is due to the fact that only the inner expectation depends on future...
4.3 Dynamic programming in discrete-time models

consumption rates and portfolios. Noting that the inner supremum is by definition the indirect utility at time \( t + \Delta t \), we arrive at

\[
J_t = \sup_{c_t, \pi_t} \mathbb{E}_t \left[ u(c_t) \Delta t + e^{-\delta \Delta t} J_{t + \Delta t} \right] = \sup_{c_t, \pi_t} \left\{ u(c_t) \Delta t + e^{-\delta \Delta t} \mathbb{E}_t \left[ J_{t + \Delta t} \right] \right\} . \tag{4.7}
\]

This equation is called the Bellman equation, and the indirect utility \( J \) is said to have the dynamic programming property. The decision to be taken at time \( t \) is split up in two: (1) the consumption and portfolio decision for the current period and (2) the consumption and portfolio decisions for all future periods. We take the decision for the current period assuming that we will make optimal decisions in all future periods. Note that this does not imply that the decision for the current period is taken independently from future decisions. We take into account the effect that our current decision has on the maximum expected utility we can get from all future periods. The expectation \( \mathbb{E}_t [ J_{t + \Delta t} ] \) will depend on our choice of \( c_t \) and \( \pi_t \).

The dynamic programming property is the basis for a backward iterative solution procedure. First, we choose \( c_{t_{N-1}} \) and \( \pi_{t_{N-1}} \) to maximize

\[
u(c_{t_{N-1}}) \Delta t + e^{-\delta \Delta t} \mathbb{E}_{t_{N-1}} \left[ \bar{u}(W_T) \right] ,
\]

where

\[
W_T = (W_{t_{N-1}} + y_{t_{N-1}} \Delta t - c_{t_{N-1}} \Delta t) \left( 1 + r_{t_{N-1}} \Delta t + \pi_{t_{N-1}}^\top (R_T - r_{t_{N-1}} \Delta t 1) \right).
\]

This is done for each possible state at time \( t_{N-1} \) and gives us \( J_{t_{N-1}} \). Then we choose \( c_{t_{N-2}} \) and \( \pi_{t_{N-2}} \) to maximize

\[
u(c_{t_{N-2}}) \Delta t + e^{-\delta \Delta t} \mathbb{E}_{t_{N-2}} \left[ J_{t_{N-1}} \right] ,
\]

and so on until we reach time zero. Since we have to perform a maximization for each state of the world at every point in time, we have to make assumptions on the possible states at each point in time before we can implement the recursive procedure. The optimal decisions at any time are expected to depend on the wealth level of the agent at that date, but also on the value of other time-varying state variables that affect future returns on investment (e.g., the interest rate level) and future income levels. To be practically implementable only a few state variables can be incorporated. Also, these state variables must follow Markov processes so only the current values of the variables are relevant for the maximization at a given point in time.

Suppose that the relevant information is captured by a one-dimensional Markov process \( x = (x_t) \) so that the indirect utility at any time \( t \in \{ 0, \Delta t, \ldots, N \Delta t \} \) can be written as \( J_t = J(W_t, x_t, t) \).

Then the dynamic programming equation (4.7) becomes

\[
J(W_t, x_t, t) = \sup_{c_t, \pi_t} \left\{ u(c_t) \Delta t + e^{-\delta \Delta t} \mathbb{E}_t \left[ J(W_{t+\Delta t}, x_{t+\Delta t}, t+\Delta t) \right] \right\} , \quad t \in T.
\]

Doing the maximization we have to remember that \( W_{t+\Delta t} \) will be affected by the choice of \( c_t \) and \( \pi_t \). From our analysis of the wealth dynamics we have that

\[
W_{t+\Delta t} = (W_t + y_t \Delta t - c_t \Delta t)R^W_{t+\Delta t}, \quad R^W_{t+\Delta t} = 1 + r_t \Delta t + \pi_t^\top (R_{t+\Delta t} - r_t \Delta t 1),
\]

1 Readers familiar with option pricing theory may note the similarity to the problem of determining the optimal exercise strategy of a Bermudan/American option. However, for that problem the decision to be taken is much simpler (exercise or not) than for the consumption/portfolio problem.
In particular, we see that
\[
\frac{\partial W_{t+\Delta t}}{\partial c_t} = -R_t^W \Delta t, \quad \frac{\partial W_{t+\Delta t}}{\partial \pi_t} = (W_t + y_t \Delta t - c_t \Delta t) (R_{t+\Delta t} - r_t \Delta t). \]

The first-order condition for the maximization with respect to \(c_t\) is
\[
u'(c_t) \Delta t + e^{-\delta \Delta t} \mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) \frac{\partial W_{t+\Delta t}}{\partial c_t} \right] = 0,
\]
which implies that
\[
u'(c_t) = e^{-\delta \Delta t} \mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) R_{t+\Delta t}^W \right]. \tag{4.8}
\]

The first-order condition for the maximization with respect to \(\pi_t\) is
\[
\mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) \frac{\partial W_{t+\Delta t}}{\partial \pi_t} \right] = 0,
\]
which implies that
\[
\mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) (R_{t+\Delta t} - r_t \Delta t) \right] = 0. \tag{4.9}
\]

While we cannot generally solve for the optimal decisions, we can show an interesting and important result, the so-called envelope condition. First note that for the optimal choice \(\hat{c}_t, \hat{\pi}_t\) we have that
\[
J(W_t, x_t, t) = u(\hat{c}_t) \Delta t + e^{-\delta \Delta t} \mathbb{E}_t \left[ J(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) \right],
\]
where \(W_{t+\Delta t}\) is next period’s wealth using \(\hat{c}_t, \hat{\pi}_t\). Taking derivatives with respect to \(W_t\) in this equation, and acknowledging that \(\hat{c}_t\) and \(\hat{\pi}_t\) will in general depend on \(W_t\), we get
\[
J_W(W_t, x_t, t) = u'(\hat{c}_t) \frac{\partial \hat{c}_t}{\partial W_t} \Delta t + e^{-\delta \Delta t} \mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) \frac{\partial W_{t+\Delta t}}{\partial W_t} \right],
\]
where
\[
\frac{\partial \hat{W}_{t+\Delta t}}{\partial W_t} = R_{t+\Delta t}^W \left(1 - \frac{\partial \hat{c}_t}{\partial W_t} \Delta t\right) + (W_t + y_t \Delta t - \hat{c}_t \Delta t) \left( \frac{\partial \hat{\pi}_t}{\partial W_t} \right)^\top (R_{t+\Delta t} - r_t \Delta t). \]

Inserting this and rearranging terms, we get
\[
J_W(W_t, x_t, t) = e^{-\delta \Delta t} \mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) R_{t+\Delta t}^W \right] + \left(u'(\hat{c}_t) - e^{-\delta \Delta t} \mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) R_{t+\Delta t}^W \right] \right) \frac{\partial \hat{c}_t}{\partial W_t} \Delta t + (W_t + y_t \Delta t - c_t \Delta t) e^{-\delta \Delta t} \left( \frac{\partial \hat{\pi}_t}{\partial W_t} \right)^\top \mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) R_{t+\Delta t}^W \right].
\]

On the right-hand side the last two terms are zero due to the first-order conditions (4.8) and (4.9) so only the leading term remains, i.e.,
\[
J_W(W_t, x_t, t) = e^{-\delta \Delta t} \mathbb{E}_t \left[ J_W(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) R_{t+\Delta t}^W \right].
\]
Combining this with (4.8) we obtain
\[
u'(c_t) = J_W(W_t, x_t, t), \tag{4.10}
\]
which is the so-called **envelope condition**. As we will see, the condition also holds in the continuous-time models. The intuition of the envelope condition is that the optimal decision must be such that the marginal utility from consumption a bit more must be identical to the marginal utility from investing that bit in an optimal way. If that was not the case the allocation of wealth between consumption and investment should be reconsidered. For example, if \( u'(c_t) > J_W(W_t, x_t, t) \), the consumption \( c_t \) should be increased and the amount invested should be decreased.

Under some simplifying assumptions on the precise form of the utility functions \( u \) and \( \bar{u} \) and on the dynamics of asset returns and income, the backward iterative procedure yields an explicit solution to the maximization problem in the form of the optimal (possibly state- and time-dependent) consumption rate and portfolio process (and also the indirect utility of wealth \( J_t \)). Since we can obtain similar (and often clearer) results under similar assumptions in the more elegant and realistic continuous-time setting, we will not go into these discrete-time examples.
CHAPTER 5

Introduction to continuous-time modelling

5.1 Introduction

An introduction to stochastic processes and stochastic calculus is given in Appendix B...

5.2 The basic continuous-time setting

The basic elements of mainstream continuous-time models can be seen as the limit of the multi-period discrete-time model elements. The basis is a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with an associated filtration \(\mathcal{F} = (\mathcal{F}_t)_{t \in [0,T]}\) which is the formal model of the evolution of the relevant uncertainty for the investor.

The agent now has to choose a continuous-time process of consumption rates \(c = (c_t)_{t \in [0,T]}\) and a continuous-time portfolio process. The portfolio process can be represented by \(\theta = (\theta_t)_{t \in [0,T]}\), where \(\theta_t\) is the \(d\)-dimensional vector of amounts invested at time \(t\) in the \(d\) risky assets, or—at least when wealth is non-zero—by \(\pi = (\pi_t)_{t \in [0,T]}\), where \(\pi_t\) is the \(d\)-dimensional vector of fractions of wealth invested at time \(t\) in the \(d\) risky assets. The remaining financial wealth is invested in the locally risk-free asset so \(\theta^0_t = W_t - \theta^\top_t \mathbf{1} = W_t - \sum_{i=1}^d \theta^i_t\) and \(\pi^0_t = 1 - \pi^\top_t \mathbf{1}\). We assume that there is a single consumption good in the economy and this good is used as a numeraire so that all prices are measured in units of this consumption good, i.e., in real terms. We will always require that \(c_t \geq 0\) with probability one. We focus on unconstrained investors so that there are no constraints on the values \(\theta_t\) or \(\pi_t\) may have, i.e., they can take any value in \(\mathbb{R}^d\); see references in Section 18.1 to problems with constraints on the portfolios, e.g., short-selling constraints or portfolio mix constraints. The stochastic variables \(c_t\) and \(\theta_t\) (or \(\pi_t\)) must be \(\mathcal{F}_t\)-measurable, i.e., they can only depend on information available at time \(t\). In other words, the processes \(c\) and \(\theta\) (or \(\pi\)) are adapted. Other technical requirements should be added.\(^1\) A consumption and investment

\(^1\) The consumption process \(c\) must be an \(\mathcal{L}^1\)-process, i.e., \(\int_0^T \|c_t\| \, dt < \infty\) with probability one. The portfolio strategy \(\theta\) must satisfy that \(\theta^\top \mu\) is an \(\mathcal{L}^1\)-process and that \(\theta^\top \sigma\) is an \(\mathcal{L}^2\)-process, i.e., \(\int_0^T \|\theta^\top_t \sigma_t\|^2 \, dt < \infty\) with probability one. Finally, \(\theta\) must be a progressively measurable process which generally involves a bit more
strategy must also satisfy that the wealth process induced by the strategy always stays above a lower bound, say $K$, where $K \in \mathbb{R}$. This rules out doubling strategies, cf. the discussion in Duffie (2001, Ch. 6). In fact, we will typically require that wealth stays non-negative at all times, corresponding to $K = 0$. This is a natural requirement, at least for the case where the investor does not receive a minimum income from non-financial sources (labor). The set of all consumption and investment strategies that satisfy all these requirements on the interval $[t, T]$ is denoted by $A_t$.

Preferences: The objective is to maximize the expected life-time utility which is assumed to be on the additively time-separable form

$$ E \left[ \int_0^T e^{-\delta(t-s)}u(c_s) \, ds + e^{-\delta(T-t)}\bar{u}(W_T) \right], \quad (5.1) $$

where $u$ and $\bar{u}$ are increasing and concave von Neumann-Morgenstern utility functions. We will assume that $u$ and $\bar{u}$ are twice continuously differentiable on their domain. We will define the indirect utility process $J_t = (J_t)$ as

$$ J_t = \sup_{(c, \theta) \in A_t} E_t \left[ \int_t^T e^{-\delta(s-t)}u(c_s) \, ds + e^{-\delta(T-t)}\bar{u}(W_T) \right]. \quad (5.2) $$

An optimal consumption and investment strategy $(c^*, \theta^*)$ has the property that it provides at least as high an expected life-time utility as any other feasible strategy. In particular,

$$ J_0 = E \left[ \int_0^T e^{-\delta t}u(c_t^*) \, dt + e^{-\delta T}\bar{u}(W_T^*) \right], $$

where $W_T^*$ is the terminal wealth level that follows from the strategy $(c^*, \theta^*)$. In other words, when an optimal strategy exists the supremum in the definition of $J$ is attained. Of course, $J_0$ will depend on the initial wealth $W_0$ of the investor. We shall assume that $J_0 < \infty$ for all $W_0 < \infty$. It can be shown that $J_0$ is an increasing and concave function of initial wealth $W_0$. See Exercise 5.1 at the end of the chapter.

Dynamics of prices and wealth: When the investor is about to choose consumption and investment strategies she has to deal with a number of variables that can evolve stochastically over time such as:

- the (locally) risk-free rate $r_t$ (i.e., the short-term interest rate),
- the prices, the expected rates of returns, the variance-covariance matrix of rates of return on the risky assets,
- the expected rate of change and variation in her income rate,
- covariances or correlations between all these variables.

Of course, in a fuller model we should also include uncertainty e.g., about the time of death of the investor, relative prices of different consumption goods, etc., but we ignore such issues at this point.

than just being adapted.
5.2 The basic continuous-time setting

We shall assume that all exogenous shocks to these variables can be represented by standard Brownian motions. A direct consequence is that we do not allow for any jumps in prices, except for points in time where the asset provides its owner with a lump-sum payment, e.g., a dividend payment of a stock or a coupon payment of a bond.\(^\text{2}\) For simplicity, we assume that the assets provide no payments in the life of the investor and that the vector of risky asset prices \(P_t\) follows a stochastic process of the form

\[
dP_t = \text{diag}(P_t) \left[ \mu_t dt + \sigma_t \, d\mathbf{z}_t \right],
\]

where \(\mathbf{z} = (z_1, \ldots, z_d)^\top\) is a \(d\)-dimensional standard Brownian motion, i.e., a vector of \(d\) independent one-dimensional standard Brownian motions. The term \(\text{diag}(P_t)\) denotes the \((d \times d)\)-matrix with the vector \(P_t\) along the main diagonal and zeros off the diagonal. We can write this componentwise as

\[
dP_{it} = P_{it} \left[ \mu_{it} \, dt + \sum_{j=1}^d \sigma_{ijt} \, dz_{jt} \right], \quad i = 1, \ldots, d.
\]

The instantaneous rate of return on asset \(i\) is given by \(dP_{it}/P_{it}\). The \(d\)-vector \(\mu_t = (\mu_{1t}, \ldots, \mu_{dt})^\top\) contains the expected rates of return and the \((d \times d)\)-matrix \(\sigma_t = (\sigma_{ijt})_{i,j=1}^d\) measures the sensitivities of the risky asset prices with respect to exogenous shocks so that the \((d \times d)\)-matrix \(\sigma_t\sigma_t^\top\) contains the variance and covariance rates of instantaneous rates of return. We assume that \(\sigma_t\) is non-singular. Of course, \(\mu\) and \(\sigma\) must be adapted to the information filtration \(\mathbb{F} = (\mathcal{F}_t)\).\(^\text{3}\) This way of modeling price dynamics in continuous-time can be seen as the limit of (4.4) when \(\varepsilon_{t+\Delta t}\) in that expression is assumed to be multivariate standard normally distributed.

Taking the limit of the wealth dynamics in (4.5) we get

\[
dW_t = \left[ \theta_0^0 r_t + \theta_t^\top \mu_t + y_t - c_t \right] \, dt + \theta_t^\top \sigma_t \, d\mathbf{z}_t.
\]

The amount invested in the (locally) risk-free asset can be expressed as total wealth minus the amounts invested in the risky assets,

\[
\theta_0^0 = W_t - \theta_t^\top \mathbf{1}.
\]

Substituting this into the wealth dynamics above, we obtain

\[
dW_t = \left[ r_t W_t + \theta_t^\top (\mu_t - r_t \mathbf{1}) + y_t - c_t \right] \, dt + \theta_t^\top \sigma_t \, d\mathbf{z}_t.
\]

Since \(\sigma_t\) is assumed to be a non-singular \((d \times d)\)-matrix, we can define the \(d\)-dimensional process \(\lambda = (\lambda_t)\) by

\[
\lambda_t = \sigma_t^{-1} (\mu_t - r_t \mathbf{1}),
\]

so that

\[
\mu_t = r_t \mathbf{1} + \sigma_t \lambda_t,
\]

i.e., \(\mu_t = r_t + \sum_{j=1}^d \sigma_{ijt} \lambda_{jt}\). \(\lambda\) has the interpretation of a vector of market prices of risk (corresponding to the shock process \(\mathbf{z}\)) since it measures the excess rate of return relative to the standard

\(^{2}\)See, e.g., Bardhan and Chao (1995), Wu (2006), and Jeanblanc-Picqué and Pontier (1990) for utility maximization problems involving jump processes.

\(^{3}\)Further technical requirements should be imposed, e.g., that the processes \(r, \mu,\) and \(\sigma\) are progressively measurable, that \(\text{diag}(P_t)\mu_t\) is an \(\mathcal{L}^1\)-process, and that \(\text{diag}(P_t)\sigma_t\) is an \(\mathcal{L}^2\)-process; cf. footnote 1.
deviation. For example, if asset $i$ is only sensitive to the first component of the exogenous shock $z_t$, it will have $\sigma_{it} = \cdots = \sigma_{idt} = 0$ and hence an expected rate of return of $\mu_{it} = r_t + \sigma_{i1t}\lambda_{it}$ so that $\lambda_{it} = (\mu_{it} - r_t)/\sigma_{i1t}$, where $\sigma_{i1t}$ is identical to the volatility of the asset. We can now rewrite the price dynamics as

$$dP_t = \text{diag}(P_t) \left[ (r_t1 + \varpi_{t}\lambda_{t}) \right] dt + \varpi_{t} dz_t.$$ 

The wealth dynamics can be rewritten as

$$dW_t = \left[ r_t W_t + \theta_t^\top \varpi_{t}\lambda_{t} + y_t - c_t \right] dt + \theta_t^\top \varpi_{t} dz_t. \tag{5.4}$$

In terms of the portfolio weights $\pi$, the wealth dynamics can be written as

$$dW_t = W_t \left[ r_t + \pi_t^\top \varpi_{t}\lambda_{t} \right] dt + [y_t - c_t] \ dt + W_t \pi_t^\top \varpi_{t} dz_t. \tag{5.5}$$

**Solution techniques:** There are two major questions to be answered: (i) Under which assumptions do optimal strategies exist, and (ii) How can optimal strategies (and the indirect utility function) be computed. In these notes we will focus on the second question. There are two major approaches for solving this type of optimization problems: the dynamic programming approach (also known as the stochastic control approach) and the martingale approach. In the following section we consider the dynamic programming approach, while the martingale approach is introduced in Section 8.1.

### 5.3 Dynamic programming in continuous-time models

In Section 4.3 we introduced the dynamic programming approach in a discrete-time multi-period setting. Apparently, Merton (1969, 1971) was the first to apply the dynamic programming approach to a continuous-time optimal consumption/investment problem. The dynamic programming approach requires that a (possibly multi-dimensional) state variable exists so that this variable follows a Markov process and all relevant objects can be written as functions of this state variable and time. The theory of dynamic programming contains some results on the existence of optimal strategies, but they often require that all admissible strategies take values in a compact set, an assumption which is certainly unsuitable for most portfolio problems. Therefore, verification theorems are typically applied. This involves solving the so-called Hamilton-Jacobi-Bellman (HJB) equation associated with the control problem. Under some technical conditions the solution to the HJB equation will give us both the optimal strategies and the indirect utility function. The HJB equation is a fully non-linear second-order partial differential equation. Despite the complexity of the equation, explicit solutions have been found in many interesting settings, as we shall see in the following chapters.

Surely we must include the wealth $W_t$ of the agent as a state variable and then look for a process $x = (x_t)$, possibly multi-dimensional, such that the pair $(W_t, x_t)$ captures all relevant information for the agent’s decision at time $t$. Basically, the pair of stochastic processes $(W, x)$ must constitute a Markov system, for any given consumption-portfolio choice $(c, \pi)$. If both $r$, $\lambda$, $\varpi$, and $y$ are constant (or at least deterministic functions of time), then the wealth process is by itself a Markov process and we need not add some $x$. We will refer to this situation as the case of *constant investment opportunities*. We study portfolio and consumption choice under that assumption in detail in Chapter 6. However, we do know that for example the short-term interest
rate varies stochastically over time. If \( r = (r_t) \) is in itself a Markov process, we should include \( r \) as a state variable, i.e., one of the elements of \( x \) should be \( r \). Maybe multiple state variables are needed to capture the interest rate dynamics. Then these variables should be included in \( x \). We will study examples of such so-called stochastic investment opportunities in Chapters 7–13.

For simplicity we assume in the following that the agent receives no labor income, i.e., \( y_t \equiv 0 \). We assume further that there is a stochastically evolving state variable \( x = (x_t) \) that captures the variations in \( r, \mu, \) and \( \sigma \) over time, i.e.,

\[
r_t = r(x_t), \quad \mu_t = \mu(x_t, t), \quad \sigma_t = \sigma(x_t, t),
\]

where \( r, \mu, \) and \( \sigma \) now (also) denote sufficiently well-behaved functions. The variations in the state variable \( x \) determine the future expected returns and covariance structure in the financial market. The market price of risk is also given by the state variable:

\[
\lambda(x_t) = \sigma(x_t, t)^{-1} (\mu(x_t, t) - r(x_t) 1).
\]

Note that we have assumed that the short-term interest rate \( r_t \) and the market price of risk vector \( \lambda_t \) do not depend on calendar time directly. The fluctuations in \( r_t \) and \( \lambda_t \) over time are presumably not due to the mere passage of time, but rather due to variations in some more fundamental economic variables. In contrast, the expected rates of returns and the price sensitivities of some assets will depend directly on time, e.g., the volatility and the expected rate of return on a bond will depend on the time-to-maturity of the bond and therefore on calendar time.

For simplicity we will first assume that the state variable is one-dimensional and write it as \( x \). Afterwards we turn to the case of multi-dimensional state variables. The wealth process for a given portfolio and consumption strategy now evolves as

\[
dW_t = W_t \left[ r(x_t) + \pi^T_t \sigma(x_t, t) \lambda(x_t) \right] dt - c_t dt + W_t \pi^T_t \sigma(x_t, t) dz_t.
\]

The state variable \( x \) is assumed to follow a one-dimensional diffusion process

\[
dx_t = m(x_t) dt + \psi(x_t) dz_t + \Phi(x_t) d\tilde{z}_t,
\]

where \( \tilde{z} = (\tilde{z}_t) \) is a one-dimensional standard Brownian motion independent of \( z = (z_t) \). Hence, if \( \Phi(x_t) \neq 0 \), there is an exogenous shock to the state variable that cannot be hedged by investments in the financial market. In other words, the financial market is incomplete. Conversely, if \( \Phi(x_t) \) is identically equal to zero, the financial market is complete. We shall consider examples of both cases later. The \( d \)-vector \( \psi(x_t) \) represents the sensitivity of the state variable with respect to the exogenous shocks to market prices. Note that the \( d \)-vector \( \sigma(x, t) \psi(x) \) is the vector of instantaneous covariance rates between the returns on the risky assets and the state variable.

The pair \((W_t, x_t)\) forms a two-dimensional Markov diffusion process that contains all the information the investor needs for making her consumption/investment decision. The indirect utility at time \( t \) is therefore \( J_t = J(W_t, x_t, t) \), where the function \( J \) is given by

\[
J(W_t, x_t) = \sup_{(c_t, \pi_t) \in [r, T]} E_{W,x,t} \left[ \int_t^T e^{-\delta(s-t)} u(c_s) ds + e^{-\delta(T-t)} \bar{u}(W_T) \right],
\]

where \( E_{W,x,t}[\cdot] \) denotes the expectation given that \( W_t = W \) and \( x_t = x \). In a discrete-time approximation of this setting, it follows from (4.7) that

\[
J(W, x) = \sup_{c_t \geq 0, \pi_t \in \mathbb{R}^d} \left\{ u(c_\Delta) \Delta t + e^{-\delta \Delta t} \bar{u}_{W,x,t} [J(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t)] \right\},
\]
Where \( c_t \) and \( \pi_t \) is held fixed over the interval \([t, t + \Delta t]\). If we multiply by \( e^{\delta \Delta t} \), subtract \( J(W, x, t) \), and then divide by \( \Delta t \), we get

\[
\frac{e^{\delta \Delta t} - 1}{\Delta t} J(W, x, t) = \sup_{c_t \geq 0, \pi_t \in \mathbb{R}^d} \left\{ e^{\delta \Delta t} u(c_t) + \frac{1}{\Delta t} E_{W,x,t} [J(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) - J(W, x, t)] \right\}.
\]

(5.6)

When we let \( \Delta t \to 0 \), we have that (by l'Hôpital's rule)

\[
\frac{e^{\delta \Delta t} - 1}{\Delta t} \to \delta,
\]

and that (by definition of the drift of a process)

\[
\frac{1}{\Delta t} E_{W,x,t} [J(W_{t+\Delta t}, x_{t+\Delta t}, t + \Delta t) - J(W, x, t)]
\]

will approach the drift of \( J \) at time \( t \), which according to Itô's Lemma is given by

\[
\frac{\partial J}{\partial t}(W, x, t) + J_W(W, x, t) \left( W \left[ r(x) + \pi^\top \sigma(x, t) \lambda(x) \right] - c_t \right) + \frac{1}{2} J_{WW}(W, x, t)W^2 \pi^\top \sigma(x, t) \sigma(x, t)^\top \pi + J_x(W, x, t)m(x) + \frac{1}{2} J_{xx}(W, x, t)(v(x)^\top v(x) + \hat{v}(x)^2) + J_{Wx}(W, x, t)W \pi^\top \sigma(x, t)v(x).
\]

(5.7)

This is called the Hamilton-Jacobi-Bellman (HJB) equation corresponding to the dynamic optimization problem. Subscripts on \( J \) denote partial derivatives, however we will write the partial derivative with respect to time as \( \partial J/\partial t \) to distinguish it from the value \( J_t \) of the indirect utility process. The HJB equation involves the supremum over the feasible time \( t \) consumption rates and portfolios \( \text{(not the supremum over the entire processes!)} \) and is therefore a highly non-linear second-order partial differential equation.

Note that we can split up the maximization over \( c \) and \( \pi \) into separate maximization terms and rewrite the HJB equation (5.7) as

\[
\delta J(W, x, t) = L^c J(W, x, t) + L^\pi J(W, x, t) + \frac{\partial J}{\partial t}(W, x, t) + r(x)W J_W(W, x, t) + J_x(W, x, t)m(x) + \frac{1}{2} J_{xx}(W, x, t)(v(x)^\top v(x) + \hat{v}(x)^2),
\]

(5.8)

where

\[
L^c J(W, x, t) = \sup_{c_t \geq 0} \left\{ u(c_t) - c_t J_W(W, x, t) \right\},
\]

\[
L^\pi J(W, x, t) = \sup_{\pi_t \in \mathbb{R}^d} \left\{ W J_W(W, x, t)\pi^\top \sigma(x, t) \lambda(x) + \frac{1}{2} J_{WW}(W, x, t)W^2 \pi^\top \sigma(x, t) \sigma(x, t)^\top \pi + J_{Wx}(W, x, t)W \pi^\top \sigma(x, t)v(x) \right\}.
\]
From the analysis above we will expect that the indirect utility function $J(W,x,t)$ solves the HJB equation for all possible values of $W$ and $x$ and all $t \in [0,T)$ and that it satisfies the terminal condition

$$J(W,x,T) = \bar{u}(W)$$  \hspace{1cm} (5.9)

for all $W$ and $x$. In the mathematical literature on stochastic control problems like the one we are looking at, there are a few results concerning when a solution to the HJB equation exists. However, these results are only valid under restrictive conditions, e.g., that the controls ($c$ and $\pi$ in our case) can only take values in a compact set. This is generally not true for the consumption/investment problems. We are mostly interested in finding a solution. Here, we can apply a verification result.

Let us formulate the result for the problem with a one-dimensional state variable:

**Theorem 5.1.** Assume that $V(W,x,t)$ solves the HJB equation (5.8) with the terminal condition (5.9) and satisfies some technical conditions. Let $C(W,x,t)$ and $\Pi(W,x,t)$ be given by

$$C(W,x,t) = \arg \max_{c \geq 0} \{ u(c) - cV_W(W,x,t) \} ,$$

$$\Pi(W,x,t) = \arg \max_{\pi \in \mathbb{R}^d} \left\{ WV_W(W,x,t)\pi^\top \lambda(x) + \frac{1}{2} VW(W,x,t)W^2 \pi^\top \sigma(x,t)\sigma(x,t)^\top \pi + W_{xW}(W,x,t)W\pi^\top \sigma(x,t)\sigma(x,t)\pi + V_{Wx}(W,x,t)W \pi^\top \sigma(x,t)v(x) \right\}$$

If the strategies $c^*_t = C(W^*_t,x_t,t), \quad \pi^*_t = \Pi(W^*_t,x_t,t), \quad$ where $(W^*_t)$ is the wealth process that $(c^*, \pi^*)$ induces, are feasible (i.e., $(c, \pi) \in A_0$), then they are optimal, and $V$ equals the indirect utility function, i.e.

$$J(W,x,t) = V(W,x,t) = E_{W,x,t} \left[ \int_t^T e^{-\delta(s-t)} u(c^*_s) \, ds + e^{-\delta(T-t)} \bar{u}(W^*_T) \right].$$

The verification theorem suggests a two-step procedure. First, solve the maximization problem embedded in the HJB-equation giving a candidate for the optimal strategies expressed in terms of the yet unknown indirect utility function and its derivatives. Second, substitute the candidate for the optimal strategies into the HJB-equation, ignore the sup-operator, and solve the resulting partial differential equation for $J(W,x,t)$. Such a solution will then also give the candidate optimal strategies in terms of $W$, $x$, and $t$. However, there is really also a third step, namely to check that the assumptions made along the way and the technical conditions needed for the verification theorem to apply are all satisfied. The standard version of the verification theorem is precisely stated and proofed in Øksendal (2003) or Fleming and Soner (1993). The technical conditions of the standard version are not always satisfied in concrete consumption-portfolio problems, however, but at least for some concrete problems a version with an appropriate set of conditions can be found; see, e.g., Korn and Kraft (2001) and Kraft (2009). In the current version of these lecture notes, we will generally ignore these technicalities and trust that a suitable verification theorem applies.

Suppose now that the state variable $x$ is $k$-dimensional and follows the diffusion process

$$dx_t = m(x_t) \, dt + v(x_t)^\top dz_t + \tilde{v}(x_t) \, d\tilde{z}_t,$$
where \( m \) now is a \( k \)-vector valued function, \( v \) is a \((d \times k)\)-matrix valued function, and \( \hat{z} \) is a \( k \)-dimensional standard Brownian motion independent of \( z \). The basic derivation is the same as with a one-dimensional state variable, but the drift of \( J \) now becomes more complicated and so does the HJB equation:

\[
\delta J(W, x, t) = \mathcal{L}^c J(W, x, t) + \mathcal{L}^\pi J(W, x, t) + \frac{\partial J}{\partial t} (W, x, t) + r(x)W J_W(W, x, t)
\]

\[
+ J_x(W, x, t)\Tr m(x) + \frac{1}{2} \text{tr} \left( J_{xx}(W, x, t) [v(x)]^\top [v(x)] + [\hat{v}(x)]^\top [\hat{v}(x)]^\top \right),
\]

where

\[
\mathcal{L}^c J(W, x, t) = \sup_{c \geq 0} \{ u(e) - c J_W(W, x, t) \},
\]

\[
\mathcal{L}^\pi J(W, x, t) = \sup_{\pi \in \mathbb{R}^d} \left\{ W J_W(W, x, t) \pi^\top \sigma(x) \lambda(x) + \frac{1}{2} J_{WW}(W, x, t) W^2 \pi^\top \sigma(x) \sigma(x) \pi \right\}
\]

Now, \( J_x \) and \( J_{WW} \) are \( k \)-vectors and \( J_{xx} \) is a \((k \times k)\)-matrix. The notation \( \text{tr}(A) \) stands for the trace of the square matrix \( A = (A_{ij}) \), which is defined as the sum of the diagonal elements, \( \text{tr}(A) = \sum_i A_{ii} \).

In the special case of constant investment opportunities, the indirect utility is given by \( J_t = J(W_t, t) \) and the corresponding HJB equation is simply

\[
\delta J(W, t) = \mathcal{L}^c J(W, t) + \mathcal{L}^\pi J(W, t) + \frac{\partial J}{\partial t} (W, t) + r W J_W(W, t)
\]

with

\[
\mathcal{L}^c J(W, t) = \sup_{c \geq 0} \{ u(e) - c J_W(W, t) \},
\]

\[
\mathcal{L}^\pi J(W, t) = \sup_{\pi \in \mathbb{R}^d} \left\{ W J_W(W, t) \pi^\top \lambda + \frac{1}{2} J_{WW}(W, t) W^2 \pi^\top \sigma \sigma^\top \pi \right\}.
\]

The terminal condition is

\[
J(W, T) = \bar{u}(W).
\]

In the next chapter we study this case in detail.

### 5.4 Loss from suboptimal strategies

The utility induced by the application of any given admissible strategy \((c, \pi)\) from time \( t \) on is

\[
V_t^{c, \pi} = \mathbb{E}_t \left[ \int_t^T e^{-\delta(s-t)} u(c_s) ds + e^{-\delta(T-t)} \bar{u}(W_T^{c, \pi}) \right],
\]

where \( W_T^{c, \pi} \) is the terminal wealth generated by the strategy \((c, \pi)\). Suppose the dynamics of the investment opportunities is captured by a one-dimensional diffusion \( x = (x_t) \) and that the strategy at any time \( s \) at most depends on wealth \( W_s \), on time, and on \( x_s \). Then \( V_t^{c, \pi} = V^{c, \pi}(W_t, x_t, t) \).

By definition, the application of a suboptimal strategy \((c, \pi)\) leads to a lower level of utility, i.e.,

\[
V^{c, \pi}(W_t, x_t, t) \leq J(W_t, x_t, t) \equiv V^{c, \pi}(W_t, x_t, t) + V^{c, \pi}(W_t, x_t, t).
\]

---

4In this multi-dimensional setting it would be natural to write the \( dz_t \)-term in the state dynamics on the form \( v(x_t) dz_t \), but this would conflict with our notation in the one-dimensional case, where we have used the term \( v(x_t)^\top dz_t \).
If we want to measure how bad the strategy \((c, \pi)\) is compared to the optimal strategy, we cannot just use the distance in utility \(J(W_t, x_t, t) - V^{c, \pi}(W_t, x_t, t)\) since that distance is not stable to positive affine transformation of the utility function. A better measure is the wealth-equivalent percentage loss \(\ell_t\) defined implicitly by
\[
V^{c, \pi}(W_t, x_t, t) = J(W_t[1 - \ell_t], x_t, t).
\]
(5.11)

We can interpret \(\ell_t\) as the percentage of time \(t\) wealth that the individual is willing to sacrifice in order to be able to apply the optimal strategy \((c^*, \pi^*)\) instead of the strategy \((c, \pi)\) from time \(t\) on. Of course, \(\ell_t\) depends on \((c, \pi)\) and generally also on \(W_t, x_t,\) and \(t\), i.e., \(\ell_t = \ell^{c, \pi}(W_t, x_t, t)\).

An equivalent measure would be the percentage of extra wealth, \(\tilde{\ell}_t\), needed to obtain the same utility with the suboptimal strategy \((c, \pi)\) as with the optimal strategy, i.e.,
\[
V^{c, \pi}(W_t[1 + \tilde{\ell}_t], x_t, t) = J(W_t, x_t, t).
\]
Again, \(\tilde{\ell}_t = \tilde{\ell}^{c, \pi}(W_t, x_t, t)\).

5.5 Exercises

**Exercise 5.1.** Show that the indirect utility, \(J_t\), defined in (5.2) is an increasing and concave function of wealth, \(W_t\). *Hint:* To show concavity, let \((c_1, \theta_1)\) be the optimal strategy with initial wealth \(W_1\) and let \((c_2, \theta_2)\) be the optimal strategy with initial wealth \(W_2\). Here, \(c_i\) is the consumption rate and \(\theta_i\) the vector of dollar amounts invested in the risky assets. The corresponding terminal wealth levels are denoted \(W_{1T}\) and \(W_{2T}\), respectively. For any \(\alpha \in (0, 1)\), you should first show that the strategy \((\alpha c_1 + (1-\alpha)c_2, \alpha \theta_1 + (1-\alpha)\theta_2)\) is a feasible strategy with initial wealth \(\alpha W_{1T} + (1-\alpha)W_{2T}\) that results in the terminal wealth \(\alpha W_{1T} + (1-\alpha)W_{2T}\). Then apply that \(u\) and \(\bar{u}\) are assumed concave.

---

5 An early example of calculations of the monetary costs associated with suboptimal intertemporal behavior was given by Cochrane (1989).
Asset allocation with constant investment opportunities

6.1 Introduction

In this chapter we will consider the relatively simple case in which the short-term interest rate \( r \), the expected rates of return \( \mu \), and the volatility matrix \( \sigma \) of the risky assets are all assumed to be constant through time. The market price of risk vector \( \lambda \) is therefore also a constant. We shall also assume that the investor has no income other than the returns on the financial investments, i.e., \( y = 0 \). This is the problem originally considered by Merton (1969). A direct consequence of these additional assumptions is that the risky asset price processes in (5.3) become geometric Brownian motions so that future risky asset prices are lognormally distributed, as is well-known from the Black-Scholes model for stock option pricing; see, e.g., Hull (2009). In this case the wealth dynamics for a given consumption strategy \( c \) and a given portfolio weight process \( \pi \) is

\[
dW_t = \left( W_t \left[ r + \pi_t \lambda \right] - c_t \right) dt + W_t \pi_t \sigma dW_t,
\]

(6.1)

and the indirect utility function (sometimes called the value function) is a function of only current wealth and time

\[
J(W, t) = \sup_{(c, \pi) \in \mathcal{H}} \mathbb{E}_W \left[ \int_t^T e^{-\delta(s-t)} u(c_s) ds + e^{-\delta(T-t)} \bar{u}(W_T) \right],
\]

where \( \mathbb{E}_W \) denotes the expectations operator given \( W_t = W \) (and given the chosen consumption and investment strategies).

We will first attack this problem applying the dynamic programming approach and try to solve the HJB equation associated with the utility maximization problem. From (5.10), we have that the HJB equation is given by

\[
\delta J(W, t) = \mathcal{L}^c J(W, t) + \mathcal{L}^\pi J(W, t) + \frac{\partial J}{\partial t}(W, t) + rW J_W(W, t)
\]

(6.2)
with

$$L^c J(W, t) = \sup_{c \geq 0} \{u(c) - cJ_W(W, t)\},$$  \hspace{1cm} \text{(6.3)}

$$L^\pi J(W, t) = \sup_{\pi \in \mathbb{R}^d} \left\{ W J_W(W, t) \pi^\top \lambda + \frac{1}{2} J_W(W, t) W^2 \sigma\sigma^\top \pi \right\}.$$  \hspace{1cm} \text{(6.4)}

The terminal condition is

$$J(W, T) = \bar{u}(W).$$

In Section 6.2 we will see how far we can get for a general utility function. Then in Sections 6.3 and 6.4 we specialize to CRRA and logarithmic utility, respectively, for which explicit solutions can be obtained (in Section 8.2 we derive the same results using the martingale approach). Section 6.6 discusses how wealth, investments, and consumption vary over the life-cycle. In Section 6.5 we analyze further the optimal investment strategy for the CRRA investors. Section 6.7 explains how to quantify the loss from following a suboptimal strategy. Finally, Section 6.8 considers the importance of the frequency of portfolio rebalancing.

### 6.2 General utility function

We will try to solve our consumption and investment problem by an application of the verification theorem, Theorem 5.1, i.e., by solving the HJB equation (6.2). The first-order condition for the maximization in (6.3) leads to

$$u'(c) = J_W(W, t),$$

where we have used the fact that the non-negativity constraint on consumption will not be binding under the assumption that marginal utility is infinite for zero consumption (or even at a positive subsistence level of consumption). This optimality condition is called the \textit{envelope condition}, which we also derived in a discrete-time framework in Chapter 4, cf. Equation (4.10). The condition says that the marginal utility from currently consuming one unit more must equal the marginal utility from investing that unit optimally. This is an intuitive optimality condition for intertemporal choice. If we let $I_u$ denote the inverse of marginal utility $u'(c)$, we can write our candidate for the optimal consumption strategy as

$$c^*_t = C(W^*_t, t),$$

where

$$C(W, t) = I_u(J_W(W, t)).$$  \hspace{1cm} \text{(6.5)}

Substituting the maximizing $c$ back into (6.3), we get

$$L^c J(W, t) = u(I_u(J_W(W, t))) - I_u(J_W(W, t))J_W(W, t).$$

The first-order condition for the (unconstrained) maximization in (6.4) leads to

$$J_W(W, t) W \sigma \lambda + J_W(W, t) W^2 \sigma \sigma^\top \pi = 0.$$  \hspace{1cm} \text{(6.6)}

Isolating $\pi$, we get

$$\pi = -\frac{J_W(W, t)}{W J_W(W, t) \sigma^\top} (\sigma^\top)^{-1} \lambda,$$
so that our candidate for the optimal investment strategy can be written as

$$\pi^*_t = \Pi(W^*_t, t),$$

where

$$\Pi(W, t) = -\frac{J_W(W, t)}{W J_{WW}(W, t)} (\sigma^\top)^{-1} \lambda = -\frac{J_W(W, t)}{W J_{WW}(W, t)} (\sigma \sigma^\top)^{-1} (\mu - r 1). \quad (6.6)$$

Note that the fraction $-J_W(W, t)/[W J_{WW}(W, t)]$ is the relative risk tolerance (i.e., the reciprocal of the relative risk aversion) of the indirect utility function. The optimal risky investment is therefore given by the relative risk tolerance of the investor times a vector that is the same for all investors (assuming they have the same perceptions about $\sigma$, $\mu$, and $r$), namely the inverse of the variance-covariance matrix multiplied by the vector of excess expected rates of return. The second-order conditions for a maximum are satisfied since $J$ is concave in $W$ and $u$ is concave in $c$. Substituting the maximizing $\pi$ back into (6.4) and simplifying, we get

$$L^\pi J(W, t) = -\frac{1}{2} \|\lambda\|^2 \frac{J_W(W, t)^2}{J_{WW}(W, t)},$$

where $\|\lambda\|^2 = \lambda^\top \lambda$.

The HJB equation is thus transformed into the second order PDE

$$\delta J(W, t) = u(I_u(J_W(W, t))) - J_W(W, t) I_u(J_W(W, t)) + \frac{\partial J}{\partial t}(W, t)$$
$$+ r W J_W(W, t) - \frac{1}{2} \|\lambda\|^2 \frac{J_W(W, t)^2}{J_{WW}(W, t)}. \quad (6.7)$$

If this PDE has a solution $J(W, t)$ such that the strategy defined by (6.5) and (6.6) is feasible (satisfies the technical conditions), then we know from the verification theorem that this strategy is indeed the optimal consumption and investment strategy and the function $J(W, t)$ is indeed the indirect utility function. We shall sometimes consider problems with no utility from intermediate consumption, i.e., $u \equiv 0$. In that case, it is of course optimal not to consume, and it is relatively easy to see that the first two terms of the right-hand side of (6.7) will vanish, i.e., the equation simplifies to

$$\delta J(W, t) = \frac{\partial J}{\partial t}(W, t) + r W J_W(W, t) - \frac{1}{2} \|\lambda\|^2 \frac{J_W(W, t)^2}{J_{WW}(W, t)}.$$

In the following sections we shall obtain simple, closed-form solutions for problems with CRRA and logarithmic utility. In Exercise 6.4 at the end of the chapter we will consider the problem with a subsistence HARA utility function, where a simple solution also can be obtained. Semi-explicit solutions for other utility functions have been given by Karatzas, Lehoczky, Sethi, and Shreve (1986). Merton (1971, Sec. 6) claimed to have found a solution for the general class of HARA functions but as noted by Sethi and Takas (1988), this solution does not satisfy the non-negativity constraints on wealth and consumption.

Without further computations we can already note an important result: With constant $r$, $\mu$, and $\sigma$, 

**two-fund separation** obtains in the continuous-time setting. This is obvious from the optimal investment strategy in (6.6).

**Theorem 6.1 (Two-fund separation).** In a financial market with constant $r$, $\mu$, and $\sigma$, the optimal investment strategy of any unconstrained investor with time-separable utility of the form (5.1) and
no non-financial income is a combination of the risk-free asset and a single portfolio of risky assets
given by the weights

\[
\pi^{\text{tan}} = \frac{1}{1^{\top}(\sigma^{\top})^{-1}\lambda^{\top}} - 1^{\top}\frac{1}{1^{\top}(\sigma^{\top})^{-1}(\mu - r 1^{\top})^{\top}}(\sigma^{\top})^{-1}(\mu - r 1^{\top}).
\] (6.8)

The investor will invest the fraction

\[
\frac{J W(W,t)}{W(W,t)} \pi^{\text{tan}} 1^{\top}\frac{1}{1^{\top}(\sigma^{\top})^{-1}{\lambda}^{\top}} - 1^{\top}(\sigma^{\top})^{-1}(\mu - r 1^{\top})
\]
of her wealth in the risky fund and the remaining wealth in the risk-free asset.

The portfolio \(\pi^{\text{tan}}\) is almost indistinguishable from the tangency portfolio (3.19) of the one-period
mean-variance analysis, but in the continuous-time case the relevant expected rates of return and
variances and covariances are measured over the next infinitesimal period of time. With this little
modification of the interpretation we can again look at the investment problem graphically in a
(standard deviation, mean)-diagram as we are used to from the static one-period setting. Also, we
again have the conclusion that all investors should hold risky assets in the same proportion, i.e.,
\(\pi_i / \pi_j\) is the same for all investors. Note that the necessary assumption of lognormal prices is much
more realistic than the normality assumption in the one-period model. Analogous to the one-
period setting, the two-fund separation result above is the basis for a capital market equilibrium
result, which in the continuous-time case is referred to as the Intertemporal Capital Asset Pricing
Model (ICAPM) or the Continuous-time CAPM; see, e.g., Merton (1973b), Duffie (2001), Cochrane
(2005), and Munk (2012) for more on equilibrium asset pricing.

6.3 CRRA utility function

We will now focus on the case where the utility function exhibits constant relative risk aversion.
We are interesting in three types of problems:

1. utility from consumption only,
2. utility from terminal wealth only,
3. utility both from consumption and terminal wealth.

We can solve all three problems simultaneously by introducing two non-negative coefficients \(\varepsilon_1\) and
\(\varepsilon_2\) and letting

\[
u(c) = \varepsilon_1 \frac{c^{1-\gamma}}{1-\gamma}, \quad \bar{u}(W) = \varepsilon_2 \frac{W^{1-\gamma}}{1-\gamma}.
\]

Situation (1) above corresponds to \(\varepsilon_2 = 0\) and \(\varepsilon_1 > 0\). The exact value of \(\varepsilon_1\) has no impact on
optimal decisions, but \(\varepsilon_1 = 1\) would be the natural choice as notation is then simpler. Similarly,
situation (2) corresponds to \(\varepsilon_1 = 0\) and \(\varepsilon_2 > 0\) with \(\varepsilon_2 = 1\) being the natural choice (in that
case we can disregard discounting and put \(\delta = 0\)). Finally, situation (3) requires both \(\varepsilon_1 > 0\) and
\(\varepsilon_2 > 0\). The ratio \(\varepsilon_2/\varepsilon_1\) determines the relative importance of terminal wealth and intermediate
consumption and will therefore in general affect the optimal decisions, but we could fix one of the
coefficients (to 1, for example) without loss of generality. In order to encompass all three situations,
we will allow for general \(\varepsilon_1 \geq 0\) and \(\varepsilon_2 \geq 0\) with \(\varepsilon_1 + \varepsilon_2 > 0\). The indirect utility function is

\[
J(W,t) = \sup_{(c,\pi)\in[\tau,T]} E_{W,t}\left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} \frac{c^{1-\gamma}}{1-\gamma} ds + \varepsilon_2 e^{-\delta(T-t)} \frac{W_{T}^{1-\gamma}}{1-\gamma} \right].
\]
The marginal utility for consumption is \( u'(c) = \varepsilon_1 c^{-\gamma} \). If \( \varepsilon_1 > 0 \), marginal utility has the inverse function \( I_u(a) = \varepsilon_1^{1/\gamma} a^{-1/\gamma} \). Consequently, we have that

\[
I_u(a) = \frac{\varepsilon_1^{1/\gamma} a^{-1/\gamma}}{1 - \gamma}
\]

and

\[
u(I_u(a)) = \varepsilon_1 \frac{I_u(a)^{1-\gamma}}{1-\gamma} = \varepsilon_1^{1/\gamma} a^{1-1/\gamma}.
\]

The first two terms on the right-hand side of Eq. (6.7) are thus equal to \( \varepsilon_1^{1/\gamma} \frac{a^{1-1/\gamma}}{1-\gamma} \). Substituting into (6.9) and gathering terms, we get

\[
\delta J(W, t) = \varepsilon_1^{1/\gamma} \frac{\gamma}{1-\gamma} J_W(W, t)^{1-\gamma} + \frac{\partial J}{\partial t} (W, t) + r W J_W(W, t) - \frac{1}{2} \| \lambda \|^2 \frac{J_W(W, t)^2}{J_W(W, t)}.
\]

(6.9)

The terminal condition is that \( J(W, T) = \varepsilon_2 W^{1-\gamma} / (1 - \gamma) \).

Due to the linearity of the wealth dynamics in (6.1) it seems reasonable to conjecture that if the strategy \((c^*, \pi^*)\) is optimal with time \( t \) wealth \( W \) and the corresponding wealth process \( W^* \), then the strategy \((k_{c^*}, \pi^*)\) will be optimal with time \( t \) wealth \( kW \) and the corresponding wealth process \( kW^* \). If this is true, then

\[
J(kW, t) = E_t \left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} \frac{(kc^*_s)^{1-\gamma}}{1-\gamma} ds + \varepsilon_2 e^{-\delta(T-t)} \frac{(kW^*_s)^{1-\gamma}}{1-\gamma} \right]
\]

\[
= k^{1-\gamma} E_t \left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} \frac{(c^*_s)^{1-\gamma}}{1-\gamma} ds + \varepsilon_2 e^{-\delta(T-t)} \frac{(W^*_s)^{1-\gamma}}{1-\gamma} \right]
\]

\[
= k^{1-\gamma} J(W, t),
\]

i.e., the indirect utility function \( J(W, t) \) is homogeneous of degree \( 1 - \gamma \) in the wealth \( W \). Inserting \( k = 1/W \) and rearranging, we get

\[
J(W, t) = \frac{g(t)^\gamma W^{1-\gamma}}{1-\gamma},
\]

where \( g(t)^\gamma = (1 - \gamma) J(1, t) \). From the terminal condition \( J(W, T) = \varepsilon_2 W^{1-\gamma} / (1 - \gamma) \), we have that \( g(T)^\gamma = \varepsilon_2 \), hence \( g(T) = \varepsilon_2^{1/\gamma} \).

The relevant derivatives of our guess \( J(W, t) \) are

\[
J_W(W, t) = g(t)^\gamma W^{-\gamma}, \quad J_{WW}(W, t) = -\gamma g(t)^\gamma W^{-\gamma - 1},
\]

\[
\frac{\partial J}{\partial t} (W, t) = \frac{\gamma}{1-\gamma} g(t)^{\gamma - 1} g'(t) W^{1-\gamma}.
\]

Substituting into (6.9) and gathering terms, we get

\[
\left\{ \left( \frac{\delta}{1-\gamma} - r - \frac{1}{2\gamma} \| \lambda \|^2 \right) g(t) - \frac{\varepsilon_1^{1/\gamma} \gamma}{1-\gamma} - \gamma g'(t) \right\} g(t)^{\gamma - 1} W^{1-\gamma} = 0.
\]

Since this equation should hold for all \( W \) and all \( t \in [0, T] \), the term in the brackets must be equal to zero for all \( t \), i.e., the function \( g \) must satisfy the ordinary differential equation

\[
g'(t) = Ag(t) - \varepsilon_1^{1/\gamma}
\]

(6.10)
with the terminal condition $g(T) = \varepsilon_2^{1/\gamma}$. Here $A$ is the constant

$$A = \frac{\delta + r(\gamma - 1)}{\gamma} + \frac{1}{2} \frac{\gamma - 1}{\gamma^2} \| \lambda \|^2$$

which we assume is different from zero. It can be checked that the solution is given by

$$g(t) = \frac{1}{A} \left( \varepsilon_1^{1/\gamma} + \left[ \varepsilon_2^{1/\gamma} A - \varepsilon_1^{1/\gamma} \right] e^{-A(T-t)} \right),$$

We will generally assume that the relative risk aversion $\gamma$ exceeds 1 and that $\delta$ and $r$ are non-negative, and in that case we have $A > 0$.

Let us show that $g(t) \geq 0$ for all $t \in [0, T]$. It is sufficient to demonstrate that the function $G(t) = \frac{1}{A} \left( \varepsilon_1^{1/\gamma} + \left[ \varepsilon_2^{1/\gamma} A - \varepsilon_1^{1/\gamma} \right] e^{-A\tau} \right)$ is non-negative for all $\tau \geq 0$. Note that $G(0) = \varepsilon_2^{1/\gamma} > 0$ and $G'(\tau) = (\varepsilon_1^{1/\gamma} - \varepsilon_2^{1/\gamma} A) e^{-A\tau}$. We split the analysis into three cases:

1. Suppose $\varepsilon_1^{1/\gamma} = \varepsilon_2^{1/\gamma} A$. Since $\varepsilon_1$ and $\varepsilon_2$ are not allowed both to be zero, this case is only possible if both $\varepsilon_1$ and $\varepsilon_2$ are strictly positive. The function $G$ is then constant, $G(\tau) = \varepsilon_1^{1/\gamma} / A = \varepsilon_2^{1/\gamma} > 0$ for all $\tau$.

2. Suppose $\varepsilon_1^{1/\gamma} > \varepsilon_2^{1/\gamma} A$. Then $G'(\tau) > 0$ for all $\tau$ so that $G$ is monotonically increasing and, since $G(0) \geq 0$, we have $G(\tau) > 0$ for $\tau > 0$. For $A > 0$, the limit is $\lim_{\tau \to \infty} G(\tau) = \varepsilon_1^{1/\gamma} / A > \varepsilon_2^{1/\gamma}$. For $A < 0$, $G(\tau) \to \infty$ for $\tau \to \infty$.

3. Suppose $\varepsilon_1^{1/\gamma} < \varepsilon_2^{1/\gamma} A$. Since both $\varepsilon_1$ and $\varepsilon_2$ are non-negative, this can only happen if $A > 0$.

We have $G'(\tau) < 0$ so that $G$ is monotonically decreasing, but the limit $\lim_{\tau \to \infty} G(\tau) = \varepsilon_1^{1/\gamma} / A$ is non-negative. Hence, $G(\tau)$ stays non-negative.

We summarize our findings in the following theorem:

**Theorem 6.2.** Assume that the constant $A$ defined in (6.11) is different from zero. For the CRRA utility maximization problem in a market with constant $r$, $\mu$, and $\sigma$, we then have that the indirect utility function is given by

$$J(W, t) = \frac{g(t)^{\gamma W^{1-\gamma}}}{1 - \gamma}$$

with

$$g(t) = \frac{1}{A} \left( \varepsilon_1^{1/\gamma} + \left[ \varepsilon_2^{1/\gamma} A - \varepsilon_1^{1/\gamma} \right] e^{-A(T-t)} \right).$$

The optimal investment strategy is given by

$$\Pi(W, t) = \frac{1}{\gamma} (\sigma \sigma^\top)^{-1} \lambda = \frac{1}{\gamma} (\sigma \sigma^\top)^{-1} (\mu - r 1).$$

If the agent has utility from intermediate consumption ($\varepsilon_1 > 0$), her optimal consumption rate is

$$C(W, t) = \varepsilon_1^{1/\gamma} \frac{W}{g(t)} = A \left( 1 + \left[ (\varepsilon_2/\varepsilon_1)^{1/\gamma} A - 1 \right] e^{-A(T-t)} \right)^{-1} W.$$

---

For $A = 0$, the ODE (6.10) simplifies to $g'(t) = -\varepsilon_1^{1/\gamma}$ which with the terminal condition $g(T) = \varepsilon_2^{1/\gamma}$ has the solution $g(t) = \varepsilon_2^{1/\gamma} + \varepsilon_1^{1/\gamma} (T - t)$. 
A similar result was first demonstrated by Merton (1969).

The optimal consumption strategy is to consume a time-varying fraction of wealth. It is easy to show that when \( \varepsilon_2 > 0 \), the consumption/wealth ratio approaches \( \frac{\varepsilon_1}{\varepsilon_2} \) as \( t \to T \), whereas \( c/W \to \infty \) for \( t \to T \) when \( \varepsilon_2 = 0 \).

The higher the risk aversion coefficient \( \gamma \), the lower the investment in the risky assets and the higher the investment in the risk-free asset. The optimal investment strategy is independent of the horizon of the investor. The fraction of wealth invested in each asset is to be kept constant over time. Note that this requires continuous rebalancing of the portfolio since the prices of individual assets vary all the time. Consider an asset which enters the optimal portfolio with a positive weight. If the price of this asset increases more than the prices of the other assets in the portfolio, the fraction of wealth made up by that asset will increase. Hence, the investor should reduce the number of units of that particular asset. So the optimal investment strategy is a “sell winners, buy losers” strategy. The fact that this asset has given a high return in the previous period has no consequence for the optimal position in that asset since the distribution of future returns is assumed to be constant over time. If the investor does not sell a recent winner stock, he will be too exposed to the risk of that stock.

Inserting the optimal strategy into the general expression for the dynamics of wealth, we find that

\[
dW^*_t = W^*_t \left[ \left( r + \frac{1}{\gamma} \|\lambda\|^2 - \varepsilon_1^{1/\gamma} g(t)^{-1} \right) dt + \frac{1}{\gamma} \lambda^T dZ_t \right].
\]

Therefore, optimal wealth evolves as a geometric Brownian motion (although with a time-dependent drift). Future values of wealth are lognormally distributed. In particular, wealth stays positive. The optimal strategy is to be further analyzed in Exercise 6.1 at the end of the chapter.

For the case where the agent only gets utility from terminal wealth (\( \varepsilon_1 = 0, \varepsilon_2 = 1 \) and \( \delta = 0 \)), the function \( g \) reduces to \( g(t) = e^{-(T-t)A} \) and

\[
A = \frac{\gamma - 1}{\gamma} \left( r + \frac{1}{2\gamma} \|\lambda\|^2 \right).
\]

Hence, the indirect utility function can be written as

\[
J(W, t) = \frac{1}{1-\gamma} e^{-(T-t)A^{1-\gamma}} W^{1-\gamma} = \frac{1}{1-\gamma} e^{-\left(\gamma - 1\right)\left(r + \frac{1}{2\gamma} \|\lambda\|^2\right)(T-t)} W^{1-\gamma}.
\]

The optimal investment strategy is unaltered. Exactly the same portfolio should be held whether or not the agent has utility from intermediate consumption. With constant investment opportunities and time-additive CRRA utility there is no clear link between investment and consumption. Of course, wealth will evolve differently over time if the agent withdraws money for consumption. Consequently, ceteris paribus, the value of the portfolio and the number of units held of the different assets will be different (smaller) with utility from intermediate consumption.

### 6.4 Logarithmic utility

The solution for the case of logarithmic utility is obtained by a similar procedure. This is the subject of Exercise 6.2 at the end of the chapter. The indirect utility function is here defined as

\[
J(W, t) = \sup_{(c, \pi) \in [t, T]} \mathbb{E}_{W, t} \left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} \ln c_s \, ds + \varepsilon_2 e^{-\delta(T-t)} \ln W_T \right].
\]

The result is:
Theorem 6.3. For the logarithmic utility maximization problem in a market with constant $r$, $\mu$, and $\sigma$, we have that the indirect utility function is given by

$$J(W, t) = g(t) \ln W + b(t),$$

with

$$g(t) = \frac{1}{\delta} \left( \varepsilon_1 + \left[ \varepsilon_2 \delta - \varepsilon_1 \right] e^{-\delta(T-t)} \right)$$

(6.14)

and, for $t < T$,

$$b(t) = \left( r + \frac{1}{2} \|\lambda\|^2 - \delta \right) \left( \frac{\varepsilon_1}{\delta^2} - e^{-\delta(T-t)} \left[ \frac{\varepsilon_1}{\delta} (T-t) - \varepsilon_2 (T-t) \right] \right) - g(t) \ln g(t).$$

The optimal investment strategy is given by

$$\Pi(W, t) = (\sigma^\top)^{-1} \lambda = (\sigma \sigma^\top)^{-1} (\mu - r 1),$$

and if the agent has utility from intermediate consumption ($\varepsilon_1 > 0$) the optimal consumption strategy is

$$C(W, t) = \varepsilon_1 g(t)^{-1} W = \delta \left( 1 + \left[ (\varepsilon_2/\varepsilon_1) \delta - 1 \right] e^{-\delta(T-t)} \right)^{-1} W.$$ 

Note that if we take the limit of $g(t)$ defined in Eq. (6.12) as $\gamma \to 1$, we get the expression given in Eq. (6.14). Also note that the optimal strategy for the logarithmic utility case can be obtained by taking limits of the optimal strategy for the CRRA case as $\gamma \to 1$.

6.5 Discussion of the optimal investment strategy for CRRA utility

Many empirical studies have documented that in the past century long-term stock investments have in most cases outperformed (i.e., have given a higher return than) a long-term bond investment. Over short investment horizons, the dominance of stock investments is less clear. Referring to these empirical facts, many investment consultants recommend that long-term investors should place a large part of their wealth in stocks and then gradually shift from stocks to bonds as they get older and their investment horizon shrinks. This recommendation conflicts with the optimal portfolio strategy we have derived above. According to our analysis, the optimal portfolio weights of CRRA investors are independent of the investment horizon. Is this because our model of the financial asset prices is inconsistent with the empirical facts mentioned before? The answer is no. To see this let us consider the simplest case with a single stock (representing the stock index) with price dynamics

$$dP_t = P_t [\mu dt + \sigma dz_t],$$

where $\mu$ and $\sigma$ as well as the interest rate $r$ are constants. In other words, the price process is a geometric Brownian motion. This implies that

$$P_T = P_0 e^{(\mu - \frac{1}{2} \sigma^2) T + \sigma z_T}.$$
Since \( z_T \sim N(0, T) \), the probability that a stock investment outperforms a risk-free investment over a period of \( T \) years is equal to

\[
\Pr\left( \frac{P_T}{P_0} > e^{rT} \right) = \Pr\left( \left( \mu - \frac{1}{2} \sigma^2 \right) T + \sigma z_T > rT \right)
\]

\[
= \Pr\left( z_T > -\frac{(\mu - r - \frac{1}{2} \sigma^2) T}{\sigma} \right)
\]

\[
= \Pr\left( z_T < \frac{(\mu - r - \frac{1}{2} \sigma^2) T}{\sigma} \right)
\]

\[
= N\left( \frac{(\mu - r - \sigma^2/2) \sqrt{T}}{\sigma} \right),
\]

where \( N(\cdot) \) is the cumulative distribution function for a standard normally distributed random variable.

Figure 6.1 illustrates the relation between the outperformance probability and the investment horizon. The curves differ with respect to the presumed expected rate of return on the stock, i.e., \( \mu \), whereas the interest rate is 4% and the volatility of the stock is 20% for all curves. Empirical studies indicate that U.S. stocks over a 100-year period have had an average excess rate of return of 8-9% per year. A \( \mu \)-value of 15% corresponds to an expected excess rate of return of 9% per year since \( 0.15 - 0.04 - (0.20)^2/2 = 0.09 \). However, it should be emphasized that historical estimates of expected rates of return, volatilities, and correlations are not necessarily good predictors of the future values of these quantities. In particular, the value of the excess expected rate of return on the stock market is frequently discussed both among practitioners and academics. There are several reasons to believe that the average return on the US stock market over the past century is higher than what the stock market is currently offering in terms of expected returns. This discussion is also closely linked to the so-called equity premium puzzle. See, e.g., Mehra and Prescott (1985), Weil (1989), Welch (2000), and Mehran (2003), Shiller (2000), and Ibbotson and Chen (2003). Probably the curves labeled \( \mu = 9\% \) and \( \mu = 12\% \) are more representative of the current investment opportunities. In any case, it is tempting to conclude from the graph that long-term investors should invest more in stocks than short-term investors. Why does the optimal portfolio derived previously not reflect this property?

It is important to realize that the optimal decision cannot be based just on the probabilities of gains and losses. After all most individuals will reject a gamble with a 99\% probability of winning 1 dollar and a 1\% probability of losing a million dollars. The magnitudes of gains and losses are also important for the optimal investment decision. Let us look at the probability that a stock investment will provide a return which is \( K \) percentage points lower than a risk-free investment over the same period, i.e.,

\[
\Pr\left( \frac{P_T}{P_0} < e^{rT} - K \right) = \Pr\left( \left( \mu - \frac{1}{2} \sigma^2 \right) T + \sigma z_T < \ln (e^{rT} - K) \right)
\]

\[
= \Pr\left( z_T < \frac{\ln (e^{rT} - K) - (\mu - \frac{1}{2} \sigma^2) T}{\sigma} \right)
\]

\[
= N\left( \frac{\ln (e^{rT} - K) - (\mu - \frac{1}{2} \sigma^2) T}{\sigma \sqrt{T}} \right).
\]

Table 6.1 shows such probabilities for various combinations of the return shortfall constant \( K \).
and the investment horizon. (The numbers in the row labeled 0% are equal to 100% minus the outperformance probabilities shown in Figure 6.1.) Over a 10-year period the return on a risk-free investment at a rate of 4% per year is

\[(e^{0.04 \cdot 10} - 1) \cdot 100\% \approx 49.1\%.

The table shows that with a 22.2% probability a stock investment over a 10-year period will give a return which is lower than 49.1% − 25% = 24.1%, and there is a 5.7% probability that the stock return will be lower than 49.1% − 75% = −25.9%. Over a 40-year period the risk-free return is 395%. There is a 13% probability that a stock investment will give a return which is at least 100 percentage points lower, i.e., lower than 295%. Over longer periods the probability that stocks underperform bonds is lower, but the probability of extremely bad stock returns is larger than over short periods. The expected excess return on the stock increases with the length of the investment horizon, but so does the variance of the return. Any risk-averse investor has to consider this trade-off. For a CRRA investor in our simple financial model, the two effects offset each other exactly so that the optimal portfolio is independent of the investment horizon.

6.6 The life-cycle

Let us look at how wealth, consumption, and investments vary over the life-cycle. Of course, these quantities all depend on the future shocks to the prices of the financial assets and thus to the wealth of the individual, but we can compute the expected future wealth, consumption, and investment given the initial wealth.

First, consider consumption. Optimal consumption at time \( t \) is given in terms of wealth and
6.6 The life-cycle

<table>
<thead>
<tr>
<th>Excess return on bond</th>
<th>1 year</th>
<th>10 years</th>
<th>40 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>44.0%</td>
<td>31.8%</td>
<td>17.1%</td>
</tr>
<tr>
<td>25%</td>
<td>6.4%</td>
<td>22.2%</td>
<td>16.1%</td>
</tr>
<tr>
<td>50%</td>
<td>0.0%</td>
<td>13.1%</td>
<td>15.1%</td>
</tr>
<tr>
<td>75%</td>
<td>0.0%</td>
<td>5.7%</td>
<td>14.0%</td>
</tr>
<tr>
<td>100%</td>
<td>0.0%</td>
<td>1.3%</td>
<td>13.0%</td>
</tr>
</tbody>
</table>

Table 6.1: Underperformance probabilities. The table shows the probability that a stock investment over a period of 1, 10, and 40 years provides a percentage return which is at least 0, 25, 50, 75, or 100 percentage points lower than the risk-free return. The numbers are computed using the parameter values \( \mu = 9\% \), \( r = 4\% \), and \( \sigma = 20\% \).

With the wealth dynamics in (6.13), the consumption dynamics follows from an application of Itô’s Lemma

\[
dc^*_t = \varepsilon_1^{1/\gamma} \frac{W^*_t}{g(t)} dt
\]

With the wealth dynamics in (6.13), the consumption dynamics follows from an application of Itô’s Lemma

\[
dc^*_t = \varepsilon_1^{1/\gamma} \frac{W^*_t}{g(t)} dt = c^*_t \left( r + \frac{1}{\gamma} \| \lambda \|^2 - A \right) dt + \frac{1}{\gamma} \lambda^\top dz_t
\]

where we have applied (6.10) and (6.11). Consequently, optimal consumption is a geometric Brownian motion. In particular, the initial expectation of the future consumption is (see properties of the geometric Brownian motion in Section B.8.1 of the appendix)

\[
E[c^*_0] = c^*_0 \exp \left\{ \frac{1}{\gamma} \left( r - \delta + \frac{\gamma + 1}{2\gamma} \| \lambda \|^2 \right) t \right\} = W_0 \frac{A}{1 + \left( \varepsilon_2 / \varepsilon_1 \right)^{1/\gamma} A - 1} e^{-\lambda^\top dz} \exp \left\{ \frac{1}{\gamma} \left( r - \delta + \frac{\gamma + 1}{2\gamma} \| \lambda \|^2 \right) t \right\}.
\]

Clearly, consumption is expected to increase with age, decrease with age, or to be age-independent depending on whether \( r - \delta + \frac{\gamma + 1}{2\gamma} \| \lambda \|^2 \) is positive, negative, or zero. With realistic parameters, the constant is positive so that consumption should increase, on average, over life.

Empirical studies show a hump-shaped consumption pattern over the life-cycle (Browning and Crossley 2001, Gourinchas and Parker 2002) so that consumption typically increases up to around age 40-45 and then drops throughout the rest of life. The simple model considered in this chapter cannot generate such a pattern. In fact, the more advanced models with closed-form solutions that we will look at in subsequent chapters cannot match the hump either. Several explanations of the hump have been suggested in the literature, including mortality risk (Hansen and Imrohoroglu 2008, Feigenbaum 2008), borrowing constraints (Thurow 1969, Gourinchas and Parker 2002), and endogenous labor supply with a hump-shaped wage profile (Bullard and Feigenbaum 2007). However, none of these additional features would preserve the explicitness of our solutions in this.
Numerical solutions that include mortality risk and borrowing constraints in a setting with labor income can generate the consumption hump, cf., for example, Cocco, Gomes, and Maenhout (2005). Next, consider wealth. From (6.13) it is clear that expected future wealth is

\[
E[W^*_t] = W^*_0 \exp \left\{ \left( r + \frac{1}{\gamma} \| \lambda \|^2 \right) t - \xi_{1/\gamma}^1 \int_0^t \frac{1}{g(u)} \, du \right\},
\]

and it can be shown that

\[
\xi_{1/\gamma}^1 \int_0^t \frac{1}{g(u)} \, du = A t - \ln \left( \frac{1 + [\epsilon_2/\epsilon_1]^{1/\gamma} A - 1}{1 + [\epsilon_2/\epsilon_1]^{1/\gamma} A - 1} e^{-A[T-t]} \right)
\]

so that

\[
E[W^*_t] = W^*_0 \exp \left\{ \frac{1}{\gamma} \left( r - \delta + \frac{1}{2\gamma} \| \lambda \|^2 \right) t \right\} \frac{1 + [\epsilon_2/\epsilon_1]^{1/\gamma} A - 1}{1 + [\epsilon_2/\epsilon_1]^{1/\gamma} A - 1} e^{-A[T-t]} - A.
\]

One can show that the sign of the derivative \( \partial E[W^*_t] / \partial t \) is equal to the sign of

\[
\left( r + \frac{1}{\gamma} \| \lambda \|^2 \right) \left( 1 + [\epsilon_2/\epsilon_1]^{1/\gamma} A - 1 \right) e^{-A[T-t]} = A.
\]

For the special case with no utility of terminal wealth, \( \epsilon_2 = 0 \), the sign will be negative at least for \( t \) very close to \( T \), which makes sense since in that case the individual will consume all wealth before the terminal date. More generally, the behavior of \( E[W^*_t] \) over life depends both on the relative weights on consumption and terminal wealth, on the time preference rate and relative risk aversion (\( \delta \) affects \( A \)), and on the investment opportunities (via \( r \) and \( \| \lambda \|^2 \)).

The expected amounts invested in the financial assets in the future is simply \( \frac{1}{\gamma} (\sigma^\top)^{-1} \lambda E[W^*_t] \) which obviously follows the same life-cycle pattern as wealth itself.

### 6.7 Loss due to suboptimal investments

In the section we want to assess the importance of getting the portfolio exactly right, so we disregard consumption and put \( \delta = 0, \epsilon_1 = 0, \) and \( \epsilon_2 = 1 \). We focus on the case with a single risky asset in addition to the riskfree asset. For any fixed portfolio weight \( \pi \) in the risky asset, the wealth dynamics will be

\[
dW_t^\pi = W_t^\pi \left[ (r + \pi \sigma \lambda) \, dt + \pi \sigma \, dz_t \right],
\]

so that wealth follows a geometric Brownian motion. It can be shown (see Exercise 6.3) that the expected utility for a given \( \pi \) is

\[
V^\pi(W, t) \equiv E_t \left[ \frac{1}{1-\gamma} (W_t^\pi)^{1-\gamma} \right] = \frac{1}{1-\gamma} (g^\pi(t))^{\gamma} W^{1-\gamma}, \tag{6.15}
\]

\footnote{Labor supply flexibility is limited and thus induces constraints that, like borrowing constraints, prevent closed-form solutions. Mortality risk effectively implies an increasing time preference rate over life which may produce a consumption hump, but it also adds unspanned risk to the labor income impeding the computation of human wealth in closed form, unless the investor can purchase full insurance against the loss of income in case of death (Kraft and Steffensen 2008). However, the actual demand for such insurance contracts is much smaller than a theoretical model would suggest, even for the simple constant-income life annuities relevant in retirement as reflected by the discussion of the so-called annuity puzzle (Davidoff, Brown, and Diamond 2005, Inkmann, Lopes, and Michaelides 2011).}
6.8 Infrequent rebalancing of the portfolio

The optimal investment strategy with CRRA utility and constant investment opportunities is to keep a fixed portfolio weight in each asset. However, that requires continuous rebalancing of the portfolio as the prices of the different assets do not move in parallel. Continuous rebalancing is not practically possible. Moreover, even with tiny trading costs per transaction, continuous rebalancing
Chapter 6. Asset allocation with constant investment opportunities

Figure 6.3: Welfare losses for different investment horizons. The figure shows the percentage wealth-equivalent utility loss $\ell_t^{\pi}$ from applying a suboptimal constant portfolio weight instead of the optimal portfolio weight. The loss is depicted as a function of the suboptimal portfolio weight with different curves for different investment horizons $T-t$. The relative risk aversion is $\gamma = 2$, the Sharpe ratio of the stock is $\lambda = 0.3$, and the volatility of the stock is $\sigma = 0.2$.

would be infinitely expensive. It is therefore interesting to see how bad it is to rebalance in a non-continuous way. Let us disregard consumption in the following considerations and assume a single risky asset.

A very simple strategy is to predetermine a finite number of trading dates. At each trading date the portfolio is rebalanced so that the portfolio weights coincide with the solution for the continuous time case. In between trading dates, the portfolio weights will deviate from the truly optimal weights. Suppose that $\Delta t > 0$ is the time period between any two adjacent trading dates. Suppose the portfolio is rebalanced at time $t$ so that the total wealth $W_t$ is split into the amount $\pi W_t$ invested in the stock and the amount $(1 - \pi) W_t$ in the riskfree asset. The gross return on the stock until the next rebalancing is

$$
\frac{S_{t+\Delta t}}{S_t} = \exp \left\{ \left( r + \sigma \lambda - \frac{1}{2} \sigma^2 \right) \Delta t + \sigma (z_{t+\Delta t} - z_t) \right\},
$$

and the gross return on the riskfree investment is $\exp \{ r \Delta t \}$. The wealth at time $t + \Delta t$ is therefore

$$
W_{t+\Delta t} = \pi W_t \exp \left\{ \left( r + \sigma \lambda - \frac{1}{2} \sigma^2 \right) \Delta t + \sigma (z_{t+\Delta t} - z_t) \right\} + (1 - \pi) W_t \exp \{ r \Delta t \}
$$

$$
= W_t e^{r \Delta t} \left\{ 1 + \pi \left[ \exp \left\{ \left( \sigma \lambda - \frac{1}{2} \sigma^2 \right) \Delta t + \sigma (z_{t+\Delta t} - z_t) \right\} - 1 \right] \right\}.
$$

Seen at time $t$, the only random variable on the right-hand side is $z_{t+\Delta t} - z_t \sim N(0, \Delta t)$. The discrete rebalancing strategy can be evaluated by Monte Carlo simulation.³ The wealth can be simulated forward using the above relation by replacing $z_{t+\Delta t} - z_t$ by $\varepsilon_{t+\Delta t} \sqrt{\Delta t}$, where $\varepsilon_{t+\Delta t}$

³Monte Carlo simulation is described in most derivatives textbooks, e.g., Hull (2009) and Munk (2011).
6.8 Infrequent rebalancing of the portfolio

is a draw from the standard normal distribution, \( N(0, 1) \), with independent draws for different time steps as the increments to the standard Brownian motion over non-overlapping intervals are independent.\(^4\) We can generate a simulated value of the terminal wealth \( W_T \) and compute the utility \( u(W_T) = \frac{1}{1-\gamma} W_T^{1-\gamma} \). By generating a large number, \( M \), of samples \( W_T^m \) of terminal wealth, we can take the average utility as an approximation of the expected utility of terminal wealth for this discrete rebalancing strategy:

\[
E[u(W_T)] \approx \frac{1}{M} \sum_{m=1}^{M} u(W_T^m)
\]

We can then compare that (approximation of the) expected utility with the value function and compute a percentage wealth-equivalent loss \( \ell \), as defined in (5.11) and used above.

