On some mathematical aspects of dynamic financial analysis

Author(s):
Blum, Peter

Publication Date:
2005

Permanent Link:
https://doi.org/10.3929/ethz-a-004948972

Rights / License:
In Copyright - Non-Commercial Use Permitted
On Some Mathematical Aspects of Dynamic Financial Analysis

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
DOKTOR DER MATHEMATIK

presented by
PETER BLUM
Dipl. Math. ETH
born May 16th, 1970
citizen of Buch & Beggingen, SH

accepted upon the recommendation of
Prof. Dr. Paul Embrechts, examiner
Prof. Dr. Andrew Cairns, co-examiner
Dr. Michel Dacorogna, co-examiner

2005
Seite Leer / Blank leaf
Acknowledgments

First and foremost, I would like to thank my supervisors Prof. Paul Embrechts and Dr. Michel Dacorogna. Their open-mindedness, their enthusiasm and their constant encouragement and support were essential for the progress of this research project at the intersection of academia and industry. Thanks also to Prof. Andrew Cairns for acting as a co-referee and to Mrs. Gabriele Baltes for being a reliable guide through all administrative complications. Moreover, I am grateful to my colleagues, at the Financial Modelling group of Converium as well as at the Mathematics department of ETH, for their support, understanding and encouragement throughout the entire project.

The research work that lead to this thesis took place while I was on a position sponsored and managed by Converium (the former Zurich Re), and the related financial and institutional support is gratefully acknowledged. Thanks are also due to RiskLab for hosting me during the completion of this thesis.

Schliesslich danke ich ganz herzlich meiner Familie für die materielle und moralische Unterstützung, die ich von ihrer Seite über all die Jahre erfahren durfte.
Seite Leer / Blank leaf
Contents

Acknowledgment iii

Abstract xi

Kurzfassung xiii

Introduction xv

I Dynamic Financial Analysis 1

1 An Introduction to DFA 3

1.1 Overview ........................................ 3

1.2 Origins and Value Proposition ................... 5

1.3 Areas of Applications ............................ 6

1.4 A Survey of the Field ............................ 8

2 A Generic DFA System 11

2.1 Overall Structure ............................... 11
# Contents

2.2 Formal Specification .................................. 13

3 The Elements of DFA ..................................... 19

3.1 Scenario Generation ................................... 19

3.2 Company and Strategy Modelling ...................... 25

3.3 Output and Evaluation .................................. 28

3.4 Control and Optimization ............................ 34

4 Assessment and Outlook .................................. 37

4.1 Assessment of Present State .......................... 37

4.2 Outlook on Future Developments .................... 38

II Scenario Generation ..................................... 41

5 Introduction ............................................. 43

5.1 Problem Statement and Requirements ................ 43

5.2 Economic Scenario Generators ........................ 46

5.3 Models for Interest Rates and FX Rates ............. 48

6 Phenomenology .......................................... 53

6.1 Terminology ........................................... 53

6.1.1 Interest Rates ....................................... 53

6.1.2 Foreign Exchange Rates ............................ 55

6.2 A Survey of Stylized Facts ............................ 56

6.3 Evaluation of Data Sets Used ......................... 60
### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.4.2</td>
<td>Interest Rates</td>
<td>118</td>
</tr>
<tr>
<td>9.4.3</td>
<td>Exchange Rates</td>
<td>121</td>
</tr>
<tr>
<td>9.5</td>
<td>The Context of the Model</td>
<td>122</td>
</tr>
<tr>
<td>10</td>
<td>Calibration of Models</td>
<td>127</td>
</tr>
<tr>
<td>10.1</td>
<td>Problem Statement and Basics</td>
<td>127</td>
</tr>
<tr>
<td>10.2</td>
<td>Generalized Method of Moments</td>
<td>130</td>
</tr>
<tr>
<td>10.3</td>
<td>Survey of Alternative Methods</td>
<td>134</td>
</tr>
<tr>
<td>10.4</td>
<td>Moment Conditions</td>
<td>139</td>
</tr>
<tr>
<td>10.5</td>
<td>The Weighting Matrix</td>
<td>141</td>
</tr>
<tr>
<td>10.6</td>
<td>Diagnostics</td>
<td>142</td>
</tr>
<tr>
<td>10.7</td>
<td>Fitting the Initial Yield Curve</td>
<td>143</td>
</tr>
<tr>
<td>10.8</td>
<td>Case Studies</td>
<td>144</td>
</tr>
<tr>
<td>10.8.1</td>
<td>Single-Currency Case – USD</td>
<td>144</td>
</tr>
<tr>
<td>10.8.2</td>
<td>Double-Currency Case – CHF and USD</td>
<td>155</td>
</tr>
<tr>
<td>11</td>
<td>Validation of Models</td>
<td>167</td>
</tr>
<tr>
<td>11.1</td>
<td>Problem Statement</td>
<td>167</td>
</tr>
<tr>
<td>11.2</td>
<td>The PIT: Probability Integral Transform</td>
<td>170</td>
</tr>
<tr>
<td>11.3</td>
<td>PIT: The Non-Parametric Case</td>
<td>174</td>
</tr>
<tr>
<td>11.4</td>
<td>PIT: Evaluation</td>
<td>178</td>
</tr>
<tr>
<td>11.4.1</td>
<td>Parametric Evaluation</td>
<td>179</td>
</tr>
<tr>
<td>11.4.2</td>
<td>Non-parametric Evaluation</td>
<td>186</td>
</tr>
<tr>
<td>11.5</td>
<td>Backtesting: Case Studies</td>
<td>189</td>
</tr>
</tbody>
</table>
11.5.1 Single-Currency Case – USD .......................... 189
11.5.2 Double-Currency Case – CHF and USD .............. 193

12 Case Study – Funds Withheld Asset ............................. 199
12.1 Introduction .................................................. 199
12.2 Some Projections ............................................ 203
12.3 Valuation and Embedded Options ......................... 208
12.4 Currency hedging ............................................. 212

13 Assessment and Outlook ............................................. 219
13.1 Assessment of Present State ................................. 219
13.2 Outlook on Future Developments ............................ 222

Appendices ............................................................... 226

A Statistical Methods ................................................ 227
A.1 Descriptive Statistics and Normality Tests .................. 227
A.2 Intertemporal Dependence Diagnostics ...................... 229

B Numerical Methods ................................................ 233
B.1 Simulating the Driver Process .................................. 233
B.1.1 The Ornstein-Uhlenbeck case .............................. 234
B.1.2 The general case ............................................ 235
B.1.3 Random Number Generation ............................... 237
B.2 Numerical Optimization Issues ................................. 237
<table>
<thead>
<tr>
<th>x</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bibliography</td>
<td>240</td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>251</td>
</tr>
</tbody>
</table>
Abstract

The leitmotiv of this thesis is Dynamic Financial Analysis (DFA), that is, large-scale computer simulation methods for the holistic modelling of insurance business with the aim of finding strategies that have a favourable impact on the overall risk and profitability of a company. Part I provides a comprehensive overview of DFA and its elements in mathematical language, thus identifying the interfaces between DFA and the various disciplines of applied mathematics, in particular, financial and insurance mathematics.

One of the most challenging parts of DFA is the modelling of economic risk factors, in particular, interest rates. Part II provides a model for the full term structures of interest rates of several economies together with the related currency exchange rates, satisfying the very specific requirements posed by DFA. The development includes not only the formulation of the model as such, but an entire, engineering-like process starting with the requirements and including a sound theoretical evaluation of the model, development of methods for its calibration and validation, implementation of all elements and their use in representative practical settings. Special emphasis is put on the use of the model in the realm of market-consistent actuarial valuation. The main aim and achievement of Part II is twofold. On the one hand, it provides a complete, implemented and proven model ready for deployment in practice. On the other hand, since all aspects are fully documented and transparent, this thesis can also serve as an engineering guide for the development of further models.
Kurzfassung


Introduction

The leitmotiv of this thesis is Dynamic Financial Analysis (DFA), that is, large-scale computer simulation methods for the holistic modelling of insurance business with the aim of finding strategies that have a favourable impact on the overall risk and profitability of the company.

The need for DFA was raised by developments in the insurance industry – globalization, deregulation and the ensuing fierce competition – and the development of DFA was mainly driven by the needs of practitioners for appropriate tools for quantitative analysis. Therefore, DFA is mainly an integration platform for a number of techniques from actuarial and financial mathematics. However, it often suffers from the use of techniques inappropriate for the specific task at hand and, in general, from some lack of overall rigour and coherence. This thesis is aimed at addressing these issues in several ways.

Part I provides a comprehensive overview of DFA and its elements in mathematical language. It is a somewhat formalized version of the related article [14] in the Encyclopedia of Actuarial Science [117]. This part does not contain mathematical achievements in the strict sense, but it identifies the interfaces of DFA with several disciplines of applied mathematics, thus providing guidance to the researcher wishing to contribute to DFA and to the practitioner in need for new modelling elements.

From among the elements of DFA, the most important one for the ap-
plied mathematician is the *scenario generator* that produces Monte Carlo scenarios for a large number of dependent risk factors and for long time horizons. Economic variables have a special status due to their intricate relations and the need for the absence of arbitrage. Therefore, the related scenario generation is usually bundled in an *Economic Scenario Generator (ESG)*.

The most difficult element within an ESG is the model for the *term structure of interest rates*. Many standard models from mathematical finance fail to satisfy the DFA-specific requirements like positive interest rates, long-term stability and faithful reproduction of empirical features. The task becomes even more demanding if interest rates for several economies, together with the currency exchange rates, have to be modelled simultaneously. Moreover, in the advent of new financial reporting standards (IFRS) and new regulations (Solvency II), the demand for *market consistent valuation* becomes paramount, and an interest rate model has to cater for this demand as well.

Part II deals with the development of an interest rate and exchange rate model consistent with these requirements. Development means not only the formulation of the model as such, but an entire, engineering-like process starting with the requirements and including a sound theoretical evaluation of the model, development of methods for its calibration and validation, implementation of all elements and their use in representative practical settings.

The main aim and achievement of Part II is twofold. On the one hand, it provides a complete, implemented and proven model ready for deployment in practice. On the other hand, since all aspects are fully documented and transparent, this thesis can also serve as an engineering guide for the development of further models.

Part II has become more voluminous than desirable. This is partly due to the presence of many introduction, survey and excursion sections that grew along the research path. Moreover, it unites a large number of ele-
ments from different areas, with the related need to introduce everything rigorously. As the research was driven by practical needs, some of the elements used in Part II are more or less known. But, in order to achieve a seamlessly integrated system, these elements had to be adapted and combined in a meaningful way, and new elements were created where necessary. For these reasons, it may be difficult to extract details of the scientific contribution out of the text as it stands. As a partial remedy, the remainder of this introduction provides a reading guide to Part II.

Starting point is a review of the requirements and the motivation of the price kernel model in Section 5.3. Then, Section 6.3 verifies the presence of the stylized empirical properties of interest rates and FX rates in the data sample used later on.

Section 7.2 re-develops the basically well-known arbitrage theory of the price kernel, but in a formalism adapted to the purpose. The really interesting part of Chapter 7 is Section 7.3 which provides a comprehensive assessment of the capabilities and limitations of the price kernel in the realm of market-consistent valuation. Moreover, Section 7.3 also introduces an algorithm for the Monte Carlo valuation of fairly general American options.

Section 8.4 develops a new, simple and workable time-inhomogeneous extension of the well-known potential approach introduced earlier in Chapter 8.

The exponential-quadratic model with an underlying multivariate Ornstein-Uhlenbeck process is constructed in Sections 9.2 and 9.3. The presentation there is more detailed than in the literature and works out some new aspects. Section 9.4 provides a generic representation of the model, allowing to obtain the well-known properties plus a few new ones.

Calibration by Generalized Method of Moments (GMM), adapted to the particularities of the model, is introduced in Sections 10.2, 10.5 and 10.6.
Section 11.3 develops some results that allow to use the Probability Integral Transform (PIT) method, reviewed in Section 11.2, in situations where only Monte Carlo scenarios are available. Section 11.4 then refines a non-parametric approach and implements and investigates a recent idea in the literature on copula-based parametric evaluations of PIT output.

Everything investigated and developed in the aforementioned sections was fully implemented in Matlab, and the usefulness of the models and methods in practical settings is proved in a number of case studies:

In Section 10.8, a single- and a double-currency version of the exponential-quadratic model are calibrated to real-world data, including the use of a genetic algorithm. The properties of the calibrated models are discussed and practical experience is summarized. In Section 11.5, the calibrated models are evaluated by using the PIT and the related parametric and non-parametric evaluation methods.

Finally, the double-currency model is used for the valuation of an international reinsurance deal involving both underwriting and financial risk. Section 12.2 features some purely actuarial investigations, whereas Section 12.3 is an example of market-consistent valuation in the presence of embedded options; the type of evaluation that will keep actuaries busy for several years to come.
Part I

Dynamic Financial Analysis
Chapter 1

An Introduction to DFA

The purpose of this chapter is to give a non-formal overview of Dynamic Financial Analysis (DFA) in order to set the scene for the subsequent chapters of Part I that describe DFA in a more formal way and put it into the context of the various mathematical disciplines involved. This chapter, as well as the non-technical parts of the subsequent chapters of Part I, are quite closely based on the related article [14] in the Encyclopedia of Actuarial Science [117].

1.1 Overview

Dynamic Financial Analysis (DFA) is a systematic approach based on large-scale computer simulations for the integrated financial modelling of insurance and reinsurance companies aimed at assessing the risks and the benefits associated with strategic decisions. Originally, the term was mainly used in non-life insurance, but meanwhile it has also become popular in life insurance.

The most important characteristic of DFA is that it takes an integrated, holistic point of view – contrary to classic financial or actuarial analysis
where different aspects of one company were considered in isolation from each other. Specifically, DFA models the reactions of the company in response to a large number of interrelated risk factors including both underwriting risks – usually from several different lines of business – as well as investment risks. In order to account for the long time horizons that are typical in insurance and reinsurance, DFA allows to make dynamic projections for several time periods into the future, where one time period is usually one year, sometimes also one quarter. DFA models normally reflect the full financial structure of the modelled company, including the impact of accounting and tax structures. Thus, DFA allows to make projections for the balance sheet and the profit-and-loss account (P&L) of the company.

Technically, DFA is a platform using various models and techniques from finance and actuarial science by integrating them into one multivariate dynamic simulation model. Given the complexity and the long time horizons of such a model, it is not anymore possible to make analytical evaluations. Therefore, DFA is based on stochastic simulation (also called Monte Carlo), where large numbers of random scenarios are generated, the reaction of the company on each one of the scenarios is evaluated, and the resulting outcomes are then analyzed statistically. Chapter 2 provides an overall view of a generic DFA system and identifies its elements, and Chapter 3 gives an in-depth description of these elements.

With this setup, DFA provides insight into the sources of creation and destruction of value and into the impact of external risk factors as well as internal strategic decisions on the bottom line of the company, i.e. on its financial statements. An important virtue of DFA is that it allows to gain insight into various kinds of dependencies that affect the company and that would be hard to grasp without a holistic approach. Thus, DFA is a tool for integrated risk management and strategic decision support. Section 1.2 describes the problem space that gave rise to the genesis of DFA, and Section 1.3 provides more information on the uses of DFA.
The aim of this section is to describe the developments in the insurance and reinsurance market that gave rise to the genesis of DFA. For a long time – up until the 1980's or 1990's, depending on the country – insurance business used to be a fairly quiet area, characterized by little strategic flexibility and innovation. Regulations heavily constrained the insurers in the types of business they could assume, and also in the way they had to do the business. Relatively simple products were predominant, each one addressing a specific type of risk, and underwriting and investment were separated, within the (non-life) insurance companies themselves and also in the products they offered to their clients. In this rather static environment, there was no particular need for sophisticated analytics: actuarial analysis was carried out on the underwriting side – without linkage to the investment side of the company which was analyzed separately. Reinsurance as the only means of managing underwriting risks was acquired locally per line of business, whereas there were separate hedging activities for financial risks. Basically, quantitative analysis amounted to modelling a group of isolated silos, without taking a holistic view.

However, insurance business is no longer a quiet area. Regulations were loosened and gave more strategic flexibility to the insurers, leading to new types of complicated products and to fierce competition in the market. The traditional separation between banking and insurance business became increasingly blurred, and many companies developed into integrated financial service providers through mergers and acquisitions. Moreover, the risk landscape was also changing, due to demographic, social and political changes, and due to new types of insured risks or changes in the characteristics of already-insured risks (e.g. liability). The boom in the financial markets in the late 1990's also affected the insurers. On the one hand, it opened up opportunities on the investment side. On the other hand, insurers themselves faced shareholders that became more attentive and demanding. Achieving a sufficient return on the capital provided by the investors was suddenly of paramount importance in order to avoid a capital drain into more profitable market segments. A detailed account of these developments, including case studies on some of their victims, can
be found in [19].

As a consequence of these developments, insurers have to select their strategies in such a way that they have a favorable impact on the bottom line of the company, and not only relative to some isolated aspect of the business. This is the domain of Integrated Risk Management, see [38]. Clearly, this calls for corresponding tools and methods that permit an integrated and holistic quantitative analysis of the company relative to all relevant risk factors and their interrelations. In non-life insurance, the term "DFA" was coined for such tools and methods. On the technical level, Monte Carlo simulation was selected because it is basically the only means that allows to deal with the long time horizons present in insurance, and with the combination of models for a large number of interacting risk factors.

1.3 Areas of Applications

At a very general level, DFA is used to determine how an insurer might fare under a range of future possible environment conditions and strategies. Here, environment conditions are topics that are not under the control of management, whereas strategies are topics that are under the control of management. Typical strategy elements whose impact is explored by DFA studies include:

**Business mix:** relative and absolute volumes in the different lines of business, premium and commission level, etc.

**Reinsurance:** reinsurance structures per line of business and on the entire book, including contract types, dependencies between contracts, parameters (quota, deductibles, limits, reinstatements, etc.), cost of reinsurance.

**Asset allocation:** normally only on a strategic level; allocation of the company’s assets to the different investment asset classes, overall or per currency; portfolio re-balancing strategies.
1.3. Areas of Applications

Capital: level and structure of the company's capital; equity and debt of all kinds, including dividend payments for equity, coupon schedules and values, redemption and embedded options for debt, allocation of capital to lines of business, return on capital.

DFA can investigate these strategy elements under all those environment conditions that the scenario generator can model; see Section 3.1. The analysis capabilities of DFA include:

Profitability: Profitability can be analyzed on a cash flow basis or on a return-on-capital basis. DFA allows to measure profitability per line of business or for the entire company.

Solvency: DFA allows to measure the solvency and the liquidity of the company or parts of it; be it on an economic or on a statutory basis. DFA can serve as an early warning tool for future solvency and liquidity gaps.

Compliance: A DFA company model can implement regulatory or statutory standards and mechanisms. In this way, the compliance of the company with regulations, or the likelihood of regulatory interventions can be assessed. Besides legal ones, the standards of rating agencies are of increasing importance for insurers.

Sensitivity: One of the most important virtues of DFA is that it allows to explore how the company reacts to a change in strategy (or also a change in environmental conditions) relative to the situation where the current strategy pertains also to the future.

Dependency: Probably the most important benefit of DFA is that it allows to discover and analyze dependencies of all kinds that are hard to grasp without a holistic modelling and analysis tool. A very typical application here is to analyze the interplay of assets and liabilities, i.e. strategic asset/liability management ("ALM").

These analytical capabilities can then be used for a number of specific tasks, either on a permanent basis or for one-time dedicated studies of special issues. If a company has set up a DFA model, it can reuse it on a regular basis, e.g. quarterly or yearly, in order to evaluate the in-force
strategy and possible improvements to this strategy. In this way, DFA can be an important part of the company’s business planning and risk management setup. On the other hand, DFA studies can also be made on a one-time basis if strategic decisions of great significance are to be made. Examples for such decisions include mergers and acquisitions, entry in or exit from some business, thorough re-balancing of reinsurance structures or investment portfolios, or capital market transactions. Basically, DFA can be used for assessing strategic issues that affect the company as a whole.

The main users of DFA are insurance and reinsurance companies themselves. They normally use DFA models on a permanent basis as a part of their risk management and planning process; [91] describes such a system. DFA systems in this context are usually of substantial complexity, and only their continued use justifies the substantial costs and efforts for their construction. Another type of users are consulting companies and brokers who use dedicated – usually less complex – DFA studies for special tasks, e.g. the structuring of large and complicated deals. An emerging class of users are regulatory bodies and rating agencies; they normally set up relatively simple models that are general enough to fit on a broad range of insurance companies and that allow to conduct regulation or rating in a quantitatively more sophisticated, transparent and standardized way, see e.g. [4]. A detailed account of the most important uses and users of DFA is given in [29]; some new perspectives are outlined in [68].

1.4 A Survey of the Field

A survey of the surroundings of DFA in a technical sense, that is, the linkages of DFA to various areas of applied mathematics, will be provided in Section 2.2 and Chapter 3. A fairly extensive survey of how DFA integrates into the actuarial and financial landscape is provided in [14]. This survey is current as of 2003, and we confine this section to pointing out some recent developments.
1.4. A Survey of the Field

The DFA committee of the Casualty Actuarial Society (CAS) has changed its name into *Dynamic Risk Modeling* (DRM) committee, but keeps on being a major driving force in the development of the field. Current activities include

1. Standardizing and improving the way in which DFA results are used for executive-level decision making.

2. Re-writing the DFA Handbook [29], with emphasis on adding case studies and fully integrating coherent risk measures [5].

3. Updating the public-access DFA – that is, Dynamo [40] – so as to have an open-source DFA framework.

Another major driving force for DFA in the future will be the regulatory authorities. The assessment of insurer solvency is currently undergoing major changes. The aim is to establish a risk capital-based framework similar to the one from the Basel II accord in the banking world. Related research from the actuarial viewpoint on a global level is described in [73]. In many countries, specific solvency assessment approaches in this new spirit have been or are being developed. In Switzerland, for instance, this is the *Swiss Solvency Test (SST)* [58]. Two features that are common to all modern solvency assessment frameworks deserve special attention from the DFA viewpoint:

1. Insurers can use their *internal models* for the determination of risk and risk-based capital, provided that these models satisfy some acceptability criteria.

2. Assets and liabilities are subject to *market-consistent valuation*. This means that assets are valued at their market price, while liabilities are valued based on the price that financial markets would place on them, taking into account all embedded options and financial guarantees [58].

DFA, as a holistic modelling approach taking into account both financial and underwriting risks lends itself quite naturally to filling these tasks. Although, if it comes to valuation, some developments beyond the present
state are necessary. Part II will develop some of the necessary ingredients.
Chapter 2

A Generic DFA System

2.1 Overall Structure

The purpose of this chapter is to identify the components of DFA and their interrelations. The structure referred to here is generic in that it does not describe a specific DFA system, but it identifies all those elements that are typical for any DFA. Figure 2.1 gives an overview of such a generic DFA system.

The scenario generator comprises stochastic models for the risk factors affecting the company. Risk factors typically include economic risks, underwriting risks, asset risks, and business risks (e.g. underwriting cycles). The output of the scenario generator is a large number of Monte Carlo scenarios for the joint behavior of all modelled risk factors over the full time range of the study, representing possible future “states-of-nature” (where “nature” is meant in a wide sense). Calibration means the process of finding suitable parameters for the models to produce sensible scenarios; it is an integral part of any DFA. Scenario generation is described in more detail in Section 3.1. If the Monte Carlo scenarios were replaced by a small set of constructed deterministic scenarios, then the DFA study would be equivalent to classical scenario testing of business plans.
Each one of the scenarios is then fed into the company model or model office that reflects the reaction of the company on the behavior of the risk factors as suggested by the scenarios. The company model reproduces the internal financial and operating structure of the company.