As an example, assume \( r = 0.02, \sigma = 0.2, \) and \( \lambda = 0.3 \), and consider an investor with a relative risk aversion of \( \gamma = 2 \) and an investment horizon of \( T - t = 10 \) years. The optimal strategy is to have \( \pi = 0.75 = 75\% \) of the wealth invested in the stock at any point in time. If we fix initial wealth to 1, the indirect utility will be \(-0.65377\). In a Monte Carlo simulation procedure implemented in Microsoft Excel, 2000 “antithetic” pairs of terminal wealth were simulated using quarterly rebalancing.\(^5\) The average utility was \(-0.65547\), which corresponds to a wealth-equivalent loss of only 0.26% (in Exercise 6.5 you are asked to do similar experiments). This experiment indicates that it is not important to rebalance the portfolio very frequently. Between two adjacent rebalancing dates the portfolio weight of the stock can deviate somewhat from the optimal weight, but the deviation is typically rather small, and we have already seen in the previous section that expected utility is relatively insensitive to small deviations from the optimal strategy.

Rogers (2001) provides a more formal analysis of the impact of infrequent portfolio rebalancing. Branger, Breuer, and Schlag (2010) perform a detailed Monte Carlo simulation study, also for some models with stochastic investment opportunities that we will discuss in later chapters. Their study

\(^4\)Some spreadsheet applications, programming environments, and other software tools may have a built-in procedure for generating such draws, but not all of them are of a good quality, i.e., if you use the procedure for generating a number of such draws, the distribution of these draws may be quite different from the standard normal distribution. Alternatively, you can generate draws from the \( N(0, 1) \) distribution by transforming draws from a uniform distribution on the unit interval, a distribution we will denote by \( U[0, 1] \). Most computer tools used for financial applications have a built-in generator of random numbers from the \( U[0, 1] \) distribution, but there are also algorithms for generating these draws that can easily be implemented in any programming environment, cf., e.g., Press, Teukolsky, Vetterling, and Flannery (2007, Ch. 7). A popular choice is the so-called Box-Muller transformation suggested by Box and Muller (1958). Given two draws \( U_1 \) and \( U_2 \) from the uniform \( U[0, 1] \) distribution, \( \varepsilon_1 \) and \( \varepsilon_2 \) defined by

\[
\varepsilon_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2), \quad \varepsilon_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)
\]

are two independent draws from the standard normal distribution. An alternative approach is to transform a draw \( U \) from the \( U[0, 1] \) distribution into a draw \( \varepsilon \) from the \( N(0, 1) \) distribution by

\[
\varepsilon = N^{-1}(U),
\]

where \( N^{-1}(\cdot) \) denotes the inverse of the probability distribution function \( N(\cdot) \) associated with the standard normal distribution, i.e., \( N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) dz \). This follows from the fact that \( P(\varepsilon < a) = P(N^{-1}(U) < a) = P(U < N(a)) = N(a) \). Of course, this approach requires an implementation of the inverse normal distribution \( N^{-1}(\cdot) \), which is not known in closed form. Again, some software tools (such as Microsoft Excel) have a built-in algorithm for computing the inverse normal distribution, but the precision of the algorithm is generally unknown to the user, and the computation is bound to be more time-consuming than when using the Box-Muller transformation.

\(^5\)The idea of antithetic variates is explained in most textbook presentations of Monte Carlo simulation, including Hull (2009) and Munk (2011).
confirms that for investment problems involving only stocks and bonds, relatively infrequent rebalancing induces small wealth-equivalent losses. However, when derivatives are included, frequent rebalancing is sometimes important.

6.9 Exercises

Exercise 6.1. Consider the optimal consumption and investment strategy for a CRRA investor (with no labor income) in a market with constant \( r, \mu, \) and \( \sigma \), cf. Theorem 6.2. How does the optimal strategy depend on time and the parameters of the model? (You may assume that only one risky asset is traded.)

Exercise 6.2. Give a proof of Theorem 6.3.

Exercise 6.3. Verify the expressions (6.15) and (6.16). Try to create figures like Figures 6.2–6.3. Show that the alternative loss measure \( \tilde{\ell}_t \) under the given assumptions becomes

\[
\tilde{\ell}_t = e^{\frac{1}{2}(\lambda - \gamma \pi \sigma)^2 (T-t)} - 1 \approx \frac{1}{2\gamma}(\lambda - \gamma \pi \sigma)^2 (T-t),
\]

so that the two loss measures are approximately the same for small deviations from the optimal strategy.

Exercise 6.4. Assume a financial market with a constant risk-free rate \( r \) and risky assets with constant \( \mu \) and \( \sigma \). Consider an investor with no income from non-financial sources and an indirect utility function

\[
J(W,t) = \sup_{(c_s, \pi_s)_{s \in [t,T]}} E_{W,t} \left[ \int_t^T e^{-\delta (s-t)} u(c_s) \, ds \right],
\]

where \( u \) now is a subsistence HARA function,

\[
u(c) = \frac{(c - \bar{c})^{1-\gamma}}{1-\gamma}
\]

with \( \bar{c} \) being the subsistence level of consumption. What is the optimal consumption and investment strategy for this investor? Compare with the standard CRRA solution. Hint: How do you invest to finance the subsistence level of consumption in the rest of your life? What is the cost of that investment? The remaining wealth can be invested “freely”.

Exercise 6.5. Implement a Monte Carlo simulation to study the impact of infrequent trading as explained in Section 6.8. Consider an investor with utility of terminal wealth only, a constant relative risk aversion \( \gamma \), and an investment horizon of \( T - t \). The market consists of a risk-free asset with a constant rate of return \( r \) and a single risky asset with volatility \( \sigma \) and a Sharpe ratio \( \lambda \), both assumed constant. Experiment with the frequency of trading, e.g., by considering 1, 4, 12, and 52 trading dates per year. Compute wealth-equivalent losses for the discrete-trading strategies compared to the continuous-time solution. How sensitive is the wealth-equivalent losses to the parameters \( r, \sigma, \lambda, \gamma, \) and \( T - t \)?
CHAPTER 7

Stochastic investment opportunities: the general case

7.1 Introduction

In the previous chapter we analyzed the optimal investment/consumption decision under the assumption of constant investment opportunities, i.e., constant interest rates, expected rates of return, volatilities, and correlations. However, it is well-documented that some, if not all, of these quantities vary over time in a stochastic manner. This situation is referred to as a stochastic investment opportunity set. In this chapter we will study the dynamic investment/consumption choice in a general financial market with stochastic investment opportunities. In later chapters we will then focus on concrete models in which, for example, interest rates or expected excess stock returns follow some specific dynamics.

The main effect of allowing investment opportunities to vary over time is easy to explain. Risk-averse investors with time-additive utility are reluctant to substitute consumption over time, as discussed in Section 2.7. To keep consumption stable across states and time, a (sufficiently) risk-averse investor will therefore choose a portfolio with high positive returns in states with relatively bad future investment opportunities (or bad future labor income) and conversely. This is what is known as intertemporal hedging. The optimal investment strategy will thus be different from the case with constant investment opportunities. From this argument, we also see that there will be a close link between the optimal consumption strategy and the intertemporal hedging part of the optimal investment strategy.

In the rest of this chapter we will formalize these issues in a general modeling framework. We will continue to assume that the investor receives no non-financial income, i.e., no labor income, and refer to Chapter 13 for the extension to the case with labor income. Throughout the chapter we apply the dynamic programming approach, i.e., we focus on solving the Hamilton-Jacobi-Bellman equation associated with the utility maximization problem.
7.2 General utility functions

7.2.1 One-dimensional state variable

As in Section 5.3 we assume that there is a stochastically evolving state variable \( x = (x_t) \) that captures the variations in \( r, \mu, \) and \( \sigma \) over time. The variations in the state variable \( x \) determine the future expected returns and covariance structure in the financial market. For simplicity we will first consider the case where \( x \) is one-dimensional and afterwards turn to the multi-dimensional case.

The dynamics of the \( d \) risky asset prices is in this setting given by

\[
dP_t = \text{diag}(P_t) [\mu(x_t, t) \, dt + \sigma(x_t, t) \, dz_t]
\]

\[
= \text{diag}(P_t) \left[ (r(x_t) \mathbf{1} + \sigma(x_t, t) \lambda(x_t)) \, dt + \sigma(x_t, t) \, dz_t \right].
\]

We assume that \( x \) follows a one-dimensional diffusion process

\[
dx_t = m(x_t) \, dt + \nu(x_t)^\top \, dz_t + \hat{v}(x_t) \, d\hat{z}_t,
\]

where \( \hat{z} \) is a one-dimensional standard Brownian motion independent of \( z \). If \( \hat{v}(x_t) \neq 0 \), the market is incomplete; otherwise, it is complete. Let

\[
\Sigma^x(x) = \nu(x)^\top \nu(x) + \hat{v}(x)^2
\]

denote the instantaneous variance of the state variable. For a given consumption strategy \( c = (c_t) \) and investment strategy \( \pi = (\pi_t) \) the wealth evolves as

\[
dW_t = W_t \left[ r(x_t) + \pi_t^\top \sigma(x_t, t) \lambda(x_t) \right] \, dt - c_t \, dt + W_t \pi_t^\top \sigma(x_t, t) \, dz_t,
\]

and the indirect utility function is defined by

\[
J(W, x, t) = \sup_{(c_t, \pi_t) \in [0, T]} E_{W, x, t} \left[ \int_t^T e^{-\delta(s-t)} u(c_s) \, ds + e^{-\delta(T-t)} \bar{u}(W_T) \right].
\]

The HJB equation associated with this problem is

\[
\delta J(W, x, t) = \mathcal{L}_c J(W, x, t) + \mathcal{L}_\pi J(W, x, t) + \frac{\partial J}{\partial t}(W, x, t) + r(x) W J_W(W, x, t)
+ J_x(W, x, t) m(x) + \frac{1}{2} J_{xx}(W, x, t) \Sigma^x(x),
\]

with the terminal condition \( J(W, x, T) = \bar{u}(W) \). Here

\[
\mathcal{L}_c J(W, x, t) = \sup_{c \geq 0} \left\{ u(c) - c J_W(W, x, t) \right\},
\]

\[
\mathcal{L}_\pi J(W, x, t) = \sup_{\pi \in \mathbb{R}^d} \left\{ W J_W(W, x, t) \pi^\top \sigma(x, t) \lambda(x) + \frac{1}{2} J_{WW}(W, x, t) W^2 \pi^\top \sigma(x, t) \sigma(x, t)^\top \pi
\right. \]

\[
\left. + J_{xW}(W, x, t) W \pi^\top \sigma(x, t) \nu(x) \right\},
\]

The first-order condition with respect to \( c \) is

\[
u'(c) = J_W(W, x, t)
\]

so that the (candidate) optimal consumption strategy is

\[c^*_t = C(W^*_t, x_t, t),\]
and (3) the hedge portfolio given by the weights

$$C(W, x, t) = I_u(J_W(W, x, t))$$  \hfill (7.5)

and, as before, $I_u(\cdot)$ is the inverse of $u(\cdot)$. Substituting the maximizing $c$ back into (7.3), we get

$$\mathcal{L}^c J(W, x, t) = u(I_u(J_W(W, x, t))) - I_u(J_W(W, x, t))J_W(W, x, t).$$  \hfill (7.6)

Note that these relations are exactly as in the case with constant investment opportunities studied in Section 6.2 with the only exception that the indirect utility function now depends on the state variable $x$.

The first-order condition with respect to $\pi$ is different than with constant investment opportunities:

$$J_W(W, x, t)W \sigma(x, t)\lambda(x) + J_{WW}(W, x, t)W^2 \sigma(x, t)\sigma(x, t)\pi + J_{Wx}(W, x, t)W \sigma(x, t)v(x) = 0$$

so that the candidate optimal portfolio is

$$\pi^*_t = \Pi(W^*_t, x_t, t),$$

where

$$\Pi(W, x, t) = -\frac{J_W(W, x, t)}{W J_{WW}(W, x, t)} (\sigma(x, t)\pi) - \frac{J_{Wx}(W, x, t)}{W J_{WW}(W, x, t)} (\sigma(x, t)\pi)^{-1} v(x).$$  \hfill (7.7)

Substituting the maximizing $\pi$ back into (7.4) and simplifying, we get

$$\mathcal{L}^\pi J(W, x, t) = -\frac{1}{2} \|\lambda(x)\|^2 \frac{J_{WW}(W, x, t)^2}{J_{WW}(W, x, t)} - \frac{1}{2} \|v(x)\|^2 \frac{J_{Wx}(W, x, t)^2}{J_{WW}(W, x, t)}$$

$$- v(x)^\top \lambda(x) \frac{J_{Wx}(W, x, t)}{J_{WW}(W, x, t)}. \hfill (7.8)$$

Let us take a closer look at the portfolio (7.7). As the horizon shrinks, the indirect utility function $J(W, x, t)$ approaches the terminal utility function $\bar{u}(W)$ which is independent of the state $x$. Consequently, the derivative $J_{Wx}(W, x, t)$ and hence the last term of the portfolio will approach zero as $t \to T$. In other words, very short-term investors do not hedge. The last term will also disappear for “non-instantaneous” investors in two special cases:

1. $J_{Wx}(W, x, t) \equiv 0$: The state variable does not affect the marginal utility of the investor. As we shall see below this is always true for investors with logarithmic utility. Such an investor is not interested in hedging changes in the state variable.

2. $v(x) \equiv 0$: The state variable is uncorrelated with instantaneous returns on the traded assets.

In this case the investor is not able to hedge changes in the state variable.

In all other cases the state variable induces an additional term to the optimal portfolio relative to the case of constant investment opportunities. From (7.7) we have the following important result:

**Theorem 7.1** (Three-fund separation). All investors will combine (1) the locally risk-free asset (“the bank account”), (2) the tangency portfolio given by the weights

$$\pi^\text{tan}_t = \frac{1}{1^\top (\sigma(x_t, t)^\top)^{-1} \lambda(x_t)} (\sigma(x_t, t)^\top)^{-1} \lambda(x_t),$$

and (3) the hedge portfolio given by the weights

$$\pi^\text{hhd}_t = \frac{1}{1^\top (\sigma(x_t, t)^\top)^{-1} v(x_t)} (\sigma(x_t, t)^\top)^{-1} v(x_t).$$
Note that the composition of the two risky funds varies over time due to fluctuations in the state variable. It is no longer true that all investors will hold different risky assets in the same proportion, i.e., the fractions \( \pi_i/\pi_j \) will be investor-specific since different investors may put different weights on the two portfolios of risky assets. The tangency portfolio has the same interpretation as previously. The position in the portfolio \( \pi^{hdg} \) is the change in the optimal investment strategy due to the stochastic variations in the investment opportunity set, hence the name “hedge portfolio”. The next theorem shows that among all portfolios the hedge portfolio has the maximal absolute correlation with the state variable. In that sense it is the portfolio that is best at hedging changes in the state variable. In a complete market the maximal correlation is one and the hedge portfolio basically replicates the dynamics of the state variable.

**Theorem 7.2.** The absolute value of the instantaneous correlation between the change in the value of an investment strategy and the change in the state variable is maximized for the investment strategy \( \pi_t = \pi^{hdg}_t \).

**Proof.** The value process of an investment strategy \( \pi = (\pi_t) \) has dynamics

\[
dV^\pi_t = V^\pi_t \left( r(x_t) + \pi^\top_t \sigma(x_t, t) \lambda(x_t) \right) dt + V^\pi_t \pi^\top_t \sigma(x_t, t) dz_t.
\]

The instantaneous variance rate is \( (V^\pi_t)^2 \pi^\top_t \sigma(x_t, t) \sigma(x_t, t)^\top \pi_t \), and the instantaneous covariance rate with the state variable is \( V^\pi_t \pi^\top_t \sigma(x_t, t) v(x_t) \). Hence, the square of the instantaneous correlation is

\[
\rho^2 = \frac{(V^\pi_t \pi^\top_t \sigma(x_t, t) v(x_t))^2}{(V^\pi_t)^2 \pi^\top_t \sigma(x_t, t) \sigma(x_t, t)^\top \pi_t} \Sigma^2(x_t)
= \frac{(\pi^\top_t \sigma(x_t, t) v(x_t))^2}{\pi^\top_t \sigma(x_t, t) \sigma(x_t, t)^\top \pi_t} \Sigma^2(x_t).
\]

The portfolio that maximizes \( \rho^2 \) will also maximize the absolute correlation \( |\rho| \). The first-order condition for the maximization implies that

\[
\sigma(x_t, t) v(x_t) (\pi^\top_t \sigma(x_t, t) \sigma(x_t, t)^\top \pi_t) = (\pi^\top_t \sigma(x_t, t) v(x_t)) \sigma(x_t, t) \sigma(x_t, t)^\top \pi_t.
\]

Multiplying through by the inverse of \( \sigma(x_t, t) \sigma(x_t, t)^\top \), we arrive at

\[
(\sigma(x_t, t)^\top)^{-1} v(x_t) (\pi^\top_t \sigma(x_t, t) \sigma(x_t, t)^\top \pi_t) = (\pi^\top_t \sigma(x_t, t) v(x_t)) \pi_t,
\]
which we want to solve for \( \pi_t \). The sum of the elements of the vector on the left-hand side is \( 1^\top (\sigma(x_t, t)^\top)^{-1} v(x_t) (\pi^\top_t \sigma(x_t, t) \sigma(x_t, t)^\top \pi_t) \), while the sum of the elements of the right-hand side vector is \( \pi^\top_t \sigma(x_t, t) v(x_t) \) since \( 1^\top \pi_t = 1 \). Dividing each side by the sum of the elements, we obtain

\[
\frac{(\sigma(x_t, t)^\top)^{-1} v(x_t)}{1^\top (\sigma(x_t, t)^\top)^{-1} v(x_t)} = \pi_t,
\]
as was to be shown. \( \square \)

Let us focus for a moment on the case with a single risky asset so that both \( \sigma(x, t) \) and \( v(x) \) are scalars. The hedge term in \( \pi^*_t \) can then be written as \( -\frac{WJ}{\sigma v} \). Note that \( J_{WW} < 0 \) by concavity. If \( v \) and \( \sigma \) have the same sign, then the return of the risky asset will be positively
correlated with changes in the state variable. In this case we see that the hedge demand on the asset is positive if marginal utility $J_W$ is increasing in $x$ so that $J_{Wx} > 0$. This makes good sense: relative to the situation with a constant investment opportunity set, the agent will devote a larger fraction of wealth to a risky asset that has a high return in states of the world where marginal utility is high. Conversely, if $v$ and $\sigma$ have opposite signs so that they are negatively correlated.

Here is another interpretation of the optimal portfolio strategy, following Ingersoll (1987, p. 282):

**Theorem 7.3.** The optimal portfolio strategy $\pi^*$ is the one that minimizes fluctuations in consumption over time among all portfolio strategies with the same expected rate of return as $\pi^*$.

**Proof.** The expected rate of return on the optimal portfolio in (7.7) is

$$\mu^*(x, t) = r(x) + (\pi^*_t)^\top (\mu(x, t) - r(x)) \mathbf{1}.$$ 

The consumption rate is given by

$$c^*_t = C(W_t, x_t, t).$$

An application of Itô’s Lemma yields

$$dc^*_t = \ldots \, dt + \{C_W(W_t, x_t, t)W_t \pi^*_t \underline{\varepsilon}(x_t, t) + C_x(W_t, x_t, t)\underline{\psi}(x_t)\underline{\psi}^\top(x_t)\} \, dz_t + C_x(W_t, x_t, t)\underline{\psi}(x_t) \, d\tilde{z}_t,$$

where we leave the drift term unspecified and the subscripts on $C$ denote partial derivatives. It follows that the instantaneous variance rate of consumption is equal to

$$\sigma^2_{c_t} \equiv C_W(W, x, t)^2 W^2 \pi^\top \underline{\varepsilon}(x, t) \underline{\varepsilon}^\top(x, t) \pi + C_x(W, x, t)^2 \Sigma^x(x) + 2C_W(W, x, t)C_x(W, x, t)W \pi^\top \underline{\varepsilon}(x, t)\underline{\psi}(x).$$

Now consider the problem of minimizing $\sigma^2_{c_t}$ over all portfolios $\pi$ that have an expected rate of return equal to $\mu^*(x, t)$, i.e., portfolios $\pi$ with $r(x) + \pi^\top \underline{\varepsilon}(x, t)\lambda(x) = \mu^*(x, t)$. Forming the Lagrangian

$$\mathcal{L} = \sigma^2_{c_t} + \psi [\mu^*(x, t) - r(x) - \pi^\top \underline{\varepsilon}(x, t)\lambda(x)]$$

we find the optimality condition

$$\pi^{**} = -\frac{\psi}{2C_W(W, x, t)^2 W^2} \left(\underline{\varepsilon}(x, t)\underline{\varepsilon}^\top(x, t)\right)^{-1} \lambda(x) - \frac{C_x(W, x, t)}{WC_W(W, x, t)} \left(\underline{\varepsilon}(x, t)\underline{\varepsilon}^\top(x, t)\right)^{-1} \underline{\psi}(x).$$

Differentiating the envelope condition $u'(C(W, x, t)) = J_W(W, x, t)$ along the optimal consumption path with respect to $W$ we get

$$u''(C(W, x, t))C_W(W, x, t) = J_{WW}(W, x, t)$$

and by differentiating with respect to $x$ we get

$$u''(C(W, x, t))C_x(W, x, t) = J_{Wx}(W, x, t).$$

Hence,

$$\frac{C_x(W, x, t)}{WC_W(W, x, t)} = \frac{J_{Wx}(W, x, t)}{J_{WW}(W, x, t)}$$

so that the second terms in $\pi^*$ and $\pi^{**}$ are identical. The first term in $\pi^{**}$ is proportional to the first term in $\pi^*$ and since $\pi^{**}$ is chosen such that it has the same expected rate of return as $\pi^*$, the first terms must also coincide. In total, $\pi^{**} = \pi^*$, which was to be shown. \qed
On the other hand, if we minimize the instantaneous variance rate of wealth, i.e., \( \sigma_W^2 = \pi^\top \sigma(x,t) \sigma(x,t)^\top \pi \), over all portfolios \( \pi \) having the same expected rate of return as \( \pi^\star \), we get

\[
\pi^{**} = \psi(\sigma(x,t)^\top) - 1 \lambda(x).
\]

This only involves the tangency portfolio. We can conclude that the investor is concerned about fluctuations over time in consumption, not in wealth.

Above, we discussed the general expressions for the optimal consumption and investment strategy in the presence of a state variable. But these were expressed in terms of the unknown indirect utility function. How do we proceed to find concrete solutions?

Substituting (7.6) and (7.8) back into the HJB equation (7.2) and gathering terms, we get the second order PDE

\[
\delta J(W,x,t) = u(Iu(JW(W,x,t))) - J_W(W,x,t)I_W(JW(W,x,t)) + \frac{\partial J}{\partial W}(W,x,t) + r(x)WJ_W(W,x,t)
\]

\[
- \frac{1}{2} J_{WW}(W,x,t)\|\lambda(x)\|^2 + J_x(W,x,t)m(x) + \frac{1}{2} J_{xx}(W,x,t)\Sigma^x(x)
\]

\[
- \frac{1}{2} J_{WW}(W,x,t)\|v(x)\|^2 - \frac{J_W(W,x,t)J_W(W,x,t)}{J_{WW}(W,x,t)}\lambda(x)^\top v(x).
\]

(7.9)

If this PDE has a solution \( J(W,x,t) \) satisfying the terminal condition \( J(W,x,T) = \bar{u}(W) \) and the strategy defined by (7.5) and (7.7) is feasible (satisfies the technical conditions), then we know from the verification theorem that this strategy is indeed the optimal consumption and investment strategy and the function \( J(W,x,t) \) is indeed the indirect utility function. With no utility from intermediate consumption, i.e., \( u \equiv 0 \), the first two terms of the right-hand side of (7.9) vanish.

Although the PDE (7.9) looks very complicated, closed-form solutions can be found for a number of interesting model specifications as we shall see later in this chapter and in other chapters.

### 7.2.2 Multi-dimensional state variable

Suppose now that the state variable \( x \) is \( k \)-dimensional and follows the diffusion process

\[
dx_t = m(x_t) dt + v(x_t)^\top dz_t + \hat{v}(x_t) \hat{z}_t,
\]

where \( m \) now is a \( k \)-vector valued function, \( v \) is a \((d \times k)\)-matrix valued function, \( \hat{v} \) is a \((k \times k)\)-matrix valued function, and \( \hat{z} \) is a \( k \)-dimensional standard Brownian motion independent of \( z \).

The instantaneous variance-covariance matrix of the state variable is the \((k \times k)\) matrix

\[
\Sigma^x(x) = v(x)^\top v(x) + \hat{v}(x)\hat{v}(x)^\top.
\]

denote the instantaneous variance of the state variable. As explained in Section 5.3, the HJB equation is then

\[
\delta J(W,x,t) = L^cJ(W,x,t) + L^xJ(W,x,t) + \frac{\partial J}{\partial t}(W,x,t) + r(x)WJ_W(W,x,t)
\]

\[
+ J_x(W,x,t)^\top m(x) + \frac{1}{2} \text{tr} \left( J_{xx}(W,x,t)\Sigma^x(x) \right),
\]

\[
\delta J(W,x,t) = L^cJ(W,x,t) + L^xJ(W,x,t) + \frac{\partial J}{\partial t}(W,x,t) + r(x)WJ_W(W,x,t)
\]

\[
+ J_x(W,x,t)^\top m(x) + \frac{1}{2} \text{tr} \left( J_{xx}(W,x,t)\Sigma^x(x) \right),
\]
where
\[
\mathcal{L}^{c} J(W, x, t) = \sup_{c \geq 0} \{ u(c) - c J_W(W, x, t) \},
\]
\[
\mathcal{L}^{\pi} J(W, x, t) = \sup_{\pi \in \mathbb{R}^d} \left\{ W J_W(W, x, t) \pi^\top \underline{g}(x, t) \lambda(x) + \frac{1}{2} J_{WW}(W, x, t) W^2 \pi^\top \underline{g}(x, t) \underline{g}(x, t)^\top \pi \\
+ W \pi^\top \underline{g}(x, t)\underline{v}(x, t) J_{Wx}(W, x, t) \right\}.
\]

Analogously to the case with a one-dimensional state variable discussed in the previous section, the (candidate) optimal consumption strategy is
\[
c^*_t = C(W^*_t, x_t, t),
\]
where
\[
C(W, x, t) = I_u(J_W(W, x, t)),
\]
so that
\[
\mathcal{L}^c J(W, x, t) = u(I_u(J_W(W, x, t))) - I_u(J_W(W, x, t)) J_W(W, x, t).
\]
Likewise, the candidate optimal portfolio is
\[
\pi^*_t = \Pi(W^*_t, x_t, t),
\]
where
\[
\Pi(W, x, t) = -\frac{J_W(W, x, t)}{W J_{WW}(W, x, t)} \left( \underline{g}(x, t)^\top \right)^{-1} \lambda(x) - \left( \underline{g}(x, t)^\top \right)^{-1} \underline{v}(x) \frac{J_{Wx}(W, x, t)}{J_{WW}(W, x, t)},
\]
and
\[
\mathcal{L}^{\pi} J(W, x, t) = -\frac{1}{2} \frac{J_W(W, x, t)^2}{J_{WW}(W, x, t)} \left\| \lambda(x) \right\|^2 - \lambda(x)^\top \underline{v}(x) \frac{J_{Wx}(W, x, t)}{J_{WW}(W, x, t)} - \frac{1}{2} \frac{J_{Wx}(W, x, t)^\top \underline{v}(x)^\top \underline{v}(x) J_{Wx}(W, x, t)}{J_{WW}(W, x, t)}.
\]
We can split up the last term of the optimal portfolio into \(k\) terms, one for each element of the state variable:
\[
\Pi(W, x, t) = -\frac{J_W(W, x, t)}{W J_{WW}(W, x, t)} \left( \underline{g}(x, t)^\top \right)^{-1} \lambda(x) - \sum_{j=1}^{k} \left( \underline{g}(x, t)^\top \right)^{-1} \begin{pmatrix} \underline{v}_1(x) \\ \underline{v}_2(x) \\ \vdots \\ \underline{v}_d(x) \end{pmatrix} J_{Wx_j}(W, x, t) W J_{WW}(W, x, t).
\]
Each of the terms in the sum has the interpretation as a fund hedging changes in one element of the state variable. Therefore, we have \((k + 2)\)-fund separation: all investors are satisfied with access to trade in the risk-free asset, the tangency portfolio, and \(k\) hedge funds.

Substituting \(L^c J\) and \(L^{\pi} J\) back into the HJB equation and gathering terms, we get the second-order PDE
\[
\delta J(W, x, t) = u(I_u(J_W(W, x, t))) - J_W(W, x, t) I_u(J_W(W, x, t)) + \frac{\partial J}{\partial t}(W, x, t) \\
+ r(x) W J_W(W, x, t) - \frac{1}{2} \frac{J_W(W, x, t)^2}{J_{WW}(W, x, t)} \left\| \lambda(x) \right\|^2 + J_x(W, x, t)^\top m(x) \\
+ \frac{1}{2} \operatorname{tr} \left( J_{xx}(W, x, t) \underline{g}(x)^\top \lambda(x) \underline{v}(x) \right) \frac{J_{Wx}(W, x, t)}{J_{WW}(W, x, t)} \\
- \frac{1}{2} \frac{J_{Wx}(W, x, t)^\top \underline{v}(x)^\top \underline{v}(x) J_{Wx}(W, x, t)}{J_{WW}(W, x, t)}.
\]

(7.11)
As before, the first two terms on the right-hand side are not present when the agent has no utility from intermediate consumption.

### 7.2.3 What risks are to be hedged?

It may appear from the analysis above that investors would want to hedge all variables affecting \( r_t, \mu_t, \) and \( \sigma_t \), but this is actually not so. We will show that the only risks the agent will want to hedge are those affecting \( r_t \) and \( \lambda_t \).

Since \( \sigma_t \) and thus \( \sigma_t^\top \) are assumed to be non-singular, we can think of the investor choosing the "volatility vector of wealth" \( \varphi_t = \sigma_t^\top \pi_t \) directly rather than \( \pi_t \). In these terms wealth evolves as

\[
dW_t = W_t \left[ r_t + \varphi_t^\top \lambda_t \right] dt - c_t \, dt + W_t \varphi_t^\top \, dz_t.\]

The indirect utility function is

\[
J_t = \sup_{(\pi, \varphi)} \mathbb{E}_t \left[ \int_t^T e^{-\delta(s-t)} u(c_s) \, ds + e^{-\delta(T-t)} \bar{u}(W_T) \right].
\]

Note that this optimization problem does not involve \( \mu_t \) or \( \sigma_t \). Assuming now that there is a variable \( x_t \) so that

\[
r_t = r(x_t), \quad \lambda_t = \lambda(x_t),
\]

then \( J_t = J(W_t, x_t, t) \) and we can use the dynamic programming approach.

For a multidimensional \( x \) we will get the optimal wealth volatility vector

\[
\varphi_t = -\frac{J_W(W_t, x_t, t)}{W_t J_{WW}(W_t, x_t, t)} \lambda(x_t) - \frac{J_{Wx}(W_t, x_t, t)}{W_t J_{WW}(W_t, x_t, t)}. \]

Hence, the optimal portfolio strategy is

\[
\pi_t = -\frac{J_W(W_t, x_t, t)}{W_t J_{WW}(W_t, x_t, t)} \left( \sigma_t^\top \right)^{-1} \lambda(x_t) - \frac{J_{Wx}(W_t, x_t, t)}{W_t J_{WW}(W_t, x_t, t)}. \]

We can conclude from this analysis that the investor will only hedge the variables that affect the short-term interest rate and the market prices of risk (this is of course only true within the present framework; e.g., an investor with stochastic income will also want to hedge the income risk). Stochastic variations in \( \mu_t \) and \( \sigma_t \) are only interesting to the extent that they cause stochastic variations in the market price of risk! One could imagine a market where volatilities vary stochastically but expected rates of return follow the variations in volatilities so that the market price of risk is constant over time. In such a market no agent would hedge the variations in volatilities and expected rates of return. Similar observations were made by Detemple, Garcia, and Rindisbacher (2003) and Munk and Sørensen (2004). The volatility matrix \( \sigma_t \) of the risky assets becomes relevant when the agent wants to find a portfolio \( \pi_t \) that will generate the desired wealth volatility vector \( \varphi_t \).

In fact, the statement above can be strengthened slightly. Look at the PDE (7.11). Suppose that both \( r \) and \( \| \lambda \|^2 \) are independent of \( x \). Then the function \( J(W, t) \) that satisfies the simple PDE

\[
\delta J(W, t) = u(I_u(J_W(W, t)) - J_W(W, t) I_u(J_W(W, t)) + \frac{\partial J}{\partial t}(W, t)
\]

\[
+ r W J_W(W, t) - \frac{1}{2} J_{WW}(W, t) \| \lambda \|^2.
\]
with \( J(W, T) = \bar{u}(W) \) will also solve the full HJB equation (7.9) as all derivatives with respect to \( x \) will be zero. Consequently, the hedge term in (7.10) disappears. In other words, the investor will only hedge stochastic variations that affect the short-term interest rate \( r_t \) and the squared market prices of risk

\[
\|\lambda_t\|^2 = (\mu_t - r_t 1)^\top (\sigma_t \sigma_t^\top)^{-1} (\mu_t - r_t 1).
\]

Nielsen and Vassalou (2006) show that this result is also true for non-Markov dynamics of prices. We summarize this in the following theorem:

**Theorem 7.4.** Investors with time-additive utility functions and no income from non-financial sources will only hedge stochastic variations in the short-term interest rate \( r_t \) and in the squared market prices of risk \( \|\lambda_t\|^2 \).

There is a very intuitive interpretation of this result, which we can see after a few computations:

The tangency portfolio is in general given by [see (6.8)]

\[
\pi_t^{\text{tan}} = \frac{1}{1^\top (\sigma_t^\top)^{-1} \lambda_t} (\sigma_t^\top)^{-1} \lambda_t.
\]

The expected excess rate of return on the tangency portfolio is

\[
(\pi_t^{\text{tan}})^\top (\mu_t - r_t 1) = \frac{1}{1^\top (\sigma_t^\top)^{-1} \lambda_t} \|\lambda_t\|^2.
\]

The volatility (instantaneous standard deviation) of the tangency portfolio is

\[
\sqrt{(\pi_t^{\text{tan}})^\top \sigma_t \sigma_t^\top \pi_t^{\text{tan}}} = \frac{1}{1^\top (\sigma_t^\top)^{-1} \lambda_t} \|\lambda_t\|.
\]

The slope of the instantaneous capital market line is therefore equal to \( \|\lambda_t\| \). (In a setting with a single risky asset, \( \lambda_t = (\mu_t - r_t)/\sigma_t \) and \( \|\lambda_t\| = \lambda_t \).) In a static framework the optimal portfolio is determined by the position of the capital market line, i.e., (1) the intercept which is equal to the risk-free rate of return and (2) the slope which is the Sharpe ratio of the tangency portfolio. It is therefore natural that investors in a dynamic framework only are concerned about the variations in these two variables.

### 7.3 CRRA utility

In this section we assume that the investor has time-additive expected CRRA utility with a constant relative risk aversion \( \gamma > 1 \). The case \( \gamma = 1 \) that corresponds to logarithmic utility has to be analyzed separately (see Section 7.4). However, it turns out that when \( \gamma \) is put equal to 1 in the optimal strategies derived for \( \gamma > 1 \) we obtain the optimal strategies derived for logarithmic utility.

#### 7.3.1 One-dimensional state variable

Consider the indirect utility function with CRRA utility:

\[
J(W, x, t) = \sup_{(c_s, \pi_s) \in [s, T]} E_{W, x, t} \left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} \frac{\varepsilon_1^{1-\gamma}}{1-\gamma} ds + \varepsilon_2 s e^{-\delta(T-t)} \frac{W_{1-\gamma}}{1-\gamma} \right],
\]

\footnote{Examples where \( \|\lambda_t\|^2 \) is constant, but \( \lambda_t \) itself is not, can be given [see Nielsen and Vassalou (2006)], but seem rather contrived.}
where $\varepsilon_1$ and $\varepsilon_2$ are greater than or equal to zero with at least one of them being non-zero. We set up a conjecture for the form of $J$ using the same arguments as we did in the case of constant investment opportunities. Due to the linearity of the wealth dynamics it seems reasonable to guess that if the strategy $({c^*, \pi^*})$ is optimal with time $t$ wealth $W$ and state $x$ and the corresponding wealth process $W^*$, then the strategy $({kc^*, \pi^*})$ will be optimal with time $t$ wealth $kW$ and state $x$ and the corresponding wealth process $kW^*$. If this is true, then

$$J(kW, x, t) = E_t \left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} \frac{(kc_s^*)^{1-\gamma}}{1-\gamma} ds + \varepsilon_2 e^{-\delta(T-t)} \frac{(kW_T^*)^{1-\gamma}}{1-\gamma} \right]$$

$$= k^{1-\gamma} E_t \left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} \frac{(c_s^*)^{1-\gamma}}{1-\gamma} ds + \varepsilon_2 e^{-\delta(T-t)} \frac{(W_T^*)^{1-\gamma}}{1-\gamma} \right]$$

$$= k^{1-\gamma} J(W, x, t),$$

i.e., the indirect utility function is homogeneous of degree $1 - \gamma$ in the wealth level. Inserting $k = 1/W$ and rearranging, we get

$$J(W, x, t) = \frac{1}{1-\gamma} g(x, t)^{\gamma} W^{1-\gamma},$$

where $g(x, t)^{\gamma} = (1-\gamma)J(1, x, t)$. From the terminal condition $J(W, x, T) = \varepsilon_2 W^{1-\gamma}/(1 - \gamma)$, we have that $g(x, T)^{\gamma} = \varepsilon_2$.

The relevant derivatives of $J$ are

$$J_W(W, x, t) = g(x, t)^{\gamma} W^{-\gamma},$$

$$J_{WW}(W, x, t) = -\gamma g(x, t)^{\gamma} W^{-\gamma - 1},$$

$$J_x(W, x, t) = \frac{\gamma}{1-\gamma} g(x, t)^{\gamma - 1} g_x(x, t) W^{1-\gamma},$$

$$J_{xx}(W, x, t) = -\gamma g(x, t)^{\gamma - 2} g_x(x, t)^2 W^{1-\gamma} + \frac{\gamma}{1-\gamma} g(x, t)^{\gamma - 1} g_{xx}(x, t) W^{1-\gamma},$$

$$J_{WX}(W, x, t) = \gamma g(x, t)^{\gamma - 1} g_x(x, t) W^{-\gamma},$$

$$\frac{\partial J}{\partial t}(W, x, t) = \frac{\gamma}{1-\gamma} g(x, t)^{\gamma - 1} \frac{\partial g}{\partial t}(x, t) W^{1-\gamma}.$$
The dynamics of the value of a given portfolio $\pi$ is

$$dV_t^\pi = V_t^\pi \left[ (r(x_t) + \pi_t^\top \varepsilon(x_t, t) \lambda(x_t)) \, dt + \pi_t^\top \varepsilon(x_t, t) \, dz_t \right].$$

We see that the hedge portfolio is matching the sensitivity of the optimal wealth-to-consumption ratio with respect to the hedgeable shocks represented by $dz_t$.

Inserting the derivatives above into (7.9) and simplifying, we get that $g(x, t)$ must solve the PDE

$$0 = \varepsilon_1^{1/\gamma} - \left( \delta + \frac{\gamma - 1}{\gamma} \lambda(x) \right) g(x, t) + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x)^\top \nu(x) \right) g_x(x, t) + \frac{\partial g}{\partial t}(x, t) + \frac{\nu(x)^2}{2} g_{xx}(x, t)^2$$

with the terminal condition $g(x, T) = \varepsilon_2^{1/\gamma}$. In the case with no intermediate consumption we have $\varepsilon_1 = 0$, and we can without loss of generality assume $\varepsilon_2 = 1$ and $\delta = 0$. If we write

$$g(x, t) \equiv g(x, t; T) = \exp \left\{ - \frac{\gamma - 1}{\gamma} H(x, T - t) \right\},$$

then $H(x, \tau)$ has to solve the simpler PDE

$$0 = \frac{\partial H}{\partial \tau}(x, \tau) + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x)^\top \nu(x) \right) H_x(x, \tau) + \frac{\nu(x)^2}{2} H_{xx}(x, \tau)$$

with the condition $H(x, 0) = 0$.

**Theorem 7.5.** With CRRA utility of terminal wealth only ($\varepsilon_1 = 0$, $\varepsilon_2 = 1$, $\delta = 0$), the indirect utility function is

$$J(W, x, t) = \frac{1}{1 - \gamma} e^{-(\gamma - 1)H(x, T - t)} W^{1 - \gamma} = \frac{1}{1 - \gamma} \left( W e^{H(x, T - t)} \right)^{1 - \gamma},$$

and the optimal investment strategy in (7.12) can be rewritten as

$$\Pi(W, x, t) = \frac{1}{\gamma} ( \gamma_0(x, t)^\top )^{-1} \lambda(x) - \frac{\gamma - 1}{\gamma} H_x(x, T - t) \left( \gamma_0(x, t)^\top \right)^{-1} \nu(x),$$

where $H(x, \tau)$ solves the PDE (7.14) with initial condition $H(x, 0) = 0$.

When the market is complete so that $\hat{\nu}(x) \equiv 0$, the next theorem shows that the solution to the utility maximization problem with intermediate consumption follows from the solution to the problem of maximizing utility of wealth at a single point in time. The proof is left for Exercise 7.1.

**Theorem 7.6.** Let $\tilde{H}(x, \tau)$ be the solution to the PDE

$$0 = \frac{\partial \tilde{H}}{\partial \tau}(x, \tau) + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x)^\top \nu(x) \right) \tilde{H}_x(x, \tau) + \frac{\nu(x)^2}{2} \tilde{H}_{xx}(x, \tau)$$

with terminal condition $\tilde{H}(x, 0) = 0$. Define

$$\tilde{g}(x, t; s) = \exp \left\{ - \frac{\delta}{\gamma} (s - t) - \frac{\gamma - 1}{\gamma} \tilde{H}(x, s - t) \right\}.$$
Then the solution to the PDE (7.13) with \( \hat{v}(x) \equiv 0 \) is

\[
g(x, t) = \varepsilon_1^{\frac{1}{\gamma}} \int_t^T \tilde{g}(x, t; s) \, ds + \varepsilon_2^{\frac{1}{\gamma}} \tilde{g}(x, t; T).
\]

In a complete market \( \hat{v}(x) \equiv 0 \), the maximization of CRRA utility of intermediate consumption and/or terminal wealth leads to the indirect utility

\[
J(W, x, t) = \frac{1}{1-\gamma} g(x, t) W^{1-\gamma},
\]

the optimal consumption strategy is

\[
C(W, x, t) = \varepsilon_1^{1/\gamma} \frac{W}{g(x, t)} = \left( \int_t^T \tilde{g}(x, t; s) \, ds + \left( \frac{\varepsilon_2}{\varepsilon_1} \right)^\frac{1}{\gamma} \tilde{g}(x, t; T) \right)^{-1} W,
\]

and the optimal investment strategy is

\[
\Pi(W, x, t) = \frac{1}{\gamma} (\sigma(x, t)^\gamma)^{-1} \lambda(x) - \frac{1}{\gamma} D(x, t, T) (\sigma(x, t)^\gamma)^{-1} v(x),
\]

where

\[
D(x, t, T) = \frac{\int_t^T \tilde{H}_x(x, s - t) \tilde{g}(x, t; s) \, ds + (\varepsilon_2/\varepsilon_1)^{\frac{1}{\gamma}} \tilde{H}_x(x, T - t) \tilde{g}(x, t; T)}{\int_t^T \tilde{g}(x, t; s) \, ds + (\varepsilon_2/\varepsilon_1)^{\frac{1}{\gamma}} \tilde{g}(x, t; T)}.
\]

The solution for the case with utility of intermediate consumption is thus obtained by simply integrating up the solution for the case with utility of wealth at each of the fixed time horizons over the remaining life-time \([t, T]\). In any specific case with complete markets, the key challenge is therefore to solve the PDE (7.15).

The PDE (7.14)—and thus the special case (7.15)—has a nice solution in a large class of interesting models as we will show below. This leads to closed-form solutions to the power utility maximization problem with terminal wealth only and—if the market is complete—with intermediate consumption. If the market is incomplete, the power utility maximization problem with intermediate consumption is generally intractable, but the PDE (7.13) can be solved numerically.

### 7.3.2 Affine models

In this and the following subsection we will look at models in which the optimal portfolio and consumption strategies of a CRRA investor can be derived in closed-form. In some of these cases we can obtain explicit solutions, in other cases the solution involves time-dependent functions that can be found by numerically solving ordinary differential equations. Many of our concrete examples in the following chapters are special cases of these models. In this section we will discuss so-called affine models, while the next section focuses on the so-called quadratic models. The results presented are similar to those obtained by Liu (1999, 2007). For notational simplicity we shall assume that the state variable is one-dimensional with dynamics given by (7.1). We will briefly discuss solutions to problems with a multi-dimensional state variable in Section 7.3.4.

As explained above, the key is to solve the PDE (7.14) for \( H(x, \tau) \) with the initial condition \( H(x, 0) = 0 \). Let us consider when we can find a solution of the affine form

\[
H(x, \tau) = A_0(\tau) + A_1(\tau)x,
\]

where \( A_0 \) and \( A_1 \) are real-valued deterministic functions that have to satisfy \( A_0(0) = A_1(0) = 0 \).
to meet the initial condition. Substituting into (7.14), we find

$$0 = r(x) + \frac{1}{2\gamma} \|\lambda(x)\|^2 - A'_0(\tau) - A'_1(\tau) x + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x)^\top v(x) \right) A_1(\tau)$$

$$- \frac{\gamma - 1}{2\gamma} (\Sigma^2(x) + (\gamma - 1) \hat{v}(x)^2) A_1(\tau)^2. \tag{7.16}$$

If \(r(x), \|\lambda(x)\|^2, m(x), v(x)^\top \lambda(x), \|v(x)\|^2, \text{ and } \hat{v}(x)^2\) are all affine functions\(^2\) of \(x\), then we can find two ordinary differential equations for \(A_0\) and \(A_1\). In order to see this, suppose that

\[
\begin{align*}
    r(x) &= r_0 + r_1 x, \\
    m(x) &= m_0 + m_1 x, \\
    \hat{v}(x) &= \sqrt{\hat{v}_0 + \hat{v}_1 x}
\end{align*}
\]

for some constants \(r_0, r_1, m_0, m_1, \hat{v}_0, \text{ and } \hat{v}_1\). Of course, we should have that \(\hat{v}_0 + \hat{v}_1 x \geq 0\) for all possible values of \(x\), which is easily satisfied if either \(\hat{v}_0\) or \(\hat{v}_1\) are zero and the other parameter is positive. The term \(\|\lambda(x)\|^2\) will be affine in \(x\) if each element of the vector \(\lambda(x) = (\lambda_1(x), \ldots, \lambda_d(x))^\top\) is of the form \(\lambda_i(x) = \sqrt{\lambda_{i0} + \lambda_{i1} x}\) since then

\[
\|\lambda(x)\|^2 = \sum_{i=1}^d \lambda_i(x)^2 = \sum_{i=1}^d (\lambda_{i0} + \lambda_{i1} x) = \left( \sum_{i=1}^d \lambda_{i0} \right) + \left( \sum_{i=1}^d \lambda_{i1} \right) x \equiv \Lambda_0 + \Lambda_1 x. \tag{7.19}
\]

Similarly, the term \(\|v(x)\|^2\) will be affine in \(x\) if each element of the vector \(v(x) = (v_1(x), \ldots, v_d(x))^\top\) is of the form \(v_i(x) = \sqrt{v_{i0} + v_{i1} x}\). Then we have

\[
\|v(x)\|^2 = \sum_{i=1}^d v_i(x)^2 = \sum_{i=1}^d (v_{i0} + v_{i1} x) = \left( \sum_{i=1}^d v_{i0} \right) + \left( \sum_{i=1}^d v_{i1} \right) x \equiv V_0 + V_1 x. \tag{7.20}
\]

In addition, we must have that \(v(x)^\top \lambda(x)\) is affine in \(x\). With the specifications of \(\lambda(x)\) and \(v(x)\) just given, we have

\[
v(x)^\top \lambda(x) = \sum_{i=1}^d v_i(x) \lambda_i(x) = \sum_{i=1}^d \sqrt{(v_{i0} + v_{i1} x)(\lambda_{i0} + \lambda_{i1} x)}.
\]

This will only be affine in \(x\) if, for each \(i\), we have either

(i) \(v_{i0} = \lambda_{i0} = 0\), or

(ii) \(v_{i1} = \lambda_{i1} = 0\), or

(iii) \(v_{i0} = \lambda_{i0}\) and \(v_{i1} = \lambda_{i1}\).

To encompass all possible situations let us write

\[
v(x)^\top \lambda(x) = K_0 + K_1 x, \tag{7.21}
\]

where \(K_0\) and \(K_1\) are real-valued parameters. If we substitute (7.17)–(7.21) into (7.16) and use the fact that (7.16) must hold for all values of \(x\) and all \(\tau\), we obtain a system of two ordinary

\(^2\)A real-valued function is said to be an affine function of the \(k\)-vector \(x\), if it can be written as \(a_1 + a_2^\top x\), where \(a_2\) is a constant scalar and \(a_2\) is a constant \(k\)-vector (possibly zero so that a constant is also included in the set of affine functions). A vector- or matrix-valued function is said to be affine if all its elements are affine.
Theorem 7.7. Assume that 
and define

\[ A_0'(\tau) = r_0 + \frac{A_0}{2\gamma} + \left( m_0 - \frac{\gamma - 1}{\gamma} K_0 \right) A_1(\tau) - \frac{\gamma - 1}{2\gamma} (V_0 + \gamma \tilde{v}_0) A_1(\tau)^2, \]
\[ A_1'(\tau) = r_1 + \frac{A_1}{2\gamma} + \left( m_1 - \frac{\gamma - 1}{\gamma} K_1 \right) A_1(\tau) - \frac{\gamma - 1}{2\gamma} (V_1 + \gamma \tilde{v}_1) A_1(\tau)^2. \] (7.22)

These equations are to be solved with the initial conditions \( A_0(0) = A_1(0) = 0 \).

First (7.22) is solved for \( A_1(\tau) \). From Theorem C.2, we can make the following conclusion. Suppose that

\[ \left( m_1 - \frac{\gamma - 1}{\gamma} K_1 \right)^2 + 2 \frac{\gamma - 1}{\gamma} \left( r_1 + \frac{A_1}{2\gamma} \right) (V_1 + \gamma \tilde{v}_1) > 0 \] (7.23)

and define

\[ \nu = \sqrt{\left( m_1 - \frac{\gamma - 1}{\gamma} K_1 \right)^2 + 2 \frac{\gamma - 1}{\gamma} \left( r_1 + \frac{A_1}{2\gamma} \right) (V_1 + \gamma \tilde{v}_1)}. \]

Then the solution to (7.22) with \( A_1(0) = 0 \) is

\[ A_1(\tau) = \frac{2 \left( r_1 + \frac{A_1}{2\gamma} \right) (e^{\nu \tau} - 1)}{\left( \nu + \frac{\gamma - 1}{\gamma} K_1 - m_1 \right) (e^{\nu \tau} - 1) + 2\nu}. \] (7.24)

Since \( A_0(\tau) = A_0(\tau) - A_0(0) = \int_0^\tau A_0'(s) \, ds \), we can afterwards compute \( A_0(\tau) \) as

\[ A_0(\tau) = \left( r_0 + \frac{A_0}{2\gamma} \right) \tau + \left( m_0 - \frac{\gamma - 1}{\gamma} K_0 \right) \int_0^\tau A_1(s) \, ds - \frac{\gamma - 1}{2\gamma} (V_0 + \gamma \tilde{v}_0) \int_0^\tau A_1(s)^2 \, ds. \] (7.25)

Also from Theorem C.2, we have that

\[ \int_0^\tau A_1(s) \, ds = -\frac{2\gamma}{(\gamma - 1)(V_1 + \gamma \tilde{v}_1)} \left\{ \frac{1}{2} \left( \nu + \frac{\gamma - 1}{\gamma} K_1 - m_1 \right) \tau + \ln \left( \frac{2\nu}{\left( \nu + \frac{\gamma - 1}{\gamma} K_1 - m_1 \right) (e^{\nu \tau} - 1) + 2\nu} \right) \right\} \]

and

\[ \int_0^\tau A_1(s)^2 \, ds = \text{- ugly expression to be filled in -} \]

Combining these findings with Theorem 7.5, we arrive at the following conclusion: 3

**Theorem 7.7.** Assume that \( r(x), \|x\|^2, m(x), v(x)^T \lambda(x), \|v(x)\|^2, \) and \( \hat{v}(x)^2 \) are all affine functions of \( x \) and given by (7.17)–(7.21), and that the parameter condition (7.23) holds. For an investor with CRRA utility from terminal wealth only, the indirect utility function is then given by

\[ J(W, x, t) = \frac{1}{1 - \gamma} \left( W e^{A_0(T-t)+A_1(T-t)x} \right)^{1-\gamma}, \]

where \( A_1 \) is given by (7.24) and \( A_0 \) is given by (7.25). The optimal investment strategy is given by

\[ \Pi(W, x, t) = \frac{1}{\gamma} \left( \sigma(x, t)^T \right)^{-1} \lambda(x) - \frac{\gamma - 1}{\gamma} \left( \sigma(x, t)^T \right)^{-1} v(x) A_1(T-t). \]

In some important special cases, \( A_0 \) and \( A_1 \) simplify considerably. For example, if \( V_1 + \gamma \tilde{v}_1 = 0 \) so that the second-order term in (7.22) vanishes then (again see Theorem C.2)

\[ \nu = -\left( m_1 - \frac{\gamma - 1}{\gamma} K_1 \right) = \frac{\gamma - 1}{\gamma} K_1 - m_1. \]

---

3Note the close connection between the analysis above and the analysis for so-called affine models of the term structure of interest rates, see e.g., Duffie and Kan (1996), Dai and Singleton (2000), or Munk (2011).
so that $A_1(\tau)$ reduces to
$$A_1(\tau) = \frac{r_1 + \frac{A_1}{\nu}}{\nu} (1 - e^{-\nu \tau}).$$

In this case the integrals in (7.25) are also relatively simple:

$$\int_0^\tau A_1(u) \, du = \frac{1}{m_1 - \frac{2-\gamma}{\gamma} K_1} \left( \left( r_1 + \frac{A_1}{2\gamma} \right) \tau - A_1(\tau) \right),$$
$$\int_0^\tau A_1(u)^2 \, du = \frac{1}{(r_1 + \frac{A_1}{2\gamma}) \left( m_1 - \frac{2-\gamma}{\gamma} K_1 \right)} \left[ \left( r_1 + \frac{A_1}{2\gamma} \right)^3 \tau - A_1(\tau) - \frac{A_1(\tau)^2}{2 (r_1 + \frac{A_1}{2\gamma})} \right].$$

This special case is relevant in Chapter 10.

For the problem with utility of intermediate consumption, we can provide a solution for the complete markets case ($\hat{v}(x) \equiv 0$) by combining the above computations with Theorem 7.6. The only difference in the relevant ODEs, and thus in their solutions, is that we have to impose the restriction $\hat{v}_0 = \hat{v}_1 = 0$ because of the complete market assumption.

**Theorem 7.8.** Assume a complete financial market ($\hat{v}(x) \equiv 0$) in which $r(x)$, $\|\lambda(x)\|^2$, $m(x)$, $v(x)^T \lambda(x)$, and $\|v(x)\|^2$ are all affine functions of $x$ and given by (7.17), (7.18), (7.19), (7.20), and (7.21). Imposing the restriction $\hat{v}_0 = \hat{v}_1 = 0$, assume that the parameter condition (7.23) holds, and let $A_1$ and $A_0$ be given by (7.24) and (7.25). Define

$$\tilde{g}(x; t; s) = \exp \left\{ -\frac{\delta}{\gamma} (s - t) - \frac{\gamma - 1}{\gamma} (A_0(s - t) + A_1(s - t)x) \right\}.$$ 

For an investor with CRRA utility from intermediate consumption and possibly terminal wealth, the indirect utility function is then given by

$$J(W, x, t) = \frac{1}{1 - \gamma} \left( \frac{\delta}{\gamma} \int_0^T \tilde{g}(x, t; s) \, ds + \frac{\delta}{\gamma} \cdot \tilde{g}(x, t; T) \right)^{-1} W^{1-\gamma},$$
the optimal consumption strategy is

$$C(W, x, t) = \left( \int_0^T \tilde{g}(x, t; s) \, ds + \frac{\delta}{\gamma} \tilde{g}(x, t; T) \right)^{-1} W,$$
and the optimal investment strategy is

$$\Pi(W, x, t) = \frac{1}{\gamma} (\tilde{g}(x, t; T))^{-1} \lambda(x) - \frac{\gamma - 1}{\gamma} D(x, t; T) (\tilde{g}(x, t; T))^{-1} v(x),$$

where

$$D(x, t; T) = \int_t^T \frac{A_1(s - t) \tilde{g}(x, t; s) \, ds + (\frac{\delta}{\gamma})^2 A_1(T - t) \tilde{g}(x, t; T)}{\int_t^T \tilde{g}(x, t; s) \, ds + (\frac{\delta}{\gamma})^2 \tilde{g}(x, t; T)}.$$ 

Assuming for simplicity that $\varepsilon_2 = 0$ and $\varepsilon_1 = 1$, we can rewrite the ratio in the hedge term of the optimal investment strategy as

$$\int_t^T A_1(s - t) \tilde{g}(x, t; s) \, ds \left/ \int_t^T \tilde{g}(x, t; s) \, ds \right. = \int_t^T w(x, s - t) A_1(s - t) \, ds,$$

where we have defined $w(x, s - t) = \tilde{g}(x, t; s) / \int_t^T \tilde{g}(x, t; s) \, ds$. Since $w(x, s - t) > 0$ and $\int_t^T w(x, s - t) \, ds = 1$, we may interpret the hedging demand of an investor with utility of consumption and a
time horizon of $T$ as a weighted average of the hedging demands of investors with time horizons of $s \in [t, T]$ and utility of terminal wealth only. If $\bar{A}_1$ is either monotonically increasing or decreasing (as will be the case in many concrete settings), there will exist a $T^* \in [t, T]$ such that

$$\int_t^T w(x, s-t)A_1(s-t) \, ds = A_1(T^* - t),$$

in which case we can represent the hedging demand as $(\sigma(x,t)\top)^{-1}v(x)A_1(T^*-t)$. Since this is exactly the hedging demand of an investor with time horizon $T^*$ and utility of terminal wealth only, we may interpret $T^*$ as the effective time horizon of the investor with time horizon $T$ and utility of consumption. Note the similarity to the concept of duration for fixed-income securities, cf. Munk (2011).

### 7.3.3 Quadratic models

The assumptions of the affine models cover some interesting settings, but not all. In this section we shall see that under another set of assumptions on the market parameter functions $r, m, v, \lambda$, and $\dot{v}$, we obtain an exponential-quadratic expression for the function $g(x,t)$. In Chapter 11, we will study an important example which is covered by these assumptions.

As before, the key is to solve the PDE (7.14) for $H(x, \tau)$ with the initial condition $H(x,0) = 0$. Let us consider when we can find a solution of the quadratic form

$$H(x, \tau) = A_0(\tau) + A_1(\tau)x + \frac{1}{2}A_2(\tau)x^2,$$

where $A_0, A_1$, and $A_2$ are real-valued deterministic functions that have to satisfy $A_0(0) = A_1(0) = A_2(0)$ to ensure that $H(x,0) = 0$ for all $x$. Substituting the relevant derivatives into (7.14), we arrive at

$$0 = r(x) + \frac{1}{2\gamma}||\lambda(x)||^2 + \left(m(x) - \frac{\gamma-1}{\gamma}v(x)\top\lambda(x)\right)(A_1(\tau) + A_2(\tau)x)
- A'_0(\tau) - A'_1(\tau)x - \frac{1}{2}A'_2(\tau)x^2 + \frac{1}{2}||v(x)||^2 + \dot{v}(x)^2)A_2(\tau)
- \frac{\gamma-1}{\gamma}(||v(x)||^2 + \dot{v}(x)^2)(A_1(\tau) + A_2(\tau)x)^2. \tag{7.26}$$

To ensure that we only have powers of $x$ of order zero, one, and two, we can allow (i) $r(x)$ and $||\lambda(x)||^2$ to be quadratic\(^4\) in $x$, (ii) $m(x)$ and $v(x)\top\lambda(x)$ can be affine in $x$, while (iii) $||v(x)||^2$ and $\dot{v}(x)^2$ have to be constant. Therefore, write $v(x) = v = (v_1, \ldots, v_d)\top$, $\dot{v}(x) = \dot{v}$, and

$$r(x) = r_0 + r_1x + r_2x^2, \tag{7.27}$$
$$m(x) = m_0 + m_1x, \tag{7.28}$$
$$\lambda_i(x) = \lambda_{i0} + \lambda_{i1}x. \tag{7.29}$$

\(^4\) A real-valued function is said to be a quadratic function of the $k$-vector $x$, if it can be written as $a_1 + a_2\top x + x\top a_3 x$, where $a_1$ is a constant scalar, $a_2$ is a constant $k$-vector, and $a_3$ is a constant ($k \times k$)-matrix (either $a_2$ or $a_3$ or both can be zero so that a constant and an affine function are also considered quadratic. A vector- or matrix-valued function is said to be quadratic if all its elements are quadratic.
for some constants $r_0, r_1, r_2, m_0, m_1, m_2, \lambda_{i0}, \lambda_{i1}, \lambda_{i2}$. Consequently,

$$\|x(x)\|^2 = \sum_{i=1}^{d} \lambda_i(x)^2 = \left( \sum_{i=1}^{d} \lambda_{i0}^2 \right) + 2 \left( \sum_{i=1}^{d} \lambda_{i0}\lambda_{i1} \right) x + \left( \sum_{i=1}^{d} \lambda_{i1}^2 \right) x^2$$

$$\equiv \Lambda_0 + \Lambda_1 x + \Lambda_2 x^2,$$ \hspace{1cm} (7.30)

$$\nu(x)^T \lambda(x) = \sum_{i=1}^{d} v_i(x) \lambda_i(x) = \left( \sum_{i=1}^{d} v_i \lambda_{i0} \right) + \left( \sum_{i=1}^{d} v_i \lambda_{i1} \right) x \equiv K_0 + K_1 x.$$ \hspace{1cm} (7.31)

If we substitute (7.27)–(7.31) into (7.26) and use the fact that (7.26) must hold for all values of $x$ and all $t$, we obtain a system of three ordinary differential equations for $A_0, A_1,$ and $A_2$:

$$A_0' (\tau) = r_0 + \frac{\Lambda_0}{2\gamma} + \left( m_0 - \frac{\gamma - 1}{\gamma} K_0 \right) A_1 (\tau)$$

$$+ \frac{1}{2} \left( \|v\|^2 + \hat{\nu}^2 \right) A_2 (\tau) - \frac{\gamma - 1}{2\gamma} \left( \|v\|^2 + \gamma \hat{\nu}^2 \right) A_1 (\tau)^2,$$ \hspace{1cm} (7.32)

$$A_1' (\tau) = r_1 + \frac{\Lambda_1}{2\gamma} + m_0 - \frac{\gamma - 1}{\gamma} K_0 A_2 (\tau)$$

$$+ \left[ m_1 - \frac{\gamma - 1}{\gamma} K_1 - \frac{\gamma - 1}{\gamma} \left( \|v\|^2 + \gamma \hat{\nu}^2 \right) A_2 (\tau) \right] A_1 (\tau),$$ \hspace{1cm} (7.33)

$$A_2' (\tau) = 2r_2 + \frac{\Lambda_2}{\gamma} + 2 \left[ m_1 - \frac{\gamma - 1}{\gamma} K_1 \right] A_2 (\tau) - \frac{\gamma - 1}{\gamma} \left( \|v\|^2 + \gamma \hat{\nu}^2 \right) A_2 (\tau)^2.$$ \hspace{1cm} (7.34)

These equations are to be solved with the initial conditions $A_0(0) = A_1(0) = A_2(0) = 0$.

The equations (7.33) and (7.34) can be solved using Theorem C.3. Suppose that

$$\left( m_1 - \frac{\gamma - 1}{\gamma} K_1 \right)^2 + \frac{\gamma - 1}{\gamma} \left( 2r_2 + \frac{\Lambda_2}{\gamma} \right) \left( \|v\|^2 + \hat{\nu}^2 \right) > 0$$ \hspace{1cm} (7.35)

and define

$$\nu = 2 \sqrt{\left( m_1 - \frac{\gamma - 1}{\gamma} K_1 \right)^2 + \frac{\gamma - 1}{\gamma} \left( 2r_2 + \frac{\Lambda_2}{\gamma} \right) \left( \|v\|^2 + \hat{\nu}^2 \right)}.$$ \hspace{1cm} (7.36)

Then the solution to (7.34) with $A_2(0) = 0$ is

$$A_2 (\tau) = \frac{2 \left( 2r_2 + \frac{\Lambda_2}{\gamma} \right) (e^{\nu \tau} - 1)}{\nu + 2 \frac{\gamma - 1}{\gamma} K_1 - 2m_1 (e^{\nu \tau} - 1) + 2\nu}.$$ \hspace{1cm} (7.37)

The solution to (7.33) with $A_1(0) = 0$ is

$$A_1 (\tau) = \frac{r_1 + \frac{\Lambda_1}{2\gamma}}{2r_2 + \frac{\Lambda_2}{\gamma}} A_2 (\tau) + \frac{4q}{\nu + 2 \frac{\gamma - 1}{\gamma} K_1 - 2m_1 (e^{\nu \tau} - 1) + 2\nu} \left( e^{\nu \tau/2} - 1 \right)^2.$$ \hspace{1cm} (7.38)

Finally, we can compute $A_0 (\tau)$ by integrating up (7.32):

$$A_0 (\tau) = \left( r_0 + \frac{\Lambda_0}{2\gamma} \right) \tau + \left( m_0 - \frac{\gamma - 1}{\gamma} K_0 \right) \int_0^\tau A_1 (s) \, ds$$

$$+ \frac{1}{2} \left( \|v\|^2 + \hat{\nu}^2 \right) \int_0^\tau A_2 (s) \, ds - \frac{\gamma - 1}{2\gamma} \int_0^\tau \left( \|v\|^2 + \gamma \hat{\nu}^2 \right) A_2 (s) \, ds.$$ \hspace{1cm} (7.39)

These integrals can be calculated explicitly and are generally quite complex, but simplify somewhat in relevant special cases.
We summarize our findings in the following theorem.\footnote{Note the close connection to the so-called quadratic models of the term structure of interest rates, see e.g., Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2003).}

**Theorem 7.9.** Assume that \( v(x) = v \), \( \dot{v}(x) = \dot{v} \), and that \( r(x) \), \( m(x) \), and \( \lambda(x) \) are given as in (7.27)–(7.29), and that the parameter condition (7.35) holds. For an investor with CRRA utility from terminal wealth only, the indirect utility function is then given by

\[
J(W, x, t) = \frac{1}{1 - \gamma} \left( W e^{A_0(T-t)+A_1(T-t)x + \frac{1}{2} A_2(T-t)x^2} \right)^{1-\gamma},
\]

where \( A_2 \), \( A_1 \), and \( A_0 \) are given by (7.36), (7.37), and (7.38). The optimal investment strategy is given by

\[
\Pi(W, x, t) = \frac{1}{\gamma} \left( \sigma(x,t)^\top \right)^{-1} \lambda(x) - \frac{\gamma - 1}{\gamma} \left( \sigma(x,t)^\top \right)^{-1} v \left( A_1(T-t) + A_2(T-t)x \right).
\]

For a complete market we can generalize the above results to encompass investors with utility from intermediate consumption. The relevant ODEs are the same, and thus their solutions are the same as above, except that we impose the condition \( \dot{v} = 0 \).

**Theorem 7.10.** Assume that the market is complete (\( \dot{v}(x) \equiv 0 \)), that \( v(x) = v \), and that \( r(x) \), \( m(x) \), and \( \lambda(x) \) are given as in (7.27)–(7.29). Imposing the restriction \( \dot{v} = 0 \), assume that the parameter condition (7.35) holds, and let \( A_2 \), \( A_1 \), and \( A_0 \) be given by (7.36), (7.37), and (7.38). Define

\[
\tilde{g}(x, t; s) = \exp \left\{ -\frac{\delta}{\gamma} (s - t) - \frac{\gamma - 1}{\gamma} \left( A_0(s-t) + A_1(s-t)x + \frac{1}{2} A_2(s-t)x^2 \right) \right\}.
\]

For an investor with CRRA utility from intermediate consumption and possibly terminal wealth, the indirect utility function is then given by

\[
J(W, x, t) = \frac{1}{1 - \gamma} \left( \frac{1}{\epsilon_1} \int_t^T \tilde{g}(x, t; s) ds + \frac{1}{\epsilon_2} \tilde{g}(x, t; T) \right)^{\gamma} W^{1-\gamma},
\]

the optimal consumption strategy is

\[
C(W, x, t) = \left( \int_t^T \tilde{g}(x, t; s) ds + \frac{1}{\epsilon_1} \tilde{g}(x, t; T) \right)^{-1} W,
\]

and the optimal investment strategy is

\[
\Pi(W, x, t) = \frac{1}{\gamma} \left( \sigma(x,t)^\top \right)^{-1} \lambda(x) - \frac{\gamma - 1}{\gamma} D(x, t, T) \left( \sigma(x,t)^\top \right)^{-1} v,
\]

where

\[
D(x, t, T) = \frac{\int_t^T \left( A_1(s-t) + A_2(s-t)x \right) \tilde{g}(x, t; s) ds + \left( \frac{\epsilon_2}{\epsilon_1} \right)^{\frac{1}{2}} \left( A_1(T-t) + A_2(T-t)x \right) \tilde{g}(x, t; T)}{\int_t^T \tilde{g}(x, t; s) ds + \left( \frac{\epsilon_2}{\epsilon_1} \right)^{\frac{1}{2}} \tilde{g}(x, t; T)}.
\]
7.3.4 Multi-dimensional state variable

With a multi-dimensional state variable \( \mathbf{x} \), a qualified guess on the indirect utility function is

\[
J(W, \mathbf{x}, t) = \frac{1}{1-\gamma} g(\mathbf{x}, t) W^{1-\gamma},
\]

which indeed is a solution to the HJB equation (7.11) if the function \( g(\mathbf{x}, t) \) solves the PDE

\[
0 = \varepsilon_1^{1/\gamma} - \left( \frac{\delta}{\gamma} + \frac{\gamma - 1}{\gamma} \mathbf{r}(\mathbf{x}) + \frac{\gamma - 1}{2\gamma^2} ||\lambda(\mathbf{x})||^2 \right) g(\mathbf{x}, t) + \frac{\partial g}{\partial t}(\mathbf{x}, t)
\]

\[
+ \left( \mathbf{m}(\mathbf{x}) - \frac{\gamma - 1}{\gamma} \mathbf{v}(\mathbf{x})^\top \lambda(\mathbf{x}) \right)^\top g_x(\mathbf{x}, t) + \frac{1}{2} \text{tr} \left( g_{xx}(\mathbf{x}, t) \Sigma^x(\mathbf{x}) \right)
\]

\[
+ \frac{1}{2} (\gamma - 1) g(\mathbf{x}, t)^{-1} g_x(\mathbf{x}, t)^\top \mathbf{v}(\mathbf{x}) \mathbf{v}(\mathbf{x})^\top g_x(\mathbf{x}, t)
\]

with the terminal condition \( g(\mathbf{x}, T) = \varepsilon_2^{1/\gamma} \). The optimal investment strategy is

\[
\Pi(W, \mathbf{x}, t) = \frac{1}{\gamma} \left( g(\mathbf{x}(t)\right)^{-1} \lambda(\mathbf{x}) + \frac{1}{g(\mathbf{x}, t)} \left( g(\mathbf{x}, t)^{-1} \mathbf{v}(\mathbf{x}) g_x(\mathbf{x}, t),
\]

and with intermediate consumption the optimal consumption rate is given by

\[
C(W, \mathbf{x}, t) = \varepsilon_1^{1/\gamma} \frac{W}{g(\mathbf{x}, t)}.
\]

With no intermediate consumption (\( \varepsilon_1 = 0, \varepsilon_2 = 1, \delta = 0 \)), we can write

\[
g(\mathbf{x}, t) \equiv g(\mathbf{x}, t; T) = \exp \left\{ -\frac{\gamma - 1}{\gamma} H(\mathbf{x}, T - t) \right\},
\]

and \( H(\mathbf{x}, \tau) \) then has to solve

\[
0 = \mathbf{r}(\mathbf{x}) + \frac{1}{2\gamma} ||\lambda(\mathbf{x})||^2 - \frac{\partial H}{\partial \tau}(\mathbf{x}, \tau) + \left( \mathbf{m}(\mathbf{x}) - \frac{\gamma - 1}{\gamma} \mathbf{v}(\mathbf{x})^\top \lambda(\mathbf{x}) \right)^\top H_x(\mathbf{x}, \tau)
\]

\[
+ \frac{1}{2} \text{tr} \left( \Sigma^x(\mathbf{x}) H_{xx}(\mathbf{x}, \tau) \right) - \frac{\gamma - 1}{2\gamma} H_x(\mathbf{x}, \tau)^\top \left( \Sigma^x(\mathbf{x}) + (\gamma - 1) \mathbf{v}(\mathbf{x}) \mathbf{v}(\mathbf{x})^\top \right) H_x(\mathbf{x}, \tau)
\]

with the condition \( H(\mathbf{x}, 0) = 0 \). In this case, the indirect utility function is

\[
J(W, \mathbf{x}, t) = \frac{1}{1-\gamma} \left( W e^{H(\mathbf{x}, T - t)} \right)^{1-\gamma},
\]

and the optimal investment strategy is

\[
\Pi(W, \mathbf{x}, t) = \frac{1}{\gamma} \left( g(\mathbf{x}(t)\right)^{-1} \lambda(\mathbf{x}) - \frac{1}{\gamma} \left( g(\mathbf{x}, t)^{-1} \mathbf{v}(\mathbf{x}) H_x(\mathbf{x}, T - t).\right)
\]

As in the case of a one-dimensional state variable, the solution to the problem with utility of consumption can be stated in terms of various integrals involving \( H \) under the assumption that the market is complete.

The results for affine and quadratic models with a one-dimensional state variable can be generalized to settings with a multi-dimensional state variable. We get exactly the same results as in Theorems 7.7–7.10 except that the \( A_1 \)-function is now vector-valued and the \( A_2 \)-function is matrix-valued. We get a larger system of differential equations to solve.

Let us briefly summarize the results for the multi-dimensional affine case. The short rate is of the form

\[
r(\mathbf{x}) = r_o + r_1^\top \mathbf{x},
\]
the dynamics of the state variable $x$ is
\[ dx_t = \left( m_0 + mx_t \right) dt + D \sqrt{\Var(x_t)} \, dz_t + \tilde{D} \sqrt{\Var(x_t)} \, d\tilde{z}_t, \]
where $m_0$ is a $k$-vector, $m$ is a $k \times k$-matrix, $D$ is a $k \times d$-matrix, $\tilde{D}$ is a $k \times k$-matrix, and the $d \times d$-matrix $\Var(x)$ and the $k \times k$-matrix $\hat{\Var}(x)$ are diagonal matrices with elements
\[ [\Var(x)]_{ii} = \nu_i + \hat{V}_i^T x, \quad [\hat{\Var}(x)]_{ii} = \hat{\nu}_i + \hat{V}_i^T x. \]

Furthermore, we must have
\[ \hat{\Var}(x) \lambda(x) = D \sqrt{\Var(x)} \lambda(x) = K_0 + K_1 x \quad (7.42) \]
for some $k$-vector $K_0$ and $(k \times k)$-matrix $K_1$, and
\[ \|\lambda(x)\|^2 = \Lambda_0 + \Lambda_1^T x \quad (7.43) \]
for some scalar $\Lambda_0$ and $k$-vector $\Lambda_1$. Eqs. (7.42) and (7.43) are satisfied if $\lambda(x) = \sqrt{\Var(x)} \xi$ for some $d$-vector $\xi$ but slightly more general specifications of $\lambda(x)$ are also possible. In this case, the PDE (7.40) has a solution of the affine form
\[ H(x, \tau) = A_0(\tau) + A_1(\tau)^T x, \]
where $A_1(\tau)$ satisfies $A_1(0) = 0$ and the ODE
\[ A_1'(\tau) = r_1 + \frac{A_1}{2\gamma} + \left( m_{\parallel \xi} - \frac{\gamma - 1}{\gamma} K_{\parallel \xi} \right)^T A_1(\tau) \]
\[ - \frac{\gamma - 1}{2\gamma} \left( \sum_{i=1}^d |D_{\parallel \xi}^T A_1(\tau)|^2 \hat{V}_i + \sum_{i=1}^k |\tilde{D}_{\parallel \xi}^T A_1(\tau)|^2 \tilde{V}_i \right), \]
and $A_0(\tau)$ satisfies $A_0(0) = 0$ and the ODE
\[ A_0'(\tau) = r_0 + \frac{A_0}{2\gamma} + \left( m_0 - \frac{\gamma - 1}{\gamma} K_0 \right)^T A_1(\tau) \]
\[ - \frac{\gamma - 1}{2\gamma} \left( \sum_{i=1}^d |D_{\parallel \xi}^T A_1(\tau)|^2 \nu_i + \sum_{i=1}^k |\tilde{D}_{\parallel \xi}^T A_1(\tau)|^2 \hat{\nu}_i \right). \]

Given $A_1$, $A_0$ can be computed by integration:
\[ A_0(\tau) = \int_0^\tau A_0'(s) \, ds = \left( r_0 + \frac{A_0}{2\gamma} \right) \tau + \left( m_0 - \frac{\gamma - 1}{\gamma} K_0 \right) \tau \int_0^\tau A_1(s) \, ds \]
\[ - \frac{\gamma - 1}{2\gamma} \left( \sum_{i=1}^d \nu_i \int_0^\tau |D_{\parallel \xi}^T A_1(s)|^2 ds + \sum_{i=1}^k \hat{\nu}_i \int_0^\tau |\tilde{D}_{\parallel \xi}^T A_1(s)|^2 ds \right). \]

The optimal portfolio with utility of terminal wealth only is
\[ \pi(x, t) = \frac{1}{\gamma} \left( g(x, t)^T \right)^{-1} \lambda(x) - \frac{1}{\gamma} \left( g(x, t)^T \right)^{-1} \sqrt{\Var(x)D^T A_1(T - t)}. \]
These results can be extended to utility of intermediate consumption as long as the market is complete.
There are also cases in which the function \( H(x,t) \) is the sum of a function which is affine in some of the individual state variables and quadratic in the others. For example, with a two-dimensional state variable \( x = (x_1, x_2)^T \), we will under some conditions get a solution of the form

\[
H(x_1, x_2, t) = A_0(T - t) + A_{11}(T - t)x_1 + A_{12}(T - t)x_2 + \frac{1}{2}A_2(T - t)x_2^2,
\]

and, consequently, the investment strategy

\[
\Pi(W, x, t) = \frac{1}{\gamma} \left( \sigma(x, t)^T \right)^{-1} \lambda(x)
\]

\[
- \frac{\gamma - 1}{\gamma} \left( \sigma(x, t)^T \right)^{-1} [v_1(x)A_{11} (T - t) + v_2(x) (A_{12} (T - t) + A_2 (T - t)x_2)],
\]

where \( v_i \) is the \( d \)-vector of sensitivities of \( x_i \) with respect to the “traded” risks \( dz_t \).

### 7.4 Logarithmic utility

Logarithmic utility is the special case of CRRA utility in which the relative risk aversion equals one. For notational simplicity, let us assume a one-dimensional state variable. Applying the same procedure to the problem with log utility as we did for CRRA utility, one can show that (this is Exercise 7.2)

\[
J(W, x, t) = g(t) \ln W + h(x, t)
\]

where \( g(t) \) is again given by (6.14) and where \( h(x, t) \) must satisfy a certain PDE. Since the cross derivative \( J_{Wx}(W, x, t) = 0 \), the optimal risky portfolio in (7.7) reduces to

\[
\Pi(W, x, t) = \left( \sigma(x, t)^T \right)^{-1} \lambda(x).
\]

We can conclude that a logarithmic investor does not hedge stochastic variations in the investment opportunity set. She behaves myopically, i.e., as in a static one-period framework. Optimal consumption is again given by

\[
C(W, x, t) = \frac{1}{\gamma} W \frac{1}{g(t)}.
\]

Letting \( \Pi_0(W, x, t) \) denote the fraction of wealth optimally invested in the instantaneously risk-free asset, we can summarize the entire investment strategy as

\[
\begin{pmatrix}
\Pi_0(W, x, t) \\
\Pi(w, x, t)
\end{pmatrix}
=\begin{pmatrix}
1-1^T \left( \frac{\sigma(x, t)^T}{\sigma(x, t)^T} \right)^{-1} \lambda(x) \\
\left( \sigma(x, t)^T \right)^{-1} \lambda(x)
\end{pmatrix}
\]

This portfolio is sometimes referred to as the log portfolio or the growth-optimal portfolio, since it is also the portfolio with the highest expected average compound growth rate of portfolio value. This average growth rate is defined as \( \frac{1}{T-t} \ln (W_T/W_t) \).

### 7.5 How costly are deviations from the optimal investment strategy?

The following results are taken from Larsen and Munk (2012).

We consider an investor with a power utility function of wealth at some future date \( T \) and ignore both intermediate consumption and income other than financial returns. Any combination of an initial wealth \( W \) and an investment strategy \( \pi \) will give rise to a terminal wealth \( W_T^\pi \) (a partially
controlled random variable) and the expected utility associated with that investment strategy is thus
\[ J^\pi(W, x, t) = E_t \left[ \frac{1}{1 - \gamma} (W_T^\pi)^{1 - \gamma} \right], \]
where \( W \) is the initial (time \( t \)) wealth and \( \gamma > 1 \) is the constant relative risk aversion coefficient.

It is well-known that no matter what assumptions are made about the dynamics of investment opportunities, the optimal investment strategy for a CRRA investor will be independent of her wealth level. Hence, we will focus on strategies of the form \( \pi(x, t) \) that only depends on the state variable and time (and not on wealth). The next theorem characterizes the expected utility generated by such an investment strategy.

**Theorem 7.11.** The expected utility generated by the investment strategy \( \pi_t = \pi(x_t, t) \) is
\[ J^\pi(W, x, t) = \frac{1}{1 - \gamma} \left( W e^{H^\pi(x, t)} \right)^{1 - \gamma}, \tag{7.44} \]
where the function \( H^\pi(x, t) \) satisfies the PDE
\begin{align*}
\frac{\partial H^\pi}{\partial t} + (m(x) - (\gamma - 1)\mu(x)) \pi(x, t)^\top \pi(x, t) + \frac{1}{2} \operatorname{tr} \left( H^\pi \frac{\partial x}{\partial x} \Sigma(x) \right) \\
- \frac{\gamma - 1}{2} \left( H^\pi \frac{\partial x}{\partial x} \right)^\top \Sigma(x) H^\pi + r(x) + \pi(x, t)^\top g(x, t) \left[ \lambda(x) - \frac{\gamma}{2} g(x, t)^\top \pi(x, t) \right] &= 0
\end{align*}
with the terminal condition \( H^\pi(x, T) = 0 \).

As explained in Section 5.4, we can associate a percentage wealth loss \( \ell_t \) with any given suboptimal investment strategy \( \pi \). The loss is implicitly defined by the relation
\[ J^\pi(W_t, x_t, t) = J(W_t[1 - \ell_t], x_t, t) \]
With \( \pi = \pi(x, t) \) and CRRA utility of terminal wealth only, it follows from Eqs. (7.41) and (7.44) that the loss can be stated as
\[ \ell_t = 1 - \exp \left\{ - |H(x_t, t) - H^\pi(x_t, t)| \right\} \approx H(x_t, t) - H^\pi(x_t, t). \]

Using these results, one can investigate various interesting suboptimal strategies, e.g.,

(i) the optimal strategy given that some assets are omitted from the portfolio,

(ii) the myopic, “no hedge” strategy, and

(iii) a certain absolute deviation from the optimal portfolio weights.

When the return dynamics have an affine or quadratic structure, the utility losses associated with these three suboptimal strategies can be derived from solving appropriate ordinary differential equations (ODEs). Obviously, case (i) allows us to evaluate the benefits of adding an extra asset class to the portfolio decision problem. Various recent academic papers have investigated portfolio choice models with various derivatives, corporate bonds, or other assets not traditionally included in a Merton-style model. From time to time innovative members of the financial industry promote investments in asset classes typically ignored. We provide a framework for a well-founded analysis of the investor welfare gains from expanding the investment universe. Case (ii) allows us to address the importance of intertemporal hedging. Some authors report that, for the specific model of return
dynamics they consider, the intertemporal hedging demand is quite small; see, e.g., Aït-Sahalia and Brandt (2001), Ang and Bekaert (2002), Brandt (1999), and Chacko and Viceira (2005). However, it is not clear that a small change in the long-term investment strategy cannot have a significant impact on the expected life-time utility. In fact, in a model with a constant risk-free rate and a single stock index with constant expected return and time-varying volatility, Gomes (2007) reports small intertemporal hedging demands and significant—although not dramatically large—utility losses from ignoring the hedge term. Case (iii) allows us to gauge the robustness of the optimal investment strategy, e.g., deviations from the truly optimal strategy due to applying a slightly mis-specified model or slightly inaccurate parameter values. The size of the utility loss from small perturbations of the optimal strategy will also indicate how frequent the portfolio should be rebalanced in practical implementations. Exercise 7.3 deals with case (iii).

For further discussions and examples see Larsen and Munk (2012).

7.6 Exercises

Exercise 7.1. Give a proof of Theorem 7.6.

Exercise 7.2. Verify the results stated in Section 7.4.

Exercise 7.3. Consider a trading strategy $\pi^{\varepsilon}$ which is a perturbation of the optimal strategy $\pi^*$ in the sense that

$$\pi^{\varepsilon}(x_t, t) = \pi^*(x_t, t) + (\varepsilon(x_t, t)^\top)^{-1} \varepsilon(x_t, t)$$

for some $\varepsilon(x, t)$ that can be interpreted as the error made in the assessment of the optimal sensitivity of wealth with respect to the shocks to asset prices. Let $\Delta^{\varepsilon}(x, t) = H(x, t) - H^{\pi^*}(x, t)$ so that the wealth loss is $\ell^{\pi^*}(x, t) = 1 - \exp\{-\Delta^{\varepsilon}(x, t)\} \approx \Delta^{\varepsilon}(x, t)$. Show that $\Delta^{\varepsilon}$ satisfies the PDE

$$\begin{align*}
(m(x) - (\gamma - 1)g(x)\left[\frac{1}{\gamma} \lambda(x) + \varepsilon(x, t)\right] - (\gamma - 1) \left[\frac{1}{\gamma} \varepsilon(x)^\top \eta(x) + \frac{\varepsilon(x) \eta(x)^\top}{\gamma}\right] H_x^* \right)^\top \Delta^{\varepsilon}_x \\
+ \frac{\partial \Delta^{\varepsilon}}{\partial t} + \frac{1}{2} \text{tr} \left(\Delta^{\varepsilon}_x \Sigma(x)\right) + \frac{\gamma - 1}{2} \left(\Delta^{\varepsilon}_x\right)^\top \Sigma(x) \Delta^{\varepsilon}_x + \frac{\gamma}{2} \|\varepsilon(x, t)\|^2 = 0
\end{align*}$$

(7.45)

with the terminal condition $\Delta^{\varepsilon}(x, T) = 0$. In particular, show that if $\varepsilon(x, t)$ is independent of $x$, the solution $\Delta^{\varepsilon}(x, t) = \Delta^{\varepsilon}(t)$ to

$$(\Delta^{\varepsilon})'(t) + \frac{\gamma}{2} \|\varepsilon(t)\|^2 = 0, \quad \Delta^{\varepsilon}(T) = 0,$$

will also solve the full PDE (7.45). Hence, the solution is

$$\Delta^{\varepsilon}(t) = \frac{\gamma}{2} \int_t^T \|\varepsilon(s)\|^2 ds.$$

Observe that the loss is increasing in the risk aversion, the time horizon, and the “squared error” $\|\varepsilon(s)\|^2$.

Exercise 7.4. In the models considered so far we have assumed a single consumption good, but modern economics offer an enormous variety of different consumption goods. The purpose of this exercise is to perform a preliminary analysis of how the presence of multiple consumption goods may affect the optimal consumption and investment strategies of an individual investor.
For simplicity, assume that the investor cares about only two consumption goods and both goods are perishable (non-storable). For $i = 1, 2$, let $c_{it}$ denote that units of good $i$ consumed at time $t$. Let good 1 be the numeraire so that its price is normalized to one at all times. The time $t$ price of good 2 is denoted by $\varphi_t$. To focus on the impact of multiple consumption goods, let us assume constant investment opportunities, i.e., we assume that the investor can invest in a risk-free asset with a constant annualized rate of return equal to $r$ and in $d$ risky assets with price dynamics

$$dP_t = \text{diag}(P_t) [(r1 + \sigma \lambda) \ dt + \sigma \ dz_t]$$

in the usual notation. Furthermore, assume that the price of good 2 follows a diffusion process

$$d\varphi_t = \mu_{\varphi}(\varphi_t) \ dt + \sigma_{\varphi}(\varphi_t)^\top dz_t + \hat{\sigma}_\varphi(\varphi_t) d\hat{z}_t.$$ 

Here $\hat{z}$ is a one-dimensional standard Brownian motion independent of the $d$-dimensional standard Brownian motion $z$.

We consider an individual with time-additive expected utility (and, for simplicity, we disregard any utility of terminal wealth) so that the indirect utility function is

$$J(W, \varphi, t) = \sup_{(c_1, c_2, \pi_\varphi) \in [t, T]} E_t \left[ \int_t^T e^{-\delta(s-t)} u(c_1(s), c_2(s)) \ ds \right].$$

(a) Explain why the HJB-equation associated with this problem can be written as

$$\delta J(W, \varphi, t) = \mathcal{L}^c J(W, \varphi, t) + \mathcal{L}^\pi J(W, \varphi, t) + \frac{\partial J}{\partial t}(W, \varphi, t) + rW J(W, \varphi, t)$$

$$+ \mu_{\varphi}(\varphi) J(W, \varphi, t) + \frac{1}{2} (\|\sigma_{\varphi}(\varphi)\|^2 + \hat{\sigma}_{\varphi}(\varphi)^2) J_{\varphi\varphi}(W, \varphi, t),$$

where

$$\mathcal{L}^c J = \sup_{c_1, c_2} \{ u(c_1, c_2) - (c_1 + c_2 \varphi) J_W \},$$

$$\mathcal{L}^\pi J = \sup_{\pi} \left\{ W J_W \pi^\top \sigma + \frac{1}{2} W^2 J_W W \pi^\top \sigma \pi + W J_W \pi \sigma \right\}.$$ 

(b) Show that the optimal consumption decisions at any point in time have the property that

$$\frac{u_2(c_1, c_2)}{u_1(c_1, c_2)} = \varphi,$$

where $u_i$ denotes the derivative of $u$ with respect to $c_i$. Interpret this result.

In the remainder of the exercise assume the Cobb-Douglas style utility function

$$u(c_1, c_2) = \frac{1}{1-\gamma} (c_1^{\alpha} c_2^{1-\alpha})^{1-\gamma},$$

where $\gamma > 0$ is the relative risk aversion and $\alpha \in (0, 1)$ captures the relative preference weights of the two goods.

(c) Show that the optimal consumption decisions imply that $c_2 \varphi = \frac{1-\alpha}{\alpha} c_1$ and interpret that result.

(d) Show that $\mathcal{L}^c J(W, \varphi, t) = \eta \xi J_W^{1-\gamma}$ for some constants $\eta$ and $\xi$ and determine those constants.
(e) Express the optimal portfolio $\pi$ in terms of relevant derivatives of $J$ and interpret your findings. How does the presence of two consumption goods affect the optimal portfolio?

(f) Show that

$$\mathcal{L}^\pi J = -\frac{1}{2} J_{W}^2 \|\lambda\|^2 - \frac{1}{2} J_{W\varphi}^2 \|\sigma_{\varphi}\|^2 - \frac{J_{W} J_{W\varphi}}{J_{WW}} \lambda^\top \sigma_{\varphi}.$$ 

(g) Conjecture that $J(W, \varphi, t) = \frac{1}{1 - \gamma} g(\varphi, t) W^{1-\gamma}$ and derive a partial differential equation for $g$.

(h) Is the market complete or incomplete?

In the remainder of the exercise assume that the price process for good 2 is a geometric Brownian motion spanned by the traded assets, i.e.,

$$d\varphi_t = \varphi_t [\tilde{\mu} dt + \tilde{\sigma}^\top dz_t],$$

where $\tilde{\mu}$ is a constant scalar and $\tilde{\sigma}$ a constant vector.

(i) Show that

$$g(\varphi, t) = \hat{\eta}_\varphi (1 - \alpha)^{\frac{1}{\gamma}} h(t)$$

solves the relevant partial differential equation for some constant $\hat{\eta}$ and some function $h(t)$.

(j) What is the optimal consumption and investment strategy in this case?
The martingale approach

8.1 The martingale approach in complete markets

The dynamic programming approach requires the existence of a finite-dimensional Markov process \( x = (x_t) \) such that the indirect utility function of the investor can be written as \( J_t = J(W_t, x_t, t) \). In contrast, the martingale approach does not require additional assumptions on the stochastic processes that the investor cannot control beyond those outlined in Section 5.2. In particular, we do not have to assume that the interest rates, price variances etc. are fully described by a finite-dimensional Markov process. The dynamic programming approach does not allow many conclusions on problems where the PDE cannot be solved explicitly. For example, it is hard to tell whether an optimal strategy actually exists. This question is easier to study with the martingale approach. In this section we consider the case where the market is complete. The subsequent section incorporates various portfolio constraints.

We go back to the general model for risky asset prices stated in (5.3). We consider a complete market so that the variations in the risk-free rate of return \( r_t \), expected rates of return \( \mu_t \), and variances and covariances defined by \( \sigma_t \) between rates of return are caused by the same \( d \)-dimensional standard Brownian motion \( z \) that affects the risky asset prices. Therefore, the market price of risk vector \( \lambda_t \) defined by

\[
\lambda_t = \sigma_t^{-1} (\mu_t - r_t 1)
\]

summarizes the risk-return tradeoff of all risks. In a complete market there is a unique state-price deflator process (a.k.a. the pricing kernel) \( \zeta = (\zeta_t) \) given by

\[
\zeta_t = \exp \left\{ - \int_0^t r_s \, ds - \int_0^t \lambda_s^\top d z_s - \frac{1}{2} \int_0^t \| \lambda_s \|^2 \, ds \right\}, \tag{8.1}
\]

Consequently (to be shown in Exercise 8.1), the state-price deflator evolves as

\[
d\zeta_t = -\zeta_t [r_t \, dt + \lambda_t^\top \, dz_t]. \tag{8.2}
\]

We also have a unique equivalent martingale measure (also known as the risk-neutral probability measure) \( Q \) defined by the Radon-Nikodym derivative \( dQ/dP = \exp \{ \int_0^T r_s \, ds \} \zeta_T \). We assume that
\( \lambda \) is an \( \mathcal{L}^2[0,T] \) process. The time zero price of a stochastic payoff \( X_T \) at some point \( T \) is given by
\[
E^Q \left[ e^{-\int_0^T r_s \, ds} X_T \right] = E \left[ \mathcal{Q}_T X_T \right].
\]
Similarly, the time \( t \) price is
\[
E^Q_t \left[ e^{-\int_t^T r_s \, ds} X_T \right] = E_t \left[ \mathcal{Q}_T X_T \right].
\]
For more information about state-price deflators, market prices of risk, and risk-neutral probabilities, see Björk (2009), Duffie (2001), Munk (2012) or other textbook presentations of modern asset pricing theory.

For simplicity we assume that the investor receives no income from non-financial sources. Then a natural constraint on the investor’s choice of consumption and portfolio strategy \((c, \pi)\) at time 0 is that
\[
E \left[ \int_0^T \zeta_t c_t \, dt + \zeta_T W_T \right] \leq W_0,
\]
where \( W_T \) is the terminal wealth induced by \((c, \pi)\) and \( W_0 \) is the initial wealth of the investor. This simply says that the time zero “price” of the strategy cannot exceed the initial wealth available. This is shown rigorously in the following theorem. But first we recall from (5.5) that wealth evolves as
\[
dW_t = W_t \left[ r_t + \pi_t^\top \sigma_t \lambda_t \right] \, dt - c_t \, dt + W_t \pi_t^\top \sigma_t \, dz_t.
\]
From this, (8.2), and Itô’s Lemma we get that
\[
d(\zeta_t W_t) = -\zeta_t c_t \, dt + \zeta_t W_t \left( \pi_t^\top \sigma_t - \lambda_t^\top \right) \, dz_t,
\]
or equivalently
\[
\zeta_t W_t + \int_0^t \zeta_s c_s \, ds = W_0 + \int_0^t \zeta_s W_s \left( \pi_s^\top \sigma_s - \lambda_s^\top \right) \, dz_s. \tag{8.3}
\]

**Theorem 8.1.** If \((c, \pi)\) is a feasible strategy, then
\[
E \left[ \int_0^T \zeta_t c_t \, dt + \zeta_T W_T \right] \leq W_0,
\]
where \( W_T \) is the terminal wealth induced by \((c, \pi)\).

**Proof.** Define the stopping times \((\tau_n)_{n \in \mathbb{N}}\) by
\[
\tau_n = T \wedge \inf \left\{ t \in [0,T] \left| \int_0^t \| \zeta_s W_s \left[ \pi_s^\top \sigma_s - \lambda_s \right] \|^2 \, ds \geq n \right. \right\}.
\]
Then the stochastic integral on the right-hand side of (8.3) is a martingale on \([0,\tau_n]\). Taking expectations in (8.3) leaves us with
\[
E[\mathcal{Q}_{\tau_n} W_{\tau_n}] + E \left[ \int_0^{\tau_n} \zeta_t c_t \, dt \right] = W_0.
\]
Letting \( n \uparrow \infty \), we have \( \tau_n \uparrow T \), and it can be shown by use of Lebesgue’s monotone convergence theorem that
\[
E \left[ \int_0^T \zeta_t c_t \, dt \right] = E \left[ \int_0^T \zeta_t c_t \, dt \right].
\]
Furthermore, Fatou’s lemma can be applied to show that
\[ \liminf_{n \to \infty} E [\zeta_n W_{\tau_n}] \geq E [\zeta_T W_T]. \]
The claim now follows. \hfill \Box

The idea of the martingale approach is to focus on the static optimization problem
\[ \sup_{(c, W)} \mathbb{E} \left[ \int_0^T e^{-\delta t} u(c_t) \, dt + e^{-\delta T} \bar{u}(W) \right], \quad (8.4) \]

rather than the original dynamic problem
\[ \sup_{(c, \pi)} \mathbb{E} \left[ \int_0^T e^{-\delta t} u(c_t) \, dt + e^{-\delta T} \bar{u}(W_T) \right], \]

s.t. \( dW_t = W_t \left[ r_t + \pi_t^\top \sigma_t \lambda_t \right] \, dt - c_t \, dt + W_t \pi_t^\top \sigma_t \, dz_t. \)

In the static problem the agent chooses the terminal wealth directly, whereas in the dynamic problem the terminal wealth follows from the portfolio strategy (and the consumption strategy). For the terminal wealth variable \( W \), the agent is allowed to choose among the non-negative, integrable and \( \mathcal{F}_T \)-measurable random variables. This approach was suggested by Karatzas, Lehoczky, and Shreve (1987) and Cox and Huang (1989, 1991). Some preliminary aspects were addressed by Pliska (1986).

The Lagrangian for the constrained optimization problem (8.4) is given by
\[ \mathcal{L} = \mathbb{E} \left[ \int_0^T e^{-\delta t} u(c_t) \, dt + e^{-\delta T} \bar{u}(W) \right] + \psi \left( W_0 - \mathbb{E} \left[ \int_0^T \zeta_t c_t \, dt + \zeta_T W \right] \right), \]

where \( \psi \) is a Lagrange multiplier. We can maximize the expectation in the last line by maximizing \( (e^{-\delta T} \bar{u}(W) - \psi \zeta_T W) \) with respect to \( W \) for each possible value of \( \zeta_T \) and maximizing \( (e^{-\delta t} u(c_t) - \psi \zeta_t c_t) \) with respect to \( c_t \) for each \( t \) and each possible value of \( \zeta_t \). This results in the first-order conditions
\[ e^{-\delta t} u'(c_t) = \psi \zeta_t, \quad e^{-\delta T} \bar{u}'(W) = \psi \zeta_T, \]

where \( \psi \) is then chosen such that the inequality constraint holds as an equality. Let \( I_u(\cdot) \) denote the inverse of the marginal utility function \( u'(\cdot) \) and \( I_{\bar{u}}(\cdot) \) the inverse of \( \bar{u}'(\cdot) \). Then the candidates for the optimal consumption and the optimal terminal wealth can be written as
\[ c_t = I_u(e^{\delta t} \psi \zeta_t), \quad W = I_{\bar{u}}(e^{\delta T} \psi \zeta_T). \]

The present value of this choice depends on the Lagrange multiplier \( \psi \):
\[ \mathcal{H}(\psi) = \mathbb{E} \left[ \int_0^T \zeta_t I_u(\psi e^{\delta t} \zeta_t) \, dt + \zeta_T I_{\bar{u}}(\psi e^{\delta T} \zeta_T) \right]. \quad (8.5) \]

We look for a multiplier \( \psi \) such that \( \mathcal{H}(\psi) = W_0 \) so that the entire budget is spend. Since marginal utility is decreasing, this is also the case for the inverse of marginal utility and hence also for the
function $\mathcal{H}$. We will assume that $\mathcal{H}(\psi)$ is finite for all $\psi > 0$. This condition should be verified in concrete applications. Under this assumption, $\mathcal{H}$ has an inverse denoted by $\psi$, and the appropriate Lagrange multiplier is $\psi = \psi(W_0)$. The next theorem says that the optimal policy in the static problem is feasible and optimal in the dynamic problem.

**Theorem 8.2.** Assume that $\mathcal{H}(\psi) < \infty$ for all $\psi > 0$. The optimal consumption rate is given by

$$c^*_t = u'(\psi(W_0)e^{\delta t} \zeta_t).$$

Under the optimal portfolio strategy the terminal wealth level is

$$W^* = I_u(\psi(W_0)e^{\delta T} \zeta_T).$$

The wealth process under the optimal policy is given by

$$W^*_t = \frac{1}{\zeta_t} E_t \left[ \int_t^T \zeta_t c^*_s ds + \zeta_T W^* \right]. \quad (8.6)$$

**Proof.** First note that for a concave and differentiable function $u$ we have that

$$\frac{u(\hat{c}) - u(c)}{\hat{c} - c} \geq u'(\hat{c})$$

for any $\hat{c} > c$ since the left-hand side is the slope of the line through the points $(c, u(c))$ and $(\hat{c}, u(\hat{c}))$ and the right-hand side is the slope of the tangent at $\hat{c}$. It follows immediately that

$$u(\hat{c}) - u(c) \geq u'(\hat{c})(\hat{c} - c).$$

A moment of reflection (maybe supported by a sketch of a graph) will convince you that the inequality holds even if $\hat{c} \leq c$. Let us take $\hat{c} = u(I_u(z))$ for some $z$. Then $u'(\hat{c}) = z$ so that we can conclude that

$$u(I_u(z)) - u(c) \geq z (I_u(z) - c), \forall c, z > 0.$$  

Analogously, we have

$$\bar{u}(I_u(z)) - \bar{u}(W) \geq z (I_u(z) - W), \forall W, z > 0.$$  

Hence, for any feasible strategy $(c, \pi)$ with associated terminal wealth $W$, we have that

$$E \left[ \int_0^T e^{-\delta t} (u(c^*_t) - u(c_t)) dt + e^{-\delta T} (\bar{u}(W^*) - \bar{u}(W)) \right]$$

$$\geq E \left[ \int_0^T \bar{y}(W_0) \zeta_t (c^*_t - c_t) dt + \bar{y}(W_0) \zeta_T (W^* - W) \right]$$

$$\geq 0,$$

where the last inequality follows from the fact that, by Theorem 8.1,

$$E \left[ \int_0^T \zeta_t c_t dt + \zeta_TW \right] \leq W_0,$$

and, per construction,

$$E \left[ \int_0^T \zeta_t c^*_t dt + \zeta_TW^* \right] = W_0.$$
Thus, if there is a portfolio strategy $\pi^*$ such that $(c^*, \pi^*)$ is feasible and gives a terminal wealth of $W^*$, then the strategy $(c^*, \pi^*)$ will be optimal. Define the process $W^*$ by (8.6). Obviously, 

$$
\zeta_t W_t^* + \int_0^t \zeta_s c_s^* ds = E_t \left[ \int_0^T \zeta_s c_s^* ds + \zeta_T W_T^* \right]
$$

defines a martingale, so by the martingale representation theorem, an adapted $L^2[0, T]$ process $\eta$ exists such that

$$
\zeta_t W_t^* + \int_0^t \zeta_s c_s^* = W_0 + \int_0^t \eta_s^* dz_s.
$$

(8.7)

Define a portfolio process $\pi$ by

$$
\pi_t = \left( \sigma_t^\top \right)^{-1} \left( \frac{\eta_t}{\zeta_t W_t^*} \lambda_t + \lambda_t \right)
$$

(with the remaining wealth $W_t^*(1 - \pi_t^\top 1)$ invested in the bank account). A comparison of (8.7) and (8.3) shows that the wealth process corresponding to this strategy together with the consumption strategy $c^*$ is exactly $(W_t^*)$. From (8.6), it is clear that terminal wealth is $W_T^* = W^*$.

Note that the indirect utility at time 0 as a function of initial wealth $W_0$ is

$$
J(W_0) = E \left[ \int_0^T e^{-\delta_t} u(c_s^*) ds + e^{-\delta T} \bar{u}(W^*) \right] = E \left[ \int_0^T e^{-\delta_t} \left( I_u(y(W_0)e^{\delta_t} \zeta_t) \right) dt + e^{-\delta T} \bar{u} \left( I_u(y(W_0)e^{\delta T} \zeta_T) \right) \right].
$$

We shall demonstrate how to apply the martingale approach on concrete consumption and investment choice problems in Sections 8.2 and 8.3. The martingale approach is in many aspects more elegant and it is better suited for answering the existence question under general conditions, cf. Cuoco (1997). However, the existence of an optimal portfolio strategy is based on the martingale representation theorem, which in itself does not give an explicit representation of the optimal portfolio, nor a way to compute it. In some settings the martingale approach can give an abstract characterization of both the optimal consumption and portfolio strategy even for non-Markov dynamics, but in order to obtain explicit expressions for the optimal strategies the setting is typically specialized to a Markov setting. So far, there are only a few examples of explicit solutions computed with the martingale approach where the solution could not have been easily found by an application of the dynamic programming approach. (See Munk and Sørensen (2004) for one example.) However, in some of the relatively simple problems, such as the complete markets case studied by Cox and Huang (1989), it can be shown that the optimal portfolio policies can be found by solving a partial differential equation (PDE), which has a simpler structure than the HJB equation.

### 8.2 Complete markets and constant investment opportunities

As discussed in Section 8.1 portfolio/consumption problems can also be analyzed using the so-called martingale approach instead of the dynamic programming approach used above. Recall that the application of the martingale approach is considerably more complex for incomplete markets, so we assume a complete market setting. We will try to get as far as possible without imposing
constant investment opportunities so that we will not have to start all over when we generalize to stochastic investment opportunities.

According to Theorem 8.2, if \( \varepsilon_1 > 0 \), the optimal consumption rate is given by

\[
c_t^* = I_u (\hat{y}(W_t)e^{\delta_t \zeta_t})
\]

and, if \( \varepsilon_2 > 0 \), the optimal level of terminal wealth level is

\[
W^* = I_u (\hat{y}(W_T)e^{\delta T \zeta_T})
\]

For the case of CRRA utility

\[
u(c) = \varepsilon_1 \frac{c^{1-\gamma}}{1-\gamma}, \quad \tilde{u}(W) = \varepsilon_2 \frac{W^{1-\gamma}}{1-\gamma},
\]

we have

\[
u'(c) = \varepsilon_1 c^{-\gamma}, \quad \tilde{u}'(W) = \varepsilon_2 W^{-\gamma}
\]

with inverse functions

\[
I_u(z) = \varepsilon_1^{1/\gamma} z^{-\frac{1}{\gamma}}, \quad I_u(z) = \varepsilon_2^{1/\gamma} z^{-\frac{1}{\gamma}},
\]

assuming that \( \varepsilon_1, \varepsilon_2 > 0 \). It turns out to be useful to define a process \( g = (g_t) \) by

\[
g_t = E_t \left[ \int_t^T \varepsilon_1^{1/\gamma} e^{-\frac{1}{\gamma}(u-st)} \left( \frac{\zeta_t}{S_t} \right)^{-1-1/\gamma} ds + \varepsilon_2^{1/\gamma} e^{-\frac{1}{\gamma}(T-t)} \left( \frac{\zeta_T}{S_T} \right)^{-1-1/\gamma} \right].
\]

Consequently, the function \( \mathcal{H} \) defined in (8.5) can be computed as

\[
\mathcal{H}(\psi) = E \left[ \int_0^T \varepsilon_1^{1/\gamma} e^{-\frac{1}{\gamma}(u-st)} \left( \frac{\zeta_t}{S_t} \right)^{-1-1/\gamma} dt + \varepsilon_2^{1/\gamma} e^{-\frac{1}{\gamma}(T-t)} \left( \frac{\zeta_T}{S_T} \right)^{-1-1/\gamma} \right]
\]

with inverse function

\[
\hat{y}(W_0) = W_0^{-\gamma} g_0.
\]

Therefore, the optimal consumption policy is

\[
c_t^* = \varepsilon_1^{1/\gamma} e^{-\frac{1}{\gamma}T} \hat{y}(W_0)^{-\frac{1}{\gamma}} \zeta_t^{-\frac{1}{\gamma}} = \varepsilon_1^{1/\gamma} \frac{W_0}{g_0} e^{-\frac{1}{\gamma}T} \zeta_t^{-\frac{1}{\gamma}}
\]

\[
e^{-\frac{1}{\gamma}T} \zeta_t^{-\frac{1}{\gamma}} W_0 \left( E \left[ \int_0^T e^{-\frac{1}{\gamma}t} \zeta_t^{-1-1/\gamma} dt + \left( \frac{\varepsilon_2}{\varepsilon_1} \right)^{1/\gamma} e^{-\frac{1}{\gamma}T} \zeta_T^{-1-1/\gamma} \right] \right)^{-1}, \tag{8.8}
\]

and the optimal terminal wealth level is

\[
W^* = \varepsilon_2^{1/\gamma} e^{-\frac{1}{\gamma}T} \hat{y}(W_0)^{-\frac{1}{\gamma}} \zeta_T^{-\frac{1}{\gamma}} = \varepsilon_2^{1/\gamma} \frac{W_0}{g_0} e^{-\frac{1}{\gamma}T} \zeta_T^{-\frac{1}{\gamma}}
\]

\[
e^{-\frac{1}{\gamma}T} \zeta_T^{-\frac{1}{\gamma}} W_0 \left( E \left[ \int_0^T \left( \frac{\varepsilon_1}{\varepsilon_2} \right)^{1/\gamma} e^{-\frac{1}{\gamma}t} \zeta_t^{-1-1/\gamma} dt + e^{-\frac{1}{\gamma}T} \zeta_T^{-1-1/\gamma} \right] \right)^{-1}.
\]
The wealth process under the optimal policy is given by

\[ W_t^* = \frac{1}{\zeta_t} E_t \left[ \int_t^T \zeta_s c_s^* ds + \zeta_T W^* \right] \]

\[ = \frac{W_0}{g_t} \frac{1}{\zeta_t} E_t \left[ \int_t^T \varepsilon_1^{1/\gamma} e^{-\frac{\gamma}{2} \zeta_s} s^{1-\frac{1}{\gamma}} ds + \varepsilon_2^{1/\gamma} e^{-\frac{T}{2} \zeta_T} \right] \]

\[ = \frac{W_0}{g_t} e^{-\frac{T}{2} \zeta_T} e^{-\frac{1}{2} \zeta_s} E_t \left[ \int_t^T \varepsilon_1^{1/\gamma} e^{-\frac{\gamma}{2} (s-t)} \left( \frac{\zeta_s}{\zeta_t} \right)^{1-\frac{1}{\gamma}} ds + \varepsilon_2^{1/\gamma} e^{-\frac{T}{2} (T-t)} \left( \frac{\zeta_T}{\zeta_t} \right)^{1-\frac{1}{\gamma}} \right] \]

\[ = \frac{W_0}{g_t} e^{-\frac{T}{2} \zeta_T} e^{-\frac{1}{2} \zeta_s} g_t. \]  

(8.9)

Consequently,

\[ \frac{W_t^*}{g_t} = \frac{W_0}{g_t} e^{-\frac{T}{2} \zeta_T} e^{-\frac{1}{2} \zeta_s}. \]

We see immediately from (8.8) that we can rewrite the optimal time \( t \) consumption rate as

\[ c_t^* = \varepsilon_1^{1/\gamma} W_t^* \]

so that \( g_t \) is proportional to the optimal wealth-to-consumption ratio. Moreover, for \( s > t \), we have

\[ c_s^* = \frac{W_0}{g_t} \varepsilon_1^{1/\gamma} e^{-\frac{\gamma}{2} \zeta_s} \left( \frac{\zeta_s}{\zeta_t} \right)^{-\frac{1}{\gamma}} = \frac{W_0}{g_t} \varepsilon_1^{1/\gamma} e^{-\frac{\gamma}{2} (s-t)} \left( \frac{\zeta_s}{\zeta_t} \right)^{-\frac{1}{\gamma}}, \]

which states the uncertain consumption rate at time \( s \) given information available at time \( t \).

Similarly, we can express the optimal terminal wealth as

\[ W^* = \frac{W_t^*}{g_t} \varepsilon_2^{1/\gamma} e^{-\frac{T}{2} (T-t)} \left( \frac{\zeta_T}{\zeta_t} \right)^{-\frac{1}{\gamma}}. \]

(8.11)

The indirect utility at time \( t \) is

\[ J_t = E_t \left[ \int_t^T e^{-\delta (s-t)} u(c_s^*) ds + e^{-\delta (T-t)} u(W^*) \right] \]

\[ = \frac{1}{1-\gamma} E_t \left[ \int_t^T e^{-\delta (s-t)} \varepsilon_1 (c_s^*)^{1-\gamma} ds + e^{-\delta (T-t)} \varepsilon_2 (W^*)^{1-\gamma} \right] \]

\[ = \frac{1}{1-\gamma} \left( \frac{W_t^*}{g_t} \right)^{1-\gamma} E_t \left[ \int_t^T e^{-\frac{T}{2} (s-t)} \varepsilon_1^{1/\gamma} \left( \frac{\zeta_s}{\zeta_t} \right)^{1-1/\gamma} ds + e^{-\frac{T}{2} (T-t)} \varepsilon_2^{1/\gamma} \left( \frac{\zeta_T}{\zeta_t} \right)^{1-1/\gamma} \right] \]

\[ = \frac{1}{1-\gamma} g_t^{\gamma} (W_t^*)^{1-\gamma}, \]

where the third equality is due to (8.10) and (8.11), whereas the last equality follows from the definition of \( g_t \).

The equations above are generally valid for CRRA utility. Now let us specialize to the case of constant investment opportunities, where the state-price deflator is

\[ \zeta_t = e^{-rt - \lambda^T z_t - \frac{t}{2} |\lambda|^2 t}. \]
Consequently, future values of the state-price deflator are lognormally distributed. Note that for any $s > t$, we have\footnote{The third equality is due to the following result: For a random variable $x \sim N(m, s^2)$, $E[e^{-ax}] = e^{-am + \frac{1}{2}a^2s^2}$. In our case $a = 1 - \frac{1}{\gamma}$ and $x = \lambda^\top(z_s - z_t) = \sum_{i=1}^d \lambda_i(z_{is} - z_{it})$ is normally distributed with mean zero and variance $\sum_{i=1}^d \lambda_i^2(s - t) = ||\lambda||^2(s - t)$.}

$$
E_t \left[ e^{-\frac{1}{\gamma}(s-t)} \left( \frac{\zeta_s}{\zeta_t} \right)^{1-1/\gamma} \right] = E_t \left[ e^{-\frac{1}{\gamma}(s-t)} \left( e^{-r(s-t) - \lambda^\top(z_s - z_t) - \frac{1}{2}||\lambda||^2(s-t)} \right)^{1-1/\gamma} \right]
$$

$$
= e^{-\frac{1}{\gamma}(s-t)} e^{-\frac{1}{2}(1-\frac{1}{\gamma})r(s-t) - \frac{1}{2}(1-\frac{1}{\gamma})||\lambda||^2(s-t)} E_t \left[ e^{-\frac{1}{2}(1-\frac{1}{\gamma})\lambda^\top(z_s - z_t)} \right]
$$

$$
= e^{-\frac{1}{\gamma}(s-t)} e^{-\frac{1}{2}(1-\frac{1}{\gamma})r(s-t) - \frac{1}{2}(1-\frac{1}{\gamma})||\lambda||^2(s-t)} e^{\frac{1}{2}(1-\frac{1}{\gamma})||\lambda||^2(s-t)}
$$

$$
= e^{-\left(\frac{1}{\gamma}(1-\frac{1}{\gamma}) - \frac{1}{\gamma}\frac{1}{\gamma}\right)||\lambda||^2(s-t)}
$$

$$
= e^{-A(s-t)},
$$

where $A$ is again the constant given by (6.11). Now we can compute $g_t$ in closed form:

$$
g_t = E_t \left[ \int_t^T e^{1/\gamma} e^{-\frac{1}{\gamma}(s-t)} \left( \frac{\zeta_s}{\zeta_t} \right)^{1-1/\gamma} ds + e^{1/\gamma} e^{-\frac{1}{\gamma}(T-t)} \left( \frac{\zeta_T}{\zeta_t} \right)^{1-1/\gamma} \right]
$$

$$
= \int_t^T e^{1/\gamma} E_t \left[ e^{-\frac{1}{\gamma}(s-t)} \left( \frac{\zeta_s}{\zeta_t} \right)^{1-1/\gamma} \right] ds + e^{1/\gamma} E_t \left[ e^{-\frac{1}{\gamma}(T-t)} \left( \frac{\zeta_T}{\zeta_t} \right)^{1-1/\gamma} \right]
$$

$$
= \int_t^T e^{1/\gamma} e^{-A(s-t)} ds + e^{1/\gamma} e^{-A[T-t]}
$$

$$
= \frac{1}{A} \left( e^{1/\gamma} + [Ae^{1/\gamma} - e^{1/\gamma}] e^{-A[T-t]} \right),
$$

which is deterministic and identical to the function $g(t)$ defined in (6.12). Hence, for the case of constant investment opportunities, the formulas for the optimal consumption rate and the indirect utility derived above coincide with the results obtained by use of the dynamic programming approach.

It remains to derive the optimal investment strategy. The optimal wealth process is given in (8.9). Since we know by now that $g_t$ is deterministic, the only stochastic process on the right-hand side is the state-price deflator $\zeta_t$. With constant investment opportunities the dynamics of the state-price deflator is

$$
d\zeta_t = -\zeta_t [r dt + \lambda^\top dz_t].
$$

Applying Itô’s Lemma we can now derive the dynamics of the optimal wealth. Focusing on the volatility term, we get

$$
dW^*_t = \ldots dt - \frac{W^*_t}{\zeta_t} d\zeta_t
$$

$$
= \ldots dt + W^*_t \frac{1}{\gamma} \lambda^\top dz_t.
$$

If we compare with the dynamics of the wealth for any given investment strategy $\pi = (\pi_t)$ stated in (6.1), we see that the optimal wealth process is obtained with the investment strategy

$$
\pi^*_t = \frac{1}{\gamma} (g^\top)^{-1} \lambda,
$$

as we found out using the dynamic programming approach.
8.3 Complete markets and stochastic investment opportunities

In this section we will apply the martingale approach to solve the consumption/portfolio problem in a situation with stochastic investment opportunities. The martingale approach was introduced in Section 8.1. In Section 8.2 we used the martingale approach to solve the consumption-portfolio problem of a CRRA investor in the case of constant investment opportunities. Also in this section we will assume complete markets and CRRA preferences for both intermediate consumption and terminal wealth corresponding to $\varepsilon_1 = \varepsilon_2 = 1$.

We know already from Section 8.2 that the optimal time $t$ consumption rate is

$$c^*_t = \frac{W_0}{g_0} e^{-\frac{1}{\gamma} \int_0^t \zeta_s ds} = \frac{W^*_t}{g_t},$$

where $W^*_t$ is the wealth at time $t$ if the optimal strategies are pursued, and the process $g = (g_t)$ is defined by

$$g_t = \mathbb{E}_t \left[ \int_t^T e^{-\frac{1}{\gamma} (s-t)} \left( \frac{\zeta_s}{\zeta_t} \right)^{1-1/\gamma} ds + e^{-\frac{1}{\gamma} (T-t)} \left( \frac{\zeta_T}{\zeta_t} \right)^{1-1/\gamma} \right].$$

The optimal terminal wealth level is

$$W^* = \frac{W_0}{g_0} e^{-\frac{1}{\gamma} \int_0^T \zeta_t^{-\frac{1}{\gamma}}}.$$

The indirect utility at time $t$ is

$$J_t = \frac{1}{1-\gamma} g_t (W^*_t)^{1-\gamma}.$$

Furthermore, the wealth process under the optimal policy is given by

$$W^*_t = \frac{W_0}{g_0} e^{-\frac{1}{\gamma} t \zeta_t^{-\frac{1}{\gamma}}} g_t.$$

If $r$ and $\lambda$ are constant, $g_t$ is a deterministic function of time and the optimal investment strategy is given in Section 8.2. If the investment opportunities are stochastic in the sense that $r$ or $\lambda$ or both are stochastic processes, then $g$ is a stochastic process. Write the dynamics of $g$ as

$$dg_t = g_t \left[ \mu_{gt} dt + \sigma_{gt}^\top dz_t \right],$$

for some drift process $\mu_g = (\mu_{gt})$ and some sensitivity process $\sigma_g = (\sigma_{gt})$. The optimal wealth is a function of $t$, $\zeta_t$, and $g_t$. Recall that the dynamics of the state-price deflator $\zeta_t$ is

$$d\zeta_t = -\zeta_t \left[ r_t dt + \lambda_{zt}^\top dz_t \right].$$

An application of Itô’s Lemma gives that the dynamics of optimal wealth is

$$dW^*_t = \ldots dt - \frac{1}{\gamma} \frac{W^*_t}{\zeta_t} d\zeta_t + \frac{W^*_t}{g_t} dg_t = \ldots dt + W^*_t \left( -\frac{1}{\gamma} \lambda_t + \sigma_{gt} \right)^\top dz_t.$$

Comparing with the dynamics of wealth for any given portfolio, we can conclude that an optimal investment strategy is

$$\pi^*_t = \frac{1}{\gamma} \left( \sigma_{gt}^{-\top} \right)^{-1} \lambda_t + \left( \sigma_{gt}^{-\top} \right)^{-1} \sigma_{gt}. $$
This result was first derived by Munk and Sørensen (2004). It is a natural generalization of the results obtained in Markov settings using the dynamic programming approach. The hedge term of the portfolio is matching the volatility of the process \( g \) which is important for consumption. Looking at the definition of \( g \), we can see that only variations in the state-price deflator, i.e., in interest rates and market prices of risk, will be hedged. This is also in line with findings in Markov set-ups. Of course, \( \sigma_g \) has to be identified in order for this result to be of practical relevance.

This is possible in many concrete cases, primarily cases with Markov dynamics where the dynamic programming approach also applies, i.e., in affine or quadratic diffusion models. But Munk and Sørensen (2004) consider a relevant and non-trivial example with non-Markov dynamics.

For investors with logarithmic utility (\( \gamma = 1 \)), we see that the process \((g_t)\) is always deterministic so that the volatility \( \sigma_g \) is zero. The optimal portfolio of a log investor is therefore \( \pi^*_t = \frac{1}{\gamma} \left( \xi_t^T \right)^{-1} \lambda_t \) as has already been shown for Markov settings.

### 8.4 The martingale approach with portfolio constraints

This note provides a short introduction to the martingale approach to dynamic consumption and portfolio choice problems in the case with constraints on the allowed portfolios. For details and further results, see the original work by He and Pearson (1991), Karatzas, Lehoczky, Shreve, and Xu (1991), Cvitanić and Karatzas (1992), Xu and Shreve (1992a, 1992b), Cuoco (1997), and Munk (1997b, Ch. 3), as well as the textbook presentations by Korn (1997, Ch. 4) and Karatzas and Shreve (1998, Ch. 6). Warning: all these references employ a lot of high-level mathematics.

#### 8.4.1 A general representation of portfolio constraints

We consider a financial market where \( d + 1 \) assets can potentially be traded, possibly with some constraints on the portfolios allowed. One of the asset will be denoted by asset 0 and represents a locally risk-free asset with return process \( r = (r_t) \), i.e., price process

\[
P_{0t} = \exp\left\{ \int_0^t r_u du \right\}.
\]

The other \( d \) assets are risky with prices given by the vector \( P_t = (P_{1t}, \ldots, P_{dt})^\top \) satisfying

\[
dP_t = \text{diag}(P_t)[\mu_t dt + \sigma_t dz_t],
\]

where \( z_t \) is a \( d \)-dimensional standard Brownian motion. \( \sigma_t \) is assumed to have full rank \( d \) implying the dynamic completeness of the market, at least potentially. None of the assets pay dividends over the period \([0, T]\) of interest to the investor considered below. Alternatively, we can think of \( P_{it} \) as the time \( t \) value that is obtained by purchasing one unit of asset \( i \) at time 0 and reinvesting any dividends received from asset \( i \) by purchasing additional units of the same asset.

A trading strategy is a pair \((\theta_0, \theta_t)\), where \( \theta_0 \) is a one-dimensional (adapted) and \( \theta_t = (\theta_1, \ldots, \theta_d)^\top \) is a \( d \)-dimensional (progressively measurable) stochastic process. \( \theta_0 \) denotes the dollar amount invested in the savings account at time \( t \). \( \theta_t \) is the dollar amount invested at time \( t \) in the \( i \)th risky asset, \( i = 1, \ldots, d \).

Let \( \mathcal{K} \) be a non-empty, closed, convex subset of \( \mathbb{R}^{d+1} \). A trading strategy \((\theta_0, \theta_t)\) is called \( \mathcal{K} \)-admissible if \((\theta_0, \theta_t)^\top \in \mathcal{K} \) for all \( t \in [0, T] \) and all states and \((\theta_0, \theta_t)\) satisfies some integrability
conditions ensuring that the value of the trading strategy is well-defined. \( \mathcal{K} \) is called the portfolio constraint set. Various interesting specifications of \( \mathcal{K} \) are listed below. The set of \( \mathcal{K} \)-admissible trading strategies is denoted by \( \mathcal{P}(\mathcal{K}) \). A consumption process is a non-negative (progressively measurable) process \( c \) in \( \mathcal{L}^1[0, T] \). The set of consumption processes is denoted by \( \mathcal{C} \).

Given a trading strategy \( (\theta_0, \theta) \in \mathcal{P}(\mathcal{K}) \) and a consumption process \( c \in \mathcal{C} \), the dynamics of the investor’s wealth \( W_t = W_{t_0}^{\theta_0, \theta, c} \) is

\[
dW_t = [\theta_0^T r_t + \theta_t^T \mu_t + y_t - c_t] \, dt + \theta_t^T \sigma_t \, dz_t.
\]

(8.12)

Initial wealth is \( W_0 = w \). Here \( y \) is a non-negative (progressively measurable) stochastic process representing the endowment stream of the agent, e.g., labor income. Since \( \theta_0^T = W_0 - \theta_1^T 1 \), we can rewrite the wealth dynamics as

\[
dW_t = [r_t W_t + \theta_t^T (\mu_t - r_t 1) + y_t - c_t] \, dt + \theta_t^T \sigma_t \, dz_t,
\]

which does not involve \( \theta_0 \) explicitly. Note, however, that there may be constraints on the investment in the instantaneously risk-free asset.

A triple \( (\theta_0, \theta, c) \) is called \( \mathcal{K} \)-admissible given the initial wealth \( w \) if

(i) \( (\theta_0, \theta) \in \mathcal{P}(\mathcal{K}), c \in \mathcal{C}, \)

(ii) \( W_t^{\theta_0, \theta, c} \geq -\mu \) at all times \( t \in T \) for some positive constant \( K \),

(iii) \( W_T^{\theta_0, \theta, c} \geq 0 \).

Let \( \mathcal{A}(w; \mathcal{K}) \) denote the set of triples \( (\theta_0, \theta, c) \), which are \( \mathcal{K} \)-admissible with initial wealth \( w \).

In some situations, it is advantageous to let the agent choose a terminal wealth \( W \) directly instead of choosing a trading strategy \( (\theta_0, \theta) \). A consumption/terminal wealth pair \((c, W)\), where \( c \in \mathcal{C} \) and \( W \) is a non-negative \( T \)-measurable random variable with finite expectations, is called \( \mathcal{K} \)-admissible with initial wealth \( w \), if there exists a trading strategy \( (\theta_0, \theta) \) such that \( (\theta_0, \theta, c) \) is \( \mathcal{K} \)-admissible with \( W_0^{\theta_0, \theta, c} = w \) and \( W_T^{\theta_0, \theta, c} = W \). In that case \( (\theta_0, \theta) \) is said to finance \((c, W)\).

Let \( \mathcal{A}'(w; \mathcal{K}) \) denote the set of \( \mathcal{K} \)-admissible consumption/terminal wealth pairs \((c, W)\). Clearly, if \( (\theta_0, \theta, c) \in \mathcal{A}(w; \mathcal{K}) \), then \( (c, W_0^{\theta_0, \theta, c}) \in \mathcal{A}'(w; \mathcal{K}) \).

Note that we can model situations, where the endowment stream is not spanned by traded assets, i.e., where \( y \) is not adapted to the filtration generated by traded assets, by letting \( y \) depend on, say, \( P_d \) and then restricting the investor to a policy with values in (a subset of) \( \mathbb{R} \times \mathbb{R}^{d-1} \times \{0\} \).

By restricting the individuals to \( \mathcal{K} \)-admissible processes, a number of interesting situations can be examined. It turns out that the so-called support function of \( -\mathcal{K} \) plays an important role. Let \( \nu = (\nu_0, \nu) \in \mathbb{R} \times \mathbb{R}^d \). Then the support function \( \delta : \mathbb{R}^{d+1} \to \mathbb{R} \cup \{-\infty, +\infty\} \) of \( -\mathcal{K} \) is defined by

\[
\delta(\nu) = \sup_{(\theta_0, \theta) \in \mathcal{K}} (-\theta_0 \nu_0 - \theta^T \nu).
\]

The effective domain of \( \delta \), i.e. the set of \( \nu \in \mathbb{R}^{d+1} \) for which \( \delta(\nu) < \infty \), is denoted by \( \tilde{\mathcal{K}} \). Next, we list a few interesting properties of \( \delta \) and \( \tilde{\mathcal{K}} \). See, e.g., Rockafellar (1970, Sect. 13) for more on support functions.

(i) \( \tilde{\mathcal{K}} \) is a closed convex cone\(^2\), called the barrier cone of \( -\mathcal{K} \).

\(^2\)A set \( D \subseteq \mathbb{R}^N \) is called a cone if \( ax \in D \) whenever \( x \in D \) and \( a > 0 \).
(ii) If $K$ is a cone, then $\delta \equiv 0$ on $\tilde{K}$.

(iii) $\delta$ is sub-additive, that is
\[
\delta(\nu_1) + \delta(\nu_2) \geq \delta(\nu_1 + \nu_2),
\]
which follows from the corresponding property of the supremum operator.

(iv) If $(\theta_0, \theta) \in K$ and $\nu \in \tilde{K}$, then
\[
\theta_0 \nu_0 + \theta^\top \nu + \delta(\nu) \geq 0.
\]
(8.13)

Of course, this follows trivially from the definition of $\delta$.

It turns out that we need to impose the following assumption on $K$.

**Assumption 8.1.** $K$ is such that $\delta$ is bounded from above on $\tilde{K}$, or, equivalently, $\delta$ is non-positive on $\tilde{K}$ and $\nu_0 \geq 0$ for all $\nu \in \tilde{K}$.

Note that we are considering constraints on the amounts invested in the different assets. Cvitanić and Karatzas (1992) started all this, but considered constraints on portfolio weights, which is less general than constraints on amounts invested. Munk (1997b) extended/adapted the results of Cvitanić and Karatzas (1992) to constraints on amounts invested, which is particularly important to cover labor income where portfolio weights might not be well-defined. Here are examples of interesting constraint sets:

**Example 8.1.** [Complete market] A complete market corresponds to having $K = \mathbb{R}^{d+1}$. This implies that $\tilde{K} = \{0\}^{d+1}$ and $\delta(\nu) = 0$ for all $\nu \in \tilde{K}$. This is the standard market structure, in which (in various degrees of generality) consumption/portfolio problems are studied by, e.g., Merton (1969, 1971), Karatzas, Lehoczky, and Shreve (1987), and Cox and Huang (1989, 1991). $\square$

**Example 8.2.** [Non-traded assets] A situation where there are only $m < d$ tradable risky assets, but otherwise no constraints on the tradable assets, can be modeled by letting $K = \mathbb{R} \times \mathbb{R}^m \times \{0\}^{d-m}$. In that case, $\tilde{K} = \{0\} \times \{0\}^m \times \mathbb{R}^{d-m}$ and $\delta(\nu) = 0$ on $\tilde{K}$. $\square$

**Example 8.3.** [Short-sale constraints] To model prohibition of short-selling the risky assets number $m + 1, \ldots, d$, let $K = \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^{d-m}$. Then $\tilde{K} = \{0\} \times \{0\}^m \times \mathbb{R}^{d-m}$ and again $\delta(\nu) = 0$ on $\tilde{K}$. $\square$

**Example 8.4.** [Buying constraints] With $K = \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^{d-m}$, the investor is not allowed to have positive amounts invested in the last $d - m$ risky assets. Then $\tilde{K} = \{0\} \times \{0\}^m \times \mathbb{R}^{d-m}$ and $\delta(\nu) = 0$ on $\tilde{K}$. $\square$

**Example 8.5.** [Portfolio mix constraints] $K = \{ (\theta_0, \theta) \in \mathbb{R}^{d+1} \mid x \equiv \theta_0 + \theta^\top \mathbf{1} \geq 0 \text{ and } \pi \in \hat{K}(x) \}$, where $\hat{K}(x)$ is some non-empty, closed, convex subset of $\mathbb{R}^d$ containing the origin, and $v\pi = \theta/x$
for $x > 0$ and $v \pi = 0$ for $x = 0$, models a portfolio mix constraint. In this case

$$\tilde{K} = \{ \nu \in \mathbb{R}^{d+1} \mid \nu^T(\theta_0, \theta) \geq 0, \forall (\theta_0, \theta) \in \mathcal{K} \}$$

and $\delta(\nu) = 0$ on $\tilde{K}$. \hfill \square

**Example 8.6.** [Collateral constraints] With $\mathcal{K} = \{ (\theta_0, \theta) \in \mathbb{R}^{d+1} \mid \psi^T(\theta_0, \theta) \geq 0 \}$, where $\psi \in [0,1]^{d+1}$, we can model the situation, where, using the $j$’th security ($j = 0, 1, \ldots, d$) as collateral, it is only possible to borrow the fraction $\psi_j$ of its value. In this case $\tilde{K} = \psi \mathbb{R}_+ = \{ k\psi \mid k \geq 0 \}$ and $\delta(\nu) = 0$ on $\tilde{K}$. \hfill \square

**Example 8.7.** [Minimum capital requirements] Let $\mathcal{K} = \{ (\theta_0, \theta) \in \mathbb{R}^{d+1} \mid \theta_0 + \theta^T 1 \geq k \}$, where $k \in \mathbb{R}_+$. Then $\tilde{K} = \mathbb{R}_+ 1_{t+1} = \{ (\psi, \ldots, \psi) \in \mathbb{R}^{d+1} \mid \psi \geq 0 \}$, and $\delta(\nu) = -k\nu_0$ for $\nu = (\nu_0, \nu) \in \mathcal{K}$. The special minimum capital requirement $k = 0$ represents a borrowing constraint. \hfill \square

**Example 8.8.** [Combinations of constraints] Any combination of the above constraints, i.e., where $\mathcal{K}$ is the intersection of some of the $\mathcal{K}$’s of the previous examples. \hfill \square

### 8.4.2 The problem to solve

The general utility maximization problem to solve is

$$J(w) = \sup_{(\theta_0, \theta, c) \in \mathcal{A}(w; \mathcal{K})} V^{\theta_0, \theta, c}(w),$$

where

$$V^{\theta_0, \theta, c}(w) = E \left[ \int_0^T U_1(c, s) \, ds + U_2(W_T^{\theta_0, \theta, c}, T) \right]$$

and it is understood that the wealth process starts at $W_0^{\theta_0, \theta, c} = w$. Equivalently, we can solve

$$J(w) = \sup_{(c, W) \in \mathcal{A}'(w; \mathcal{K})} V^{c, W}(w),$$

where

$$V^{c, W}(w) = E \left[ \int_0^T U_1(c, s) \, ds + U_2(W, T) \right].$$

We assume that the utility functions $U_1(\cdot, t)$ and $U_2(\cdot, T)$ have infinite marginal utility at zero, i.e., $U_1'(0, t) \equiv \lim_{x \downarrow 0} U_1'(c, t) = \infty$ and similarly $U_2'(0, T) \equiv \lim_{w \downarrow 0} U_2'(W, T) = \infty$. A technical aside: we have to modify the definition of the set of admissible policies such that now $\mathcal{A}(w; \mathcal{K})$ denotes the set of strategies $(\theta_0, \theta, c)$ which are admissible in the sense explained above and, further, satisfy the condition

$$E \left[ \int_0^T U_1(c, t)^{\pi} \, dt + U_2(W_T^{\theta_0, \theta, c}, T)^{\pi} \right] < \infty$$

and similarly for $\mathcal{A}'(w; \mathcal{K})$.

---

$^3$Here $X^- = \max\{0, -X\}$. 

---
8.4.3 Auxiliary unconstrained problems

We will define a set of artificial, auxiliary unconstrained markets. Given a process \((\nu_0, \nu_t)\), where \((\nu_0, \nu_t) \in \tilde{\mathcal{K}}\) for any \(t \in [0, T]\), we define a market \(\mathcal{M}_\nu\) where the short-term risk-free rate, the expected returns on the risky assets, and the income rate are perturbed relative to the true market:

(i) the risk-free rate at time \(t\) is \(r_t + \nu_0 t\),

(ii) the drift vector of the risky asset prices is \(\mu_t + \nu_t\),

(iii) the income rate is \(y_t + \delta(\nu_t)\).

There are no portfolio constraints in the artificial market \(\mathcal{M}_\nu\), i.e., it is a complete market. The unique market price of risk is

\[
\lambda_{\nu t} = \frac{\sigma_t^{-1}}{2} (\mu_t + \nu_t - (r_t + \nu_0 t)) \mathbf{1},
\]

the change of measure to the unique risk-neutral measure \(Q_\nu\) is captured by

\[
dQ_\nu dP = Z_{\nu T},
\]

where

\[
Z_{\nu t} = \exp \left\{ -\int_0^t \lambda_{\nu s}^\top d\mathbf{z}_s - \frac{1}{2} \int_0^t \lambda_{\nu s}^\top \lambda_{\nu s} ds \right\},
\]

and the unique state-price deflator is given by

\[
\zeta_{\nu t} = \exp \left\{ -\int_0^t (r_s + \nu_0 s) ds \right\} Z_{\nu t}.
\]

In general, \(Z_\nu\) is a local martingale. For technical reasons, we have to restrict ourselves to \(\nu\)'s for which \(Z_\nu\) is a true martingale. Let \(\mathcal{N}^*\) be the set of such processes \(\nu\), i.e.,

\[
\mathcal{N}^* = \left\{ \nu \in \mathcal{L}^2[0, T] \mid \nu(t, \omega) \in \tilde{\mathcal{K}}, \forall (t, \omega) \in [0, T] \times \Omega \text{ and } Z_\nu \text{ is a martingale} \right\}.
\]

The wealth process in the auxiliary market \(\mathcal{M}_\nu\) corresponding to any investment/consumption policy \((\theta_0, \theta, c)\) is the process \(W_{\nu t}^{\theta_0, \theta, c}\) given by

\[
dW_{\nu t}^{\theta_0, \theta, c} = (\theta_0 [r_t + \nu_0 t] + \theta_t^\top [\mu_t + \nu_t]) dt - (c_t - y_t - \delta(\nu_t)) dt + \theta_t^\top \sigma_t d\mathbf{z}_t = (\theta_0 r_t + \theta_t^\top \mu_t) dt - (c_t - y_t) dt + \theta_t^\top \sigma_t d\mathbf{z}_t + (\delta(\nu_t) + \theta_0 \nu_0 + \theta_t^\top \nu_t) dt. \quad (8.14)
\]

Note that, from (8.13),

\[
\delta(\nu_t) + \theta_0 \nu_0 + \theta_t^\top \nu_t \geq 0,
\]

so a comparison of (8.14) and (8.12) yields that

\[
W_{\nu t}^{\theta_0, \theta, c} \geq W_t^{\theta_0, \theta, c} \quad (8.15)
\]

path-by-path: following a given strategy you will always end up with at least as high a terminal wealth in any artificial market as in the true market.

A triple \((\theta_0, \theta, c)\) consisting of a trading strategy \((\theta_0, \theta)\) and a consumption process \(c\) is called admissible in \(\mathcal{M}_\nu\) [with initial wealth \(w\)] if \((\theta_0, \theta, c)\) and \(W_{\nu t}^{\theta_0, \theta, c}\) satisfy the same conditions as a \(\mathcal{K}\)-admissible triple in the original market except for the requirement \((\theta_0 t, \theta_t) \in \mathcal{K}, \forall t\). The set of
triples \((\theta_0, \theta, c)\) admissible in \(M_\nu\) is denoted \(A_\nu(w)\), i.e.,
\[
A_\nu(w) = \left\{ (\theta_0, \theta, c) \in \mathcal{P}(\mathbb{R}^{d+1}) \times \mathcal{C} \left| W_{\nu t}^{\theta_0, \theta, c} \geq -K, t \in [0, T], \int_0^T U_1(c_t, t)\, dt + U_2(W_{\nu T}^{\theta_0, \theta, c}, T) < \infty \right. \right\}.
\]

The unconstrained utility maximization problem in \(M_\nu\) is
\[
J_\nu(w) = \sup_{(\theta_0, \theta, c) \in A_\nu(w)} V^{\theta_0, \theta, c}(w).
\]

We let \((\theta_\nu^0, \theta_\nu, c_\nu)\) denote the optimal strategy in the market \(M_\nu\), i.e., \(J_\nu(w) = V^{\theta_\nu^0, \theta_\nu, c_\nu}(w)\). As before, we can also maximize over consumption and terminal wealth:
\[
J_\nu(w) = \sup_{(c, W) \in A_\nu'(w)} V^{c, W}(w).
\]

Let \((c^\nu, W^\nu)\) denote the optimal consumption process and terminal wealth in the market \(M_\nu\), i.e., \(J_\nu(w) = V^{c^\nu, W^\nu}(w)\). Admissibility means budget-feasible in the sense that
\[
E \left[ \int_0^T \zeta_\nu(c_t - y_t - \delta(\nu_t))\, dt + \zeta_\nu W_T \right] \leq w,
\]
plus some technical integrability conditions.

### 8.4.4 Linking the artificial markets to the true market

Due to (8.15), we can conclude that \((\theta_0, \theta, c) \in A(w; \mathcal{K}) \Rightarrow (\theta_0, \theta, c) \in A_\nu(w)\). Consequently,
\[
J(w) \leq J_\nu(w), \quad \forall \nu \in \mathbb{N}^*.
\]

The indirect utility obtainable in any of the artificial markets is at least as high as the indirect utility in the true market. The main result of Cvitanić and Karatzas (1992) and Munk (1997b, Ch. 3) is to provide the following four ways to characterize optimality in the true market via the artificial markets:

1. **Minimality of \(\nu\):** The optimal trading strategy in an artificial market is not necessarily \(\mathcal{K}\)-valued and is therefore not necessarily admissible in the true market. If we can find an artificial market \(M_\nu\) in which the optimal strategy \((\theta_\nu^0, \theta_\nu, c_\nu)\) is also admissible in the true market, then it is clear that
\[
J(w) \geq V^{\theta_\nu^0, \theta_\nu, c_\nu}(w) = J_\nu(w).
\]

Combining that with (8.16), we can conclude that
\[
J(w) = J_\nu(w) = V^{\theta_\nu^0, \theta_\nu, c_\nu}(w)
\]
so that \((\theta_\nu^0, \theta_\nu, c_\nu)\) is the optimal strategy also in the true market. It is clear that \(J(w) = J_\nu(w)\) can only be satisfied in the least favorable artificially unconstrained market, i.e., we should minimize the indirect utility over all artificial markets.
2. **Financiability of** $(c^\nu, W^\nu)$: Suppose we can find a $\nu$ so that the optimal consumption and terminal wealth $(c^\nu, W^\nu)$ is financed by a trading strategy $(\theta^\nu_0, \theta^\nu)$, which is $\mathcal{K}$-valued and satisfies
\[
\delta(\nu_t) + \theta^\nu_{0t} \nu_{0t} + (\theta^\nu_t)^T \nu_t = 0
\]
for all $t$ and all states. Then it follows from (8.14) that the strategy will generate the same terminal wealth in the true market as in the artificial market $\mathcal{M}_\nu$. Since the strategy is admissible in the true market, we have
\[
J(w) \geq V^{\theta^\nu_0, \theta^\nu, c^\nu}(w) = J_\nu(w),
\]
and again we can combine that with (8.16) and conclude that $(\theta^\nu_0, \theta^\nu, c^\nu)$ is optimal in the true market.

3. **Parsimony of** $\nu$: If we can find a $\nu \in \mathbb{N}^*$ such that $(c^\nu, W^\nu) \in \mathcal{C} \times \mathcal{L}_1^\nu$ satisfies
\[
\mathbb{E} \left[ \int_0^T \zeta_{\nu t} (c^\nu_t - y_t - \delta(\nu_t)) \, dt + \zeta_{\nu T} W^\nu \right] \leq w, \quad \forall \tilde{\nu} \in \mathbb{N}^*,
\]
then $(c^\nu, W^\nu)$ and the corresponding strategy $(\theta^\nu_0, \theta^\nu, c^\nu)$ are optimal in the true market. This proof is complicated and will be skipped here. Note that the left-hand side of the above inequality is the cost of implementing $(c^\nu, W^\nu)$ in the artificial market $\mathcal{M}_\nu$. For $\tilde{\nu} = \nu$, the above inequality will be satisfied as an equality. The intuition is that if we can find an artificial market for which the optimal strategy is budget-feasible in all other artificial markets, then it is the least expensive and hence the least favorable of the solutions to the artificial market problems.

4. **Dual optimality of** $\nu$: The unconstrained maximization problem
\[
J_\nu(w) = \sup_{(c, W)} \mathbb{E} \left[ \int_0^T U_1(c_s, s) \, ds + U_2(W, T) \right],
\]
s.t. $\mathbb{E} \left[ \int_0^T \zeta_{\nu t} (c_t - y_t - \delta(\nu_t)) \, dt + \zeta_{\nu T} W \right] \leq w,$

can be solved with Lagrangian technique. If $\psi$ denotes the Lagrange multiplier on the budget constraint, the solution can be written as $c_t = I_1(\psi \zeta_{\nu t}, t)$, $W = I_2(\psi \zeta_{\nu T}, T)$, where $I_1(\cdot, t)$ and $I_2(\cdot, T)$ are the inverse functions of $U_1(\cdot, t)$ and $U_2(\cdot, T)$, respectively. Substituting the solution back into the objective function, we obtain $\tilde{V}^\nu(\psi) + \psi w$, where
\[
\tilde{V}^\nu(\psi) = \mathbb{E} \left[ \int_0^T \tilde{U}_1(\psi \zeta_{\nu t}, t) \, dt + \tilde{U}_2(\psi \zeta_{\nu T}, T) \right] + \psi \mathbb{E} \left[ \int_0^T \zeta_{\nu t} (y_t + \delta(\nu_t)) \, dt \right],
\]
and $\tilde{U}_1$ and $\tilde{U}_2$ are the convex conjugates of $U_1$ and $U_2$, respectively, i.e.
\[
\tilde{U}_1(x, t) = \sup_{q > 0} \{ U_1(q, t) - qx \} = U_1(I_1(x, t), t) - xI_1(x, t),
\]
and similarly for $\tilde{U}_2$. The problem
\[
\tilde{J}(\psi) = \inf_{\nu \in \mathbb{N}^*} \tilde{V}^\nu(\psi)
\]
is called the dual problem. The Lagrange multiplier in $M_\nu$ is related to initial wealth $w$ via some function $\psi = Y_\nu(w)$ which ensures that the budget constraint is satisfied as an equality. It can then be shown that the dual problem is linked to the original problem as follows: if we can find a $\nu \in \mathbb{N}^\ast$ such that

$$\tilde{J}(\psi) = \tilde{V}_\nu(\psi) \quad \text{for} \quad \psi = Y_\nu(w),$$

then $M_\nu$ is the least favorable market and $(\theta_0^\nu, \theta^\nu, c^\nu)$ is optimal for the original constrained problem in the true market.

The dual problem leads to a way of proving the existence of an optimal consumption and investment strategy in the constrained true market. If there is a solution to the dual problem, then there is also a solution to the primal problem, i.e., the utility maximization problem in the true market. Cvitanić and Karatzas (1992) state sufficient conditions for the existence of an optimal solution to the dual problem, which are then also sufficient conditions for the existence of an optimal consumption and investment strategy in the constrained true market. However, one of the conditions is that the Arrow-Pratt relative risk aversion measures corresponding to the utility functions $U_1$ and $U_2$ are smaller than or equal to one, whereas individuals are generally believed to have a relative risk aversion greater than one. Cuoco (1997) attacks the primal problem directly using alternative methods and is able to establish less restrictive conditions for the existence of an optimal solution.

The results above provide important intuition for constrained utility maximization problems. The results have been used to provide explicit solutions to some constrained utility maximization problems, but only with simple constraints and simple preferences such as logarithmic utility. The ideas of considering the dual problem and the artificially unconstrained markets have also been used recently in various numerical solution techniques, cf. Haugh, Kogan, and Wang (2006) and Bick, Kraft, and Munk (2012).

### 8.5 Exercises

**Exercise 8.1.** Show that (8.2) follows from (8.1).
CHAPTER 9

Numerical methods for solving dynamic asset allocation problems

If the problem features CRRA utility, no portfolio constraints, and return dynamics that do not fit into neither the affine nor the quadratic class: with (i) utility of terminal wealth only or (ii) complete markets and utility of intermediate consumption and/or terminal wealth, it suffices to solve a PDE like (7.40) numerically. This can be done (when the dimension of the state variable is three or lower) using standard methods, like finite difference methods. See, e.g., Wilmott, Dewynne, and Howison (1993), Thomas (1995), Wilmott (1998), Tavella and Randall (2000), Seydel (2009), and Munk (2011). With incomplete markets and intermediate consumption, the more complicated PDE (7.39) has to be solved numerically. For more general preferences, we would normally have to solve the even more complicated PDE (7.9) numerically.

In many realistic cases, the portfolios are constrained and these constraints have to be taken into account when solving the HJB equation, and then closed-form solutions are generally impossible to find. It is still possible (at least, for low-dimensional problems) to implement a finite difference type recursive solution method to solve the relevant HJB equation (a variant is called the Markov Chain Approximation Approach). See, e.g., Brennan, Schwartz, and Lagnado (1997), Fitzpatrick and Fleming (1991), Munk (1997a, 2003), Van Hemert (2010), and Munk and Sørensen (2010). A more or less equivalent approach used in some papers is to assume a discrete-time setting from the beginning and then solve the Bellman equation by backwards recursive dynamic programming. However, some authors use large time steps (e.g., allow only annual consumption and investment decisions) and assume very simple distributions (binomial, trinomial) of the relevant state variables over these long time steps. See, e.g., Campbell and Cocco (2003), Cocco (2005), Cocco, Gomes, and Maenhout (2005), and Yao and Zhang (2005a).

An alternative which, at least potentially, can handle higher-dimensional problems is Monte Carlo simulation based approaches to the HJB equation. Various versions have been proposed. See, e.g., Detemple, Garcia, and Rindisbacher (2003, 2005), Cvitanić, Goukasian, and Zapatero (2003), Brandt, Goyal, Santa-Clara, and Stroud (2005), van Binsbergen and Brandt (2007), and Kooijen, Nijman, and Werker (2007, 2010).
Yet another alternative for the solution of some consumption and portfolio choice problems involving portfolio constraints and/or incomplete markets is suggested by Bick, Kraft, and Munk (2012). The method applies to CRRA utility and return dynamics of the affine-quadratic type. The method combines (i) the idea of artificially unconstrained and complete markets introduced in connection with the martingale approach in Section 8.4 and (ii) the results on closed-form solutions for unconstrained affine-quadratic settings and CRRA utility. The method considers a subset of artificially unconstrained and complete markets for which relatively simple closed-form solutions exist. Each of these strategies is transformed into a feasible strategy in the true, constrained market. This gives a set of feasible strategies parameterized by a number of parameters. If this number is fairly low, one can search for the best of the strategies, where the evaluation of the strategy is done via Monte Carlo simulation. The method also provides an upper bound for the true, unknown optimal expected utility (given by the worst of the considered artificial markets) and thus an upper bound on the wealth-equivalent loss the individual might suffer by following the best of the feasible strategies considered. Another numerical method building on the martingale techniques was suggested by Haugh, Kogan, and Wang (2006).
CHAPTER 10

Asset allocation with stochastic interest rates

10.1 Introduction

It is an empirical fact that both nominal and real interest rates and bond yields vary stochastically over time. It is therefore natural to include the short-term interest rate $r_t$ as a state variable. This was first done in a portfolio-choice context by Merton (1973b) who considered a general one-factor dynamics for $r_t$, but he was not able to go beyond the general characterization of the investment strategy in (7.7). We will focus on individuals with CRRA utility and on models in which the interest rate dynamics is of an affine form, since then we can obtain closed-form solutions for the optimal strategies as explained in Section 7.3.2. The affine class includes the well-known models of Vasicek (1977) and Cox, Ingersoll, and Ross (1985). See, e.g., Munk (2011) for a comprehensive analysis of dynamic models of the term structure of interest rates. We can also apply the general results of Section 7.3 to cases where the dynamics of the term structure of interest rate is given by a multi-factor affine or quadratic model.