Each company model comprises a number of parameters that are under the control of management, e.g. investment portfolio weights or reinsurance retentions. A set of values for these parameters corresponds to a
strategy, and DFA is a means for comparing the effectiveness of different strategies under the projected future course of events. The output of a DFA study consists of the results of the application of the company model, parameterized by a strategy, on each of the generated scenarios. So, each risk scenario fed into the company model is mapped onto one resulting scenario. Company and strategy modelling is described in more detail in Section 3.2.

Given the Monte Carlo setup, there is a large number of output scenarios, so that sophisticated analysis and presentation facilities become necessary for extracting information from the output; see Section 3.3. The extracted information can then be used to readjust the strategy for the optimization of the target values of the company; see Section 3.4.

### 2.2 Formal Specification

In Section 2.1, we have given a functional overview of a generic DFA and its elements. Now, we are going to formulate this setup in mathematical language in order to embed DFA in the realm of applied mathematics. The terminology employed here loosely follows [62].

Since uncertainty is involved, we introduce a probability space $(\Omega, \mathcal{A}, P)$ on which all random variables are defined. We let $t \in \mathbb{Z}$ denote time on a discrete time scale. This simplification is admissible since time scales in DFA are typically equispaced. One time step is usually one year, sometimes one quarter, and rarely one month. The meaning of different values for $t$ is as follows:

- $t < 0$ denotes the past, assumed to be known.
- $t = 0$ denotes the present, i.e. the starting point of the DFA.
- $t > 0$ denotes the future, assumed to be uncertain.

Moreover, we let $T \in \mathbb{N}$ denote the time horizon, that is, the number of time steps in the future over which the DFA is to be performed. Hence,
the time frame of the DFA is \( t \in \{0, 1, \ldots, T\} \).

Recall from Figure 2.1 and Section 2.1 that the future fate of the company depends on a number \( N \) of risk factors. Let the random variable \( X_n(t) \) denote the value of the \( n \)-th risk factor at time \( t \). The joint behavior of all risk factors over time is then modeled by the multivariate stochastic process

\[
X = (X(t))_{t=0}^{T} \quad \text{where} \quad X(t) = (X_1(t), \ldots, X_N(t))'.
\] (2.1)

The space of possible realizations of \( X \) is denoted by \( \mathcal{X} \). We will generally have \( \mathcal{X} \subset \mathbb{R}^{N \times (T+1)} \). The value of \( X \) at the initial time \( t = 0 \) is assumed to be known, that is \( X(0) = x \).

As \( X \) is random, we have to introduce a model for this random behavior. A pair \((\mathcal{X}, \mathcal{P})\), where \( \mathcal{X} \) is the space of possible realizations of \( X \), and \( \mathcal{P} \) is a family of probability distributions on \( \mathcal{X} \), is called a stochastic model for \( X \). Since the observations of \( X \) are indexed by time, and since we cannot generally assume independence over time, \((\mathcal{X}, \mathcal{P})\) is a multivariate dynamic model in the sense of \([62]\).

We further assume that the family \( \mathcal{P} \) of probability distributions on the space \( \mathcal{X} \) is of the form \( \mathcal{P} = \{P_\theta : \theta \in \Theta\} \), where \( \theta \) is the parameter, belonging to the parameter space \( \Theta \). While the whole family \( \mathcal{P} \) is called stochastic model, a specific instance \( P_\theta \) is called a stochastic scenario.

As will become clear below, one DFA simulation is always based on one stochastic scenario. It follows from the setup above that the process of finding a stochastic scenario involves several stages:

1. The activity of selecting the risk factors \( X_n \) to include in the risk factor process \( X \) is called model specification.

2. The activity of selecting a stochastic model \( \mathcal{P} = \{P_\theta : \theta \in \Theta\} \) is called model selection.
2.2. Formal Specification

3. Given a selected stochastic model \( \mathcal{P} \), the process of identifying a suitable parameter \( \theta \in \Theta \), that is the process of selecting a specific stochastic scenario \( P_\theta \in \mathcal{P} \), is called calibration.

In addition to the known initial value \( X(0) = x(0) \) of the risk factor process, past values \( X(-s) = x(-s), \ldots, X(-1) = x(-1) \) may also be known. The sample \( \{x(t)\}_{t=0}^s \) is called historical data, and \( s \in \mathbb{N}_0 \) is called backward horizon. Historical data is often required by the company model (see further below), and it may also be used for model selection and calibration.

When the risk factors \( X \) are specified, the stochastic model \( \mathcal{P} \) is selected and the stochastic scenario \( P_\theta \) is calibrated, the actual scenario generation, that is, the simulation can be done. This consists of generating Monte Carlo realizations

\[
X^k := \{X^k(t)\}_{t=0}^T
\]

(2.2)

of the trajectories of the risk factor process \( X \) over the simulation period \([0,T]\). We note that \( X^k(0) = x \) for all \( k \) and that any two scenarios \( X^j \) and \( X^k \) are independent conditional on \( X(0) \). The result of the simulation is the scenario set

\[
\{X^k\}_{k=1}^K = \left\{ \{X^k(t)\}_{t=0}^T \right\}_{k=1}^K
\]

(2.3)

where \( K \in \mathbb{N} \) is the number of simulations. Hence, the simulation turns the stochastic scenario \( P_\theta \) into the numerically tractable form of a scenario set. The specifications up until here cover the scenario generator and calibration parts according to Figure 2.1. More specific information on the risk factors included in \( X \) and on the internal structure of stochastic models \( \mathcal{P} \) and stochastic scenarios \( P_\theta \) is given in Section 3.1. Some indications on how to select \( K \) such that the scenario set is a sufficiently good approximation of the stochastic scenario are given in Section 11.3.

Each scenario \( X^k \) represents one possible future course of events for the risk factors modelled. According to Figure 2.1, each scenario must be fed into a model of the working structure of the company, the company model.
If we assume that the result of the company in period \( t \) only depends on the realizations \( X(t) \) of the risk factors in the same time period, then the company model can be formalized as a mapping

\[
f : \mathbb{R}^N \times \mathbb{R}^M \rightarrow \mathbb{R}^L \\
(X(t), y(t)) \mapsto Z(t).
\]  

(2.4)

The vector \( Z(t) \) contains the output variables, that is, one or several variables representing the operational result of the company. The vector \( y(t) \) represents values that are under the control of the management of the company. The set \( y := \{y(t)\}_{t=1}^T \) of these control values is called a strategy. A strategy \( y \) may be fully deterministic, i.e. pre-specified as of \( t = 0 \), or it may depend on past values of the risk factors, in the sense

\[
y^k(t) = g(X^k(t-1), \ldots, X^k(t-s)).
\]  

(2.5)

That is, the strategy must be previsible in stochastic process terminology. In many real-life situations, the result of the company in period \( t \) will also depend on risk factor values in earlier periods. Hence, the company model must often be formalised as

\[
f : \mathbb{R}^{N \times (q+1)} \times \mathbb{R}^{M \times (q+1)} \rightarrow \mathbb{R}^L \\
(X(t), \ldots, X(t-q); y(t), \ldots, y(t-q)) \mapsto Z(t),
\]  

(2.6)

where we must have \( q \leq s \). The issue of strategy modeling will be explored in more detail in Section 3.2.

The company model, parameterized with the strategy, transforms each risk factor scenario into an output scenario that describes the operating results of the company under the given realization of the risk factors and with the specified strategy; formally

\[
Z^k = \{Z^k(t)\}_{t=1}^T = \left\{ f\left( X^k(t), y^k(t) \right) \right\}_{t=1}^T, \quad k = 1, \ldots, K.
\]  

(2.7)

The collected output scenarios \( \{Z^k\}_{k=1}^K \) form the result set. Since a result set always pertains to a specific strategy \( y \) and to a specific stochastic scenario \( P_\theta \), we may also denote it by \( \{Z^k; y, P_\theta\}_{k=1}^K \). The problems around company and strategy modelling will be dealt with in more depth in Section 3.2.
2.2. Formal Specification

The result set \( \{Z^k\} \) is the Monte Carlo representation of the result process \( Z = \{Z(t)\}_{t=1}^T \). This means that we have to analyze an entire stochastic process that may, moreover, be multivariate. Hence, there is a strong need for transformations that reduce this vast amount of information to a form that a human analyst or a formal optimizer can grasp, see again Figure 2.1. There are different approaches to this task, and also many caveats, which we will analyze in more detail in Section 3.3.

For the moment, we remain on a rather abstract level. Let \( Z \) be the space of possible realizations of the output process \( Z \). Then we assume that we have transformations of the form

\[
\epsilon : \mathcal{Z}^K \rightarrow \mathbb{R} \\
\{Z^k; y, P_0\}_{k=1}^K \mapsto \epsilon(y, P_0).
\]

(2.8)

Here, \( \epsilon \) is a measure of either risk or reward associated with the strategy \( y \) under the stochastic scenario \( P_0 \). Letting \( \mu(\cdot, \cdot) \) denote a measure of reward, and \( \mathcal{Y} \) the set of admissible strategies (which may be heavily structured or constrained), we can formulate DFA as the following abstract, optimization problem:

\[
y^* = \arg\max_{y \in \mathcal{Y}} \mu(y, P_0) \quad \text{s.t.} \quad \epsilon(y, P_0) \in \mathcal{C}.
\]

(2.9)

Here, \( \mu \) denotes a function as in (2.8) measuring reward, \( \epsilon \) is a vector of one or several functions as in (2.8) measuring risk, and \( \mathcal{C} \) is the set of admissible values for these risks. Hence, (2.9) specifies the control and optimization element in Figure 2.1. The possibilities of control and optimization in DFA will be further explored in Section 3.4.
Seite Leer / Blank leaf
Chapter 3

The Elements of DFA

3.1 Scenario Generation

Scenario generation lies at the basis of DFA; see Figure 2.1. The constituents of scenario generation are formally specified and embedded into the context in Section 2.2. The overall task of scenario generation is to produce Monte Carlo scenarios for the risk factors \( \mathbf{X}(t) = (X_1(t), \ldots, X_N(t))' \) over the simulation period \( t = 1, \ldots, T \). This entails the tasks of model specification (determining which factors to include into \( \mathbf{X} \)), model selection (finding a family \( \{P_\theta : \theta \in \Theta\} \) of probability laws), calibration (finding a suitable \( \theta \in \Theta \)) and Monte Carlo simulation (producing a scenario set \( \{\{\mathbf{X}^{(k)}(t)\}_{t=0}^T\}_{k=1}^K \) ). In many situations, an additional validation of the scenario set produced is necessary.

The scenario generator has to satisfy a number of particular requirements: First of all, it does not only have to produce scenarios for each individual risk factor, but must also allow, specify and account for dependencies between the risk factors (contemporaneous dependencies) and dependencies over time (intertemporal dependencies). Neglecting these dependencies means underestimating the risks since the model would suggest diversification opportunities where, actually, none are present. Moreover, the
### Table 3.1: Overview of the risk factors included in DFA scenario generators. Variables in parentheses are only included in more sophisticated systems.

<table>
<thead>
<tr>
<th>Economic</th>
<th>Underwriting</th>
<th>Investment</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per economy:</td>
<td>Per LOB:</td>
<td>gov’t bonds</td>
<td>(U/W cycles)</td>
</tr>
<tr>
<td>- inflation</td>
<td>- attritional losses</td>
<td>stocks</td>
<td>(R/I cycles)</td>
</tr>
<tr>
<td>- interest rates</td>
<td>- large losses</td>
<td>real estate</td>
<td>(OP risk)</td>
</tr>
<tr>
<td>(FX rates)</td>
<td>- loss development</td>
<td>(corporate bonds)</td>
<td>(etc.)</td>
</tr>
<tr>
<td>(credit spreads)</td>
<td>Across LOBs:</td>
<td>(ABS)</td>
<td></td>
</tr>
<tr>
<td>(GDP)</td>
<td>- CAT losses</td>
<td>(ILS)</td>
<td></td>
</tr>
<tr>
<td>(wage levels)</td>
<td>(reserve uncertainty)</td>
<td>(etc.)</td>
<td></td>
</tr>
<tr>
<td>(etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

scenarios should not only reproduce the "usual" behavior of the risk factors, but they should also sufficiently account for their extreme individual and joint outcomes. Given the holistic point of view of DFA, the scenario generator has to contain stochastic models for a large number of quite diverse risk factors, belonging to different groups. Table 3.1 shows an overview.

The economic risk factors according to Table 3.1, particularly inflation and interest rates, are the most basic ones in the sense that both underwriting and investment outcomes depend on them rather than the converse. The relations between the economic risk factors are more intricate than between the underwriting ones, so that their modelling is usually bundled into a separate, closed unit, a so-called economic scenario generator (ESG). The modelling of economic risk factors is quite extensively discussed in Chapter 5.

The value of investable assets does not necessarily correspond to the value of the financial risk factors from Table 3.1. This is particularly the case for fixed-income securities, and it adds an additional model layer. However, as long as the fixed income securities do not contain significant embedded derivatives, there are no further stochastic modelling problems. A portfolio of coupon-paying bonds can always be expressed by an equivalent portfolio of zero-coupon bonds, and the prices of the latter follow directly
from an interest rate model, as do, if necessary, prices of possible embedded derivatives. Some challenge arises from the modelling of the portfolio composition over time. The most popular modelling approach here is to keep the portfolio duration constant by applying certain rebalancing rules at each time step. Details on investment management approaches for insurers can be found in [6].

Underwriting risks are usually modelled per line of business. The claims per line of business are usually separated into attritional losses (high frequency / low severity) and large losses (low frequency / high severity). Attritional losses can be modelled as annual aggregate losses or loss ratios, that is \( X_A(t) \sim F \) for some suitable distribution \( F \). Large losses must be modelled on a frequency-severity basis, that is, \( X_L(t) = \sum_{i=1}^{N(t)} X_i \), where \( N(t) \) is the random number of losses in period \( t \), and \( X_1, X_2, \ldots \sim \text{iid} \ G \) are the severities of the single losses. Only in this way can one properly reflect the impact of non-proportional insurance. Catastrophe losses are special in that one event (an earthquake, for instance) affects several lines of business. Catastrophe modelling can also be done through stochastic models (see [79], for instance), but – for the perils covered by them – it is fairly commonplace to rely on scenario output from special CAT modelling tools such as CATrader, RiskLink or EQEcat. Actuarial science provides a wide range of models for both loss ratios and frequency-severity approaches; see [41] for an overview. Although these models were created with pricing in mind, they can be reused quite directly for simulation.

The considerations above cover the incurred losses \( X_I(t) \) in period \( t \). For certain lines of business (liability or aviation, for instance), a loss incurred in a certain time period, \( X_I(t) \), is only paid gradually over an extended amount of time, according to a payout pattern \( (p(s))_{s=0}^{S} \), where \( p(s) \) denotes the percentage of the loss \( X_I(t) \) that is effectively paid in period \( t + s \). Hence, the effective cash outflow in period \( t \) is

\[
X_C(t) = \sum_{s=0}^{S} X_I(t - s) p(s)
\]

rather than \( X_I(t) \). Since DFA models are cash flow-based, and since the speed of payout influences the investment income on the funds held to
cover the losses, payout patterns must not be neglected in DFA models. Payout patterns are usually modelled in a deterministic manner and provided by the reserving department of the company. However, in reality there is also uncertainty as to the loss development, and, accordingly, it is desirable to incorporate this uncertainty into a DFA model. A comprehensive treatment of both deterministic and stochastic models for payout patterns is provided in [116].

If losses $X(t)$ are modelled over an extended time span $t = 0, 1, \ldots T$, their dependence on economic factors has to be accounted for as well. Loss costs (house repairs or therapies, for instance) can depend on inflation or other economic indicators, as can loss frequencies (credit insurance, for instance). Money amounts can be adjusted by simple indexing, whereas claims frequencies are usually adjusted by mixing the claims arrival process with the related economic indicator (inflation, for instance); Chapter 9 of [41] provides full details.

Besides the underwriting risks and the basic economic risk factors such as inflation, (government) interest rates and equities, sophisticated DFA scenario generators may contain models for various further risk factors. In international setups, foreign exchange rates have to be incorporated, and an additional challenge is to let the model also reflect the international dependencies. Additional risk factors for one economy may include Gross Domestic Product (GDP) or specific relevant types of inflation as e.g. wage or medical cost inflation. Increasingly important are also models for credit defaults and credit spreads which must, of course, properly reflect the dependencies on other economic variables. This, subsequently, allows to model investments like asset-backed securities and corporate bonds that are extremely important for insurers, see [6]. The modelling of operational risks (see [38], which also provides a very general overview and classification of all risks affecting financial companies), which are a current area of concern in banking regulation, is not yet very wide spread in DFA.

An important problem specific to insurance and reinsurance is the presence of underwriting cycles ("hard" and "soft" markets), which have a
3.1. Scenario Generation

non-negligible business impact on the long time horizons considered by DFA. These cycles and their origins and dependencies are not very well understood and very difficult to model; see [55] for a survey of the current state of knowledge. Nevertheless, simple models for business cycles are in place; see [79] or Chapter 9 of [41], for instance.

The real challenge of DFA scenario generation lies in the composition of the component models into an integrated model, that is, in the modelling of the ubiquitous contemporaneous and intertemporal dependencies among the risk factors. These dependencies come in many different, more or less strong forms, many of which are non-linear. A particular challenge in this context is the adequate assessment of the impact of extreme events, when the historically observable dependency becomes much stronger and risk factors appear much more interrelated (the so-called tail dependency). A comprehensive taxonomy of dependency concepts and modelling approaches is provided in [37]. Different approaches for dependency modelling are pursued, namely:

- **Deterministic** modelling by postulating functional relations between risk factors, e.g. mixture models or regression-type models; see [41].

- Statistical modelling of dependencies, with linear correlation being the most popular concept. However, linear correlation has some serious limitations when extreme values are important; see [54] for a related study and possible modelling approaches.

The discussions above make it clear that a stochastic model \( \{P_{\theta} : \theta \in \Theta\} \) for the risk factor process \( \{X(t)\}_{t=0}^{T} \) will hardly ever have a tractable, coherent form. Rather, it consists of a number of intricately entangled component models, and the number of parameters will usually be quite high. This makes calibration, that is finding suitable parameters \( \theta \in \Theta \), a daunting task, further complicated by the fact that, for most risk factors, only small amounts of data are available. Parsimony and transparency are, therefore, crucial requirements for models being used in DFA scenario generation.
Under the circumstances outlined above, it must be feared that the stochastic scenarios $P_\theta$ selected are subject to considerable parameter uncertainty. An additional validation is, therefore, necessary. If sufficient amounts of historical data are available, formal backtesting procedures – as described in Chapter 11 – can be applied. This is, however, not the case for many risk factors. The only possibility then is to do either peer reviews of the model or statistical evaluations of the scenario output $\{\{X^{(k)}(t)\}_{t=0}^{T}\}_{k=1}^{K}$ and to compare the results with external forecasts or expert opinions. Validation will become even more important in the future if DFA models are to be used as internal models in a regulatory context; see Section 1.4.

The generation of Monte Carlo scenarios, that is, the transformation of a stochastic scenario $P_\theta$ into a scenario set $\{\{X^{(k)}(t)\}_{t=0}^{T}\}_{k=1}^{K}$, does not affect most DFA practitioners as they rely on commercially available software packages. Nevertheless, one should bear in mind that generating thousands of Monte Carlo scenarios for a large number of dependent risk factors over many time periods poses some non-trivial problems; see [59]. The most elementary part is to have a random number generator that can produce thousands or even millions of iid $U(0,1)$ pseudo random numbers. Indeed, some of the commercially available random number generators fail miserably in this task, so that related evaluations are a must; see [12]. These random numbers are then used to produce realizations of univariate or multivariate random variables with other distributions. The accuracy of this second stage can be tested by using the Probability Integral Transform described in Section 11.2, where it is particularly important to check if there are sufficiently many replicates far out in the tails of the distribution. One problem that each DFA analyst has is to determine the number $K$ of Monte Carlo replicates in such a way that the range of possible outcomes is well covered. Some answers to this question are given in Section 11.3.

Finally, it is fundamentally difficult to make judgments on the plausibility of scenarios for the expanded time horizons often present in DFA studies. Fitting a stochastic model either to historical or current market data implies the assumption that history or current expectations are a reliable
prediction for the future. While this may be true for short time horizons, it is definitely questionable for time horizons as long as five to 20 years, as they are quite commonplace in insurance. There are regime switches or other hitherto unexperienced events that are not reflected by historical data or current market expectations. Past examples include asbestos liabilities or the events of September 11th, 2001. An interesting case study on the issue is [16], whereas [110] explores in very general the limitations of risk management based on stochastic models and argues that the latter must be complemented with some judgmental crisis scenarios.

3.2 Company and Strategy Modelling

The scenario generator produces a scenario set \( \{\{X^k(t)\}_{t=0}^{T}\}_{k=1}^{K} \) representing the possible future courses of events of the risk factors affecting the company. The next step, according to (2.6), is then to model the reaction of the company under each one of the scenarios. That is, we have a transformation

\[
f : \mathbb{R}^{N \times (q+1)} \times \mathbb{R}^{M \times (q+1)} \rightarrow \mathbb{R}^L
\]

\[
(X(t), \ldots, X(t-q); y(t), \ldots, y(t-q)) \mapsto Z(t),
\]

assuming that historical data is available back to \( t = -q \). This transformation is called company model or model office. Following Section 2.2, the vector \( y(t) \) represents values that are under the control of the management of the company, and the set \( \{y(t)\}_{t=0}^{T} \) is thus a strategy. The vector \( Z(t) \) consists of variables representing the financial condition of the company. Typical examples include earnings before or after interest and tax, or the level of shareholders' equity. More information on the selection of target variables is given in [29]. Given the aim of DFA to model the bottom line impact of strategic decisions, elements of \( Z(t) \) are usually balance sheet or income statement positions.

(2.6) alludes that \( f \) is some mathematically tractable function. The reality, however, is different. The aim of DFA is to make projections of the bottom line of the company, that is, its financial statements. This requires a careful representation of the internal operation and consolidation
structure of the company, including the effects of investment, reinsurance and even some accounting rules. So, besides mathematically tractable elements, \( f \) generally also contains decision rules of the if-then type. Think, for instance, of a multi-layer excess-of-loss reinsurance program with limited reinstatements and a drawdown clause on the upper layers. Hence, we must consider \( f \) as a function that generally has none of the desirable mathematical properties like continuity or linearity. This distinguishes DFA company models from the ones used in dynamic programming-based ALM models which are kept simple enough to allow the use of formal optimization techniques, at the possible expense of a biased representation of the operating structure of the company.

Company models can be relatively simple, as the ones in [79] or [41] which basically consolidate in a purely technical way the outcomes of the various risks. However, many DFA company models used in practice tend to be highly complex. In particular, they also incorporate the effects of regulation, accounting and taxation since these issues have an important impact on the behavior and the financial results of insurance companies. Examples of detailed models for US property-casualty insurers are described in [70] and [40]. The partial model shown in Figure 3.1 represents just one line of business of a company; the full model would then contain several other lines of business, plus the entire investment side of the company, plus the top level structure consolidating everything into the balance sheet. This gives a good idea of the actual complexity of real-world DFA models. Modelling of non-life insurance companies is somewhat more recent than related efforts in life insurance, usually termed life office modelling; see [65] or [93], for instance. Synoptical overviews of company modelling – for both life and non-life – are provided by [67] and [94].

Company models used in DFA are usually very cash flow-oriented, i.e. they try to imitate the cash flows of the company, or, more specifically, the technical and financial accounting structures. Alternatively, it would be imaginable to structure a company model along the lines of economic value creation. The problem with this approach is, however, that this issue is not very well understood in insurance; see [63] for a survey of the current state of the knowledge.
Figure 3.1: Extract from a DFA company model. The underlying DFA software is ReMetrica II from Benfield Group.
The strategy vector \( y(t) \) represents those parameters that are under the control of the company at time \( t \). In order for a strategy \( \{y(t)\}_{t=0}^{T} \) to be plausible, it must be \textit{previsible}, that is, \( y(t) \) must be known \textit{before} time \( t \). Typical elements in a strategy are reinsurance retentions or investment portfolio weights. The simplest way of strategy modelling is to use values predetermined at the start of the simulation for all \( t > 0 \). This is not very realistic in many situations, since DFA simulations cover time periods of several years. On this time scale, management is well capable of adapting a strategy according to the development of the risk factors. So, a more realistic approach is to let the strategy depend on past values, for instance

\[
y^k(t) = g\left(X^k(t-1), y^k(t-1)\right), \quad t = 1, \ldots, T,
\]

that is, the strategy becomes path-dependent. Strategy formulations of this type are called \textit{decision rules}. They can account for managerial flexibility to some extent, but not perfectly. In order to obtain optimal strategy choices fully reflecting managerial flexibility, one has to do formal dynamic optimization. This is, however, difficult in realistic DFA settings given the above-mentioned analytical intractability of company models. The topic of optimization is further discussed in Section 3.4.

### 3.3 Output and Evaluation

The result of a DFA simulation is a \textit{result set} \( \{Z^k\}_{k=1}^{K} \) where

\[
Z^k = \left\{ f(X^k(t), y^k(t)) \right\}_{t=0}^{T}
\]

representing the probability law of the result process \( \{Z(t)\}_{t=0}^{T} \) under the stochastic scenario \( P_\theta \) and the strategy \( y \). Given that the result set stretches over time, over scenarios and, usually, across several variables, its evaluation requires some sophistication. For the time being, we limit the attention to univariate outcomes \( Z(t) := Z_i(t) \), and we let \( Z := \{Z(t)\}_{t=1}^{T} \).

Classically, an output variable \( \{Z(t)\}_{t=1}^{T} \) is analyzed for each \( t \) separately, or at the horizon \( T \) only. The most general approach is to analyze the
empirical distribution

$$
\hat{F}_Z(t)(z) = \frac{1}{K} \sum_{k=1}^{K} 1\{Z^k(t) \leq z\},
$$

If $Z(t)$ represents a financial result, then this distribution is called profit-and-loss distribution or P&L for short. For comparison and decision making, it is more desirable to characterize the P&L by a few numbers, typically by one measure for risk and one measure for reward. These measures are either estimates of transformed moments of $Z(t)$, that is,

$$
\hat{E}[g(Z(t))|P_\theta, y] = \frac{1}{K} \sum_{k=1}^{K} g(Z^k(t)),
$$

or estimates of the empirical quantiles, that is,

$$
\hat{q}_\alpha[Z(t)|P_\theta, y] := \hat{F}^{-1}_{Z(t)}(\alpha) := \min \left\{ Z^{(k)}(t) : \frac{k}{K} > \alpha \right\},
$$

where $Z^{(k)}(t)$ denotes the $k$-th order statistic of $\{Z^k(t)\}_{k=1}^{K}$, or combinations thereof. The measure of reward is usually the expectation. As for the measure of risk, the choice is somewhat richer. The most classical one is the standard deviation which equally accounts for upside and downside deviations. In modern risk management, however, the trend goes towards risk measures for the downside only, with VaR, that is $\hat{q}_\alpha(Z(t))$, being the most prominent one. The problem with VaR and many other risk measures is, however, that they fail to satisfy the subadditivity property

$$
\rho(X + Y) \leq \rho(X) + \rho(Y),
$$

where $\rho$ stands for some measure of risk. That is, they are not coherent in the sense of [5]. Subadditivity, however, is particularly important in the context of DFA where one often has to relate results of parts of the company to the result of the entire company. If there is no subadditivity, risk for the parts stand in no sensible relation with total risk, thereby failing to account for the benefits of diversification that are so crucial for insurance business. Luckily, the use of coherent measures of risk is becoming the standard in DFA. This is also reflected by the fact that the new version of [29] will contain a full chapter dedicated to them. The most popular coherent measure of risk in DFA is the TailVaR, often
referred to as \textit{Expected Shortfall (ES)}, although they are only equivalent for continuous distributions. It is defined as

\[
\text{TailVaR}_\alpha[Z(t)] = -\mathbb{E}[Z(t) | Z(t) \leq q_\alpha[Z(t)]]
\]

which can be easily computed from a Monte Carlo sample. The left panel of Figure 3.2 shows a typical evaluation of a DFA result set with measures of risk and reward evolving over time. The right panel illustrates an important caveat in this type of evaluation: the target variable may temporarily assume values that correspond to a disruption of the ordinary course of business such as ruin or regulatory intervention. Such degenerate trajectories have to be accounted for in suitable ways, otherwise the terminal results may no longer be realistic. One possibility is to censor these trajectories, to compute the probability of coming into distress and to state measures of risk and reward conditional on not coming into distress. Probabilities that the target variables exceed certain thresholds $\bar{z}$, e.g. for bankruptcy or regulatory intervention, are particularly popular when DFA results must be reported to executive management. Such probabilities are easily computed by

\[
\hat{p} = \frac{1}{K} \text{card} \{k : Z^k(t) \leq \bar{z} \text{ for some } t\}
\]

By repeating the simulation and computing the target values for several different strategies one can compare these strategies in terms of their risks and rewards, determine ranges of feasible results and, finally, select the best among the feasible strategies. Figure 3.3 shows such a comparison, conceptually very similar to risk-return analysis in classical portfolio theory. It is, however, important to notice that DFA does not normally allow for the use of formal optimization techniques (such as convex optimization), since the structure of the model is too irregular. The optimization rather consists of educated guesses for better strategies and subsequent evaluations thereof by carrying out a new simulation run. Such repeated simulation runs with different strategy settings (or also with different calibrations of the scenario generator) are often used for exploring the sensitivities of the business against strategy changes or against changes in the environment, i.e. for exploring relative rather than absolute impacts in order to see what strategic actions do actually have a substantial leverage.
3.3. Output and Evaluation

Figure 3.2: Left panel: evolution of risk and reward measures over time. Right panel: an illustration of trajectories that distort the evaluation.

A common feature of all evaluation methods shown above is that they heavily depend on the probability law $P_{\theta}$ assigned to the risk factor process $\{X(t)\}_{t=0}^{T}$. Given the high model and parameter uncertainty (see Section 3.1), this is problematic. One way to deal with this is sensitivity analysis. If $\theta_0 \in \Theta$ is the best estimate for the parameter, then sensitivity analysis is the attempt to compute

$$\frac{\partial}{\partial \theta} E[g(Z(t)) | P_{\theta}, y] \Big|_{\theta = \theta_0} \quad \text{or} \quad \frac{\partial}{\partial \theta} q_{\alpha} [Z(t) | P_{\theta}, y] \Big|_{\theta = \theta_0}.$$ 

Given the dimension of $\theta$, the non-tractability of the function $f$ relating the risk factors $X(t)$ to the output variables $Z(t)$ and the computational efforts required to evaluate the model, this is practically impossible. However, by educated guessing and some experiments, one can identify at least some key parameters with particularly high impact on the results and focus calibration efforts on these parameters.
An analysis technique that partly circumvents model and parameter risk is \textit{drill-down analysis}. Consider the set \( \{Z^k\}_{k=1}^K \) of simulated trajectories \( Z^k = (Z^k(1), ..., Z^k(T))' \) and let \( \mathcal{A} \subseteq \mathbb{R}^T \) be a set of particularly desirable or undesirable results. Then, one can easily identify all those trajectories that fall into \( \mathcal{A} \), yielding an index set \( I = \{ k : Z^k \in \mathcal{A} \} \). Then, one can identify the related scenarios \( X^k \) for which \( k \in I \) and analyze these scenarios statistically in order to identify environment conditions that give rise to outcomes in \( \mathcal{A} \). In many cases, this is more easily said than done due to the dimensionality and the dynamic nature of the risk factor process \( \{X(t)\}_{t=1}^T \). However, this type of analysis mitigates model and parameter risk since it basically attempts to compute \( f^{-1}(\mathcal{A}) \subseteq \mathcal{X} \), that is, the subset of the state space \( \mathcal{X} \) that gives rise to results in \( \mathcal{A} \), by using a brute-force
exhaustion of $\mathcal{X}$. In this context, the probability law $P_0$ is only an auxiliary to speed up this exhaustion.

Up until here, we have considered only projections, that is, we have explored the space of possible future courses of events of the risk factors and the behaviour of the company under a certain strategy. In some situations, however, we have to do valuation, that is, we have to assign a monetary value to the future business. This is the case if the business under consideration is for sale, an ART deal [113] or a takeover, for instance.

Assume that $Z$ is a suitable monetary output variable such as the dividend stream or the net cash flow of the business. In principle, one can then proceed along the lines of classical corporate finance [18] by computing the net present value (NPV) for each scenario path

$$\text{NPV}[Z^k] = \sum_{t=0}^{T} e^{-R(0,t)t} Z^k(t),$$

where $R(0,t)$ is the relevant market interest rate for maturity $t$ at valuation time 0. Then, one can proceed as shown earlier in this chapter. There is, however, one problem. Usually, $Z$ depends on financial risk factors such as equity markets or interest rates. That is, some of the cash flows stem from embedded derivatives. Under these circumstances, the NPV with a fixed interest rate is meaningless, as it fails to incorporate the value of the flexibility granted by the option; see [118]. In order to account for the value of the flexibility, we must have a more general way of discounting, in other words, a deflator:

$$\pi_0(Z^k) = \sum_{t=0}^{T} \xi(t) Z^k(t).$$

The deflator $\{\xi(t)\}_{t=0}^{T}$ is some kind of a stochastic discount factor that leads to a value $\pi_0(Z^k)$ which is adjusted for all financial risk. A detailed discussion of the deflator is provided in Section 7.2.

The value $\pi_0(Z^k)$ is still a random variable, but adjusted for all systematic risk from the financial markets. It can then be valuated along the classical
lines, by computing a risk-adjusted reward like RORAC or RAROC or EVA. That is, the expected gross reward is reduced by the cost for the risk capital needed to support the business. Sections 12.2 and 12.3 provide examples and comparisons of projections and valuations.

3.4 Control and Optimization

As was argued in Section 2.2, DFA is essentially an optimization problem of the form

$$
\begin{align*}
y^* &= \arg\max_{y \in \mathcal{Y}} \mu(y, P_\theta) \quad \text{s.t.} \quad \epsilon(y, P_\theta) \in C.
\end{align*}
$$

That is, under the given stochastic scenario $P_\theta$, we have to find the strategy $y^*$ that maximizes the reward as measured by the criterion $\mu$ over all admissible strategies in the set $\mathcal{Y}$ while respecting the constraints specified by the criterion $\epsilon$ and the set $C$. Given the time resolution of DFA, we are actually mainly interested in the initial value $y^*(0)$ of the strategy, but under the assumption that we are able to adapt it as time evolves.

The use of formal optimization techniques in DFA is generally limited by the sheer size and complexity of the risk factor process $\{X(t)\}_{t=0}^T$ (see Section 3.1) and of the company model $f$ (see Section 3.2). A formal optimization over the set of admissible strategies usually requires thousands, if not millions, of evaluations of the full model, and this is often beyond the feasible for large-scale DFA models.

The most prevalent optimization technique in practical DFA is, therefore, educated guessing. In the process of model construction and evaluation, the analysts involved usually acquire a deep understanding of the structure of the business under consideration. Based on this understanding, one is usually well capable of formulating alternative strategies. The DFA model can then be re-evaluated under these alternative strategies, and results can be compared. Figure 3.3 shows the results of such a comparison: the re-structured reinsurance strategy may not be the absolute optimum,
but it is shown to bring a considerable improvement.

If the business under consideration is sufficiently simple or well-structured so that a formal optimization appears feasible, then there are two basic approaches:

1. Combine repeated Monte Carlo evaluation of the DFA model with some sufficiently general optimization method.

2. Transform the DFA model into a dynamic stochastic optimization model.

For the former approach, there exists a number of optimization procedures that can deal with target functions that are not linear, quadratic or convex. Examples include the \texttt{fmincon} function of Matlab and the genetic algorithm; see Appendix B for both. \texttt{fmincon} can only find a strategy $y^*$ that is locally optimal around some given base strategy $y_B$. Moreover, some problems are to be expected if the target function is not continuous, which may well be the case in DFA; see Section 3.2. The genetic algorithm, on the other hand, can find a global optimum independent of an initial guess and for rather ill-behaved target functions, though at very high computational expenditure.

There are some possibilities to improve the efficiency of the optimization. First, since we optimize the strategy $y$ without changing the stochastic scenario $P_\theta$, the scenario set $\{\{X^k(t)\}_{t=0}^{T_k}\}_{k=1}^K$ needs only be generated once for the entire simulation. This considerably reduces the evaluation time of the model. Moreover, some of the parameters in the strategy $y$ may be kept constant, or the strategy may be formulated in terms of a parametric decision rule; see Section 3.2. That is, if

$$y(t + 1) = g_\alpha(X(t), y(t))$$

for some function with a parameter $\alpha$, one can considerably reduce the dimensionality of the search and, hence, the number of model evaluations. For limited problems like the optimization of reinsurance programmes or
strategic asset allocation, such formal optimizations are feasible. A partial example is provided in Section 12.3.

An example of how the second approach from the list above, that is, dynamic stochastic programming, can be related to DFA is provided in [43]. More details on the use of stochastic programming in asset liability management are provided in [121]. The distinctive feature of stochastic programming is that it requires a scenario tree rather than a set of Monte Carlo scenarios. Such a scenario tree can be constructed directly, but there are also ways to obtain it from a model otherwise designed for Monte Carlo simulation; see [43] or [121].

However, there are some problems with scenario trees in the DFA environment. One problem is that a scenario tree for a large number of risk factors quickly becomes intractable. Hence, the stochastic programming approach is only applicable to problems with small numbers of risk factors, such as the strategic asset allocation for pension funds. Moreover, it is hard to keep scenario trees arbitrage-free; see [121]. The presence of arbitrage opportunities, however, should be discovered by a properly implemented optimizer and, hence, leads to meaningless optimization results; see the discussion section in [43].
Chapter 4

Assessment and Outlook

4.1 Assessment of Present State

In view of the developments in the insurance markets as outlined in Section 1.2, the approach taken by DFA is undoubtedly appropriate. DFA is a means for addressing those topics that really matter in the modern insurance world, in particular the management of risk capital and its structure, the analysis of overall profitability and solvency, cost-efficient integrated risk management aimed at optimal bottom line impact and the addressing of regulatory, tax and rating agency issues. Moreover, DFA takes a sensible point of view in addressing these topics, namely a holistic one that makes no artificial separation of aspects that actually belong together.

This stance is further supported by the trend in the regulatory environment towards risk-based supervision; see [73] and [58]. If properly evolving, DFA models can fill the role of internal models in this new regulatory setup, leading to a unified process for internal risk management and external solvency reporting. A side effect of a credible risk-based supervision could be that the assessment of insurer financial strength is in the public domain again, thus limiting the influence of rating agencies.
Chapter 4. Assessment and Outlook

The genesis of DFA was driven by the industry rather than academia. The downside of this very market-driven development is that many features of practically used DFA systems lack a certain scientific soundness. The complexity of the risk factor models, which are often ad-hoc combinations of modelling elements, often leads to high model risk, and this model risk is carried forward into the results by the use of evaluations that fully depend on the stochastic model. In line with risk factor models, company models are also very detailed but often fail to incorporate properly the value of managerial flexibility. If DFA models are to gain acceptance as internal models in a regulatory environment, this lack has to be overcome.

4.2 Outlook on Future Developments

For the near and mid-term future, we can expect that company-level effectiveness will remain the main yardstick for managerial decisions, and the advent of risk-based supervision underpins this. Hence, the trend towards integrated risk management will prevail. Technically, Monte Carlo methods have become ubiquitous in quantitative analysis, and they will remain so, since they are easy to implement and easy to handle, and they allow for an easy combination of models. The availability of ever more computing power will make DFA even less computationally demanding in the future. However, as compared to the present state, DFA models will have to become more sophisticated in a number of ways:

Driven by the new regulations and financial reporting standards, but also by traumatic experience from the past [19], market-consistent valuation of insurance business [21] is gaining importance. Until now, most DFA studies have focused on projections only, but in the future, DFA will also have to cater for valuation. While the cash flow-based structure of the company models is compatible with this requirement, some of the underlying economic models need a higher degree of sophistication. Company models, in turn, must better reflect the value of managerial flexibility. This may entail that they become somewhat simpler than the ones currently in use, but the potential for simplification is limited if balance sheet
impact is to be modelled.

On a technical level, the issue of optimization has to be addressed. The potential of dynamic stochastic programming is somewhat limited in this respect, because of the limitations of scenario trees in terms of number of variables, absence of arbitrage and proper reflection of extreme outcomes. Research efforts should be concentrated on methods that can be used in conjunction with Monte Carlo scenario output.

In scenario generation, there are two main challenges. The first one is to develop models for the financial risk factors that cater for market-consistent valuation; for instance, by featuring a deflator. The other one is dependence modelling, so as to have sound methods for combining partial models for different risk factors in a way that accounts for all inherent dependencies. A side issue of particular importance in a regulatory context is the development of validation methods for risk factor models.

In general, the DFA approach has the potential of becoming the state-of-the-industry for integrated risk management, strategic decision support and even reporting to the public domain. If this vision is to become reality, much interdisciplinary research and development is still necessary.
Part II

Scenario Generation
Seite Leer / Blank leaf
Chapter 5

Introduction

5.1 Problem Statement and Requirements

The scenario generator, including facilities for calibration and validation, is the basis of each DFA system. The purpose of the scenario generator is to produce Monte Carlo scenarios spanning the range of possible outcomes for all variables of interest. The role of the scenario generator within DFA is identified in Chapter 2, and relevant features are discussed in Section 3.1.

Among the risk factors stated in Table 3.1, the group of economic and financial risk factors plays a somewhat special role. This is due to the strong functional and statistical dependence between economic variables, and because additional economic properties like no-arbitrage or covered interest parity must be satisfied. Therefore, the scenario generation for economic and financial risk factors is usually grouped into one highly integrated model for all factors, called Economic Scenario Generator (ESG). In the remainder of this section, we briefly review the requirements to an ESG; they largely correspond to those stated in [14], [69], [29] or [2].
The variables to be covered by an ESG are listed in Table 3.1. Within one economy, the most important ones are inflation, interest rates and the equity market. Inflation is important mainly because the development of future claims heavily depends on it. Inflation is generally represented by a consumer price index (CPI), but an inflation model can also be calibrated to more specific types of inflation like wage inflation or medical cost inflation. The most general measure of inflation is the GDP deflator. Some economies (mainly UK; US to some extent) have markets for inflation-adjusted bonds. Then, instead of a univariate inflation rate, one can model the term structure of real (inflation-adjusted) interest rates and retrieve the inflation from the difference to the nominal interest rates.

Interest rates are usually the risk-free ones. For modern DFA, it is crucial that the ESG models the full term structure of interest rates and not just a few selected key rates. Interest rates are important in view of the fact that bonds are generally the biggest part of the investment portfolio of an insurer, but they also play a role in the modelling of life insurance and ART contracts. Corporate bond yields are rarely modelled in ESGs, although they would be useful given the large corporate bond portfolios that many insurers possess.

Equity markets are modelled through an equity index. As DFA company models are usually cash flow-based, it is desirable to have a model that includes both the prices and the dividends. Equity models should be calibrated to indices that reflect the point of view of an institutional investor. The best choice here are probably the indices of Morgan Stanley Capital International (MSCI), which are market capitalization-weighted, freefloat-adjusted and distinguish between prices and dividends. Moreover, MSCI indices exist for all relevant markets.

If an ESG covers several economies, then it has to contain the above-mentioned risk factors for each one of the economies. Moreover, it has to include the currency exchange rates between the economies.

Other economic risk factors listed in Table 3.1 are not included in all
common ESGs. The gross domestic product (GDP) can be useful for indexing losses, or as a leading indicator for other quantities, including economic cycles. Property is an important investment for many insurers, but a property model is also useful for indexing losses in property-related lines of business. Property markets should be modelled in the same way as equity markets, that is, with separate prices and rental yields. Unemployment rates may be interesting for the indexing of claims in certain lines of business.

The models requirement for these economic risk factors according to the references cited above are the following:

1. The scenarios must be representative for the empirical behaviour of the risk factors modelled. Given the multivariate and dynamic structure of DFA, this includes, in particular, contemporaneous and intertemporal dependence. Given the focus on risk management, extreme outcomes must also be reflected properly. The models should also have structures that can cope with the long time horizons prevalent in DFA.

2. The models must respect generally accepted economic principles. For financial risk factors, this means in particular that the models must be arbitrage-free. This requirement becomes even more important if the models are also used in the realm of market-consistent valuation; see below.

3. The models must be simple and transparent. Given the large number of variables to be included, this is essential for tractability. Moreover, it must be possible to explain the model to external parties in order to assure acceptance of its use. In statistical terms, this entails the requirement of parsimony, that is, low numbers of parameters. This is essential given the sparseness of data generally available for calibration.

4. Methods and tools for calibration and validation must be an integral part of the model. A model without a sensible parametrization is useless, and sensibility of a parametrization is only recognized if it has passed validation procedures.
The model and its implementation must have good numerical qualities. This means in the first place the use of sound stochastic simulation methodology and other numerical methods. The most refined mathematical model is useless if the quality of the Monte Carlo scenarios is hampered by a careless implementation. Though not of utmost importance, computational efficiency should not be neglected. The generation of some tens of thousands of scenarios should not take more than a few hours.

Up until recently, requirements for an ESG focused exclusively on its projection capabilities. However, with the advent of fair valuation or market-consistent valuation (see [21] or [75]) in the context of IFRS and Solvency II, the valuation capabilities become more important. This reinforces, in particular, the requirement for no-arbitrage, and favours modelling approaches from mathematical finance over purely statistical ones.

The sequel of this thesis will focus on models for interest rates and FX rates. Some of the requirements can be made more specific in this context. This will be done in Section 5.3 connected with a taxonomy of possible modelling approaches.

### 5.2 Economic Scenario Generators

There exists a considerable number of ESG models for actuarial use corresponding more or less to the profile outlined in Section 5.1. Some of them are well documented, others – usually developed by private companies – are considered proprietary and, therefore, only imperfectly documented so that no real evaluation is possible. Since most of the models, to the extent that their structure is known, are quite similar, we focus in this presentation on a description of the common features.

The earliest ESG model for actuarial use was Wilkie’s [119], also described in [41]. Although outdated today due to the lack of a full term structure model and the presence of arbitrage, it is still an important reference. An-
other important model, marketed by the actuarial consulting Firm Till- 
inghast Towers Perrin and widely used in the industry, is CAP:Link; see [121]. Some versions of CAP:Link can model several economies, making it Global CAP:Link. A class of interesting but imperfectly documented models is due to Andrew Smith and B&W Deloitte; see [114]. Very well documented, on the other hand, is Andrew Cairns’ model [23]. The interest rate part of it (see also [24]) is considered in more detail in Section 8.3. Influenced by Cairns’ model, and influential among UK actuaries, is the model by Barrie & Hibbert Ltd; see [69]. This model features, in particular, a regime-switching model for the equity returns; see [66]. The Barrie & Hibbert model was also the basis for the economic models developed in response to a joint initiative of the CAS and the SOA; see [2]. This model is likely to become widely used in the US, given the influence of the sponsoring institutions.

Basically all ESG models are structured as cascade models, that is, they consist of modules for the different risk factors, and these modules have a hierarchical dependence structure. This approach allows to combine existing models for risk factors, and it also leads to flexibility and transparency (see Requirement 3 above). The alternative to cascade models would be fully integrated multivariate time series models like VAR or VECM; see [35]. However, such models cannot cope with the dimensions necessary for a full term structure, and they are not arbitrage-free either.