Recall that investors are (or should be) concerned about real interest rates and hence they would want to invest in real bonds. Indeed, we will assume that investors have access to trade in a complete market of real bonds (Exercise 10.1 at the end of the chapter discusses an optimal investment problem with stochastic interest rates when no bonds are traded.) We will focus on determining the optimal bond/stock mix so we assume that only a single stock is traded. We interpret this stock as the entire stock market index. The results can be generalized to the case with multiple stocks.

Investors with non-log utility will hedge variations in interest rates. Bonds carry a build-in hedge against interest rate risk since bond prices are inversely related to interest rates. Over a period where interest rates have fallen, indicating that future investment opportunities are worsened, bond prices have risen and generated a positive return. The converse is also true. We will therefore expect that interest rates are hedged by investing in bonds, but precisely how many bonds and which bonds this hedge should involve has to be computed using concrete models.

In Section 10.2 we study the case where the real short-term interest rate behaves according to
the Vasicek model. Section 10.3 considers the CIR model of interest rates. Section 10.4 gives a numerical example in which the quantitative effects of interest rate uncertainty on optimal portfolios can be assessed. Section 10.5 briefly looks at optimal portfolio choice when interest rate dynamics is given by a two-factor version of the Vasicek model. Other studies with stochastic interest rates are briefly discussed in Section 10.6. In many countries there are no liquid markets for real bonds, only for nominal bonds. Then we have to take the dynamics of consumer prices and inflation into account. We consider those issues in Chapter 12.

10.2 One-factor Vasicek interest rate dynamics

Following Vasicek (1977), assume that \( r_t \) follows the Ornstein-Uhlenbeck process

\[
\frac{dr_t}{r_t} = \kappa [\bar{r} - r_t] dt - \sigma_r dz_t,
\]

with an associated constant market price of risk \( \lambda_1 \). We assume that \( \kappa, \bar{r}, \) and \( \sigma \) are positive constants. The process exhibits mean reversion in the sense that, if \( r_t < \bar{r} \), the short rate is expected to increase over the next instant, whereas if \( r_t > \bar{r} \), the short rate is expected to fall. This is a very realistic feature of the model. Future values of the short rate are normally distributed so, in particular, short rates can take on any negative value, which is not realistic.

It is a consequence of these assumptions that the price of a zero-coupon bond with maturity \( \bar{T} \) is given by

\[
B_{t\bar{T}} = e^{-a(T-t)-b(T-t)r_t},
\]

where

\[
b(\tau) = \frac{1}{\kappa} \left( 1 - e^{-\kappa \tau} \right),
\]

\[
a(\tau) = y_\infty (\tau - b(\tau)) + \frac{\sigma^2}{4\kappa} b(\tau)^2,
\]

where \( y_\infty = \left( \bar{r} + \frac{\lambda_1 \sigma_r}{\kappa} - \frac{\sigma^2}{2\kappa^2} \right) \) is the asymptotic zero-coupon yield as time-to-maturity goes to infinity. From Itô’s Lemma it follows that the dynamics of the zero-coupon bond price is

\[
\frac{dB_{t\bar{T}}}{B_{t\bar{T}}} = B_{t\bar{T}} \left[ (r_t + \lambda_1 \sigma_r b(\bar{T}-t)) dt + \sigma_r b(\bar{T}-t) dz_t \right],
\]

and similarly the dynamics of any bond (or any other fixed-income security) is of the form

\[
\frac{dB_t}{B_t} = B_t \left[ (r_t + \lambda_1 \sigma_B(r_t,t)) dt + \sigma_B(r_t,t) dz_t \right]. \tag{10.1}
\]

It is well-known that any bond (or other fixed-income security) can be generated from an appropriate dynamic investment strategy in the bank account and in just one (arbitrary) bond (or other long-lived term structure derivative). Let us for the present take an arbitrary bond with price \( B_t \) and dynamics given by (10.1).

The price of the single stock (representing the stock market index) is assumed to follow the process

\[
\frac{dS_t}{S_t} = S_t \left[ (r_t + \psi \sigma_S) dt + \rho \sigma_S dz_1 + \sqrt{1-\rho^2} \sigma_S dz_2 \right].
\]

The parameter \( \rho \) is the correlation between bond market returns and stock market returns, \( \sigma_S \) is the volatility of the stock, and \( \psi \) is the Sharpe ratio of the stock which we assume constant.
The asset allocation problem of a CRRA investor under these assumptions was studied by Serensen (1999) and Bajeux-Besnainou, Jordan, and Portait (2001) for utility of terminal wealth only. Korn and Kraft (2001) discuss some technical issues related to the application of the verification theorem to this problem.

To get this into the notation applied so far, we rewrite the price dynamics as

\[
\begin{pmatrix}
 dB_t \\
 dS_t
\end{pmatrix} = \begin{pmatrix}
 B_t & 0 \\
 0 & S_t
\end{pmatrix} \left[ \begin{pmatrix}
 r_1 \mathbf{1} + \left( \frac{\sigma_B(r_1, t)}{\rho \sigma_S} \right) \left( \frac{1 - r^2 \sigma^2}{\sqrt{1 - r^2 \sigma^2}} \right) \left( \frac{\lambda_1}{\lambda_2} \right) \\
 0 & \frac{\sigma_B(r_1, t)}{\rho \sigma_S} \left( \frac{1 - r^2 \sigma^2}{\sqrt{1 - r^2 \sigma^2}} \right) \left( \frac{dz_1t}{dz_2t} \right)
\end{pmatrix} \right] dt
\]

where

\[
\lambda_2 = (\psi - \rho \lambda_1) / \sqrt{1 - \rho^2}.
\] (10.2)

We are therefore in a complete market model with a single state variable \((x = r)\). We can rewrite the dynamics of \(r\) as

\[
dr_t = \kappa [\bar{r} - r_t] dt + (-\sigma_r, 0) d\mathbf{z}_t,
\]

where \(\mathbf{z} = (z_1, z_2)^T\). In this model the state variable has an affine drift and a constant volatility, and the market price of risk vector \(\mathbf{\lambda} = (\lambda_1, \lambda_2)^T\) is also constant. Hence, Theorem 7.7 applies with CRRA utility from terminal wealth only and Theorem 7.8 applies with CRRA utility from intermediate consumption and possibly terminal wealth. In the notation used there, we have

\[
\begin{align*}
\tau_0 &= 0, \\
r_1 &= 1, \\
m_0 &= \kappa \bar{r}, \\
m_1 &= -\kappa, \\
\Lambda_0 &= \lambda_1^2 + \lambda_2^2, \\
\Lambda_1 &= 0, \\
\bar{v}_0 &= 0, \\
\bar{v}_1 &= 0, \\
V_0 &= \sigma^2_r, \\
V_1 &= 0, \\
K_0 &= -\sigma_r \lambda_1, \\
K_1 &= 0.
\end{align*}
\]

In this case the ordinary differential equation (7.22) reduces to

\[
A'_1(\tau) = 1 - \kappa A_1(\tau),
\]

which with the initial condition \(A_1(0) = 0\) has the unique solution

\[
A_1(\tau) = \frac{1}{\kappa} (1 - e^{-\kappa \tau}) = b(\tau).
\]

This result also follows from the discussion below Theorem 7.7. Next, \(A_0\) follows from (7.25):

\[
A_0(\tau) = \frac{1}{2\gamma} (\lambda_1^2 + \lambda_2^2) \tau + \left( \kappa \bar{r} + \frac{\gamma - 1}{\gamma - \sigma_r \lambda_1} \right) \int_0^\tau b_s ds - \frac{\gamma - 1}{2\gamma} \sigma^2_s \int_0^\tau b_s ds
\]

\[
= \frac{1}{2\gamma} (\lambda_1^2 + \lambda_2^2) \tau + \left( \bar{r} - \frac{\gamma - 1}{2\kappa^2 \gamma} \left( \sigma^2_r - 2\kappa \sigma_r \lambda_1 \right) \right) (\tau - b(\tau)) + \frac{\gamma - 1}{4\kappa \gamma} \sigma^2_r b(\tau)^2,
\]

where we have used that

\[
\int_0^\tau b_s ds = \frac{1}{\kappa} (\tau - b(\tau)), \quad \int_0^\tau b_s^2 ds = \frac{1}{\kappa^2} (\tau - b(\tau)) - \frac{1}{2\kappa} b(\tau)^2.
\]

For the case with utility from terminal wealth only we have from Theorem 7.7 that the optimal investment strategy is

\[
\Pi(W, r, t) \equiv \begin{pmatrix}
\Pi_B(W, r, t) \\
\Pi_S(W, r, t)
\end{pmatrix} = \frac{1}{\gamma} \left( \sigma(r, t)^\tau \right)^{-1} \mathbf{\lambda} - \frac{\gamma - 1}{\gamma} \left( \sigma(r, t)^\tau \right)^{-1} \left( \frac{-\sigma_r}{0} \right) b(T - t)
\]

\[
= \frac{1}{\gamma} \left( \sigma(r, t)^\tau \right)^{-1} \mathbf{\lambda} + \frac{\gamma - 1}{\gamma} \left( \sigma(r, t)^\tau \right)^{-1} \left( \frac{\sigma_r}{0} \right) b(T - t).
\] (10.3)
We can see that the hedge portfolio only involves the bond, not the stock, which should not come as a surprise since bonds seem more appropriate for hedging interest rate risk than stocks. The higher the risk aversion $\gamma$, the lower the investment in the tangency portfolio and the higher the investment in the hedge bond. The inverse of the transposed volatility matrix is

$$
\left( \begin{array}{cc}
\sigma_B(r,t) & \rho \sigma_S \\
0 & \sqrt{1-\rho^2} \sigma_S
\end{array} \right)^{-1} = \frac{1}{\sqrt{1-\rho^2} \sigma_B(r,t) \sigma_S} \left( \begin{array}{cc}
\sqrt{1-\rho^2} \sigma_S & -\rho \sigma_S \\
0 & \sigma_B(r,t)
\end{array} \right)
$$

so that we can write out the fraction of wealth invested in the stock and the bond as

$$
\Pi_S(W, r, t) = \frac{\lambda_2}{\gamma \sigma_S \sqrt{1 - \rho^2}},
$$

$$
\Pi_B(W, r, t) = \frac{1}{\gamma \sigma_B(r,t)} \left( \lambda_1 - \frac{\rho}{\sqrt{1-\rho^2}} \lambda_2 \right) + \frac{\gamma - 1}{\gamma} \frac{\sigma_r b(T-t)}{\sigma_B(r,t)}.
$$

If the bond in the portfolio is the zero-coupon bond maturing at the end of the investment horizon of the investor, i.e., at time $T$, then $\sigma_B(r,t) = \sigma_r b(T-t)$, and we see that the hedge term simply consists of a fraction $(\gamma - 1)/\gamma$ in the zero-coupon bond. This is a natural choice of hedge instrument since it is exactly the truly risk-free asset for an investor only interested in time $T$ wealth. The log utility investor ($\gamma = 1$) does not hedge. The hedge position of a less risk averse investor ($\gamma < 1$) is negative, while a more risk averse investor ($\gamma > 1$) takes a long position in the bond in order to hedge interest rate risk. An infinitely risk averse investor ($\gamma \to \infty$) will invest her entire wealth in the zero-coupon bond maturing at $T$.

If we continue to use the zero-coupon bond maturing at $T$ as the bond instrument, we see from (10.3) that we can write the risky part of the optimal investment strategy as

$$
\Pi(W, r, t) \equiv \begin{pmatrix}
\Pi_B(W, r, t) \\
\Pi_S(W, r, t)
\end{pmatrix} = \frac{1}{\gamma} \left( \sigma(t)^\tau \right)^{-1} \lambda + \frac{\gamma - 1}{\gamma} \begin{pmatrix}
1 \\
0
\end{pmatrix}.
$$

Consequently, the fraction of wealth invested in the bank account (i.e., the locally risk-free asset) is

$$
\Pi_0(W, r, t) = 1 - \Pi_B(W, r, t) - \Pi_S(W, r, t) = 1 - \frac{1}{\gamma} \left( \sigma(t)^\tau \right)^{-1} \lambda - \frac{\gamma - 1}{\gamma} = \frac{1}{\gamma} \left( 1 - \left( \sigma(t)^\tau \right)^{-1} \lambda \right).
$$

Note that the term in the parenthesis is exactly what a log investor would hold in the bank account. The entire investment strategy can be written as

$$
\begin{pmatrix}
\Pi_0 \\
\Pi_B \\
\Pi_S
\end{pmatrix} = \frac{1}{\gamma} \begin{pmatrix}
\Pi_0^{\log} \\
\Pi_B^{\log} \\
\Pi_S^{\log}
\end{pmatrix} + \frac{\gamma - 1}{\gamma} \begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix}.
$$

The strategy is hence a simple combination of the log investor’s portfolio and the zero-coupon bond maturing at the investment horizon of the investor. Note that as the risk aversion $\gamma$ increases, the position in stocks will decrease, while the position in bonds will increase. Hence, the bond/stock ratio increases with risk aversion consistent with popular advice. However, the allocation to stock
is still independent of the investment horizon which conflicts with traditional advice that the stock weight should increase with the investment horizon.

With utility from intermediate consumption only, it follows from Theorem 7.8 that the hedge term of the optimal bond investment strategy is

\[
\frac{\gamma - 1}{\gamma} \sigma_B(r, t) \frac{\int_t^T e^{-\frac{1}{2}(s-t) - \frac{1}{2}A_0(s-t) - \frac{1}{2}b(s-t)r} b(s-t) ds}{\int_t^T e^{-\frac{1}{2}(s-t) - \frac{1}{2}A_0(s-t) - \frac{1}{2}b(s-t)r} ds},
\]

(10.6)

where \(\sigma_B(r, t)\) again represents the volatility of the bond chosen for implementing the strategy. It can be shown that the time \(t\) volatility of a coupon bond paying a continuous coupon at a deterministic rate \(K(s)\) up to time \(T\) is given by

\[
\sigma_B(r, t) = \frac{\int_t^T K(s)B_t^s \sigma_r b(s-t) ds}{\int_t^T K(s)B_t^s ds}.
\]

Hence, we can interpret the time \(t\) interest rate hedge as the fraction \((\gamma - 1)/\gamma\) of wealth invested in a bond with continuous coupon

\[
K(s) = e^{a(s-t) - \frac{1}{2}A_0(s-t) - \frac{1}{2}(s-t) + \frac{1}{2}b(s-t)r}.
\]

Munk and Sørensen (2004) show that this coupon is closely related to the expected consumption rate at time \(s\). For an investor with utility from consumption over the entire period \([t, T]\), the zero-coupon bond maturing at \(T\) is no longer the truly risk-free asset. Since the investor is interested in payments at all dates in \([t, T]\), he hedges interest rate risk by investing in a combination of all zero-coupon bonds maturing in this interval, i.e., in some sort of coupon bond.

### 10.3 One-factor CIR dynamics

Consider the same set-up as above except that the short-term interest rate now is assumed to follow the square-root process

\[
dr_t = \kappa[\bar{r} - r_t] dt - \sigma_r \sqrt{r_t} dz_{1t},
\]

(10.7)

where \(\kappa, \bar{r},\) and \(\sigma_r\) are positive constants. The market price of the risk represented by \(z_1\) is assumed to be given by \(\lambda_1(r, t) = \lambda_1 \sqrt{r_t}/\sigma_r\). As shown by Cox, Ingersoll, and Ross (1985), zero-coupon bond prices are on the form

\[
B_t^T = e^{-a(T-t) - b(T-t)r_t},
\]
as in the Vasicek model, but \(a\) and \(b\) are now given by

\[
b(\tau) = \frac{2(e^{\alpha \tau} - 1)}{(\alpha + \bar{\kappa})(e^{\alpha \tau} - 1) + 2\alpha},
\]

\[
a(\tau) = -\frac{2\kappa \bar{r}}{\sigma_r^2} \left(\frac{1}{2}(\bar{\kappa} + \alpha)\tau + \ln \frac{2\alpha}{(\alpha + \bar{\kappa})(e^{\alpha \tau} - 1) + 2\alpha}\right),
\]

where \(\bar{\kappa} = \kappa - \lambda_1\) and \(\alpha = \sqrt{\bar{\kappa}^2 + 2\sigma_r^2}\). The price evolves as

\[
dB_t^T = B_t^T \left[ (r_t + b(T-t)\lambda_1 r_t) dt + b(T-t)\sigma_r \sqrt{r_t} dz_{1t} \right].
\]

We assume that the investor can also trade in a single stock with price \(S_t\) evolving as

\[
dS_t = S_t \left[(r_t + \psi(r_t)\sigma_S) dt + \rho \sigma_S dz_{1t} + \sqrt{1 - \rho^2} \sigma_S dz_{2t}\right].
\]
Here \( \sigma_S \) is a positive constant, and \( z_2 \) is a one-dimensional standard Brownian motion independent of \( z_1 \) so that the constant \( \rho \) is the instantaneous correlation between stock returns and bond returns. We assume that the market price of risk associated with \( z_2 \) is a constant \( \lambda_2 \) so that

\[
\psi(r) = \rho \frac{\lambda_1}{\sigma_r} \sqrt{r} + \sqrt{1 - \rho^2} \lambda_2. \tag{10.8}
\]

Again we have an affine, complete market model of the type studied in Section 7.3.2. In this case we have

\[
\begin{align*}
    r_0 &= 0, & r_1 &= 1, & m_0 &= \kappa r, & m_1 &= -\kappa, \\
    \Lambda_0 &= \frac{\lambda_2^2}{\sigma_r^2}, & \Lambda_1 &= \frac{\lambda_1^2}{\sigma_r^2}, & \hat{v}_0 &= 0, & \hat{v}_1 &= 0, \\
    V_0 &= 0, & V_1 &= \sigma_r^2, & K_0 &= 0, & K_1 &= -\lambda_1.
\end{align*}
\]

The solution is stated in terms of two deterministic functions \( A_0 \) and \( A_1 \). Let

\[
\bar{\kappa} = \kappa - \frac{\gamma - 1}{\gamma} \lambda_1.
\]

The ordinary differential equation (7.22) for \( A_1 \) becomes

\[
A'_1(\tau) = \left( 1 + \frac{\lambda_1^2}{2 \gamma \sigma_r^2} \right) - \bar{\kappa} A_1(\tau) - \frac{\gamma - 1}{2 \gamma} \sigma_r^2 A_1(\tau)^2
\]

with the initial condition \( A_1(0) = 0 \). Assuming

\[
\bar{\kappa}^2 + 2 \sigma_r^2 \frac{\gamma - 1}{\gamma} \left( 1 + \frac{\lambda_1^2}{2 \gamma \sigma_r^2} \right) > 0,
\]

which is certainly satisfied for \( \gamma \geq 1 \), the unique solution is follows immediately from (7.24):

\[
A_1(\tau) = \frac{2 \left( 1 + \frac{\lambda_1^2}{2 \gamma \sigma_r^2} \right) (e^{\nu \tau} - 1)}{\left( \nu + \bar{\kappa} \right) (e^{\nu \tau} - 1) + 2 \nu},
\]

and we have introduced the additional auxiliary parameters

\[
\nu = \sqrt{\bar{\kappa}^2 + 2 \sigma_r^2 \frac{\gamma - 1}{\gamma} \left( 1 + \frac{\lambda_1^2}{2 \gamma \sigma_r^2} \right)}.
\]

\( A_0 \) can then be computed from (7.25):

\[
A_0(\tau) = \frac{\lambda_2^2}{2 \gamma} \tau + \kappa \bar{\kappa} \int_0^\tau A_1(s) \, ds = \frac{\lambda_2^2}{2 \gamma} \tau - \frac{\gamma - 1}{\gamma} \frac{2 \kappa \bar{\kappa}}{\sigma_r^2} \left( \frac{1}{2} (\nu + \kappa) \tau + \ln \frac{2 \nu}{(\nu + \kappa) (e^{\nu \tau} - 1) + 2 \nu} \right).
\]

It follows from Theorem 7.7 that the optimal investment strategy for an investor with CRRA utility from terminal wealth only is

\[
\Pi_B(W, r, t) = \frac{1}{\gamma \sigma_B(r, t)} \left( \frac{\lambda_1}{\sigma_r} \sqrt{\tau} - \frac{\rho \lambda_2^2}{\sigma_r \sqrt{1 - \rho^2}} \right) + \frac{\gamma - 1}{\gamma} \frac{\sigma_r \sqrt{\tau}}{\sigma_B(r, t)} A_1(T - t),
\]

\[
\Pi_S(W, r, t) = \frac{\lambda_2}{\gamma \sigma_S \sqrt{1 - \rho^2}}.
\]

If the bond instrument used is the zero-coupon bond maturing at the end of the investor’s horizon, we have \( \sigma_B(r, t) = \sigma_r \sqrt{tb(T - t)} \), and the hedge component will simplify to \( \frac{\gamma - 1}{\gamma} A_1(t - t)/b(T - t) \).
As opposed to the Vasicek case we do not have \( A_1(T-t) = b(T-t) \). This implies that the optimal hedge consists of investing the time-varying fraction \( \frac{\gamma - 1}{\gamma} A_1(T-t)/b(T-t) \) in the zero-coupon bond maturing at the end of the investor’s horizon. A similar result was obtained by Deelstra, Grasselli, and Koehl (2000) and Grasselli (2000) using the martingale approach for the case of utility from terminal wealth only.

For an investor with CRRA utility of intermediate consumption only, Theorem 7.8 applies. The fraction of wealth optimally invested in the stock is the same as above, while the fraction of wealth optimally invested in the bond instrument changes to

\[
\Pi_B(W,r,t) = \frac{1}{\gamma \sigma_B(r,t)} \left( \frac{\lambda_1}{\sigma_r} \sqrt{\tau} - \frac{\rho \lambda_2}{\sqrt{1-\rho^2}} \right) + \frac{\gamma - 1}{\gamma} \frac{\sigma_r \sqrt{\tau}}{\sigma_B(r,t)} \int_t^T A_1(s-t) e^{-\frac{1}{2}(s-t)\sigma^2_A(s-t) - \frac{\gamma}{2} A_1(s-t)\rho \lambda} ds.
\]

10.4 A numerical example

We will take historical estimates of mean returns, standard deviations, and correlations as representative of future investment opportunities. These estimates are taken from Dimson, Marsh, and Staunton (2002). All returns are measured per year. The historical average real return on the U.S. stock market is \( \mu_S = 8.7\% \) with a standard deviation of \( \sigma_S = 20.2\% \), while the average real return on bonds is \( \mu_B = 2.1\% \) with a standard deviation of \( \sigma_B = 10.0\% \). The average real U.S. short-term interest rate is \( \bar{r} = 1.0\% \). The correlation between stock returns and bond returns is \( \rho = 0.2 \). Different bonds will have different average returns and different standard deviation of the return. Similarly, the correlation between the return on a bond and the return on the stock market index may not be identical for all bonds. It is not clear exactly what bond or bond index, the above estimates are based on, but we will assume that the estimates for \( \mu_B \) and \( \sigma_B \) apply to a 10-year zero-coupon bond.

The volatility matrix of the bond and the stock is

\[
\Sigma = \begin{pmatrix} 0.1 & 0 \\ 0.0404 & 0.1979 \end{pmatrix}.
\]

The (average) Sharpe ratio of the bond is \( \lambda_1 = (2.1 - 1.0)/10.0 = 0.11 \) and the (average) Sharpe ratio of the stock market is \( \psi = (8.7 - 1.0)/20.2 \approx 0.3812 \). Using (10.2) this corresponds to a market price of risk of \( \lambda_2 \approx 0.3666 \) on the exogenous shock that only affects the stock market. The variance-covariance matrix of returns is \( \Sigma = \sigma \sigma^T \). From (6.8), the tangency portfolio of the bond and the stock is given by

\[
\pi^{\text{tan}} = \begin{pmatrix} \pi_B^{\text{tan}} \\ \pi_S^{\text{tan}} \end{pmatrix} = \begin{pmatrix} 0.1596 \\ 0.8404 \end{pmatrix},
\]

so that the bond/stock ratio is approximately 0.19. Recall that this will be true for all agents who have time-additive utility and who believe that investment opportunities are constant over time. The tangency portfolio has a mean return of 7.65% and a standard deviation of 17.37%.

CRRRA investors ignoring the fluctuations of interest rates will choose a portfolio of risky assets given by \( \pi = \frac{1}{\gamma} (\pi^{\text{tan}})^{-1} \lambda \pi^{\text{tan}} \), where \( \gamma \) is the relative risk aversion of the agent. The portfolio is independent of the investment horizon. In Table 10.1 we show the portfolio allocation for
Chapter 10. Asset allocation with stochastic interest rates

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>tangency</th>
<th>bond</th>
<th>stock</th>
<th>cash</th>
<th>exp. return</th>
<th>volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>4.4079</td>
<td>0.7034</td>
<td>3.7045</td>
<td>-3.4079</td>
<td>0.3030</td>
<td>0.7655</td>
</tr>
<tr>
<td>1</td>
<td>2.2039</td>
<td>0.3517</td>
<td>1.8522</td>
<td>-1.2039</td>
<td>0.1565</td>
<td>0.3827</td>
</tr>
<tr>
<td>2</td>
<td>1.1020</td>
<td>0.1758</td>
<td>0.9261</td>
<td>-0.1020</td>
<td>0.0832</td>
<td>0.1914</td>
</tr>
<tr>
<td>2.2039</td>
<td>1.0000</td>
<td>0.1596</td>
<td>0.8404</td>
<td>0.0000</td>
<td>0.0765</td>
<td>0.1737</td>
</tr>
<tr>
<td>3</td>
<td>0.7346</td>
<td>0.1172</td>
<td>0.6174</td>
<td>0.2654</td>
<td>0.0588</td>
<td>0.1276</td>
</tr>
<tr>
<td>4</td>
<td>0.5510</td>
<td>0.0879</td>
<td>0.4631</td>
<td>0.4490</td>
<td>0.0466</td>
<td>0.0957</td>
</tr>
<tr>
<td>5</td>
<td>0.4408</td>
<td>0.0703</td>
<td>0.3704</td>
<td>0.5592</td>
<td>0.0393</td>
<td>0.0765</td>
</tr>
<tr>
<td>6</td>
<td>0.3673</td>
<td>0.0586</td>
<td>0.3087</td>
<td>0.6327</td>
<td>0.0344</td>
<td>0.0638</td>
</tr>
<tr>
<td>8</td>
<td>0.2755</td>
<td>0.0440</td>
<td>0.2315</td>
<td>0.7245</td>
<td>0.0283</td>
<td>0.0478</td>
</tr>
<tr>
<td>10</td>
<td>0.2204</td>
<td>0.0352</td>
<td>0.1852</td>
<td>0.7796</td>
<td>0.0246</td>
<td>0.0383</td>
</tr>
<tr>
<td>20</td>
<td>0.1102</td>
<td>0.0176</td>
<td>0.0926</td>
<td>0.8898</td>
<td>0.0173</td>
<td>0.0191</td>
</tr>
<tr>
<td>50</td>
<td>0.0441</td>
<td>0.0070</td>
<td>0.0370</td>
<td>0.9559</td>
<td>0.0129</td>
<td>0.0077</td>
</tr>
<tr>
<td>200</td>
<td>0.0110</td>
<td>0.0018</td>
<td>0.0093</td>
<td>0.9890</td>
<td>0.0107</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

Table 10.1: Portfolio weights for CRRA investors ignoring interest rate fluctuations.

Various $\gamma$-values. The numbers in the column “tangency” denotes the fraction of wealth invested in the tangency portfolio. This investment is divided into the bond and the stock in the following two columns. The cash position is determined residually so that weights sum to one. The last two columns show the instantaneous expected rate of return and volatility of the portfolio. In Figure 10.1 the curved line shows the mean-variance efficient portfolios of risky assets, i.e., the combinations of expected returns and volatility that can be obtained by combining the bond and the stock. The straight line corresponds to the optimal portfolios for investors assuming constant investment opportunities with an interest rate equal to the long-term average.

Now let us look at investors who realize that interest rates vary over time and consequently alter their investment strategy (except for log-utility investors). First, we assume that the real short-term interest rate $r_t$ follows the one-factor Vasicek model so that the analysis and results of Section 10.2 applies. The long-term average interest rate is $\bar{r} = 1.0\%$ and we take a short-rate volatility of $\sigma_r = 5\%$, which is also consistent with the U.S. historical estimate. We use the same values of the market prices of risk as above. We set the value of the mean reversion rate $\kappa = 0.4965$ so that the volatility of a 10-year zero-coupon bond according to the model is equal to the historical estimate of 10.0%. The current short rate is assumed to equal the long-term level, $r_t = \bar{r}$.

Let us first consider investors with utility of terminal wealth only. Their optimal portfolios are given by (10.4) and (10.5). Table 10.2 shows the optimal portfolios for CRRA investors with different combinations of risk aversion and investment horizon. The numbers under the column heading ‘hedge’ are $\frac{2-\gamma}{\gamma}b(T)/b(10)$, which is the hedge demand for the 10-year zero-coupon bond which the investors are allowed to trade in. While the weight on the tangency portfolio and thus the stock is independent on the investment horizon, this is not true for the weight on the hedge portfolio and hence not true for the total weight on the bond and on cash. The ratio of the bond weight to the stock weight is shown in the column ‘bond/stock’. The bond-stock ratio increases considerably with the risk aversion and, for investors with $\gamma > 1$, with the investment horizon.
The investor with a horizon of $T$ will want to hedge interest rate risk by investing in the $T$-period zero-coupon bond. That bond is replicated by a portfolio of $b(T)/b(10)$ units of the 10-year zero-coupon bond and a cash position. Since $b$ is increasing in $T$, the hedge demand for the 10-year bond increases with the horizon $T$. It is important to emphasize that the portfolio weights on the bond and thus the bond/stock ratio will depend on the maturity (and payment schedule) of the bond, the investor is trading in. In particular, a recommendation of a particular bond weight or bond/stock ratio should always be accompanied by a specification of what bond the recommendation applies to.

Next, we consider investors with utility from intermediate consumption and no utility from terminal wealth. In this case the hedge term in the bond weight (10.5) is replaced by (10.6). Now the hedge demand depends on the current interest rate level, which we assume is equal to the long-term average of 1%. Table 10.3 shows the optimal portfolios for investors with a 1-year and a 30-year horizon. We see the same overall picture as for investors with utility from terminal wealth only, but for a given investment horizon the hedge demand for bond and hence the bond/stock ratio are smaller with utility of consumption since the optimal bond for hedging has a smaller duration than the investment horizon.

Let us now compare the current mean/variance tradeoff chosen by different investors. As discussed above, CRRA investors that either have a zero (or very, very short) investment horizon or do not take interest rate risk into account will pick a portfolio that corresponds to a point on the straight line in Figure 10.2. This is the instantaneous mean-variance efficient frontier. Similarly, each of the other curves corresponds to the combinations chosen by CRRA investors with a given
Table 10.2: Portfolio weights for CRRA investors who assume Vasicek interest rate dynamics and have utility from terminal wealth only.
Table 10.3: Portfolio weights for CRRA investors who assume Vasicek interest rate dynamics and have utility from intermediate consumption only.

<table>
<thead>
<tr>
<th>horizon</th>
<th>$\gamma$</th>
<th>tangency</th>
<th>hedge</th>
<th>bond</th>
<th>stock</th>
<th>bond/stock</th>
<th>cash</th>
<th>exp. return</th>
<th>volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 1$</td>
<td>0.5</td>
<td>4.4079</td>
<td>-0.2253</td>
<td>0.4780</td>
<td>3.7045</td>
<td>0.1290</td>
<td>-3.1825</td>
<td>0.3005</td>
<td>0.7593</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2.2039</td>
<td>0.0000</td>
<td>0.3517</td>
<td>1.8522</td>
<td>0.1899</td>
<td>-1.2039</td>
<td>0.1565</td>
<td>0.3827</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.1020</td>
<td>0.1114</td>
<td>0.2872</td>
<td>0.9261</td>
<td>0.3101</td>
<td>-0.2134</td>
<td>0.0845</td>
<td>0.1949</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.4408</td>
<td>0.1787</td>
<td>0.2490</td>
<td>0.3704</td>
<td>0.6722</td>
<td>0.3805</td>
<td>0.0413</td>
<td>0.0835</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.2204</td>
<td>0.2013</td>
<td>0.2365</td>
<td>0.1852</td>
<td>1.2766</td>
<td>0.5783</td>
<td>0.0269</td>
<td>0.0481</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.1102</td>
<td>0.2126</td>
<td>0.2302</td>
<td>0.0926</td>
<td>2.4856</td>
<td>0.6772</td>
<td>0.0197</td>
<td>0.0324</td>
</tr>
<tr>
<td>$T = 30$</td>
<td>0.5</td>
<td>4.4079</td>
<td>-0.9639</td>
<td>-0.2605</td>
<td>3.7045</td>
<td>-0.0699</td>
<td>-2.4440</td>
<td>0.2924</td>
<td>0.7435</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2.2039</td>
<td>0.0000</td>
<td>0.3517</td>
<td>1.8522</td>
<td>0.1899</td>
<td>-1.2039</td>
<td>0.1565</td>
<td>0.3827</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.1020</td>
<td>0.4428</td>
<td>0.6187</td>
<td>0.9261</td>
<td>0.6677</td>
<td>-0.5448</td>
<td>0.0881</td>
<td>0.2085</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.4408</td>
<td>0.7254</td>
<td>0.7957</td>
<td>0.3704</td>
<td>2.1445</td>
<td>-0.1662</td>
<td>0.0473</td>
<td>0.1196</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.2204</td>
<td>0.8245</td>
<td>0.8597</td>
<td>0.1852</td>
<td>4.6319</td>
<td>-0.0449</td>
<td>0.0337</td>
<td>0.1004</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.1102</td>
<td>0.8751</td>
<td>0.8927</td>
<td>0.0926</td>
<td>9.6176</td>
<td>0.0147</td>
<td>0.0270</td>
<td>0.0948</td>
</tr>
</tbody>
</table>

Table 10.4: Portfolio weights for investors with a constant relative risk aversion of $\gamma = 2$.

<table>
<thead>
<tr>
<th>horizon</th>
<th>tangency</th>
<th>hedge</th>
<th>bond</th>
<th>stock</th>
<th>bond/stock</th>
<th>cash</th>
<th>exp. return</th>
<th>volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 0$</td>
<td>1.1020</td>
<td>0.1758</td>
<td>0.9261</td>
<td>0.19</td>
<td>-0.1020</td>
<td>0.0832</td>
<td>0.1914</td>
<td></td>
</tr>
<tr>
<td>$T = 1$, wealth</td>
<td>1.1020</td>
<td>0.1970</td>
<td>0.3729</td>
<td>0.9261</td>
<td>0.40</td>
<td>-0.2990</td>
<td>0.0854</td>
<td>0.1979</td>
</tr>
<tr>
<td>$T = 5$, wealth</td>
<td>1.1020</td>
<td>0.4615</td>
<td>0.6373</td>
<td>0.9261</td>
<td>0.69</td>
<td>-0.5634</td>
<td>0.0883</td>
<td>0.2094</td>
</tr>
<tr>
<td>$T = 10$, wealth</td>
<td>1.1020</td>
<td>0.5000</td>
<td>0.6758</td>
<td>0.9261</td>
<td>0.73</td>
<td>-0.6020</td>
<td>0.0887</td>
<td>0.2112</td>
</tr>
<tr>
<td>$T = 30$, wealth</td>
<td>1.1020</td>
<td>0.5035</td>
<td>0.6794</td>
<td>0.9261</td>
<td>0.73</td>
<td>-0.6055</td>
<td>0.0888</td>
<td>0.2114</td>
</tr>
<tr>
<td>$T = 1$, cons.</td>
<td>1.1020</td>
<td>0.1114</td>
<td>0.2872</td>
<td>0.9261</td>
<td>0.3101</td>
<td>-0.2134</td>
<td>0.0845</td>
<td>0.1949</td>
</tr>
<tr>
<td>$T = 30$, cons.</td>
<td>1.1020</td>
<td>0.4428</td>
<td>0.6187</td>
<td>0.9261</td>
<td>0.6677</td>
<td>-0.5448</td>
<td>0.0881</td>
<td>0.2085</td>
</tr>
</tbody>
</table>

non-zero horizon who take interest rate risk into account. Since these curves lie to the right of the instantaneous mean-variance frontier, all these investors could obtain a higher instantaneous expected rate of return for the same volatility by choosing a different portfolio. But the long-term investors are willing to sacrifice some expected return in the short term in order to hedge changes in interest rates and place themselves in a better position if interest rates should decline.

Table 10.4 shows the optimal portfolios for investors with a constant relative risk aversion equal to 2, but with different investment horizons. Here we can clearly see the effect of the investment horizon on the optimal bond holdings and the bond/stock ratio. Relative to the extreme short-term investor, long-term investors have the same stock weight but shifts wealth from cash to bonds. If we look at the instantaneous risk/return trade-off, the longer-term investors choose more risky portfolios, i.e., they take on more short-term risk. But the main point is that long-term investors do not choose their portfolio according to the short-term risk/return trade-off.

Next, we want to investigate how sensitive the asset allocation choice is to the assumed interest
Figure 10.2: Optimal frontiers with Vasicek interest rates. Each curve contains the combinations of current expected rate of return and volatility for CRRA investors with a given investment horizon $T$. From the left, the curves represent (a) $T = 0$ (black straight line; identical to the mean-variance frontier), (b) $T = 1$ and utility of consumption (blue curve), (c) $T = 1$ and utility from terminal wealth only (red curve), (d) $T = 30$ and utility from consumption (grey curve), and (e) $T = 30$ and utility from terminal wealth only (green curve).

rate model. We do that by computing the optimal portfolios when interest rates follow the CIR model (10.7). We want to make a reasonably fair comparison between the two models. For that purpose we choose $\sigma_r = 0.5$ in the CIR model so that the average short rate volatility is $\sigma_r \sqrt{\bar{r}} = 0.05$ as in the Vasicek model. We set $\lambda_1 = 0.55$ and $\lambda_2 = 0.3666$ so that the model is consistent with the estimated mean stock and bond returns when $r = \bar{r}$ is used to compute the Sharpe ratios of the bond market ($\lambda_1(r)$) and the stock market ($\psi(r)$ in (10.8)). The mean reversion rate is set at $\kappa = 0.7994$ so that the volatility of a 10-year zero-coupon bond according to the model is equal to the historical estimate of 10.0%. The optimal portfolio in the CIR setting depends on the current interest rate level. In the computations we put this equal to the long-term average of 1%.

In Table 10.5 we list the optimal portfolios for investors with CRRA utility of terminal wealth both for the Vasicek and the CIR setting. We consider an investor with a 1-year horizon and an investor with a 30-year horizon. The stock weight is identical in the two models. The hedge demand for bonds and hence the total bond demand (and the cash position) do depend on the interest rate model, but the differences are relatively small. The yield curves of the two models are almost identical. The long-term yield is 1.601% in the Vasicek model and 1.600% in the CIR model. With a current short rate of 1%, the yield curve is uniformly increasing in both models, cf. the results on the shape of the yield curve in the two models reported by, e.g., Munk (2011).
### 10.5 Two-factor Vasicek model

Brennan and Xia (2000) study a two-factor Vasicek interest rate model with utility from terminal wealth only. Assume that the dynamics of the short-term interest rate $r_t$ is

$$dr_t = (\varphi_r + u_t - \kappa_r r_t) dt - \sigma_r dz_{1t},$$

$$du_t = -\kappa_u u_t dt - \sigma_u \rho_{ru} dz_{1t} - \sigma_u \sqrt{1 - \rho_{ru}^2} dz_{2t},$$

where $z_1 = (z_{1t})$ and $z_2 = (z_{2t})$ are independent one-dimensional standard Brownian motions. The one-factor Vasicek model is the special case where $u_t \equiv 0$, and then the short rate $r_t$ is expected to move towards the long-run level $\varphi_r / \kappa_r$. The new state variable $u$ allows for variations in this long-run target for the short-term interest rate. Note that we can rewrite the drift rate of $u$ as $\kappa_u (0 - u_t)$ which shows that $u$ exhibits mean reversion around the long-run level 0. Future values of $r$ and $u$ are normally distributed so it is a Gaussian model. The market prices of risk associated with the shocks represented by $z_1$ and $z_2$ are denoted by $\lambda_1$ and $\lambda_2$, respectively, and are assumed constant.

Beaglehole and Tenney (1991) and Hull and White (1994) studied such a model and its implications for the pricing of bonds. They have shown that the time $t$ price of the zero-coupon bond maturing at time $\bar{T}$ is given by

$$B^\bar{T}(r, u, t) = e^{-a(\bar{T}-t)-b_1(\bar{T}-t)r-b_2(\bar{T}-t)u},$$

where

$$b_1(\tau) = \frac{1}{\kappa_r} \left(1 - e^{-\kappa_r \tau}\right),$$

$$b_2(\tau) = \frac{1}{\kappa_r \kappa_u} + \frac{1}{\kappa_r (\kappa_r - \kappa_u)} e^{-\kappa_r \tau} - \frac{1}{\kappa_u (\kappa_r - \kappa_u)} e^{-\kappa_u \tau},$$

Table 10.5: Portfolio weights with Vasicek or CIR dynamics for CRRA investors with utility from terminal wealth only.

<table>
<thead>
<tr>
<th>horizon</th>
<th>$\gamma$</th>
<th>tangency</th>
<th>stock</th>
<th>Vasicek model</th>
<th>CIR model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hedge</td>
<td>bond</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>0.5</td>
<td>4.4079</td>
<td>3.7045</td>
<td>-0.3941</td>
<td>0.3093</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2.2039</td>
<td>1.8522</td>
<td>0.0000</td>
<td>0.3517</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.1020</td>
<td>0.9261</td>
<td>0.1970</td>
<td>0.3729</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.4408</td>
<td>0.3704</td>
<td>0.3153</td>
<td>0.2439</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.2204</td>
<td>0.1852</td>
<td>0.3547</td>
<td>0.2439</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.1102</td>
<td>0.0926</td>
<td>0.3744</td>
<td>0.2439</td>
</tr>
<tr>
<td>$T = 30$</td>
<td>0.5</td>
<td>4.4079</td>
<td>3.7045</td>
<td>-1.0070</td>
<td>-0.3036</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2.2039</td>
<td>1.8522</td>
<td>0.0000</td>
<td>0.3517</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.1020</td>
<td>0.9261</td>
<td>0.5035</td>
<td>0.6794</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.4408</td>
<td>0.3704</td>
<td>0.8056</td>
<td>0.8760</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.2204</td>
<td>0.1852</td>
<td>0.9063</td>
<td>0.9415</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.1102</td>
<td>0.0926</td>
<td>0.9567</td>
<td>0.9743</td>
</tr>
</tbody>
</table>
and \( a(\cdot) \) is a quite complicated function which is not important for what follows. The dynamics of the price of the zero-coupon bond maturing at time \( \bar{t} \) is then
\[
d B^\bar{t}_t = B^\bar{t}_t \left[ (r_t + \psi_B(\bar{t} - t)) \, dt + \sigma_B1(\bar{t} - t) \, dz_1 + \sigma_B2(\bar{t} - t) \, dz_2 \right],
\]
where, for all \( \tau \geq 0 \), we have defined
\[
\psi_B(\tau) = \lambda_1 \sigma_B1(\tau) + \lambda_2 \sigma_B2(\tau),
\]
\[
\sigma_B1(\tau) = \sigma_r b_1(\tau) + \sigma_u \rho_{ru} b_2(\tau),
\]
\[
\sigma_B2(\tau) = \sigma_u \sqrt{1 - \rho^2_{ru}} b_2(\tau).
\]

Consider an investor with utility of wealth at time \( T \) exhibiting a constant relative risk aversion \( \gamma > 1 \). The investor earns no labor income, has no preferences for consumption before time \( T \), and is not subject to portfolio constraints. The investor can trade in a single non-dividend paying stock (representing the stock market index) with time \( t \) price \( S_t \), which evolves according to
\[
d S_t = S_t \left[ (r_t + \psi_S) \, dt + \sigma_S k_1 \, dz_1 + \sigma_S k_2 \, dz_2 + \sigma_S k_3 \, dz_3 \right],
\]
where \( z_3 = (z_3t) \) is a one-dimensional standard Brownian motion independent of \( z_1 \) and \( z_2 \), and \( k_3 = \sqrt{1 - k_1^2 - k_2^2} \) so that \( \sigma_S \) is the volatility of the stock. The constant \( \lambda_3 \) is the market price of risk associated with \( z_3 \) so the Sharpe ratio of the stock is
\[
\psi_S = k_1 \lambda_1 + k_2 \lambda_2 + \sqrt{1 - k_1^2 - k_2^2} \lambda_3.
\]

In total we assume that the investor can invest in the following four assets:

1. the locally risk-free asset (aka. the bank account or cash deposits) providing a net rate of return of \( r_t \),
2. a zero-coupon bond maturing at time \( T_1 \),
3. a zero-coupon bond maturing at time \( T_2 \neq T_1 \),
4. the stock.

Of course, both \( T_1 \) and \( T_2 \) must be greater than current time \( t \), but they can be smaller or larger than the investment horizon \( T \). If one or both bonds mature before \( T \), the investor will then have to replace the matured bond with a new bond maturing further into the future. As the term structure of interest rates is driven by two Brownian motions, it would not help the investor to trade in additional default-free bonds. The dynamics of the three risky assets can be written compactly as
\[
d \begin{pmatrix} B^{T_1}(r,u,t) \\ B^{T_2}(r,u,t) \\ S_t \end{pmatrix} = \begin{pmatrix} B^{T_1}(r,u,t) & 0 & 0 \\ 0 & B^{T_2}(r,u,t) & 0 \\ 0 & 0 & S_t \end{pmatrix} \left[ (r_t \mathbf{1} + \sigma(t) \lambda) \, dt + \sigma(t) \, dz_1 \right],
\]
where
\[
\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix}, \quad \sigma(t) = \begin{pmatrix} \sigma_{B1}(T_1 - t) & \sigma_{B2}(T_1 - t) & 0 \\ \sigma_{B1}(T_2 - t) & \sigma_{B2}(T_2 - t) & 0 \\ \sigma_S k_1 & \sigma_S k_2 & \sigma_S k_3 \end{pmatrix}.
\]
This is a two-dimensional affine asset allocation model as defined in Section 7.3.4. By solving the relevant ODEs, the indirect utility function turns out to be (the following results are to be verified in Exercise 10.3)

\[ J(W, r, u, t) = \frac{1}{1 - \gamma} g(r, u, t)^\gamma W^{1-\gamma}, \]

\[ g(r, u, t) = \exp \left\{ -\frac{\gamma - 1}{\gamma} A_0(T - t) - \frac{\gamma - 1}{\gamma} A_1r(T - t) - \frac{\gamma - 1}{\gamma} A_1u(T - t) \right\}, \]

\[ A_1r(\tau) \equiv b_1(\tau), \quad A_1u(\tau) \equiv b_2(\tau), \]

where \( b_1 \) and \( b_2 \) are defined above. \( A_0(\tau) \) is another deterministic function, which is not important for the optimal portfolio if we disregard intermediate consumption. The optimal investment strategy is\(^1\)

\[
\begin{pmatrix}
\pi_{B1}(t) \\
\pi_{B2}(t) \\
\pi_S
\end{pmatrix} = \frac{1}{\gamma} \left( \Omega(t)^\top \right)^{-1} \lambda - \frac{\gamma - 1}{\gamma} \left( \Omega(t)^\top \right)^{-1} \begin{pmatrix}
-\sigma_r & -\sigma_u \rho_{ru} \\
0 & -\sigma_u \sqrt{1 - \rho_{ru}^2} \\
0 & 0
\end{pmatrix} \begin{pmatrix}
A_1r(T - t) \\
A_1u(T - t)
\end{pmatrix}
\]

The optimal investment in the stock reduces to

\[ \pi_S = \frac{1}{\gamma} \frac{\lambda_3}{\sigma_S k_3}. \]

The optimal investments in the two bonds can be rewritten as

\[ \pi_{B1}(t) = \frac{1}{\gamma \sigma_r \sigma_u \sqrt{1 - \rho_{ru}^2} d(t)} \left( \sigma_r b_1(T_2 - t) \left[ k_2 \frac{\lambda_3}{k_3} - \lambda_2 \right] + \sigma_u b_2(T_2 - t) \left[ \left( \lambda_1 - \frac{k_1 \lambda_3}{k_3} \right) \sqrt{1 - \rho_{ru}^2} + \left( \frac{k_2 \lambda_3}{k_3} - \lambda_2 \right) \rho_{ru} \right] \right) + \frac{\gamma - 1}{\gamma d(t)} \left( b_2(T_2 - t) b_1(T - t) - b_1(T_2 - t) b_2(T - t) \right), \]

\[ \pi_{B2}(t) = \frac{1}{\gamma \sigma_r \sigma_u \sqrt{1 - \rho_{ru}^2} d(t)} \left( \sigma_r b_1(T_1 - t) \left[ \lambda_2 - \frac{k_2 \lambda_3}{k_3} \right] + \sigma_u b_2(T_1 - t) \left[ \left( \frac{k_1 \lambda_3}{k_3} - \lambda_1 \right) \sqrt{1 - \rho_{ru}^2} + \left( \lambda_2 - \frac{k_2 \lambda_3}{k_3} \right) \rho_{ru} \right] \right) + \frac{\gamma - 1}{\gamma d(t)} \left( b_1(T_1 - t) b_2(T - t) - b_1(T - t) b_2(T_1 - t) \right), \]

where

\[ d(t) = b_1(T_1 - t) b_2(T_2 - t) - b_1(T_2 - t) b_2(T_1 - t). \]

Note that the portfolio weights are purely deterministic and thus independent of the state variables \( r \) and \( u \). Also note that if \( T = T_1 \), the hedge term for the \( T_1 \)-bond vanishes, whereas the hedge term for the \( T_2 \)-bond reduces to \( \frac{\gamma - 1}{\gamma} \), whereas the hedge term for the \( T_2 \)-bond vanishes. Conversely, if \( T = T_2 \).

\(^1\)When computing \( (\Omega(t)^\top)^{-1} \), it is useful to know that

\[
\begin{pmatrix}
M_{11} & M_{12} & \cdots & M_{13} \\
M_{21} & M_{22} & \cdots & M_{23} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & M_{33}
\end{pmatrix}^{-1} = \begin{pmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22} \\
\vdots & \vdots \\
0 & 0
\end{pmatrix}^{-1} - \frac{1}{M_{33}} \begin{pmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22} \\
\vdots & \vdots \\
0 & 0
\end{pmatrix}^{-1} \begin{pmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22} \\
\vdots & \vdots \\
0 & 0
\end{pmatrix}^{-1} \begin{pmatrix}
M_{11} & M_{12} \\
M_{21} & M_{22} \\
\vdots & \vdots \\
0 & 0
\end{pmatrix}^{-1}
\]
Now assume the following parameters:

\[
\begin{align*}
\varphi_r &= 0.02, & \kappa_r &= 0.5, & \sigma_r &= 0.022, & \lambda_1 &= 0.11, \\
\kappa_u &= 0.1, & \sigma_u &= 0.005, & \rho_{ru} &= 0, & \lambda_2 &= 0.07, \\
\sigma_S &= 0.141, & \psi_S &= 0.284, & k_1 &= 0.15, & k_2 &= 0.2.
\end{align*}
\]

Suppose in the following that \textit{at any point in time} the two zero-coupon bonds which the individual invests in mature in 5 years and 20 years, respectively. Hence, the sensitivity matrix \( \frac{\partial S(t)}{\partial t} \) is constant and so is the speculative part of the optimal investment strategy. With the listed parameters, the 5-year bond has a volatility of 4.82\% and an expected excess rate of return of 0.63\%, whereas the 20-year bond has a volatility of 9.40\% and an excess expected rate of return of 1.07\%. The instantaneous correlations between the stock and the 5-year bond and 20-year bond are 0.235 and 0.247, respectively. The instantaneous correlation between the two bonds is 0.874. The tangency portfolio (normalized so that the weights sum to 100\%) consists of 62.5\% in the 5-year bond, \(-14.5\%\) in the 20-year bond, and 52.0\% in the stock.

The Figures 10.3 and 10.4 depict the optimal investments in the risky assets as a function of the investment horizon for a relative risk aversion of \( \gamma = 2 \) and \( \gamma = 4 \), respectively. The lines corresponding to the speculative demands are flat as explained above, and the figures confirm that for each asset the speculative demand for \( \gamma = 4 \) is exactly half of the speculative demand for \( \gamma = 2 \). There is no hedge demand for the stock so the speculative stock demand equals the total stock demand. The hedge demand for the two bonds are highly dependent on the investment horizon of the investor. With a 5-year horizon, the 5-year bond is the perfect hedge instrument so the 20-year bond drops out of the hedge portfolio. Conversely for a 20-year horizon. For a horizon between 5 and 20 years, the hedge portfolio consists of long positions in both bonds in order to replicate a zero-coupon bond with a maturity identical to the horizon. The same argument explains the composition of the hedge portfolio for other horizons. For a horizon shorter than 5 years, a large long position in the 5-year bond and a small short position in the 20-year bond to emulate the desired bond maturity. For a horizon longer than 20 years, this requires a large long position in the 20-year bond and a small short position in the 5-year bond. A consequence of these results is that the optimal portfolio weight in a bond of a given maturity is non-monotonic in the horizon of the investor.

### 10.6 Other studies with stochastic interest rates

Brennan, Schwartz, and Lagnado (1997) apply the two-factor Brennan-Schwartz interest rate dynamics in a model that also has stochastic dividends on stocks. They study the effect the length of the investment horizon has for an investor with utility from terminal wealth only. Due to the complexity of their model they must resort to numerical solution techniques.

Wachter (2003) shows that as risk aversion approaches infinity, an investor with utility only of wealth at time \( T \) will invest solely in the real zero-coupon bond maturing at time \( T \). This holds for any utility function and for all well-behaved Itô-processes for the returns of the available assets. With utility of intermediate consumption, the infinitely risk-averse individual should invest in a certain coupon bond with coupons related to expected future consumption.

Munk and Sørensen (2004) study the asset allocation problem when the term structure of interest
10.6 Other studies with stochastic interest rates

Figure 10.3: Optimal portfolios in the two-factor Vasicek model: risk aversion of 2.

Figure 10.4: Optimal portfolios in the two-factor Vasicek model: risk aversion of 4.
rates evolve according to models in the Heath-Jarrow-Morton (HJM) class. As shown by Heath, Jarrow, and Morton (1992), any dynamic interest rate model is fully specified by the current term structure and the forward rate volatilities. Therefore the HJM modeling framework is natural when comparing the separate effects of the current term structure and the dynamics of the term structure on the optimal interest rate hedging strategy. Term structure models in the HJM class are not necessarily Markovian, but the class includes the well-known Markovian models such as the Vasicek model. To cover the non-Markovian models the authors apply the martingale approach to solve the utility maximization problem instead of the dynamic programming approach. Within the HJM framework one may fix the current yield curve and vary its future dynamics to gauge the effect of the interest rate dynamics. As in all term structure models one can fix the dynamics and vary the initial yield curve (for absolute pricing models, such as the Vasicek and CIR models, not all initial yield curves are possible). The paper compares the optimal portfolio and consumption strategies for a standard one-factor Vasicek and a three-factor model where the term structure can exhibit three kinds of changes: A parallel shift, a slope change, and a curvature change. The authors find that the form of the initial term structure is of crucial importance for the certainty equivalents of future consumption and, hence, important for the relevant interest rate hedge, while the specific dynamics of the term structure is of minor importance. Of course, further studies of this kind is needed to find out whether this conclusion is generally valid.

Detemple and Rindisbacher (2010) derive a new and very general portfolio decomposition result. Assuming utility of time $T$ wealth only, the optimal portfolio is decomposed into three terms: (i) the speculative term, (ii) a term hedging variations in the price of the zero-coupon bond maturing at time $T$, and (iii) a term hedging against fluctuations in the density of the so-called $T$-forward probability measure, i.e., the equivalent martingale measure corresponding to the use of the zero-coupon bond maturing at $T$ as the numeraire, cf. Björk (2009) or Munk (2011).

It has long been recognized that the volatility of interest rates varies over time in a non-deterministic way. This is a key motivation behind models in which one or several of the state variables follow square-root processes. In the basic interest rate models with stochastic volatility, the zero-coupon bond prices will depend non-trivially on all state variables and thus in particular on the volatility-determining state variables. Because the first-order partial derivatives of the zero-coupon bond price with respect to these volatility factors are generally non-zero, it is possible to set up a trading strategy in bonds of different maturities which is completely hedged against volatility shocks, i.e., the stochastic volatility is spanned by the traded bonds. However, some recent empirical studies document unspanned stochastic volatility in the sense that a part of the stochastic volatility in the yield curve cannot be hedged away using only bonds. Simple fixed-income derivatives like caps and swaptions, which obviously depend on the volatility of interest rates, cannot be perfectly replicated by trading even a larger number of bonds. Bond markets are incomplete.\footnote{For example, based on 1995-2000 data from the U.S., the U.K., and Japan, Collin-Dufresne and Goldstein (2002) find that only a (small) part of the returns on at-the-money straddles can be explained by changes in the underlying swap rates in a regression analysis. An at-the-money straddle is a portfolio consisting of an at-the-money cap and an at-the-money floor. By construction, such a straddle is neutral to small changes in the interest rate level, but very sensitive to changes in volatility. The results thus show that variations in interest rate volatility are only partly due to variations in the level of interest rates. Note that this is model-independent evidence of unspanned stochastic volatility: no model is assumed for the pricing of the caps and floors involved. For further empirical support of} Trolle (2009) studies the optimal demand for bonds and interest rate derivatives in a
model featuring unspanned stochastic volatility (such a model is necessarily quite complex). Since interest rate derivatives (options, caps, floors, swaptions, etc.) will depend on the volatility factors not spanned by the bonds, investing in derivatives allow you to pick up the market price of risk associated with those factors and also to hedge against adverse shifts in the factors. Trolle’s empirical investigation shows that the market prices of risk of the unspanned volatility factors and thus the Sharpe ratios of the interest rate derivatives are high (compared to bonds). As a consequence, he finds substantial welfare gains from including interest rate derivatives in the portfolio.

A number of papers explore models with both stochastic interest rates and stochastic inflation rates. We will study such a setting in Chapter 12. As an example, Campbell and Viceira (2001) study a discrete-time consumption and portfolio choice problem of an infinitely-lived investor with recursive utility of the Epstein-Zin type. They assume that the real short-term interest rate and the expected inflation rate follow correlated AR(1) processes, i.e., discrete-time versions of the Ornstein-Uhlenbeck process, similar to the dynamics of \( r \) and \( u \) in the two-factor Vasicek model above. They derive an approximate analytic solution to the problem and compare the optimal bond demand for a long-term inflation-indexed bond to the optimal bond demand for a long-term nominal bond.

Another study with both stochastic interest rates and inflation is Sangvinatsos and Wachter (2005). They assume that the nominal short-term interest rate and the expected inflation rate are affine functions of a three-dimensional state variable, which follows an Ornstein-Uhlenbeck process. The market prices of risk are allowed to be affine in the state variable as well so that excess expected bond returns vary with the state variables in contrast to the one- and two-factor Vasicek models considered earlier in the chapter. The model is still affine, so they can derive a closed-form solution for the optimal portfolio of an investor with CRRA utility of terminal wealth. Among other things they show that, when the investor has access to several long-term bonds of different maturities, the optimal portfolio typically involves relatively extreme long and short positions.

### 10.7 Exercises

**Exercise 10.1.** Consider a financial market where the only two assets traded are (1) a bank account with a rate of return of \( r_t \) and (2) a risky asset with price \( P_t \) following the geometric Brownian motion,

\[
dP_t = P_t [\mu dt + \sigma dz_t].
\]

The short-term interest rate is assumed to follow a Vasicek process:

\[
dr_t = \kappa [\bar{r} - r_t] dt + \rho \sigma r dz_t + \sqrt{1 - \rho^2} \sigma \hat{r} d\hat{z}_t.
\]

(a) Describe the model!

We look at an investor with CRRA utility of terminal wealth only,

\[
J(W,r,t) = \sup_{\pi} E_{W,r,t} \left[ \frac{W_t^{1-\gamma}}{1-\gamma} \right],
\]

where the process \( \pi \) denotes the fraction of wealth invested in the risky asset.

unspanned stochastic volatility, see Heidari and Wu (2003), Li and Zhao (2006), Jarrow, Li, and Zhao (2007), and Trolle and Schwartz (2009).
(b) State the HJB equation corresponding to this problem.
(c) Find the first-order condition for \( \pi \).
(d) Show that the indirect utility function is of the form
\[
J(W,r,t) = \frac{1}{1-\gamma} \left( We^{A_0(T-t)+A_1(T-t)r+\frac{1}{2}A_2(T-t)r^2} \right)^{1-\gamma}.
\]
What can you say about the functions \( A_i \)?
(e) Find the optimal portfolio strategy. Compare it with the solution for constant \( r \).

Exercise 10.2. Consider an economy with a single agent. The agent owns a production plant that generates units of the consumption good of the economy. The agent can choose to withdraw consumption goods from the production or reinvest them in the production process. The productivity of her plant depends on a state variable \( Y_t \) that follows the process
\[
dY_t = (b - \kappa Y_t) dt + \kappa \sqrt{Y_t} dZ_t, \quad Y_0 = y,
\]
where \( b, \kappa \) and \( k \) are positive constants with \( 2b > k^2 \). Let \( c_t \geq 0 \) denote the rate by which the agent withdraws consumption goods from the production plant and let \( X_t^c \) be the value of the plant at time \( t \) given the consumption process \( c_t \). We assume that
\[
dX_t^c = (X_t^c h Y_t - c_t) dt + X_t^c \varepsilon \sqrt{Y_t} dz_t, \quad X_0^c = x,
\]
where \( h \) and \( \varepsilon \) are positive constants with \( h > \varepsilon^2 \). The agent has a log utility of consumption over her life-time \( T \), so that the indirect utility function is
\[
V(x,y,t) = \sup_{c} \mathbb{E}_{x,y,t} \left[ \int_t^T e^{-\delta(s-t)} \ln c_s ds \right].
\]
(a) State the HJB equation corresponding to the problem and find the first-order condition for the optimal consumption rate.
(b) Verify that the function
\[
V(x,y,t) = A_0(t) \ln x + A_1(t)y + A_2(t)
\]
satisfies the HJB equation and find ordinary differential equations that the functions \( A_0, A_1 \) and \( A_2 \) must solve. Show that \( A_0(t) = \frac{1}{\delta}(1 - e^{-\delta(T-t)}) \). Find an explicit expression for the optimal consumption rate, \( c^*_t \).
(c) We know from the martingale approach that the state-price deflator \( \zeta_t \) satisfies \( \psi \zeta_t = u'(c^*_t, t) \), where \( \psi \) is a constant, and where \( u(c, t) = e^{-\delta t} \ln c \) in our case. Use this and the expression for optimal consumption to show that
\[
\zeta_t = \frac{1}{\psi} e^{-\delta t} \frac{A_0(t)}{X_t^*},
\]
where \( X_t^* \) is the optimal value of the production plant, i.e., \( X_t^* = X_t^{c^*} \). Apply Itô’s Lemma in order to find the dynamics of \( \zeta_t \).
(d) We also know that
\[
d\zeta_t = -\zeta_t [r_t dt + \lambda_t dz_t],
\]
where $r_t$ is the short-term interest rate. Conclude that $r_t = (h - \varepsilon^2)Y_t$. Show that the dynamics of $r_t$ is on the form

$$dr_t = \kappa [\bar{r} - r_t] dt + \sigma_{r_t} \sqrt{r_t} dz_t,$$

where $\kappa$, $\bar{r}$ and $\sigma_{r_t}$ are positive constants. Appreciate this result!

**Exercise 10.3.** Verify the expressions stated in Section 10.5 for the indirect utility function and the optimal investment strategy for the two-factor Vasicek model. If preferences for intermediate consumption are included, how will the optimal consumption and investment strategy look like?
11.1 Introduction

In this chapter we consider models where interest rates are constant, but the market price of stock market risk varies stochastically over time.

11.2 Mean reversion in stock returns

Several empirical studies provide evidence of mean reversion in stock returns so that expected stock returns are high after a period of low realized returns and vice versa. See, e.g., Poterba and Summers (1988), Fama and French (1989), Campbell, Lo, and MacKinlay (1997, Ch. 7), and Cochrane (2005, Ch. 20). Formulated differently, stock returns appear to be predictable by factors related to the current stock price, such as the earnings/price ratio or the dividend/price ratio.\footnote{There is also evidence that stock returns can be predicted by the current level of interest rates, cf., e.g., Ang and Bekaert (2007).}

We have seen earlier that CRRA investors should have a constant fraction of wealth invested in the stock market index if the stock market risk premium is constant over time. Mean reversion in stock returns leads to lower variance of long-term stock returns, which intuitively should lead to larger investments in the stock. Moreover, we expect that CRRA investors should invest more [less] in the stock in periods where the expected future stock return is high [low].

Some recent papers have set up formal models studying the implications for portfolio decisions of mean reversion in stock returns. Both Kim and Omberg (1996) and Wachter (2002) obtain closed-form expressions for the optimal investment strategy in a set-up with a constant risk-free interest rate $r$ and a single risky asset (representing the stock market) with price $P_t$ evolving as

$$dP_t = P_t [(r + \sigma \lambda_t) \, dt + \sigma \, dz_t],$$

(11.1)

where the volatility $\sigma$ is assumed to be a positive constant, but the market price of risk $\lambda_t$ follows a mean-reverting process. Note that in this setting the market price of risk is identical to the
Sharpe ratio of the stock. Kim and Omberg (1996) consider an investor with a CRRA utility of terminal wealth only, which allows them to let $\lambda_t$ have an undiversifiable risk component. On the other hand, Wachter (2002) considers a time-separable CRRA utility function of consumption, so to obtain explicit solutions she assumes that the market price of risk is perfectly (negatively) correlated with the price level. Wachter argues that the assumption of a correlation of $-1$ is empirically not unreasonable. To allow for non-perfect correlation we write the dynamics of $\lambda$ as

$$d\lambda_t = \kappa \left[ \bar{\lambda} - \lambda_t \right] dt + \rho \sigma \lambda_t z_t + \sqrt{1 - \rho^2 \sigma} \lambda_t \, dz_t. \quad (11.2)$$

All constants are assumed positive, except the correlation parameter $\rho$. The market price of risk is assumed to follow an Ornstein-Uhlenbeck process with long-term average $\bar{\lambda}$, mean reversion speed $\kappa$, and volatility $\sigma$. A negative value of the correlation $\rho$ will represent mean reversion in the returns on the stock in the following sense. A positive shock $dz_t$ will then affect the current stock return $\frac{dP_t}{P_t} = \frac{P_t+dt-P_t}{P_t}$ positively and the market price of risk $\lambda_{t+dt} = \lambda_t + d\lambda_t$ negatively. Hence the market price of risk and the expected stock return for a short period starting at $t + dt$ will be lower. So high realized return in the current period will be followed by a low expected return in the following period. Likewise, low realized return in the current period will be followed by a high expected return in the subsequent period.

Let us first study how the distribution of future prices is affected by the mean reversion property. It follows from the price dynamics (11.1) that

$$P_T = P_t \exp \left\{ \int_t^T \left[ r - \frac{1}{2} \sigma^2 + \sigma \lambda_s \right] ds + \int_t^T \sigma \, dz_u \right\} = P_t \exp \left\{ \left[ r - \frac{1}{2} \sigma^2 \right] (T - t) + \sigma \int_t^T \lambda_s ds + \int_t^T \sigma \, dz_u \right\}. \quad (11.3)$$

From (11.2), it follows that

$$\lambda_s = \tilde{\lambda} + e^{-\kappa(s-t)} (\lambda_t - \tilde{\lambda}) + \int_t^s \rho \sigma \lambda e^{-\kappa(s-u)} \, dz_u + \int_t^s \sqrt{1 - \rho^2 \sigma} \lambda e^{-\kappa(s-u)} \, dz_u$$

and, hence,

$$\begin{align*}
\int_t^T \lambda_s &\, ds = \tilde{\lambda} (T - t) + (\lambda_t - \tilde{\lambda}) \int_t^T e^{-\kappa(s-t)} \, ds \\
&\quad + \int_t^T \left[ \int_t^s \rho \sigma \lambda e^{-\kappa(s-u)} \, dz_u \right] \, ds + \int_t^T \left[ \int_t^s \sqrt{1 - \rho^2 \sigma} \lambda e^{-\kappa(s-u)} \, dz_u \right] \, ds.
\end{align*}$$

To proceed, we interchange the order of integration in the two double integrals, which leaves us with

$$\begin{align*}
\int_t^T \lambda_s &\, ds = \tilde{\lambda} (T - t) + (\lambda_t - \tilde{\lambda}) \int_t^T e^{-\kappa(s-t)} \, ds \\
&\quad + \int_t^T \left[ \int_u^T \rho \sigma \lambda e^{-\kappa(s-u)} \, ds \right] \, dz_u + \int_t^T \left[ \int_u^T \sqrt{1 - \rho^2 \sigma} \lambda e^{-\kappa(s-u)} \, ds \right] \, dz_u \\
&= \tilde{\lambda} (T - t) + (\lambda_t - \tilde{\lambda}) b(T - t) + \int_t^T \rho \sigma \lambda b(T - u) \, dz_u + \int_t^T \sqrt{1 - \rho^2 \sigma} \lambda b(T - u) \, dz_u \\
&= \tilde{\lambda} (T - t) + (\lambda_t - \tilde{\lambda}) b(T - t) + \int_t^T \rho \sigma \lambda b(T - s) \, dz_s + \int_t^T \sqrt{1 - \rho^2 \sigma} \lambda b(T - s) \, dz_s,
\end{align*}$$

where $b(s) = \int_s^T e^{-\kappa(t-s)} \, dt$ is the discount factor for the future stock return.
where we have introduced \( b(\tau) = (1 - e^{-\kappa \tau})/\kappa \), and where the last line simply replaces \( u \) by \( s \) in the integrals. Next, we substitute this expression into (11.3) and combine the two \( z \)-integrals so that we end up with

\[
P_T = P_t \exp \left\{ \left( r - \frac{\sigma^2}{2} + \sigma \lambda \right) (T - t) + \sigma b(T - t) (\lambda_t - \lambda) \right. \\
+ \left. \sigma \int_t^T (1 + \rho \sigma \lambda b(T - s)) \, dz_s + \sigma \sigma \lambda \sqrt{1 - \rho^2} \int_t^T b(T - s) \, dz_s \right\}.
\]

Only the last two terms are stochastic and since the integrands are deterministic functions of time, the two stochastic integrals are normally distributed random variables. Hence, \( P_T \) is lognormally distributed. Since the integrals have mean zero, we get

\[
\mathbb{E}_t[\ln P_T] = \ln P_t + \left( r - \frac{\sigma^2}{2} + \sigma \lambda \right) (T - t) + \sigma b(T - t) (\lambda_t - \lambda).
\]

The variance is

\[
\text{Var}_t[\ln P_T] = \text{Var}_t \left[ \sigma \int_t^T (1 + \rho \sigma \lambda b(T - s)) \, dz_s + \sigma \sigma \lambda \sqrt{1 - \rho^2} \int_t^T b(T - s) \, dz_s \right] \\
= \sigma^2 \left( \int_t^T (1 + \rho \sigma \lambda b(T - s))^2 \, ds + \sigma^2 \lambda^2 (1 - \rho^2) \int_t^T b(T - s)^2 \, ds \right) \\
= \sigma^2 \int_t^T (1 + 2 \rho \sigma \lambda b(T - s) + \sigma^2 \lambda^2 b(T - s)^2) \, ds \\
= \sigma^2 \left( 1 + \frac{2 \rho \sigma \lambda}{\kappa} + \frac{\sigma^2 \lambda^2}{\kappa^2} \right) (T - t) - \left( \frac{2 \rho \sigma \lambda}{\kappa} + \frac{\sigma^2 \lambda^2}{\kappa^2} \right) b(T - t) - \frac{\sigma^2 \lambda^2}{2\kappa} b(T - t)^2,
\]

where the last equality follows from the integrals

\[
\int_t^T b(T - u) \, du = \frac{1}{\kappa} (T - t - b(T - t)), \quad \int_t^T b(T - u)^2 \, du = \frac{1}{\kappa^2} (T - t - b(T - t))^2 - \frac{1}{2\kappa} b(T - t)^2.
\]

With a constant Sharpe ratio \( \lambda \), the stock price would follow a geometric Brownian motion so that the future price would be lognormally distributed with \( \mathbb{E}_t[\ln P_T] = \ln P_t + \left( r - \frac{\sigma^2}{2} + \sigma \lambda \right) (T - t) \) and \( \text{Var}_t[\ln P_T] = \sigma^2 (T - t) \). If we take the ratio of the variance of \( \ln P_T \) with the mean reversion feature to the variance of \( \ln P_T \) without mean reversion, we get

\[
\frac{\text{Var}_t[\ln P_T]}{\sigma^2 (T - t)} = 1 + \frac{2 \rho \sigma \lambda}{\kappa} + \frac{\sigma^2 \lambda^2}{\kappa^2} - \left( \frac{2 \rho \sigma \lambda}{\kappa} + \frac{\sigma^2 \lambda^2}{\kappa^2} \right) \frac{b(T - t)}{T - t} - \frac{\sigma^2 \lambda^2}{2\kappa} \frac{b(T - t)^2}{T - t} \\
\rightarrow 1 + \frac{2 \rho \sigma \lambda}{\kappa} + \frac{\sigma^2 \lambda^2}{\kappa^2} \quad \text{for} \ T \to \infty.
\]

The variations in the Sharpe ratio will therefore decrease the variance in the long run if

\[
\frac{2 \rho \sigma \lambda}{\kappa} + \frac{\sigma^2 \lambda^2}{\kappa^2} < 0 \quad \Leftrightarrow \quad \rho < -\frac{\sigma \lambda}{2\kappa},
\]

i.e., if the correlation between the Sharpe ratio and the stock price is sufficiently negative.

Figure 11.1 illustrates the effects of the mean reversion feature on the distribution of \( \ln(P_T/P_0) \) for \( T = 5 \) and \( T = 30 \) years by comparing with the distribution under the assumption of the standard Merton model in which \( \lambda_t \) is constant and the stock price follows a geometric Brownian motion (GBM). As expected, the distribution with mean reversion has thinner tails and a higher
Figure 11.1: Effects of mean reversion on the distribution of the log-return, \( \ln(P_T/P_0) \). The graphs show the distribution of log-return with mean reversion (black curve) and without mean reversion, i.e., assuming the stock price follows a geometric Brownian motion (red curve). The parameter values are \( r = 0.03, \sigma = 0.2, \kappa = 0.02, \bar{\lambda} = 0.3, \sigma_{\lambda} = 0.01, \rho = -0.8, \) and \( \lambda_t = \bar{\lambda} \).

Given the seemingly reasonable parameter values assumed when generating the figure, the differences between the two distributions are not visible for horizons lower than one year (not illustrated), still quite small for the 5-year horizon, while very clear for the 30-year horizon. This suggests that it is more important to take the mean reversion property of stock returns into account for investors with relatively long investment horizons. Figure 11.2 shows that the mean reversion feature increases the probability that a 100% stock market position outperforms a 100% risk-free position, but—with the given parameters—the increase is rather limited even for long investment horizons.