The term structure of interest rates is the most complicated one among the risk factors to be modelled. Therefore, term structure models are usually at the basis of the cascade. These models usually come from the realm of mathematical finance and belong to the classes of equilibrium models or price kernel models (see Section 5.3 below), often with time-inhomogeneous extensions. In order to cope with the statistical properties of term structures, they usually have two or more random factors.

Models mainly intended for the use in the UK, but also the new CAS/SOA model for the US, feature a term structure of real interest rates instead of a univariate inflation rate process. This makes sense as long as there
is an active market for inflation-linked bonds from which the real term structure can be estimated.

The sources of randomness are usually multivariate Gaussian processes, thus keeping the models (particularly for interest rates) within the realm of complete markets. Some members of Smith’s model family may be an exception, but this feature is undocumented.

Most models do not put emphasis on the use for market-consistent valuation which, in a Monte Carlo setup, amounts to modelling the state price deflator (hereinafter called price kernel). Notable exceptions are Smith’s models [114] which pioneered the use of this valuation technique in the realm of actuarial work. Cairns’ models [23, 24] also have a price kernel structure and are, therefore, suitable for valuation.

The method proposed in [102] is radically different from all those hitherto mentioned. It is based on a refined re-sampling of historical data instead of parametric models for the risk factors. This approach preserves empirical features like contemporaneous dependence or heavy-tailed innovations even with very large numbers of variables. In order to assure consistency over time, re-sampling is done on carefully transformed variables, and some extensions like a GARCH volatility generator are added.

5.3 Models for Interest Rates and FX Rates

This section provides an overview of the different approaches to interest rate modelling with the main aim of motivating the choice of the modelling approach that will be pursued in the sequel of this thesis. Interest rate modelling is a very wide field with a number of quite different schools of thought and hundreds of specific models. A concise introduction is [25], a textbook-level description of modelling approaches is provided by [26], whereas [74] is a comprehensive reference manual. Aspects of multi-economy term structure models are investigated in [84], and [71] puts the emphasis on distinguishing the different schools.
The baseline is that we need a model class that produces positive interest rates, that is suited for projections far into the future and that can cope with multi-economy settings. It—almost—goes without saying that the model class should be arbitrage-free within one economy as well as across economies. The terminology related to interest rates is defined in detail in Section 6.1.1.

The most classical interest rate models belong to the class of equilibrium models. That is, there is an underlying equilibrium of a representative agent economy from which the interest rates are derived; see Section 7.1 for some related considerations. This economy and its equilibrium need not be known explicitly, but they can be shown always to exist. Implied interest rate processes have mean-reverting features, thus assuring numerical stability for long-horizon projections. Classical examples include the Vasicek and CIR models (single factor) or, more generally, the class of affine models (including multi-factor models). Single-factor models are often also referred to as short rate models, since the only random factor is identified with the short rate process, that is

\[ dr(t) = \alpha(r(t))dt + \sigma(r(t))dW(t). \]

This is somewhat misleading as there are always more or less explicit forms for the ZCB prices via

\[ P(t,T) = \mathbb{E}^Q \left[ \exp \left\{ - \int_t^T r(s)ds \right\} \big| \mathcal{F}_t \right], \]

where \( Q \) is the equivalent martingale measure. Equilibrium models in their pure form are time-homogeneous. Thus, the term structure is a function of a finite number of variables, so that the term structure of the model cannot generally fit an observed initial term structure, and, hence, cannot fully incorporate the information contained therein on the anticipated future course of events.

This gave rise to the development of models that can exactly reproduce the initial term structure. Such models are termed no-arbitrage models. This term is somewhat misleading as time-homogeneous models are
also internally arbitrage-free. Only the mismatch between modelled and actual initial term structures may cause arbitrage opportunities in practice. At first, no-arbitrage models consisted of equilibrium models with a deterministic time-inhomogeneous extension, for instance the Hull-White model used in the Barrie & Hibbert ESG; see [69].

Then came Heath-Jarrow-Morton (HJM). This was a major revolution mainly driven by practitioners who needed better methods for the valuation of derivatives while not caring too much about macroeconomic foundations. The basic idea of HJM is to consider the evolution of the entire forward curve, that is

\[ df(t, T) = \alpha(t, T)dt + \sigma(t, T)dW(t). \]

The conceptual difference to equilibrium models is that in general no underlying economic equilibrium is assumed. Notice however that, under some regularity conditions, most interest rate models can be represented in the HJM framework, including equilibrium models. The HJM approach has become extremely popular for derivatives valuation and for interest rate risk management, but less so for scenario generation, since it does not put emphasis on long-term stability.

The class of market models, most prominently represented by Brace-Gatarek-Musiela (BGM) grew out of the HJM approach. The key idea is to model directly observable quantities like LIBOR rates or swap rates instead of the short rates or instantaneous forward rates that are only observable through proxies. BGM then assumes that these rates follow a log-normal law. Market models, like HJM, have become popular for derivatives valuation and for short-term risk management, but they have not yet made their way into DFA economic scenario generation. For the latter task, one would have to use swap market models rather than LIBOR market models because of the focus on longer terms to maturity.

The class of price kernel models, which is related to equilibrium models, grew in parallel to HJM and market models. The base element is a positive diffusion process \((\xi(t))_{t \geq 0}\) adapted to some filtration \((\mathcal{F}_t)_{t \geq 0}\), and ZCB
prices are defined by
\[ P(t,T) = \mathbb{E}^P[\xi(T)|\mathcal{F}_t]/\xi(t), \]
where \( P \) is some probability measure, for instance, the real-world measure. If \( \xi(t) \geq 0 \) is, moreover, a supermartingale, then interest rates remain positive almost surely, thus establishing the positive interest framework. Contrary to the other model classes, where only very specific models assure positive interest, we have here an entire, fairly generic model class which satisfies this requirement that is crucial for long-term projections. On top of this, it can be shown that, under some regularity conditions, the FX rate between two economies is given by the ratio of the price kernels. Given all these facts, price kernel models are the natural choice for interest rate and FX rate modelling in DFA, and this choice is reinforced by further advantages. First, price kernels can be generated from underlying diffusions with flexible dimension, thus allowing for multi-factor models. And second, the price kernel provides a generic means for the valuation of all kinds of interest rate-contingent claims under the real world measure; Chapter 7 provides the details. There are two particular construction methods for price kernels, namely the martingale or Flesaker-Hughston approach and the potential approach due to Rogers. Within the realm of DFA scenario generation, the former approach is explored by Cairns [23, 24], whereas this thesis deals with the latter approach.

Current modelling approaches beyond the ones above include Markov-functional models and the chaotic approach. Most of the models belonging to the aforementioned classes feature bond prices that are some function of an underlying Markov process, and the Markov-functional approach is aimed at generalizing this to find new and richer model structures. The chaotic approach takes the price kernel as its starting point and uses a Wiener chaos representation in order to find more general and richer models than can be found with conventional construction methods. So, both approaches essentially try to extend the range of models based on Brownian motion. On the other hand, little is done to incorporate other types of basic processes, like \( \alpha \)-stable Lévy motion into interest rate modelling. This may be due to the fact that, in these cases, one comes into the realm of incomplete markets, and much of the convenient arbitrage theory would have to be reworked thoroughly.
Chapter 6

Phenomenology

6.1 Terminology

6.1.1 Interest Rates

The purpose of this section is to introduce the terminology for the fixed-income quantities, as well as some general relations that hold independently of model assumptions. We let $t$ denote time, and we assume a continuous-time setup, i.e. $t \in \mathbb{R}_0^+$. General reference is [74].

For all $0 < t < T < \infty$, let $P(t, T)$ denote the price at time $t$ of a zero-coupon bond (ZCB), i.e. the price of a security that pays for sure one unit of currency at the time of maturity $T \geq t$. Sometimes, it may be more appropriate to consider the time to maturity, i.e. $\tau := T - t$.

Unless otherwise stated, we always consider continuous compounding of interest. Given this, for all $0 \leq t \leq T < \infty$, let $R(t, T)$ denote the yield to maturity of the zero-coupon bond $P(t, T)$, defined by

$$P(t, T) \cdot e^{(T-t)R(t, T)} = 1,$$
which is equivalent to

$$R(t, T) = -\frac{\log P(t, T)}{T - t}.$$ 

For every fixed $t \geq 0$, the set $\mathcal{R}(t) := \{(t, R(t, T)) : T > t\}$ is the term structure of interest rates, or the yield curve.

For $t \geq 0$, let $r(t)$ denote the instantaneous, continuously compounded risk-free interest rate, commonly referred to as the short rate; it is defined as

$$r(t) := \lim_{T \to t} R(t, T).$$

For all $t > 0$, let $B(t)$ denote the value of the money market account at time $t$, defined as

$$B(t) := B(0) \exp \left\{ \int_0^t r(s) ds \right\}.$$

We make two important assumptions on the functioning of the underlying markets. First, we assume absence of arbitrage and second, we also assume absence of credit risk. These assumptions imply the following rules to hold, irrespective of more specific modelling assumptions:

1. Any ZCB price process has a deterministic terminal value, i.e.

$$P(T, T) = 1, \quad T \geq 0.$$

2. A ZCB price will always have a positive value, and it will never exceed its terminal value:

$$0 < P(t, T) \leq 1, \quad 0 \leq t \leq T < \infty.$$

3. A ZCB price cannot exceed the price of a ZCB with shorter maturity:

$$P(t, T_2) \leq P(t, T_1), \quad 0 \leq t \leq T_1 \leq T_2 < \infty.$$

4. Interest rates shall not be negative:

$$R(t, T) \geq 0, \quad 0 \leq t \leq T < \infty.$$
6.1. Terminology

5. The value of a ZCB must always be equal to the value of a replicating portfolio of ZCBs and forward rate agreements.

For all $0 \leq t \leq T_1 \leq T_2 < \infty$, let $f(t, T_1, T_2)$ denote the forward rate - or the yield to maturity of a forward rate agreement - i.e. the yield to maturity, agreed-upon at time $t$, for a zero-coupon bond starting at time $T_1$ and maturing at time $T_2$. The law of one price then implies the following relation:

$$f(t, T_1, T_2) = \frac{\log P(t, T_1) - \log P(t, T_2)}{T_2 - T_1},$$

where $f(t, T_1, T_2)$ is called the forward rate. For $0 \leq t \leq T$, let $f(t, T)$ denote the instantaneous forward rate, i.e. the instantaneous, continuously compounded interest rate, agreed-upon at time $t$, valid for time $T$. It is defined as

$$f(t, T) := \lim_{\delta \to 0} f(t, T, T + \delta), \quad 0 \leq t \leq T < \infty.$$

We find the following relations for the instantaneous forward rate:

$$f(t, T) = -\frac{\partial}{\partial \tau} \log P(t, \tau) \bigg|_{\tau=T} = -\frac{1}{P(t, T)} \frac{\partial}{\partial T} P(t, T),$$

$$P(t, T) = \exp \left\{ - \int_t^T f(t, s) ds \right\},$$

$$r(t) = f(t, t).$$

When we are in an international setup, where the yield curves of several economies are considered, we simply add an index denoting the respective economy to each of the objects defined above: $P^{(k)}(t, T)$, $R^{(k)}(t, T)$, $f^{(k)}(t, T)$, $r^{(k)}(t)$ and $B^{(k)}(t)$. By convention, $k = 1$ denotes the domestic economy.

6.1.2 Foreign Exchange Rates

Assume that we consider $K$ different economies. We use here the term "economy" rather than "country", since there is not necessarily a one-to-one connection between country and currency, the most prominent example being the Euro. We assume that each one of the $K$ economies has its
own distinct currency. Let \( i \) and \( j \) denote two of these economies. The \textit{spot exchange rate} \( C_{ij}(t) \) denotes the spot price as of time \( t \), expressed in units of currency \( i \), to be paid for one unit of currency \( j \). Absence of arbitrage implies the following rules to hold for all \( t \in \mathbb{R}_0^+ \) independent of any model:

\[
C_{ij}(t) > 0, \quad i, j \in \{1, \ldots, K\},
\]

\[
C_{ii}(t) = 1, \quad i \in \{1, \ldots, K\},
\]

\[
C_{ij}(t) = 1 / C_{ji}(t), \quad i \in \{1, \ldots, K\},
\]

\[
C_{ij}(t) = C_{ik}(t)C_{kj}(t), \quad i,j,k \in \{1, \ldots, K\}.
\]

These rules give rise to a considerable simplification in the modelling of the FX rates. Indeed, if we select some \textit{domestic} or \textit{pivotal} currency, denoted by \( i = 1 \), then it suffices to set up stochastic models for the exchange rates \( C_{1j}(t) \) for all \( t = 2, \ldots, K \), whereas all other rates follow deterministically from the above-stated relations.

Besides the spot exchange rate \( C_{ij}(t) \), we will also consider the \textit{forward exchange rate}, denoted by \( F_{C_{ij}}(t, T) \), where \( 0 \leq t \leq T < \infty \). This is the number of units of currency \( i \) that is paid at time \( t \) for one unit of currency \( j \) to be delivered at the future time \( T \). Absence of arbitrage implies that the \textit{covered interest parity} holds, see [18] (p.793ff):

\[
F_{C_{ij}}(t, T) = C_{ij}(t)e^{(T-t)(R_i(t,T)-R_j(t,T))}.
\]

Two important quantities that are often used in the statistical analysis of the FX markets are the \textit{depreciation rate}

\[
c_{ij}(t, T) = \log C_{ij}(T) - \log C_{ij}(t),
\]

which is the log-return over the period \([t, T]\), and the \textit{forward premium}

\[
p_{ij}(t, T) = \log F_{C_{ij}}(t, T) - \log C_{ij}(t).
\]

### 6.2 A Survey of Stylized Facts

It was emphasized in Section 3.1 and Section 5.1 that one of the most important requirements for a DFA scenario generator — and, hence, its
interest rate component – is the faithful reproduction of the empirical
behaviour of the variables modelled. It is difficult to make specific state-
ments on the behaviour of "the interest rates" or "the FX rates", since,
depending on the specific sample used, one might obtain rather diverse
results. However, there is a number of **stylized facts** quite generally ob-
servable in interest rate and FX data. We give a summary of these stylized
facts in this section, whereas the next section verifies their prevalence in
the specific data samples used in this thesis.

We start with a remark on the use of historical data. Calibrating pro-
jection models to historical data makes the implicit assumption that the
past of a variable is a reliable prediction of its future. This assumption
is sensible as long as there are no **structural breaks** in the mechanisms
governing the behaviour of the data. However, such structural breaks
occur sometimes. An important example is the Bretton-Woods system
which pegged the exchange rates of the major currencies in a 1%-band.
This system broke down in 1971, and exchange rates were allowed to float
freely afterwards. Another structural break occurred in the late 1970ies,
when the US Federal Reserve introduced new policies for setting their key
interest rates with the aim of controlling inflation. These policies were
subsequently also adopted by other central banks. For these reasons, it is
not advisable to use historical interest rate and FX rate data from before
the early 1980ies. Another problem is the introduction of the Euro in
1999, leading to the lack of sufficiently long historical data series. There
are, however, ways to construct artificial Euro data for times before 1999
if needed; for instance, Bloomberg’s Theoeuro index. More information
on the historical background behind interest rates and FX rates can be
found in [92] and [74].

The stylized facts on interest rates relevant for DFA scenario generation
are discussed in [24], [69] and, to some extent, in Chapter 8 of [41]. Here
is a summary, including also FX rates:

1. There are no negative interest rates, but interests can vary within
a broad range over time. Values from close-to-zero (Japan) up to
almost 20% (UK, late 1980ies) could be observed. On the other
hand, interest rates do not grow to arbitrarily high levels, except in countries prone to hyperinflation.

2. Interest rates are very persistent, that is, the autocorrelation is close to one, and they tend to fluctuate around local equilibria for extended amounts of time. However, they can also move towards the next local equilibrium quite swiftly, particularly under the influence of external shocks. So, there is mean reversion in interest rates, but the equilibrium can fluctuate slowly over long time spans.

3. The shape of the yield curve is upward-sloping most of the time; the slope can be more or less steep. There is, however, a non-negligible probability that the yield curve is downward-sloping (inverted) or hump-shaped.

4. Short-term rates are significantly more volatile than long-term ones. In some economies, rates with time-to-maturity around one to three years are also more volatile than rates with maturities below one year; see [17].

5. There is high, but imperfect positive contemporaneous correlation between interest rates of different maturities. Correlation tends to be highest for rates with similar maturities and for long maturities in general.

6. There is also high, but imperfect correlation between interest rates of different economies, and interest rate levels or movements have some explanatory power for the FX rate. This is related to weakened forms of the Purchasing Power Parity (PPP); see [15] for further investigations.

7. FX rates are always positive, and they are also quite persistent. Given their dependence on interest rates, FX rates can also fluctuate around local equilibria for extended amounts of time. Moreover, FX rates may have drifts over longer time spans, so-called Peso effects.

8. Heavy-tailed innovations are an issue for short maturities and low time aggregation (e.g. daily). For longer maturities and higher time aggregation, heavy tails tend to be less of a problem.
6.2. A Survey of Stylized Facts

There is a vast econometric literature investigating any conceivable empirical aspect of interest rates and FX rates. It is beyond the scope of this thesis to provide even a brief overview. Good starting points are [28], [74] and [39]. We only take a quick look at a few additional aspects that may be of some importance in the realm of this thesis.

Numerous forecasting relations are postulated between interest rates and FX rates. The most popular one is the expectation hypothesis which states that the forward rate is a predictor of the future short rate, that is $E[r(T)|\mathcal{F}_t] = f(t, T)$ for $t \leq T$. In this basic form, the hypothesis is rejected by almost every data set. However, versions including constant or time-varying term premia like $E[r(T)|\mathcal{F}_t] = f(t, T) + p(t, T)$ can be reconciled with the data. If one knows the empirical term premium, one can compare it with its theoretical counterpart in the model in order to assess the suitability of the latter; see [7], for instance.

An analogous approach can also be taken with spot and forward FX rates, postulating that $E[C_{ij}(T)|\mathcal{F}_t] = F_{C_{ij}}(t, T)$. This hypothesis too is rejected by the data; this phenomenon is called forward premium anomaly. See [8] or [84] for more details.

Given the low amounts of historical data that we have due to the high time aggregation, it would be useful to extract information from data with lower time aggregation. If it comes to volatility structures, this can be achieved through scaling laws. A scaling law is a relation between the observation time scale $\Delta t$ and the powers of the absolute returns of some traded asset, formally

$$(E[|r_{\Delta t}|^p])^{1/p} = c(p)(\Delta t)^{D(p)},$$

where $c(p)$ and $D(p)$ are deterministic functions. Scaling laws were extensively investigated for FX rates in the realm of high-frequency finance; see Section 5.5 of [39]. Empirical investigations revealed tractable relations between $\Delta t$ and $(E[|r_{\Delta t}|^p])^{1/p}$ for $t = 1, 2$ and $\Delta t$ ranging from 10 minutes up to two months. Such relations can be used for obtaining additional input for calibration, or they can be related to their empirical counterparts for assessment purposes.
6.3 Evaluation of Data Sets Used

While Section 6.2 gives an overview of general stylized facts of yield curve and FX data, we investigate in this section the related properties of the specific data sets that will be used later in this thesis for the calibration and backtesting of our models.

The sample used includes 201 monthly observations starting in January 1988 and ranging up to September 2004. It consists of government bond yield curve data for Switzerland (henceforth denoted CHF) and the United States (USD), and exchange rate data for the two currencies. For both economies, we have yields for 32 times-to-maturity, namely 3 months, 6 months and 1 year up to 30 years. The yield curve data for after 05/1994 (CHF) and 01/1991 (USD) stems from the Fair Market Curve Index series of Bloomberg which are computed from a wide range of traded instruments and which are considered as state-of-the-art among practitioners. Earlier observations were taken from the National Bank (CHF) and the Federal Reserve (USD). A smooth transition was assured by mixing the two data sources over a one-year intermediate period, which may sound problematic. However, it is not, since the differences between the two data sources were found to be minimal in the full overlapping period. All exchange rate data is from Bloomberg.

Figure 6.1 shows the trajectories of interest rates from both economies for a representative set of times-to-maturity, and Table 6.1 presents the related numbers. We can see that the average term structure is upward-sloping with a tendency to flatten above five years. There are, however, instances where the yield curve is downward-sloping (10.5% of observed months for CHF, 3.5% for USD), and there are instances where the yield curve is neither upward- nor downward-sloping (15.24% for CHF, 14.93% for USD). Moreover, the shapes have a strong tendency to cluster.

In general, interest rates are highly persistent. Table 6.1 indicates that autocorrelations at lag 1 are close to one for all maturities. The autocorrelation diagnostic plots (see Appendix A for a description) in Figure 6.2
### 6.3. Evaluation of Data Sets Used

#### CHF Yields: Levels

<table>
<thead>
<tr>
<th>TTM</th>
<th>Mean</th>
<th>StDev</th>
<th>Skew</th>
<th>Kurt</th>
<th>ACF(1)</th>
<th>Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>3m</td>
<td>3.62%</td>
<td>2.79%</td>
<td>0.76</td>
<td>2.31</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>1y</td>
<td>3.54%</td>
<td>2.43%</td>
<td>0.60</td>
<td>2.11</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>3y</td>
<td>3.68%</td>
<td>1.81%</td>
<td>0.44</td>
<td>2.04</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>5y</td>
<td>3.87%</td>
<td>1.55%</td>
<td>0.41</td>
<td>2.06</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>10y</td>
<td>4.24%</td>
<td>1.26%</td>
<td>0.45</td>
<td>2.11</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>20y</td>
<td>4.64%</td>
<td>1.00%</td>
<td>0.55</td>
<td>2.22</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
</tbody>
</table>

#### CHF Yields: 1st Differences

<table>
<thead>
<tr>
<th>TTM</th>
<th>Mean</th>
<th>StDev</th>
<th>Skew</th>
<th>Kurt</th>
<th>ACF(1)</th>
<th>Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>3m</td>
<td>-0.01%</td>
<td>0.33%</td>
<td>0.78</td>
<td>5.20</td>
<td>0.29</td>
<td>0.09</td>
</tr>
<tr>
<td>1y</td>
<td>-0.01%</td>
<td>0.27%</td>
<td>0.51</td>
<td>3.95</td>
<td>0.34</td>
<td>&lt;0</td>
</tr>
<tr>
<td>3y</td>
<td>-0.01%</td>
<td>0.21%</td>
<td>0.34</td>
<td>3.03</td>
<td>0.45</td>
<td>&lt;0</td>
</tr>
<tr>
<td>5y</td>
<td>-0.01%</td>
<td>0.19%</td>
<td>0.19</td>
<td>2.93</td>
<td>0.46</td>
<td>&lt;0</td>
</tr>
<tr>
<td>10y</td>
<td>-0.01%</td>
<td>0.16%</td>
<td>-0.16</td>
<td>3.22</td>
<td>0.41</td>
<td>&lt;0</td>
</tr>
<tr>
<td>20y</td>
<td>-0.01%</td>
<td>0.16%</td>
<td>-0.23</td>
<td>4.76</td>
<td>0.31</td>
<td>&lt;0</td>
</tr>
</tbody>
</table>

#### USD Yields: Levels

<table>
<thead>
<tr>
<th>TTM</th>
<th>Mean</th>
<th>StDev</th>
<th>Skew</th>
<th>Kurt</th>
<th>ACF(1)</th>
<th>Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>3m</td>
<td>4.70%</td>
<td>2.11%</td>
<td>-0.16</td>
<td>2.35</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>1y</td>
<td>5.06%</td>
<td>2.13%</td>
<td>-0.24</td>
<td>2.34</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>3y</td>
<td>5.68%</td>
<td>1.86%</td>
<td>-0.17</td>
<td>2.48</td>
<td>0.99</td>
<td>&lt;0</td>
</tr>
<tr>
<td>5y</td>
<td>6.05%</td>
<td>1.62%</td>
<td>0.00</td>
<td>2.34</td>
<td>0.98</td>
<td>&lt;0</td>
</tr>
<tr>
<td>10y</td>
<td>6.48%</td>
<td>1.42%</td>
<td>0.20</td>
<td>2.15</td>
<td>0.98</td>
<td>&lt;0</td>
</tr>
<tr>
<td>20y</td>
<td>6.84%</td>
<td>1.23%</td>
<td>0.24</td>
<td>1.99</td>
<td>0.98</td>
<td>&lt;0</td>
</tr>
</tbody>
</table>

#### USD Yields: 1st Differences

<table>
<thead>
<tr>
<th>TTM</th>
<th>Mean</th>
<th>StDev</th>
<th>Skew</th>
<th>Kurt</th>
<th>ACF(1)</th>
<th>Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>3m</td>
<td>-0.02%</td>
<td>0.21%</td>
<td>-0.59</td>
<td>4.01</td>
<td>0.53</td>
<td>&lt;0</td>
</tr>
<tr>
<td>1y</td>
<td>-0.02%</td>
<td>0.25%</td>
<td>0.01</td>
<td>3.09</td>
<td>0.45</td>
<td>&lt;0</td>
</tr>
<tr>
<td>3y</td>
<td>-0.02%</td>
<td>0.27%</td>
<td>0.19</td>
<td>2.54</td>
<td>0.38</td>
<td>&lt;0</td>
</tr>
<tr>
<td>5y</td>
<td>-0.02%</td>
<td>0.27%</td>
<td>0.23</td>
<td>2.56</td>
<td>0.35</td>
<td>&lt;0</td>
</tr>
<tr>
<td>10y</td>
<td>-0.02%</td>
<td>0.23%</td>
<td>0.27</td>
<td>2.59</td>
<td>0.29</td>
<td>&lt;0</td>
</tr>
<tr>
<td>20y</td>
<td>-0.02%</td>
<td>0.20%</td>
<td>0.27</td>
<td>2.66</td>
<td>0.24</td>
<td>&lt;0</td>
</tr>
</tbody>
</table>

#### USD / CHF Exchange Rate

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>StDev</th>
<th>Skew</th>
<th>Kurt</th>
<th>ACF(1)</th>
<th>Tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{ij}$</td>
<td>0.70</td>
<td>0.08</td>
<td>0.35</td>
<td>2.50</td>
<td>0.96</td>
<td>&lt;0</td>
</tr>
<tr>
<td>$c_{ij}$</td>
<td>0.00</td>
<td>0.03</td>
<td>0.59</td>
<td>3.19</td>
<td>0.31</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 6.1: Descriptive statistics for the yield curve and FX rate data for CHF and USD. Values in italics are not significantly different from zero on the 5%-confidence level. The statistics used are described in Appendix A.1. Further discussions are provided throughout Section 6.3.
Figure 6.1: Trajectories of selected interest rates over time: 1y (solid), 5y (dashed), 10y (dash-dotted) and 20y (dotted). The dots below the trajectories indicate months with downward-sloping yield curve; the dots above the trajectories indicate months with yield curve that is neither upward-nor downward-sloping. See p. 60 for further discussions.
Figure 6.2: Autocorrelation diagnostic plots (according to Appendix A.2) of the 5y-rate rates of USD (top) and CHF (bottom) for both levels (left) and first differences (right). The plots clearly confirm the very high persistence of the interest rate time series; see p. 60 for further discussions. Corresponding plots for other times to maturity show very similar results.

provide a more detailed picture: there is very strong autoregressive dependence, but also some moving-average component. Moreover, since there is even significant autocorrelation in the first differences, we must assume that the series are integrated of order greater than one.

Besides being highly persistent, it also appears that the rates have a general downward trend over the observation period. However, the average of the first differences is not significantly different from zero. Anyway, the trajectories in Figure 6.1 suggest that it is problematic to use historical data for estimating expected levels of future interest rates. Depending on
Figure 6.3: As a proxy for the conditional volatility, the annualized variances of lagged differences, $h^{-1} \text{Var}[R(X_{t+h}, \tau) - R(X_t, \tau)]$ are plotted against the time lag $h$ for different times to maturity $\tau$. The hump-shaped volatility structure is clearly visible; see p. 64 for further discussions.

The sample length used, one can obtain a wide range of different estimates.

Volatilities, as expressed by the standard deviations of levels and first differences in Table 6.1 are highest around one year and decrease towards longer maturities. Also, they tend to be somewhat lower for maturities below one year. To get more information on the conditional dynamics, we investigate the variance of lagged differences of yields, i.e.

$$h^{-1} \text{Var}[R(X_{t+h}, \tau) - R(X_t, \tau)],$$

annualized to put all measurements on equal footing. As is argued in [88], this is a good proxy for the conditional variance

$$h^{-1} \text{Var}[R(X_{t+h}, \tau)|R(X_t, \tau)]$$

(which is itself hard to measure) in the case of time series with very high autocorrelation. Moreover, the theoretical counterpart is readily available;
6.3. Evaluation of Data Sets Used

![CHF Levels and USD Levels](image1)

![CHF 1st Differences and USD 1st Differences](image2)

**Figure 6.4:** Contemporaneous correlations (z-axis) between interest rates of different times to maturity (x- and y-axis) within each economy for both levels and first differences. The correlation is very high, but clearly not perfect; see p. 65 for further discussions.

see (9.35). Figure 6.3 shows estimates across time lags of one month to five years for a representative set of times to maturity for both economies. One can clearly see the phenomenon known as the *hump-shaped term structure of conditional volatility* which is widely discussed in the interest rate literature; see e.g. [88] or [74]. The differences in the general levels of the graphs for different maturities reflect the fact the unconditional variances are the highest for short times to maturity, decreasing considerably for longer times to maturity; see also Table 6.1.

The contemporaneous correlation between interest rates of different ma-
Yields at equal TTM FX rate against yields

Figure 6.5: Contemporaneous correlations of interest rates of equal maturity across economies (left) and between FX rates and interest rates (right). Solid lines represent levels, dashed lines represent first differences. In the right panel '*' denotes USD, and '+' denotes CHF; see p. 66 for further discussions.

The distributions of the interest rate levels tend to be significantly skewed for CHF, but less so for USD, and excess kurtosis is absent throughout. The distributions of the first differences are only significantly skewed for some shorter maturities for which there is also some excess kurtosis. There are no significantly heavy tails, with the exception of the 3m-rate for CHF. The FX rate, both levels and log-returns, shows features similar to the
6.3. Evaluation of Data Sets Used

Figure 6.6: Some graphical evaluations of the FX rates: trajectories for levels (NW) and log-returns (SW), autocorrelation diagnostic plots (according to Appendix A.2) for levels (NE) and log-returns (SE).

ones described for the interest rate. Figure 6.6 and Table 6.1 provide more information. Note that the FX log-returns show significantly heavy tails.

As we are operating in the realm of multi-factor models, the number of independent random factors to be used is an important issue. It is addressed by a Principal Component Analysis (PCA); see [115] as a general reference or [74] for the use of PCA with yield curve data. The first differences of the yield curve samples as described above were used, and there was one PCA for the CHF data, one for the USD data and one on the combined sample. The results are summarised in Figure 6.7. The bar graphs on the upper and middle left confirm the conventional wisdom that most of the yield curve variability is explained by three random factors (CHF: 97.12%, USD: 98.66%), and even two factors might be sufficient (CHF: 92.87%,...
USD: 97.21%). For the combined sample, it takes four (95.18%) or five (97.20%) factors, with three factors appearing as the absolute minimum (92.15%). The panels on the right-hand side of Figure 6.7 show Hotelling's $t^2$ statistic, which can be interpreted as a measure of distance from the average in the high-dimensional sample space, plotted against time. This information can be used in model validation in order to explain predictive failures of calibrated models. For the USD data, in particular, we can see that the extreme observations are quite strongly clustered in the late nineties.

In conclusion, the evaluations of our data set confirm the stylized facts on interest and FX rates from the literature cited in Section 6.2. Regarding model selection, the evaluations clearly reject single-factor models such as CIR, as they are never capable of reproducing the observed behaviour. Moreover, a good model should also include some possibility to add time-inhomogeneity. This is because we cannot devise sensible estimates for the expected future interest rate levels from the data. Depending on the time lag, we can have a wide range of mean values. Therefore, it should be possible to account for market expectations as expressed e.g. by the initial yield curve. On the other hand, heavy tails are not very much of a problem on this level of time aggregation. Nevertheless, we should not forget that there is always a possibility for extreme events, and we should at least add a few stress scenarios in practical evaluations.
Figure 6.7: Results of the PCA of the yield curve data for CHF (top), USD (middle) and CHF and USD jointly (bottom). The bars in the left panel show the percentage of variance explained by the respective principal component, whereas the solid line shows the cumulated sum. The dashed and dotted lines are at 90% and 95%. The right panels show Hotelling’s $t^2$ against time. The solid, dashed and dotted lines represent the 75%, 90% and 95% levels respectively. For more information see p. 67.
Chapter 7

The Theory Behind the
Price Kernel

7.1 Introduction, Background, Motivation

7.1.1 Introduction and Motivation

The modelling approach used in this thesis belongs to the general class of price kernel models, sometimes also called state price deflator models. Price kernel models have a somewhat marginal status in the wide universe of interest rate models. The more popular classes of models are short rate models (e.g. Vasicek or Cox-Ingersoll-Ross), forward rate models (e.g. general Heath-Jarrow-Morton models), and market models (e.g. Brace-Gatarek-Musiela); see Part II of [74]. There are, however, good reasons for this choice, and there also exist close relations between price kernel models and the other model classes.

The main motivation for the choice of the price kernel approach arises from the requirement that we have to set up an international term structure model that must simultaneously incorporate the yield curves of several
economies and the related currency exchange rates. Price kernel models provide a tractable way to satisfy these requirements in an economically sensible, i.e. arbitrage-free way. Moreover, state price models can be formulated in terms of the objective probability measure rather than the risk-neutral one. This is an advantage given that the model must be optimised so as to faithfully reproduce the observable behaviour of real-world yield curves. Moreover, a model formulated without the risk neutral measure may find more acceptance in the actuarial world, where the concept of risk-neutral pricing is not very well introduced.

The concept of the price kernel has been known in economics since the 1950s, when it was introduced by Arrow and Debreu in the context of von Neumann-Morgenstern utility theory. A comprehensive reference on the general theory in both discrete and continuous time is [48]. Despite the old age of the general concept of price kernels, its direct application to interest rate modelling is a more recent development. It started in 1992 with Constantinides' SAINTS model, see [36]. The development then continued on two paths: The first one was started by Chris Rogers and is generally referred to as the potential approach; see [106]. The second approach is the so-called positive interest or Flesaker-Hughston framework after its original creators; see e.g. [60]. Though different in appearance, the two approaches have the same underlying concepts and are basically equivalent. This is emphasized in some of the more recent publications on price kernel models, especially in [76] and [84]. The former is a main reference for the sequel of this chapter. There have already been applications of price kernel models in the realm of DFA scenario generation, notably Andrew Cairns' model [23] and the Smith-model (see www.thesmithmodel.com).

7.1.2 Digression: Economic Background

This section provides a short derivation of the price kernel in the classic setting of equilibrium theory and von Neumann-Morgenstern utility theory. The presentation is informal and for illustration only. The general reference is [48]; the specific presentation follows [84]. Consider an economy where a representative agent optimizes the expected discounted
utility of future consumption:

\[
\max_c \mathbb{E} \left[ \int_t^\infty e^{-\rho s} \mathcal{U}(c_s) \, ds \bigg| \mathcal{F}_t \right]. \tag{7.1}
\]

Here \( \mathcal{U}(\cdot) \) is the von Neumann-Morgenstern utility function of the representative agent, \( c = (c_t)_{t \geq 0} \) is some consumption stream, and \( \rho \) is a constant discount factor. In equilibrium, the first-order condition for a specific consumption stream \( c^* \) to be optimal is

\[
\mathbb{E} \left[ \frac{\partial \mathcal{U}(c^*(T))}{\partial c^*(T)} \cdot \frac{K(T)}{e^{\rho T}} - \frac{\partial \mathcal{U}(c^*(t))}{\partial c^*(t)} \cdot \frac{\pi(t, K(T))}{e^{\rho T}} \bigg| \mathcal{F}_t \right] = 0, \tag{7.2}
\]

where \( \pi(t, K(T)) \) is the price at time \( t \) for buying a contingent claim that pays \( K(T) \) at time \( T \geq t \). In words, (7.2) means that the utility forfeited by spending \( \pi(t, K(T)) \) for buying a contingent claim must be compensated by the added utility of the future payoff \( K(T) \). Introducing a process \( \xi = (\xi(t))_{t \geq 0} \) defined as

\[
\xi(t) = e^{-\rho t} \frac{\partial \mathcal{U}(c^*(t))}{\partial c^*(t)}, \tag{7.3}
\]

one can restate (7.2) as

\[
\pi(t, K(T)) = \mathbb{E} \left[ \frac{\xi(T)}{\xi(t)} K(T) \bigg| \mathcal{F}_t \right]. \tag{7.4}
\]

Von Neumann-Morgenstern utility functions are characterized by a positive first derivative, hence we have \( \xi(t) > 0 \) for all \( t > 0 \). Moreover, it follows from (7.4) that the deflated price process \( \xi(t)\pi(t, K(T)) \) is a martingale. A process \( \xi = (\xi(t))_{t \geq 0} \) with these properties is called state price deflator or price kernel. Under the technical conditions used in Section 7.2, it is shown in [48] that there exists a price kernel if and only if there is no arbitrage.

Given this result, one could now set up a specific model for the underlying economy and the utility function of the representative agent and then solve (7.1) to obtain the price kernel \( \xi \) and, in turn, the term structure of interest rates by observing that \( K(T) \equiv 1 \) for a ZCB maturing at time \( T \). This path was actually followed in the construction of the famous
Cox-Ingersoll-Ross (CIR) model; see e.g. Section 11.5.9 of [74]. The problem with this explicit approach is that it is extremely difficult to find a setup that results in a model with sufficient statistical properties. For instance the CIR model, while providing a fairly good model for the short rate, cannot properly reproduce observed real-world yield curves; see [74].Moreover, there are also some problems with the validity of the concept of a representative agent economy; see [80]. That is, we could not take advantage of the fact that the processes driving such a model correspond to some real-world quantities.

It is, however, possible to specify directly a model for the price kernel $\xi = (\xi(t))_{t \geq 0}$, as long as one respects the required properties of the latter. This is the path followed by [36] and all the subsequent approaches to price kernel models stated above. In the sequel of this thesis, we will also move along these lines.

### 7.2 Specification of the Price Kernel

#### 7.2.1 Single-economy Case

While Section 7.1.2 is mainly for illustration and to put things into a historical context, this section provides a formal specification of the price kernel for a single economy. A comprehensive presentation of the theory is given in Chapters 5 and 6 of [48]; this presentation loosely follows [76]. Let $t \in \mathbb{R}_0^+$ denote time and let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, where $\mathbb{P}$ is the objective probability measure. All the uncertainty in the model is modelled by a $d$-dimensional standard Brownian motion $\mathbf{B} = (B^1, \ldots, B^d)'$, where $B^i = (B^i_t)_{t \geq 0}$. We specify the filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ as the natural filtration of $\mathbf{B}$ augmented by all $\mathbb{P}$-null sets. Let $(-\mu_t^\xi)_{t \geq 0}$ be a positive, $\mathcal{F}$-adapted process with continuous paths and define

$$A_t := -\int_0^t \mu_s^\xi \, ds, \quad t \geq 0, \quad (7.5)$$

which is an increasing $\mathcal{F}$-adapted process starting at $A_0 = 0$. We assume

$$\mathbb{E} [A_\infty^2] < \infty \quad \text{where} \quad A_\infty = \lim_{t \to \infty} A_t, \quad (7.6)$$
and we note that, through Jensen’s inequality, this also implies that $E[A_{\infty}] < \infty$. From the construction of $A_t$ it then also follows that $E[A_t] < \infty$ for all $t \geq 0$. Now we can define the process $\xi = (\xi(t))_{t \geq 0}$ as

$$\xi(t) := E[A_{\infty}|\mathcal{F}_t] - A_t, \quad t \geq 0.$$  \hspace{1cm} (7.7)

Before using $\xi$ as the price kernel, we explore some of its properties:

**Proposition 7.1** Let the process $\xi$ be defined as above. Then we have:

1. $\xi(t) > 0, \quad t \geq 0$,
2. $\xi$ is a $(\mathbb{P}, \mathcal{F})$-supermartingale,
3. $E[\sup_{t \geq 0} \xi^2(t)] < \infty$, and
4. $\lim_{t \to \infty} E[\xi(t)] = 0$.

**Proof.** (1) Using Equations 7.7 and 7.5 we obtain $\xi(t) = E[-\int_t^\infty \mu_s^\xi \, ds|\mathcal{F}_t]$, which is positive for all $t$ given the assumption that $-\mu_s^\xi$ is positive.

(2) By construction, $\xi$ is $\mathcal{F}$-adapted. Moreover, for all $0 < s < t < \infty$ we have

$$E[\xi(t)|\mathcal{F}_s] = E[A_{\infty}|\mathcal{F}_s] - E[A_t|\mathcal{F}_s] \leq E[A_{\infty}|\mathcal{F}_s] - A_s = \xi(s),$$

since $E[A_t|\mathcal{F}_s] = A_s + E[\int_s^t (-\mu_s^\xi) \, ds|\mathcal{F}_s] \geq A_s$. The $L^1$-property of $\xi$ then follows from the construction and from (7.6).

(3) See [76]

(4) Obvious, given (7.7).

**Remark 7.2** Given Proposition 7.1 we can assume, without loss of generality, that $(A_t)_{t \geq 0}$ is normalized such that $\xi(0) = E[A_{\infty}] = 1$.

**Remark 7.3** If a positive supermartingale with the limit property (4) of Proposition 7.1 is, moreover, Markovian, then it is called a potential, thus justifying the name potential approach for the method to be pursued in the sequel; see [106].
Remark 7.4 In a specific modelling effort, it is not necessary to specify explicitly the process \((-\mu_t^\xi)t\geq 0\) resp. \((\xi_t)t\geq 0\). One can start right away with a process \((\xi(t))t\geq 0\) which is a positive \((\mathbb{P},\mathcal{F})\)-supermartingale with \(\lim_{t\to\infty} \mathbb{E}[\xi(t)] = 0\) and \(\mathbb{E}[\sup_{0\leq t\leq \infty} \xi^2(t)] \leq \infty\). The Doob-Meyer decomposition (see Chapter 1.4 of [78]) then assures that \(\xi\) has a representation (unique up to \(\mathbb{P}\)-null sets) as in (7.7).

Let now \(K(T)\) be some \(\mathcal{F}_T\)-measurable random variable with \(\mathbb{E}[K(T)^2] < \infty\), interpreted as the payout of some contingent claim maturing at time \(0 < T < \infty\). For \(0 \leq t \leq T\), let \(\pi(t, K(T))\) denote the time-\(t\) price of the contingent claim \(K(T)\). We define

\[
\pi(t, K(T)) := \frac{\mathbb{E}[\xi(T)K(T)|\mathcal{F}_t]}{\xi(t)}, \quad 0 \leq t \leq T < \infty.
\]  

(7.8)

In the special case where the contingent claim is a ZCB, we have \(K(T) \equiv 1\), and the pricing relation from (7.8) simplifies to

\[
P(t, T) = \pi(t, K(T) \equiv 1) = \frac{\mathbb{E}[\xi(T)|\mathcal{F}_t]}{\xi(t)}, \quad 0 \leq t \leq T < \infty.
\]  

(7.9)

Lemma 7.5 Let the process \(\xi\) be defined as in (7.7), and let the ZCB prices be defined as in (7.9). Then we have for all \(0 \leq t \leq T < \infty\):

1. \(f(t, T) = \frac{1}{\mathbb{E}[\xi(T)|\mathcal{F}_t]} \frac{\partial \mathbb{E}[A_T|\mathcal{F}_t]}{\partial T} = \frac{1}{\mathbb{E}[\xi(T)|\mathcal{F}_t]} \mathbb{E} \left[ \frac{\partial A_T}{\partial T} | \mathcal{F}_t \right],\)

2. \(r(t) = \frac{1}{\xi(t)} \frac{\partial A_t}{\partial t} = -\frac{\mu_t^\xi}{\xi(t)}\).

PROOF. By using the definitions from Section 6.1.1 and calculating. The second equality in (1) is by Fubini. \(\square\)

Proposition 7.6 Let the process \(\xi\) be defined as above, and let the ZCB prices be defined as in (7.9). Then we have:

1. \(0 < P(t, T) \leq 1, \quad 0 \leq t \leq T < \infty,\)
2. \(P(T, T) = 1, \quad 0 \leq T < \infty,\)
3. \(P(t, T_2) \leq P(t, T_1), \quad 0 \leq t \leq T_1 \leq T_2 < \infty,\)
7.2. Specification of the Price Kernel

4. \( R(t,T) \geq 0 \) and \( f(t,T) \geq 0, \ 0 \leq t \leq T < \infty, \) and

5. \( \lim_{T \to \infty} P(0,T) = 0. \)

**Proof.** Recall the definition of \( P(t,T) \) from (7.9). Then:

(1) "> 0" follows from the positivity of \( \xi \) and "<< 1" from the supermartingale property of \( \xi; \) see Proposition 7.1 for both.

(2) Follows from the fact that \( \mathbb{E}[\xi(T)|\mathcal{F}_T] = \xi(T). \)

(3) Using the properties of conditional expectation and the supermartingale property of \( \xi \) from Proposition 7.1 yields

\[
\mathbb{E} [\xi(T_2)|\mathcal{F}_1] = \mathbb{E} [\mathbb{E} [\xi(T_2)|\mathcal{F}_{T_1}]|\mathcal{F}_1] \leq \mathbb{E} [\xi(T_1)|\mathcal{F}_1].
\]

(4) Recall the definitions of \( R(t,T) \) and \( f(t,T) \). The claims then directly follow from (1) and (3). (Note that \( P(t,T) \) is differentiable in \( T \).)

(5) Follows immediately from (4) of Proposition 7.1. \( \square \)

**Remark 7.7** The conclusion from Proposition 7.6 is that a price kernel model can produce yield curves that satisfy the basic requirements set forth in Section 6.1.1. This does, however, not yet mean that the model is arbitrage-free; this remains to be verified.

We move on to prove that the present modelling approach is arbitrage-free. The derivation of the no-arbitrage property will also establish the connection to the generally more popular modelling setup based on the equivalent martingale measure. We let

\[
M_t := \mathbb{E} [A_\infty|\mathcal{F}_t], \quad t \geq 0,
\]

and we note that \( (M_t)_{t \geq 0} \) is a martingale, square integrable due to the assumption stated in (7.6), and with càdlàg paths due to the definition of the filtration \( \mathcal{F} \). The martingale representation theorem ([78], Theorem 4.15) yields the representation

\[
M_t = \mathbb{E}[A_\infty] + \sum_{i=1}^{d} \int_0^t Y^i(s) dB^i(s) \quad \mathbb{P} - a.s. \quad t \geq 0,
\]

where, for all \( i = 1, \ldots, d, \) the processes \( (Y^i(t))_{t \geq 0} \) are unique and \( \mathcal{F} \)-adapted with \( \mathbb{E} \left[ \int_0^T (Y^i(t))^2 \ dt \right] < \infty \) for all \( T \geq 0. \) We also note that

\[
d\xi(t) = dM_t - dA_t = dM_t - r(t) \xi(t) \ dt,
\]
where the second equality is by Lemma 7.5. Now, the scene is set for

**Theorem 7.8** For each fixed $T \in [0, \infty)$, assume that (7.6) holds and that

$$
E \left[ \exp \left\{ \sum_{i=1}^{d} \int_{0}^{T} \frac{Y^i(t)}{\xi(t)} dB^i(t) - \frac{1}{2} \sum_{i=1}^{d} \int_{0}^{T} \left( \frac{Y^i(t)}{\xi(t)} \right)^2 dt \right\} \right] = 1. \quad (7.13)
$$

Then there exists a probability measure $Q$ equivalent to $P$ and described by

$$
\frac{dQ}{dP}\bigg|_{\mathcal{F}_T} = \xi(T)B(T) = \xi(T) \exp \left\{ \int_{0}^{T} r(t) \, dt \right\} \quad (7.14)
$$

under which the price $\pi(t, K(T))$ of any $\mathcal{F}_T$-measurable and square-integrable contingent claim $K(T)$ can be expressed as

$$
\pi(t, K(T)) = B(t) \mathbb{E}^Q \left[ \frac{K(T)}{B(T)} \big| \mathcal{F}_t \right] = \mathbb{E}^Q \left[ \exp \left\{ - \int_{t}^{T} r(s) \, ds \right\} K(T) \big| \mathcal{F}_t \right] \quad (7.15)
$$

for $0 \leq t \leq T < \infty$. In particular, the discounted price process $(\pi(t, K(T)))_{0 \leq t \leq T}$ is a $Q$-martingale.