Now, let us turn to the effect of mean reversion on the optimal investment strategy for investors with a constant relative risk aversion \( \gamma > 1 \). In the model introduced above the market price of risk is the only state variable, i.e., we put \( x = \lambda \) in the notation of Chapter 7. For CRRA utility we have from Section 7.3 that the indirect utility function will be of the form

\[
J(W, \lambda, t) = \frac{1}{1 - \gamma} g(\lambda, t)^\gamma W^{1 - \gamma}.
\]

Since \( \lambda \) is an affine function of itself, we have a “quadratic” model according to the classification in Chapter 7. For an investor with CRRA utility of terminal wealth only, we get from Theorem 7.9 that the indirect utility function is given by

\[
J(W, \lambda, t) = \frac{1}{1 - \gamma} \left( W e^{A_0 (T-t) + A_1 (T-t) \lambda + \frac{1}{2} A_2 (T-t) \lambda^2} \right)^{1 - \gamma}
\]

and the optimal investment strategy in the stock is

\[
\Pi(W, \lambda, t) = \frac{1}{\gamma} \frac{\lambda}{\sigma} - \frac{1}{\gamma} \frac{\rho \sigma \lambda}{\sigma} (A_1 (T-t) + A_2 (T-t) \lambda).
\]

In the notation of Section 7.3.3 we have

\[
\begin{align*}
    r_0 &= r, & r_1 &= 0, & r_2 &= 0, & m_0 &= \kappa \bar{\lambda}, \\
    m_1 &= -\kappa, & \Lambda_0 &= 0, & \Lambda_1 &= 0, & \Lambda_2 &= 1, \\
    K_0 &= 0, & K_1 &= \rho \sigma \lambda, & \|v\|^2 &= \rho^2 \sigma_\lambda^2, & \hat{v}^2 &= (1 - \rho^2) \sigma_\lambda^2.
\end{align*}
\]
If we define
\[ \bar{\kappa} = \kappa + \gamma - \frac{1}{\gamma} \rho \sigma_{\lambda}, \]
assume that\(^2\)
\[ \bar{\kappa}^2 + \sigma_{\lambda}^2 (\rho^2 + \gamma (1 - \rho^2)) \frac{\gamma - 1}{\gamma^2} > 0 \]
and define
\[ \nu = 2 \sqrt{\bar{\kappa}^2 + \frac{\gamma - 1}{\gamma^2} \sigma_{\lambda}^2 (\rho^2 + \gamma (1 - \rho^2))}, \]
it follows from Section 7.3.3 that
\[
A_2(\tau) = \frac{2}{\gamma} \left( \frac{e^{\nu \tau} - 1}{(\nu + 2 \bar{\kappa})(e^{\nu \tau} - 1) + 2\nu} \right),
\]
\[
A_1(\tau) = \frac{4 \kappa \lambda}{\gamma \nu} \left( \frac{e^{\nu \tau/2} - 1}{(\nu + 2 \bar{\kappa})(e^{\nu \tau} - 1) + 2\nu} \right),
\]
and
\[
A_0(\tau) = r \tau + \kappa \lambda \int_0^\tau A_1(s) \, ds + \frac{1}{2} \sigma_{\lambda}^2 \int_0^\tau A_2(s) \, ds - \frac{\gamma - 1}{2\gamma} \sigma_{\lambda}^2 (\rho^2 + \gamma (1 - \rho^2)) \int_0^\tau A_1(s)^2 \, ds
\]
\[
= r \tau + \frac{1}{\gamma} \left( 2 \kappa^2 \lambda^2 \nu^2 + \frac{\sigma_{\lambda}^2}{\nu + 2 \bar{\kappa}} \right) \frac{\nu}{\nu + 2 \bar{\kappa}} \frac{\kappa^2 \lambda^2}{\gamma} \nu^3 + \frac{4 \kappa^2 \lambda^2 (\nu - 4 \bar{\kappa}) e^{-\nu \tau} + 8 \kappa e^{-\nu \tau/2} - 4 \bar{\kappa} - \nu}{2 \nu - (\nu - 2 \bar{\kappa}) (1 - e^{-\nu \tau})}
\]
\[
+ \frac{\gamma}{2(\gamma - 1)} \rho^2 + \gamma (1 - \rho^2) \ln \left( \frac{2\nu - (\nu - 2 \bar{\kappa})(1 - e^{-\nu \tau})}{2\nu} \right),
\]
where the last equality is adapted from Kim and Omberg (1996).

---

\(^2\)This condition will be satisfied except for “extreme” combinations of \(\kappa\), \(\sigma_{\lambda}\), \(\rho\), and \(\gamma\). A discussion of the solution if this condition is not satisfied can be found in Kim and Omberg (1996).
With $\gamma > 1$, it can be shown that $A_1(\tau)$ and $A_2(\tau)$ are positive\(^3\) and increasing.\(^4\) If the current value of the market price of risk is positive and the correlation is negative (consistent with empirical observations), it follows that the hedge term of the optimal portfolio is positive and increasing with the horizon of an investor with $\gamma > 1$. An investor with a long horizon should therefore invest a larger fraction of wealth in stocks than an investor with the same risk aversion, but a shorter horizon. This is consistent with typical recommendations of investment advisors. This is illustrated by Figure 11.3 for reasonable parameter values. Note, however, that the extra fraction of wealth invested in the stock due to the mean reversion of returns is relatively small even for long horizons. Figure 11.4 shows how the optimal stock allocation depends on the current market price of risk, both in the model with mean reversion and in the model where the market price of risk is assumed to be constant.

With utility from intermediate consumption and possibly terminal wealth, we must assume that either $\rho = 1$ or $\rho = -1$. We will stick to the latter, more realistic case. The restriction $\rho = -1$ affects all the functions $A_0$, $A_1$, and $A_2$ due to the presence of $\rho$ in $\bar{\kappa}$ and $q$. For notational simplicity let us consider an investor with utility stemming only from intermediate consumption, i.e., $\varepsilon_2 = 0$. From Theorem 7.10, we get that the optimal investment strategy is

$$
\Pi(W, \lambda, t) = \frac{1}{\gamma} \frac{\lambda}{\bar{\kappa}} + \frac{\gamma - 1}{\gamma} D(\lambda, t, T) \frac{\sigma_{\lambda}}{\bar{\kappa}},
$$

where

\[\begin{align*}
\rho^2 + \gamma(1 - \rho^2) > \rho^2 + 1 - \rho^2 = 1 \quad \text{and, hence,} \quad \nu + 2\bar{\kappa} > 2\sqrt{\nu^2} + 2\bar{\kappa} \geq 0. \quad \text{It is then clear that } A_1 \text{ and } A_2 \text{ are positive.}
\end{align*}\]

\[\begin{align*}
A_1'(\tau) &= 4\gamma^{-1} \kappa \lambda e^{\nu/2} [(\nu + 2\bar{\kappa})(e^{\nu/2} - 1) + 2\nu]/[\nu + 2\bar{\kappa} + 2\nu] > 0, \quad \text{which is also positive.}
\end{align*}\]
11.2 Mean reversion in stock returns

Figure 11.4: Optimal portfolio weight of stock as a function of the current market price of risk. The parameter values are $\gamma = 2$, $T - t = 30$, $r = 0.03$, $\kappa = 0.02$, $\bar{\lambda} = 0.3$, $\sigma = 0.01$, $\rho = -0.8$, and $\lambda_t = \bar{\lambda}$.

where

$$D(\lambda, t, T) = \int_t^T \left( A_1(s - t) + A_2(s - t)\lambda \right) \tilde{g}(\lambda, t; s) \, ds,$$

and we must insert $\rho = -1$ in the expressions of the $A_i$’s. Again it can be shown that, for $\gamma > 1$ and $\lambda > 0$, the hedging component is positive and increasing with the time horizon $T$. With intermediate consumption the horizon effect on the stock investment is dampened relative to the case with utility from terminal wealth only since the “effective” investment horizon is lower than $T$.

The optimal consumption rate is

$$C(W, \lambda, t) = \left( \int_t^T \tilde{g}(\lambda, t; s) \, ds \right)^{-1} W.$$
in states with relatively bad future investment opportunities, i.e., low $\lambda$. With $\rho = -1$, stocks have high returns exactly when $\lambda$ is low so the investor will hold more stocks relative to the case with constant investment opportunities.

Various empirical studies show that value (high book-to-market) stocks and growth (low book-to-market) stocks have risk-return characteristics that deviate considerably from the general stock market, cf., e.g., Fama and French (1992, 2007) and Campbell and Vuolteenaho (2004). In particular, short-term returns of value stocks have a higher average and lower standard deviation than growth stocks, but value stocks are riskier than growth stocks in the long run. Although the practical interest in value and growth stocks is immense, only few papers have studied dynamic portfolio choice models taking into account the special characteristics of value and growth stocks. Lynch (2001) and Jurek and Viceira (2011) use a vector auto regression model of return dynamics incorporating predictive variables and allow for infrequent rebalancing of portfolios. While Lynch solves for the optimal portfolios by numerical dynamic programming, Jurek and Viceira suggest a recursive approach based on an approximation. Larsen and Munk (2012) set up a continuous-time model that leads to exact and relative simple closed-form expressions for the optimal strategies and for the losses associated with selected suboptimal strategies. The model allows both for special return characteristics in growth and value stocks and for mean reversion in returns. They derive simple expressions (involving solutions to some Ricatti-type ODEs) for the optimal investments in the different types of stocks and a risk-free asset for a power utility maximizing investor. The model is estimated using U.S. return data.


11.3 Stochastic volatility

As discussed in Section 7.2.3, stochastic volatility will only induce hedging to the extent that it affects the market prices of risk. Suppose, for example, that a CRRA investor can trade in a risk-free asset with a constant interest rate and in the stock market index following the process

$$dP_t = P_t \left[ (r + \lambda_1 \sigma_t) \ dt + \sigma_t \ dz_{1t} \right],$$

where the volatility $\sigma_t$ can follow any stochastic process. The Sharpe ratio of the stock, $\lambda_1$, is assumed constant. Then the optimal fraction of wealth invested in the stock index is

$$\pi_t = \frac{1}{\gamma} \frac{\lambda_1}{\sigma_t}$$

without any hedge term. The dynamics of wealth is then

$$dW_t = \left( W_t \left[ r + \pi_t \sigma_t \lambda_1 \right] - c_t \right) dt + W_t \pi_t \sigma_t \ dz_{1t}$$

$$= \left( W_t \left[ r + \frac{\lambda_1^2}{\gamma} \right] - c_t \right) dt + W_t \frac{\lambda_1}{\gamma} \ dz_{1t}$$

with the constant relative risk exposure that the CRRA investor prefers. The optimal combination of the risk-free asset and the stock index varies over time as the volatility varies, corresponding to movements along the instantaneous mean-variance frontier. Of course to know exactly how the portfolio is to be rebalanced, it is necessary to model the fluctuations in volatility.
The more interesting case is when the Sharpe ratio of the stock depends on the level of the volatility. A tractable model is the Heston model, introduced by Heston (1993) for the pricing of stock options in the presence of stochastic volatility. The dynamics of the stock price and the instantaneous variance \( V_t = \sigma_t^2 \) of the stock is assumed to be

\[
dP_t = P_t \left( (r + \lambda V_t) dt + \sqrt{V_t} d z_{1t} \right),
\]
\[
dV_t = \kappa (\bar{V} - V_t) dt + \rho \sigma \sqrt{V_t} d z_{1t} + \sqrt{1 - \rho^2} \sigma \sqrt{V_t} d z_{2t},
\]

where \( z_1 \) and \( z_2 \) are independent standard Brownian motions. Hence, \( \sigma \sqrt{V_t} \) is the volatility of the variance, \( \rho \) is the instantaneous correlation between the stock and the variance, and the variance is assumed to mean-revert around a long-term level of \( \bar{V} \) with \( \kappa \) reflecting the speed of mean reversion. The market price of risk associated with \( z_1 \), which is identical to the Sharpe ratio of the stock, is \( \lambda (V_t) = \bar{\lambda} \sqrt{V_t} \). Of course, we can safely assume \( \bar{\lambda} > 0 \).

If \(|\rho| \neq 1\) and the investor can only trade in the stock and the riskfree asset, the market is incomplete. The optimal portfolio choice of a CRRA investor in this framework was derived by Liu (1999, 2007) and Kraft (2005) and follows as a special case of our analysis of affine models. The state variable is \( V_t \), which obviously follows an affine process, and the squared market price of risk is proportional to \( V_t \) and thus also affine. More precisely, in the notation of Section 7.3.2, we have

\[
\begin{align*}
  r_0 &= r, \quad r_1 = 0, \quad m_0 = \kappa \bar{V}, \quad m_1 = -\kappa, \quad V_0 = 0, \quad V_1 = \rho^2 \sigma_V^2, \\
  \dot{v}_0 &= 0, \quad \dot{v}_1 = (1 - \rho^2) \sigma_V^2, \quad A_0 = 0, \quad A_1 = \lambda_1^2, \quad K_0 = 0, \quad K_1 = \rho \sigma \lambda_1.
\end{align*}
\]

Let

\[
\bar{\kappa} = \kappa + \gamma - \frac{1}{\gamma} \rho \sigma \lambda_1.
\]

The condition (7.23) becomes

\[
\bar{\kappa}^2 + \frac{\gamma - 1}{\gamma^2} \lambda_1^2 \sigma_V^2 \left( \rho^2 + \gamma [1 - \rho^2] \right) > 0,
\]

which is certainly satisfied for \( \gamma > 1 \). Defining

\[
\nu = \sqrt{\bar{\kappa}^2 + \frac{\gamma - 1}{\gamma^2} \lambda_1^2 \sigma_V^2 \left( \rho^2 + \gamma [1 - \rho^2] \right)},
\]

the key function \( A_1(\tau) \) follows from (7.24):

\[
A_1(\tau) = \frac{\lambda_1^2}{\gamma} e^{\nu \tau} - 1 \over \gamma (\nu + \bar{\kappa}) (e^{\nu \tau} - 1) + 2 \nu.
\]

\( A_1 \) is positive and increasing in \( \tau \). For a CRRA investor with utility of time \( T \) wealth only, the optimal fraction of wealth invested in the stock is then

\[
\pi(t) = \frac{\lambda_1}{\gamma} - \frac{\gamma - 1}{\gamma} \rho \sigma \lambda_1 (T - t) = \frac{\lambda_1}{\gamma} - \frac{\gamma - 1}{\gamma^2} \lambda_1^2 \rho \sigma \left( \frac{e^{\nu (T - t)} - 1}{\nu (\nu + \bar{\kappa}) (e^{\nu (T - t)} - 1) + 2 \nu} \right).
\]

cf. Theorem 7.7. Note that the portfolio weight does not vary with the current volatility level.

Empirical estimates of the correlation between the stock and its instantaneous variance are negative. Volatility tends to go up, when stock prices go down. Consequently, the hedge term is
positive. A low variance represents a situation of bad investment opportunities since the market prices of risk are then also low. Due to the negative correlation, stocks have a built-in hedge: should investment opportunities deteriorate (falling variance), the stock will typically increase substantially in price.

When the volatility of the stock is stochastic and imperfectly correlated with the stock, options on the stock are non-redundant assets. By investing in an option, which is sensitive to the shock $z_2$, the investor can improve his welfare. Let $O_t = f(P_t, V_t, t)$ denote the price of such an option (or any asset/portfolio with non-zero exposure to $z_2$), where $f$ is assumed to be sufficiently differentiable. By Itô’s Lemma

$$dO_t = \ldots dt + \frac{\partial f}{\partial P_t} \frac{\partial P_t}{\partial t} + \frac{\partial f}{\partial V_t} \frac{\partial V_t}{\partial t} \left( \rho \sigma_V \sqrt{V_t} dz_t + \sqrt{1 - \rho^2 \sigma_V^2} \sqrt{V_t} dz_{2t} \right)$$

An important element of the model, which has to be specified at this point, is the market price of risk $\lambda_{2t}$ associated with $z_2$. Following Liu and Pan (2003), assume that $\lambda_{2t} = \hat{\lambda}_2 \sqrt{V_t}$, which will keep us in the affine model class. The expected rate of return of the option is then

$$\mu_t = r + \lambda_t (f(P_t, V_t, t) P_t + f_V(P_t, V_t, t) \rho \sigma_V) \sqrt{V_t} O_t^{1/2} + \lambda_{2t} f_V(P_t, V_t, t) \sqrt{1 - \rho^2 \sigma_V} \sqrt{V_t} O_t^{1/2} \lambda_{2t}.$$

We now have a complete markets model with two risky assets having a volatility matrix of

$$\Sigma = \begin{pmatrix} \sqrt{V_t} & 0 \\ \frac{\partial f}{\partial P_t} \frac{\partial P_t}{\partial t} & \frac{\partial f}{\partial V_t} \frac{\partial V_t}{\partial t} \sqrt{1 - \rho^2 \sigma_V} \sqrt{V_t} \end{pmatrix},$$

and

$$\lambda(V_t) = \begin{pmatrix} \hat{\lambda}_1 \sqrt{V_t} \\ \hat{\lambda}_2 \sqrt{V_t} \end{pmatrix}, \quad \nu(V_t) = \begin{pmatrix} \rho \sigma_V \sqrt{V_t} O_t^{1/2} \\ \sqrt{1 - \rho^2 \sigma_V} \sqrt{V_t} \end{pmatrix}, \quad \nu^t(V_t) = 0.$$

In the notation and terminology of Section 7.3.2, this is an affine model with

$$r_0 = r, \quad r_1 = 0, \quad m_0 = \kappa \hat{V}, \quad m_1 = -\kappa, \quad V_0 = 0, \quad V_1 = \sigma_V^2, \quad \hat{v}_0 = 0, \quad \hat{v}_1 = 0, \quad \Lambda_0 = 0, \quad \Lambda_1 = \hat{\lambda}_1^2 + \hat{\lambda}_2^2, \quad K_0 = 0, \quad K_1 = \sigma_V \left( \rho \bar{\lambda}_1 + \sqrt{1 - \rho^2 \bar{\lambda}_2} \right).$$

Let

$$\hat{\kappa} = \kappa + \frac{\gamma - 1}{\gamma} \sigma_V (\rho \bar{\lambda}_1 + \sqrt{1 - \rho^2 \bar{\lambda}_2})$$

and note that the condition (7.23) is satisfied as long as $\gamma > 1$. Define

$$\hat{\nu} = \sqrt{\hat{\kappa}^2 + \frac{\gamma - 1}{\gamma^2} \sigma_V^2 (\hat{\lambda}_1^2 + \hat{\lambda}_2^2)}.$$

From (7.24) we get the relevant version of the $A_1$-function, which we now denote by $\hat{A}_1$ to distinguish it from the $A_1$-function earlier in this section:

$$\hat{A}_1(\tau) = \frac{\hat{\lambda}_1^2 + \hat{\lambda}_2^2}{\gamma} \frac{e^{\hat{\nu} \tau} - 1}{(\hat{\nu} + \hat{\kappa})(e^{\hat{\nu} \tau} - 1) + 2\hat{\nu}}.$$

Just like $A_1$ above, $\hat{A}_1$ is also positive and increasing.
According to Theorem 7.7, the optimal portfolio for an investor with CRRA utility of time $T$ wealth is then

$$
\begin{pmatrix}
\pi_{St} \\
\pi_{Ot}
\end{pmatrix} = \frac{1}{\gamma} \left( g_t^\gamma \right)^{-1} \begin{pmatrix}
\hat{\lambda}_1 \sqrt{V_t} \\
\hat{\lambda}_2 \sqrt{V_t}
\end{pmatrix} - \frac{\gamma - 1}{\gamma} \left( g_t^\gamma \right)^{-1} \begin{pmatrix}
\rho \sigma V \sqrt{V_t} \\
\sqrt{1 - \rho^2} \sigma V \sqrt{V_t}
\end{pmatrix} \hat{A}_1(T - t),
$$

which implies that the fraction of wealth optimally invested in the option is

$$\pi_{Ot} = \frac{O_t}{f_V(P_t, V_t, t)} \left( \frac{\hat{\lambda}_2}{\gamma \sigma V \sqrt{1 - \rho^2}} - \frac{\gamma - 1}{\gamma} \hat{A}_1(T - t) \right),$$

and the fraction of wealth optimally invested in the stock is

$$\pi_{St} = \frac{1}{\gamma} \left( \hat{\lambda}_1 - \hat{\lambda}_2 \left[ \frac{\rho}{\sqrt{1 - \rho^2}} + \frac{f_P(P_t, V_t, t)P_t}{\sigma V \sqrt{1 - \rho^2} f_V(P_t, V_t, t)} \right] \right) + \frac{\gamma - 1}{\gamma} \frac{f_P(P_t, V_t, t)P_t}{f_V(P_t, V_t)} \hat{A}_1(T - t)$$

$$= \frac{\hat{\lambda}_1}{\gamma} - \frac{\rho \hat{\lambda}_2}{\gamma \sqrt{1 - \rho^2}} - f_P(P_t, V_t, t)P_t \pi_{Ot}.$$

Let us assume that the option price is positively related to the stock volatility so that $f_V(P_t, V_t, t) > 0$. The hedge demand for the option is then negative. The hedge portfolio should increase in value when the variance $V_t$ drops as this implies deteriorating market prices of risk. As the option price increases with the variance, a short position in the option will give the desired hedge. The sign of the speculative demand for the option equals the sign of the constant $\hat{\lambda}_2$ in the market price of $z_2$-risk. According to most empirical studies, this market price of risk is negative; see, e.g., Bakshi and Kapadia (2003) and Chernov and Glysels (2000). A negative position in the option will give a negative exposure to the volatility-specific risk represented by $z_2$, which leads to a positive risk premium. Both the speculative demand and the hedging demand for the option are thus negative.

In their illustration of the solution, Liu and Pan (2003) assumes that the “option” the investor trades in is a so-called delta-neutral straddle. A straddle is a combination of a long position in a call and a long position in a put with the same strike prices and maturity dates. The strike price is determined so that the delta of the call, i.e., the derivative of the call price with respect to the stock price, equals $\frac{1}{2}$. Then it follows from the put-call parity that the delta of the put equals $-\frac{1}{2}$ so that the delta of the straddle is equal to zero. The value of the straddle is thus insensitive to small changes in the stock price. On the other hand, the straddle will be highly sensitive to changes in the volatility, so it is an obvious instrument for “trading volatility.” In their numerical illustrations, Liu and Pan (2003) find for example that the optimal portfolio of an investor with a relative risk aversion of 3 and a horizon of 5 years consists of (approximately) 24% in the stock and -54% in the straddle, and thus 130% in the riskfree asset. This is certainly a non-standard investment recommendation.

Liu and Pan (2003) and Larsen and Munk (2012) compute utility losses from ignoring options completely when determining the optimal investment or from including options in a suboptimal way. Both studies conclude that the utility losses from excluding options can be substantial. The results of Larsen and Munk (2012) indicate that inclusion of the option is mainly important because it gives access to the apparently sizeable volatility risk premium, whereas the benefits from

---

5There is less consensus about the magnitude of the market price of volatility risk. Liu and Pan (2003) refers to $\hat{\lambda}_2 = 0 - 6$ as a conservative estimate. Note that it remains unclear whether the assumption that $\lambda_{2t}$ is proportional to $\sigma_t = \sqrt{V_t}$ is appropriate.
volatility hedging are smaller. Of course, the attractiveness of the option depends heavily on the estimate of the parameter $\lambda_2$.

Liu and Pan (2003) extend the above setting to include jumps of a given size in the stock price, motivated by the observed stock market crashes. With the assumption that the intensity of jump arrivals is be proportional to $V_t$, they are able to find a closed-form solution (this is an affine jump-diffusion setting). Their results indicate that the estimates of the jump size, the jump risk premium, and the jump intensity are highly important for the optimal option position. A put option provides protection against big drops in the stock price and thus becomes more attractive due to the inclusion of negative jumps in the model. Liu, Longstaff, and Pan (2003) and Branger, Schlag, and Schneider (2008) consider extensions where both the stock price and the variance may jump.

Chacko and Viceira (2005) consider a quite spurious model with stochastic volatility that does not fit into the cases where we have explicit solutions. They find explicit, approximate solutions for an investor with Epstein-Zin utility.

11.4 More

Mean reversion and momentum in stock prices: Koijen, Rodriguez, and Sbuelz (2009)
Correlation risk: Buraschi, Porchia, and Trojani (2010)

11.5 Exercises

**Exercise 11.1.** Throughout this exercise consider an individual with a time-additive expected power utility of consumption and/or terminal wealth, so that the objective of the individual at any time $t \leq T$ is to maximize

$$
E_t \left[ \int_t^T e^{-\delta(s-t)} \varepsilon_1 u(c_s) \, ds + e^{-\delta(T-t)} \varepsilon_2 u(W_T) \right],
$$

where $\varepsilon_1, \varepsilon_2 \geq 0$ with $\varepsilon_1 + \varepsilon_2 > 0$, $\delta > 0$ is the subjective time preference rate, and $u(c) = \frac{1}{1-\gamma} c^{1-\gamma}$, where $\gamma > 1$ is the constant relative risk aversion. The individual has an initial (time $t$) wealth of $W_t$ and earns no income from non-financial sources. The individual has access to a financial market with a risk-free asset and a risky asset (a stock). The risk-free asset pays a constant continuously compounded annualized rate of return of $r$. The risky asset has a price process $P = (P_t)$ with dynamics

$$
dP_t = P_t [\mu_t \, dt + \sigma_t \, dz_t],
$$

where $z = (z_t)$ is a one-dimensional standard Brownian motion, and $\mu_t$ and $\sigma_t$ are well-behaved stochastic processes. Below you are going to consider various models for $\mu_t$ and $\sigma_t$.

First, consider “Model 1” in which $\mu_t = r + \sigma_t \lambda$ for a constant $\lambda > 0$ and any well-behaved $\sigma_t$.

(a) What is the optimal consumption and investment strategy of the individual?

Next, consider “Model 2” in which $\mu_t = \mu > r$ and $\sigma_t = k P_t^\beta$ for some constants $k > 0$ and $\beta \in \mathbb{R}$.

(b) Show that the instantaneous return variance rate of the stock, $\frac{1}{P_t} \text{Var}_t[dP_t/P_t]$, has a constant
11.5 Exercises

elasticity with respect to the stock price.\(^6\) (The elasticity of a function \(f(x)\) is defined as \(\frac{df}{f} = \frac{df}{dx} \frac{x}{f(x)}\).) Describe the impact of the sign of \(\beta\) on the relation between the stock price and the volatility.

(c) Determine and describe the market price of the (stock) risk in this model.

The natural next step would be to note that the indirect utility must be depending on the stock price, i.e., of the form \(J(W_t, P_t, t)\), and then write down and try to solve the associated HJB-equation. However, the HJB-equation for \(J(W, P, t)\) turns out to appear quite complicated (try it yourself!). A change-of-variable simplifies the equation to be solved. Define the process \(x_t = (x_t)\) by

\[ x_t = k^{-2} P_t^{-2 \beta}. \]

(d) What is the dynamics of \(x\) (if possible, express \(dx_t\) without any \(P_t\) in the equation)?

(e) Argue that the model fits into the affine setting of Chapter 7. State the optimal consumption and investment strategy. Find explicit expressions for the two deterministic functions entering the solution, i.e., \(A_0\) and \(A_1\) in the notation of Section 7.3 of the lecture notes.

(f) How does the optimal consumption and investment at a given point in time depend on the stock price at that date? Explain! How does the optimal investment depend on the time horizon? (When considering the optimal investment here, you may assume \(\varepsilon_1 = 0, \varepsilon_2 > 0\).)

Next, consider “Model 3”: suppose that \(\mu_t = \mu > r\) and \(\sigma_t = 1/\sqrt{\bar{x}_t}\), where

\[ dx_t = \kappa(\bar{x} - x_t) dt + \sigma_x \sqrt{x_t} \left( \rho dz_t + \sqrt{1 - \rho^2} d\hat{z}_t \right), \]

where \(\hat{z} = (\hat{z}_t)\) is a one-dimensional standard Brownian motion independent of \(z\) and \(\rho \in [-1, +1]\).

(g) Determine the dynamics of the instantaneous stock variance \(V_t = \sigma_t^2\); if possible, express \(dV_t\) without any \(x_t\) in the equation. What is the instantaneous correlation between the stock price and the instantaneous stock variance?

(h) For the case \(\varepsilon_1 = 0, \varepsilon_2 > 0\), find the optimal investment strategy. Provide explicit expressions for any functions entering the optimal investment strategy.

(i) Compare models 2 and 3.

Assume the following parameter values:

\[ r = 0.02, \quad \mu = 0.10, \quad \kappa = 0.34, \quad \bar{x} = 28, \quad \sigma_x = 0.65, \quad \rho = 0.5. \]

(j) Assume that the current (time 0) value of the state variable is equal to the long-run level, i.e., \(x_0 = \bar{x}\). For all combinations of \(\gamma \in \{1.01, 2, 4, 10, 20\}\) and \(T \in \{1, 5, 10, 30\}\), compute the optimal fraction of wealth invested in the stock. How big is the hedge demand compared to the myopic demand?

(k) How sensitive is the optimal portfolio to the current volatility of the stock? Provide a few graphs illustrating your answer.

(l) For each of the three models, discuss whether the individual would benefit from having access to trade in an option on the stock.

\(^6\) Therefore the stock price process is called a CEV process (CEV: Constant Elasticity of Variance).
Exercise 11.2. In Vasicek’s original model the excess expected return of any zero-coupon bond is given by \( \lambda_1 \sigma_r b(T-t) \) and thus deterministic. However, empirical studies indicate that excess bond returns vary with the level of interest rates. We can obtain that by generalizing the Vasicek model to the so-called essentially affine Vasicek model in which the real-world short rate dynamics is still

\[
dr_t = \kappa [\bar{r} - r_t] \, dt - \sigma_r \, dz_{1t},
\]

but the market price of risk associated with \( z_1 \) is now allowed to be an affine function of the short rate:

\[
\lambda_{1t} = \bar{\lambda}_1 + \tilde{\lambda}_1 r_t,
\]

where \( \bar{\lambda}_1 \) and \( \tilde{\lambda}_1 \) are constants. It turns out that the price of a zero-coupon bond is still of the exponential-affine form

\[
B^T_t = e^{-a(T-t) - b(T-t)r_t},
\]

but \( a \) and \( b \) are different from the original Vasicek model. In particular,

\[
b(\tau) = \frac{1}{\tilde{\kappa}} \left( 1 - e^{-\tilde{\kappa} \tau} \right), \quad \tilde{\kappa} = \kappa - \sigma_r \bar{\lambda}_1.
\]

(a) State the bond price dynamics in this model.

There is also evidence that the excess expected return on the stock market vary (negatively) with the level of short-term interest rates. Write the stock price dynamics as

\[
dS_t = S_t \left[ (r_t + \sigma_S \psi_t) \, dt + \rho \sigma_S \, dz_{1t} + \sqrt{1 - \rho^2} \sigma_S \, dz_{2t} \right].
\]

Here \( \psi_t \) is the instantaneous Sharpe ratio of the stock. Assume that the market price of risk associated with \( z_2 \) is of the affine form

\[
\lambda_{2t} = \bar{\lambda}_2 + \tilde{\lambda}_2 r_t,
\]

where \( \bar{\lambda}_2 \) and \( \tilde{\lambda}_2 \) are constants.

(b) Determine \( \psi_t \) as a function of \( r_t \) and check that the model can potentially capture the explained predictability pattern.

Now think of the asset allocation model in which a CRRA investor (with no labor income and utility of terminal wealth only) can invest in the bank account, a single bond, and the stock index with price dynamics as stated above.

(c) Verify that the model fits into the quadratic framework.

(d) Determine the optimal investment strategy (including the necessary \( A_i \)-functions).

(e) How does the optimal strategy in this model differ from the strategy found in Section 10.2, where we assumed the original Vasicek model and no return predictability?
12.1 Introduction

We should recognize that the models discussed in the previous chapters really use the consumption good as the numeraire and, hence, all asset prices are assumed to be formulated in real terms, i.e., the price of an asset is the number of units of the consumption good into which the asset can be exchanged. In particular, the bonds considered in the models have been real bonds that pay out in units of the consumption good. However, in many markets real bonds are not traded (at least not at a volume ensuring liquid prices). Furthermore, the risk-free asset in the previous models is assumed to be risk-free in real terms and the short-term interest rate is the real short rate. The risk-free asset has been modeled as a continuous roll-over strategy in deposits over infinitesimal short periods. While such a strategy is of course quite extreme, it may be seen as a reasonable approximation to a strategy of frequently rolling over short-term deposits. While it may be possible to lock in a risk-free nominal return over a short period, it seems to be more questionable to get someone to promise you a return which is risk-free in real terms. In this chapter we will discuss effects of inflation risk on asset allocation, optimal asset allocation involving nominal bonds and asset allocation without a truly risk-free asset.

12.2 Real and nominal price dynamics

In order to study the link between the real and the nominal return on an asset we have to model the dynamics of the nominal asset price and the price of the consumption good. Let $\tilde{P}_t$ denote the nominal price of asset $i$ at time $t$, i.e., the price in monetary units (like dollars). Let $\Phi_t$ denote the monetary price of the consumption good at time $t$. (With many consumption goods we may loosely think of $\Phi_t$ as the consumer price index.) Then the real price of asset $i$ is $P_{rt} = \tilde{P}_t/\Phi_t$.

Assume that

$$d\tilde{P}_t = \tilde{P}_t [\tilde{\mu}_t dt + \tilde{\sigma}_t^\tau d\tilde{z}_t]$$
and that
\[ d\Phi_t = \Phi_t [\varphi_t dt + \sigma_{\Phi t}^T d\tilde{z}_t + \tilde{\sigma}_{\Phi t} dz_{\Phi t}], \]
where \( z_{\Phi} \) is a one-dimensional standard Brownian motion independent of \( \tilde{z} \). We can interpret
\( d\Phi_t/\Phi_t \) as the realized inflation over the period \([t, t+dt]\), which in general will not be known before \( t + dt \), and interpret \( \varphi_t \) as the expected rate of inflation per year. Using Itô’s Lemma, we can derive the dynamics of the real price of asset \( i \) as
\[ dP_{it} = P_{it} \left[ (\tilde{\mu}_{it} - \varphi_t - \tilde{\sigma}_{\Phi t}^T \sigma_{\Phi t} + \| \sigma_{\Phi t} \|^2 + \tilde{\sigma}_{\Phi t}^2) dt + (\tilde{\sigma}_{\Phi t} - \sigma_{\Phi t})^T d\tilde{z}_t - \tilde{\sigma}_{\Phi t} dz_{\Phi t} \right]. \]

It is important to realize that if there is uncertainty about the change in consumer prices over the deposit period, the real value of a nominal deposit will not be risk-free. Let \( \tilde{r}_t \) denote the nominal risk-free short rate at time \( t \). Then the nominal value of the nominally risk-free asset satisfies
\[ d\tilde{A}_t = \tilde{r}_t \tilde{A}_t dt. \]
The real value of the nominally risk-free asset is again obtained by deflating with the price of consumption, \( A_t = \tilde{A}_t/\Phi_t \), and applying Itô’s Lemma we get
\[ dA_t = A_t \left[ (\tilde{r}_t - \varphi_t + \| \sigma_{\Phi t} \|^2 + \tilde{\sigma}_{\Phi t}^2) dt - \sigma_{\Phi t}^T d\tilde{z}_t - \tilde{\sigma}_{\Phi t} dz_{\Phi t} \right]. \]

Unless there is no uncertainty about the realized inflation, the nominally risk-free asset is risky in real terms.

Assume now that we have \( m \) nominally risky assets with nominal price dynamics of the form
\[ d\tilde{P}_t = \text{diag}(\tilde{P}_t) \left[ \tilde{\mu}_t dt + \tilde{\sigma}_{\tilde{P}_t} d\tilde{z}_t \right], \]
where \( \tilde{z} \) is an \( m \)-dimensional standard Brownian motion. Suppose that we invest fractions of wealth given by the vector \( \pi_t \) in the nominally risky assets and, consequently, the fraction \( 1 - \pi_t^T 1 \) in the nominally risk-free asset. Then the nominal wealth \( \tilde{W}_t \) will evolve as
\[ d\tilde{W}_t = (\tilde{r}_t \tilde{W}_t + \tilde{W}_t \pi_t^T (\tilde{\mu}_t - \tilde{r}_t 1) - c_t \pi_t) dt + \tilde{W}_t \pi_t^T \tilde{\sigma}_{\tilde{P}_t} d\tilde{z}_t, \]
where \( c_t \) is the number of units consumed of the consumption good. The real wealth is \( W_t = \tilde{W}_t/\Phi_t \), which evolves as
\[ dW_t = \left( W_t \left[ \tilde{r}_t - \varphi_t + \| \sigma_{\Phi t} \|^2 + \sigma_{\Phi t}^2 \right] + W_t \pi_t^T \left( \tilde{\mu}_t - \tilde{r}_t 1 - \tilde{\sigma}_{\tilde{P}_t} \sigma_{\Phi t} \right) - c_t \right) dt + W_t \left( \pi_t^T \tilde{\sigma}_{\tilde{P}_t} - \sigma_{\Phi t}^T \right) d\tilde{z}_t - W_t \sigma_{\Phi t} dz_{\Phi t}. \]

Of course, we could also derive the dynamics of real wealth directly from the dynamics of real prices. We can see that any asset will have the same real sensitivity towards the shock process \( z_{\Phi} \) so that it will be impossible to hedge against such a shock.

If \( \tilde{\sigma}_{\tilde{P}_t} \) is non-singular, we can define
\[ \tilde{\lambda}_t = \tilde{\sigma}_{\tilde{P}_t}^{-1} (\tilde{\mu}_t - \tilde{r}_t 1), \]
which has the interpretation as the nominal market price of risk vector. Then the dynamics of real wealth can be rewritten as
\[ dW_t = \left( W_t \left[ \tilde{r}_t - \varphi_t + \| \sigma_{\Phi t} \|^2 + \sigma_{\Phi t}^2 \right] + W_t \pi_t^T \tilde{\sigma}_{\tilde{P}_t} \left( \tilde{\lambda}_t - \sigma_{\Phi t} \right) - c_t \right) dt + W_t \left( \pi_t^T \tilde{\sigma}_{\tilde{P}_t} - \sigma_{\Phi t}^T \right) d\tilde{z}_t - W_t \sigma_{\Phi t} dz_{\Phi t}. \]
12.3 Constant investment opportunities

If $\dot{\sigma}_{qt} = 0$ and $\ddot{\sigma}_{t}$ is non-singular, the inflation uncertainty is spanned by the traded assets and we can obtain a risk-free real return by investing in the portfolio given by $\pi_{t}^{\text{safe}} = \left(\ddot{\sigma}_{t}^{-1}\right)^{-1} \sigma_{qt}$. The rate of return in this portfolio is then the real short-term interest rate which will be

$$r_t = \dot{r}_t - \varphi_t + \|\sigma_{qt}\|^{2} + \ddot{\sigma}_{t} \mu_t - \ddot{r}_t, $$

$$= \dot{r}_t - \varphi_t + \sigma_t^{\top} \lambda_t.$$

The above set-up involves $m + 1$ assets in total, all of them being risky in an inflation-adjusted sense except when $\dot{\sigma}_{qt} = 0$ and $\ddot{\sigma}_{t}$ is non-singular. We have given special attention to one of these assets, namely the nominally risk-free asset. Since we can loosely interpret this asset as “cash”, it may make sense to include that asset in an asset allocation framework. However, there is really nothing especially attractive about that asset. Therefore we might as well collect all real risky asset prices in a vector $P_t$ with dynamics of the form

$$dP_t = \text{diag}(P_t) \left[ \mu_t dt + \sigma_t dz_t \right].$$

(12.3)

and no other assets, in particular no risk-free asset. If we let $\omega_t$ denote the vector of portfolio weights invested in these assets, we have to require that $\omega_t^{\top} 1 = 1$. The real wealth dynamics is

$$dW_t = [W_t \omega_t^{\top} \mu_t - c_t] dt + W_t \omega_t^{\top} \sigma_t dz_t.$$  

(12.4)

The earlier formulation of the price dynamics can be fitted into this more general framework by letting the nominally risk-free asset be one of the assets, say the one corresponding to the last element in the price vector. Furthermore, we must let

$$\omega_t = \begin{pmatrix} \pi_t \\ 1 - \pi_t^{\top} 1 \end{pmatrix},$$

$$z_t = \begin{pmatrix} \dot{z}_t \\ \ddot{z}_{qt} \end{pmatrix},$$

$$\mu_t = \begin{pmatrix} \ddot{r}_t - \varphi_t + \|\sigma_{qt}\|^{2} + \ddot{\sigma}_{t} \mu_t \\ \ddot{r}_t - \varphi_t + \|\sigma_{qt}\|^{2} + \ddot{\sigma}_{t} 1 \end{pmatrix},$$

$$\sigma_t = \begin{pmatrix} \ddot{\sigma}_{qt} - \ddot{\sigma}_{qt} 1 \\ -\ddot{\sigma}_{qt} 1 \end{pmatrix}. $$

(12.5)

12.3 Constant investment opportunities

In this section we solve our general utility maximization problem in the case where no asset is risk-free in real terms and real investment opportunities are constant.

12.3.1 General formulation

First, we consider the case where the real price dynamics of all available assets are given by (12.3) so that the real wealth dynamics for a given consumption process $c = (c_t)$ and a given portfolio process $\omega = (\omega_t)$ is represented by (12.4). The indirect utility function is

$$J(W_t, t) = \sup_{(c_t, \omega_t), c_t \in [t, T]} \mathbb{E}_{W_t} \left[ \int_t^T e^{-\delta(s-t)} \varepsilon_1 u(c_s) ds + e^{-\delta(T-t)} \varepsilon_2 \tilde{u}(W_T) \right],$$

(12.6)
where the indicators \( \varepsilon_1 \) and \( \varepsilon_2 \) are either zero or one with at least one of them being equal to one, and where it is implicitly understood that \( \omega_s^1 \mathbf{1} = 1 \) for all \( s \). The HJB-equation associated with the utility maximization problem is

\[
\delta J(W, t) = \sup_{c \geq 0, \omega^\top \mathbf{1} = 1} \left\{ \varepsilon_1 \left( u(c) - c J_W(W, t) \right) + \frac{\partial J}{\partial \omega}(W, t) + J_W(W, t)W \omega^\top \mu \right. \\
+ \left. \frac{1}{2} J_{WW}(W, t) W^2 \omega^\top \sigma \sigma^\top \omega \right\}
\]

with the terminal condition \( J(W, T) = \varepsilon_2 \tilde{u}(W) \). The first-order condition for consumption is the usual envelope condition. The first-order condition for the portfolio is now different since we have to maximize under the constraint \( \omega^\top \mathbf{1} = 1 \). The Lagrangian for this constrained maximization problem is

\[
\mathcal{L} = J_W(W, t)W \omega^\top \mu + \frac{1}{2} J_{WW}(W, t) W^2 \omega^\top \sigma \sigma^\top \omega - \nu \left( 1 - \omega^\top \mathbf{1} \right),
\]

where \( \nu \) is the Lagrange multiplier. Solving for \( \omega \), we get

\[
\omega = -\frac{J_W(W, t)}{W J_{WW}(W, t)} \left( \sigma \sigma^\top \right)^{-1} \mu - \frac{\nu}{W^2 J_{WW}(W, t)} \left( \sigma \sigma^\top \right)^{-1} \mathbf{1}.
\]

The constraint \( \omega^\top \mathbf{1} = 1^\top \omega = 1 \) implies that

\[
1 = -\frac{J_W(W, t)}{W J_{WW}(W, t)} 1^\top \left( \sigma \sigma^\top \right)^{-1} \mu - \frac{\nu}{W^2 J_{WW}(W, t)} 1^\top \left( \sigma \sigma^\top \right)^{-1} \mathbf{1},
\]

so that

\[
-\frac{\nu}{W^2 J_{WW}(W, t)} = \frac{1 - \left( -\frac{J_W(W, t)}{W J_{WW}(W, t)} \right) 1^\top \left( \sigma \sigma^\top \right)^{-1} \mu}{1^\top \left( \sigma \sigma^\top \right)^{-1} \mathbf{1}}.
\]

The optimal portfolio is therefore

\[
\omega = -\frac{J_W(W, t)}{W J_{WW}(W, t)} \left( \sigma \sigma^\top \right)^{-1} \mu + \frac{1 - \left( -\frac{J_W(W, t)}{W J_{WW}(W, t)} \right) 1^\top \left( \sigma \sigma^\top \right)^{-1} \mu}{1^\top \left( \sigma \sigma^\top \right)^{-1} \mathbf{1}} \left( \sigma \sigma^\top \right)^{-1} \mathbf{1}.
\]

This is a combination of two portfolios, namely the portfolio

\[
\omega_{\text{slope}} = \frac{1}{1^\top \left( \sigma \sigma^\top \right)^{-1} \mu} \left( \sigma \sigma^\top \right)^{-1} \mu
\]

and the portfolio

\[
\omega_{\text{min}} = \frac{1}{1^\top \left( \sigma \sigma^\top \right)^{-1} \mathbf{1}} \left( \sigma \sigma^\top \right)^{-1} \mathbf{1}
\]

since the optimal portfolio can be written as

\[
\omega = -\frac{J_W(W, t)}{W J_{WW}(W, t)} \left( 1^\top \left( \sigma \sigma^\top \right)^{-1} \mu \right) \omega_{\text{slope}} + \left( 1 - \left( -\frac{J_W(W, t)}{W J_{WW}(W, t)} \right) \frac{1^\top \left( \sigma \sigma^\top \right)^{-1} \mu}{1^\top \left( \sigma \sigma^\top \right)^{-1} \mathbf{1}} \right) \omega_{\text{min}}.
\]

(12.7)

Again, we have a two-fund separation result. As discussed in the one-period mean-variance framework, see Equations (3.16) and (3.17), we can interpret \( \omega_{\text{min}} \) as the minimum-variance portfolio and \( \omega_{\text{slope}} \) as the portfolio with the largest mean-to-standard deviation ratio (i.e., the maximum slope in a \((\sigma, \mu)\)-diagram).

With CRRA utility of both consumption and terminal wealth, it can be shown (by solving the HJB-equation) that the indirect utility function is given by

\[
J(W, t) = \frac{1}{1 - \gamma} \hat{g}(t)^{\gamma} W^{1 - \gamma},
\]

(12.8)
where
\[
g(t) = \frac{1}{\tilde{A}} \left( \varepsilon_1^{1/\gamma} + (\varepsilon_2^{1/\gamma} \tilde{A} - \varepsilon_1^{1/\gamma}) e^{-\tilde{A}[T-t]} \right),
\]
and
\[
\tilde{A} = \frac{\delta}{\gamma} - \frac{1 - \gamma}{2\gamma} \left( \mu^\top (\sigma \sigma^\top)^{-1} \mu - \gamma k^2 (\sigma \sigma^\top)^{-1} \mu \right),
\]

\[
k = \frac{1}{1^\top (\sigma \sigma^\top)^{-1} 1} \left( 1 - \frac{1}{\gamma} 1^\top (\sigma \sigma^\top)^{-1} \mu \right).
\]

The optimal portfolio is then
\[
\omega = \frac{1}{\gamma} (\sigma \sigma^\top)^{-1} \mu + k (\sigma \sigma^\top)^{-1} 1,
\]
while the optimal consumption rate is
\[
c = \varepsilon_1^{1/\gamma} W \frac{g(t)}{\tilde{g}(t)}.
\]

The general structure of the solution is thus the same as for the case with a traded risk-free asset.

**12.3.2 Formulation with a nominally risk-free asset**

Now let us consider the formulation where one of the assets is the nominally risk-free asset so that the dynamics of real wealth is of the form (12.1). The indirect utility function is now
\[
J(W, t) = \sup_{(c, \pi) \in [t,T]} E_{W,t} \left[ \int_t^T e^{-\delta(s-t)} \varepsilon_1 u(c_s) ds + e^{-\delta(T-t)} \varepsilon_2 \tilde{u}(W_T) \right],
\]
and the associated HJB-equation is
\[
\delta J(W, t) = \sup_{c, \pi} \left\{ \varepsilon_1 \left( u(c) - c J_W(W, t) \right) + \frac{\partial J}{\partial t}(W, t) \right. \\
+ W J_W(W, t) \left[ \tilde{c} - \varphi + \| \sigma \phi \|^2 + \tilde{\phi}^\top \mu - \tilde{\phi}^\top 1 - \tilde{\sigma} \sigma \phi \right] \\
\left. + \frac{1}{2} W^2 J_{WW}(W, t) \left( \tilde{\phi}^\top \tilde{\sigma} \tilde{\phi}^\top + 2 \| \sigma \phi \|^2 + \tilde{\phi}^\top \tilde{\sigma} \tilde{\phi} \right) \right\}
\]
with terminal condition \( J(W, T) = \varepsilon_2 u(W) \). Here there is no constraint on the portfolio vector \( \pi \).

As always, if \( \varepsilon_1 = 1 \), the first-order condition for \( c \) is the envelope condition, which implies that \( c = I_u(J_W(W, t)) \), where \( I_u \) is the inverse of \( u' \). The first-order condition for \( \pi \) implies that
\[
\pi = (\tilde{\sigma} \tilde{\sigma}^\top)^{-1} \tilde{\sigma} \sigma \phi - \frac{J_W(W, t)}{W J_{WW}(W, t)} (\tilde{\sigma} \tilde{\sigma}^\top)^{-1} \left( \tilde{\mu} - \tilde{\phi}^\top 1 - \tilde{\sigma} \sigma \phi \right),
\]
which gives the optimal portfolio weights in the nominally risky assets so that the optimal weight in the nominally risk-free asset is
\[
\pi_t^0 = 1 - \pi^\top 1 = 1 - 1^\top (\tilde{\sigma} \tilde{\sigma}^\top)^{-1} \tilde{\sigma} \sigma \phi + \frac{J_W(W, t)}{W J_{WW}(W, t)} 1^\top (\tilde{\sigma} \tilde{\sigma}^\top)^{-1} \left( \tilde{\mu} - \tilde{\phi}^\top 1 - \tilde{\sigma} \sigma \phi \right). \tag{12.10}
\]

If \( \tilde{\sigma} \) is non-singular, we may simplify these expressions to
\[
\pi = (\tilde{\sigma}^\top)^{-1} \sigma \phi - \frac{J_W(W, t)}{W J_{WW}(W, t)} (\tilde{\sigma}^\top)^{-1} \left( \tilde{\lambda} - \sigma \phi \right), \tag{12.11}
\]
\[
\pi_t^0 = 1 - \pi^\top 1 = 1 - 1^\top (\tilde{\sigma}^\top)^{-1} \sigma \phi + \frac{J_W(W, t)}{W J_{WW}(W, t)} 1^\top \left( \tilde{\lambda} - \sigma \phi \right).
\]

Apparently, the optimal portfolio exhibits three-fund separation with the three funds being
172 Chapter 12. Inflation risk and asset allocation with no risk-free asset

(1) the portfolio of nominally risky assets given by the weights
\[ \pi = \frac{1}{1^\top \left( \hat{\sigma} \hat{\sigma}^\top \right)^{-1} \sigma} \left( \hat{\sigma} \hat{\sigma}^\top \right)^{-1} \hat{\sigma} \sigma \];
this basically mimics the inflation process as well as possible,

(2) the portfolio of nominally risky assets given by the weights
\[ \pi = \frac{1}{1^\top \left( \hat{\sigma} \hat{\sigma}^\top \right)^{-1} \left( \hat{\mu} - \hat{r}1 - \hat{\sigma} \sigma \right)} \left( \hat{\sigma} \hat{\sigma}^\top \right)^{-1} \left( \hat{\mu} - \hat{r}1 - \hat{\sigma} \sigma \right), \]

(3) the nominally risk-free asset.

However, since the formulation in this subsection is just a special case of that in the previous subsection, we know that two-fund separation obtains. In fact, using the link between the two formulations given by (12.5)–(12.6), one can verify (after some hours work!) that the portfolio vector \( \left( \pi_t, \pi^0_t \right) \) defined by (12.9) and (12.10) is identical to the portfolio vector \( \omega_t \) defined by (12.7).

For CRRA utility, the indirect utility function is, of course, given by (12.8), but we can rewrite the constant \( \hat{A} \) as
\[
\hat{A} = \frac{\delta}{\gamma} + \frac{\gamma - 1}{\gamma} \left[ \hat{r} - \varphi + \frac{1}{2\gamma} \left( \hat{\mu} - \hat{r}1 \right)^\top \left( \hat{\sigma} \hat{\sigma}^\top \right)^{-1} \left( \hat{\mu} - \hat{r}1 \right) \right.
\]
\[
+ \frac{\gamma - 1}{\gamma} \left( \hat{\mu} - \hat{r}1 \right)^\top \left( \hat{\sigma} \hat{\sigma}^\top \right)^{-1} \hat{\sigma} \sigma + \left( 1 - \frac{\gamma}{2} \right) \left( \| \sigma \|^2 + \hat{\sigma}^2 \right)
\]
\[
- \left( 1 - \frac{\gamma}{2} - \frac{1}{2\gamma} \right) \sigma^\top \left( \hat{\sigma} \hat{\sigma}^\top \right)^{-1} \hat{\sigma} \sigma \left. \right].
\]

If \( \hat{\sigma} \) is non-singular, we can write \( \hat{A} \) as
\[
\hat{A} = \frac{\delta}{\gamma} + \frac{\gamma - 1}{\gamma} \left[ \hat{r} - \varphi + \frac{1}{2\gamma} \| \hat{\lambda} \|^2 + \frac{\gamma - 1}{\gamma} \hat{\lambda}^\top \sigma + \left( 1 - \frac{\gamma}{2} \right) \hat{\sigma}^2 + \frac{1}{2\gamma} \| \sigma \|^2 \right].
\]

where \( \hat{\lambda} = \hat{\sigma}^{-1} (\hat{\mu} - \hat{r}1) \) as defined earlier.

12.4 General stochastic investment opportunities

We now turn to the case where investment opportunities are stochastic. Let us consider the setting in which the investor will invest in a nominally risk-free asset and a number of nominally risky assets so that the dynamics of his real wealth for a given consumption process \( c = (c_t) \) and a given portfolio process \( \pi = (\pi_t) \) is represented by (12.1).

MORE TO COME LATER – go to specific case in next subsection...

12.5 Hedging real interest rate risk without real bonds

It is sometimes claimed that stocks are appropriate for hedging inflation uncertainty so that the real returns on stocks are quite stable relative to the real returns on long-term nominal bonds. This could explain the popular advice that long-term investors should invest more in stocks than short-term investors.

If only nominal bonds are traded, the optimal investment strategy of an investor with utility of terminal wealth only is to combine the mean-variance portfolio and the portfolio that has the
highest correlation with the return on an indexed bond with a maturity equal to the remaining horizon. The hedge portfolio generally involves both stocks and nominal bonds, the precise mix will be determined by the correlation structure. If inflation uncertainty is modest, nominal bonds are good substitutes for real bonds (true in the U.S. for the period 1983-2000; not true for 1950-1982) and nominal bonds will dominate the hedge portfolio. Estimates on U.S. data over the period of approximately 1950-2000 show that the stock index is slightly positively correlated with the real interest rate. Hence the stock will enter the hedge portfolio with a negative weight unlike the popular advice.

General aspects of the portfolio choice problem with uncertain inflation are discussed by Munk and Sørensen (2007). The effects of uncertain inflation on portfolio choice have been studied in concrete settings by e.g., Brennan and Xia (2002), Munk, Sørensen, and Vinther (2004), and Campbell and Viceira (2001). Both Brennan and Xia (2002) and Munk, Sørensen, and Vinther (2004) consider investors with CRRA utility of wealth at the end of a finite horizon, whereas Campbell and Viceira (2001) allow for intermediate consumption and a more general recursive utility specification in an infinite horizon setting. The infinite horizon assumption, however, makes it difficult to address effects due to investors having different investment horizons. In both Brennan and Xia (2002) and Campbell and Viceira (2001) (a proxy for) the real interest rate is described by a one-factor Vasicek model and the expected inflation dynamics is given by an Ornstein-Uhlenbeck process. The term structure of nominal interest rates is therefore described by a two-factor model. Munk, Sørensen, and Vinther (2004) differ slightly by assuming a one-factor Vasicek model for the nominal interest rates, while the implied term structure of real interest rates is described by a two-factor model. In the model of Munk, Sørensen, and Vinther it is impossible to replicate a real bond by trading in any number of nominal bonds whereas this is possible in the other models. The main conclusions of Brennan and Xia (2002) and Munk, Sørensen, and Vinther (2004) are very close, however. For concreteness, let us follow the set-up of Munk, Sørensen, and Vinther.

We consider the investment problem of an investor who has CRRA utility of terminal (time $T$) real wealth only. As before $\gamma$ represents the relative risk aversion of the agent. The investor can hold cash (i.e., a money market bank account), nominal bonds, and stocks. The nominal short rate dynamics is described by an Ornstein-Uhlenbeck process,

$$d\tilde{r}_t = \kappa(\bar{r} - \tilde{r}_t) dt - \sigma_r \, dz_{1t},$$

as we have previously assumed to hold for the real short rate. The dynamics of the nominal price $\tilde{B}_t$ of any bond (or other fixed-income securities) is of the form

$$d\tilde{B}_t = \tilde{B}_t \left[ (\tilde{r}_t + \tilde{\lambda}_1 \tilde{\sigma}_B(\tilde{r}_t, t)) \, dt + \sigma_B(\tilde{r}_t, t) \, dz_{1t} \right],$$

where $\tilde{\lambda}_1$ is the (nominal) market price of risk induced by the exogenous shock process $z_1$. The nominal stock price or stock index value (with dividends reinvested) is assumed to evolve according to the stochastic differential equation

$$d\tilde{S}_t = \tilde{S}_t \left[ (\tilde{r}_t + \tilde{\psi} \tilde{\sigma}_S) \, dt + \rho_{BS} \tilde{\sigma}_S \, dz_{1t} + \sqrt{1 - \rho_{BS}^2} \tilde{\sigma}_S \, dz_{2t} \right].$$

1The model of Munk, Sørensen, and Vinther (2004) also allows for mean reversion in stock returns in a similar way as studied in Chapter 11. We ignore that feature in the discussion here.
The parameter $\rho_{BS}$ is the correlation between bond market returns and stock market returns, $\hat{\sigma}_S$ is the volatility of the nominal stock price, and $\hat{\psi}$ is the Sharpe ratio of the stock which we assume constant. In total, the dynamics of nominal asset prices can be written as

$$\begin{pmatrix} d\tilde{B}_t \\ d\tilde{S}_t \end{pmatrix} = \begin{pmatrix} \tilde{B}_t & 0 \\ 0 & \tilde{S}_t \end{pmatrix} \begin{pmatrix} \tilde{r}_t 1 + \frac{\tilde{\sigma}_B(\tilde{r}_t, t)}{\rho_{BS}\hat{\sigma}_S} \frac{\hat{\lambda}_1}{\sqrt{1 - \rho_{BS}^2}} \\ \frac{\hat{\lambda}_2}{\sqrt{1 - \rho_{BS}^2}} \end{pmatrix} \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} dt + \begin{pmatrix} \tilde{\sigma}_B(\tilde{r}_t, t) \\ \rho_{BS}\hat{\sigma}_S \end{pmatrix} \begin{pmatrix} 0 \\ \frac{\hat{\lambda}_2}{\sqrt{1 - \rho_{BS}^2}} \end{pmatrix} \begin{pmatrix} dz_{1t} \\ dz_{2t} \end{pmatrix},$$

where $\hat{\lambda}_2 = (\hat{\psi} - \rho_{BS}\hat{\lambda}_1)/\sqrt{1 - \rho_{BS}^2}$. Letting $\pi = (\pi_B, \pi_S)^T$ denote the fractions of wealth invested in the bond and the stock, the nominal wealth $\tilde{W}_t$ will evolve as

$$d\tilde{W}_t = \tilde{W}_t \begin{pmatrix} \tilde{r}_t + \pi_1^T \tilde{\sigma}(\tilde{r}_t, t) \hat{\lambda} \\ \pi_2^T \tilde{\sigma}(\tilde{r}_t, t) \end{pmatrix} dt + \pi_1^T \tilde{\sigma}(\tilde{r}_t, t) \begin{pmatrix} dz_{1t} \\ dz_{2t} \end{pmatrix},$$

where

$$\tilde{\sigma}(\tilde{r}_t, t) = \begin{pmatrix} \tilde{\sigma}_B(\tilde{r}_t, t) & 0 \\ \rho_{BS}\hat{\sigma}_S & \frac{\hat{\lambda}_2}{\sqrt{1 - \rho_{BS}^2}} \end{pmatrix}, \quad \hat{\lambda} = \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix}.$$

The dynamics of the nominal price of the consumption good is given by the following system of differential equations:

$$\frac{d\Phi_t}{\Phi_t} = \varphi_t dt + \sigma_{\varphi_1} dz_{1t} + \sigma_{\varphi_2} dz_{2t} + \sigma_{\varphi_3} dz_{3t},$$

and

$$d\varphi_t = \beta(\bar{\varphi} - \varphi_t) dt + \sigma_{\varphi_1} dz_{1t} + \sigma_{\varphi_2} dz_{2t} + \sigma_{\varphi_3} dz_{3t} + \sigma_{\varphi_4} dz_{4t},$$

where $\varphi_t$ is the expected rate of inflation, $\bar{\varphi}$ describes the long-run mean of the rate of inflation, $\beta$ describes the degree of mean-reversion, and the volatility coefficients $\sigma_{\varphi_k}$ and $\sigma_{\varphi_k}$ are all constant. Define $\Sigma_{\varphi}^2 = \sigma_{\varphi_1}^2 + \sigma_{\varphi_2}^2 + \sigma_{\varphi_3}^2$ and $\Sigma_{\varphi}^2 = \sigma_{\varphi_1}^2 + \sigma_{\varphi_2}^2 + \sigma_{\varphi_3}^2 + \sigma_{\varphi_4}^2$. The instantaneous variance rates of the price index and the expected inflation rate are then $\Sigma_{\Phi}^2$ and $\Sigma_{\varphi}^2$, respectively.

The real wealth of the investor at time $t$ is $W_t = \tilde{W}_t/\Phi_t$, which by Itô’s Lemma has the dynamics

$$dW_t = W_t \begin{pmatrix} \tilde{r}_t - \varphi_t + \Sigma_{\Phi}^2 + \pi_1^T \tilde{\sigma}(\tilde{r}_t, t) \left( \hat{\lambda} - \begin{pmatrix} \sigma_{\varphi_1} \\ \sigma_{\varphi_2} \end{pmatrix} \right) \\ \pi_1^T \tilde{\sigma}(\tilde{r}_t, t) \end{pmatrix} \begin{pmatrix} dz_{1t} \\ dz_{2t} \end{pmatrix} - \begin{pmatrix} \sigma_{\varphi_1} \\ \sigma_{\varphi_2} \\ \sigma_{\varphi_3} \end{pmatrix}^T \begin{pmatrix} dz_{1t} \\ dz_{2t} \\ dz_{3t} \end{pmatrix},$$

which is just how the equation (12.2) looks like in this specific model. The variables $W$, $\tilde{r}$, and $\varphi$ form a Markov system and provide sufficient information for the decisions of the investor. Hence, the indirect utility is given as a function $J(W, \tilde{r}, \varphi, t)$.

Let us focus on the utility maximization problem of an investor with utility of terminal wealth so that the indirect utility function is

$$J(W, \tilde{r}, \varphi, t) = \sup_{(\pi_k)_{k \in \{1, 2\}}} E_W \tilde{r}, \varphi, t [u(W_T)]$$.
The associated HJB-equation is

\[
0 = \sup_{\pi = (\pi_B, \pi_S) \in \mathbb{R}^2} \left\{ \frac{\partial J}{\partial t} + W J_W \left\{ \tilde{r} - \phi + \Sigma_\phi^2 + \pi^T \tilde{\sigma}(\tilde{r}, t) \left( \lambda - \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \end{pmatrix} \right) \right\} + \frac{1}{2} W^2 J_{WW} \left( \pi^T \tilde{\sigma}(\tilde{r}, t) \tilde{\sigma}(\tilde{r}, t)^T \pi + \Sigma_\phi^2 - 2 \pi^T \tilde{\sigma}(\tilde{r}, t) \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \end{pmatrix} \right) + \kappa (\tilde{r} - \bar{r}) J_{\tilde{r}} + \frac{1}{2} \sigma_\phi^2 J_{\phi \phi} + \frac{1}{2} \Sigma_\phi^2 J_{\phi \phi} \right\}
\]

\[
+ W J_{W \phi} (\pi_B \bar{\sigma}_B - \sigma_{\phi_1}) - J_{\phi \phi} \sigma_{\phi_1} \sigma_{\phi_1} + W J_{W \phi} \begin{pmatrix} \pi^T \tilde{\sigma}(\tilde{r}, t) \left( \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \end{pmatrix} \right)^T \end{pmatrix} - \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \\ \sigma_{\phi_3} \end{pmatrix} \right) \right\}.
\]

The boundary condition is \( J(W, \tilde{r}, \phi, T) = u(W) \). The first-order condition of the maximization problem in (12.12) provides the following characterization of the optimal risky asset proportions \( \pi \):

\[
\pi = \begin{pmatrix} \pi_B \\ \pi_S \end{pmatrix} = \left( \tilde{\sigma}(\tilde{r}, t)^T \right)^{-1} \left( \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \end{pmatrix} \right) - \frac{J_W}{W J_{WW}} \left( \tilde{\sigma}(\tilde{r}, t)^T \right)^{-1} \left( \lambda - \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \end{pmatrix} \right) + \frac{J_{W \phi}}{W J_{WW} \sigma_B(\tilde{r}, t)} \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \frac{J_{W \phi}}{W J_{WW}} \left( \tilde{\sigma}(\tilde{r}, t)^T \right)^{-1} \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \end{pmatrix}.
\]

The first two terms are also present in the setting with constant investment opportunities, cf. (12.11). The last two terms hedge variations in the two state variables, i.e., the nominal short-term interest rate \( \tilde{r} \) and the expected inflation rate \( \phi \). Since the nominal bond price is perfectly correlated with the nominal short rate, only the nominal bond is used for hedging those variations. This is similar to the analysis in Chapter 10. On the other hand, both the nominal bond and the stock are generally used for hedging variations in the expected inflation rate with the weights determined by

\[
\left( \tilde{\sigma}(\tilde{r}, t)^T \right)^{-1} \begin{pmatrix} \sigma_{\phi_1} \\ \sigma_{\phi_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_B(\tilde{r}, t)} \left( \sigma_{\phi_1} - \frac{\rho_{BS}}{\sqrt{1 - \rho_{BS}^2}} \sigma_{\phi_2} \right) \\ \frac{\sigma_{\phi_1} - \frac{\rho_{BS}}{\sqrt{1 - \rho_{BS}^2}} \sigma_{\phi_2}}{\sigma_B(\tilde{r}, t)^{1/2}} \end{pmatrix}.
\]

Note that the values of \( \sigma_{\phi_1} \) and \( \sigma_{\phi_2} \) capture the correlations between the asset returns and the expected inflation:

\[
\sigma_{\phi_1} = \rho_{BS} \Sigma_\phi, \quad \rho_{BS} \sigma_{\phi_2} + \sqrt{1 - \rho_{BS}^2} \sigma_{\phi_2} = \rho_{BS} \Sigma_\phi.
\]

Now let us specialize to the case of CRRA utility, \( u(W) = \frac{1}{1 - \gamma} W^{1-\gamma} \). Note that the dynamics of the state variables \( \tilde{r} \) and \( \phi \) have an “affine” structure. Given the analysis of Chapter 7, it should therefore come as no surprise that the indirect utility function of the CRRA investor is given by

\[
J(W, \tilde{r}, \phi, t) = \frac{1}{1 - \gamma} \left( W e^{A_0(T-t) + A_1(T-t)\tilde{r} + A_2(T-t)\phi} \right)^{1-\gamma},
\]

where

\[
A_1(\tau) = \frac{1}{\kappa} (1 - e^{-\kappa \tau}), \quad A_2(\tau) = -\frac{1}{\beta} (1 - e^{-\beta \tau}),
\]

and \( A_0 \) can be found explicitly, but is not important for the optimal portfolio choice. By substitution of the relevant derivatives into (12.13), the vector of optimal risky asset allocations at time \( t \)
is given by

\[
\begin{pmatrix}
\pi_B \\
\pi_S
\end{pmatrix} = \left( \tilde{\sigma}(\tilde{r}, t)^\top \right)^{-1} \begin{pmatrix}
\sigma_{\Phi_1} \\
\sigma_{\Phi_2}
\end{pmatrix} + \frac{1}{\gamma} \left( \tilde{\sigma}(\tilde{r}, t)^\top \right)^{-1} \left( \lambda - \begin{pmatrix}
\sigma_{\Phi_1} \\
\sigma_{\Phi_2}
\end{pmatrix} \right) \\
+ \left( 1 - \frac{1}{\gamma} \right) A_1(T-t) \frac{\sigma_r}{\tilde{\sigma}_B(\tilde{r}, t)} \begin{pmatrix} 1 \\
0 \end{pmatrix} - \left( 1 - \frac{1}{\gamma} \right) A_2(T-t) \left( \tilde{\sigma}(\tilde{r}, t)^\top \right)^{-1} \begin{pmatrix}
\sigma_{\varphi_1} \\
\sigma_{\varphi_2}
\end{pmatrix}
\]

\[
= \frac{1}{\gamma} \left( \tilde{\sigma}(\tilde{r}, t)^\top \right)^{-1} \lambda + \left( 1 - \frac{1}{\gamma} \right) A_1(T-t) \frac{\sigma_r}{\tilde{\sigma}_B(\tilde{r}, t)} \begin{pmatrix} 1 \\
0 \end{pmatrix} \\
+ \left( 1 - \frac{1}{\gamma} \right) \left( \tilde{\sigma}(\tilde{r}, t)^\top \right)^{-1} \left[ \begin{pmatrix}
\sigma_{\Phi_1} \\
\sigma_{\Phi_2}
\end{pmatrix} - A_2(T-t) \begin{pmatrix}
\sigma_{\varphi_1} \\
\sigma_{\varphi_2}
\end{pmatrix} \right].
\] (12.14)

The residual \(1 - \pi_S - \pi_B\) is invested in the nominally risk-free bank account.

The optimal portfolio weights for CRRA investors are linear combinations of the speculative portfolio and the different hedge portfolios. In particular, for investors with the same investment horizon \(T\) the optimal portfolios are linear combinations of the speculative portfolio and a single hedge portfolio; the relative risk tolerance, \(1/\gamma\), describes the weights on the two relevant portfolios.

The second term in (12.14) describes the hedge against changes in the nominal interest rate and consists entirely of a position in the bond. As noted in Section 10.2, the occurrence of this hedge term implies that the bond/stock ratio will increase with the risk aversion consistent with popular recommendations. If the bond is a zero-coupon bond of the same maturity as the horizon setting. Moreover, while the last term in (12.14) can potentially explain the typically recommended horizon dependence for stocks, it may also change the ratio between bonds and stocks. However, whether this horizon effect implies more or fewer stocks for the long-term investor depends on the sign of the correlation \(\rho_{S\varphi}\) between stock returns and inflation, that is whether the stock serves as a relatively good substitute for the real bond that should ideally be used for hedging changes in real rates in a complete market setting. Moreover, while the last term in (12.14) can potentially explain the typically recommended horizon dependence for stocks, it may also change the ratio between bonds and stocks.

Munk, Sørensen, and Vinther calibrate the model using historical US data from the period 1951–2001. The estimation is based on maximum likelihood and an application of the Kalman filter. The point estimate of the correlation parameter \(\rho_{S\varphi}\) is slightly decreasing so that the optimal stock weight for \(\gamma > 1\) is slightly decreasing with the investment horizon in contrast to popular investment advice. The stock index is, in fact, positively correlated with the (proxy for the) real short-term interest rate and is therefore a bad substitute for the relevant real bond that should ideally be used as the instrument for hedging long term inflation risk and real interest rate risk. However, when the capital market parameters are allowed to vary within intervals of plus-minus two standard deviations on the estimates (which could reflect reasonable uncertainty on the parameter estimates), the theoretical asset allocation results can closely mimic popular asset allocation advice. In particular, the model

---

2Under the assumptions of the model, the (proxy for the) real short-term interest rate is given by the nominal interest rate minus the expected inflation rate plus a constant.
can generate both a bond/stock ratio which is increasing in the risk aversion coefficient and a stock investment that increases with the length of the investment horizon. The recommendations are quantitatively very difficult to match, however.
13.1 Introduction

In the general description of the continuous-time model in Section 5.2 we allowed for the case where the agent receives income from non-financial sources at a rate $y_t$. But in all the concrete problems studied until now we have assumed $y_t \equiv 0$. We shall refer to income from non-financial sources as labor income although this may in general include gifts, welfare payments, etc. In this section we will study the influence of labor income on optimal portfolio and consumption choice. Intuitively, the effects of labor income will depend on the present value of the future labor income, which we will refer to as the human wealth, and on the riskiness of labor income. An investor will focus on the magnitude and riskiness of his total wealth, i.e., the sum of the current financial and the human wealth. The size of the human wealth will therefore affect how much to consume and how much to invest and the riskiness of the human wealth will affect the allocation of financial wealth between risky assets and the risk-free asset.

13.2 A motivating example

Let us look at a small numerical example illustrating the main effects of labor income.\(^1\) Assume that investment opportunities are constant and that a single risky financial asset (representing the stock market index) is traded. With constant interest rates the risk-free asset is equivalent to a bond. Consider an investor with a financial wealth of 500,000 dollars and a constant relative risk aversion of $\gamma = 2$. Assume that the risk-free interest rate is $r = 4\%$, the expected rate of return on stocks is $\mu = 10\%$, and the volatility of the stock is $\sigma = 20\%$. (The market price of risk is $\lambda = (\mu - r)/\sigma = 0.3$.) We know from the analysis in Chapter 6 that, in the absence of labor income, it is optimal for the investor to have 75% of his wealth invested in stocks and 25% in the risk-free asset, i.e., the bond. When the investor receives labor income it seems fair to conjecture that he will invest his financial wealth such that the riskiness of his total position corresponds to

\(^1\)The example is inspired by Jagannathan and Kocherlakota (1996).
Chapter 13. Labor income

Table 13.1: Investments with a relatively short horizon. The table shows the optimal investment strategy for three types of labor income. The financial wealth is 500,000 and the capitalized labor income is 500,000 corresponding to a relatively short investment horizon.

<table>
<thead>
<tr>
<th>Risk-free income</th>
<th>Stock investment</th>
<th>Bond investment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial inv.</td>
<td>750,000</td>
<td>(150%)</td>
</tr>
<tr>
<td>Total position</td>
<td>750,000</td>
<td>(75%)</td>
</tr>
<tr>
<td>Quite risky income</td>
<td>250,000</td>
<td>(50%)</td>
</tr>
<tr>
<td>Financial inv.</td>
<td>500,000</td>
<td>(100%)</td>
</tr>
<tr>
<td>Total position</td>
<td>750,000</td>
<td>(75%)</td>
</tr>
<tr>
<td>Very risky income</td>
<td>500,000</td>
<td>(100%)</td>
</tr>
<tr>
<td>Financial inv.</td>
<td>250,000</td>
<td>(50%)</td>
</tr>
<tr>
<td>Total position</td>
<td>750,000</td>
<td>(75%)</td>
</tr>
</tbody>
</table>

 investing 75% of his total wealth in stocks. We will verify this in a following section.

Let us first assume that the investor has a labor income stream with a present value of 500,000 dollars and, hence, a total wealth of one million. It is then optimal to have a total position of 750,000 dollars in stocks and 250,000 dollars in the risk-free asset. How the financial wealth is to be allocated depends on the riskiness of his labor income. In Table 13.1 we consider three cases:

(a) If the labor income is completely risk-free, it is equivalent to a position of 0 dollars in stocks and 500,000 dollars in the risk-free asset. To obtain the desired overall riskiness, he has to allocate his financial wealth of 500,000 by investing 750,000 dollars in stocks and -250,000 dollars in the risk-free asset. This corresponds to a stock investment of 150% of the financial wealth, financed in part by borrowing 50% of the financial wealth. The certain labor income corresponds to the returns of a risk-free investment. Hence the financial wealth (and more) has to be invested in stocks to achieve the desired balance between risky and risk-free returns.

(b) If the labor income is quite risky and corresponds to an equal combination of stocks and bonds, the entire financial wealth (100%) is to be invested in stocks.

(c) If the labor income is extremely risky and corresponds to a 100% investment in stocks, the financial wealth is to be split equally between stocks and bonds.

Clearly, the optimal allocation of financial wealth is highly dependent on the risk profile of labor income.

Next, let us consider an investor with the same risk aversion, but a longer investment horizon and, consequently, a higher capitalized labor income, namely 1,500,000 dollars. Table 13.2 shows the allocation of the financial wealth that is needed to obtain the desired 75-25 split between risky and risk-free returns. Comparing with Table 13.1 we see that the younger investor in Table 13.2 will have a significantly higher fraction of financial wealth invested in stocks than the older investor in Table 13.1, except for the case where the income is extremely uncertain. The optimal stock weight in the portfolio is clearly depending on the investment horizon.
Table 13.2: Investments with a relatively long horizon. The table shows the optimal investment strategy for three types of labor income. The financial wealth is 500,000 and the capitalized labor income is 1,500,000 corresponding to a relatively long investment horizon.

<table>
<thead>
<tr>
<th>Risk type</th>
<th>Stock investment</th>
<th>Bond investment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk-free</td>
<td>0 (0%)</td>
<td>1,500,000 (100%)</td>
</tr>
<tr>
<td>Financial inv.</td>
<td>1,500,000 (300%)</td>
<td>-1,000,000 (-200%)</td>
</tr>
<tr>
<td>Total position</td>
<td>1,500,000 (75%)</td>
<td>500,000 (25%)</td>
</tr>
<tr>
<td>Quite risky</td>
<td>750,000 (50%)</td>
<td>750,000 (50%)</td>
</tr>
<tr>
<td>Financial inv.</td>
<td>750,000 (150%)</td>
<td>-250,000 (-50%)</td>
</tr>
<tr>
<td>Total position</td>
<td>1,500,000 (75%)</td>
<td>500,000 (25%)</td>
</tr>
<tr>
<td>Very risky</td>
<td>1,500,000 (100%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Financial inv.</td>
<td>0 (0%)</td>
<td>500,000 (100%)</td>
</tr>
<tr>
<td>Total position</td>
<td>1,500,000 (75%)</td>
<td>500,000 (25%)</td>
</tr>
</tbody>
</table>

According to empirical studies, the correlation between labor income and the stock market index is very small for most individuals. In that case, labor income resembles a risk-free investment more than a stock investment, and the fraction of financial wealth invested in stocks should increase with the length of the investment horizon—in line with typical investment advice. However, for some investors the labor income may be highly correlated with the stock market, or at least some individual stocks, and in that case the weight of stocks in the financial portfolio should decrease with the length of the horizon.