**PROOF.** Essentially an application of Girsanov’s Theorem (see [78]); for details see Appendix A of [76].

**Remark 7.9**

1. Theorem 7.8 particularly applies to ZCBs, i.e. the situation where $K(T) \equiv 1$ and hence $\pi(t, K(T)) = P(t, T)$.

2. If Theorem 7.8 is applicable, then there is no arbitrage.

3. Verifying the assumption in (7.13) is non-trivial. A sufficient condition is that

$$
\left( \frac{Y^1(t)}{\xi(t)}, \ldots, \frac{Y^d(t)}{\xi(t)} \right)' \in \mathbb{R}^d
$$

is bounded for all $0 \leq t \leq T < \infty$; see [76]. Alternatively, one can revert to the Novikov condition, see [78].
7.2. Specification of the Price Kernel

By using Equations 7.11, 7.12 and 7.5, we obtain the following representation of the price kernel

\[
d\xi(t) = \mu^\xi_t \, dt + \sum_{i=1}^{d} Y^i(t) \, dB^i(t).
\]

(7.16)

We observe that the process \((\mu^\xi_t)_{t\geq 0}\) introduced at the very beginning is actually the drift of the price kernel. Moreover, under the setup prevailing here the volatility of the price kernel is completely determined by its drift: the processes \(Y^i(t)\) from (7.11) provide a unique representation of the martingale \(M_t\) which is, in turn, completely specified by \(\mu^\xi_t\). Applying Lemma 7.5 to (7.16) and re-arranging terms, we obtain

\[
\frac{d\xi(t)}{\xi(t)} = -r(t) \, dt - \sum_{i=1}^{d} \rho^i(t) \, dB^i(t),
\]

(7.17)

where we use the definition

\[
\rho^i(t) := -\frac{Y^i(t)}{\xi(t)}, \quad i = 1, \ldots, d, \quad t \geq 0,
\]

for the market price of risk with respect to the \(i\)-th risk factor \((B^i(t))_{t\geq 0}\). Defining the vector \(\rho(t) := (\rho^1(t), \ldots, \rho^d(t))^\prime\), we can write

\[
\frac{d\xi(t)}{\xi(t)} = -r(t) \, dt - \rho(t)^\prime dB(t),
\]

(7.18)

or, in explicit form

\[
\xi(t) = \exp\left\{ -\int_0^t r(s) \, ds - \frac{1}{2} \int_0^t \|\rho(s)\|^2 \, ds - \int_0^t \rho(s)^\prime dB(s) \right\}
\]

\[
= \exp\left\{ -\int_0^t r(s) \, ds \right\} \eta(t),
\]

(7.19)

where \((\eta(t))_{t\geq 0}\) is the Radon-Nikodym derivative of \(Q\) with respect to \(P\), that is,

\[
\eta(t) = E \left[ \frac{dQ}{dP} \bigg| \mathcal{F}_t \right] = E \left[ \exp \left\{ -\frac{1}{2} \int_0^t \|\rho(s)\|^2 \, ds - \int_0^t \rho(s)^\prime dB(s) \right\} \bigg| \mathcal{F}_t \right].
\]

(7.20)

Hence, under some regularity conditions, the price kernel and the equivalent martingale measure are just two different ways of expressing the same things.
Chapter 7. The Theory Behind the Price Kernel

7.2.2 Price Kernels in the Multi-Economy Case

The extension of the price kernel concept to a multi-economy setup is straightforward. We assume that we have \( n \) economies – denoted by \((1), \ldots, (n)\) – each one with its own currency. We assume that the uncertainty of all economies is modelled by one \( d \)-dimensional Brownian motion \( B = (B_t)_{t \geq 0} \) on the probability space \((\Omega, \mathcal{A}, P)\), where \( P \) is the global objective measure. The filtration \( \mathcal{F} = (\mathcal{F}_t)_{t \geq 0} \) is the \( P \)-augmented natural filtration of \( B \).

Now, we consider any two economies \((i)\) and \((j)\). The spot exchange rate between currencies \((i)\) and \((j)\) is denoted by \( C_{ij}(t) \), interpreted as the time-\(t\) price, expressed in units of currency \((i)\), for one unit of currency \((j)\); see Section 6.1.2 for more details and some model-independent no-arbitrage properties. Each economy has its own price kernel, namely

\[
\frac{d\xi^{(i)}(t)}{\xi^{(i)}(t)} = -r^{(i)}(t) \, dt - \rho^{(i)}(t)' dB_t \tag{7.21}
\]

and

\[
\frac{d\xi^{(j)}(t)}{\xi^{(j)}(t)} = -r^{(j)}(t) \, dt - \rho^{(j)}(t)' dB_t \tag{7.22}
\]

where \( r^{(\cdot)}(t) \) and \( \rho^{(\cdot)}(t) \) are the short rate and the market price of risk for the respective economies.

Now, assume that \( S^{(j)}(t) \) is the price of some traded asset in economy \((j)\), denominated in units of currency \((j)\). Then, according to the results of Section 7.2.1, \((\xi^{(j)}(t) \, S^{(j)}(t))_{t \geq 0}\) is a \( P \)-local martingale. Assuming complete markets, \( \xi^{(j)} \) is the only process (up to scaling) with this property. On the other hand, \( C_{ij}(t) \, S^{(j)}(t) \) is a traded asset in economy \((i)\), denominated in units of currency \((i)\). Therefore, we also have that \((\xi^{(i)}(t) \, C_{ij}(t) \, S^{(j)}(t))_{t \geq 0}\) is a \( P \)-local martingale. Hence, the exchange rate must equal the scaled ratio of the two price kernels, that is \( C_{ij}(t) = C_{ij}(0) \xi^{(j)}(t)/\xi^{(i)}(t) \). More formally, we have the following result:

**Proposition 7.10** Assume the probabilistic setup of Section 7.2.1 and Section 7.2.2, and let \( \xi^{(i)} \) as in (7.21), and \( \xi^{(j)} \) as in (7.22) be the price
kernels of two economies \((i)\) and \((j)\). Then the spot exchange rate \(C_{ij}(t)\) between the currencies of \((i)\) and \((j)\) is given by

\[
C_{ij}(t) = C_{ij}(0) \frac{\xi^{(j)}(t)}{\xi^{(i)}(t)} \frac{\xi^{(i)}(0)}{\xi^{(j)}(0)}, \quad t \geq 0.
\]

The \(P\)-dynamics of the exchange rate are given by

\[
\frac{dC_{ij}(t)}{C_{ij}(t)} = \left(r^{(i)}(t) - r^{(j)}(t) + \rho^{(i)}(t)\sigma_{C_{ij}}(t)\right) dt + \sigma_{C_{ij}}(t) dB_t,
\]

where the volatility of the exchange rate is given by

\[
\sigma_{C_{ij}}(t) = \rho^{(i)}(t) - \rho^{(j)}(t).
\]

**Proof.** See [84] or [106].

The consequences of Proposition 7.10 are discussed in [84]. We just note here that the characteristics of the exchange rates are fully determined by the characteristics of the involved price kernels, so that there is no need for the estimation of extra parameters.

7.3 Valuation by Using the Price Kernel

7.3.1 The Base Case

In the preceding sections of Chapter 7, the emphasis was mainly on the *projection* capabilities of the model class, that is, on its capability of producing scenarios that have certain desired properties like absence of arbitrage or positive interest rates. However, the model can also be used for *valuation*, that is, assigning fair present values to some future, uncertain cash flows. The theoretical foundations of the valuation approach are provided in Section 7.2; the present section mainly focuses on applications. Useful further references are [48] and [75].

Recall that all uncertainty in the model follows from a Markovian driver process \((X_t)_{t\geq 0}\) with natural filtration \((\mathcal{F}_t)_{t\geq 0}\). The price kernel is then given by \(\xi(t) = f(t, X_t)\) for some measurable function \(f(t, x)\). If \(K(T)\) is
an $\mathcal{F}_T$-measurable cash flow occurring at some given future time $T \geq t$, then, according to Section 7.2, the time-$t$ value of $K(T)$ is given by
\[
\pi(t, K(T)) = \frac{\mathbb{E}[\xi(T)K(T) | \mathcal{F}_t]}{\xi(t)}. \tag{7.23}
\]
We will usually let the valuation time be $t = 0$. The approach naturally extends to sequences of future cash flows. Let $t_0 \leq t_1 \leq \ldots, t_N < \infty$ be some sequence of time points, and let $K(t_n)$ be an $\mathcal{F}_{t_n}$-measurable cash flow for each $n = 1, \ldots, N$. Then we have
\[
\pi\left(t_0, (K(t_n))_{n=1}^N\right) = \sum_{n=1}^N \frac{\mathbb{E}[\xi(t_n)K(t_n) | \mathcal{F}_{t_0}]}{\xi(t_0)}. \tag{7.24}
\]
If we have a cash flow $K(T)$ that is deterministic or $\mathcal{F}_t$-measurable, then (7.23) turns into
\[
\pi(t, K(T)) = K(T)\frac{\mathbb{E}[\xi(T) | \mathcal{F}_t]}{\xi(t)} = K(T)\mathcal{P}(t, T), \tag{7.25}
\]
as one would expect from the classical theory. Finally, if we have a cash flow $K(T)$ that is independent of $\mathcal{F}_t$, then (7.23) becomes
\[
\pi(t, K(T)) = \mathbb{E}[K(T)] \frac{\mathbb{E}[\xi(T) | \mathcal{F}_t]}{\xi(t)} = \mathbb{E}[K(T)] \mathcal{P}(t, T). \tag{7.26}
\]
This means that non-systematic risk, i.e. risk that is not related to the uncertainty inherent in the modelled financial market, is not rewarded by the price kernel. The reason for this is that the whole theory is based on the assumption that the agent is capable of diversifying any non-systematic risk perfectly. If this assumption is not fulfilled, e.g. in the case of insurance risk, then this non-systematic risk has to be accounted for by other means; see Section 7.3.3 for some ideas.

The extension of the basic valuation principle in (7.23) to a multi-currency setting is straightforward. Suppose that we have two economies $(i)$ and $(j)$ with their respective pricing kernels $(\xi^{(i)}(t))_{t \geq 0}$ and $(\xi^{(j)}(t))_{t \geq 0}$ that are both adapted to the same filtration $(\mathcal{F}_t)_{t \geq 0}$. Recall from Section 6.1.2 that $C_{ij}(t)$ denotes the spot price in units of currency $(i)$ for one unit of
7.3. Valuation by Using the Price Kernel

currency \((j)\). Recall from Proposition 7.10 that the spot exchange rate is given by

\[
C_{ij}(t) = C_{ij}(0) \frac{\xi^{(j)}(t)}{\xi^{(i)}(t)} \frac{\xi^{(i)}(0)}{\xi^{(j)}(0)}.
\]

Let \(K^{(j)}(t)\) denote a cash flow denominated in currency \((j)\). The reporting currency, however, is \((i)\). So, the cash flow to be valued is \(C_{ij}(t) \cdot K^{(j)}(t)\). Combining the different elements, we obtain

\[
\pi^{(i)}(0, K(t)) = \mathbb{E} \left[ \xi^{(i)}(t) C_{ij}(t) K^{(j)}(t) \bigg| \mathcal{F}_0 \right] / \xi^{(i)}(0)
\]
\[
= \mathbb{E} \left[ \xi^{(i)}(t) C_{ij}(0) \frac{\xi^{(j)}(t)}{\xi^{(i)}(t)} \frac{\xi^{(i)}(0)}{\xi^{(j)}(0)} K^{(j)}(t) \bigg| \mathcal{F}_0 \right] / \xi^{(i)}(0)
\]
\[
= C_{ij}(0) \mathbb{E} \left[ \xi^{(j)}(t) K^{(j)}(t) \bigg| \mathcal{F}_0 \right] / \xi^{(j)}(0)
\]
\[
= C_{ij}(0) \pi^{(j)}(0, K(T)).
\] (7.27)

This means that the variability of the FX rate over time is not explicitly part of the valuation. The reason for this is that we assume a complete, frictionless and perfectly liquid financial market in which the agent can perfectly hedge the risk coming from the FX rate.

It is easy to see that the basic valuation formula in (7.23) can be evaluated by means of Monte Carlo simulation as long as the cash flow \(K(T)\) can be expressed as \(K(T) = g(T, X_T)\) for some sufficiently nice and explicitly known function. Hence, this valuation approach already covers a wide range of securities, including coupon bonds, floating rate bonds and all kinds of European interest rate derivatives. There are, however, notable cases where the basic valuation formula from (7.24) does not apply, namely

1. The cash flows \((K(t_n))_{n=1}^{N}\) depend on optimal stopping time, with the main example being American options. This case will be dealt with in Section 7.3.2.

2. The cash flow \(K(t)\) depends on both \(X_t\) and some exogenous random variable independent of \(\mathcal{F}_t\). Some ideas on how to deal with this case are given in Section 7.3.3.
7.3.2 Valuation of cash flows with optimal stopping

The valuation method set forth in Section 7.3.1 works well as long as cash flows \( K(t_n) \) can be expressed as an explicit function of the driver \( X_t \). There is, however, an additional degree of complication that requires special attention if the specification of the cash flows depends on an optimal stopping time, as e.g. in the case of American or Bermudian options. Then, the valuation problem to be solved is tantamount to the optimization

\[
\max_{\tau \in \mathcal{T}_{[0,T]}} \pi(0, K(\tau)), \quad (7.28)
\]

where \( \mathcal{T}_{[0,T]} \) is the set of all stopping times defined on the interval \([0,T]\). In order to be able to use the Monte Carlo scenarios, we introduce here a version of the Longstaff - Schwartz algorithm. This algorithm was originally presented in [90]. A more formal presentation plus proofs that the algorithm actually converges are given in [34]. The version of the algorithm developed here is adapted to the specific conditions of a price kernel model.

We consider a discrete, equispaced time frame \((t\Delta t)_{t=0}^T\). For notational simplicity, we let \( K_t \equiv K(t\Delta t) \), \( \xi_t \equiv \xi(t\Delta t) \), \( \mathcal{F}_t \equiv \mathcal{F}_{t\Delta t} \) and \( \pi_t(K) \equiv \pi(t, K) \). Moreover, we let \( \mathcal{T}_{[t,T]} \) denote the set of all stopping times taking values in \( \{t, t+1, \ldots, T\} \). The optimization problem from (7.28) translates into

\[
\max_{\tau \in \mathcal{T}_{t,T}} \pi_0(K(\tau)), \quad (7.29)
\]

We let \( (V_t)_{t=0}^T \) denote the value function of the optimization problem. At each time \( t \), this value function is either the value of the immediate payoff (termination value) or the present value of the remainder of the sequence, whichever is greater. As the problem has a finite time horizon, the value function can be constructed through backward induction. Formally:

\[
\begin{align*}
V_T & = K_T, \\
V_t & = \max(K_t, \pi_t(V_{t+1})), \quad 0 \leq t < T.
\end{align*}
\]

This also amounts to

\[
V_t = \pi_t(K_{\tau_t}) \quad \text{where} \quad \tau_t = \min\{s \geq t \mid V_s = K_s\} \quad (7.31)
\]
and, in particular, we have

\[ E[V_0] = \sup_{\tau \in T_{[0,T]}} E[K_\tau] = E[K_{\tau_0}], \quad (7.32) \]

where \( \tau_0 \) is as in (7.31) with \( t = 0 \). This means that we have managed to give an explicit expression for the optimal stopping time, so that we can now re-formulate the backward induction in terms of stopping times:

\[
\begin{aligned}
\tau_T &= T, \\
\tau_t &= t \cdot 1_{A_t} + \tau_{t+1} \cdot 1_{A_t^c}, \quad 0 \leq t < T,
\end{aligned}
\]

where

\[ A_t = \{ K_t \geq \pi_t (K_{\tau_{t+1}}) \}, \quad 0 \leq t < T. \]

The formulation of the backward induction in terms of optimal stopping times as given in (7.33) is one of the cornerstones of the Longstaff-Schwartz method. The other one is the assumption that the underlying processes are Markovian. Recall that all uncertainty emanates from the Markovian driver process \( (X_t)_{t \geq 0} \) as set forth in Section 9.2. The price kernel and the payoff both have representations in terms of measurable functions, say \( \xi_t = f(t, X_t) \) and \( K_t = g(t, X_t) \). Let \( s \geq t \), we then have

\[
\pi_t(K_s) = E[\xi_s K_s | F_t] / \xi_t
\]

\[
= E[f(s, X_s) g(s, X_s) | X_t] / f(t, X_t)
\]

\[
= h(s, t, X_t)
\]

for some measurable but not necessarily explicitly known function. The idea is now to approximate the conditional expectation by an orthogonal projection on a subspace of \( L^2 \) generated by a finite number of functions. To this end, we consider a sequence \( (e_i(x))_{i \geq 1} \) of real-valued measurable functions on the state space of the driver process \( (X_t)_{t \geq 0} \).

**Assumption 7.11** Following [34] we assume that the functions \( (e_i(x))_{i \geq 1} \) satisfy the following conditions.

1. For each \( t = 1, \ldots, T - 1 \), the sequence \( (e_i(X_t))_{i \geq 1} \) is total in \( L^2 \).

2. For each \( t = 1, \ldots, T - 1 \) and \( m \geq 1 \), if \( \sum_{i=1}^{m} \lambda_i e_i(X_t) = 0 \) a.s., then \( \lambda_i = 0 \) for \( i = 1, \ldots, m \).
The choice of the approximating functions can pose some problems, as the dimension of $X_t$ may be rather high in realistic settings. This issue will be dealt with below. The estimation procedure is operationalized in the following way: Suppose we have $N$ simulated trajectories $(X_t^{(1)})_{t=1}^T$, ..., $(X_t^{(n)})_{t=1}^T$, ..., $(X_t^{(N)})_{t=1}^T$. Let $K_t^n = g(t, X_t^n)$ denote the associated payoffs, and let $e_m(x) = (e_1(x), ..., e_1(x))' \in \mathbb{R}^m$. Then, for each trajectory $n$, we can recursively estimate the associated realization $\tau_t^n$ of the optimal stopping time:

$$n \in \{1, ..., N\} : \begin{cases} \tau_T^n = T, \\ \tau_t^n = t \cdot 1_{A_t} + \tau_{t+1}^n \cdot 1_{A_t}, \end{cases} \quad 1 \leq t < T, \quad (7.34)$$

where

$$A_t = \{K_t^n \geq \alpha_t' e^m(X_t^n)\}, \quad 1 \leq t < T,$$

and $\alpha_t$ is the least-squares estimate

$$\alpha_t = \arg\min_{a \in \mathbb{R}^m} \sum_{n=1}^N \left( K_{\tau_{t+1}}^n - a'e^m(X_t^n) \right)^2, \quad 1 \leq t < T.$$  

Once all realizations of the optimal stopping time are known, the valuation can be done in the obvious way:

$$V_0 = \max \left( K_0, 1 - \frac{1}{N} \sum_{n=1}^N \frac{\xi_t^n K_{\tau_t^n}}{\xi_0^n} \right). \quad (7.35)$$

The authors of [34] show that, if Assumption 7.11 holds and if $K_t \in L^2$, then the above algorithm converges almost surely for each choice of $m$ as $N$ tends to infinity.

We now return to the problem of selecting an appropriate sequence $(e_k(x))_{k \in \mathbb{N}}$ of approximating functions $e_k : \mathbb{R}^d \rightarrow \mathbb{R}$ for the regression in (7.34). This task is relatively simple in the one-dimensional case, as there exist many families of suitable functions like power series, Laguerre or Hermite polynomials, splines, or even wavelets. The situation becomes more difficult in a multi-dimensional setting. We present here a simple generic approach for coping up with this situation.
We start in the one-dimensional case, \( x \in \mathbb{R} \). We assume that we have managed to identify a series \( (\tilde{e}_k(x))_{k \in \mathbb{N}} \) of suitable functions and an appropriate degree \( \tilde{m} \). Specifically, we always assume that \( \tilde{e}_k(x) \) is a polynomial of degree \( k - 1 \), e.g. Laguerre or power series. We can now apply this one-dimensional approximation to each dimension separately, yielding the following information

\[
\begin{pmatrix}
\tilde{e}_1(x_1) & \tilde{e}_2(x_1) & \cdots & \tilde{e}_{\tilde{m}}(x_1) \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{e}_1(x_d) & \tilde{e}_1(x_d) & \cdots & \tilde{e}_{\tilde{m}}(x_d)
\end{pmatrix}
\]

The presentation in matrix form is only for convenience. Now, there are different possibilities to turn the data from (7.36) into a vector-valued function as required in (7.34). The most simple one is to turn \( E(x) \) into a vector:

\[
e^m(x) = \text{vec}(E(x)) \in \mathbb{R}^{d\tilde{m}}.
\]

This choice is simple, but does not comprise any product terms between the different components of \( x \), which may be problematic if the different components of the driver process \( (X_t)_{t \geq 0} \) are strongly correlated. To cope up with this problem, we can select combinations of length \( k \) of the elements of \( E(x) \) and let the product of each combination form one component of the vector \( e^m(x) \). For \( d = 2 \) and \( \tilde{m} = 2 \) this looks like

\[
e^m(x) = (e_1(x_1)e_2(x_1), e_1(x_1)e_1(x_2), e_1(x_1)e_2(x_2),
\]

\[
e_2(x_1)e_1(x_2), e_2(x_1)e_2(x_2), e_1(x_2)e_2(x_2))' \in \mathbb{R}^m.
\]

The problem with this approach is that the length of \( e^m(x) \) increases very quickly, indeed,

\[
m = \dim(e^m(x)) = \binom{\tilde{m}d}{k}, \quad k \leq \tilde{m}d.
\]

A first measure to control the dimension is to let always \( k = 2 \), i.e. to consider only products of two base functions. Recall also that we assume the functions \( e_k(x) \) to be polynomials of degree up to \( \tilde{m} - 1 \), it seems plausible to include only those products with a combined degree not greater than \( \tilde{m} - 1 \). This considerably reduces the dimensionality. Moreover, it should be noticed that the least-squares computation only has to be carried out once per time step, which will amount to a relatively low number
of executions, and the related Matlab function lsqlin can deal with large problems quite efficiently.

The method set forth above is fairly generic, but will lead to rather expensive computations. Here is a more simple alternative that is, however, not universally applicable. In many cases, the payout $K(t)$ will be defined as a function of one of the observables, $R(t, \tau)$ say, which is one-dimensional and a function of $X_t$ itself:

$$K(t) = \tilde{g}(t, R(t, \tau)) = \tilde{g}(t, r_\tau(t, X_t)) = g(t, X_t).$$

The idea is then to select again a series of univariate approximating functions $(\tilde{e}_k(x))_{k \in \mathbb{N}}$ and to concatenate them with the functional representation of the one observable that the payout depends on:

$$e_K(x) = \tilde{e}_k(r_\tau(t, x)), \quad k = 1, \ldots, m.$$ 

Intuitively, this approach can lead to a considerable reduction in the dimensionality of the regression problem in (7.34) without loss of relevant information. But it still has to be verified whether this intuition is actually correct and whether the convergence results of [34] still hold. Instead of using parametric regression as discussed above, one could also use non-parametric regression methods; see [95]. This works fine in the one-dimensional case. For the multi-dimensional case, a combination technique as set forth further above would have to be used.