13.3 Exogenous income in a complete market

13.3.1 General income and price dynamics

Now we will look at the problems more formally. We take our standard setting with an instantaneously risk-free asset with a rate of return of \( r_t \) and \( d \) risky assets with price dynamics

\[
\begin{align*}
   dP_t = \text{diag}(P_t) \left[ (r_t \mathbf{1} + \sigma_t \lambda_t) \, dt + \sigma_t \, dz_t \right],
\end{align*}
\]

where \( z = (z_1, \ldots, z_d)^\top \) is a \( d \)-dimensional standard Brownian motion. We let \( \theta_t \) be the vector of amounts invested in these risky assets at time \( t \). The labor income rate is given by the process \( y = (y_t) \). From (5.4) we have that wealth evolves as

\[
\begin{align*}
   dW_t = [r_t W_t + \theta_t^\top \sigma_t \lambda_t + y_t - c_t] \, dt + \theta_t^\top \sigma_t \, dz_t,
\end{align*}
\]

where \( c = (c_t) \) is the consumption rate process. We take a Markovian framework so that we can apply the dynamic programming approach. In this section we consider the case where the labor

---

\(^2\)Davis and Willen (2000) find that – depending on the individual’s sex, age, and educational level – the correlation between aggregate stock market returns and labor income shocks is between -0.25 and 0.3, while the correlation between industry-specific stock returns and labor income shocks is between -0.4 and 0.1. Campbell and Viceira (2002) report that the correlation between aggregate stock market returns and labor income shocks is between 0.328 and 0.516. Heaton and Lucas (2000) find that the labor income of entrepreneurs typically is more highly correlated with the overall stock market (0.14) than with the labor income of ordinary wage earners (-0.07).
income rate is exogenously given. In a later section we incorporate explicitly the labor supply
decision of the agent.

Most studies of the effect of labor income on consumption and portfolio choice assume a process
for the labor income rate such as

\[ dy_t = y_t \left( \alpha(y_t, t) \, dt + \xi(y_t, t)^T \, dz_t + \tilde{\xi}(y_t, t) \, d\tilde{z}_t \right). \]

If \( \tilde{\xi} \neq 0 \), the income risk is not fully hedgeable in the financial market, which seems to be the
realistic situation. However, this is a more difficult problem to analyze, so let us first look at
the complete market case where \( \tilde{\xi} = 0 \) so that the labor income process is spanned by the price
processes of traded assets. It is well-documented that typical income growth rates and income
volatility depend on the age of the individual, so that \( \alpha, \xi, \) and \( \tilde{\xi} \) in general should depend on
time. Income growth rates tend to be high for young individuals and then slow down and eventually
become slightly negative with age. See, e.g., Cocco, Gomes, and Maenhout (2005).

In the complete market case where \( \tilde{\xi} \equiv 0 \), the income stream is fully hedgeable and can be valued
as any financial asset. We can think of the income as the dividend stream from some (possibly
strange) trading strategy in the traded financial assets. The time \( t \) value of the income stream
\( (y_s)_{s \in [t,T]} \) must be

\[ H(x, y, t) = E_{x,y,t} \left[ \int_t^T e^{-\int_t^s r(x_u) \, du} \, y_s \, ds \right] \]

where \( Q \) is the risk-neutral probability measure, and \( x \) is a state variable affecting the short-term
interest rate \( r \) and the market price of risk vector \( \lambda \). We refer to \( H(x, y, t) \) as the human wealth
of the agent at time \( t \). In this situation we can think of the agent “selling” his future income at
the financial market in the exchange of the payment \( H(x, y, t) \) so that he has a total wealth of
\( W + H(x, y, t) \) to invest. Intuitively, he will invest in a financial portfolio such that the riskiness of
his total position of financial investments and labor income is similar to the riskiness of his optimal
financial portfolio in the absence of labor income.

### 13.3.2 Constant investment opportunities and GBM income

For simplicity, we will in the following consider the classical Merton setting with constant in-
vestment opportunities, i.e., a constant interest rate \( r \) and a constant market price of risk \( \lambda \). Then
the human wealth expression simplifies to

\[ H(y, t) = E_{y,t} \left[ \int_t^T \exp \left\{ - \left( r + \frac{1}{2} \| \lambda \|^2 \right) \, (s - t) - \lambda^T (z_s - z_t) \right\} \, y_s \, ds \right] \quad (13.1) \]

and it is known from the Feynman-Kac theorem frequently applied in the option pricing literature
that the function \( H(y, t) \) satisfies the PDE

\[ \frac{\partial H}{\partial t}(y, t) + (\alpha(y, t) - \xi(y, t)^T \lambda) y H_y(y, t) + \frac{1}{2} \| \xi(y, t) \|^2 y^2 H_{yy}(y, t) - r H(y, t) + y = 0. \quad (13.2) \]

If we further assume that \( \alpha \) and \( \xi \) are constants, the labor income process is a geometric Brownian
motion so that

\[ y_s = y_t \exp \left\{ \left( \alpha - \frac{1}{2} \| \xi \|^2 \right) (s - t) + \xi^T (z_s - z_t) \right\}. \]
If we substitute this into (13.1), we can compute the human wealth in closed form as

\[
H(y, t) = y \mathbb{E}_{y,t} \left[ \int_t^T \exp \left\{ - \left( r - \alpha + \frac{1}{2} \| \lambda \|^2 + \frac{1}{2} \| \xi \|^2 \right) (s - t) + (\xi - \lambda)\top (z_s - z_t) \right\} \, ds \right]
\]

\[
= y \int_t^T \mathbb{E}_{y,t} \left[ \exp \left\{ - \left( r - \alpha + \frac{1}{2} \| \lambda \|^2 + \frac{1}{2} \| \xi \|^2 \right) (s - t) + (\xi - \lambda)\top (z_s - z_t) \right\} \right] \, ds
\]

\[
= y \int_t^T e^{-(r-\alpha+\xi\top \lambda)(s-t)} \, ds
\]

\[
= \begin{cases} 
\frac{y}{r-\alpha+\xi\top \lambda} \left( 1 - e^{-(r-\alpha+\xi\top \lambda)(T-t)} \right), & \text{if } r - \alpha + \xi\top \lambda \neq 0, \\
y(T - t), & \text{if } r - \alpha + \xi\top \lambda = 0,
\end{cases}
\]

\[
\equiv yM(t),
\]

(13.3)
i.e., the present value of the future income stream is given by the product of the current income and a time-dependent multiplier. The third equality in the above computation is due to the fact that \((\xi - \lambda)\top (z_s - z_t) \sim N(0, \| \xi - \lambda \|^2 (s - t))\) and that for a random variable \(\tilde{x} \sim N(m, s^2)\), we have \(\mathbb{E}[\exp(-a\tilde{x})] = \exp(-am + \frac{1}{2}a^2s^2)\). Note that the human wealth itself depends on the riskiness of the labor income stream, in contrast to our numerical example in the previous section.

Let us study the human wealth in a simple numerical example. The risk-free rate is \(r = 0.02\), and we assume a single risky asset (the stock market index) with a volatility of \(\sigma = 0.2\) and a Sharpe ratio of \(\lambda = 0.3\). With a single risky asset, the sensitivity of the income rate is just a scalar, \(\xi\), and since the income rate is assumed to be spanned, it must be either perfectly positively or perfectly negatively correlated with the price of the risky asset. It will perfectly positively correlated with the asset price if \(\xi\) is positive, and perfectly negatively correlated with the asset price if \(\xi\) is negative. The volatility of the income rate is the absolute value, \(|\xi|\). Let us assume that \(\xi\) is either +0.1 or −0.1 so that the income rate volatility is 10%. In Figure 13.1 we illustrate how the income multiplier \(M(t)\) and, hence, the human wealth depends on the time horizon for various values of the expected income growth rate \(\alpha\) ranging from 1% to 6%. The human wealth naturally increases significantly with the expected growth rate. The left panel is for the case with an income-asset correlation of +1, while the right panel is for an income-asset correlation of −1. For young individuals with a long time horizon, the income multiplier is in all cases very large. For an individual with a 40-year income ahead with an expected annual growth rate of 4%, the human wealth is 33 times his current annual income if the correlation is +1 and 128 times current income if the correlation is −1! Clearly, the human wealth will dominate financial wealth for many young individuals. The income stream is more valuable if it is negatively correlated with the stock market than if it is positively correlated. The income is like the dividends from a traded asset and from the basic CAPM we know that assets that are positively correlated with the overall stock market have a high required expected return and a low present value. It follows from (13.3) that human wealth is decreasing in the term \(\xi\top \lambda\), which in the one asset framework is equal to \(\rho|\xi|\lambda\), where the correlation \(\rho\) is either −1 or +1. Consequently, the human wealth will be increasing in the income volatility \(|\xi|\) if the income-asset correlation is negative.

We have from Theorem 6.2 that without labor income it is optimal for a CRRA utility investor to invest the proportions \(\pi_t = \frac{1}{\gamma} (\xi\top)^{-1} \lambda\) or, equivalently, the amounts \(\theta_t = \frac{W_t}{\gamma} (\xi\top)^{-1} \lambda\) in the
The figures show how the present value of future income per unit of current income, i.e., \( M(t) \), depends on the time horizon with either perfectly positive or perfectly negative correlation between the income rate and the asset price. The lowest curve is for \( \alpha = 1\% \), the one just above is for \( \alpha = 2\% \), etc., so that the top curve is for \( \alpha = 6\% \).

With the optimal investment strategy the wealth will evolve as

\[
dW_t = \dot{\theta}_t^\top \sigma dz_t,
\]

cf. (6.13). An investor with labor income has a total wealth of \( W_t + H(y_t,t) \). We conjecture that the investor will seek to invest such that the dynamics of total wealth is

\[
d(W_t + H(y_t,t)) = \dot{\theta}_t^\top \sigma dz_t.
\]

By Itô’s Lemma, the dynamics of human wealth is

\[
dH(y_t,t) = \dot{\theta}_t^\top \sigma dz_t.
\]

So the dynamics of the optimally invested financial wealth must be given by

\[
dW_t = \dot{\theta}_t^\top \sigma dz_t - H(y_t,t) y_t \dot{\theta}_t^\top \sigma dz_t
\]

This is the case for an investment strategy \( \theta_t \) that satisfies

\[
\dot{\theta}_t^\top \sigma = \left[ (W_t + H(y_t,t)) \frac{1}{\gamma} \lambda - H_y(y_t,t) y_t \frac{1}{\gamma} \xi \right]^\top,
\]

i.e., the optimal amounts invested in the risky financial assets are given by the vector \( \theta_t = \Theta(W_t, y_t, t) \), where

\[
\Theta(W, y, t) = \frac{1}{\gamma} \left( W + H(y, t) \right) \frac{1}{\gamma} \lambda - H_y(y, t) y \left( \frac{1}{\gamma} \xi \right)^{-1} \xi.
\]

Since \( H_y(y, t) = \frac{1}{\gamma} H(y, t) \) under our assumptions, we can rewrite the optimal investment strategy as

\[
\Theta(W, y, t) = \frac{1}{\gamma} W \frac{1}{\gamma} \lambda + H(y, t) \left( \frac{1}{\gamma} \xi \right)^{-1} \left( \frac{1}{\gamma} \lambda - \xi \right). \tag{13.4}
\]
The first term is identical to the optimal investment without labor income so that the second term represents the effect of labor income on the optimal investment strategy. The indirect utility function of the investor with constant relative risk aversion $\gamma$ is

$$ J(W,y,t) = \frac{1}{1-\gamma} g(t)^\gamma (W + H(y,t))^{1-\gamma}, \quad (13.5) $$

where, exactly as in Chapter 6, $g(t)$ is given by

$$ g(t) = \frac{1}{A} \left( \varepsilon_1^{1/\gamma} + \left[ \varepsilon_2^{1/\gamma} A - \varepsilon_1^{1/\gamma} \right] e^{-A(T-t)} \right) \tag{13.6} $$

with

$$ A = \frac{\delta - r(1-\gamma)}{2} \frac{1 - \gamma^2 \|\lambda\|^2}{2}. $$

The optimal consumption rate is $c^*_t = C(W_t,y_t,t)$, where

$$ C(W,y,t) = \varepsilon_1^{1/\gamma} \frac{W + H(y,t)}{g(t)} = \frac{A}{1 + (\varepsilon_2/\varepsilon_1)^{1/\gamma} A - 1} e^{-A(T-t)} (W + H(y,t)). $$

Let us outline how to verify these findings. The indirect utility function is defined as

$$ J(W,y,t) = \sup_{\{c_t, \theta_t\}_{t \in [t,T]}} \mathbb{E}_{W,y,t} \left[ \varepsilon_1 \int_t^T e^{-\delta(s-t)} u(c_s) ds + \varepsilon_2 e^{-\delta(T-t)} u(W_T) \right], $$

where $u(c) = c^{1-\gamma}/(1-\gamma)$. Given the dynamics of wealth and income, the associated HJB equation is

$$ \delta J = \sup_{c, \theta} \left\{ \varepsilon_1^{1-\gamma} J_W^{1-\gamma} + \frac{\partial J}{\partial t} + J_W (rW + \theta^\top \sigma \lambda + y - c) + \frac{1}{2} J_{WW} \theta^\top \sigma \sigma^\top \theta ight. $$

$$ + J_{yy} \gamma \frac{1}{2} J_{yy} y^2 \|\xi\|^2 + J_{wy} \theta^\top \sigma \xi \right\} $$

with the terminal condition $J(W,y,T) = \varepsilon_2 W^{1-\gamma}/(1-\gamma)$. The first-order conditions for the optimal controls imply that

$$ c = \varepsilon_1^{1/\gamma} J_W^{1-\gamma}, \quad \theta = -\frac{J_W}{J_{WW}} (\sigma^\top)^{-1} \lambda - \frac{y J_{Wy}}{J_{WW}} (\sigma^\top)^{-1} \xi. \tag{13.7} $$

Substituting these relations back into the HJB equation and removing the sup-operator, we arrive after some simplifications at the PDE

$$ \delta J = \varepsilon_1^{1/\gamma} \frac{\gamma}{1-\gamma} J_W^{1-\gamma} + \frac{\partial J}{\partial t} + J_W J_W - \frac{1}{2} J_{WW} \|\lambda\|^2 $$

$$ + y J_W - \frac{1}{2} J_{WW} y^2 \|\xi\|^2 - J_{Wy} \frac{1}{2} J_{WW} y^2 \xi^\top \lambda + J_{yy} \gamma + \frac{1}{2} J_{yy} y^2 \|\xi\|^2, $$

which extends (6.9) to the case with labor income. Then substitute in $J$ from (13.5) and the relevant partial derivatives. Rewrite the term $rW$ as $r(W + H(y,t)) - rH(y,t)$. There will now be two terms involving $(W + H)^{-\gamma-1}$, but these terms cancel. There will be a number of terms involving $g^\gamma (W + H)^{-\gamma}$, but these cancel since

$$ \frac{\partial H}{\partial t} + (\alpha - \xi^\top \lambda) y H_y + \frac{1}{2} \|\xi\|^2 y^2 H_{yy} - r H + y = 0, $$

cf. the PDE (13.2). By dividing all the remaining terms by $\varepsilon_1^{1/\gamma} g(t)^{\gamma-1} (W + H(y,t))^{1-\gamma}$, we arrive at

$$ g'(t) = \frac{1}{\gamma} \left( \delta + r(\gamma - 1) + \frac{\gamma - 1}{2\gamma} \|\lambda\|^2 \right) g(t) - \varepsilon_1^{1/\gamma}. $$
Chapter 13. Labor income

and the associated terminal condition is \( g(T) = \varepsilon_2^{1/\gamma} \). This is equivalent to the ODE (6.10) in the case without labor income. Hence, the solution for \( g(t) \) is the same and therefore given by (13.6). The expressions for the optimal strategies follow from substituting (13.5) into the first-order conditions (13.7).

Note that we have expressed the investment strategy in terms of the amounts invested rather than in terms of portfolio weights. The reason is that portfolio weights are not suitable for the case where financial wealth does not stay strictly positive under all circumstances. The portfolio weights are undefined when financial wealth is zero and hard to relate to when financial wealth is negative. In the present case we can only be sure that the sum of financial wealth and human wealth stays positive, but the financial wealth by itself may very well be negative. For example, if you initially have no financial wealth but a sure future labor income, you will probably want to borrow funds in order to be able to consume goods right now.

If financial wealth is positive, we see from (13.4) that the optimal portfolio weights can be written as

\[
\Pi(W, y, t) = \frac{1}{\gamma} \left( \frac{\sigma}{\gamma} \right)^{-1} \lambda + \frac{H(y, t)}{W} \left( \frac{\sigma}{\gamma} \right)^{-1} \left( \frac{1}{\gamma} \lambda - \xi \right). 
\]

With a single risky asset this reduces to

\[
\Pi(W, y, t) = \frac{1}{\gamma} \left( \frac{\sigma}{\gamma} \right)^{-1} \lambda + \frac{H(y, t)}{W} \left( \frac{1}{\gamma} \lambda - \xi \right). 
\]

The human wealth is increasing in the horizon so the optimal portfolio weights will generally depend on the horizon of the investor. If \( \lambda > \gamma \xi \), we see that the optimal portfolio weight will be increasing in the horizon. In that case it is optimal for investors to decrease their fraction of financial wealth invested in the stock market as they grow older. This is consistent with popular investment advice – but not with the explanation that usually accompanies the advice, cf. the discussion in Section 6.5. And note that if \( \lambda < \gamma \xi \), we get the opposite conclusion.

Let us again consider a numerical example with a single risky asset and market parameters \( r = 2\% \), \( \sigma = 20\% \), \( \lambda = 0.3 \). The individual has constant relative risk aversion with a time preference rate of \( \delta = 3\% \) and an income process with an expected growth rate of \( \alpha = 4\% \) and a volatility of \( |\xi| = 10\% \). Table 13.3 illustrates the optimal strategies for the case with perfect positive correlation for the four combinations of a risk aversion of 2 or 10 and a time horizon of 10 or 30 years. For each combination the table shows the optimal fraction of financial wealth invested in the stock and in cash as well as the optimal consumption-to-income ratio for various values of the wealth-to-income ratio. When the wealth-to-income ratio is very high, human wealth becomes unimportant and the optimal portfolio is close to the one without any labor income, which is a 75-25 split between stocks and cash for a risk aversion of 2 and a 15-85 split for a risk aversion of 10. With lower and more reasonable wealth-to-income ratios, the optimal portfolios are very different from the no-income case. For \( \gamma = 2 \), the inequality \( \lambda > \gamma \xi \) is satisfied so that the optimal stock investment is higher than without income and increasing in the time horizon. With \( \gamma = 10 \), the inequality is reversed leading to a lower stock investment which decreases with the horizon. For low wealth-to-income ratios the optimal portfolios are in all cases extreme with either substantial borrowing or substantial short-selling of the stock.

In Table 13.4 we fix \( \gamma = 2 \) and \( T - t = 30 \) and compare the optimal strategies for an income-asset correlation of +1 (left panel) and −1 (right panel). The optimal portfolio is even more extreme
### Table 13.3: Optimal strategies with positive income-asset correlation.

The table shows how the optimal strategies vary with the wealth-to-income ratio $W/y$ for different combinations of the risk aversion coefficient $\gamma$ and the time horizon $T - t$. The numbers in the columns labeled stock and cash show the fractions of current financial wealth optimally invested in the stock and in cash (the bank account), respectively. The numbers in the column labeled $c/y$ show the optimal consumption-to-income ratio. The income is assumed to be perfectly positively correlated with the stock price.

<table>
<thead>
<tr>
<th>$W/y$</th>
<th>$\gamma = 2, T - t = 10, M = 9.52$</th>
<th>$\gamma = 2, T - t = 30, M = 25.92$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>stock</td>
<td>cash</td>
</tr>
<tr>
<td>0.2</td>
<td>12.6453</td>
<td>-11.6453</td>
</tr>
<tr>
<td>0.6</td>
<td>4.7151</td>
<td>-3.7151</td>
</tr>
<tr>
<td>1</td>
<td>3.1291</td>
<td>-2.1291</td>
</tr>
<tr>
<td>2</td>
<td>1.9395</td>
<td>-0.9395</td>
</tr>
<tr>
<td>5</td>
<td>1.2258</td>
<td>-0.2258</td>
</tr>
<tr>
<td>10</td>
<td>0.9879</td>
<td>0.0121</td>
</tr>
<tr>
<td>50</td>
<td>0.7976</td>
<td>0.2024</td>
</tr>
<tr>
<td>1000</td>
<td>0.7524</td>
<td>0.2476</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$W/y$</th>
<th>$\gamma = 10, T - t = 10, M = 9.52$</th>
<th>$\gamma = 10, T - t = 30, M = 25.92$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>stock</td>
<td>cash</td>
</tr>
<tr>
<td>0.2</td>
<td>-16.5035</td>
<td>17.5035</td>
</tr>
<tr>
<td>0.6</td>
<td>-5.4012</td>
<td>6.4012</td>
</tr>
<tr>
<td>2</td>
<td>-1.5153</td>
<td>2.5153</td>
</tr>
<tr>
<td>5</td>
<td>-0.5161</td>
<td>1.5161</td>
</tr>
<tr>
<td>10</td>
<td>-0.1831</td>
<td>1.1831</td>
</tr>
<tr>
<td>50</td>
<td>0.0834</td>
<td>0.9166</td>
</tr>
<tr>
<td>1000</td>
<td>0.1467</td>
<td>0.8533</td>
</tr>
</tbody>
</table>
positive correlation, \( M = 25.92 \) & negative correlation, \( M = 69.63 \) \\
\( W/y \) & stock & cash & \( c/y \) & stock & cash & \( c/y \) \\
0.2 & 12.6453 & -11.6453 & 1.0696 & 435.9611 & -434.9611 & 3.7494 & \\
0.6 & 4.7151 & -3.7151 & 1.1136 & 145.8204 & -144.8204 & 3.7709 & \\
1 & 3.1291 & -2.1291 & 1.1577 & 87.7922 & -86.7922 & 3.7924 & \\
2 & 1.9395 & -0.9395 & 1.2678 & 44.2711 & -43.2711 & 3.8461 & \\
5 & 1.2258 & -0.2258 & 1.5980 & 18.1584 & -17.1584 & 4.0072 & \\
10 & 0.9879 & 0.0121 & 2.1484 & 9.4542 & -8.4542 & 4.2756 & \\
50 & 0.7976 & 0.2024 & 6.5518 & 2.4908 & -1.4908 & 6.4233 & \\
1000 & 0.7524 & 0.2476 & 11.1318 & 0.8370 & 0.1630 & 57.4297 & \\

Table 13.4: Optimal strategies with negative income-asset correlation. The table shows how the optimal strategies vary with the wealth-to-income ratio \( W/y \) for a risk aversion of 2 and a time horizon of 30 years. The left [right] side of the table is for the case where the income and the stock price are perfectly positively [negatively] correlated. The numbers in the columns labeled stock and cash show the fractions of current financial wealth optimally invested in the stock and in cash (the bank account), respectively. The numbers in the column labeled \( c/y \) show the optimal consumption-to-income ratio.

with a negative correlation both due to the fact the human wealth is larger and because the hedge term is much larger. Basically, the individual can take on much more financial risk since the income process provides an implicit hedge.

It is not without loss of generality to assume a single risky asset. To obtain a spanned income with a single asset, it is clear that the income has to be perfectly correlated with the price of that asset. Perfect correlation is certainly unrealistic. If we add further risky assets, the income does not have to be perfectly correlated with any individual asset. Suppose \( n \) risky assets are traded and assume that (i) the prices of any two risky assets have the same correlation given by \( \rho_{PP} \), (ii) all the risky assets have the same volatility, and (iii) all assets have the same correlation \( \rho_{Py} \) with the income rate. Then it can be shown that the income risk is spanned if the condition

\[
\rho_{Py}^2 = \rho_{PP} + \frac{1 - \rho_{PP}}{n}
\]

is satisfied. Clearly, this is decreasing in \( n \). With many risky assets traded, the required correlation between the income process and each individual asset can be quite small.

Moreover, the labor income of a given individual may not be significantly correlated with the overall stock market, but highly correlated with a specific stock. One could imagine that the labor income of an employee of a corporation was positively correlated with the price of the company’s stocks and maybe also with stock prices of other companies in the same industry. If this is true, the labor income will to some extent replace a financial investment in these stocks. Consequently, the individual should invest less of his financial wealth in these stocks. Following this line of thought, a pension fund with members in a given industry should perhaps underinvest in the stocks of the corporations in which the members work - simply to give the members a better diversified total
position. The horror example is the case of the pension fund of Enron employees, which had 58% of the total fund invested in Enron stocks prior to the 98.8% drop in the Enron stock price in 2001. Not only did Enron employees lose their jobs, they also lost a major part of their pension savings.

### 13.3.3 Stochastic interest rates

Stochastic interest rates is the main source of shifts in the investment opportunity set, and the effect of interest rate uncertainty on the optimal strategies of an investor without labor income is by now relatively well-studied in the literature, cf. Chapter 10. In order to analyze how individuals should allocate their funds to various asset classes, e.g., cash, bonds, and stocks, it is important to combine stochastic interest rates and labor income. The relative allocation to bonds and stocks can be significantly affected by the presence of uncertain labor income for several reasons. First, bonds and stocks can be differently correlated with labor income shocks so that bonds may be better for hedging income rate shocks than stocks or vice versa. Second, risk-averse investors want to hedge total wealth against shifts in investment opportunities. When the short-term interest rate captures the investment opportunities, the appropriate asset for this hedging motive is the bond. Third, since human wealth is defined as the discounted value of the future income stream, it will in general be sensitive to the interest rate level like a bond and, hence, the income stream represents an implicit investment in a bond, so that the explicit bond investment is reduced. Moreover, the expected growth rate and variability of labor income may itself vary over the business cycle, which we can approximate by the level of interest rates. Such dependencies between income and interest rates will also affect the asset allocation decision. These issues are formalized by Munk and Sørensen (2010) to which the reader is referred.

### 13.4 Exogenous income in incomplete markets

As seen in the earlier numerical examples, the optimal strategy outlined above may involve extensive borrowing of young investors that anticipate high future income rates. In practice, investors cannot actually sell their future income stream as slavery is forbidden these days. Moreover, young investors will find it extremely difficult to borrow substantial amounts for risky stock investments putting up only anticipated future income as implicit collateral or the acquired stocks as explicit collateral. This can be explained by the moral hazard and adverse selection features of labor income. In reality the income rate is not exogenously given, but reflects the abilities and the effort of the investor.

Some models take these problems partially into account by still assuming an exogenous income process, but restricting the agent to consumption and investment strategies that have the property that financial wealth $W_t$ always stays positive. The future income stream will then have a lower value than in the unrestricted, complete market case. See Duffie and Zariphopoulou (1993), Duffie, Fleming, Soner, and Zariphopoulou (1997), Koo (1998), and Munk (2000). For example, Duffie, Fleming, Soner, and Zariphopoulou (1997) and Munk (2000) study the case with a single risky asset with price process

$$dP_t = P_t \left[ \mu dt + \sigma dz_t \right],$$
constant $r$, $\mu$, and $\sigma$, and where the income rate follows the geometric Brownian motion

$$dy_t = y_t \left[ \alpha dt + \rho \sigma_y dz_t + \sqrt{1 - \rho^2} \sigma_y d\tilde{z}_t \right].$$

Here $\rho$ is the correlation between the asset price and the labor income. The agent must keep financial wealth positive, $W_t > 0$, so that she faces a liquidity constraint. Furthermore, she faces undiversifiable income risk. The numerical results of Munk (2000) show that the implicit value the agent associates with her income stream can be considerably less than without the liquidity constraint and the undiversifiable part of the income risk, especially if she has a high preference for current consumption and a low current financial wealth. The results indicate that the reduction in human wealth is mainly due to the liquidity constraint, while the undiversifiability is of minor importance.

A few papers find closed-form solutions in settings with unspanned (undiversifiable) income risk, but have to assume negative exponential utility and a Gaussian income process. Svensson and Werner (1993) solve for the optimal consumption and portfolio strategies in an infinite time horizon setting, whereas Henderson (2005) assumes a finite horizon and utility of terminal wealth only. Henderson (2005) also finds near-explicit solutions for more general income processes. Duffie and Jackson (1990) and Teplá (2000) derive similar solutions for investors receiving an unspanned income only at the terminal date. Christensen, Larsen, and Munk (2012) derive optimal consumption and portfolio strategies with an unspanned income stream and a finite time horizon.

Explicit solutions have only been found in the following special cases involving CARA (negative exponential) utility, a normally distributed income stream, a constant risk-free rate, and a constant drift and volatility of the stock price. Svensson and Werner (1993) and Wang (2006) consider infinite time horizon settings where a transversality condition has to be imposed on the utility maximization problem. The models of Svensson and Werner (1993) and Wang (2006) differ slightly with respect to the specification of the income process. Furthermore, in the model of Wang (2006) only a risk-free asset is traded, whereas Svensson and Werner (1993) allow for risky assets. In similar settings, Wang (2004, 2009) investigates the impact of unobservable or partially observable income growth on consumption and investment decisions. Henderson (2005) assumes a finite time horizon with utility of terminal wealth only, and she also derives near-explicit solutions for more general income processes. Duffie and Jackson (1990) and Teplá (2000) derive similar solutions for investors receiving an unspanned income only at the terminal date. Christensen, Larsen, and Munk (2012) generalize Henderson’s explicit solution to the case of consumption over a finite lifetime. The transversality condition in the infinite horizon model mentioned above restricts the rate at which the debt of the investor can grow, but does not force the investor to ever pay back his debt and can thus lead to excessive borrowing compared to the more realistic finite horizon setting. In contrast, in their finite horizon model, Christensen, Larsen, and Munk ensure that the debt of the investor equals zero at the end of the horizon.

It seems difficult—if not impossible—to move beyond the assumptions of negative exponential utility and a Gaussian income process and still obtain closed-form solutions to the investor’s utility maximization problem with unspanned income risk. Several recent papers have numerically solved for optimal consumption and portfolio strategies in more general settings, e.g., Cocco, Gomes, and Maenhout (2005), Koijen, Nijman, and Werker (2010), Lynch and Tan (2011), Munk and Sørensen (2010), Van Hemert (2010), and Viceira (2001).
13.5 Endogenous labor supply and income

13.5.1 The model and the solution

Bodie, Merton, and Samuelson (1992) endogenize the labor supply decision of the agent. Let \( \omega_t \) denote the wage rate, which is assumed to follow the geometric Brownian motion

\[
d\omega_t = \omega_t [m \, dt + \nu^\top d z_t].
\]

In particular, the wage rate is spanned by the financial securities traded. Let \( \varphi_t \in [0, 1] \) denote the fraction of time working so that the total labor income over the interval \([t, t+dt]\) is \( \varphi_t \omega_t \, dt \).

Equivalently, we can let \( l_t \equiv 1 - \varphi_t \) denote the fraction of time not working and think of the agent receiving \( \omega_t \) and then paying \( l_t \omega_t \) on the “consumption good” leisure. The wage rate \( \omega_t \) is the unit price of leisure measured in units of the consumption good. Assuming a constant interest rate and a constant market price of risk, the wealth of the investor will then follow

\[
dW_t = (rW_t + \theta_t^\top \lambda - c_t - \varphi_t \omega_t) \, dt + \theta_t^\top \nu \, d z_t,
\]

where \( \beta \) is a constant between 0 and 1 determining the relative weights of consumption and leisure, and we can interpret \( \gamma > 0 \) as the coefficient of risk aversion with respect to the “composite consumption” \( c^\beta l^{1-\beta} \).

We ignore utility of terminal wealth and define the indirect utility function as

\[
J(W, \omega, t) = \sup_{(c, \theta, l)} E_{W, \omega, t} \left[ \int_t^T e^{-\delta(s-t)} \frac{1}{1-\gamma} \left[ c_s^\beta l_s^{1-\beta} \right]^{1-\gamma} \, ds \right],
\]

where the supremum is taken over all non-negative consumption strategies \( c \), all investment strategies \( \theta \), and all labor-leisure strategies \( l \) valued in \([0, 1]\).

We demonstrate below that the indirect utility function is given in closed-form by

\[
J(W, \omega, t) = \frac{1}{1-\gamma} \beta^{\gamma(1-\gamma)} (1-\beta)^{(1-\beta)(1-\gamma)} G(t)^{\gamma(1-\beta)(1-\gamma)} (W + \omega F(t))^{1-\gamma},
\]

where

\[
G(t) = \frac{1}{k} \left( 1 - e^{-k(T-t)} \right),
\]

\[
F(t) = \frac{1}{r - m + \nu^\top \lambda} \left[ \frac{1}{1-\gamma} (1-\beta)^{(1-\gamma)} \frac{1}{\gamma} \frac{1}{2\gamma^2} \lambda^\top \lambda + \frac{1}{\gamma} (1-\beta) \left[ m + \frac{1}{\gamma} \frac{1}{\gamma} \nu^\top \lambda - \frac{1}{2\gamma} (1-\beta(1-\gamma)) \nu^\top \nu \right] \right],
\]

where

\[
k = \frac{\delta}{r - \frac{1}{\gamma}} - \frac{1}{\gamma} - \frac{1}{2\gamma^2} \lambda^\top \lambda + \frac{1}{\gamma} (1-\beta) \left[ m + \frac{1}{\gamma} \frac{1}{\gamma} \nu^\top \lambda - \frac{1}{2\gamma} (1-\beta(1-\gamma)) \nu^\top \nu \right].
\]
The optimal strategies are \( c_t^* = C(W_t, \omega_t, t), \ l_t^* = L(W_t, \omega_t, t), \) and \( \theta_t^* = \Theta(W_t, \omega_t, t), \) where

\[
C(W_t, \omega_t, t) = \frac{\beta}{G(t)} (W + \omega F(t)), \tag{13.10}
\]

\[
L(W_t, \omega_t, t) = \frac{1 - \beta}{\omega} \frac{W + \omega F(t)}{G(t)}, \tag{13.11}
\]

\[
\Theta(W_t, \omega_t, t) = \frac{1}{\gamma} \left( W + \omega F(t) \right) \left( \frac{\sigma}{\gamma} \right)^{-1} \lambda - F(t) \omega \left( \frac{\sigma}{\gamma} \right)^{-1} v
- \frac{(1 - \beta)(1 - \gamma)}{\gamma} \left( W + \omega F(t) \right) \left( \frac{\sigma}{\gamma} \right)^{-1} v. \tag{13.12}
\]

Here \( \omega_t F(t) \) denotes the time \( t \) value of the maximum labor income that the agent can receive. To see this note that the future wage rate is

\[
\omega_s = \omega_t \exp \left\{ \left( m - \frac{1}{2} v^\top v \right) (s - t) + v^\top (z_s - z_t) \right\}.
\]

Working at a maximum rate, \( \varphi_s \equiv 1 \) for all \( s \in [t, T] \), the time \( t \) value of future labor income is

\[
E_t \left[ \int_t^T \exp \left\{ -r(s - t) - \lambda^\top [z_s - z_t] - \frac{1}{2} \| \lambda \|^2 (s - t) \right\} \omega_s \, ds \right]
= \omega_t \int_t^T E_t \left[ \exp \left\{ \left( m - r - \frac{1}{2} \| \lambda \|^2 - \frac{1}{2} \| v \|^2 \right) (s - t) + (v - \lambda)^\top (z_s - z_t) \right\} \right] \, ds
= \omega_t \int_t^T e^{(m - r - v^\top \lambda) (s - t)} \, ds
= \omega_t F(t).
\]

This solution is only valid if the leisure strategy \( L(W_t, \omega, t) \) stated above is always valued in \([0, 1]\), which is not necessarily the case.

### 13.5.2 Verifying the solution

Again we attack the problem by solving the associated HJB-equation:

\[
\delta J = \sup_{c, \theta, l} \left\{ \frac{1}{1 - \gamma} c^{\beta(1 - \gamma)} (1 - \beta)^{(1 - \gamma)} + \frac{\partial J}{\partial t} + J \omega m + \frac{1}{2} J \omega \omega \| v \|^2
+ J_W (rW + \theta^\top \sigma \lambda + \omega - c - \omega l) + \frac{1}{2} J_{WW} \theta^\top \sigma \sigma^\top \theta + J_{W, \omega} \omega \theta^\top \sigma v \right\}.
\]

The first-order conditions for \( c \) and \( l \) are

\[
\beta e^{\beta(1 - \gamma)} (1 - \beta)^{(1 - \gamma)} = J_W,
(1 - \beta) e^{\beta(1 - \gamma)} (1 - \beta)^{(1 - \gamma)} = \omega J_W,
\]

which imply the simple relation

\[
c = \frac{\beta}{1 - \beta} \omega l
\]

between the optimal consumption rate and the optimal leisure rate. This relation ensures that the ratio of (i) the marginal utility with respect to leisure and (ii) the marginal utility with respect to consumption will equal the relative price \( \omega \). Solving the two first-order conditions, we find that

\[
c = \beta^{1 + \beta(1 - \gamma)} (1 - \beta)^{(1 - \beta)(1 - \gamma)} \omega^{-\frac{(1 - \beta)(1 - \gamma)}{\gamma}} J_W^{-1/\gamma},
l = \beta^{\beta(1 - \gamma)} (1 - \beta)^{-\frac{1 - \beta(1 - \gamma)}{\gamma}} \omega^{-\frac{1 - \beta(1 - \gamma)}{\gamma}} J_W^{-1/\gamma}.
\]
Inserting the derivative of the candidate indirect utility function \((13.8)\), these expressions will give \((13.10)\) and \((13.11)\). The first-order condition with respect to \(\theta\) implies that

\[
\theta = -\frac{J_W}{J_{WW}} (\sigma^\top)^{-1} \lambda - \frac{J_{W\omega}}{J_{WW}} \omega (\sigma^\top)^{-1} v.
\]

Applying our candidate for \(J\), we have

\[
-\frac{J_W}{J_{WW}} = \frac{1}{\gamma} (W + \omega F(t)), \quad \frac{J_{W\omega}}{J_{WW}} = F(t) + \frac{(1 - \beta)(1 - \gamma)}{\gamma} \frac{W + \omega F(t)}{\omega},
\]

and we obtain \((13.12)\).

Substituting the maximizing values for \(c\), \(l\), and \(\theta\) into the HJB-equation and deleting the superoperator, we arrive after some simplifications at the PDE

\[
\delta J = \frac{\gamma}{1 - \gamma} \beta^{1 - \gamma} (1 - \beta)^{(1 - \beta)(1 - \gamma)} \omega^{-(1 - \beta)(1 - \gamma)} J_W^{1 - \frac{1}{\gamma}} + \frac{\partial J}{\partial t} + \int_{\mathcal{D}} \left[ J_{WW} \omega^2 \|v\|^2 + r(W + \omega F(t))J_W - r\omega F(t)J_W + \omega J_W \right] ds
\]

\[
- \frac{1}{2} \frac{J_W^2}{J_{WW}} \|\lambda\|^2 - \frac{1}{2} \frac{J_{W\omega}}{J_{WW}} \omega^2 \|v\|^2 - \frac{J_{W\omega}}{J_{WW}} \omega^2 v^\top \lambda.
\]

It remains to verify that our candidate \((13.8)\) satisfies this PDE. Substituting the relevant derivatives into the PDE, we get a lot of terms. They are all involving \((1 - \beta)(1 - \gamma)\) with the exception of \(\|v\|^2\) in which case the power \(1 - \gamma\) and the power \(-\gamma\) cancel. Next note that the terms with the power \(-\gamma\) cancel due to the fact that the function \(F(t)\) satisfies the ordinary differential equation \(F'(t) - (\gamma - m + v^\top \lambda)F(t) + 1 = 0\). Then only the terms involving \((W + \omega F(t))(1 - \gamma)\) are left. Dividing through by \(\frac{\gamma}{1 - \gamma} \beta^{1 - \gamma}(1 - \beta)^{(1 - \beta)(1 - \gamma)}\) we end up with the equation \(G'(t) = kG(t) - 1\) which with the terminal condition \(G(T) = 0\) has the solution stated in \((13.9)\).

### 13.5.3 Inflexible labor supply

To study the effect of labor supply flexibility on optimal investments let us look at an agent who once and for all fixes a constant labor supply rate \(\bar{\varphi} \equiv 1 - l\). For a given supply \(\bar{\varphi}\), the agent finds the optimal consumption and investment strategies by solving the optimization problem

\[
J(W, \omega, t; \bar{\varphi}) = \sup_{(c, \theta)} \mathbb{E}_{W, \omega, t} \left[ \int_t^T e^{-\delta(s-t)} \frac{1}{1 - \gamma} \left[ c_s^\beta (1 - \bar{\varphi})^{1-\beta} \right]^{1-\gamma} ds \right]
\]

\[
= \beta (1 - \bar{\varphi})^{(1-\beta)(1-\gamma)} \sup_{(c, \theta)} \mathbb{E}_{W, \omega, t} \left[ \int_t^T e^{-\delta(s-t)} \frac{1}{\beta(1 - \gamma)} c_s^{(1-\gamma)} ds \right].
\]

The supremum in the last expression equals the indirect utility of an investor with a constant relative risk aversion of \(1 - \beta(1 - \gamma)\) and an exogenously given labor income at the rate \(y_t = \bar{\varphi} \omega_t\). Clearly the present value of future labor income will be \(H(y_t, t) = \bar{\varphi} \omega_t F(t)\), where \(F(t)\) is given above. Using the previously derived results for the case with exogenous income, we get

\[
J(W, \omega, t; \bar{\varphi}) = \frac{1}{1 - \gamma} (1 - \bar{\varphi})^{(1-\beta)(1-\gamma)} g(t)^{1-\beta(1-\gamma)} (W + \bar{\varphi} \omega F(t))^{\beta(1-\gamma)},
\]

where \(g(t)\) is given by \((13.6)\). The optimal investment strategy for a given \(\bar{\varphi}\) is given by

\[
\Theta(W, \omega, t; \bar{\varphi}) = \frac{1}{1 - \beta(1 - \gamma)} (W + \bar{\varphi} \omega F(t)) \left(\sigma^\top\right)^{-1} \lambda - \bar{\varphi} \omega F(t) \left(\sigma^\top\right)^{-1} v.
\]
The optimal value of $\bar{\phi}$ is found by maximizing $J(W_0, \omega_0, 0; \bar{\phi})$ and turns out to be

$$\bar{\phi} = \beta - \frac{(1 - \beta)W_0}{\omega_0 F(0)}.$$  

13.5.4 Comparison of results

For easy comparison let us assume a deterministic wage rate, $v \equiv 0$. Then the optimal investment strategy of the agent with flexible labor supply is

$$\Theta(W, \omega, t) = \frac{1}{\gamma} (W + \omega F(t)) (\sigma^\top)^{-1} \lambda,$$

while the optimal investment strategy of the agent with fixed labor supply at a rate $\bar{\phi}$ is

$$\Theta(W, \omega, t; \bar{\phi}) = \frac{1}{1 - \beta(1 - \gamma)} (W + \bar{\phi}\omega F(t)) (\sigma^\top)^{-1} \lambda.$$

First note that the amounts invested in any given asset by each of the two agents have the same sign; if one agent is long [short] a given asset so is the other agent. There are two differences between these two expressions: the relevant risk aversion coefficient and the valuation of future income. With flexible supply the labor income enters as the maximum value of future wages, which can only be obtained by working all the time. On the other hand, the total risk aversion $\gamma$ is relevant for the flexible supplier instead of the consumption risk aversion $1 - \beta(1 - \gamma)$ relevant for the fixed supplier.

Let us consider assets with positive amounts invested. If $\gamma < 1$, then $\gamma < 1 - \beta(1 - \gamma)$, and hence the flexible supplier will unambiguously invest more in the risky assets. If $\gamma$ is sufficiently larger than 1, the relation between the amounts invested is ambiguous and will depend on the exact parameter values, the remaining life-time, and the fixed labor supply rate. For moderately risk-averse investors at an early stage in their working life, the financial investments of the flexible labor supplier tend to be more risky than those of the fixed labor supplier. The intuition is that investors incurring losses on their financial investments may compensate by working harder and drive up labor income. Labor supply flexibility serves as a kind of insurance. Changes of labor supply have the largest effect on capitalized labor income for young investors. The flexibility of labor supply may therefore amplify the horizon effect of labor income on risky investments which is present already for an exogenously given labor income stream. With an uncertain wage rate spanned by the risky financial assets, this conclusion seems to hold as long as the wage rate is not “too risky”, cf. the discussion in Bodie, Merton, and Samuelson (1992). Apparently, the effects of labor supply flexibility have not been studied in the more reasonable incomplete market setting, where the wage rate is not fully diversifiable.

13.6 More

The purchase of a house serves a dual role by both generating consumption services and by constituting an investment affecting future wealth and consumption opportunities.

Several recent papers include housing in life-cycle decision problems. Campbell and Cocco (2003) study the mortgage choice in a life-cycle framework with stochastic house price, labor income, and interest rates. They do not allow housing investment to differ from housing consumption and, furthermore, fix the house size (the number of housing units), so they cannot address the interaction between housing decisions and portfolio decisions. Cocco (2005) considers a model in which house prices and aggregate income shocks are perfectly correlated. Also in his model housing consumption and housing investment cannot be disentangled, as renting is not possible and there are no house price linked financial assets traded. The individual can only enjoy the consumption benefits of a home by buying a house and is thus forced into home ownership. Since there is a minimum choice of house size, a young individual has to tie up a large share of wealth in real estate and will invest little in stocks (also because of borrowing constraints and an imposed stock market entry cost). Cocco concludes that house price risk crowds out stock holdings and can therefore help in explaining limited stock market participation does not carry over to our setting.

Yao and Zhang (2005a) generalize Cocco’s setting to an imperfect correlation between income and house prices, and they show that there are substantial welfare gains from allowing renting and that the renting/owning decision changes the optimal investment strategy. In their model, the individual would prefer owning a house to renting, but cannot always do so because of constraints (e.g., a down payment is required to buy a house). If the individual decides to rent a house of a given size, that will be equal to his housing consumption and he will have zero wealth exposure to house price risk. If the individual decides to own a house, the size of the house determines his housing consumption and is identical to his housing investment position. Yao and Zhang (2005a)
find that home-owners invest less in stocks than home-renters.

Van Hemert (2010) generalizes the setting further by allowing for stochastic variations in interest rates and thereby introducing a role for bonds, and his focus is on the interest rate exposure and choice of mortgage over the life-cycle. Kraft and Munk (2011) disconnect the housing consumption and investment positions further, as the individual can simultaneously rent and own, and his investment position can be higher or lower than the housing consumption by renting out part of the owned property or by investing in house price linked financial contracts. In the other models, the housing investment position is closely linked to the demand for housing consumption, and that level of housing investment will affect the investments in the other risky assets to obtain the best overall level of risk-taking and exposure to different risks. In the Kraft-Munk setting, the housing investment is more freely determined and, hence, does not have similar repercussions for the stock and bond demand. Their results indicate that access to well-functioning markets for financial assets linked to house price will lead to welfare gains that are non-negligible, although of a moderate magnitude. In related work, de Jong, Driessen, and Van Hemert (2008) conclude that the welfare gains from having access to housing futures are small, but their model ignores labor income risk, does not allow for renting, fixes the housing investment, and assumes utility only from terminal wealth. Other papers addressing various aspects of housing in individual decision making include Sinai and Souleles (2005), Li and Yao (2007), Cauley, Pavlov, and Schwartz (2007), and Corradin, Fillat, and Vergara-Alert (2010).1

Most of the papers listed above impose various realistic constraints on the investment decisions of the individual and/or allow labor income to have an unspanned risk component. Therefore, they solve the decision problems by numerical dynamic programming with a coarse discretization of time and the state space (Van Hemert (2010) is able to handle a finer discretization by relying on 60 parallel computers). This computational procedure is highly time-consuming and cumbersome, and little is known about the precision of the numerical results. Kraft and Munk (2011) derive closed-form solutions that are much easier to analyze, interpret, and implement and thus facilitate an understanding and a quantification of the economic forces at play. On the other hand, Kraft and Munk (2011) must disregard unspanned components of income risk, housing transaction costs, borrowing constraints, short-sales constraint, etc.

1Papers on the impact of housing decisions and prices on financial asset prices include Piazzesi, Schneider, and Tuzel (2007), Lustig and van Nieuwerburgh (2005), and Yogo (2006).
CHAPTER 15

Other variations of the problem...

15.1 Multiple and/or durable consumption goods

References: Several perishable: Breeden (1979), Wachter and Yogo (2010)

15.2 Uncertain time of death; insurance

International asset allocation

TO COME...

17.1 Preferences with habit formation

It has long been recognized by economists that preferences may not be intertemporally separable. According to Browning (1991), this idea dates back to the 1890 book “Principles of Economics” by Alfred Marshall. See Browning’s paper for further references to the critique on intertemporally separable preferences. In particular, the utility associated with the choice of consumption at a given date may depend on past choices of consumption. This is modeled by replacing \( u(c_t, t) \) by \( u(c_t, h_t, t) \), where \( u \) is decreasing in \( h_t \), which is a measure of the standard of living or the habit level of consumption, e.g., a weighted average of past consumption rates:

\[
h_t = h_0 e^{-\beta t} + \alpha \int_0^t e^{-\beta (t-s)} c_s \, ds,
\]

where \( h_0, \alpha, \) and \( \beta \) are non-negative constants. High past consumption generates a desire for high current consumption, so that preferences display intertemporal complementarity. As additional motivation for such preferences, note that several papers have documented the importance of allowing for habit formation in utilities when it comes to equilibrium asset pricing. Empirical facts that seem puzzling relative to models with a representative agent having time-separable utility can be resolved by introducing habit formation into the utility function. For example, Constantinides (1990) and Sundaresan (1989) demonstrate that models with habit formation can obtain a high equity premium with low risk aversion. Campbell and Cochrane (1999) and Wachter (2006) construct representative agent models with habit formation that are consistent with observed variations in expected returns on stocks and bonds over time. Detemple and Zapatero (1991) also study asset pricing implications of habit formation preferences.\(^1\)

Sundaresan (1989), Constantinides (1990), and Ingersoll (1992) all derive the optimal strategies for an investor with an infinite time horizon under the assumption of a constant investment

\(^1\)Both Campbell and Cochrane (1999) and Wachter (2006) consider utility with external habit formation in the sense that the agent does not take into account the effect that the choice of current consumption has on future habit levels. In the other papers referred to, these effects are considered.
opportunity set. In addition, Ingersoll (1992) considers a finite-horizon investor with log utility.

Detemple and Zapatero (1992) derive conditions under which optimal policies exist for an investor with habit persistence in preferences. They are able to characterize the optimal consumption strategy in a general setting, but, except for the case of deterministic investment opportunities, they state the optimal portfolio in terms of an unknown stochastic process that comes out of the martingale representation theorem. Detemple and Karatzas (2003) provide a similar analysis for a preference structure that also involves habit formation but is more general in several respects.

Schroder and Skiadas (2002) show that the general decision problem of an investor with habit persistence in preferences who can trade in a given financial market is equivalent to the decision problem of an investor who does not exhibit habit formation, and who can trade in a financial market with more complex dynamics of investment opportunities.

Munk (2008) gives a precise characterization of the optimal portfolio in a general complete market setting and derive explicit results in concrete settings with stochastic investment opportunities. The assumed objective is

$$J_t = \sup_{(c, \pi) \in A(t)} \mathbb{E}_t \left[ \int_t^T e^{-\delta(t-s)} u(c_s, h_s) \, ds \right],$$

where $A(t)$ denotes the set of feasible consumption and portfolio strategies over the period $[t, T]$, and the “instantaneous” utility function $u(c, h)$ is assumed to be power-linear,

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

where the constant $\gamma > 0$ is a risk aversion parameter. With this specification the consumption rate is required to exceed the habit level, so that the habit level plays the role of a minimum or subsistence consumption rate determined by past consumption rates. Let us briefly summarize the main findings of that paper without going into the modeling details:

**Mean-reverting stock returns.** Stock returns are assumed to be predictable in the sense that the market price of risk follows a mean-reverting process. Interest rates are assumed constant. Under the assumption of perfect negative correlation between the stock price and the market price of risk, Munk finds an explicit solution for the optimal strategies. This is a generalization of the results of Wachter (2002), cf. Chapter 11, who assumes time-separable utility. The optimal fraction of wealth invested in stocks is the sum of a myopic demand and a (positive) hedge demand. Habit persistence has different effects on these two components, but in our numerical examples the differences are very small. It is argued that, contrary to the case of time-additive utility, the optimal fraction of wealth invested in stocks is not necessarily monotonically decreasing over the life of an investor with habit persistence in preferences for consumption. Finally, relative to the case of constant expected returns, mean reverting returns support a higher consumption rate, but in the numerical examples the increase is considerably smaller for investors with habit persistence than investors without.

**Stochastic interest rates.** The short-term interest rate is assumed to follow a square-root process as suggested by Cox, Ingersoll, and Ross (1985) with the market prices of risk being fully determined by the interest rate level. The assets available for investment are a stock (index), cash (i.e., the bank account), and a single bond (without loss of generality). While the optimal stock
portfolio weight can be found in closed form, the optimal allocation to the bond and cash as well as the optimal consumption rate involve a time and interest rate dependent function which is the solution to a relatively simple partial differential equation (PDE). With time-additive preferences the PDE has an explicit solution, cf. Section 10.3, but with habit preferences the PDE must be solved numerically. The bond portfolio weight has all three components identified in the general model: a myopic term, a hedge term, and a term ensuring that the future consumption at least reaches the habit level. The stock portfolio weight, on the other hand, has only the myopic component. The numerical experiments shown in the paper verify that habit formation have very different effects on stock and bond investments and show that the effects on consumption are ambiguous.

**Labor income.** The agent is assumed to receive a continuous stream of labor income. The income stream has two effects. Firstly, the initial wealth is to be increased by the present value of the future income stream, which implies that a larger fraction of financial wealth is to be invested in the risky assets. Habit persistence in preferences dampens this effect. Secondly, a labor income stream is implicitly equivalent to a stream of returns on a financial portfolio, so the explicit investment strategy must be adjusted accordingly. This adjustment is independent of the preference parameters and, hence, unaffected by habit persistence. Except for extreme habit persistence and very low present value of income (relative to financial wealth), the effects of labor income seem to dominate the effects of habit persistence.

In sum, habit persistence dampens the speculative investments of investors due to the fact that some funds must be reserved for the purpose of ensuring that consumption in the future will meet the habit level. The hedge investments may be affected differently by habit persistence, but in the numerical examples given by Munk (2008) the differences are small. The main effect on the relative allocations to different assets stems from the fact that some assets (bonds and cash) are better investment objects than others (stocks) when it comes to ensuring that future consumption will not fall below the habit level.

**Further references:** Hindy, Huang, and Zhu (1997)

### 17.2 Recursive utility

Schroder and Skiadas (1999) and Campbell and Viceira (1999, 2001) study consumption and portfolio decisions with so-called recursive utility or stochastic differential utility... See also Bhamra and Uppal (2006) and Cocco, Gomes, and Maenhout (2005).

Assume a single consumption good. We use a stochastic differential utility or recursive utility specification for the preferences of the individual so that the utility index $V_{t}^{c, \pi}$ associated at time $t$ with a given consumption process $c$ and portfolio process $\pi$ over the remaining lifetime $[t, T]$ is recursively given by

$$V_{t}^{c, \pi} = E_t \left[ \int_t^T f (c_u, V_u^{c, \pi}) \, du + V_T^{c, \pi} \right].$$  

(17.1)
We assume that the so-called normalized aggregator $f$ is defined by

$$f(c, V) = \begin{cases} \frac{\delta}{1 - \gamma / \psi} c^{1 - 1/\psi} (1 - \gamma |V|)^{1 - 1/\theta} - \delta \theta V, & \text{for } \psi \neq 1 \\ \delta (1 - \gamma) V \ln c - \delta V \ln (1 - \gamma |V|), & \text{for } \psi = 1 \end{cases}$$

where $\theta = (1 - \gamma)/(1 - \gamma / \psi)$. The preferences are characterized by the three parameters $\delta, \gamma, \psi$. It is well-known that $\delta$ is a time preference parameter, $\gamma > 1$ reflects the degree of relative risk aversion towards atemporal bets (on the composite consumption level $z$ in our case), and $\psi > 0$ reflects the elasticity of intertemporal substitution (EIS) towards deterministic consumption plans. The term $\bar{V}_c, \pi_T$ represents terminal utility and we assume that $\bar{V}_c, \pi_T = \alpha_1 - \gamma (W_c, \pi_T)^{1 - \gamma}$, where $\alpha \geq 0$ and $W_c, \pi_T$ is the terminal wealth induced by the strategies $c, \pi$. The special case where $\psi = 1/\gamma$ (so that $\theta = 1$) corresponds to the classic time-additive power utility. More precisely, with $\psi = 1/\gamma$ the recursion (17.1) is satisfied by

$$V_{t}^{c, \pi} = \delta \left( E_t \left[ \int_t^T e^{-\delta (u-t)} \frac{1}{1 - \gamma} c^{1 - \gamma} du + \frac{1}{\delta} e^{-\delta (T-t)} \frac{\alpha}{1 - \gamma} (W_T^{c, \pi})^{1 - \gamma} \right] \right),$$

which is a positive multiple of the traditional time-additive power utility specification. Note that $\alpha = \delta$ would correspond to the case where utility of a terminal wealth of $W$ will count roughly as much as the utility of consuming $W$ over the final year.

The above utility specification is the continuous-time analogue of the Kreps-Porteus-Epstein-Zin recursive utility defined in a discrete-time setting. Such utility specifications and their properties have been discussed at a general level by, e.g., Kreps and Porteus (1978), Epstein and Zin (1989), Duffie and Lions (1992), Duffie and Epstein (1992), Skiadas (1998), Schroder and Skiadas (1999), and Kraft and Seifried (2010). Both the discrete-time and the continuous-time versions have been applied in a few recent studies of utility maximization problems involving a single consumption good, cf. Campbell and Viceira (1999), Campbell, Cocco, Gomes, Maenhout, and Viceira (2001), and Chacko and Viceira (2005), and has also been applied in a two-good setting by Yao and Zhang (2005b). Bhamra and Uppal (2006) provide a detailed analysis of the effects of the relative risk aversion and the elasticity of intertemporal substitution parameters on the optimal portfolios in a two-period model with stochastic interest rates.

### 17.2.1 Solution via dynamic programming

Let $A_t$ denote the set of admissible control processes $(c, \pi)$ over the remaining lifetime $[t, T]$. Constraints on the controls are reflected by $A_t$. At any point in time $t < T$, the individual maximizes $V_{t}^{c, \pi}$ over all admissible control processes given the values of the state variables at time $t$. The indirect utility is defined as

$$J_t = \sup_{(c, \pi) \in A_t} V_{t}^{c, \pi}.$$ 

Duffie and Epstein (1992) have demonstrated the validity of the dynamic programming solution technique in the case of stochastic differential utility. For simplicity, we assume that the individual

---

2It is also possible to define a normalized aggregator for $\gamma = 1$ and for $0 < \gamma < 1$ but we focus on the empirically more reasonable case of $\gamma > 1$. 

---
does not receive any income from non-financial sources. Suppose the relevant information for the
decision problem is captured by wealth $W_t$ with dynamics
\[
dW_t = \left(W_t \left[ r(x_t) + \pi_t \sigma(x_t, t) \lambda(x_t) \right] - c_t \right) \, dt + W_t \pi_t \sigma(x_t, t) \, dW_t,
\]
and a one-dimensional Markov process $x = (x_t)$ so that $J_t = J(W_t, x_t, t)$ and the dynamics of $x$
has the form
\[
dx_t = m(x_t) \, dt + \nu(x_t) \, dW_t + \psi(x_t) \, dN_t.
\]
Then the Hamilton-Jacobi-Bellman (HJB) equation to solve is
\[
0 = \sup_{c \geq 0, \pi \in \mathbb{R}^d} \left\{ f(c, J(W, x, t)) + \frac{\partial J}{\partial t}(W, x, t) + J_W(W, x, t) \left( W \left[ r(x) + \pi \sigma(x, t) \lambda(x) \right] - c \right) \right.
\]
\[
+ \frac{1}{2} J_{WW}(W, x, t) W^2 \pi \sigma(x, t) \pi \sigma(x, t)^T \pi + J_x(W, x, t)m(x)
\]
\[
+ \frac{1}{2} J_{xx}(W, x, t) (\nu(x))^T \nu(x) + \psi(x)^2 \right\} + J_{Wx}(W, x, t) W \pi \sigma(x, t) \nu(x)
\]
with the terminal condition $J(W, x, T) = \frac{\alpha}{\gamma} W^{1-\gamma}$. We rewrite the HJB-equation as
\[
0 = \mathcal{L} \pi J(W, x, t) + \sup_{c \geq 0} \left\{ f(c, J(W, x, t)) - cJ_W(W, x, t) \right\} \frac{\partial J}{\partial t}(W, x, t)
\]
\[
+ J_W(W, x, t) W r(x) + J_x(W, x, t)m(x) + \frac{1}{2} J_{xx}(W, x, t) (\nu(x))^T \nu(x) + \psi(x)^2,
\]
where
\[
\mathcal{L} \pi J(W, x, t) = \sup_{\pi \in \mathbb{R}^d} \left\{ J_W(W, x, t) W \pi \sigma(x, t) \lambda(x) + \frac{1}{2} J_{WW}(W, x, t) W^2 \pi \sigma(x, t) \sigma(x, t)^T \pi \right.
\]
\[
+ J_{Wx}(W, x, t) W \pi \sigma(x, t) \nu(x) \right\}.
\]

The maximization with respect to $\pi$ is exactly as for the case with general time-additive expected
utility in Section 7.2.1. The maximizer is
\[
\pi = - \frac{J_W(W, x, t)}{W J_{WW}(W, x, t)} (\sigma(x, t)^T)^{-1} \lambda(x) - \frac{J_{Wx}(W, x, t)}{W J_{WW}(W, x, t)} (\sigma(x, t)^T)^{-1} \nu(x),
\]
which implies that
\[
\mathcal{L} \pi J(W, x, t) = - \frac{1}{2} \frac{J_W(W, x, t)^2}{J_{WW}(W, x, t)} \| \lambda(x) \|^2 - \frac{1}{2} \frac{J_{Wx}(W, x, t)^2}{J_{WW}(W, x, t)} \| \nu(x) \|^2
\]
\[
- \frac{J_W(W, x, t) J_{Wx}(W, x, t)}{J_{WW}(W, x, t)} \nu(x)^T \lambda(x).
\]
Note that the specification of the aggregator does not directly affect terms involving the portfolio
$\sigma$. Hence, the above expressions for $\pi$ and $\mathcal{L} \pi J$ are exactly as in the case with time-additive
power utility and is also the same whether $\psi = 1$ or not. The terms involving consumption will be
different from power utility and will depend on the value of $\psi$ and, therefore, the indirect utility
function solving the HJB-equation will also depend on the value of $\psi$, so we have to consider
different cases separately. Of course, when the indirect utility function is substituted into (17.4),
the optimal portfolio as a function of $W, x,$ and $t$ is also going to depend on the value of $\psi$.  

17.2.2 The case $\psi = 1$

When substituting the aggregator for $\psi = 1$ into (17.2), we can reformulate the HJB-equation as

$$
0 = \mathcal{L}^\psi J(W, x, t) + \mathcal{L}^c J(W, x, t) - \delta J(W, x, t) \ln \left(\left[1 - \gamma\right]J(W, x, t)\right) + \frac{\partial J}{\partial t}(W, x, t) + J_W(W, x, t)W_r(x) + J_x(W, x, t)m(x) + \frac{1}{2}J_{xx}(W, x, t)(v(x)^\top v(x) + \dot{v}(x)^2),
$$

where

$$
\mathcal{L}^c J(W, x, t) = \sup_{c \geq 0} \left\{ \delta(1 - \gamma)J(W, x, t) \ln c - cJ_W(W, x, t) \right\}.
$$

The first-order condition for the consumption choice is

$$
\delta(1 - \gamma)J(W, x, t) \frac{1}{c} = J_W(W, x, t) \iff c = \delta(1 - \gamma)J(W, x, t)J_W(W, x, t)^{-1},
$$

which implies that

$$
\mathcal{L}^c J(W, x, t) = \delta(1 - \gamma)J(W, x, t) \left[ \ln \delta + \ln \left(\left[1 - \gamma\right]J(W, x, t)\right) - \ln J_W(W, x, t) \right] - \delta(1 - \gamma)J(W, x, t) = \delta(1 - \gamma)J(W, x, t) \left[ \ln \delta + \ln \left(\left[1 - \gamma\right]J(W, x, t)\right) - \ln J_W(W, x, t) - 1 \right].
$$

Substituting (17.5) and (17.7) into (17.6), we arrive at

$$
0 = -\frac{1}{2}J_{WW}(W, x, t)\|\lambda(x)\|^2 - \frac{1}{2}J_{Wx}(W, x, t)^2\|v(x)\|^2 - \frac{J_W(W, x, t)J_{Wx}(W, x, t)}{J_{WW}(W, x, t)}v(x)^\top \lambda(x)
$$

$$
+ \delta(1 - \gamma)J(W, x, t) \left[ \ln \delta + \ln \left(\left[1 - \gamma\right]J(W, x, t)\right) - \ln J_W(W, x, t) \right] - \delta J(W, x, t) \ln \left(\left[1 - \gamma\right]J(W, x, t)\right) + \frac{\partial J}{\partial t}(W, x, t)
$$

$$
+ J_W(W, x, t)W_r(x) + J_x(W, x, t)m(x) + \frac{1}{2}J_{xx}(W, x, t)(v(x)^\top v(x) + \dot{v}(x)^2),
$$

We conjecture a solution of the form

$$
J(W, x, t) = \frac{1}{1 - \gamma}G(x, t)^\gamma W^{1 - \gamma}
$$

for some deterministic function $G$ to be determined. The terminal condition is $J(W, x, T) = \frac{1}{1 - \gamma}W^{1 - \gamma}$ so we need $G(x, T) = a^{1/\gamma}$ for all possible values of $x$. The relevant derivatives are

$$
J_W(W, x, t) = G(x, t)^\gamma W^{-\gamma},
$$

$$
J_{WW}(W, x, t) = -\gamma G(x, t)^\gamma W^{-\gamma - 1},
$$

$$
J_x(W, x, t) = \frac{\gamma}{1 - \gamma}G(x, t)^{\gamma - 1}G_x(x, t)W^{1 - \gamma},
$$

$$
J_{xx}(W, x, t) = -\gamma G(x, t)^{\gamma - 2}G_x(x, t)^2W^{1 - \gamma} + \frac{\gamma}{1 - \gamma}G(x, t)^{\gamma - 1}G_{xx}(x, t)W^{1 - \gamma},
$$

$$
J_W(W, x, t) = \frac{\gamma}{1 - \gamma}G(x, t)^{\gamma - 1}G_x(x, t)W^{-\gamma},
$$

$$
\frac{\partial J}{\partial t}(W, x, t) = \frac{\gamma}{1 - \gamma}G(x, t)^{\gamma - 1}\frac{\partial G}{\partial t}(x, t)W^{1 - \gamma}.
$$

Our candidates for the optimal decisions then become

$$
e^*_t = \delta W_t,
$$

$$
\pi^*_t = \frac{1}{\gamma} \left(\frac{G(x, t)}{G(x, t)}\right)^{-1} \lambda(x_t) + \frac{G_x(x_t, t)}{G(x, t)} \left(\frac{G(x, t)}{G(x, t)}\right)^{-1} v(x_t).
$$
Compared to the case with time-additive power utility, the portfolio seems unchanged but, of course, the function $G(x,t)$ might be different from the function $g(x,t)$ appearing with power utility. The candidate for the optimal consumption rate is very different from the power utility case. With power utility, it is optimal to consume a fraction $1/g(x,t)$ of wealth. With recursive utility and $\psi = 1$, it is optimal to consume a constant fraction of wealth equal to the subjective time preference rate $\delta$.

With the conjecture (17.8), we get

$$
\mathcal{L}^c J(W, x, t) = \delta G(x, t)^\gamma W^{1-\gamma} \left\{ \ln \delta + \ln \left( G(x, t)^\gamma W^{1-\gamma} \right) - \ln \left( G(x, t)^\gamma W^{-\gamma} - 1 \right) \right\}
$$

$$
= \delta G(x, t)^\gamma W^{1-\gamma} \left( \ln W + \ln \delta - 1 \right),
$$

$$
\mathcal{L}^\pi J(W, x, t) = G(x, t)^\gamma W^{1-\gamma} \left( \frac{1}{2\gamma} ||\lambda||^2 + \frac{\gamma}{2} \left( \frac{G_{x}(x,t)}{G(x,t)} \right)^2 ||v(x)||^2 + \frac{G_{x}(x,t)}{G(x,t)} v(x)^T \lambda(x) \right)
$$

Substitute into the HJB equation (17.6), multiply through by $\frac{1-\gamma}{\gamma} G(x, t)^{1-\gamma} W^{-\gamma}$, and simplify. Then you will get the following PDE for $G(x,t)$:

$$
0 = \frac{1}{2} \left( ||v(x)||^2 + \dot{v}(x)^T G_{xx}(x,t) + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x)^T v(x) \right) G_x(x,t) + \frac{1}{2} (\gamma - 1) \dot{v}(x)^T \frac{G_{x}(x,t)}{G(x,t)} \right) L_{W}(x,t)
$$

$$
+ \partial G \partial t (x,t) - \left( \delta \ln G(x,t) + \frac{\gamma - 1}{\gamma} \delta \ln \delta - 1 \right) + \frac{\gamma - 1}{\gamma} r(x) + \frac{\gamma - 1}{2 \gamma^2} ||\lambda(x)||^2 \right) G(x,t),
$$

(17.10)

which we have to solve with the terminal condition $G(x,T) = \alpha^{1/\gamma}$. We can obtain an explicit solution to this PDE under some assumptions on the dependence of $r$, $\lambda$, $v$, and $\dot{v}$ on $x$, whereas numerical solution techniques have to be implemented for other cases. Let us try a solution of the form

$$
G(x,t) = \alpha^{1/\gamma} e^{-D_0 (T-t) - D_1 (T-t) x}.
$$

The terminal condition implies that $D_0(0) = D_1(0) = 0$. After substitution into the PDE (17.10) and simplifications, we find that

$$
0 = \frac{1}{2} \left( ||v(x)||^2 + \gamma \dot{v}(x)^2 \right) D_1(T-t)^2 + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x)^T v(x) + \delta x \right) D_1(T-t) + D_1(T-t) x
$$

$$
+ D_0'(T-t) + \delta D_0(t) - \left( \frac{\delta}{\gamma} \ln \alpha + \frac{\gamma - 1}{\gamma} \delta \ln \delta - 1 \right) + \frac{\gamma - 1}{\gamma} r(x) + \frac{\gamma - 1}{2 \gamma^2} ||\lambda(x)||^2 \right).
$$

If $||v(x)||^2$, $\dot{v}(x)^2$, $m(x)$, $\lambda(x)^T v(x)$, $r(x)$, and $||\lambda(x)||^2$ are all affine functions of $x$, the above equation can be decomposed in a system of two ordinary differential equations for $D_0$ and $D_1$. Note that even though we are considering a case with utility of intermediate consumption, we can allow for incomplete markets (i.e., $\dot{v}(x) \neq 0$), and the solution $G(x,t)$ does not involve an integral; these findings contrast the results for time-additive power utility.

### 17.2.3 The case $\psi \neq 1$

When substituting the aggregator for $\psi \neq 1$ into (17.2), we can reformulate the HJB-equation as

$$
0 = \mathcal{L}^\pi J(W, x, t) + \mathcal{L}^c J(W, x, t) - \delta \theta J(W, x, t) + \frac{\partial J}{\partial t} (W, x, t)
$$

$$
+ J_W(W, x, t) W r(x) + J_x(W, x, t) m(x) + \frac{1}{2} J_{xx} W(x, t) v(x) + \dot{v}(x)^2,
$$

(17.11)
By substituting that equation together with (17.9) into the HJB-equation (17.11), multiplying which implies that
\[
c = \delta^\psi J_W(W, x, t)^{-\psi} ([1 - \gamma] J(W, x, t))^{\psi(1 - \frac{1}{\gamma})},
\]
which implies that
\[
\mathcal{L}^c J(W, x, t) = \frac{1}{\psi - 1} \delta^\psi J_W(W, x, t)^{1 - \psi} ([1 - \gamma] J(W, x, t))^{\psi(1 - \frac{1}{\gamma})}.
\]
Again, we conjecture that indirect utility is of the form (17.8) for some function \( G \) to be determined. If that is true, the optimal consumption rate is
\[
c_t^* = \delta^\psi (G(x_t, t)^{\gamma} W_t^{-\gamma})^{-\psi} \left( G(x_t, t)^\gamma W_t^{1 - \gamma} \right)^{\psi(1 - \frac{1}{\gamma})} = \delta^\psi G(x_t, t)^{-\psi \gamma / \theta} W_t, \tag{17.12}
\]
which implies that
\[
\mathcal{L}^c J(W, x, t) = \frac{1}{\psi - 1} \delta^\psi (G(x, t)^\gamma W^{-\gamma})^{1 - \psi} \left( G(x, t)^\gamma W^{1 - \gamma} \right)^{\psi(1 - \frac{1}{\gamma})}
= \frac{1}{\psi - 1} \delta^\psi G(x, t)^{(1 + \frac{\gamma - 1}{\gamma})} W^{1 - \gamma}.
\]
By substituting that equation together with (17.9) into the HJB-equation (17.11), multiplying through by \( \frac{1}{\gamma} G(x, t)^{1 - \gamma} W^{\gamma - 1} \), and simplifying, one arrives at the PDE
\[
0 = \frac{1}{2} \left( \| v(x) \|^2 + \hat{v}(x)^2 \right) G_{xx}(x, t) + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x) \right) v(x) G_x(x, t) + \frac{1}{2} (\gamma - 1) \hat{v}(x)^2 \frac{G_x(x, t)^2}{G(x, t)}
+ \frac{\theta}{\gamma \psi} \delta^\psi G(x, t)^{\frac{\gamma - 1}{\gamma - 1}} - \left( \frac{\delta \theta}{\gamma} + \frac{\gamma - 1}{\gamma} r(x) + \frac{\gamma - 1}{2 \gamma^2} \| \lambda(x) \|^2 \right) G(x, t),
\tag{17.13}
\]
which we have to solve with the terminal condition \( G(x, T) = \alpha^{1/\gamma} \).

The term with \( \frac{1}{\gamma - 1} \) is a potential complication. In the case of power utility, i.e., \( \psi = 1 / \gamma \), the power of \( G \) reduces to 0 so the term is simply the constant \( \delta^{1/\gamma} \). It is then well-known that we can find closed-form solutions for \( G(x, t) \) if the market is complete (so that \( \hat{v}(x) \equiv 0 \)) and the model has an affine or quadratic structure. For example, with an affine structure, the solution is of the form
\[
G(x, t) = \int_t^T \delta^{1/\gamma} \exp \left\{ -\frac{\delta}{\gamma} (s - t) + \frac{1 - \gamma}{\gamma} A_0(s - t) + \frac{1 - \gamma}{\gamma} A_1(s - t) x \right\} \, ds
+ \alpha^{1/\gamma} \exp \left\{ -\frac{\delta}{\gamma} (T - t) + \frac{1 - \gamma}{\gamma} A_0(T - t) + \frac{1 - \gamma}{\gamma} A_1(T - t) x \right\}
\]
for some deterministic functions \( A_0 \) and \( A_1 \) that solve certain ODEs.

In other cases than power utility, the PDE (17.13) does not seem to have an explicit solution. One way to proceed is to solve the PDE numerically. Another way is to introduce an approximation so that the nasty term disappears. For simplicity, consider first the case of constant investment opportunities, where we do not need any state variable \( x \). Then we are searching for the function \( G(t) \) solving the non-linear ODE
\[
0 = G'(t) + \frac{\theta}{\gamma \psi} \delta^\psi G(t)^{\frac{2 \gamma - 1}{2 \gamma - 1}} - AG(t), \quad A = \frac{\delta \theta}{\gamma} + \frac{\gamma - 1}{\gamma} r + \frac{\gamma - 1}{2 \gamma^2} \| \lambda \|^2, \tag{17.14}
\]
which we have to solve with the terminal condition $G(x, T) = \alpha^{1/\gamma}$. Following an idea originally put forward by Campbell (1993) in a discrete-time setting and adapted to a continuous-time setting by Chacko and Viceira (2005), we can obtain a closed-form approximate solution in the following way. A Taylor approximation of $z \mapsto e^z$ around $\hat{z}$ gives $e^z \approx e^{\hat{z}}(1 + z - \hat{z})$. When we apply that to $z = \frac{\gamma \psi - 1}{\gamma - 1} \ln G(t)$, we get

\[
G(t) = G(t)\frac{\gamma \psi - 1}{\gamma - 1} = G(t)e^{\frac{\gamma \psi - 1}{\gamma - 1} \ln G(t)}
\]

\[
\approx G(t)e^{\frac{\gamma \psi - 1}{\gamma - 1} \ln G(t)} \left( 1 + \frac{\gamma (\psi - 1)}{\gamma - 1} [\ln G(t) - \ln \hat{G}(t)] \right)
\]

(17.15)

Using that approximation in the ODE (17.14), we get

\[
0 = G''(t) - a(t)G(t) - b(t)G(t)\ln G(t),
\]

(17.16)

where

\[
a(t) = A - \delta \psi \hat{G}(t)\frac{\gamma \psi - 1}{\gamma - 1} \left( \frac{\theta}{\gamma \psi} + \ln \hat{G}(t) \right), \quad b(t) = \delta \psi \hat{G}(t)\frac{\gamma \psi - 1}{\gamma - 1}.
\]

The solution to (17.16) with $G(T) = \alpha^{1/\gamma}$ is

\[
G(t) = \alpha^{1/\gamma}e^{-D(t)}, \quad D(t) = \int_t^T e^{-\int_\tau^s b(u) du} \left( a(s) + b(s) \frac{1}{\gamma} \ln \alpha \right) ds.
\]

Using the approximation to $G(t)$ in the optimal consumption rule (17.12), we get

\[
e^\gamma_\delta(t) = \delta \psi \left( \alpha^{1/\gamma}e^{-D(t)} \right)^{-\psi \gamma/\theta} W_t = \delta \psi \alpha^{-\psi/\theta} e^{\psi \gamma/\theta} D(t) W_t,
\]

i.e., the optimal consumption rate is a time-dependent fraction of wealth. It remains to decide on the function $\hat{G}(t)$ in the approximation. We should make sure that $\ln G(t)$ is rather close to $\ln \hat{G}(t)$. One idea is to presume that the optimal consumption/wealth ratio from (17.12) is close to the optimal consumption/wealth ratio in the special case of $\psi = 1$, i.e.,

\[
\delta \psi G(t)^{-\psi \gamma/\theta} \approx \delta \Rightarrow G(t) \approx \delta^{-\gamma \psi - 1} \equiv \hat{G}(t).
\]

In that case, the functions $a$ and $b$ are simply constants,

\[
b = \delta, \quad a = A - \delta \left( \frac{\theta}{\gamma \psi} - \frac{1}{\gamma} \ln \delta \right),
\]

so that $D(t)$ reduces to

\[
D(t) = \left( \frac{A}{\delta} - \frac{\theta}{\gamma \psi} + \frac{1}{\gamma} \ln \delta + \frac{1}{\gamma} \ln \alpha \right) \left( 1 - e^{-\delta (T-t)} \right).
\]

In the case of stochastic investment opportunities, the approximation (17.15) becomes

\[
G(x, t) = G(x, t)\frac{\gamma \psi - 1}{\gamma - 1} \left( 1 + \frac{\gamma (\psi - 1)}{\gamma - 1} [\ln G(x, t) - \ln \hat{G}(t)] \right).
\]

By substituting that into (17.13), we obtain

\[
0 = \frac{1}{2} (\|v(x)\|^2 + \hat{v}(x)^2) G_{xx}(x, t) + \left( m(x) - \frac{\gamma - 1}{\gamma} \lambda(x)^\gamma v(x) \right) G_x(x, t) + \frac{1}{2} (\gamma - 1) \hat{v}(x)^2 \frac{G_x(x, t)^2}{G(x, t)}
\]

\[
+ \frac{\partial G}{\partial t}(x, t) + \frac{\theta}{\gamma \psi} \delta \psi G(x, t)\hat{G}(t)\frac{\gamma \psi - 1}{\gamma - 1} \left( 1 + \frac{\gamma (\psi - 1)}{\gamma - 1} [\ln G(x, t) - \ln \hat{G}(t)] \right)
\]

\[
- \left( \frac{\delta \theta}{\gamma} + \frac{\gamma - 1}{\gamma} r(x) + \frac{\gamma - 1}{2 \gamma^2} \|\lambda(x)\|^2 \right) G(x, t),
\]
which is a PDE of the same form as the relevant PDE (17.10) for the case $\psi = 1$, except that we now have an explicit time-dependence in the coefficient of the approximated term via $\hat{G}(t)$. If the model has an affine structure, the approximated PDE will therefore have a solution of the form

$$G(x, t) = \alpha^{1/\gamma} e^{-D_0(t,T) - D_1(t,T)x},$$

where the deterministic functions $D_0$ and $D_1$ now depend separately on $t$ and $T$ because of the time-dependent coefficients in the PDE. In particular, $D_0(t, T)$ and $D_1(t, T)$ will depend on the values of $\hat{G}(u)$ for $u \in (t, T)$. Again, $D_0$ and $D_1$ solve some equations that depend on the specific affine structure of the model. Intuitively, the approximation works best if $\hat{G}(t)$ is chosen so that $\ln G(x_t, t)$ stays close to $\ln \hat{G}(t)$, which is now potentially harder due to the presence of the stochastic process $x_t$. One idea is to determine $\hat{G}(t)$ so that

$$\ln \hat{G}(t) = \mathbb{E}[\ln G(x_t, t)] = \frac{1}{\gamma} \ln \alpha - D_0(t, T) - D_1(t, T) \mathbb{E}[x_t].$$

Since the right-hand side depends on all $\hat{G}(u)$ for $u \in (t, T)$, this involves a recursive procedure moving backwards from $T$.

In any case, it seems impossible to say anything concrete about the precision of the approximation. Of course, for a concrete problem the approximate solution could be compared to the solution stemming from a numerical solution of the relevant PDE for $G$, but apparently no such studies have been published.

17.3 Model/parameter uncertainty, incomplete information, learning


17.4 Ambiguity aversion

See Maenhout (2004)

17.5 Other objective functions

Portfolio choice problems of portfolio managers whose compensation depends on the performance of the portfolio chosen and a benchmark portfolio. The compensation may include option elements. See Carpenter (2000), Browne (1999).

17.6 Consumption and portfolio choice for non-price takers


17.7 Non-utility based portfolio choice

17.8 Allowing for bankruptcy

18.1 Trading constraints


18.2 Transaction costs

The simplest type of transaction costs to handle is proportional costs. Some initial, heuristic work was made by Magill and Constantinides (1976) and Constantinides (1979, 1986). A more formal analysis was provided by Davis and Norman (1990) and we follow their presentation.

Model set-up:

(1) Risk-free bank account with constant interest rate \( r \) (continuously compounded), traded without transaction costs.

(2) A single risky asset (the stock). The listed unit price \( P_t \) follows geometric Brownian motion:

\[
dP_t = P_t [\mu \, dt + \sigma \, dz_t].
\]

Buying one unit costs \( (1 + a)P_t \), selling one unit provides \( (1 - b)P_t \), where \( a, b \geq 0 \).

(3) Investment strategy in the stock is represented by the pair of processes \( (L, U) \) with \( L_t \) denoting the cumulative amounts of stock purchased on the time interval \([0, t]\) and \( U_t \) the cumulative amounts of stock sold on \([0, t]\), where the amounts are measured by the listed
(4) Let $S_0$ denote the balance of the bank account at time $t$ and let $S_{1t}$ denote the value of the stocks owned at time $t$ (measured at the listed unit price at time $t$). The dynamics is

$$
\begin{align*}
\frac{dS_0}{S_0} &= (rS_0 - c_t) \, dt - (1 + a) \, dL_t + (1 - b) \, dU_t, \quad S_{00} = x, \\
\frac{dS_{1t}}{S_{1t}} &= \mu S_{1t} \, dt + \sigma S_{1t} \, dz_t + dL_t - dU_t, \quad S_{10} = y.
\end{align*}
$$

Here $c_t$ is the consumption rate at time $t$.

(5) The individual is required to stay solvent, so that after eliminating his position in the stock, he should have non-negative wealth. If $S_{1t} > 0$, the requirement is $S_{0t} + (1 - b)S_{1t} \geq 0$, i.e., $S_{1t} \geq -\frac{1}{1 - b}S_{0t}$. If $S_{1t} < 0$, the requirement is $S_{0t} + (1 + a)S_{1t} \geq 0$, i.e., $S_{1t} \geq -\frac{1}{1 + a}S_{0t}$. The solvency region is therefore

$$
S = \{(x, y) \in \mathbb{R}^2 : x + (1 - b)y \geq 0, x + (1 + a)y \geq 0\}.
$$

(6) The set of admissible consumption and trading strategies is

$$
\mathcal{U}(x, y) = \{(c, L, U) : (S_{0t}, S_{1t}) \in S \text{ for all } t \geq 0 \text{ (a.s.), } c_t \geq 0\}
$$

(7) For preferences, assume infinite horizon and power utility with $\gamma > 1$ denoting the relative risk aversion. Let

$$
J(x, y) = \sup_{(c, L, U) \in \mathcal{U}(x, y)} \mathbb{E}_{x,y} \left[ \int_0^\infty e^{-\delta t} \frac{1}{1 - \gamma} c_t^{1-\gamma} \, dt \right].
$$

For the case without transaction costs ($a = b = 0$), we solved the similar problem for a finite time horizon in Section 6.3. Let $\lambda = (\mu - r)/\sigma$. If the constant

$$
A = \frac{\delta + r(\gamma - 1)}{\gamma} + \frac{1}{2} \frac{\gamma - 1}{\gamma^2} \lambda^2
$$

is positive, the limit as $T \to \infty$ of the solution is

$$
J(x, y) = \frac{1}{1 - \gamma} A^{-\gamma}(x + y)^{1-\gamma},
$$

$$
e^* = A[x + y],
$$

$$
\pi^* = \frac{\lambda}{\gamma\sigma},
$$

where $x + y$ is the total wealth. Since $\pi$ is the fraction of total wealth optimally invested in the stock, we have $\pi^*_t = \frac{S_{1t}}{S_{0t} + S_{1t}}$, and hence

$$
\frac{S_{1t}}{S_{0t}} = \frac{\pi^*}{1 - \pi^*} = \frac{\lambda}{\gamma\sigma - \lambda},
$$

corresponding to a straight line through the origin in the $(S_0, S_1)$-space, the so-called Merton line.

Let us turn to the case with transaction costs. Here is the first result:

**Theorem 18.1.** The value function $J(x, y)$ has the following properties:
(a) $J$ is concave, i.e., for $\theta \in [0, 1]$
\[ J(\theta x_1 + [1 - \theta]x_2, \theta y_1 + [1 - \theta]y_2) \geq \theta J(x_1, y_1) + [1 - \theta]J(x_2, y_2). \]

(b) $J$ is homogeneous of degree $1 - \gamma$, i.e., for $k > 0$
\[ J(kx, ky) = k^{1-\gamma}J(x, y). \]

Proof. (a) For any variable or process $\omega$, define $\omega^\theta = \theta \omega_1 + [1 - \theta]\omega_2$, where $\omega_1$ is associated with initial conditions $(x_i, y_i), i = 1, 2$. Let $a = (c, L, U)$ denote the control process. Then
\[
J(x^\theta, y^\theta) = \sup_{a \in \mathbb{U}(x^\theta, y^\theta)} E_{x^\theta, y^\theta} \left[ \int_0^\infty e^{-\delta t} \frac{1}{1 - \gamma} \gamma_t dt \right]
\geq \sup_{a^\theta \in \mathbb{U}(x^\theta, y^\theta)} E_{x^\theta, y^\theta} \left[ \int_0^\infty e^{-\delta t} \frac{1}{1 - \gamma} (\theta c_1 + [1 - \theta]c_2)^{1-\gamma} dt \right]
\geq \sup_{a_1 + [1 - \theta]a_2 \in \mathbb{U}(x^\theta, y^\theta)} E_{x^\theta, y^\theta} \left[ \int_0^\infty e^{-\delta t} \left\{ \theta \frac{1}{1 - \gamma} (c_1^{1-\gamma}) + [1 - \theta] \frac{1}{1 - \gamma} (c_2^{1-\gamma}) \right\} dt \right]
= \theta \sup_{a_1 \in \mathbb{U}(x_1, y_1)} E_{x_1, y_1} \left[ \int_0^\infty e^{-\delta t} \frac{1}{1 - \gamma} (c_1^{1-\gamma}) dt \right]
+ [1 - \theta] \sup_{a_2 \in \mathbb{U}(x_2, y_2)} E_{x_2, y_2} \left[ \int_0^\infty e^{-\delta t} \frac{1}{1 - \gamma} (c_2^{1-\gamma}) dt \right]
= \theta J(x_1, y_1) + [1 - \theta] J(x_2, y_2),
\]
where the first inequality holds due to the restriction to controls of the form $a^\theta$ instead of the general controls $a$, and the second inequality is due to the concavity of the power utility function.

(b) It is clear from the dynamics of $S_0$ and $S_1$ and the form of the solvency region that
\[ (c, L, U) \in \mathbb{U}(x, y) \iff (kc, kL, kU) \in \mathbb{U}(kx, ky). \]

Therefore
\[
J(kx, ky) = \sup_{(c, L, U) \in \mathbb{U}(x, y)} E_{x, y} \left[ \int_0^\infty e^{-\delta t} \frac{1}{1 - \gamma} (kc_1)^{1-\gamma} dt \right] = k^{1-\gamma}J(x, y).
\]

Of course, it follows from (b) that
\[ J(x, y) = k^{\gamma-1}J(kx, ky) \]
for any $k > 0$. Consequently,
\[ J_x(x, y) \equiv \frac{\partial J}{\partial x}(x, y) = \frac{\partial}{\partial x} (k^{\gamma-1}J(kx, ky)) = k^\gamma J_x(kx, ky) \]
and, similarly,
\[ J_y(x, y) \equiv \frac{\partial J}{\partial y}(x, y) = k^\gamma J_y(kx, ky). \]

It follows that
\[ \frac{J_y(kx, ky)}{J_x(kx, ky)} = \frac{J_y(x, y)}{J_x(x, y)} \]
for all $k > 0$. In other words, the ratio of the derivatives $J_y/J_x$ is constant along any straight line through the origin.
To derive and understand the optimal strategies, it is useful to apply some heuristic arguments by assuming that the trading strategies are of the form

$$L_t = \int_0^t l_s \, ds, \quad U_t = \int_0^t u_s \, ds; \quad l_s, u_s \in [0, K]$$

for some constant $K$. In particular, $dL_t = l_t \, dt$ and $dU_t = u_t \, dt$. The HJB equation is then

$$\delta J(x, y) = \sup_{c \geq 0 \in [0, K], u \in [0, K]} \left\{ \frac{1}{1 - \gamma} c^{1-\gamma} + J_x [r_x - c - (1 + a)l + (1 - b)u] + J_y [\mu y + l - u] + \frac{1}{2} J_{yy} \right\}$$

$$= \sup_{c \geq 0} \left\{ \frac{1}{1 - \gamma} c^{1-\gamma} - cJ_x \right\} + \sup_{l \in [0, K]} \{(J_y - (1 + a)J_x) l\} + \sup_{u \in [0, K]} \{(1 - b)J_x - J_y) u\} + r_x J_x + \mu y J_y + \frac{1}{2} \sigma^2 y^2 J_{yy}.$$  

The first-order conditions imply

$$l = \begin{cases} K, & \text{if } J_y \geq (1 + a)J_x, \\ 0, & \text{otherwise,} \end{cases} \quad u = \begin{cases} 0, & \text{if } J_y > (1 - b)J_x, \\ K, & \text{otherwise.} \end{cases}$$

Intuitively, purchasing stocks with a total listed price of one unit of account leads to an increase in utility equal to $J_y - (1 + a)J_x$. As long as this is positive, it is optimal to purchase more stocks. So the optimal strategy can be described in the following way:

$$J_y \geq (1 + a)J_x: \quad \text{buy stocks}$$

$$(1 + a)J_x > J_y > (1 - b)J_x: \quad \text{do not trade stocks}$$

$$J_y \leq (1 - b)J_x: \quad \text{sell stocks.}$$

This divides the solvency region into three regions: a buying region, a no trade region, and a selling region. The boundary $\partial B$ between the buying region and the no trade region is the set of points $(x, y)$ for which $J_y(x, y) = (1 + a)J_x(x, y)$, i.e., $J_y(x, y)/J_x(x, y) = 1 + a$. According to our analysis above, these points form a straight line in the $(x, y)$-plane through the origin. Let the slope of this line be denoted by $1/\omega_B$. The boundary $\partial S$ between the selling region and the no trade region is the set of points for which $J_y(x, y) = (1 - b)J_x(x, y)$, i.e., $J_y(x, y)/J_x(x, y) = 1 - b$, which again is true for points along a straight line through the origin. Denote the slope of this line by $1/\omega_S$. The $\partial S$ line is steeper than the $\partial B$ line, so we have $\omega_B \geq \omega_S$. The no trade region is a wedge in the $(x, y)$-plane bounded by the $\partial S$ and $\partial B$ lines.

In the selling region it is optimal to sell exactly the number of stocks needed to move to the selling boundary $\partial S$. Similarly, in the buying region it is optimal to buy the number of stocks needed to move to the buying boundary $\partial B$. If the initial holdings $(x, y)$ fall in the buying region or in the buying region, there will thus be an initial transaction to the nearest boundary. After that $(S_{0t}, S_{1t})$ will stay in the no trade region or on the boundaries $\partial S$ and $\partial B$. Even when no trades are made, the investments $(S_{0t}, S_{1t})$ will move around as the stock prices moves. As soon as the selling boundary is reached, enough stocks must be sold so that $(S_{0t}, S_{1t})$ does not move beyond the boundary $\partial S$ and into the interior of the selling region. Similarly when the buying
boundary is reached from inside the no trade region. After a potential initial trade, we will have
\[ \frac{1}{\omega_B} \leq \frac{S_{1t}}{S_{0t}} \leq \frac{1}{\omega_S}. \]
The fraction of wealth invested in the stock is \( \pi_t = \frac{S_{1t}}{S_{0t} + S_{1t}} \), which will then satisfy
\[ \frac{1}{1 + \omega_B} \leq \pi_t \leq \frac{1}{1 + \omega_S}. \]
In the case without transaction costs, transactions are made continuously to keep \( \pi_t \) constant. With transaction costs that strategy would be infinitely costly, and the solution shows that it is optimal to allow \( \pi_t \) to vary in an interval without making any transactions. Under some, apparently reasonable, conditions, the Merton portfolio weight \( \pi^* = \frac{1}{\gamma} \) will fall in the interval between \( 1/(1 + \omega_B) \) and \( 1/(1 + \omega_S) \), cf. Davis and Norman (1990). Intuitively, the investor will allow some deviation from the Merton weight before trading to save on transaction costs. There are cases, however, in which the Merton weight is outside the interval, cf. Shreve and Soner (1994).

Inside the no trade region, the HJB equation simplifies to
\[ \delta J = \sup_{c \geq 0} \left\{ \frac{1}{1 - \gamma} - cJ_x \right\} + r_x J_x + \mu y J_y + \frac{1}{2} \sigma^2 y^2 J_{yy} \]
\[ = \frac{\gamma}{1 - \gamma} J_x 1^{-\gamma} + r_x J_x + \mu y J_y + \frac{1}{2} \sigma^2 y^2 J_{yy}. \]
We can reduce the dimensionality of this partial differential equation by exploiting the homogeneity of the value function, since
\[ J \left( \frac{x}{y}, 1 \right) = \left( \frac{1}{y} \right)^{1-\gamma} J(x, y) \Rightarrow J(x, y) = y^{1-\gamma} J \left( \frac{x}{y}, 1 \right) \equiv y^{1-\gamma} \psi \left( \frac{x}{y} \right). \]
We thus have that
\[ J_x = y^{-\gamma} \psi' \left( \frac{x}{y} \right), \]
\[ J_y = (1 - \gamma)y^{-\gamma} \psi \left( \frac{x}{y} \right) - xy^{-\gamma-1} \psi' \left( \frac{x}{y} \right), \]
\[ J_{yy} = -\gamma(1 - \gamma)y^{-\gamma-1} \psi \left( \frac{x}{y} \right) + 2x\gamma y^{-\gamma-2} \psi' \left( \frac{x}{y} \right) + x^2 y^{-\gamma-3} \psi'' \left( \frac{x}{y} \right). \]
Substituting into the HJB equation, we arrive at an ordinary differential equation for \( \psi \):
\[ \frac{1}{2} \sigma^2 \omega^2 \psi''(\omega) + (r - \mu + \gamma \sigma^2) \omega \psi'(\omega) \]
\[ - \left( \delta + (\gamma - 1) \mu - \frac{1}{2} \sigma^2 \gamma (\gamma - 1) \right) \psi(\omega) + \frac{\gamma}{1 - \gamma} \psi'(\omega) 1^{-\gamma} = 0, \quad \omega \in [\omega_S, \omega_B]. \]
In the selling region, we must have \( J(x, y) \) constant along any line of slope \( -1/(1-b) \), so that \( J(x, y) = F(x + [1-b]y) \) for some function \( F \). Then \( J_x = F' \) and \( J_y = (1-b)F' \) so that \( J_y = (1-b)J_x \). Inserting the above expressions for \( J_x \) and \( J_y \), we see that
\[ \psi'(\omega)(\omega + 1 - b) = (1 - \gamma)\psi(\omega), \]
which is satisfied by
\[ \psi(\omega) = A \frac{1}{1 - \gamma} (\omega + 1 - b)^{1-\gamma} \]
for a constant $A$. Hence, $J(x, y) = y^{1-\gamma} \psi(x/y) = A \frac{1}{1-\gamma} (x + [1-b]y)^{1-\gamma}$. Using similar arguments, it can be shown that

$$\psi(\omega) = B \frac{1}{1-\gamma} (\omega + 1 + a)^{1-\gamma}$$

for some constant $B$ in the buying region, i.e., $J(x, y) = B \frac{1}{1-\gamma} (x + [1+b]y)^{1-\gamma}$.

To sum up, in order to obtain the full solution to the problem we have to find constants $\omega_B, \omega_S, A, B$ and a function $\psi$ so that

$$\frac{1}{2} \sigma^2 \omega^2 \psi''(\omega) + (r - \mu + \gamma \sigma^2) \omega \psi'(\omega) + \frac{\gamma}{1-\gamma} \psi'(\omega)^{1-\frac{1}{\gamma}}$$

$$- \left( \delta + (\gamma - 1) \mu - \frac{1}{2} \sigma^2 \gamma (\gamma - 1) \right) \psi(\omega) = 0, \ \omega \in [\omega_S, \omega_B],$$

$$\psi(\omega) = A \frac{1}{1-\gamma} (\omega + 1 - b)^{1-\gamma}, \ \omega \leq \omega_S,$n

$$\psi(\omega) = B \frac{1}{1-\gamma} (\omega + 1 + a)^{1-\gamma}, \ \omega \geq \omega_B.$$n

Theorem 4.2 in Davis and Norman (1990) shows that (under a technical condition) a solution to this problem will lead to the optimal strategies as described above. The optimal consumption rate will be

$$c_t^* = S_t \left( \frac{\psi'(S_0/S_t)}{\psi'} \right)^{-1/\gamma}.$$n

Theorem 5.1 in Davis and Norman (1990) confirms that a solution to the problem exists. At the boundaries, we have the so-called value-matching conditions

$$\psi(\omega_S) = A \frac{1}{1-\gamma} (\omega_S + 1 - b)^{1-\gamma},$$

$$\psi(\omega_B) = B \frac{1}{1-\gamma} (\omega_B + 1 + a)^{1-\gamma}.$$n

The so-called smooth-pasting conditions ensure that the derivative of $\psi$ at $\omega_S$ is the same from the left and from the right, and equivalently at $\omega_B$. Therefore

$$\psi'(\omega_S) = A (\omega_S + 1 - b)^{-\gamma},$$

$$\psi'(\omega_B) = B (\omega_B + 1 + a)^{-\gamma}.$$n

Numerical solution techniques are required!

Relevant extensions:

- proportional and fixed transaction costs: Øksendal and Sulem (2002)

A random variable $Y$ is said to be lognormally distributed if the random variable $X = \ln Y$ is normally distributed. In the following we let $m$ be the mean of $X$ and $s^2$ be the variance of $X$, so that

$$X = \ln Y \sim N(m, s^2).$$

The probability density function for $X$ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi s^2}} \exp \left\{ -\frac{(x - m)^2}{2s^2} \right\}, \quad x \in \mathbb{R}.$$

**Theorem A.1.** The probability density function for $Y$ is given by

$$f_Y(y) = \frac{1}{\sqrt{2\pi s^2 y}} \exp \left\{ -\frac{(\ln y - m)^2}{2s^2} \right\}, \quad y > 0,$$

and $f_Y(y) = 0$ for $y \leq 0$.

This result follows from the general result on the distribution of a random variable which is given as a function of another random variable; see any introductory text book on probability theory and distributions.

**Theorem A.2.** For $X \sim N(m, s^2)$ and $\gamma \in \mathbb{R}$ we have

$$\mathbb{E} [e^{-\gamma X}] = \exp \left\{ -\gamma m + \frac{1}{2}\gamma^2 s^2 \right\}.$$

**Proof.** Per definition we have

$$\mathbb{E} [e^{-\gamma X}] = \int_{-\infty}^{+\infty} e^{-\gamma x} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}} \, dx.$$

Manipulating the exponent we get

$$\mathbb{E} [e^{-\gamma X}] = e^{-\gamma m + \frac{1}{2}\gamma^2 s^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{1}{2\gamma^2} [(x-m)^2 + 2\gamma(x-m)s^2 + \gamma^2 s^2]} \, dx$$

$$= e^{-\gamma m + \frac{1}{2}\gamma^2 s^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-[m-\gamma s]^2)}{2s^2}} \, dx$$

$$= e^{-\gamma m + \frac{1}{2}\gamma^2 s^2},$$

219
where the last equality is due to the fact that the function
\[ x \mapsto \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-(m-\gamma s^2))^2}{2s^2}} \]
is a probability density function, namely the density function for an \( N(m - \gamma s^2, s^2) \) distributed random variable.

Using this theorem, we can easily compute the mean and the variance of the lognormally distributed random variable \( Y = e^X \). The mean is (let \( \gamma = -1 \))
\[
E[Y] = E[e^X] = \exp \left\{ m + \frac{1}{2} s^2 \right\}.
\]
With \( \gamma = -2 \) we get
\[
E[Y^2] = E[e^{2X}] = e^{2(m+s^2)},
\]
so that the variance of \( Y \) is
\[
\text{Var}[Y] = E[Y^2] - (E[Y])^2 = e^{2(m+s^2)} - e^{2m+s^2} = e^{2m+s^2} (e^{s^2} - 1).
\]

The next theorem provides an expression for the truncated mean of a lognormally distributed random variable, i.e., the mean of the part of the distribution that lies above some level. We define the indicator variable \( 1_{\{Y > K\}} \) to be equal to 1 if the outcome of the random variable \( Y \) is greater than the constant \( K \) and equal to 0 otherwise.

**Theorem A.3.** If \( X = \ln Y \sim N(m, s^2) \) and \( K > 0 \), then we have
\[
E[Y 1_{\{Y > K\}}] = e^{m + \frac{1}{2}s^2} N \left( \frac{m - \ln K}{s} + s \right)
= E[Y] N \left( \frac{m - \ln K}{s} + s \right).
\]

**Proof.** Because \( Y > K \iff X > \ln K \), it follows from the definition of the expectation of a random variable that
\[
E[Y 1_{\{Y > K\}}] = E[e^X 1_{\{X > \ln K\}}]
= \int_{\ln K}^{+\infty} e^x \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m-s^2)^2}{2s^2}} dx
= \int_{\ln K}^{+\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-(m+s^2))^2}{2s^2}} e^{2m+s^2} dx
= e^{m + \frac{1}{2}s^2} \int_{\ln K}^{+\infty} f_X(x) dx,
\]
where
\[
f_X(x) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-(m+s^2))^2}{2s^2}}\]
is the probability density function for an \( N(m + s^2, s^2) \) distributed random variable. The calculations

\[
\int_{\ln K}^{+\infty} f_X(x) \, dx = \text{Prob}(\bar{X} > \ln K)
\]

\[
= \text{Prob}\left( \frac{\bar{X} - [m + s^2]}{s} > \frac{\ln K - [m + s^2]}{s} \right)
\]

\[
= \text{Prob}\left( \frac{\bar{X} - [m + s^2]}{s} < -\frac{\ln K - [m + s^2]}{s} \right)
\]

\[
= N\left( -\frac{\ln K - [m + s^2]}{s} \right)
\]

\[
= N\left( \frac{m - \ln K}{s} + s \right)
\]

complete the proof.

\[\square\]

**Theorem A.4.** If \( X = \ln Y \sim N(m, s^2) \) and \( K > 0 \), we have

\[
E[\max(0, Y - K)] = e^{m + \frac{1}{2}s^2} N\left( \frac{m - \ln K}{s} + s \right) - KN\left( \frac{m - \ln K}{s} \right) - E[Y]N\left( \frac{m - \ln K}{s} + s \right) - KN\left( \frac{m - \ln K}{s} \right)
\]

**Proof.** Note that

\[
E[\max(0, Y - K)] = E[(Y - K)\mathbf{1}_{\{Y > K\}}]
\]

\[
= E[Y\mathbf{1}_{\{Y > K\}}] - K\text{Prob}(Y > K).
\]

The first term is known from Theorem A.3. The second term can be rewritten as

\[
\text{Prob}(Y > K) = \text{Prob}(X > \ln K)
\]

\[
= \text{Prob}\left( \frac{X - m}{s} > \frac{\ln K - m}{s} \right)
\]

\[
= \text{Prob}\left( \frac{X - m}{s} < -\frac{\ln K - m}{s} \right)
\]

\[
= N\left( -\frac{\ln K - m}{s} \right)
\]

\[
= N\left( \frac{m - \ln K}{s} \right).
\]

The claim now follows immediately. \[\square\]
Stochastic processes and stochastic calculus

B.1 Introduction

Most interest rates and asset prices vary over time in a non-deterministic way. We can observe the price of a given asset today, but the price of the same asset at any future point in time will typically be unknown, i.e., a random variable. In order to describe the uncertain evolution in the price of the asset over time, we need a collection of random variables, namely one random variable for each point in time. Such a collection of random variables is called a stochastic process. Modern finance models therefore apply stochastic processes to represent the evolution in prices and rates over time.

This chapter gives an introduction to stochastic processes and the mathematical tools needed to do calculations with stochastic processes, the so-called stochastic calculus. We will omit many technical details that are not important for a reasonable level of understanding and focus on processes. For more details and proofs, the reader is referred to textbooks on stochastic processes such as, for example, Øksendal (2003) and Karatzas and Shreve (1988), and to more extensive and formal introductions to stochastic processes in the mathematical finance textbooks of Dothan (1990), Duffie (2001), and Björk (2009).

The outline of the remainder of the chapter is as follows. In Section B.2 we define the concept of a stochastic process more formally and introduce much of the terminology used. We define and a particular process, the so-called Brownian motion, in Section B.3. This will be the basic building block in the definition of other processes. In Section B.4 we introduce the class of diffusion processes, which contains most of the processes used in popular fixed income models. Section B.5 gives a short introduction to the more general class of Itô processes. Both diffusions and Itô processes involve stochastic integrals, which are discussed in Section B.6. In Section B.7 we state the very important Itô’s Lemma, which is frequently applied when handling stochastic processes. Three diffusions that are widely used in finance models are introduced and studied in Section B.8. Section B.9 discusses multi-dimensional processes. Finally, Section B.10 explains the change of probability measure which is often used in financial models.
B.2 What is a stochastic process?

B.2.1 Probability spaces and information filtrations

The basic object for studies of uncertain events is a probability space, which is a triple \((\Omega, \mathcal{F}, P)\). Let us look at each of the three elements.

\(\Omega\) is the state space, which is the set of possible states or outcomes of the uncertain object. For example, if one studies the outcome of a throw of a dice (meaning the number of “eyes” on top of the dice), the state space is \(\Omega = \{1, 2, 3, 4, 5, 6\}\). In our finance models an outcome is a realization of all relevant uncertain objects over the entire time interval studied in the model. Only one outcome, the “true” outcome, will be realized.

\(\mathcal{F}\) is the set of events to which a probability can be assigned, i.e., the set of “probabilizable” events. Here, an event is a set of possible outcomes, i.e., a subset of the state space. In the example with the dice, some events are \(\{1, 2, 3\}\), \(\{4, 5\}\), \(\{1, 3, 5\}\), \(\{6\}\), and \(\{1, 2, 3, 4, 5, 6\}\). In a finance model an event is some set of realizations of the uncertain object. For example, in a model of the uncertain dynamics of a given asset price over a period of 10 years, one event is that the asset price one year into the future is above 100. Since \(\mathcal{F}\) is a set of events, it is really a set of subsets of the state space. It is required that

(i) the entire state space can be assigned a probability, i.e., \(\Omega \in \mathcal{F}\);

(ii) if some event \(F \subseteq \Omega\) can be assigned a probability, so can its complement \(F^c \equiv \Omega \setminus F\), i.e., \(F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}\); and

(iii) given a sequence of probabilizable events, the union is also probabilizable, i.e., \(F_1, F_2, \cdots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} F_i \in \mathcal{F}\).

Often \(\mathcal{F}\) is referred to as a sigma-algebra.

\(P\) is a probability measure, which formally is a function from the sigma-algebra \(\mathcal{F}\) into the interval \([0, 1]\). To each event \(F \in \mathcal{F}\), the probability measure assigns a number \(P(F)\) in the interval \([0, 1]\). This number is called the \(P\)-probability (or simply the probability) of \(F\). A probability measure must satisfy the following conditions:

(i) \(P(\Omega) = 1\) and \(P(\emptyset) = 0\), where \(\emptyset\) denotes the empty set;

(ii) the probability of the state being in the union of disjoint sets is equal to the sum of the probabilities for each of the sets, i.e., given \(F_1, F_2, \cdots \in \mathcal{F}\) with \(F_i \cap F_j = \emptyset\) for all \(i \neq j\), we have \(P(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} P(F_i)\).

Many different probability measures can be defined on the same sigma-algebra, \(\mathcal{F}\), of events. In the example of the dice, a probability measure \(P\) corresponding to the idea that the dice is “fair” is defined by

\(P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = 1/6\).

Another probability measure, \(Q\), can be defined by

\(Q(\{1\}) = 1/12, \quad Q(\{2\}) = \cdots = Q(\{5\}) = 1/6, \quad Q(\{6\}) = 3/12,\)

which may be appropriate if the dice is believed to be “unfair” in a particular way.
Two probability measures $P$ and $Q$ defined on the same state space $\Omega$ and sigma-algebra $\mathcal{F}$ are called **equivalent** if the two measures assign probability zero to exactly the same events, i.e., if $P(A) = 0 \Leftrightarrow Q(A) = 0$. The two probability measures in the dice example are equivalent. In the stochastic models of financial markets switching between equivalent probability measures turns out to be important.

In our models of the uncertain evolution of financial markets, the uncertainty is resolved gradually over time. At each date we can observe values of prices and rates that were previously uncertain so we learn more and more about the true outcome. We need to keep track of the information flow. Let us again consider the throw of a dice so that the state space is $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the set $\mathcal{F}$ of probabilizable events consists of all subsets of $\Omega$. Suppose now that the outcome of the throw of the dice is not resolved at once, but sequentially. In the beginning, at “time 0”, we know nothing about the true outcome so it can be any element in $\Omega$. Then, at “time 1”, you will be told that the outcome is either in the set $\{1, 2\}$, in the set $\{3, 4, 5\}$, or in the set $\{6\}$. Of course, in the latter case you will know exactly the true outcome, but in the first two cases there is still uncertainty about the true outcome. Later on, at “time 2”, the true outcome will be announced.

We can represent the information available at a given point in time by a **partition** of $\Omega$. By a partition of a given set, we simply mean a collection of disjoint subsets of $\Omega$ so that the union of these subsets equals the entire set $\Omega$. At time 0, we only know that one of the six elements in $\Omega$ will be realized. This corresponds to the (trivial) partition $F_0 = \{\Omega\}$. The information at time 1 can be represented by the partition

$$F_1 = \{\{1, 2\}, \{3, 4, 5\}, \{6\}\}.$$  

At time 2 we know exactly the true outcome, corresponding to the partition

$$F_2 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}.$$  

As time passes we receive more and more information about the true path. This is reflected by the fact that the partitions become finer and finer in the sense that every set in $F_1$ is a subset of some set in $F_0$ and every set in $F_2$ is a subset of some set in $F_1$. The information flow in this simple example can then be represented by the sequence $(F_0, F_1, F_2)$ of partitions of $\Omega$. In more general models, the information flow can be represented by a sequence $(F_t)_{t \in T}$ of partitions, where $T$ is the set of relevant points in time in the model. Each $F_t$ consists of disjoint events and the interpretation is that at time $t$ we will know which of these events the true outcome belongs to.

The fact that we learn more and more about the true outcome implies that the partitions will be increasingly fine meaning that, for $u > t$, every element in $F_t$ is a union of elements in $F_u$.

An alternative way of representing the information flow is in terms of an **information filtration**. Given a partition $F_t$ of $\Omega$, we can define $\mathcal{F}_t$ as the set of all unions of sets in $F_t$, including the “empty union”, i.e., the empty set $\emptyset$. Where $F_t$ contains the disjoint “decidable” events at time $t$, $\mathcal{F}_t$ contains all “decidable” events at time $t$. Each $\mathcal{F}_t$ is a sigma-algebra. For our example above we get

$$\mathcal{F}_0 = \{\emptyset, \Omega\},$$

$$\mathcal{F}_1 = \{\emptyset, \{1, 2\}, \{3, 4, 5\}, \{6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 6\}, \{3, 4, 5, 6\}, \Omega\},$$

whereas $\mathcal{F}_2$ becomes the collection of all possible subsets of $\Omega$. The sequence $\mathbf{F} = (\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2)$ is called an information filtration. In models involving the set $\mathcal{F}$ of points in time, the information
Appendix B. Stochastic processes and stochastic calculus

filtration is written as \( \mathbf{F} = (\mathcal{F}_t)_{t \in \mathcal{T}} \). We will always assume that the time 0 information is trivial, corresponding to \( \mathcal{F}_0 = \{\emptyset, \Omega\} \) and that all uncertainty is resolved at or before some final date \( T \) so that \( \mathcal{F}_T \) is equal to the set \( \mathcal{F} \) of all probabilizable events. The fact that we accumulate information dictates that \( \mathcal{F}_t \subset \mathcal{F}_{t'} \) whenever \( t < t' \), i.e., every set in \( \mathcal{F}_t \) is also in \( \mathcal{F}_{t'} \).

Above we constructed an information filtration from a sequence of partitions. We can also go from a filtration to a sequence of partitions. In each \( \mathcal{F}_t \), simply remove all sets that are unions of other sets in \( \mathcal{F}_t \). Therefore there is a one-to-one relationship between information filtration and a sequence of partitions. When we go to models with an infinite state space, the information filtration representation is preferable. Hence, our formal model of uncertainty and information is a **filtered probability space** \((\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F})\), where \((\Omega, \mathcal{F}, \mathbb{P})\) is a probability space and \( \mathbf{F} = (\mathcal{F}_t)_{t \in \mathcal{T}} \) is an information filtration. We will always assume that all the uncertainty is resolved over time. Hence, \( \mathcal{F}_T = \mathcal{F} \) in an economy where the terminal time point is \( T \). We will also assume that to begin with we know nothing about the future realizations of uncertainty, i.e., \( \mathcal{F}_0 \) is the trivial sigma-algebra consisting of only the full state space \( \Omega \) and the empty set \( \emptyset \).

It might seem frightening to have to specify a certain filtered probability space in which the behavior of interest rates, bond prices, etc., can be studied. However, in the models we are going to consider, the relevant filtered probability space will be implicitly defined via assumptions about the way the key variables can evolve over time.

In our models we will often deal with expectations of random variables, e.g., the expectation of the (discounted) payoff of an asset at a future point in time. In the computation of such an expectation we should take the information currently available into account. Hence we need to consider conditional expectations. One can generally write the expectation of a random variable \( X \) given the \( \sigma \)-algebra \( \mathcal{F}_t \) as \( \mathbb{E}[X | \mathcal{F}_t] \). For our purposes the \( \sigma \)-algebra \( \mathcal{F}_t \) will always represent the information at time \( t \) and we will write \( \mathbb{E}_t[X] \) instead of \( \mathbb{E}[X | \mathcal{F}_t] \). Since we assume that the information at time 0 is trivial, conditioning on time 0 information is the same as not conditioning on any information, hence \( \mathbb{E}_0[X] = \mathbb{E}[X] \). If we assume that all uncertainty is resolved at time \( T \), we have \( \mathbb{E}_T[X] = X \). We will sometimes use the following result:

**Theorem B.1** (The Law of Iterated Expectations). If \( \mathcal{F} \) and \( \mathcal{G} \) are two \( \sigma \)-algebras with \( \mathcal{F} \subseteq \mathcal{G} \) and \( X \) is a random variable, then \( \mathbb{E} \left[ \mathbb{E}[X | \mathcal{G}] \mid \mathcal{F} \right] = \mathbb{E}[X | \mathcal{F}] \). In particular, if \((\mathcal{F}_t)_{t \in \mathcal{T}}\) is an information filtration and \( t' > t \), we have

\[
\mathbb{E}_t \left[ \mathbb{E}_{t'}[X] \right] = \mathbb{E}_t[X].
\]

Loosely speaking, the theorem says that what you expect today of some variable that will be realized in two days is equal to what you expect today that you will expect tomorrow about the same variable. This is a very intuitive result. For a more formal statement and proof, see Øksendal (2003).

We can define conditional variances, covariances, and correlations from the conditional expectation exactly as one defines (unconditional) variances, covariances, and correlations from (unconditional) expectations:

\[
\text{Var}_t[X] = \mathbb{E}_t \left[ (X - \mathbb{E}_t[X])^2 \right] = \mathbb{E}_t[X^2] - \left( \mathbb{E}_t[X] \right)^2,
\]

\[
\text{Cov}_t[X, Y] = \mathbb{E}_t \left[ (X - \mathbb{E}_t[X])(Y - \mathbb{E}_t[Y]) \right] = \mathbb{E}_t[X Y] - \mathbb{E}_t[X] \mathbb{E}_t[Y],
\]

\[
\text{Corr}_t[X, Y] = \frac{\text{Cov}_t[X, Y]}{\sqrt{\text{Var}_t[X] \text{Var}_t[Y]}}.
\]
B.2 What is a stochastic process?

Again the conditioning on time $t$ information is indicated by a $t$ subscript.

B.2.2 Random variables and stochastic processes

A random variable is a function from $\Omega$ into $\mathbb{R}^K$ for some integer $K$. The random variable $x: \Omega \rightarrow \mathbb{R}^K$ associates to each outcome $\omega \in \Omega$ a value $x(\omega) \in \mathbb{R}^K$. Sometimes we will emphasize the dimension and say that the random variable is $K$-dimensional. With sequential resolution of the uncertainty the values of some random variables will be known before all uncertainty is resolved.

In the dice example with sequential information from before, suppose that your friend George will pay you 10 dollars if the dice shows either three, four, or five eyes and nothing in other cases. The payment from George is a random variable $x$. Of course, at time 2 you will know the true outcome, so the payment $x$ will be known at time 2. We say that $x$ is time 2 measurable or $\mathcal{F}_2$-measurable. At time 1 you will also know the payment $x$ because you will be told either that the true outcome is in $\{1, 2\}$, in which case the payment will be 0, or that the true outcome is in $\{3, 4, 5\}$, in which case the payment will be 10, or that the true outcome is 6, in which case the payment will be 0. So the random variable $x$ is also $\mathcal{F}_1$-measurable. Of course, at time 0 you will not know what payment you will get so $x$ is not $\mathcal{F}_0$-measurable. Suppose your friend John promises to pay you 10 dollars if the dice shows 4 or 5 and nothing otherwise. Represent the payment from John by the random variable $y$. Then $y$ is surely $\mathcal{F}_2$-measurable. However, $y$ is not $\mathcal{F}_1$-measurable, because if at time 1 you learn that the true outcome is in $\{3, 4, 5\}$, you still will not know whether you get the 10 dollars or not.

A stochastic process $x$ is a collection of random variables, namely one random variable for each relevant point in time. We write this as $x = (x_t)_{t \in \mathcal{T}}$, where each $x_t$ is a random variable. We still have an underlying filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F} = (\mathcal{F}_t)_{t \in \mathcal{T}})$ representing uncertainty and information flow. We will only consider processes $x$ that are adapted in the sense that for every $t \in \mathcal{T}$ the random variable $x_t$ is $\mathcal{F}_t$-measurable. This is just to say that the time $t$ value of the process will be known at time $t$. Some models consider the dynamic investment decisions of utility-maximizing investors (or other dynamic decisions under uncertainty). The investment decision is represented by a portfolio process characterizing the portfolio to be held at given points in time depending on the information of the investor at that date. Hence, it is natural to require that the portfolio process is adapted to the information filtration. You cannot base investment decisions on information you have not yet received.

By observing a given stochastic process $x$ adapted to a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F} = (\mathcal{F}_t)_{t \in \mathcal{T}})$, we obtain some information about the true state. In fact, we can define an information filtration $\mathcal{F}^x = (\mathcal{F}^x_t)_{t \in \mathcal{T}}$ generated by $x$. Here, $\mathcal{F}^x_t$ represents the information that can be deduced by knowing the values $x_s$ for $s \leq t$ (for technical reasons, this sigma-algebra is “completed” by including all sets of $\mathcal{F}$ that have zero $\mathbb{P}$-probability). $\mathcal{F}^x_t$ is the smallest sigma-algebra with respect to which $x$ is adapted. By construction, $\mathcal{F}^x_t \subseteq \mathcal{F}_t$.

B.2.3 Other important concepts and terminology

Let $x = (x_t)_{t \in \mathcal{T}}$ denote a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F} = (\mathcal{F}_t)_{t \in \mathcal{T}})$. Each possible outcome $\omega \in \Omega$ will fully determine the value of the process at all points in
time. We refer to this collection \((x_t(\omega))_{t \in \mathcal{T}}\) of realized values as a **(sample) path** of the process.

As time goes by, we can observe the evolution in the object which the stochastic process describes. At any given time \(t'\), the previous values \((x_t)_{t \leq t'}\) will be known. These values constitute the **history** of the process up to time \(t'\). The future values are (typically) still stochastic.

As time passes and we obtain new information about the true outcome, we will typically revise our expectations of the future values of the process or, more precisely, revise the probability distribution we attribute to the value of the process at any future point in time. Suppose we stand at time \(t\) and consider the value of a process \(x\) at a future time \(t' > t\). The distribution of the value of \(x_{t'}\) is characterized by probabilities \(P(x_{t'} \in A)\) for different sets \(A\). If for all \(t, t' \in \mathcal{T}\) with \(t < t'\) and all \(A\), we have that

\[
P(x_{t'} \in A \mid (x_s)_{s \in [0,t]}) = P(x_{t'} \in A \mid x_t),
\]

then \(x\) is called a **Markov process**. Broadly speaking, this condition says that, given the presence, the future is independent of the past. The history contains no information about the future value that cannot be extracted from the current value. Markov processes are often used in financial models to describe the evolution in prices of financial assets, since the Markov property is consistent with the so-called weak form of market efficiency, which says that extraordinary returns cannot be achieved by use of the precise historical evolution in the price of an asset.\(^1\) If extraordinary returns could be obtained in this manner, all investors would try to profit from it, so that prices would change immediately to a level where the extraordinary return is non-existent. Therefore, it is reasonable to model prices by Markov processes. In addition, models based on Markov processes are often more tractable than models with non-Markov processes.

A stochastic process is said to be a **martingale** if, at all points in time, the expected change in the value of the process over any given future period is equal to zero. In other words, the expected future value of the process is equal to the current value of the process. Because expectations depend on the probability measure, the concept of a martingale should be seen in connection with the applied probability measure. More rigorously, a stochastic process \(x = (x_t)_{t \geq 0}\) is a \(\mathbb{P}\)-martingale if for all \(t \in \mathcal{T}\) we have that

\[
E^\mathbb{P}_t [x_{t'}] = x_t, \quad \text{for all } t' \in \mathcal{T} \text{ with } t' > t.
\]

Here, \(E^\mathbb{P}_t\) denotes the expected value computed under the \(\mathbb{P}\)-probabilities given the information available at time \(t\), that is, given the history of the process up to and including time \(t\). Sometimes the probability measure will be clear from the context and can be notationally suppressed.

We assume, furthermore, that all the random variables \(x_t\) take on values in the same set \(S\), which we call the **value space** of the process. More precisely this means that \(S\) is the smallest set with the property that \(P(\{x_t \in S\}) = 1\). If \(S \subseteq \mathbb{R}\), we call the process a one-dimensional, real-valued process. If \(S\) is a subset of \(\mathbb{R}^K\) (but not a subset of \(\mathbb{R}^{K-1}\)), the process is called a \(K\)-dimensional, real-valued process, which can also be thought of as a collection of \(K\) one-dimensional, real-valued processes. Note that as long as we restrict ourselves to equivalent probability measures, the value space will not be affected by changes in the probability measure.

\(^1\)This does not conflict with the fact that the historical evolution is often used to identify some characteristic properties of the process, e.g., for estimation of means and variances.
B.2 What is a stochastic process?

B.2.4 Different types of stochastic processes

A stochastic process for the state of an object at every point in time in a given interval is called a **continuous-time stochastic process**. This corresponds to the case where the set $\mathcal{T}$ takes the form of an interval $[0,T]$ or $[0,\infty)$. In contrast, a stochastic process for the state of an object at countably many separated points in time is called a **discrete-time stochastic process**. This is, for example, the case when $\mathcal{T} = \{0, \Delta t, 2\Delta t, \ldots, T \equiv N\Delta t\}$ or $\mathcal{T} = \{0, \Delta t, 2\Delta t, \ldots\}$ for some $\Delta t > 0$. If the process can take on all values in a given interval (e.g., all real numbers), the process is called a **continuous-variable stochastic process**. On the other hand, if the state can take on only countably many different values, the process is called a **discrete-variable stochastic process**.

What type of processes should we use in our financial models? Our choice will be guided both by realism and tractability. First, let us consider the time dimension. The investors in the financial markets can trade at more or less any point in time. Due to practical considerations and transaction costs, no investor will trade continuously. However, it is not possible in advance to pick a fairly moderate number of points in time where all trades take place. Also, with many investors there will be some trades at almost any point in time, so that prices and interest rates etc. will also change almost continuously. Therefore, it seems to be a better approximation of real life to describe such economic variables by continuous-time stochastic processes than by discrete-time stochastic processes. Continuous-time stochastic processes are in many aspects also easier to handle than discrete-time stochastic processes.

Next, consider the value dimension. Strictly speaking, most economic variables can only take on countably many values in practice. Stock prices are multiples of the smallest possible unit (0.01 currency units in many countries), and interest rates are only stated with a given number of decimals. But since the possible values are very close together, it seems reasonable to use continuous-variable processes in the modelling of these objects. In addition, the mathematics involved in the analysis of continuous-variable processes is simpler and more elegant than the mathematics for discrete-variable processes. Integrals are easier to deal with than sums, derivatives are easier to handle than differences, etc. Some models were originally formulated using discrete-time, discrete-variable processes as, for example, the binomial option pricing model. For many years, the most significant model developments have applied continuous-time, continuous-variable processes, and such continuous-time term structure models are now standard in the financial industry and in academic work. In sum, we will use continuous-time, continuous-variable stochastic processes throughout to describe the evolution in prices and rates. Therefore the remaining sections of this chapter will be devoted to that type of stochastic processes.

It should be noted that discrete-time and/or discrete-variable processes also have their virtues. First, many concepts and results are easier understood or illustrated in a simple framework. Second, even if we have low-frequency data for many financial variables, we do not have continuous data. When it comes to estimation of parameters in financial models, continuous-time processes often have to be approximated by discrete-time processes. Third, although explicit results on asset prices, optimal investment strategies, etc. are easier to obtain with continuous-time models, not all relevant questions can be explicitly answered. Some problems are solved numerically by computer algorithms and also for that purpose it is often necessary to approximate continuous-time, continuous-variable processes with discrete-time, discrete-variable processes (see Chapter 9).
B.2.5 How to write up stochastic processes

Many financial models describe the movements and comovements of various variables simultaneously. The standard modelling procedure is to assume that there is some common exogenous shock that affects all the relevant variables and then model the response of all these variables to that shock. First, consider a discrete-time framework with time set $\mathcal{T} = \{0, t_1, t_2, \ldots, t_N \equiv T\}$ where $t_n = n \Delta t$. The shock over any period $[t_n, t_{n+1}]$ is represented by a random variable $\varepsilon_{t_{n+1}}$, which in general may be multi-dimensional, but let us for now just focus on the one-dimensional case. The sequence of shocks $\varepsilon_{t_1}, \varepsilon_{t_2}, \ldots, \varepsilon_{t_N}$ constitutes the basic or the underlying uncertainty in the model. Since the shock should represent some unexpected information, assume that every $\varepsilon_{t_n}$ has mean zero.

A stochastic process $x = (x_t)_{t \in \mathcal{T}}$ representing the dynamics of a price, an interest rate, or another interesting variable can then be defined by the initial value $x_0$ and the increments $\Delta x_{t_{n+1}} \equiv x_{t_{n+1}} - x_{t_n}, n = 0, \ldots, N-1$, which are typically assumed to be of the form

$$\Delta x_{t_{n+1}} = \mu_{t_n} \Delta t + \sigma_{t_n} \varepsilon_{t_{n+1}}.$$  

(B.1)

In general $\mu_{t_n}$ and $\sigma_{t_n}$ can themselves be stochastic, but must be known at time $t_n$, i.e., they must be $\mathcal{F}_t$-measurable random variables. In fact, we can form adapted processes $\mu = (\mu_t)_{t \in \mathcal{T}}$ and $\sigma = (\sigma_t)_{t \in \mathcal{T}}$. Given the information available at time $t_n$, the only random variable on the right-hand side of (B.1) is $\varepsilon_{t_{n+1}}$, which is assumed to have mean zero and some variance $\text{Var}[\varepsilon_{t_{n+1}}]$.

Hence, the mean and variance of $\Delta x_{t_{n+1}},$ conditional on time $t_n$ information, are

$$\mathbb{E}_{t_n}[\Delta x_{t_{n+1}}] = \mu_{t_n} \Delta t,$$

$$\text{Var}_{t_n}[\Delta x_{t_{n+1}}] = \sigma_{t_n}^2 \text{Var}[\varepsilon_{t_{n+1}}].$$

We can see that $\mu_{t_n}$ has the interpretation of the expected change in $x$ per time period.

If the shocks $\varepsilon_{t_1}, \ldots, \varepsilon_{t_N}$ are the only source of randomness in all the quantities we care about, then the relevant information filtration is exactly $\mathcal{F}^x = (\mathcal{F}_t)_{t \in \mathcal{T}}$, i.e., $\mathcal{F}_t = \mathcal{F}_{t_n}$. In that case $\mu_{t_n}$ and $\sigma_{t_n}$ are required to be measurable with respect to $\mathcal{F}_{t_n}$, i.e., they can depend on the realizations of $\varepsilon_{t_1}, \ldots, \varepsilon_{t_n}$. If $\sigma_{t_n}$ is non-zero at all times and for all states, we can invert (B.1) to get

$$\varepsilon_{t_{n+1}} = \frac{\Delta x_{t_{n+1}} - \mu_{t_n} \Delta t}{\sigma_{t_n}}.$$  

It is then clear that we learn exactly the same from observing the $x$-process as observing the exogenous shocks directly, i.e., $\mathcal{F}^x = \mathcal{F}^\varepsilon = \mathcal{F}$. We can fix the set of probabilizable events $\mathcal{F}$ to $\mathcal{F}_T = \mathcal{F}_T$. The probability measure $\mathbb{P}$ will be defined by specifying the probability distribution of each of the shocks $\varepsilon_{t_n}$.

From the sequence $\varepsilon_{t_1}, \varepsilon_{t_2}, \ldots, \varepsilon_{t_N}$ of exogenous shocks we can define a stochastic process $z = (z_t)_{t \in \mathcal{T}}$ by letting $z_0 = 0$ and $z_{t_n} = \varepsilon_{t_1} + \cdots + \varepsilon_{t_n}$. Consequently, $\varepsilon_{t_{n+1}} = z_{t_{n+1}} - z_{t_n} \equiv \Delta z_{t_{n+1}}$. Now the process $z$ captures the basic uncertainty in the model. The information filtration of the model is then defined by the information that can be extracted from observing the path of $z$. Without loss of generality we can assume that $\text{Var}[\Delta z_{t_{n+1}}] = \text{Var}[\varepsilon_{t_{n+1}}] = \Delta t$ for any period $[t_n, t_{n+1}]$. With the $z$-notation we can rewrite (B.1) as

$$\Delta x_{t_{n+1}} = \mu_{t_n} \Delta t + \sigma_{t_n} \Delta z_{t_{n+1}}$$

(B.2)

and now $\text{Var}_{t_n}[\Delta x_{t_{n+1}}] = \sigma_{t_n}^2 \Delta t$ so that $\sigma_{t_n}^2$ can be interpreted as the variance of the change in $x$ per time period.
The distribution of $\Delta x_{t_{n+1}}$ will be determined by the distribution assumed for the shocks $\varepsilon_{t_{n+1}} = \Delta z_{t_{n+1}}$. If the shocks are assumed to be normally distributed, the increment $\Delta x_{t_{n+1}}$ will be normally distributed conditional on time $t$ information, but not necessarily if we condition on earlier or no information.

We can loosely think of a continuous-time model as the result of taking a discrete-time model and let $\Delta t$ go to zero. In that spirit we will often define a continuous-time stochastic process $x = (x_t)_{t \in \mathcal{T}}$ by writing

$$dx_t = \mu_t \, dt + \sigma_t \, dz_t$$

which is to be thought of as the limit of (B.2) as $\Delta t \to 0$. Hence, $dx_t$ represents the change in $x$ over the infinitesimal (i.e., infinitely short) period after time $t$. Similarly for $dz_t$. The interpretations of $\mu_t$ and $\sigma_t$ are also similar to the discrete-time case. While (B.3) might seem very intuitive, it does not really make much sense to talk about the change of something over a period of infinitesimal length. The expression (B.3) really means that the change in the value of $x$ over any time interval $[t, t'] \subseteq \mathcal{T}$ is given by

$$x_{t'} - x_t = \int_t^{t'} \mu_u \, du + \int_t^{t'} \sigma_u \, dz_u.$$

The problem is that the right-hand side of this equation will not make sense before we define the two integrals. The integral $\int_t^{t'} \mu_u \, du$ is simply defined as the random variable whose value in any state $\omega \in \Omega$ is given by $\int_t^{t'} \mu_u(\omega) \, du$, which is an ordinary integral of real-valued function of time. If $\mu$ is adapted, the value of the integral $\int_t^{t'} \mu_u \, du$ will become known at time $t'$. The definition of the integral $\int_t^{t'} \sigma_u \, dz_u$ is much more delicate. We will return to that issue in Section B.6.

In almost all the continuous-time models studied in this book we will assume that the basic exogenous shocks are normally distributed, i.e., that the change in the shock process $z$ over any time interval is normally distributed. A process $z$ with this property is the so-called standard Brownian motion. In the next section we will formally define this process and study some of its properties. Then in later sections we will build various processes $x$ from that basic process $z$.

### B.3 Brownian motions

All the stochastic processes we shall apply in the financial models in the following chapters build upon a particular class of processes, the so-called Brownian motions. A (one-dimensional) stochastic process $z = (z_t)_{t \geq 0}$ is called a **standard Brownian motion**, if it satisfies the following conditions:

(i) $z_0 = 0$,

(ii) for all $t, t' \geq 0$ with $t < t'$: $z_{t'} - z_t \sim N(0, t' - t)$ [normally distributed increments],

(iii) for all $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $z_{t_1} - z_{t_0}, \ldots, z_{t_n} - z_{t_{n-1}}$ are mutually independent [independent increments],

(iv) $z$ has continuous paths.

Here $N(\alpha, \beta)$ denotes the normal distribution with mean $\alpha$ and variance $\beta$.

If we suppose that a standard Brownian motion $z$ represents the basic exogenous shock to an economy over a time interval $[0, T]$, then the relevant filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ is
implicitly given as follows. The state space $\Omega$ is the set of all possible paths $(z_t)_{t \in [0,T]}$. The information filtration is the one generated by $z$, i.e., $\mathcal{F} = \mathcal{F}^z$. The set of probabilizable events $\mathcal{F}$ is equal to $\mathcal{F}_T$. The probability measure $P$ is defined by the requirement that
$$
P\left(\frac{z_t - z_t'}{\sqrt{t'-t}} < h\right) = N(h) \equiv \int_{-\infty}^{h} \frac{1}{\sqrt{2\pi}} e^{-a^2/2} da$$
for all $t < t'$ and all $h \in \mathbb{R}$, where $N(\cdot)$ denotes the cumulative distribution function for an $N(0,1)$-distributed random stochastic variable.

Note that a standard Brownian motion is a Markov process, since the increment from today to any future point in time is independent of the history of the process. A standard Brownian motion is also a martingale, since the expected change in the value of the process is zero.

The name Brownian motion is in honor of the Scottish botanist Robert Brown, who in 1828 observed the apparently random movements of pollen submerged in water. The often used name Wiener process is due to Norbert Wiener, who in the 1920s was the first to show the existence of a stochastic process with these properties and who initiated a mathematically rigorous analysis of the process. As early as in the year 1900, the standard Brownian motion was used in a model for stock price movements by the French researcher Louis Bachelier, who derived the first option pricing formula, cf. Bachelier (1900).

The choice of using standard Brownian motions to represent the underlying uncertainty has an important consequence. All the processes defined by equations of the form (B.3) will then have continuous paths, i.e., there will be no jumps. Stochastic processes which have paths with discontinuities also exist. The jumps of such processes are often modeled by Poisson processes or related processes. It is well-known that large, sudden movements in financial variables occur from time to time, for example, in connection with stock market crashes. There may be many explanations of such large movements, for example, a large unexpected change in the productivity in a particular industry or the economy in general, perhaps due to a technological break-through. Another source of sudden, large movements is changes in the political or economic environment such as unforeseen interventions by the government or central bank. Stock market crashes are sometimes explained by the bursting of a bubble. Whether such sudden, large movements can be explained by a sequence of small continuous movements in the same direction or jumps have to be included in the models is an empirical question, which is still open. Large movements over a short period of time seem to be less frequent in interest rates and bond prices than in stock prices.

The defining characteristics of a standard Brownian motion look very nice, but they have some drastic consequences. It can be shown that the paths of a standard Brownian motion are nowhere differentiable, which broadly speaking means that the paths bend at all points in time and are therefore strictly speaking impossible to illustrate. However, one can get an idea of the paths by simulating the values of the process at different times. If $\varepsilon_1, \ldots, \varepsilon_n$ are independent draws from a standard $N(0,1)$ distribution, we can simulate the value of the standard Brownian motion at time $0 = t_0 < t_1 < t_2 < \cdots < t_n$ as follows:
$$z_{t_i} = z_{t_{i-1}} + \varepsilon_i \sqrt{t_i - t_{i-1}}, \quad i = 1, \ldots, n.$$  

With more time points and hence shorter intervals we get a more realistic impression of the paths of the process. Figure B.1 shows a simulated path for a standard Brownian motion over the interval $[0,1]$ based on a partition of the interval into 200 subintervals of equal length. Note that since
a normally distributed random variable can take on infinitely many values, a standard Brownian motion has infinitely many paths that each has a zero probability of occurring. The figure shows just one possible path.

Another property of a standard Brownian motion is that the expected length of the path over any future time interval (no matter how short) is infinite. In addition, the expected number of times a standard Brownian motion takes on any given value in any given time interval is also infinite. Intuitively, these properties are due to the fact that the size of the increment of a standard Brownian motion over an interval of length $\Delta t$ is proportional to $\sqrt{\Delta t}$, in the sense that the standard deviation of the increment equals $\sqrt{\Delta t}$. When $\Delta t$ is close to zero, $\sqrt{\Delta t}$ is significantly larger than $\Delta t$, so the changes are large relative to the length of the time interval over which the changes are measured.

The expected change in an object described by a standard Brownian motion equals zero and the variance of the change over a given time interval equals the length of the interval. This can easily be generalized. As before let $z = (z_t)_{t \geq 0}$ be a one-dimensional standard Brownian motion and define a new stochastic process $x = (x_t)_{t \geq 0}$ by

$$
x_t = x_0 + \mu t + \sigma z_t, \quad t \geq 0,
$$

where $x_0$, $\mu$, and $\sigma$ are constants. The constant $x_0$ is the initial value for the process $x$. It follows from the properties of the standard Brownian motion that, seen from time 0, the value $x_t$ is normally distributed with mean $x_0 + \mu t$ and variance $\sigma^2 t$, i.e., $x_t \sim N(x_0 + \mu t, \sigma^2 t)$.

The change in the value of the process between two arbitrary points in time $t$ and $t'$, where $t < t'$, is given by

$$
x_{t'} - x_t = \mu(t' - t) + \sigma(z_{t'} - z_t).
$$

The change over an infinitesimally short interval $[t, t + \Delta t]$ with $\Delta t \to 0$ is often written as

$$
dx_t = \mu \, dt + \sigma \, dz_t, \quad (B.4)
$$
Appendix B. Stochastic processes and stochastic calculus

where $dz_t$ can loosely be interpreted as a $N(0, dt)$-distributed random variable. As discussed earlier, this must really be interpreted as a limit of the expression

$$x_{t+\Delta t} - x_t = \mu \Delta t + \sigma (z_{t+\Delta t} - z_t)$$

for $\Delta t \to 0$. The process $x$ is called a **generalized Brownian motion**, or an arithmetic Brownian motion, or a generalized Wiener process. The parameter $\mu$ reflects the expected change in the process per unit of time and is called the **drift** of the process. The parameter $\sigma$ reflects the uncertainty about the future values of the process. More precisely, $\sigma^2$ reflects the variance of the change in the process per unit of time and is often called the **variance rate** of the process. $\sigma$ is a measure for the standard deviation of the change per unit of time and is referred to as the **volatility** of the process.

A generalized Brownian motion inherits many of the characteristic properties of a standard Brownian motion. For example, also a generalized Brownian motion is a Markov process, and the paths of a generalized Brownian motion are also continuous and nowhere differentiable. However, a generalized Brownian motion is not a martingale unless $\mu = 0$. The paths can be simulated by choosing time points $0 \equiv t_0 < t_1 < \cdots < t_n$ and iteratively computing

$$x_{t_i} = x_{t_{i-1}} + \mu (t_i - t_{i-1}) + \varepsilon_i \sigma \sqrt{t_i - t_{i-1}}, \quad i = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent draws from a standard normal distribution. Figures B.2 and B.3 show simulated paths for different values of the parameters $\mu$ and $\sigma$. The straight lines represent the deterministic trend of the process, which corresponds to imposing the condition $\sigma = 0$ and hence ignoring the uncertainty. Both figures are drawn using the same sequence of random numbers $\varepsilon_i$, so that they are directly comparable. The parameter $\mu$ determines the trend, and the parameter $\sigma$ determines the size of the fluctuations around the trend.

If the parameters $\mu$ and $\sigma$ are allowed to be time-varying in a deterministic way, the process $x$ is said to be a **time-inhomogeneous** generalized Brownian motion. In differential terms such a process can be written as defined by

$$dx_t = \mu(t) dt + \sigma(t) dz_t. \quad (B.5)$$

Over a very short interval $[t, t+\Delta t]$ the expected change is approximately $\mu(t) \Delta t$, and the variance of the change is approximately $\sigma(t)^2 \Delta t$. More precisely, the increment over any interval $[t, t']$ is given by

$$x_{t'} - x_t = \int_t^{t'} \mu(u) du + \int_t^{t'} \sigma(u) d\varepsilon_u.$$ 

The last integral is a so-called stochastic integral, which we will define and describe in a later section. There we will also state a theorem, which implies that, seen from time $t$, the integral $\int_t^{t'} \sigma(u) d\varepsilon_u$ is a normally distributed random variable with mean zero and variance $\int_t^{t'} \sigma(u)^2 du$.

B.4 Diffusion processes

For both standard Brownian motions and generalized Brownian motions, the future value is normally distributed and can therefore take on any real value, i.e., the value space is equal to $\mathbb{R}$. Many economic variables can only have values in a certain subset of $\mathbb{R}$. For example, prices of
Figure B.2: Simulation of a generalized Brownian motion with $\mu = 0.2$ and $\sigma = 0.5$ or $\sigma = 1.0$. The straight line shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals.

Figure B.3: Simulation of a generalized Brownian motion with $\mu = 0.6$ and $\sigma = 0.5$ or $\sigma = 1.0$. The straight line shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals.
financial assets with limited liability are non-negative. The evolution in such variables cannot be well represented by the stochastic processes studied so far. In many situations we will instead use so-called diffusion processes.

A (one-dimensional) **diffusion process** is a stochastic process \( x = (x_t)_{t \geq 0} \) for which the change over an infinitesimally short time interval \([t, t+dt]\) can be written as

\[
dx_t = \mu(x_t, t) \, dt + \sigma(x_t, t) \, dz_t, \tag{B.6}
\]

where \( z \) is a standard Brownian motion, but where the drift \( \mu \) and the volatility \( \sigma \) are now functions of time and the current value of the process.\(^2\) This expression generalizes (B.4), where \( \mu \) and \( \sigma \) were assumed to be constants, and (B.5), where \( \mu \) and \( \sigma \) were functions of time only. An equation like (B.6), where the stochastic process enters both sides of the equality, is called a **stochastic differential equation**. Hence, a diffusion process is a solution to a stochastic differential equation.

If both functions \( \mu \) and \( \sigma \) are independent of time, the diffusion is said to be **time-homogeneous**, otherwise it is said to be **time-inhomogeneous**. For a time-homogeneous diffusion process, the distribution of the future value will only depend on the current value of the process and how far into the future we are looking – not on the particular point in time we are standing at. For example, the distribution of \( x_{t+\delta} \) given \( x_t = x \) will only depend on \( x \) and \( \delta \), but not on \( t \).

This is not the case for a time-inhomogeneous diffusion, where the distribution will also depend on \( t \).

In the expression (B.6) one may think of \( dz_t \) as being \( N(0, dt) \)-distributed, so that the mean and variance of the change over an infinitesimally short interval \([t, t+dt]\) are given by

\[
E_t[dx_t] = \mu(x_t, t) \, dt, \quad \text{Var}_t[dx_t] = \sigma(x_t, t)^2 \, dt,
\]

where \( E_t \) and \( \text{Var}_t \) denote the mean and variance, respectively, conditionally on the available information at time \( t \). To be more precise, the change in a diffusion process over any interval \([t, t']\) is

\[
x_{t'} - x_t = \int_t^{t'} \mu(x_u, u) \, du + \int_t^{t'} \sigma(x_u, u) \, dz_u, \tag{B.7}
\]

where \( \int_t^{t'} \sigma(x_u, u) \, dz_u \) is a stochastic integral, which we will discuss in Section B.6. However, we will continue to use the simple and intuitive differential notation (B.6). The drift rate \( \mu(x_t, t) \) and the variance rate \( \sigma(x_t, t)^2 \) are really the limits

\[
\mu(x_t, t) = \lim_{\Delta t \to 0} \frac{E_t[x_{t+\Delta t} - x_t]}{\Delta t},
\]

\[
\sigma(x_t, t)^2 = \lim_{\Delta t \to 0} \frac{\text{Var}_t[x_{t+\Delta t} - x_t]}{\Delta t}.
\]

A diffusion process is a Markov process as can be seen from (B.6), since both the drift and the volatility only depend on the current value of the process and not on previous values. A diffusion process is not a martingale, unless the drift \( \mu(x_t, t) \) is zero for all \( x_t \) and \( t \). A diffusion process will have continuous, but nowhere differentiable paths. The value space for a diffusion process and the distribution of future values will depend on the functions \( \mu \) and \( \sigma \). If \( \sigma(x, t) \) is continuous and non-zero, the information generated by \( x \) will be identical to the information generated by \( z \), i.e., \( F^x = F^z \).

\(^2\)For the process \( x \) to be mathematically meaningful, the functions \( \mu(x, t) \) and \( \sigma(x, t) \) must satisfy certain conditions. See, e.g., Øksendal (2003, Ch. 7) and Duffie (2001, App. E).
In Section B.8 we will give some important examples of diffusion processes which we shall use in later chapters to model the evolution of some economic variables.

### B.5 Itô processes

It is possible to define even more general continuous-variable stochastic processes than those in the class of diffusion processes. A (one-dimensional) stochastic process $x_t$ is said to be an Itô process, if the local increments are on the form

$$dx_t = \mu_t \, dt + \sigma_t \, dz_t,$$

(B.8)

where the drift $\mu$ and the volatility $\sigma$ themselves are stochastic processes. A diffusion process is the special case where the values of the drift $\mu_t$ and the volatility $\sigma_t$ are given by $t$ and $x_t$. For a general Itô process, the drift and volatility may also depend on past values of the $x$ process. Or the drift and volatility can depend on another exogenous shock, for example, another standard Brownian motion than $z$. It follows that Itô processes are generally not Markov processes. They are generally not martingales either, unless $\mu_t$ is identically equal to zero (and $\sigma_t$ satisfies some technical conditions). The processes $\mu$ and $\sigma$ must satisfy certain regularity conditions for the $x$ process to be well-defined. We will refer the reader to Øksendal (2003, Ch. 4).

The expression (B.8) gives an intuitive understanding of the evolution of an Itô process, but it is more precise to state the evolution in the integral form

$$x_{t'} - x_t = \int_t^{t'} \mu_u \, du + \int_t^{t'} \sigma_u \, dz_u,$$

(B.9)

where the last term again is a stochastic integral.

### B.6 Stochastic integrals

#### B.6.1 Definition and properties of stochastic integrals

In (B.7) and (B.9) and similar expressions a term of the form $\int_t^{t'} \sigma_u \, dz_u$ appears. An integral of this type is called a stochastic integral or an Itô integral. We will only consider stochastic integrals where the “integrator” $z$ is a standard Brownian motion, although stochastic integrals involving more general processes can also be defined. For given $t < t'$, the stochastic integral $\int_t^{t'} \sigma_u \, dz_u$ is a random variable. Assuming that $\sigma_u$ is known at time $u$, the value of the integral becomes known at time $t'$. The process $\sigma$ is called the integrand.

The stochastic integral can be defined for very general integrands. The simplest integrands are those that are piecewise constant. Assume that there are points in time $t_0 < t_1 < \cdots < t_n \equiv t'$, so that $\sigma_u$ is constant on each subinterval $[t_i, t_{i+1})$. The stochastic integral is then defined by

$$\int_t^{t'} \sigma_u \, dz_u = \sum_{i=0}^{n-1} \sigma_{t_i} \left( z_{t_{i+1}} - z_{t_i} \right).$$

If the integrand process $\sigma$ is not piecewise constant, a sequence of piecewise constant processes $\sigma^{(1)}, \sigma^{(2)}, \ldots$ exists, which converges to $\sigma$. For each of the processes $\sigma^{(m)}$, the integral $\int_t^{t'} \sigma^{(m)}_u \, dz_u$ is defined as above. The integral $\int_t^{t'} \sigma_u \, dz_u$ is then defined as a limit of the integrals of the
Theorem B.2. If \( \sigma = (\sigma_t) \) satisfies some regularity conditions, the stochastic integral \( \int_t^{t'} \sigma_u \, dz_u \) has the following properties:

\[
E_t \left[ \int_t^{t'} \sigma_u \, dz_u \right] = 0,
\]

\[
\text{Var}_t \left[ \int_t^{t'} \sigma_u \, dz_u \right] = \int_t^{t'} E_t[\sigma_u^2] \, du.
\]

**Proof.** Suppose that \( \sigma \) is piecewise constant and divide the interval \( [t, t'] \) into subintervals defined by the time points \( t = t_0 < t_1 < \cdots < t_n = t' \) so that \( \sigma \) is constant on each subinterval \( [t_i, t_{i+1}) \) with a value \( \sigma_i \), which is known at time \( t_i \). Then

\[
E_t \left[ \int_t^{t'} \sigma_u \, dz_u \right] = \sum_{i=0}^{n-1} E_t [\sigma_i (z_{t_{i+1}} - z_{t_i})] = \sum_{i=0}^{n-1} E_t [\sigma_i, E_{t_i} (z_{t_{i+1}} - z_{t_i})] = 0,
\]

using the Law of Iterated Expectations. For the variance we have

\[
\text{Var}_t \left[ \int_t^{t'} \sigma_u \, dz_u \right] = E_t \left[ \left( \int_t^{t'} \sigma_u \, dz_u \right)^2 \right] - \left( E_t \left[ \int_t^{t'} \sigma_u \, dz_u \right] \right)^2 = E_t \left[ \left( \int_t^{t'} \sigma_u \, dz_u \right)^2 \right]
\]

and

\[
E_t \left[ \left( \int_t^{t'} \sigma_u \, dz_u \right)^2 \right] = E_t \left[ \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \sigma_i \sigma_j (z_{t_{i+1}} - z_{t_i})(z_{t_{j+1}} - z_{t_j}) \right] = \sum_{i=0}^{n-1} E_t [\sigma_i^2 (z_{t_{i+1}} - z_{t_i})^2] = \sum_{i=0}^{n-1} E_t [\sigma_i^2] (t_{i+1} - t_i) = \int_t^{t'} E_t[\sigma_u^2] \, du.
\]

If \( \sigma \) is not piecewise constant, we can approximate it by a piecewise constant process and take appropriate limits. We skip the details. \( \square \)

If the integrand is a deterministic function of time, \( \sigma(u) \), the integral will be normally distributed, so that the following result holds:

**Theorem B.3.** If \( \sigma(u) \) is a deterministic function of time, the random variable \( \int_t^{t'} \sigma(u) \, dz_u \) is normally distributed with mean zero and variance \( \int_t^{t'} \sigma(u)^2 \, du \).

**Proof.** We present a sketch of the proof. Dividing the interval \( [t, t'] \) into subintervals defined by the time points \( t = t_0 < t_1 < \cdots < t_n = t' \), we can approximate the integral with a sum,

\[
\int_t^{t'} \sigma(u) \, dz_u \approx \sum_{i=0}^{n-1} \sigma(t_i) (z_{t_{i+1}} - z_{t_i}).
\]
The increment of the Brownian motion over any subinterval is normally distributed with mean zero and a variance equal to the length of the subinterval. Furthermore, the different terms in the sum are mutually independent. It is well-known that a sum of normally distributed random variables is itself normally distributed, and that the mean of the sum is equal to the sum of the means, which in the present case yields zero. Due to the independence of the terms in the sum, the variance of the sum is also equal to the sum of the variances, i.e.,

$$\text{Var}_t \left( \sum_{i=0}^{n-1} \sigma(t_i) \left( z_{t_{i+1}} - z_{t_i} \right) \right) = \sum_{i=0}^{n-1} \sigma(t_i)^2 \text{Var}_t \left( z_{t_{i+1}} - z_{t_i} \right) = \sum_{i=0}^{n-1} \sigma(t_i)^2 (t_{i+1} - t_i),$$

which is an approximation of the integral $\int_t^{t'} \sigma(u)^2 \, du$. The result now follows from an appropriate limit where the subintervals shrink to zero length.