### 7.3.3 Valuation of exogenous cash flows

In practical insurance situations, we often have random cash flows that are independent of or only partly dependent on the financial market as presented by our model. How – or to what extent – can such cash flows be evaluated in our price kernel framework? Recall from (7.26) that the value according to the general valuation principle from (7.23) of some cash flow that is independent of the filtration $(\mathcal{F}_t)_{t \geq 0}$ governing the price kernel is given by the discounted expected value, that is

$$\pi(t, K_T) = \mathbb{E}[K_T] P(t, T).$$
7.3. Valuation by Using the Price Kernel

This is based on the standard assumption in mathematical finance that the agent is perfectly capable of diversifying all non-systematic risk and, hence, does not deserve a reward for the latter. This is clearly unsatisfactory for an insurer that is precisely in the business of assuming and managing non-systematic risk. Therefore, extensions of the price kernel approach are needed. But, before proceeding further, let us generalise the problem by assuming that the cash flow $K_T$ is given by

$$K_T = h(X_T, Y_T), \quad (7.38)$$

where $X = (X_t)_{t \geq 0}$ is the familiar driver process underlying the price kernel, $Y = (Y_t)_{t \geq 0}$ is some other process independent of $X$, and $h(\cdot)$ is some sufficiently regular function. Then, we can still use the general valuation principle from (7.23) to obtain

$$\pi(t, K_T) = \frac{E[K_T \xi(T) | \mathcal{F}_t]}{\xi(t)},$$

and it is no problem to evaluate this expression by Monte Carlo methods. However, $\pi(t, K_T)$ still does not contain any allowance for the non-systematic risk emanating from $Y$. Let us assume now, as Bühlmann does [21, 10], that we can decompose $K_T$ as

$$K_T = h(X_T, Y_T) = \sum_{n=1}^{N} f_n(X_T)g_n(Y_T), \quad (7.39)$$

where we assume $N = 1$ in the sequel without any loss of generality. That is, we have a decomposition into fully systematic and fully non-systematic components. Such decompositions can be achieved in fairly general situations, see Section 12.3 or [10]. Now we can condition on $Y$ to obtain

$$\tilde{K}_T = \pi (t, K_T | Y_T)$$

$$= E [f(X_T)g(Y_T)\xi(T)| \mathcal{F}_t; Y_T] / \xi(t)$$

$$= g(Y_T)\pi(t, f(X_T)). \quad (7.40)$$

That is, we obtain a new random cash flow $\tilde{K}_T$ which is adjusted for systematic risk emanating from the financial market as modelled by $X$. Note that this corresponds roughly to Bühlmann's decomposition [21] into price of a unit (here: $\pi(t, f(X_T))$) and number of units (here: $g(Y_T)$), where
the number of units is still a random quantity at this stage. Now, we can continue with the valuation of the financially adjusted risk $\hat{K}_T$ by whatever method deemed suitable, irrespective of the financial risk component.

For a first attempt, let us assume that we manage to find another price kernel $\xi^Y = (\xi^Y(t))_{t \geq 0}$ for the risk factor process $Y$. Under some regularity conditions on $Y$ and $g(.)$, this is actually no (technical) problem; see [21]. Given the independence of $X$ and $Y$, $\xi^Y$ is also independent of $X$. Applying the general valuation principle from (7.23) to $\hat{K}_T$, we obtain the total value

$$\pi(t, K_T) = \pi^Y(t, g(Y_T)) \pi^X(t, f(X_T)), \quad (7.41)$$

which corresponds again to [21]. However, depending on the type of risk that $Y$ represents, one might feel uneasy about the use of this second price kernel.

Recall the theoretical considerations in Section 7.2, where we can see that the price kernel valuation is equivalent to the expected discounted payoff under the equivalent martingale measure. This valuation, however, is not justified by the mere fact that the discounted payoff can be turned into a martingale. The crucial point is rather that we can set up a self-financing portfolio that replicates the contingent payoff. Then, the law of one price dictates that the value of the payoff be the purchase price of this replicating portfolio. The martingale property then follows as a rather technical consequence. See [118] for a discussion of this issue. The construction of the replicating portfolio, however, requires the presence of a liquid and frictionless market underlying the contingent payoff under scrutiny. Notice that this requirement also results if we motivate the price kernel along the more traditional lines presented in Section 7.1.2: it must – at least in principle – be possible to replicate the Arrow-Debreu securities.

There are cases where the assumption of such an underlying market holds – at least in good approximation. One can think of certain credit exposures, weather risk or crop yield insurance. However, for other types of insurance business it is clearly not satisfied and, therefore, one might re-
main in doubt as to whether the deflator valuation provides a viable price.

If one is in doubt on the viability of the price kernel for the non-systematic risk contribution $Y$, one can still proceed along purely actuarial lines. That is, one takes the financially adjusted payoff $\tilde{K}_T$ and applies one of the many actuarial pricing principles to it. These pricing principles are based on the assumption that the risk is fully borne by the insurer and that possible shortfalls must be covered by risk capital which, in turn, must yield a sufficient return to the investors. For an overview of actuarial valuation techniques, see e.g. [41]; a simple example of a modern actuarial valuation technique is presented in Section 12.2. An interesting point for future research would be to explore possible relations between price kernels and actuarial valuation techniques like the Esscher or Wang transforms.

It is important to notice that such purely actuarial valuation is, in turn, only viable if all systematic (hedgeable) risk has been removed. So, in general situations, the question is not whether one needs actuarial or financial valuation techniques, but rather how to combine the two.
Seite Leer / Blank leaf
Chapter 8

The Potential Approach to Modelling

8.1 The Basic Setup

The conclusion of Section 7.2 is that the price kernel $\xi = (\xi(t))_{t \geq 0}$ must be a positive supermartingale with the additional property that $\mathbb{E}[\xi(t)] \to 0$ as $t \to \infty$. We are now left with the problem of finding suitable ways for constructing price kernels with these properties, and we follow the so-called potential approach originally proposed by [106] and further explored by [84] in the multi-economy case.

We take the probabilistic setup of Section 7.2, so that the no-arbitrage results shown there remain valid. Specifically, let $t \in \mathbb{R}_0^+$ denote time and let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space for all economies, where $\mathbb{P}$ is the objective probability measure. All the uncertainty in the model is reflected by a $d$-dimensional standard Brownian motion $\mathbf{B} = (B^1, \ldots, B^d)'$, where $B^i = (B^i_t)_{t \geq 0}$. We specify the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ as the natural filtration of $\mathbf{B}$ augmented by all $\mathbb{P}$-null sets.

Let $\mathbf{X} = (X_t)_{t \geq 0}$ be a continuous-time Markov process with $\mathbb{R}^d$ as its
state space. We will refer to this process as the driver process. More specifically, we assume that $X$ is a time-homogeneous Itô diffusion:

**Definition 8.1** A time-homogeneous Itô diffusion is a stochastic process $X_t(\omega): \mathbb{R}_0^+ \times \Omega \to \mathbb{R}^n$ satisfying a stochastic differential equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad t \geq 0, \quad X_0 = x,$$

where $B_t$ is an $m$-dimensional Brownian Motion and the functions $b: \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma: \mathbb{R}^n \to \mathbb{R}^{n \times m}$ satisfy the condition

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|, \quad x, y \in \mathbb{R}^n,$$

where $|\sigma|^2 = \sum_{i,j} |\sigma_{ij}|^2$.

A detailed discussion of Itô diffusions is given in Chapter 7 of [104]; here are some properties relevant for the moment:

**Proposition 8.2** Let $X$ be an Itô diffusion as specified in Definition 8.1. Then:

1. $X$ exists; moreover, it is essentially unique (modulo $P$-null sets) and has $P$-almost surely continuous sample paths.

2. $X$ is adapted to the filtration $\mathcal{F}_t^{X_0}$ generated by $X_0$ and $\{B_s: 0 \leq s \leq t\}$, and it is also square-integrable: $E \left[ \int_0^T |X_t|^2 \, dt \right] < \infty$.

3. $X$ has the Markov property and also the strong Markov property.

**Proof.** See [104]: Theorem 5.2.1 for (1) and (2); Theorem 7.1.2 and Theorem 7.2.4 for (3). \qed

**Remark 8.3** From now on, we assume the filtration $\mathbb{F}$ to be as described in (2) of Proposition 8.2.

In order to turn the driver process $X$ into a price kernel $\xi$, we use the resolvent of a continuous-time Markov process. Whereas the resolvent is defined for a fairly general class of Markov processes, some of the properties that we will rely upon do only hold if the Markov process $X$ is more specifically an Itô diffusion.
8.1. The Basic Setup

Definition 8.4 Let $\alpha > 0$, let $(X_t)_{t \geq 0}$ be a Markov process taking values in $\mathbb{R}^d$, and let $g \in C_b(\mathbb{R}^d)$. Then the resolvent operator $R_\alpha$ is defined by

$$(R_\alpha g)(x) = \mathbb{E}^x \left[ \int_0^\infty e^{-\alpha s} g(X_s) \, ds \right].$$

Lemma 8.5 $R_\alpha : C_b(\mathbb{R}^d) \to C_b(\mathbb{R}^d)$

Proof. See Lemma 8.1.3 of [104].

From now on we will more specifically assume that $g \in C_b(\mathbb{R}^d, [0, \infty))$, and we exclude the degenerate case of $g(x) \equiv 0$. It is immediately clear from Definition 8.4 that $R_\alpha g$ is then also positive. Now we are ready to define the price kernel:

Definition 8.6 Let $X$ be a time-homogeneous Itô diffusion as specified in Definition 8.1, let $g \in C_b(\mathbb{R}^d, [0, \infty)) \setminus \{0\}$, and let $\alpha > 0$. The price kernel $\xi = (\xi(t))_{t \geq 0}$ is defined by

$$\xi(t) = e^{-\alpha t} \frac{(R_\alpha g)(X_t)}{(R_\alpha g)(X_0)}, \quad t \geq 0.$$ 

Note that $(R_\alpha g)(X_0)$ is just a norming constant that assures $\xi(0) = 1$; we may omit it for simplicity’s sake from time to time without loss of generality. We now have to show that $\xi$ actually qualifies as a price kernel in the sense of Section 7.2:

Proposition 8.7 Let $\xi$ be as specified in Definition 8.6. Then $\xi$ is a positive $(P,F)$-supermartingale and $\mathbb{E} [\xi(t)] \to 0$ as $t \to \infty$.

Proof. Positivity and convergence of the expectation are obvious given Definition 8.4 and Definition 8.6. Let $h > 0$:

$$\mathbb{E}^x [\xi(t+h)|\mathcal{F}_t] = \mathbb{E}^x \left[ e^{-\alpha (t+h)} \mathbb{E}^{X_{t+h}} \left[ \int_0^\infty e^{-\alpha s} g(X_s) \, ds \right] | \mathcal{F}_t \right]$$

$$= e^{-\alpha h} e^{-\alpha t} \mathbb{E}^{X_t} \left[ \int_0^\infty e^{-\alpha s} g(X_s) \, ds \right]$$

$$\leq e^{-\alpha t} \mathbb{E}^{X_t} \left[ \int_0^\infty e^{-\alpha s} g(X_s) \, ds \right] = \xi(t).$$
Chapter 8. The Potential Approach to Modelling

The $L^1$-property follows from Lemma 8.5 and measurability is assured by the properties of conditional expectation.

In a multi-economy setup, we still have only one driver process $X$ – in a sufficiently high-dimensional space $\mathbb{R}^d$ – but we select one specific function $g^{(k)}(.)$ for each economy $(k)$, and we obtain $\xi^{(k)}(.)$ by plugging each $g^{(k)}(.)$ into Definition 8.6.

**Proposition 8.8** Let $\xi$ be the price kernel as specified in Definition 8.6. Then, for all $t \geq 0$, the short rate $r(t)$ is given by

$$r(t) = \frac{g(X_t)}{(R_\alpha g)(X_t)}.$$

**Proof.** This result is stated in [106] without detailed proof. The latter is added here for completeness' sake. Due to Lemma 8.5 and Proposition 8.7, $\xi$ admits a Doob-Meyer decomposition $\xi(t) = M_t - A_t$, where $M_t$ is a martingale and $A_t$ is an increasing process. Assume first that $M_t$ has the following explicit form:

$$M_t = \mathbb{E}^x \left[ \int_0^\infty e^{-\alpha s} g(X_s) \, ds \bigg| \mathcal{F}_t \right].$$

This is clearly a martingale; moreover, due to the Markov property

$$M_t = \mathbb{E}^x \left[ \int_0^t e^{-\alpha s} g(X_s) \, ds + \int_t^\infty e^{-\alpha s} g(X_s) \, ds \bigg| \mathcal{F}_t \right]$$

$$= \int_0^t e^{-\alpha s} g(X_s) \, ds + \int_0^\infty e^{-\alpha s} g(X_s) \, ds$$

$$= \int_0^t e^{-\alpha s} g(X_s) \, ds + (R_\alpha g)(X_t).$$

Hence, because of Definition 8.6, we have $\xi(t) = M_t - A_t$ for $A_t = \int_0^t e^{-\alpha s} g(X_s) \, ds$ as conjectured. Since the Doob-Meyer decomposition is unique (modulo inequality on $\mathbb{P}$-null sets), we have indeed the explicit form of $M_t$. Inserting in (7.12) and comparing coefficients completes the proof.  \(\square\)
8.2 Construction Method: Inverse Resolvent

The problem with the construction of $\xi$ through Definition 8.6 is that it is hard to find explicit representations for the resolvent as given by Definition 8.4. The key ingredient for a solution to this problem is the infinitesimal generator of the Itô diffusion $X$:

**Definition 8.9** Let $X$ be a time-homogeneous Itô diffusion in $\mathbb{R}^d$ as given by Definition 8.1. The infinitesimal generator $A$ of $X$ is defined as

$$(Af)(x) = \lim_{t \searrow 0} \frac{E^x[f(X_t)] - f(x)}{t} \quad \text{for} \quad x \in \mathbb{R}^d.$$ 

The set of functions $f : \mathbb{R}^d \to \mathbb{R}$ for which the limit exists for a specific $x$ is denoted by $D_A(x)$; the set $D_A$ contains the functions for which the limit exists for all $x \in \mathbb{R}^d$.

It turns out that the infinitesimal generator of an Itô diffusion has an explicit representation:

**Theorem 8.10** Let $X$ be a time-homogeneous Itô diffusion according to Definition 8.1, and let $f \in C_0^2(\mathbb{R}^d)$. Then $f \in D_A$ and

$$(Af)(x) = \sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x) + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (\sigma \sigma')_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

**Proof.** See [104], Theorem 7.3.3.

**Theorem 8.11** Let $\alpha > 0$. Then $R_\alpha$ and $\alpha - A$ are inverse operators:

1. If $f \in C_0^2(\mathbb{R}^d)$ then $R_\alpha(\alpha - A)f = f$.
2. If $g \in C_0(\mathbb{R}^d)$ then $R_\alpha g \in D_A$ and $(\alpha - A)R_\alpha g = g$.

**Proof.** See [104], Theorem 8.1.5.

Following the path of [106], we can now circumvent the problems associated with the resolvent.
1. We select some function \( f : \mathbb{R}^d \to (0, \infty) \),

2. then we define \( R_\alpha g := f \),

3. so that we have \( g = (\alpha - A)f \).

Some care must be taken when selecting the function \( f \). On top of being positive, we should also have \( f \in C^2_c(\mathbb{R}^d) \), that is, \( f \) is twice continuously differentiable with compact support, for the infinitesimal generator to be well defined. In principle, the requirement for a compact support can be dropped, but it must be verified on a case-by-case basis whether the implied \( g = (\alpha - A) \) satisfies the requirements set forth in Section 8.1, i.e. positivity, boundedness and continuity. If we actually manage to find a suitable function \( f \), then we can set

\[
\xi(t) = e^{-\alpha t} \frac{(R_\alpha g)(X_t)}{(R_\alpha g)(X_0)} = e^{-\alpha t} \frac{f(X_t)}{f(X_0)},
\]

which also looks fairly tractable from a numerical point of view. Some choices for \( f \) will be explored further below. An alternative construction method based on eigenfunctions of the infinitesimal generator of the driver process is sketched, but not further developed, in [106]. Some preliminary studies showed that, appealing though this approach looks, it does not lead to tractable models. We will therefore stick to the inverse resolvent method in the sequel.

### 8.3 Alternative: The Flesaker and Hughston Approach

Recall from Section 7.2 that price kernel models for interest rates are based on the fundamental relation

\[
P(t, T) = \frac{\mathbb{E}[\xi(T)|\mathcal{F}]}{\xi(t)},
\]

where \( \xi = (\xi(t))_{t \geq 0} \) that is adapted to the \( d \)-dimensional Itô diffusion \( X \) and for which \( \mathbb{E}[\xi(t)] \to 0 \) as \( t \to \infty \). The question is then how to construct such processes \( \xi, \) and we have opted for the potential approach as
In this section, we briefly explore an alternative to the potential approach, which could be termed martingale approach or Flesaker-Hughston approach (after [60]) or in the context of DFA - Cairns approach (after [24, 27]). Suppose that the price kernel \( \xi \) can be represented as

\[
\xi(t) = \int_{t}^{\infty} M(t, s) \, ds, \quad t \geq 0, \tag{8.2}
\]

where \( M(0,T) \equiv 1 \) and \( (M(t,T))_{t \geq 0} \) is a positive martingale in \( t \) for all \( T \geq 0 \). Then, we have

\[
P(t,T) = \frac{\int_{T}^{\infty} M(t, s) \, ds}{\int_{t}^{\infty} M(t, s) \, ds}, \quad 0 \leq t \leq T < \infty. \tag{8.3}
\]

An interesting special case arises if we have \( M(t, s) = M(t, s)\phi(s) \), where \( M(t, s) \) has the same properties as \( M(t, s) \) above, and \( \phi(s) \) is a positive and decreasing deterministic function. Plugging this into Equation 8.4 with \( t = 0 \) and recalling that \( M(0,T) \equiv 1 \) for all \( T \geq 0 \), we obtain

\[
\phi(s) = -c_\phi \frac{\partial}{\partial T} P(0,T), \quad T \geq 0, \tag{8.4}
\]

for some constant \( c_\phi > 0 \). That is, we can incorporate the initial term structure in a very natural way if desired. On the other hand, if we let \( M(t, T) = H(T - t, X_t) \) for some time-homogeneous Itô diffusion \( X \) and some suitable function \( H(t, X) \), then we obtain time-homogeneous models with ZCB prices

\[
P(t,T) = \frac{\int_{T-t}^{\infty} H(u, X_t) \, ds}{\int_{0}^{\infty} H(u, X_t) \, ds}, \quad 0 \leq t \leq T < \infty. \tag{8.5}
\]

In the references stated above, it is shown that this class of models has the desired properties of absence of arbitrage, positive rates, flexible curve shapes and clustering of high and low interest rate levels.

The first specific instance of this model class is the Rational Log-Normal Model; see p. 227 of [74]. This model is very simple and tractable, but
not suitable for use in the context relevant here as it only admits one stochastic factor.

A multi-factor model for the use in DFA-type settings was developed and tested by Cairns; see [24] and some preceding working papers. This model features the time-homogeneous setting of (8.5) with

$$H(u, x) = \exp \left\{ -\beta u + \sum_{i=1}^{n} \sigma_i x_i e^{-\alpha_i u} - \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} \sigma_i \sigma_j e^{-(\alpha_i + \alpha_j) u} \right\}$$

and $X = (X_t)_{t \geq 0}$ being an $n$-dimensional Ornstein-Uhlenbeck process

$$dX_t = -AX_t dt + CD B_t, \quad X_0 = x,$$

with diagonal mean reversion matrix $A = \text{diag}(\alpha_1, \ldots, \alpha_n)$ and general symmetric correlation matrix $CC' = (\rho_{ij})_{i,j=1}^{n}$, which is the exact opposite of the setup chosen in Chapter 9. In its most general form, this model has $2n + 1 + n(n + 1)/2$ parameters as opposed to $n^2 + 2n$ parameters for the most general form of the exponential-quadratic model from Chapter 9. This parsimony is an advantage in view of the low amounts of data available for calibration in typical DFA situations. The integrals in (8.5) have to be evaluated numerically at each time step and for each random replicate in a Monte Carlo simulation as they depend on the value of the underlying process $X$. Notice, however, that the integrals are along one dimension only, irrespective of the dimension of $X$, so that no particular numerical problems arise. Evaluations carried out in [24] suggest that the model is well capable of reproducing the salient empirical features of interest rates.

A second model class, presented in [27], is aimed at including stochastic volatility while the ZCB prices have the integral affine form

$$P(t, T) = \frac{\int_{T-t}^{\infty} \exp \{A(u) + B(u)'X_t\} du}{\int_{0}^{\infty} \exp \{A(u) + B(u)'X_t\} du}.$$ 

It is then shown that, if the model has to have the required properties, the underlying process $X$ can have the same generic form as for general affine models (see [74], Chapter 7). The deterministic functions $A(u)$ and
8.4 A Time-inhomogeneous Extension

\[ B(u) \] can be obtained as the solution of a system of differential equations.

Although the models in [24] and [27] are only formulated in single-economy settings, they can be generalised to multi-economy settings in exactly the same way as models based on the potential approach. Since the price kernel can be computed explicitly (see (8.4), the models presented here also fit into the generic valuation approach presented in Section 7.3.

8.4 A Time-inhomogeneous Extension

The construction approach for the price kernel as introduced in Section 8.1 comes with the disadvantage that the resulting yield curves cannot be fitted exactly to some given initial term structure \( \{ P(0, t) : t \geq 0 \} \). Provided that the model order is sufficiently high, one can achieve fairly good fits; see [13]. However, the order of the model cannot be increased infinitely just to fit the initial curve, because this may result in too high numbers of parameters for the model to be calibrated meaningfully. Moreover, it may be desirable to have an exact fit to the initial term structure so as to fully incorporate the information contained therein.

To cope with this situation, we propose here a simple time-inhomogeneous extension to the potential model which is partly inspired by [36] and [106]. The distinctive feature of our method is that the function introducing the time inhomogeneity is superimposed on the price kernel rather than on the driver process as in most classical models like Hull-White. Hence, there is only one scalar function per yield curve, which can be determined unambiguously, even if the driver process has a high dimension.

Let us assume that we have already fitted a \textit{time-homogeneous model} to some set of ZCB price data; i.e., we already know the law of the time-homogeneous price kernel \( \xi \), and hence the modelled initial ZCB prices \( \{ P(0, T) : T \geq 0 \} \). These may, however, differ from the actually observed initial ZCB prices \( \{ \hat{P}(0, T) : T \geq 0 \} \). In order to overcome this discrepancy, we define a new, \textit{time-inhomogeneous model} through a
modified price kernel \( \tilde{\xi} = (\tilde{\xi}(t))_{t \geq 0} \) defined as
\[
\tilde{\xi}(t) = h(t) \xi(t), \quad t \geq 0. \tag{8.6}
\]
Recalling the general definition of the price kernel, and in particular (7.9), we obtain the following relation for the ZCB prices under the time-inhomogeneous model:
\[
\tilde{P}(0, T) = \mathbb{E}^x \left[ \tilde{\xi}(T) \right] = \mathbb{E}^x [h(T)\xi(T)] = h(T)P(0, T),
\]
so that the deterministic correction function \( h(T) \) is uniquely specified by
\[
h(T) = \frac{\tilde{P}(0, T)}{P(0, T)}, \quad T \geq 0. \tag{8.7}
\]
So far, so good. But does the so-specified time-inhomogeneous price kernel \( \tilde{\xi} \) still satisfy the general requirements of a price kernel set forth in Section 7.2, i.e. is \( \tilde{\xi} \) still a positive supermartingale with expectation tending to zero as time tends to infinity?