Note that the process $y = (y_t)_{t \geq 0}$ defined by $y_t = \int_0^t \sigma_u \, dz_u$ is a martingale (under regularity conditions on $\sigma$), since

$$E_t[y_t'] = E_t \left[ \int_0^{t'} \sigma_u \, dz_u \right] = E_t \left[ \int_0^t \sigma_u \, dz_u + \int_t^{t'} \sigma_u \, dz_u \right] = E_t \left[ \int_0^t \sigma_u \, dz_u \right] + E_t \left[ \int_t^{t'} \sigma_u \, dz_u \right] = \int_0^t \sigma_u \, dz_u = y_t,$$

so that the expected future value is equal to the current value. More generally $y_t = y_0 + \int_0^t \sigma_u \, dz_u$ for some constant $y_0$, is a martingale. The converse is also true in the sense that any martingale can be expressed as a stochastic integral. This is the so-called martingale representation theorem:

**Theorem B.4.** Suppose the process $M = (M_t)$ is a martingale with respect to a filtered probability space implicitly defined by the standard Brownian motion $z = (z_t)_{t \in [0,T]}$ so that, in particular, the information filtration is $\mathcal{F} = \mathcal{F}^z$. Then a unique adapted process $\theta = (\theta_t)$ exists such that

$$M_t = M_0 + \int_0^t \theta_u \, dz_u$$

for all $t$.

For a mathematically more precise statement of the result and a proof, see Øksendal (2003, Thm. 4.3.4).

### B.6.2 Leibnitz’ rule for stochastic integrals

Leibnitz’ differentiation rule for ordinary integrals is as follows: If $f(t,s)$ is a deterministic function, and we define $Y(t) = \int_t^T f(t,s) \, ds$, then

$$Y'(t) = -f(t,t) + \int_t^T \frac{\partial f}{\partial t}(t,s) \, ds.$$

If we use the notation $Y'(t) = \frac{dY}{dt}$ and $\frac{\partial f}{\partial t} = \frac{df}{dt}$, we can rewrite this result as

$$dY = -f(t,t) \, dt + \left( \int_t^T \frac{df}{dt}(t,s) \, ds \right) \, dt.$$
and formally cancelling the $dt$-terms, we get

$$dY = -f(t, t) dt + \int_t^T df(t, s) ds.$$  

We will now consider a similar result in the case where $f(t, s)$ and, hence, $Y(t)$ are stochastic processes.

**Theorem B.5.** For any processes $\alpha$ and $\beta$ where $\alpha \in [t_0, s]$ be the Itô process defined by the dynamics $df_t^s = \alpha_t^s dt + \beta_t^s dz_t$, where $\alpha$ and $\beta$ are sufficiently well-behaved stochastic processes. Then the dynamics of the stochastic process $Y_t = \int_t^T f_t^s ds$ is given by

$$dY_t = \left[ \left( \int_t^T \alpha_t^s ds \right) - f_t^T \right] dt + \left( \int_t^T \beta_t^s ds \right) dz_t.$$  

Since the result is usually not included in standard textbooks on stochastic calculus, a sketch of the proof is included. The proof applies the generalized Fubini-rule for stochastic processes, which was stated and demonstrated in the appendix of Heath, Jarrow, and Morton (1992). The Fubini-rule says that the order of integration in double integrals can be reversed, if the integrand is a sufficiently well-behaved function – we will assume that this is indeed the case.

**Proof.** Given any arbitrary $t_1 \in [t_0, T]$. Since

$$f_{t_1}^s = f_{t_0}^s + \int_{t_0}^{t_1} \alpha_t^s dt + \int_{t_0}^{t_1} \beta_t^s dz_t,$$  

we get

$$Y_{t_1} = \int_{t_1}^T f_t^s ds + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

$$= Y_{t_0} + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

$$= Y_{t_0} + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

$$= Y_{t_0} + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

$$= Y_{t_0} + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

$$= Y_{t_0} + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

$$= Y_{t_0} + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

$$= Y_{t_0} + \int_{t_0}^{t_1} \left[ \int_t^T \alpha_t^s ds \right] dt + \int_{t_0}^{t_1} \left[ \int_t^T \beta_t^s ds \right] dz_t$$  

where the Fubini-rule was employed in the second and fourth equality. The result now follows from the final expression. \hfill \Box
B.7 Itô’s Lemma

In our dynamic models of the term structure of interest rates, we will take as given a stochastic process for the dynamics of some basic quantity such as the short-term interest rate. Many other quantities of interest will be functions of that basic variable. To determine the dynamics of these other variables, we shall apply Itô’s Lemma, which is basically the chain rule for stochastic processes. We will state the result for a function of a general Itô process, although we will most frequently apply the result for the special case of a function of a diffusion process.

**Theorem B.6.** Let \( x = (x_t)_{t \geq 0} \) be a real-valued Itô process with dynamics

\[ dx_t = \mu_t \, dt + \sigma_t \, dz_t, \]

where \( \mu \) and \( \sigma \) are real-valued processes, and \( z \) is a one-dimensional standard Brownian motion. Let \( g(x,t) \) be a real-valued function which is two times continuously differentiable in \( x \) and continuously differentiable in \( t \). Then the process \( y = (y_t)_{t \geq 0} \) defined by

\[ y_t = g(x_t,t) \]

is an Itô process with dynamics

\[ dy_t = \left( \frac{\partial g}{\partial t}(x_t,t) + \frac{\partial g}{\partial x}(x_t,t) \mu_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(x_t,t) \sigma_t^2 \right) \, dt + \frac{\partial g}{\partial x}(x_t,t) \sigma_t \, dz_t. \]

The proof is based on a Taylor expansion of \( g(x_t,t) \) combined with appropriate limits, but a formal proof is beyond the scope of this book. Once again, we refer to Øksendal (2003, Ch. 4) and similar textbooks. The result can also be written in the following way, which may be easier to remember:

\[ dy_t = \frac{\partial g}{\partial t}(x_t,t) \, dt + \frac{\partial g}{\partial x}(x_t,t) \, dx_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(x_t,t)(dx_t)^2. \] \hspace{1cm} (B.10)

Here, in the computation of \( (dx_t)^2 \), one must apply the rules \( (dt)^2 = dt \cdot dz_t = 0 \) and \( (dz_t)^2 = dt \), so that

\[ (dx_t)^2 = (\mu_t \, dt + \sigma_t \, dz_t)^2 = \mu_t^2(dt)^2 + 2 \mu_t \sigma_t \, dt \cdot dz_t + \sigma_t^2(dz_t)^2 = \sigma_t^2 dt. \]

The intuition behind these rules is as follows: When \( dt \) is close to zero, \( (dt)^2 \) is far less than \( dt \) and can therefore be ignored. Since \( dz_t \sim N(0,dt) \), we get \( E[dt \cdot dz_t] = dt \cdot E[dz_t] = 0 \) and \( \text{Var}[dt \cdot dz_t] = (dt)^2 \text{Var}[dz_t] = (dt)^3 \), which is also very small compared to \( dt \) and is therefore ignorable. Finally, we have \( E[(dz_t)^2] = \text{Var}[dz_t] = (E[dz_t])^2 = dt \), and it can be shown that \( \text{Var}[(dz_t)^2] = 2(dt)^2 \). For \( dt \) close to zero, the variance is therefore much less than the mean, so \( (dz_t)^2 \) can be approximated by its mean \( dt \).

In standard mathematics, the differential of a function \( y = g(x,t) \) where \( x \) and \( t \) are real variables is defined as \( dy = \frac{\partial g}{\partial t} \, dt + \frac{\partial g}{\partial x} \, dx \). When \( x \) is an Itô process, (B.10) shows that we have to add a second-order term.

In Section B.8, we give examples of the application of Itô’s Lemma, which is used extensively in modern continuous-time finance.

---

3This is based on the computation \( \text{Var}[(z_{t+\Delta t}-z_t)^2] = E[(z_{t+\Delta t}-z_t)^4] - (E[(z_{t+\Delta t}-z_t)^2])^2 = 3(\Delta t)^2 - (\Delta t)^2 = 2(\Delta t)^2 \) and a passage to the limit.
Figure B.4: Simulation of a geometric Brownian motion with initial value \( x_0 = 100 \), relative drift rate \( \mu = 0.1 \), and a relative volatility of \( \sigma = 0.2 \) and \( \sigma = 0.5 \), respectively. The smooth curve shows the trend corresponding to \( \sigma = 0 \). The simulations are based on 200 subintervals of equal length, and the same sequence of random numbers has been used for the two \( \sigma \)-values.

B.8 Important diffusion processes

In this section we will discuss particular examples of diffusion processes that are frequently applied in modern financial models, as those we consider in the following chapters.

B.8.1 Geometric Brownian motions

A stochastic process \( x = (x_t)_{t \geq 0} \) is said to be a geometric Brownian motion if it is a solution to the stochastic differential equation

\[
\frac{dx_t}{x_t} = \mu dt + \sigma d z_t,
\]

where \( \mu \) and \( \sigma \) are constants. The initial value for the process is assumed to be positive, \( x_0 > 0 \).

A geometric Brownian motion is the particular diffusion process that is obtained from (B.6) by inserting \( \mu(x_t, t) = \mu x_t \) and \( \sigma(x_t, t) = \sigma x_t \). Paths can be simulated by computing

\[
x_{t_i} = x_{t_{i-1}} + \mu x_{t_{i-1}} (t_i - t_{i-1}) + \sigma x_{t_{i-1}} \varepsilon_i \sqrt{t_i - t_{i-1}}.
\]

Figure B.4 shows a single simulated path for \( \sigma = 0.2 \) and a path for \( \sigma = 0.5 \). For both paths we have used \( \mu = 0.1 \) and \( x_0 = 100 \), and the same sequence of random numbers.

The expression (B.11) can be rewritten as

\[
\frac{dx_t}{x_t} = \mu dt + \sigma dz_t,
\]

which is the relative (percentage) change in the value of the process over the next infinitesimally short time interval \([t, t + dt]\). If \( x_t \) is the price of a traded asset, then \( dx_t/x_t \) is the rate of return.
on the asset over the next instant. The constant $\mu$ is the expected rate of return per period, while
$\sigma$ is the standard deviation of the rate of return per period. In this context it is often $\mu$ which is
called the drift (rather than $\mu x_t$) and $\sigma$ which is called the volatility (rather than $\sigma x_t$). Strictly
speaking, one must distinguish between the relative drift and volatility ($\mu$ and $\sigma$, respectively) and
the absolute drift and volatility ($\mu x_t$ and $\sigma x_t$, respectively). An asset with a constant expected
rate of return and a constant relative volatility has a price that follows a geometric Brownian
motion. For example, such an assumption is used for the stock price in the famous Black-Scholes-
Merton model for stock option pricing and a geometric Brownian motion is also used to describe
the evolution in the short-term interest rate in some models of the term structure of interest rate,

Next, we will find an explicit expression for $x_t$, i.e., we will find a solution to the stochastic
differential equation (B.11). We can then also determine the distribution of the future value
$y_t$ of the process. We apply Itô’s Lemma with the function $dy_t = \ln x$ and define the process
$y_t = g(x_t, t) = \ln x_t$. Since
\[
\frac{\partial g}{\partial t}(x_t, t) = 0, \quad \frac{\partial g}{\partial x}(x_t, t) = \frac{1}{x_t}, \quad \frac{\partial^2 g}{\partial x^2}(x_t, t) = -\frac{1}{x_t^2},
\]
we get from Theorem B.6 that
\[
dy_t = \left(0 + \frac{1}{x_t} \mu x_t - \frac{1}{2} \frac{1}{x_t^2} \sigma^2 x_t^2\right) dt + \frac{1}{x_t} \sigma x_t dz_t
= \left(\mu - \frac{1}{2} \sigma^2\right) dt + \sigma dz_t.
\]
Hence, the process $y_t = \ln x_t$ is a generalized Brownian motion. In particular, we have
\[
y_{t'} - y_t = \left(\mu - \frac{1}{2} \sigma^2\right) (t' - t) + \sigma (z_{t'} - z_t),
\]
which implies that
\[
\ln x_{t'} = \ln x_t + \left(\mu - \frac{1}{2} \sigma^2\right) (t' - t) + \sigma (z_{t'} - z_t).
\]
Taking exponentials on both sides, we get
\[
x_{t'} = x_t \exp \left\{\left(\mu - \frac{1}{2} \sigma^2\right) (t' - t) + \sigma (z_{t'} - z_t)\right\}. \quad \text{(B.12)}
\]
This is true for all $t' > t \geq 0$. In particular,
\[
x_t = x_0 \exp \left\{\left(\mu - \frac{1}{2} \sigma^2\right) t + \sigma z_t\right\}.
\]
Since exponentials are always positive, we see that $x_t$ can only have positive values, so that the
value space of a geometric Brownian motion is $S = (0, \infty)$.

Suppose now that we stand at time $t$ and have observed the current value $x_t$ of a geometric
Brownian motion. Which probability distribution is then appropriate for the uncertain future
value, say at time $t'$? Since $z_{t'} - z_t \sim N(0, t' - t)$, we see from (B.12) that the future value $x_{t'}$
(given $x_t$) will be lognormally distributed. The probability density function for $x_{t'}$ (given $x_t$) is
\[
f(x) = \frac{1}{x \sqrt{2\pi \sigma^2 (t' - t)}} \exp \left\{-\frac{1}{2\sigma^2 (t' - t)} \left(\ln \frac{x}{x_t}\right)^2 - \left(\mu - \frac{1}{2} \sigma^2\right) (t' - t)\right\}, \quad x > 0,
\]
and the mean and variance are
\[
E_t[x_{t'}] = x_t e^{\mu (t' - t)},
\]
\[
\text{Var}_t[x_{t'}] = x_t^2 e^{2\mu (t' - t)} \left[e^{\sigma^2 (t' - t)} - 1\right].
\]
cf. Appendix A.

The geometric Brownian motion in (B.11) is time-homogeneous, since neither the drift nor the volatility are time-dependent. We will also make use of the time-inhomogeneous variant, which is characterized by the dynamics

$$dx_t = \mu(t)x_t \, dt + \sigma(t)x_t \, d\zeta_t,$$

where $\mu$ and $\sigma$ are deterministic functions of time. Following the same procedure as for the time-homogeneous geometric Brownian motion, one can show that the inhomogeneous variant satisfies

$$x'_t = x_t \exp\left\{ \int_t^{t'} \left( \mu(u) - \frac{1}{2} \sigma(u)^2 \right) \, du \right\}.$$

According to Theorem B.3, $\int_t^{t'} \sigma(u) \, d\zeta_u$ is normally distributed with mean zero and variance $\int_t^{t'} \sigma(u)^2 \, du$. Therefore, the future value of the time-inhomogeneous geometric Brownian motion is also lognormally distributed. In addition, we have

$$\mathbb{E}_t[x_{t'}] = x_t e^{\int_t^{t'} \mu(u) \, du},$$

$$\text{Var}_t[x_{t'}] = x_t^2 e^{2 \int_t^{t'} \mu(u) \, du} \left( e^{\int_t^{t'} \sigma(u)^2 \, du} - 1 \right).$$

### B.8.2 Ornstein-Uhlenbeck processes

Another stochastic process we shall apply in models of the term structure of interest rate is the so-called Ornstein-Uhlenbeck process. A stochastic process $x_t = (x_t)_{t \geq 0}$ is said to be an Ornstein-Uhlenbeck process, if its dynamics is of the form

$$dx_t = [\varphi - \kappa x_t] \, dt + \beta \, d\zeta_t,$$

where $\varphi$, $\beta$, and $\kappa$ are constants with $\kappa > 0$. Alternatively, this can be written as

$$dx_t = \kappa [\theta - x_t] \, dt + \beta \, d\zeta_t,$$

where $\theta = \varphi/\kappa$. An Ornstein-Uhlenbeck process exhibits mean reversion in the sense that the drift is positive when $x_t < \theta$ and negative when $x_t > \theta$. The process is therefore always pulled towards a long-term level of $\theta$. However, the random shock to the process through the term $\beta \, d\zeta_t$ may cause the process to move further away from $\theta$. The parameter $\kappa$ controls the size of the expected adjustment towards the long-term level and is often referred to as the mean reversion parameter or the speed of adjustment.

To determine the distribution of the future value of an Ornstein-Uhlenbeck process we proceed as for the geometric Brownian motion. We will define a new process $y_t$ as some function of $x_t$ such that $y_t = (y_t)_{t \geq 0}$ is a generalized Brownian motion. It turns out that this is satisfied for $y_t = g(x_t, t)$, where $g(x, t) = e^{\kappa t} x$. From Itô’s Lemma we get

$$dy_t = \left[ \frac{\partial g}{\partial t}(x_t, t) + \frac{\partial g}{\partial x}(x_t, t)(\varphi - \kappa x_t) + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(x_t, t) \beta^2 \right] \, dt + \frac{\partial g}{\partial x}(x_t, t) \beta \, d\zeta_t$$

$$= \left[ \kappa e^{\kappa t} x_t + e^{\kappa t} (\varphi - \kappa x_t) \right] \, dt + e^{\kappa t} \beta \, d\zeta_t$$

$$= \varphi e^{\kappa t} \, dt + \beta e^{\kappa t} \, d\zeta_t.$$
This implies that
\[ y_{t'} = y_t + \int_t^{t'} \varphi \kappa u \, du + \int_t^{t'} \beta \kappa u \, dz_u. \]

After substitution of the definition of \( y_t \) and \( y_{t'} \) and a multiplication by \( e^{-\kappa t'} \), we arrive at the expression
\[
x_{t'} = e^{-\kappa(t'-t)} x_t + \int_t^{t'} \varphi e^{-\kappa(t'-u)} \, du + \int_t^{t'} \beta e^{-\kappa(t'-u)} \, dz_u \\
= e^{-\kappa(t'-t)} x_t + \theta \left(1 - e^{-\kappa(t'-t)}\right) + \int_t^{t'} \beta e^{-\kappa(t'-u)} \, dz_u.
\]

This holds for all \( t' > t \geq 0 \). In particular, we get that the solution to the stochastic differential equation (B.13) can be written as
\[
x_t = e^{-\kappa t} x_0 + \theta \left(1 - e^{-\kappa t}\right) + \int_0^t \beta e^{-\kappa(t-u)} \, dz_u.
\]

According to Theorem B.3, the integral \( \int_t^{t'} \beta e^{-\kappa(t'-u)} \, dz_u \) is normally distributed with mean zero and variance \( \int_t^{t'} \beta^2 e^{-2\kappa(t'-u)} \, du = \frac{\beta^2}{2\kappa} \left(1 - e^{-2\kappa(t'-t)}\right) \). We can thus conclude that \( x_{t'} \) (given \( x_t \)) is normally distributed, with mean and variance given by
\[
E_t[x_{t'}] = e^{-\kappa(t'-t)} x_t + \theta \left(1 - e^{-\kappa(t'-t)}\right), \\
\text{Var}_t[x_{t'}] = \frac{\beta^2}{2\kappa} \left(1 - e^{-2\kappa(t'-t)}\right).
\]

The value space of an Ornstein-Uhlenbeck process is \( \mathbb{R} \). For \( t' \to \infty \), the mean approaches \( \theta \), and the variance approaches \( \beta^2/(2\kappa) \). For \( \kappa \to 0 \), the mean approaches the current value \( x_t \), and the variance approaches \( \beta^2(t' - t) \). The distance between the level of the process and the long-term level is expected to be halved over a period of \( t' - t = (\ln 2)/\kappa \), since \( E_t[x_{t'}] - \theta = \frac{1}{2} (x_t - \theta) \) implies that \( e^{-\kappa(t'-t)} = \frac{1}{2} \) and, hence, \( t' - t = (\ln 2)/\kappa \).

The effect of the different parameters can also be evaluated by looking at the paths of the process, which can be simulated by
\[
x_{t_i} = x_{t_{i-1}} + \kappa [\theta - x_{t_{i-1}}] (t_i - t_{i-1}) + \beta \varepsilon_i \sqrt{t_i - t_{i-1}}.
\]

Figure B.5 shows a single path for different combinations of \( x_0, \kappa, \theta, \) and \( \beta \). In each sub-figure one of the parameters is varied and the others fixed. The base values of the parameters are \( x_0 = 0.08 \), \( \theta = 0.08 \), \( \kappa = \ln 2 \approx 0.69 \), and \( \beta = 0.03 \). All paths are computed using the same sequence of random numbers \( \varepsilon_1, \ldots, \varepsilon_n \) and are therefore directly comparable. None of the paths shown involve negative values of the process, but other paths will (see Figure B.6). As a matter of fact, it can be shown that an Ornstein-Uhlenbeck process with probability one will sooner or later become negative.

We will also apply the time-inhomogeneous Ornstein-Uhlenbeck process, where the constants \( \varphi \) and \( \beta \) are replaced by deterministic functions:
\[
dx_t = [\varphi(t) - \kappa x_t] \, dt + \beta(t) \, dz_t = \kappa [\theta(t) - x_t] \, dt + \beta(t) \, dz_t.
\]
Figure B.5: Simulated paths for an Ornstein-Uhlenbeck process. The basic parameter values are $x_0 = \theta = 0.08$, $\kappa = \ln 2 \approx 0.69$, and $\beta = 0.03$.

Following the same line of analysis as above, it can be shown that the future value $x_{t'}$ given $x_t$ is normally distributed with mean and variance given by

$$E_t[x_{t'}] = e^{-\kappa(t'-t)}x_t + \int_t^{t'} \varphi(u)e^{-\kappa(t'-u)} \, du,$$

$$\text{Var}_t[x_{t'}] = \int_t^{t'} \beta(u)^2e^{-2\kappa(t'-u)} \, du.$$  

One can also allow $\kappa$ to depend on time, but we will not make use of that extension.

One of the earliest (but still frequently applied) dynamic models of the term structure of interest rates, the Vasicek model, is based on the assumption that the short-term interest rate follows an Ornstein-Uhlenbeck process; see Section 10.2. In an extension of that model, the short-term interest rate is assumed to follow a time-inhomogeneous Ornstein-Uhlenbeck process.

### B.8.3 Square-root processes

Another stochastic process frequently applied in term structure models is the so-called square-root process. A one-dimensional stochastic process $x = (x_t)_{t \geq 0}$ is said to be a square-root process if...
process, if its dynamics is of the form

\[ dx_t = [\varphi - \kappa x_t] \, dt + \beta \sqrt{x_t} \, dz_t = \kappa [\theta - x_t] \, dt + \beta \sqrt{x_t} \, dz_t, \]  

(B.16)

where \( \varphi = \kappa \theta \). Here, \( \varphi, \theta, \beta, \) and \( \kappa \) are positive constants. We assume that the initial value of the process \( x_0 \) is positive, so that the square root function can be applied. The only difference to the dynamics of an Ornstein-Uhlenbeck process is the term \( \sqrt{x_t} \) in the volatility. The variance rate is now \( \beta^2 x_t \) which is proportional to the level of the process. A square-root process also exhibits mean reversion.

A square-root process can only take on non-negative values. To see this, note that if the value should become zero, then the drift is positive and the volatility zero, and therefore the value of the process will with certainty become positive immediately after (zero is a so-called reflecting barrier).

It can be shown that if \( 2\varphi \geq \beta^2 \), the positive drift at low values of the process is so big relative to the volatility that the process cannot even reach zero, but stays strictly positive.\(^4\) Hence, the value space for a square-root process is either \( S = [0, \infty) \) or \( S = (0, \infty) \).

Paths for the square-root process can be simulated by successively calculating

\[ x_{t_i} = x_{t_{i-1}} + \kappa [\theta - x_{t_{i-1}}] (t_i - t_{i-1}) + \beta \sqrt{x_{t_{i-1}}} \varepsilon_i \sqrt{t_i - t_{i-1}}. \]

Variations in the different parameters will have similar effects as for the Ornstein-Uhlenbeck process, which is illustrated in Figure B.5. Instead, let us compare the paths for a square-root process and an Ornstein-Uhlenbeck process using the same drift parameters \( \kappa \) and \( \theta \), but where the \( \beta \)-parameter for the Ornstein-Uhlenbeck process is set equal to the \( \beta \)-parameter for the square-root process multiplied by the square root of \( \theta \), which ensures that the processes will have the same variance rate at the long-term level. Figure B.6 compares two pairs of paths of the processes. In part (a), the initial value is set equal to the long-term level, and the two paths continue to be very close to each other. In part (b), the initial value is lower than the long-term level, so that the variance rates of the two processes differ from the beginning. For the given sequence of random numbers, the Ornstein-Uhlenbeck process becomes negative, while the square-root process of course stays positive. In this case there is a clear difference between the paths of the two processes.

Since a square-root process cannot become negative, the future values of the process cannot be normally distributed. In order to find the actual distribution, let us try the same trick as for the Ornstein-Uhlenbeck process, that is we look at \( y_t = e^{\kappa t} x_t \). By Itô’s Lemma,

\[ dy_t = \kappa e^{\kappa t} x_t \, dt + e^{\kappa t} (\varphi - \kappa x_t) \, dt + e^{\kappa t} \beta \sqrt{x_t} \, dz_t \]

\[ = \varphi e^{\kappa t} \, dt + \beta e^{\kappa t} \sqrt{x_t} \, dz_t, \]

so that

\[ y_{t'} = y_t + \int_t^{t'} \varphi e^{\kappa u} \, du + \int_t^{t'} \beta e^{\kappa u} \sqrt{x_u} \, dz_u. \]

Computing the ordinary integral and substituting the definition of \( y \), we get

\[ x_{t'} = x_t e^{-\kappa (t' - t)} + \theta \left( 1 - e^{-\kappa (t' - t)} \right) + \beta \int_t^{t'} e^{-\kappa (t' - u)} \sqrt{x_u} \, dz_u. \]

\(^4\)To show this, the results of Karlin and Taylor (1981, p. 226ff) can be applied.
For $\kappa$ and the variance approaches $\theta \beta$ variance is more complicated for the square-root process. For $t$.

Note that the mean is identical to the mean for an Ornstein-Uhlenbeck process, whereas the fact that the stochastic integral has mean zero, cf. Theorem B.2, we easily get

$$E_t[x_{t'}] = e^{-\kappa (t'-t)} x_t + \theta \left(1 - e^{-\kappa (t'-t)}\right) = \theta + (x_t - \theta) e^{-\kappa (t'-t)}. $$

To compute the variance we apply the second equation of Theorem B.2:

$$\text{Var}_t[x_{t'}] = \text{Var}_t \left[ \beta \int_t^{t'} e^{-\kappa (t'-u)} \sqrt{T_u} \, du \right]$$

$$= \beta^2 \int_t^{t'} e^{-2\kappa (t'-u)} E_t[x_u] \, du$$

$$= \beta^2 \int_t^{t'} e^{-2\kappa (t'-u)} \left(\theta + (x_t - \theta) e^{-\kappa (u-t)}\right) \, du$$

$$= \beta^2 \theta \int_t^{t'} e^{-2\kappa (t'-u)} \, du + \beta^2 (x_t - \theta) e^{-2\kappa (t'-t)} \int_t^{t'} e^{\kappa u} \, du$$

$$= \frac{\beta^2 \theta}{2\kappa} \left(1 - e^{-2\kappa (t'-t)}\right) + \frac{\beta^2}{\kappa} (x_t - \theta) \left(e^{-\kappa (t'-t)} - e^{-2\kappa (t'-t)}\right)$$

$$= \frac{\beta^2 x_t}{\kappa} \left(e^{-\kappa (t'-t)} - e^{-2\kappa (t'-t)}\right) + \frac{\beta^2 \theta}{2\kappa} \left(1 - e^{-\kappa (t'-t)}\right)^2.$$

Note that the mean is identical to the mean for an Ornstein-Uhlenbeck process, whereas the variance is more complicated for the square-root process. For $t' \to \infty$, the mean approaches $\theta$, and the variance approaches $\theta \beta^2 / (2\kappa)$. For $\kappa \to \infty$, the mean approaches $\theta$, and the variance approaches 0. For $\kappa \to 0$, the mean approaches the current value $x_t$, and the variance approaches $\beta^2 x_t (t'-t)$.

It can be shown that, conditional on the value $x_t$, the value $x_{t'}$ with $t' > t$ is given by the non-central $\chi^2$-distribution. A non-central $\chi^2$-distribution is characterized by a number $a$ of degrees

(a) Initial value $x_0 = 0.08$, same random numbers as in Figure B.5

(b) Initial value $x_0 = 0.06$, different random numbers

Figure B.6: A comparison of simulated paths for an Ornstein-Uhlenbeck process and a square-root process. For both processes, the parameters $\theta = 0.08$ and $\kappa = \ln 2 \approx 0.69$ are used, while $\beta$ is set to 0.03 for the Ornstein-Uhlenbeck process and to $0.03/\sqrt{0.08} \approx 0.1061$ for the square-root process.
of freedom and a non-centrality parameter $b$ and is denoted by $\chi^2(a, b)$. More precisely, the distribution of $x_{t'}$ given $x_t$ is identical to the distribution of the random variable $Y/c(t' - t)$ where $c$ is the deterministic function

$$c(\tau) = \frac{4\kappa}{\beta^2 (1 - e^{-\kappa\tau})}$$

and $Y$ is a $\chi^2(a, b(t' - t))$-distributed random variable with

$$a = \frac{4\varphi}{\beta^2}, \quad b(\tau) = x_t c(\tau) e^{-\kappa\tau}.$$ 

The density function for a $\chi^2(a, b)$-distributed random variable is

$$f_{\chi^2(a, b)}(y) = \sum_{i=0}^{\infty} \frac{e^{-b/2} (b/2)^i}{i!} f_{\chi^2(a+2i)}(y) = \sum_{i=0}^{\infty} \frac{e^{-b/2} (b/2)^i (1/2)^{i+a/2}}{\Gamma(i + a/2)} y^{i-1+a/2} e^{-y/2},$$

where $f_{\chi^2(a+2i)}$ is the density function for a central $\chi^2$-distribution with $a + 2i$ degrees of freedom.

Inserting this density in the first sum will give the second sum. Here $\Gamma$ denotes the so-called gamma-function defined as $\Gamma(m) = \int_0^\infty x^{m-1} e^{-x} dx$. The probability density function for the value of $x_{t'}$ conditional on $x_t$ is then

$$f(x) = c(t' - t) f_{\chi^2(a,b(t'-t))}(c(t' - t)x).$$

The mean and variance of a $\chi^2(a, b)$-distributed random variable are $a + b$ and $2(a + b)$, respectively. This opens another way of deriving the mean and variance of $x_{t'}$ given $x_t$. We leave it for the reader to verify that this procedure will yield the same results as given above.

A frequently applied dynamic model of the term structure of interest rates is based on the assumption that the short-term interest rate follows a square-root process, cf. Section 10.3. Since interest rates are positive and empirically seem to have a variance rate which is positively correlated to the interest rate level, the square-root process gives a more realistic description of interest rates than the Ornstein-Uhlenbeck process. On the other hand, models based on square-root processes are more complicated to analyze than models based on Ornstein-Uhlenbeck processes.

### B.9 Multi-dimensional processes

So far we have only considered one-dimensional processes, i.e., processes with a value space which is $\mathbb{R}$ or a subset of $\mathbb{R}$. In many cases we want to keep track of several processes, e.g., price processes for different assets, and we will often be interested in covariances and correlations between different processes.

In a continuous-time model where the exogenous shock process $z = (z_t)_{t \in [0,T]}$ is one-dimensional, the instantaneous increments of any two processes will be perfectly correlated. For example, if we consider the two Itô processes $x$ and $y$ defined by

$$dx_t = \mu_{xt} dt + \sigma_{xt} dz_t, \quad dy_t = \mu_{yt} dt + \sigma_{yt} dz_t,$$

then $\text{Cov}_t[dx_t, dy_t] = \sigma_{xt} \sigma_{yt} dt$ so that the instantaneous correlation becomes

$$\text{Corr}_t[dx_t, dy_t] = \frac{\text{Cov}_t[dx_t, dy_t]}{\sqrt{\text{Var}_t[dx_t] \text{Var}_t[dy_t]}} = \frac{\sigma_{xt} \sigma_{yt} dt}{\sqrt{\sigma_{xt}^2 dt \sigma_{yt}^2 dt}} = 1.$$
Increments over any non-infinitesimal time interval are generally not perfectly correlated, i.e., for any \( h > 0 \) a correlation like \( \text{Corr}_t[x_{t+h} - x_t, y_{t+h} - y_t] \) is typically different from one but close to one for small \( h \).

To obtain non-perfectly correlated changes over the shortest time period considered by the model we need an exogenous shock of a dimension higher than one, i.e., a shock vector. One can without loss of generality assume that the different components of this shock vector are mutually independent and generate non-perfect correlations between the relevant processes by varying the sensitivities of those processes towards the different exogenous shocks. We will first consider the case of two processes and later generalize.

### B.9.1 Two-dimensional processes

In the example above, we can avoid the perfect correlation by introducing a second standard Brownian motion so that

\[
\begin{align*}
    dx_t &= \mu_x t\, dt + \sigma_{x1t}\, dz_{1t} + \sigma_{x2t}\, dz_{2t}, \\
    dy_t &= \mu_y t\, dt + \sigma_{y1t}\, dz_{1t} + \sigma_{y2t}\, dz_{2t},
\end{align*}
\]

where \( z_1 = (z_{1t}) \) and \( z_2 = (z_{2t}) \) are independent standard Brownian motions. This generates an instantaneous covariance of \( \text{Cov}_t[dx_t, dy_t] = (\sigma_{x1t}\sigma_{y1t} + \sigma_{x2t}\sigma_{y2t})\, dt \), instantaneous variances of \( \text{Var}_t[dx_t] = (\sigma_{x1t}^2 + \sigma_{x2t}^2)\, dt \) and \( \text{Var}_t[dy_t] = (\sigma_{y1t}^2 + \sigma_{y2t}^2)\, dt \), and thus an instantaneous correlation of

\[
    \text{Corr}_t[dx_t, dy_t] = \frac{\sigma_{x1t}\sigma_{y1t} + \sigma_{x2t}\sigma_{y2t}}{\sqrt{(\sigma_{x1t}^2 + \sigma_{x2t}^2)(\sigma_{y1t}^2 + \sigma_{y2t}^2)}},
\]

which again can be anywhere in the interval \([-1, +1]\).

The shock coefficients \( \sigma_{x1t}, \sigma_{x2t}, \sigma_{y1t}, \) and \( \sigma_{y2t} \) are determining the two instantaneous variances and the instantaneous correlation. But many combinations of the four shock coefficients will give rise to the same variances and correlation. We have one degree of freedom in fixing the shock coefficients. For example, we can put \( \sigma_{x2t} \equiv 0 \), which has the nice implication that it will simplify various expressions and interpretations. If we thus write the dynamics of \( x \) and \( y \) as

\[
\begin{align*}
    dx_t &= \mu_x t\, dt + \sigma_{x1t}\, dz_{1t}, \\
    dy_t &= \mu_y t\, dt + \sigma_{y1t}\, dz_{1t} + \sqrt{1 - \rho_t^2} d z_{2t},
\end{align*}
\]

\( \sigma_{x1t}^2 \) and \( \sigma_{y1t}^2 \) are the variance rates of \( x_t \) and \( y_t \), respectively, while the covariance is \( \text{Cov}_t[dx_t, dy_t] = \rho_t \sigma_{x1t} \sigma_{y1t} \). If \( \sigma_{x1t} \) and \( \sigma_{y1t} \) are both positive, then \( \rho_t \) will be the instantaneous correlation between the two processes \( x \) and \( y \).

In many continuous-time models, one stochastic process is defined in terms of a function of two other, not necessarily perfectly correlated, stochastic processes. For that purpose we need the following two-dimensional version of Itô’s Lemma.

**Theorem B.7.** Suppose \( x = (x_t) \) and \( y = (y_t) \) are two stochastic processes with dynamics

\[
\begin{align*}
    dx_t &= \mu_x t\, dt + \sigma_{x1t}\, dz_{1t} + \sigma_{x2t}\, dz_{2t}, \\
    dy_t &= \mu_y t\, dt + \sigma_{y1t}\, dz_{1t} + \sigma_{y2t}\, dz_{2t},
\end{align*}
\]

where \( z_1 = (z_{1t}) \) and \( z_2 = (z_{2t}) \) are independent standard Brownian motions. Let \( g(x, y, t) \) be a real-valued function for which all the derivatives \( \frac{\partial g}{\partial t}, \frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial^2 g}{\partial x^2}, \frac{\partial^2 g}{\partial y^2}, \) and \( \frac{\partial^2 g}{\partial x \partial y} \) exist and are
continuous. Then the process \( W = (W_t) \) defined by \( W_t = g(x_t, y_t, t) \) is an Itô process with

\[
dW_t = \left( \frac{\partial g}{\partial t} + \frac{\partial g}{\partial x} \mu_x t + \frac{\partial g}{\partial y} \mu_y t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} \left( \sigma^2_x t + \sigma^2_x 2t \right) + \frac{1}{2} \frac{\partial^2 g}{\partial y^2} \left( \sigma^2_y t + \sigma^2_y 2t \right) \right) \, dt \\
+ \left( \frac{\partial g}{\partial x} \sigma_x t + \frac{\partial g}{\partial y} \sigma_y t \right) \, dz_t \\
+ \left( \frac{\partial g}{\partial x} \sigma_x 2t + \frac{\partial g}{\partial y} \sigma_y 2t \right) \, dz_{2t},
\]

where the dependence of all the partial derivatives on \((x_t, y_t, t)\) has been notationally suppressed.

Alternatively, the result can be written more compactly as

\[
dW_t = \frac{\partial g}{\partial t} \, dt + \frac{\partial g}{\partial x} \, dx_t + \frac{\partial g}{\partial y} \, dy_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2} (dx_t)^2 + \frac{1}{2} \frac{\partial^2 g}{\partial y^2} (dy_t)^2 + \frac{\partial^2 g}{\partial x \partial y} (dx_t)(dy_t),
\]

where it is understood that \((dt)^2 = dt \cdot dz_{1t} = dt \cdot dz_{2t} = dz_{1t} \cdot dz_{2t} = 0.\)

**Example B.1.** Suppose that the dynamics of \( x \) and \( y \) are given by (B.17) and \( W_t = x_t y_t \). In order to find the dynamics of \( W \), we apply the above version of Itô’s Lemma with the function \( g(x, y) = xy \). The relevant partial derivatives are

\[
\frac{\partial g}{\partial t} = 0, \quad \frac{\partial g}{\partial x} = y, \quad \frac{\partial g}{\partial y} = x, \quad \frac{\partial^2 g}{\partial x^2} = 0, \quad \frac{\partial^2 g}{\partial y^2} = 0, \quad \frac{\partial^2 g}{\partial x \partial y} = 1.
\]

Hence,

\[
dW_t = y_t \, dx_t + x_t \, dy_t + (dx_t)(dy_t).
\]

In particular, if the dynamics of \( x \) and \( y \) are written on the form

\[
dx_t = x_t \left[ m_x t \, dt + v_{x 1t} \, dz_{1t} + v_{x 2t} \, dz_{2t} \right], \quad dy_t = y_t \left[ m_y t \, dt + v_{y 1t} \, dz_{1t} + v_{y 2t} \, dz_{2t} \right], \quad (B.18)
\]

we get

\[
dW_t = W_t \left[ (m_x t + m_y t + v_{x 1t} v_{y 1t} + v_{x 2t} v_{y 2t}) \, dt + (v_{x 1t} + v_{y 1t}) \, dz_{1t} + (v_{x 2t} + v_{y 2t}) \, dz_{2t} \right].
\]

For the special case, where both \( x \) and \( y \) are geometric Brownian motion so that \( m_x t, m_y, v_{x 1t}, v_{x 2t}, v_{y 1t}, \text{ and } v_{y 2t} \) are all constants, it follows that \( W_t = x_t y_t \) is also a geometric Brownian motion. \( \square \)

**Example B.2.** Define \( W_t = x_t / y_t \). In this case we need to apply Itô’s Lemma with the function \( g(x, y) = x / y \) which has derivatives

\[
\frac{\partial g}{\partial t} = 0, \quad \frac{\partial g}{\partial x} = \frac{1}{y}, \quad \frac{\partial g}{\partial y} = -\frac{x}{y^2}, \quad \frac{\partial^2 g}{\partial x^2} = 0, \quad \frac{\partial^2 g}{\partial y^2} = -2 \frac{x}{y^3}, \quad \frac{\partial^2 g}{\partial x \partial y} = -\frac{1}{y^2}.
\]

Then

\[
dW_t = \frac{1}{y_t} \, dx_t - \frac{x_t}{y_t^2} \, dy_t + \frac{x_t}{y_t^4} (dy_t)^2 - \frac{1}{y_t^3} (dx_t)(dy_t)
\]

\[
= W_t \left[ \frac{dx_t}{x_t} - \frac{dy_t}{y_t} + \left( \frac{dy_t}{y_t} \right)^2 - \frac{dx_t \, dy_t}{x_t \, y_t} \right].
\]

In particular, if the dynamics of \( x \) and \( y \) are given by (B.18), the dynamics of \( W_t = x_t / y_t \) becomes

\[
dW_t = W_t \left[ (m_x t - m_y t + (v_{x 1t}^2 + v_{y 2t}^2) - (v_{x 1t} v_{y 1t} + v_{x 2t} v_{y 2t})) \, dt \\
+ (v_{x 1t} - v_{y 1t}) \, dz_{1t} + (v_{x 2t} - v_{y 2t}) \, dz_{2t} \right].
\]
Note that for the special case, where both \( x \) and \( y \) are geometric Brownian motions, \( W = x/y \) is also a geometric Brownian motion.

We can apply the two-dimensional version of Itô’s Lemma to prove the following useful result relating expected discounted values and the drift rate.

**Theorem B.8.** Under suitable regularity conditions, the relative drift rate of an Itô process \( x = (x_t) \) is given by the process \( m = (m_t) \) if and only if \( x_t = E_t[x_T \exp\{- \int^T_t m_s \, ds\}] \).

**Proof.** Suppose first that the relative drift rate is given by \( m \) so that \( dx_t = x_t[m_t \, dt + v_t \, dz_t] \). Let us use Itô’s Lemma to identify the dynamics of the process \( W_t = x_t \exp\{- \int^t_0 m_s \, ds\} \) or \( W_t = x_t y_t \), where \( y_t = \exp\{- \int^t_0 m_s \, ds\} \). Note that \( dy_t = -y_t m_t \, dt \) so that \( y \) is a locally deterministic stochastic process. From Example B.1, the dynamics of \( W \) becomes

\[
\frac{dW_t}{W_t} = (m_t - m_t + 0) \, dt + v_t \, dz_t, \quad W_t = W_t \exp\{- \int^t_0 m_s \, ds\}.
\]

Since \( W \) has zero drift, it is a martingale. It follows that \( W_t = E_t[W_T] \), i.e., \( x_t \exp\{- \int^t_0 m_s \, ds\} = E_t[x_T \exp\{- \int^t T m_s \, ds\}] \) and hence \( x_t = E_t[x_T \exp\{- \int^t T m_s \, ds\}] \).

If, on the other hand, \( x_t = E_t[x_T \exp\{- \int^T_t m_s \, ds\}] \) for all \( t \), then the absolute drift of \( x \) follows from this computation:

\[
\frac{1}{\Delta t} E_t[x_{t+\Delta t} - x_t] = \frac{1}{\Delta t} E_t\left[\left(E_{t+\Delta t}\left[x_T e^{-\int^T_{t+\Delta t} m_s \, ds}\right]\right) - \left(E_t\left[x_T e^{-\int^T_t m_s \, ds}\right]\right)\right] = E_t\left[x_T e^{-\int^T_t m_s \, ds} e^{\int^T_t m_s \, ds} \right] = E_t\left[x_T e^{-\int^T_t m_s \, ds} \right] = m_t x_t,
\]

so that the relative drift rate equals \( m_t \).

**B.9.2 \( K \)-dimensional processes**

Simultaneously modeling the dynamics of a lot of economic quantities requires the use of a lot of shocks to those quantities. For that purpose we will work with represent shocks to the economy by a vector standard Brownian motion. We define this below and state Itô’s Lemma for processes of a general dimension.

A **\( K \)-dimensional standard Brownian motion** \( z = (z_1, \ldots, z_K)^T \) is a stochastic process for which the individual components \( z_i \) are mutually independent one-dimensional standard Brownian motions. If we let \( 0 = (0, \ldots, 0)^T \) denote the zero vector in \( \mathbb{R}^K \) and let \( I_K \) denote the identity matrix of dimension \( K \times K \) (the matrix with ones in the diagonal and zeros in all other entries), then we can write the defining properties of a \( K \)-dimensional Brownian motion \( z \) as follows:

1. \( z_0 = 0 \),
2. for all \( t, t' \geq 0 \) with \( t < t' \): \( z_{t'} - z_t \sim N(0, (t' - t)I_K) \) [normally distributed increments],
3. for all \( 0 \leq t_0 < t_1 < \cdots < t_n \), the random variables \( z_{t_0} - z_0, \ldots, z_{t_n} - z_{t_{n-1}} \) are mutually independent [independent increments],
(iv) $z$ has continuous sample paths in $\mathbb{R}^K$.

Here, $N(\mathbf{a}, \mathbf{b})$ denotes a $K$-dimensional normal distribution with mean vector $\mathbf{a}$ and variance-covariance matrix $\mathbf{b}$.

A **$K$-dimensional diffusion process** $\mathbf{x} = (x_1, \ldots, x_K)^T$ is a process with increments of the form

$$dx_t = \mathbf{\mu}(x_t, t) \, dt + \mathbf{\sigma}(x_t, t) \, dz_t,$$

where $\mathbf{\mu}$ is a function from $\mathbb{R}^K \times \mathbb{R}_+$ into $\mathbb{R}^K$, and $\mathbf{\sigma}$ is a function from $\mathbb{R}^K \times \mathbb{R}_+$ into the space of $K \times K$-matrices. As before, $z$ is a $K$-dimensional standard Brownian motion. The evolution of the multi-dimensional diffusion can also be written componentwise as

$$dx_{it} = \mu_i(x_t, t) \, dt + \mathbf{\sigma}_i(x_t, t)^T \, dz_t$$

$$= \mu_i(x_t, t) \, dt + \sum_{k=1}^K \sigma_{ik}(x_t, t) \, dz_{kt}, \quad i = 1, \ldots, K,$$

where $\mathbf{\sigma}_i(x_t, t)^T$ is the $i$'th row of the matrix $\mathbf{\sigma}(x_t, t)$, and $\sigma_{ik}(x_t, t)$ is the $(i, k)$'th entry (i.e., the entry in row $i$, column $k$). Since $dz_{kt}, \ldots, dz_{Kt}$ are mutually independent and all $N(0, dt)$ distributed, the expected change in the $i$'th component process over an infinitesimal period is

$$E_i[dx_{it}] = \mu_i(x_t, t) \, dt, \quad i = 1, \ldots, K,$$

so that $\mu_i$ can be interpreted as the drift of the $i$'th component. Furthermore, the covariance between changes in the $i$'th and the $j$'th component processes over an infinitesimal period becomes

$$\text{Cov}_i[dx_{it}, dx_{jt}] = \text{Cov}_i \left[ \sum_{k=1}^K \sigma_{ik}(x_t, t) \, dz_{kt}, \sum_{l=1}^K \sigma_{jl}(x_t, t) \, dz_{lt} \right]$$

$$= \sum_{k=1}^K \sum_{l=1}^K \sigma_{ik}(x_t, t) \sigma_{jl}(x_t, t) \text{Cov}_i[dz_{kt}, dz_{lt}]$$

$$= \sum_{k=1}^K \sigma_{ik}(x_t, t) \sigma_{jk}(x_t, t) \, dt$$

$$= \mathbf{\sigma}_i(x_t, t)^T \mathbf{\sigma}_j(x_t, t) \, dt, \quad i, j = 1, \ldots, K,$$

where we have applied the usual rules for covariances and the independence of the components of $z$. In particular, the variance of the change in the $i$'th component process of an infinitesimal period is given by

$$\text{Var}_i[dx_{it}] = \text{Cov}_i[dx_{it}, dx_{it}] = \sum_{k=1}^K \sigma_{ik}(x_t, t)^2 \, dt = \|\mathbf{\sigma}_i(x_t, t)\|^2 \, dt, \quad i = 1, \ldots, K.$$

The volatility of the $i$'th component is given by $\|\mathbf{\sigma}_i(x_t, t)\|$. The variance-covariance matrix of changes of $x_t$ over the next instant is $\sum (x_t, t) \, dt = \mathbf{\sigma}(x_t, t) \mathbf{\sigma}(x_t, t)^T \, dt$. The correlation between instantaneous increments in two component processes is

$$\text{Corr}_i[dx_{it}, dx_{jt}] = \frac{\mathbf{\sigma}_i(x_t, t)^T \mathbf{\sigma}_j(x_t, t) \, dt}{\sqrt{\|\mathbf{\sigma}_i(x_t, t)\|^2 \, dt \|\mathbf{\sigma}_j(x_t, t)\|^2 \, dt}} = \frac{\mathbf{\sigma}_i(x_t, t)^T \mathbf{\sigma}_j(x_t, t)}{\|\mathbf{\sigma}_i(x_t, t)\| \|\mathbf{\sigma}_j(x_t, t)\|},$$

which can be any number in $[-1, 1]$ depending on the elements of $\mathbf{\sigma}_i$ and $\mathbf{\sigma}_j$. 

B.9 Multi-dimensional processes 253
Similarly, we can define a $K$-dimensional Itô process $x = (x_1, \ldots, x_K)^\top$ to be a process with increments of the form

$$dx_t = \mu_t \, dt + \sigma_t \, dz_t,$$

where $\mu = (\mu_i)$ is a $K$-dimensional stochastic process and $\sigma = (\sigma_i)$ is a stochastic process with values in the space of $K \times K$-matrices.

Next, we state a multi-dimensional version of Itô's Lemma, where a one-dimensional process is defined as a function of time and a multi-dimensional process.

**Theorem B.9.** Let $x = (x_t)_{t \geq 0}$ be an Itô process in $\mathbb{R}^K$ with dynamics $dx_t = \mu_t \, dt + \sigma_t \, dz_t$, and the variance-covariance function $\Sigma$ is a stochastic process with values in the space of $K \times K$-matrices.

Next, we state a multi-dimensional version of Itô’s Lemma, where a one-dimensional process is defined as a function of time and a multi-dimensional process.

**Theorem B.9.** Let $x = (x_t)_{t \geq 0}$ be an Itô process in $\mathbb{R}^K$ with dynamics $dx_t = \mu_t \, dt + \sigma_t \, dz_t$, or, equivalently,

$$dx_t = \mu_t \, dt + \sigma_t \, dz_t = \mu_t \, dt + \sum_{k=1}^K \sigma_{ik} \, dz_{kt}, \quad i = 1, \ldots, K,$$

where $z_1, \ldots, z_K$ are independent standard Brownian motions, and $\mu_t$ and $\sigma_{ik}$ are well-behaved stochastic processes.

Let $g(x, t)$ be a real-valued function for which all the derivatives $\frac{\partial g}{\partial x}, \frac{\partial g}{\partial x_i},$ and $\frac{\partial^2 g}{\partial x_i \partial x_j}$ exist and are continuous. Then the process $y = (y_t)_{t \geq 0}$ defined by $y_t = g(x_t, t)$ is also an Itô process with dynamics

$$dy_t = \left( \frac{\partial g}{\partial t} (x_t, t) + \sum_{i=1}^K \frac{\partial g}{\partial x_i} (x_t, t) \mu_i + \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \frac{\partial^2 g}{\partial x_i \partial x_j} (x_t, t) \gamma_{ij} \right) dt + \sum_{i=1}^K \frac{\partial g}{\partial x_i} (x_t, t) \sigma_{iK} \, dz_{Kt},$$

where $\gamma_{ij} = \sigma_{i1} \sigma_{j1} + \cdots + \sigma_{iK} \sigma_{jK}$ is the covariance between the processes $x_i$ and $x_j$.

The result can also be written as

$$dy_t = \frac{\partial g}{\partial t} (x_t, t) \, dt + \sum_{i=1}^K \frac{\partial g}{\partial x_i} (x_t, t) \, dx_{it} + \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \frac{\partial^2 g}{\partial x_i \partial x_j} (x_t, t) (dx_{it}) (dx_{jt}),$$

where in the computation of $(dx_{it}) (dx_{jt})$ one must use the rules $(dt)^2 = dt \cdot dz_{it} = 0$ for all $i, j \neq i$, and $(dz_{it})^2 = dt$ for all $i$. Alternatively, the result can be expressed using vector and matrix notation:

$$dy_t = \left( \frac{\partial g}{\partial t} (x_t, t) + \left( \frac{\partial g}{\partial x_i} (x_t, t) \right)^\top \mu_t + \frac{1}{2} \text{tr} \left( \Sigma_t \left( \frac{\partial^2 g}{\partial x_i \partial x_j} (x_t, t) \right) \Sigma_t \right) \right) dt + \left( \frac{\partial g}{\partial x} (x_t, t) \right)^\top \sigma_t \, dz_t,$$

where

$$\frac{\partial g}{\partial x} (x_t, t) = \begin{pmatrix} \frac{\partial g}{\partial x_1} (x_t, t) \\ \vdots \\ \frac{\partial g}{\partial x_K} (x_t, t) \end{pmatrix}, \quad \frac{\partial^2 g}{\partial x_i \partial x_j} (x_t, t) = \begin{pmatrix} \frac{\partial^2 g}{\partial x_1^2} (x_t, t) & \frac{\partial^2 g}{\partial x_1 x_2} (x_t, t) & \cdots & \frac{\partial^2 g}{\partial x_1 x_K} (x_t, t) \\ \frac{\partial^2 g}{\partial x_2 x_1} (x_t, t) & \frac{\partial^2 g}{\partial x_2^2} (x_t, t) & \cdots & \frac{\partial^2 g}{\partial x_2 x_K} (x_t, t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial x_K x_1} (x_t, t) & \frac{\partial^2 g}{\partial x_K x_2} (x_t, t) & \cdots & \frac{\partial^2 g}{\partial x_K^2} (x_t, t) \end{pmatrix},$$

and $\text{tr}$ denotes the trace of a quadratic matrix, i.e., the sum of the diagonal elements. For example,

$$\text{tr}(A) = \sum_{i=1}^K A_{ii}.$$

The probabilistic properties of a $K$-dimensional diffusion process is completely specified by the drift function $\mu$ and the variance-covariance function $\Sigma$. The values of the variance-covariance matrix are

$$K \times K$$

matrices.
function are symmetric and positive-definite matrices. Above we had $\Sigma = \sigma \sigma^\top$ for a general $(K \times K)$-matrix $\sigma$. But from linear algebra it is well-known that a symmetric and positive-definite matrix can be written as $\hat{\sigma} \hat{\sigma}^\top$ for a lower-triangular matrix $\hat{\sigma}$, i.e., a matrix with $\hat{\sigma}_{ik} = 0$ for $k > i$. This is the so-called Cholesky decomposition. Hence, we may write the dynamics as

\[
\begin{align*}
\dot{x}_{1t} &= \mu_1(x_t, t) \ dt + \hat{\sigma}_{11}(x_t, t) \ dz_{1t} \\
\dot{x}_{2t} &= \mu_2(x_t, t) \ dt + \hat{\sigma}_{21}(x_t, t) \ dz_{1t} + \hat{\sigma}_{22}(x_t, t) \ dz_{2t} \\
&\vdots \\
\dot{x}_{Kt} &= \mu_K(x_t, t) \ dt + \hat{\sigma}_{K1}(x_t, t) \ dz_{1t} + \hat{\sigma}_{K2}(x_t, t) \ dz_{2t} + \cdots + \hat{\sigma}_{KK}(x_t, t) \ dz_{Kt}.
\end{align*}
\]

We can think of building up the model by starting with $x_1$. The shocks to $x_1$ are represented by the standard Brownian motion $z_1$ and its coefficient $\hat{\sigma}_{11}$ is the volatility of $x_1$. Then we extend the model to include $x_2$. Unless the infinitesimal changes to $x_1$ and $x_2$ are always perfectly correlated we need to introduce another standard Brownian motion, $z_2$. The coefficient $\hat{\sigma}_{21}$ is fixed to match the covariance between changes to $x_1$ and $x_2$ and then $\hat{\sigma}_{22}$ can be chosen so that $\sqrt{\hat{\sigma}_{21}^2 + \hat{\sigma}_{22}^2}$ equals the volatility of $x_2$. The model may be extended to include additional processes in the same manner.

Some authors prefer to write the dynamics in an alternative way with a single standard Brownian motion $\hat{z}_i$ for each component $x_i$ such as

\[
\begin{align*}
\dot{x}_{1t} &= \mu_1(x_t, t) \ dt + V_1(x_t, t) \ d\hat{z}_{1t} \\
\dot{x}_{2t} &= \mu_2(x_t, t) \ dt + V_2(x_t, t) \ d\hat{z}_{2t} \\
&\vdots \\
\dot{x}_{Kt} &= \mu_K(x_t, t) \ dt + V_K(x_t, t) \ d\hat{z}_{Kt}.
\end{align*}
\]

Clearly, the coefficient $V_i(x_t, t)$ is then the volatility of $x_i$. To capture an instantaneous non-zero correlation between the different components the standard Brownian motions $\hat{z}_1, \ldots, \hat{z}_K$ have to be mutually correlated. Let $\rho_{ij}$ be the correlation between $\hat{z}_i$ and $\hat{z}_j$. If (B.20) and (B.19) are meant to represent the same dynamics, we must have

\[
\begin{align*}
V_i &= \sqrt{\hat{\sigma}_{i1}^2 + \cdots + \hat{\sigma}_{ii}^2}, \quad i = 1, \ldots, K, \\
\rho_{ii} &= 1; \quad \rho_{ij} = \frac{\sum_{k=1}^{i} \hat{\sigma}_{ik} \hat{\sigma}_{jk}}{V_i V_j}, \quad \rho_{ji} = \rho_{ij}, \quad i < j.
\end{align*}
\]

### B.10 Change of probability measure

When we represent the evolution of a given economic variable by a stochastic process and discuss the distributional properties of this process, we have implicitly fixed a probability measure $\mathbb{P}$. For example, when we use the square-root process $x = (x_t)$ in (B.16) for the dynamics of a particular interest rate, we have taken as given a probability measure $\mathbb{P}$ under which the stochastic process $z = (z_t)$ is a standard Brownian motion. Since the process $x$ is presumably meant to represent the uncertain dynamics of the interest rate in the world we live in, we refer to the measure $\mathbb{P}$ as the real-world probability measure. Of course, it is the real-world dynamics and distributional properties of economic variables that we are ultimately interested in. Nevertheless, it turns out that in order to compute and understand prices and rates it is often convenient to look at the dynamics and
distributional properties of these variables assuming that the world was different from the world we live in. The prime example is a hypothetical world in which investors are assumed to be risk-neutral instead of risk-averse. Loosely speaking, a different world is represented mathematically by a different probability measure. Hence, we need to be able to analyze stochastic variables and processes under different probability measures. In this section we will briefly discuss how we can change the probability measure.

Consider first a state space with finitely many elements, \( \Omega = \{ \omega_1, \ldots, \omega_n \} \). As before, the set of events, i.e., subsets of \( \Omega \), that can be assigned a probability is denoted by \( \mathcal{F} \). Let us assume that the single-element sets \( \{ \omega_i \}, i = 1, \ldots, n \), belong to \( \mathcal{F} \). In this case we can represent a probability measure \( \mathbb{P} \) by a vector \( (p_1, \ldots, p_n) \) of probabilities assigned to each of the individual elements:

\[
p_i = \mathbb{P}(\{\omega_i\}), \quad i = 1, \ldots, n.
\]

Of course, we must have that \( p_i \in [0, 1] \) and that \( \sum_{i=1}^n p_i = 1 \). The probability assigned to any other event can be computed from these basic probabilities. For example, the probability of the event \( \{\omega_2, \omega_4\} \) is given by

\[
\mathbb{P}(\{\omega_2, \omega_4\}) = \mathbb{P}(\{\omega_2\} \cup \{\omega_4\}) = \mathbb{P}(\{\omega_2\}) + \mathbb{P}(\{\omega_4\}) = p_2 + p_4.
\]

Another probability measure \( \mathbb{Q} \) on \( \mathcal{F} \) is similarly given by a vector \( (q_1, \ldots, q_n) \) with \( q_i \in [0, 1] \) and \( \sum_{i=1}^n q_i = 1 \). We are only interested in equivalent probability measures. In this setting, the two measures \( \mathbb{P} \) and \( \mathbb{Q} \) will be equivalent whenever \( p_i > 0 \iff q_i > 0 \) for all \( i = 1, \ldots, n \). With a finite state space there is no point in including states that occur with zero probability so we can assume that all \( p_i \), and therefore all \( q_i \), are strictly positive.

We can represent the change of probability measure from \( \mathbb{P} \) to \( \mathbb{Q} \) by the vector \( \xi = (\xi_1, \ldots, \xi_n) \), where

\[
\xi_i = \frac{q_i}{p_i}, \quad i = 1, \ldots, n.
\]

We can think of \( \xi \) as a random variable that will take on the value \( \xi_i \) if the state \( \omega_i \) is realized. Sometimes \( \xi \) is called the Radon-Nikodym derivative of \( \mathbb{Q} \) with respect to \( \mathbb{P} \) and is denoted by \( d\mathbb{Q}/d\mathbb{P} \). Note that \( \xi_i > 0 \) for all \( i \) and that the \( \mathbb{P} \)-expectation of \( \xi = d\mathbb{Q}/d\mathbb{P} \) is

\[
\mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = \mathbb{E}_{\mathbb{P}}[\xi] = \sum_{i=1}^n p_i \xi_i = \sum_{i=1}^n p_i q_i = \sum_{i=1}^n q_i = 1.
\]

Consider a random variable \( x \) that takes on the value \( x_i \) if state \( i \) is realized. The expected value of \( x \) under the measure \( \mathbb{Q} \) is given by

\[
\mathbb{E}_{\mathbb{Q}}[x] = \sum_{i=1}^n q_i x_i = \sum_{i=1}^n p_i q_i x_i = \sum_{i=1}^n p_i \xi_i x_i = \mathbb{E}_{\mathbb{P}}[\xi x].
\]

Now let us consider the case where the state space \( \Omega \) is infinite. Also in this case the change from a probability measure \( \mathbb{P} \) to an equivalent probability measure \( \mathbb{Q} \) is represented by a strictly positive random variable \( \xi = d\mathbb{Q}/d\mathbb{P} \) with \( \mathbb{E}_{\mathbb{P}}[\xi] = 1 \). Again the expected value under the measure \( \mathbb{Q} \) of a random variable \( x \) is given by \( \mathbb{E}_{\mathbb{Q}}[x] = \mathbb{E}_{\mathbb{P}}[\xi x] \), since

\[
\mathbb{E}_{\mathbb{Q}}[x] = \int_{\Omega} x \, d\mathbb{Q} = \int_{\Omega} x \frac{d\mathbb{Q}}{d\mathbb{P}} \, d\mathbb{P} = \int_{\Omega} x \xi \, d\mathbb{P} = \mathbb{E}_{\mathbb{P}}[\xi x].
\]
In our economic models we will model the dynamics of uncertain objects over some time span 
[0, T]. For example, we might be interested in determining bond prices with maturities up to T years. Then we are interested in the stochastic process on this time interval, i.e., x = (x_t)_{t \in [0, T]}.

The state space Ω is the set of possible paths of the relevant processes over the period [0, T] so that all the relevant uncertainty has been resolved at time T and the values of all relevant random variables will be known at time T. The Radon-Nikodym derivative ξ = dQ/dP is also a random variable and is therefore known at time T and usually not before time T. To indicate this the Radon-Nikodym derivative is often denoted by ξ_T = \frac{dQ}{dP}.

We can define a stochastic process ξ = (ξ_t)_{t \in [0, T]} by setting

$$ξ_t = E^P_t \left[ dQ \right] = E^P_t [ξ_T].$$

This definition is consistent with ξ_T being identical to dQ/dP, since all uncertainty is resolved at time T so that the time T expectation of any variable is just equal to the variable. Note that the process ξ is a P-martingale, since for any t < t' ≤ T we have

$$E^P_t [ξ_{t'}] = E^P_t [E^P_{t'} [ξ_T]] = E^P_t [ξ_T] = ξ_t.$$

Here the first and the third equalities follow from the definition of ξ. The second equality follows from the Law of Iterated Expectations, Theorem B.1. The following result turns out to be very useful in our dynamic models of the economy. Let x = (x_t)_{t \in [0, T]} be any stochastic process. Then we have

$$E^Q_t [x_{t'}] = E^P_t \left[ \frac{ξ_{t'}}{ξ_t} x_{t'} \right].$$

(B.21)

This is called Bayes’ Formula. For a proof, see Björk (2009, Prop. B.41).

Suppose that the underlying uncertainty is represented by a standard Brownian motion z = (z_t) (under the real-world probability measure P), as will be the case in all the models we will consider. Let λ = (λ_t)_{t \in [0, T]} be any sufficiently well-behaved stochastic process. Here, z and λ must have the same dimension. For notational simplicity, we assume in the following that they are one-dimensional, but the results generalize naturally to the multi-dimensional case. We can generate an equivalent probability measure Q^λ in the following way. Define the process ξ^λ = (ξ^λ_t)_{t \in [0, T]} by

$$ξ^λ_t = \exp \left\{ - \int_0^t λ_s dz_s - \frac{1}{2} \int_0^t λ^2_s ds \right\}. \quad (B.22)$$

Then ξ^λ_0 = 1, ξ^λ is strictly positive, and it can be shown that ξ^λ is a P-martingale (see Exercise B.6) so that E^P[ξ^λ_T] = ξ^λ_0 = 1. Consequently, an equivalent probability measure Q^λ can be defined by the Radon-Nikodym derivative

$$\frac{dQ^λ}{dP} = ξ^λ_T = \exp \left\{ - \int_0^T λ_s dz_s - \frac{1}{2} \int_0^T λ^2_s ds \right\}.$$

From (B.21), we get that

$$E^Q_t \left[ x_{t'} \right] = E^P_t \left[ \frac{ξ^λ_t}{ξ^λ_{t'}} x_{t'} \right] = E^P_t \left[ x_{t'} \exp \left\{ - \int_t^{t'} λ_s dz_s - \frac{1}{2} \int_t^{t'} λ^2_s ds \right\} \right]$$

for any stochastic process x = (x_t)_{t \in [0, T]}. A central result is Girsanov’s Theorem:

5Basically, λ must be square-integrable in the sense that \int_0^T λ^2_s ds is finite with probability 1 and that λ satisfies Novikov’s condition, i.e., the expectation E^P \left[ \exp \left\{ \frac{1}{2} \int_0^T λ^2_s ds \right\} \right] is finite.
Theorem B.10 (Girsanov). The process $z^\lambda = (z^\lambda_t)_{t \in [0,T]}$ defined by

$$z^\lambda_t = z_t + \int_0^t \lambda_s ds, \quad 0 \leq t \leq T,$$

is a standard Brownian motion under the probability measure $Q^\lambda$. In differential notation,

$$dz^\lambda_t = dz_t + \lambda_t dt.$$

This theorem has the attractive consequence that the effects on a stochastic process of changing the probability measure from $P$ to some $Q^\lambda$ are captured by a simple adjustment of the drift. If $x = (x_t)$ is an Itô process with dynamics

$$dx_t = \mu_t dt + \sigma_t dz_t,$$

then

$$dx_t = \mu_t dt + \sigma_t (dz^\lambda_t - \lambda_t dt) = (\mu_t - \sigma_t \lambda_t) dt + \sigma_t dz^\lambda_t.$$

Hence, $\mu - \sigma \lambda$ is the drift under the probability measure $Q^\lambda$, which is different from the drift under the original measure $P$ unless $\sigma$ or $\lambda$ are identically equal to zero. In contrast, the volatility remains the same as under the original measure.

In many financial models, the relevant change of measure is such that the distribution under $Q^\lambda$ of the future value of the central processes is of the same class as under the original $P$ measure, but with different moments. For example, consider the Ornstein-Uhlenbeck process

$$dx_t = (\varphi - \kappa x_t) dt + \sigma dz_t$$

and perform the change of measure given by a constant $\lambda_t = \lambda$. Then the dynamics of $x$ under the measure $Q^\lambda$ is given by

$$dx_t = (\hat{\varphi} - \kappa x_t) dt + \sigma dz^\lambda_t,$$

where $\hat{\varphi} = \varphi - \sigma \lambda$. Consequently, the future values of $x$ are normally distributed both under $P$ and $Q^\lambda$. From (B.14) and (B.15), we see that the variance of $x_t$ (given $x_0$) is the same under $Q^\lambda$ and $P$, but the expected values will differ (recall that $\theta = \varphi/\kappa$):

$$E^P_t [x_{t'}] = e^{-\kappa(t'-t)} x_t + \frac{\varphi}{\kappa} \left(1 - e^{-\kappa(t'-t)} \right),$$

$$E^Q^\lambda_t [x_{t'}] = e^{-\kappa(t'-t)} x_t + \frac{\hat{\varphi}}{\kappa} \left(1 - e^{-\kappa(t'-t)} \right).$$

However, in general, a shift of probability measure may change not only some or all moments of future values, but also the distributional class.

### B.11 Exercises

**Exercise B.1.** Suppose $x = (x_t)$ is a geometric Brownian motion, $dx_t = \mu x_t dt + \sigma x_t dz_t$. What is the dynamics of the process $y = (y_t)$ defined by $y_t = (x_t)^n$? What can you say about the distribution of future values of the $y$ process?

**Exercise B.2.** Let $y$ be a random variable and define a stochastic process $x = (x_t)$ by $x_t = E_t[y]$. Show that $x$ is a martingale.
**Exercise B.3** ((Adapted from Björk (2009).)) Define the process \( y = (y_t) \) by \( y_t = z^4_t \), where \( z = (z_t) \) is a standard Brownian motion. Find the dynamics of \( y \). Show that

\[
y_t = 6 \int_0^t z^2_s \, ds + 4 \int_0^t z^3_s \, dz_s.
\]

Show that \( E[y_t] \equiv E[z^4_t] = 3t^2 \), where \( E[\cdot] \) denotes the expectation given the information at time 0.

**Exercise B.4** ((Adapted from Björk (2009).)) Define the process \( y = (y_t) \) by \( y_t = e^{az_t} \), where \( a \) is a constant and \( z = (z_t) \) is a standard Brownian motion. Find the dynamics of \( y \). Show that

\[
y_t = 1 + \frac{1}{2} a^2 \int_0^t y_s \, ds + a \int_0^t y_s \, dz_s.
\]

Define \( m(t) = E[y_t] \). Show that \( m \) satisfies the ordinary differential equation

\[
m'(t) = \frac{1}{2} a^2 m(t), \quad m(0) = 1.
\]

Show that \( m(t) = e^{a^2 t/2} \) and conclude that

\[
E[e^{az_t}] = e^{a^2 t/2}.
\]

**Exercise B.5.** Consider the two general stochastic processes \( x_1 = (x_{1t}) \) and \( x_2 = (x_{2t}) \) defined by the dynamics

\[
\begin{align*}
dx_{1t} &= \mu_{1t} \, dt + \sigma_{1t} \, dz_{1t}, \\
dx_{2t} &= \mu_{2t} \, dt + \rho_t \sigma_{2t} \, dz_{1t} + \sqrt{1 - \rho^2_t \sigma^2_{2t}} \, dz_{2t},
\end{align*}
\]

where \( z_1 \) and \( z_2 \) are independent one-dimensional standard Brownian motions. Interpret \( \mu_{1t}, \sigma_{1t}, \) and \( \rho_t \). Define the processes \( y = (y_t) \) and \( w = (w_t) \) by \( y_t = x_{1t} x_{2t} \) and \( w_t = x_{1t} / x_{2t} \). What is the dynamics of \( y \) and \( w \)? Concretize your answer for the special case where \( x_1 \) and \( x_2 \) are geometric Brownian motions with constant correlation, i.e., \( \mu_{1t} = \mu_i x_{1t}, \sigma_{1t} = \sigma_i x_{1t}, \) and \( \rho_t = \rho \) with \( \mu_i, \sigma_i, \) and \( \rho \) being constants.

**Exercise B.6.** Find the dynamics of the process \( \xi^\lambda \) defined in (B.22).
Theorem C.1. The ordinary differential equation

\[ A'(\tau) = a(\tau) - b(\tau)A(\tau), \quad A(0) = 0, \]

has the solution

\[ A(\tau) = \int_0^\tau e^{-\int_u^\tau b(s)\,ds} a(u) \, du. \]

Theorem C.2. If \( b^2 > 4ac \), then the ordinary differential equation

\[ A'(\tau) = a - bA(\tau) + cA(\tau)^2, \quad A(0) = 0, \]

has the solution

\[ A(\tau) = \frac{2a(e^{\nu\tau} - 1)}{(\nu + b)(e^{\nu\tau} - 1) + 2\nu}, \]

where \( \nu = \sqrt{b^2 - 4ac} \). Furthermore, if \( c \neq 0 \),

\[ \int_0^\tau A(u) \, du = \frac{1}{c} \left\{ \frac{1}{2} (\nu + b)\tau + \ln \left( \frac{2\nu}{(\nu + b)(e^{\nu\tau} - 1) + 2\nu} \right) \right\} \]

and

\[ \int_0^\tau A(u)^2 \, du = \text{- ugly expression to be filled in -}. \]

In the special case in which \( c = 0 \), the solution is

\[ A(\tau) = \frac{a}{b} \left( 1 - e^{-b\tau} \right), \]

and

\[ \int_0^\tau A(u) \, du = \frac{1}{b} (a\tau - A(\tau)), \]

\[ \int_0^\tau A(u)^2 \, du = \frac{1}{ab^2} (a^3\tau - A(\tau)) - \frac{1}{2a^2b} A(\tau)^2. \]
Of course, the special case \( c = 0 \) in Theorem C.2 can also be seen as the special case of Theorem C.1 in which \( a \) and \( b \) are constants.

**Theorem C.3.** If \( b^2 > 4ac \), the solution to the system of ordinary differential equations

\[
A_2' (\tau) = a - b A_2 (\tau) + c A_2 (\tau)^2, \quad A_2 (0) = 0,
\]

\[
A_1' (\tau) = d + f A_2 (\tau) - \left( \frac{1}{2} b - c A_2 (\tau) \right) A_1 (\tau), \quad A_1 (0) = 0
\]

is given by

\[
A_2 (\tau) = \frac{2a(e^{\nu \tau} - 1)}{(\nu + b)(e^{\nu \tau} - 1) + 2\nu},
\]

\[
A_1 (\tau) = \frac{d}{a} A_2 (\tau) + \frac{2}{\nu} (db + 2fa) \frac{(e^{\nu/2} - 1)^2}{(\nu + b)(e^{\nu \tau} - 1) + 2\nu} = \left( \frac{d}{a} + \frac{db + 2af}{\nu} \frac{(e^{\nu \tau} - 1)^2}{e^{\nu \tau} - 1} \right) A_2 (\tau),
\]

where \( \nu = \sqrt{b^2 - 4ac} \).

**Proof.** The expression for \( A_2 \) follows from Theorem C.2. From Theorem C.1 we get

\[
A_1 (\tau) = \int_0^\tau e^{-\int_0^x (\frac{b}{2} - ca_2 (s)) ds} (d + f A_2 (u)) du
\]

\[
= \int_0^\tau e^{-\frac{b}{2}(\tau - u) + \int_0^u A_2 (s) ds} (d + f A_2 (u)) du
\]

\[
= \int_0^\tau e^{\frac{b}{2}(\tau - u)} (\nu + b) (e^{\nu u} - 1) + 2\nu \int_0^u e^{\nu u} - 1) + 2\nu (d + f A_2 (u)) du
\]

\[
= \frac{de^{\tau \nu}}{\nu + b} (e^{\nu \tau} - 1) + 2\nu \int_0^\tau ((\nu + b) e^{\tau \nu} + (\nu - b) e^{-\tau \nu}) + \frac{2a f e^{\tau \nu}}{\nu + b} (e^{\nu \tau} - 1) + 2\nu \int_0^\tau (e^{\tau \nu} - e^{-\tau \nu}) du
\]

\[
= \frac{2d}{\nu (\nu + b) (e^{\nu \tau} - 1) + 2\nu} (\nu + b) e^{\nu \tau} - (\nu - b) e^{-\nu \tau} + 2be^{\nu \tau} + \frac{4af}{\nu (\nu + b) (e^{\nu \tau} - 1) + 2\nu} (e^{\nu \tau} - 1)^2
\]

\[
= \frac{2d}{\nu (\nu + b) (e^{\nu \tau} - 1) + 2\nu} + \frac{db + 2af}{\nu (\nu + b) (e^{\nu \tau} - 1) + 2\nu} (e^{\nu \tau} - 1)^2
\]

\[
= \frac{d}{a} A_2 (\tau) + \frac{db + 2af}{\nu} A_2 (\tau) \frac{(e^{\nu \tau} - 1)^2}{e^{\nu \tau} - 1}
\]

\[
= \left( \frac{d}{a} + \frac{db + 2af}{\nu} \frac{(e^{\nu \tau} - 1)^2}{e^{\nu \tau} - 1} \right) A_2 (\tau).
\]