If the time-homogeneous model satisfies the requirements for the price kernel, then Proposition 7.6 assures that \( 0 < P(0, T) \leq 1 \) for all \( T \geq 0 \). If the real-world ZCB prices also satisfy the elementary no-arbitrage condition \( 0 < \tilde{P}(0, T) \leq 1 \) then it follows from (8.7) that \( h(t) \) is positive for all \( t \geq 0 \) and hence, by (8.6), \( \tilde{\xi} \) is also positive. Along the same line of argument, we can postulate that both \( \tilde{P}(0, T) \) and \( P(0, T) \) tend to zero as \( T \) tends to infinity. This is, of course, not sufficient for assuring that \( h(T) \) tends to 1. However, real-world ZCB prices are only observable up to some finite maximum time \( T_{\text{max}} \), beyond which \( \tilde{P}(0, T) \) must be obtained by extrapolation. And we can then select the extrapolation method in such a way that \( \lim_{T \to \infty} h(T) = 1 \), and hence \( \lim_{T \to \infty} \mathbb{E}^x[\tilde{\xi}(T)] = 0 \).

What about the supermartingale property? First, we note that \( h(T) \) is a deterministic function, hence measurability of \( \tilde{\xi} \) is no problem. Moreover, it is realistic to assume that both \( \tilde{P}(0, T) \) and \( P(0, T) \) are sufficiently well-behaved for \( h(T) \) to be strictly bounded, so that \( \tilde{\xi}(t) \) is also \( L^1 \) for all \( t \). The crucial topic is to specify conditions under which we have
\[
\mathbb{E}^x \left[ \tilde{\xi}(T) \big| \mathcal{F}_t \right] \leq \tilde{\xi}(t), \quad 0 \leq t \leq T < \infty.
\]
8.4. A Time-inhomogeneous Extension

By (8.6) and (7.9) this is equivalent to

\[
\frac{h(T)}{h(t)} \leq \frac{\xi(t)}{\mathbb{E}^x[\xi(T)|\mathcal{F}_t]} = \frac{1}{P(t, T)}, \quad 0 \leq t \leq T < \infty. \tag{8.8}
\]

A sufficient condition for this inequality to be satisfied is easily found: since \(P(t, T)\) is always in \((0, 1]\), the upper bound is never below one. Hence, the inequality is satisfied if \(h(t)\) is non-increasing in \(t\).

For the general case, we will see in Section 9.3 that the ZCB prices \(P(t, T)\) can be expressed as \(P(t, T) = F(X_t, T - t)\) where \(X_t\) is the time-\(t\) value of the Markovian driver process \(X\), and \(F(x, \tau)\) is function explicitly known in all cases of interest. Then, Inequality 8.8 amounts to

\[
\frac{h(T)}{h(t)} \leq \frac{1}{F(x, T - t)}, \quad x \in \mathbb{R}^d, \quad 0 \leq t \leq T < \infty. \tag{8.9}
\]

This evaluation may be tedious, but it is feasible.

The approach presented in this section, though perhaps not very elegant, yields a feasible time-inhomogeneous extension of the time-homogeneous generic model introduced in Section 8.1. This approach allows to fit a model to any given initial term structure. An alternative would be to introduce time inhomogeneity into the driver process \(X\) for the price kernel. This approach may be less desirable because it may be very hard to calibrate. A final statement on the usability can only be made when specific models are applied to real-world data; this will be done in Section 10.8.
Chapter 9

The Exponential-Quadratic Model Class

9.1 Introduction

We briefly summarise the current situation. In Chapter 7, we have introduced the price kernel $\xi = (\xi(t))_{t \geq 0}$ as a means for the valuation of future contingent cash flows under the real-world measure $P$. We have shown that the value of a cash flow $K(T)$ that satisfies some measurability and integrability conditions is given by

$$V_T(K(T)) = \mathbb{E}_{\xi} \left[ \xi(T) K(T) \mid \mathcal{F}_t \right].$$

This particularly includes the case of zero-coupon bonds, where $K(T) \equiv 1$, and hence

$$P(t, T) = \pi(t, K(T)) = \frac{\mathbb{E}_{\xi} \left[ \xi(T) \mid \mathcal{F}_t \right]}{\xi(t)}.$$

which opens the way for the construction of interest rate models. It was also shown that, if this setup is to be arbitrage-free, the price kernel $\xi$ must be a positive supermartingale adapted to a $d$-dimensional Brownian Motion $B = (B_t)_{t \geq 0}$ and having the property $\mathbb{E}[\xi(t)] \to 0$ as $t \to \infty$. 

105
In Chapter 8, we have then introduced a generic method for constructing such price kernels: If $X = (X_t)_{t \geq 0}$ is a $d$-dimensional, time-homogeneous Itô diffusion with resolvent operator $R_\alpha$, and if $g$ is a sufficiently regular function mapping $\mathbb{R}^d$ into $[0, \infty]$, then the process

$$\xi(t) = e^{-\alpha t} \frac{(R_\alpha g)(X_t)}{(R_\alpha g)(X_0)}$$

has the desired properties for $\alpha > 0$. As the resolvent $(R_\alpha g)(x)$ is often not easy to compute, one can first postulate a function $f(x) = (R_\alpha g)(x)$. The implied function $g$ can then be recovered through $g = (\alpha - \mathcal{A}) f$, where $\mathcal{A}$ is the infinitesimal generator of $X$.

Hence, we are left with the task of making suitable choices for the process $X$ and for the function $g$ or $f$, respectively. In [106], Rogers explores a number of possible choices, some of which are further investigated in [84]. For completeness’ sake, and to motivate the choices relevant in the rest of this thesis, we give a brief overview here.

For the driver process $X$, the first choice is a multivariate Ornstein-Uhlenbeck process of the form

$$dX_t = \Gamma(\theta - X_t)dt + \Sigma dB_t. \quad (9.1)$$

This class of processes can reproduce the essential feature of mean reversion (see Chapter 6), and is analytically well tractable. A possible alternative is the multivariate version of the Cox-Ingersoll-Ross (CIR) process, that is

$$dX_t = \Gamma(\theta - X_t)dt + \Sigma \text{diag}(X_t)^{\frac{1}{2}} dB_t. \quad (9.2)$$

This process also has mean reversion. Moreover, it has the advantage that it only takes positive values and that it has level-dependent volatility. The downside is that the analytical properties are hard to grasp in the case of a general mean reversion matrix $\Gamma$.

Geometric Brownian Motion would be another possibility, but it is less suitable in the context of interest rate models as it is lacking the essential
feature of mean reversion. The Markovian setup of Chapter 8 would also allow for more general processes like jump diffusions. However, the arbitrage theory developed in Section 7.2 is valid for Brownian Motion only and is not applicable to any deviation from the latter.

As for the choice of the function \( f \), a wide range of possibilities exists. In an earlier phase of my research, the following choices, all suggested by [106] were investigated:

- **Exponential-quadratic**: \( f(x) = \exp \left\{ \frac{1}{2}(x - c)'Q(x - c) \right\} \), (9.3)
- **Exponential-linear**: \( f(x) = \exp\{\gamma'x\} \), (9.4)
- **Quadratic**: \( f(x) = \gamma + \frac{1}{2}(x - c)'Q(x - c) \), (9.5)
- **Linear**: \( f(x) = \gamma'x \). (9.6)

First of all, the pure quadratic and linear models were soon found to be inferior to their exponential counterparts in that they required more complicated constraints while not offering advantages in terms of lower numbers of parameters or simpler structures and, hence, dropped out of the competition.

The exponential-linear function gives rise to the well-known class of *affine models* (see [84]), where ZCB prices have the form

\[
P(t, \tau) = \exp \{a(\tau) + b(\tau)'X_t\}
\]

for some functions \( a(\tau) \) and \( b(\tau) \). Hence, negative interest rates can only be avoided if the driver process \( X \) is of the less tractable CIR type according to (9.2). Moreover, except for some very specific cases, the functions \( a(\tau) \) and \( b(\tau) \) are not explicitly known and have to be obtained through the numerical solution of partial differential equations; see [74]. In a typical DFA context, where tens of thousands of single rates have to be computed, this appears less desirable. Moreover, explicit forms for ZCB prices and yields are also an advantage if the model is to be calibrated by using methods like GMM or MLE.
The exponential-quadratic function from (9.3) was found to yield models most suitable for our purposes. With some constraints on the parameters, it is possible to produce positive interest rates even when using the more tractable Ornstein-Uhlenbeck process from (9.1) for the driver $X$. In this case, we also have explicit formulae for the ZCB prices and yields, and we can work out moment conditions for the calibration with GMM.

Hence, the ultimate decision was to select the exponential quadratic function according to (9.3) in conjunction with the multivariate Ornstein-Uhlenbeck process according to (9.1). The rest of this chapter is organized as follows: Section 9.2 investigates the properties of the multivariate Ornstein-Uhlenbeck process, Section 9.3 presents the detailed construction of the model, Section 9.4 explores the analytical properties of interest rates and FX rates under the model, and Section 9.5 puts the model into a wider context.

9.2 The Driver Process

In Section 8.1 we only required the driver process $X$ to be a time-homogeneous Itô diffusion according to Definition 8.1. Now, following [106] and [84], we propose the following specific and more tractable structures:

**Definition 9.1** Let the $\mathbb{R}^d$-valued process $X = (X_t)_{t \geq 0}$ be defined as the solution of the SDE

$$dX_t = -\Gamma X_t dt + dB_t, \quad X_0 = x.$$  

The matrix $\Gamma \in \mathbb{R}^{d \times d}$ governs mean reversion, and, and we assume the real parts of all its eigenvalues to be positive. $B = (B_t)_{t \geq 0}$ is a $d$-dimensional standard Brownian Motion.

This definition follows [106] rather than [84] in that the implied long-run mean is zero. As will become clear in the applications of the process, this comes with no loss of generality. Moreover, this setup with $X_t \in \mathbb{R}^d$, $B_t \in \mathbb{R}^d$ and full-rank volatility matrix also assures market completeness,
which is necessary for cross-country no-arbitrage; see [48], Section 6.H.

The process $X$ as specified here is a time-homogeneous Itô diffusion in the sense of Definition 8.1, so that the existence and uniqueness results of Proposition 8.2 are also valid. Then, according to Theorem 8.10, the infinitesimal generator $(Af)(x)$ is given by

\[
(Af)(x) = -(\Gamma x)'(\nabla f)(x) + \frac{1}{2} \text{tr} [(Hf)(x)] = \frac{1}{2}(\Delta f)(x), \quad (9.8)
\]

where $(\nabla f)$ and $(Hf)$ are the gradient vector and the Hessian matrix, respectively, of $f \in C^2_0(\mathbb{R}^d)$, and $(\Delta f)$ is the Laplacian of $f$, supposing Cartesian coordinates. In the sequel, we explore some useful properties of the Ornstein-Uhlenbeck process:

**Proposition 9.2** Let $0 \leq t \leq T < \infty$. In the process as given by Definition 9.1, $X_t$ given $X_0 = x$ is normally distributed with conditional mean

\[
\mu(x,t) := \mathbb{E}^x [X_t] = e^{-\Gamma t} x
\]

and conditional covariance matrix

\[
V(x,t) := \mathbb{E}^x [(X_t - \mu(x,t))(X_t - \mu(x,t))']
= \int_0^t e^{-\Gamma s} (e^{-\Gamma s})' \, ds
= : V(t).
\]

Due to the condition on the eigenvalues of $\Gamma$ in Definition 9.1, the limit $V(\infty) := \lim_{\tau \to \infty} V(\tau)$ exists, and the process $X$ has a stationary distribution which is Gaussian with zero mean and covariance matrix $V(\infty)$.

**PROOF.** We condition on $X_0 = x$. Following Chapter 5.6 of [78], $X$ then admits the representation

\[
X_t = e^{-\Gamma t} \left[ x + \int_0^t e^{-\Gamma s} dB_s \right].
\]

It follows from this representation that $X_t$ is normally distributed. Continuing along the lines of Chapter 5.6 of [78] we immediately obtain

\[
\mu(x,t) = \mathbb{E}^x [X_t] = e^{-\Gamma t} x.
\]
Similarly, using the above representation and the Itô isometry, we obtain for the variance

\[ V(x, t) = \mathbb{E}^x [(X_t - \mu(x, t))(X_t - \mu(x, t))'] \]

\[ = e^{-\Gamma t} \int_0^t e^{\Gamma s} (e^{\Gamma s})' ds (e^{-\Gamma t})' \]

\[ = \int_0^t (e^{-\Gamma t} e^{\Gamma s}) (e^{-\Gamma t} e^{\Gamma s})' ds \]

\[ = \int_0^t e^{-\Gamma(t-s)} (e^{-\Gamma(t-s)})' ds \]

\[ = \int_0^t e^{-\Gamma s} (e^{-\Gamma s})' ds. \]

Existence of the stationary distribution is by Theorem 6.7 of [78].

Under some regularity conditions, we can even obtain a closed-form expression for \( V(t) \):

**Proposition 9.3** Assume that the matrix \( \Gamma \) is symmetric, non-singular and the real parts of all its eigenvalues are positive. Then we have

\[ V(t) = \frac{1}{2} \Gamma^{-1} \left( 1 - e^{-2\Gamma t} \right), \]

and the covariance matrix of the stationary distribution of \( X \) is given by

\[ V(\infty) = \frac{1}{2} \Gamma^{-1}. \]

**Proof.** Recall that the matrix exponential is defined as \( e^{At} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k \).

Therefore, we can write \( \int_0^t e^{-\Gamma s} (e^{-\Gamma s})' ds = \int_0^t e^{-(\Gamma+s)'s} ds = \int_0^t e^{As} ds \) for the non-singular matrix \( A = -2\Gamma \). This leads to

\[ \int_0^t e^{As} ds = \int_0^t \sum_{k=0}^{\infty} \frac{s^k}{k!} A^k ds = \sum_{k=0}^{\infty} \frac{t^{k+1}}{(k+1)!} A^k = A^{-1} \left( \sum_{k=1}^{\infty} \frac{t^k}{k!} A^k \right) \]

\[ = A^{-1} (e^{At} - 1). \]

Given the existence of the stationary distribution in the general case (shown in Proposition 9.2), we have that \( e^{-\Gamma t} \rightarrow 0 \) as \( t \rightarrow \infty \), and hence \( V(\infty) = \frac{1}{2} \Gamma^{-1} \).
Let $X$ be as specified in Definition 9.1, and let $f \in C^{1,2}(\mathbb{R}_0^+ \times \mathbb{R}^d, \mathbb{R})$. Then the general version of Itô's Lemma as given in Theorem 3.3.6 of [78] simplifies to

$$
df(t, X_t) = f_t(t, X_t)\, dt - (\nabla_x f)(t, X_t) \cdot (\Gamma X_t)\, dt + \frac{1}{2} \text{tr} [(H_x f)(t, X_t)]\, dt + (\nabla_x f)(t, X_t) \cdot dB_t,
$$

where $f_t = \frac{\partial f}{\partial t}$, $\nabla_x f = \left(\frac{\partial f}{\partial x_i}\right)_{i=1}^d$, and $H_x f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j=1}^d$.

9.3  The Model

The model specified below, which is the richest one among those summarized in Section 9.1, is suggested by both [106] and [84]. The construction draws on the inverse resolvent approach described in Section 8.2. We let $f(x) := (R_x g)(x)$ have the exponential-quadratic form

$$
f(x) = \exp \left\{ \frac{1}{2} (x - c)'Q(x - c) \right\}, \tag{9.10}
$$

where $c \in \mathbb{R}^d$ is a constant vector and $Q \in \mathbb{R}^{d \times d}$ a symmetric and positive definite matrix. The gradient $(\nabla f)(x)$ and the Hessian $(H f)(x)$ are then given by

$$
(\nabla f)(x) = f(x)\, Q(x - c),

(\nabla^2 f)(x) = f(x)\, [(Q(x - c))(Q(x - c))' + Q]. \tag{9.11}
$$

We let the driver $X$ be a multivariate Ornstein-Uhlenbeck process as given in Definition 9.1:

$$
dX_t = -\Gamma X_t\, dt + dB_t, \quad X_0 = x.
$$

According to (8.1), the price kernel $\xi(t)$ then has the form

$$
\xi(t) = \exp \left\{ -\alpha t + \frac{1}{2} (X_t - c)'Q(X_t - c) \right\}. \tag{9.12}
$$

We can recover the implied function $g(x)$ by using Theorem 8.11 and the specific form of the infinitesimal generator given in (9.8):

$$
g(x) = f(x) \left( \alpha + x'\Gamma'Q(x - c) - \frac{1}{2} \text{tr} ((Q(x - c))(Q(x - c))' + Q) \right).
$$
We make the following definitions suggested by [106]:

\[ S := \Gamma'Q + Q\Gamma - Q^2, \quad v := (\Gamma' - Q)Qc. \]  \hfill (9.13)

**Lemma 9.4** If \( S \) and \( v \) are defined as in (9.13), we then have that

\[
g(x) = f(x) \left( \alpha + \frac{1}{2} (x - S^{-1}v)'S(x - S^{-1}v) 
- \frac{1}{2} v'S^{-1}v - \frac{1}{2} tr[Q] - \frac{1}{2} |Qc|^2 \right).\]

**Proof.** Some linear algebra including the fact that \( tr[xx'] = x'x \) for any \( x \in \mathbb{R}^d \). \( \square \)

If we manage to calibrate \( \Gamma \) and \( Q \) in such a way that \( S \) is positive definite, and if we let moreover

\[
\alpha = \frac{1}{2} \left( v'S^{-1}v + tr[Q] + |Qc|^2 \right), \]  \hfill (9.14)

then we obtain a version of \( g(x) \) that is non-negative for all \( x \in \mathbb{R}^d \), namely

\[
g(x) = f(x) \frac{1}{2} (x - S^{-1}v)'S(x - S^{-1}v).\]

We also observe that, by making the assumptions in Equations 9.13 and 9.14, we automatically satisfy the requirement that \( \alpha > 0 \): If \( S \) is positive definite, then so is \( S^{-1} \). Moreover, the trace of a symmetric matrix equals the sum of its eigenvalues, and the eigenvalues of a symmetric positive definite matrix are all positive.

By applying Itô's Lemma (9.9) to \( \xi(t) = e^{-at}f(X_t) \) and then using (7.18) we obtain the representation

\[
\frac{d\xi(t)}{\xi(t)} = -r(t)dt - \rho(t)'dB_t,
\]
where $r(t)$ is the short rate and $\rho(t)$ is the market price of risk, specifically:

\[
\begin{align*}
    r(t) &= \alpha + (Q(X_t - c))' (\Gamma X_t) \\
    &\quad - \frac{1}{2} \text{tr} \left( (Q(x - c))(Q(x - c))' + Q \right), \quad (9.15) \\
    \rho(t) &= -Q(X_t - c). \quad (9.16)
\end{align*}
\]

As explained by Proposition 8.8, the short rate $r(t)$ equals $g(X_t)/f(X_t)$ with $g(.)$ as given in (9.3) above. Re-applying the assumptions from Equations 9.13 and 9.14 we obtain

\[
r(t) = \frac{1}{2} (X_t - S^{-1}v)'S(X_t - S^{-1}v). \quad (9.17)
\]

By Proposition 9.2, $r(t)$ is a squared Gaussian process.

To recover the ZCB prices we have to calculate $E^X[\xi(T)|\mathcal{F}_t]$, which equals $e^{-\alpha(T-t)}E^X_t[f(X_{T-t})]$ due to the Markov property of $X$. We introduce $\tau := T - t \geq 0$. As shown in [85], we have

\[
E^X[f(X_\tau)] = |1 - QV(\tau)|^{-\frac{1}{2}} \\
\exp \left\{ \frac{1}{2} (\mu(x, \tau) - c)' (1 - QV(\tau))^{-1} Q(\mu(x, \tau) - c) \right\},
\]

where $|A|$ denotes the determinant of a matrix $A \in \mathbb{R}^{d \times d}$, and the conditional mean $\mu(x, t)$ and the conditional covariance $V(t)$ are as specified in Proposition 9.2, that is

\[
\begin{align*}
    \mu(x, \tau) &= e^{-\Gamma \tau} x \\
    V(\tau) &= \int_0^\tau e^{-\Gamma s} (e^{-\Gamma s})' ds.
\end{align*}
\]

The ZCB prices are then given by

\[
\begin{align*}
P(t, T) &= e^{-\alpha \tau} |1 - QV(\tau)|^{-\frac{1}{2}} \\
&\quad \exp \left\{ \frac{1}{2} (\mu(X_t, \tau) - c)' (1 - QV(\tau))^{-1} Q(\mu(X_t, \tau) - c) \\
&\quad \quad - \frac{1}{2} (X_t - c)'Q(X_t - c) \right\}. \quad (9.18)
\end{align*}
\]
For the yields, defined as $R(t, T) = -\frac{1}{\tau} \log P(t, t + \tau)$, we obtain

$$R(t, T) = \alpha + \frac{1}{2\tau} \log |1 - QV(\tau)|
- \frac{1}{2\tau} \left\{ (\mu(X_t, \tau) - c)' (1 - QV(\tau))^{-1} Q (\mu(X_t, \tau) - c)
- (X_t - c)' Q (X_t - c) \right\}. \quad (9.19)$$

The instantaneous forward rates $f(t, T)$ are somewhat tedious to compute. Recall from Section 6.1.1 that

$$f(t, T) = -\frac{\partial}{\partial T} \log P(t, T), \quad 0 \leq t \leq T.$$  

Applying this to the ZCB prices from (9.18), we obtain

$$f(t, T) = \alpha + \frac{1}{2 |A(\tau)|} \frac{\partial}{\partial \tau} |A(\tau)|
- \frac{1}{2} \frac{\partial}{\partial \tau} (\mu(X_t, \tau) - c)' A(\tau)^{-1} Q (\mu(X_t, \tau)), \quad (9.20)$$

where we use the definition

$$A(\tau) := 1 - QV(\tau), \quad \tau \geq 0, \quad (9.21)$$

which is a matrix-valued function of $\mathbb{R}^+_0$. Following Section 8.3 of [122], we have

$$\frac{\partial}{\partial \tau} |A(\tau)| = \sum_{i=1}^{d} \sum_{j=1}^{d} \hat{A}_{ij}(\tau) |A_{ij}(\tau)|, \quad (9.22)$$

where $A_{ij}(\tau)$ denotes the complement matrix of $A_{ij}(\tau)$, that is, the matrix $A(\tau)$ without its $i$-th row and $j$-th column, and

$$\hat{A}(\tau) := \frac{\partial}{\partial \tau} A(\tau) = -Q (e^{-\Gamma\tau}) (e^{-\Gamma\tau})', \quad (9.23)$$
9.3. The Model

where the second equality is by the representation of $V(\tau)$ from Proposition 9.2. For the third term on the right-hand side of (9.20), we obtain

$$\frac{\partial}{\partial \tau} (\mu(X_t, \tau) - c)' A^{-1}(\tau) Q (\mu(X_t, \tau)) =$$

$$- 2(\Gamma \mu(X_t, \tau))' A^{-1}(\tau) Q \mu(X_t, \tau) + 2c' A^{-1}(\tau) Q (\Gamma \mu(X_t, \tau))$$

$$- \mu(X_t, \tau)' A^{-1}(\tau) \hat{A}(\tau) A^{-1}(\tau) Q \mu(X_t, \tau) + 2c' A^{-1}(\tau) \hat{A}(\tau) A^{-1}(\tau) Q c,$$

where we use the obvious fact that

$$\frac{\partial}{\partial \tau} \mu(X_t, \tau) = -\Gamma \mu(X_t, \tau), \quad (9.25)$$

and the less obvious one that

$$\frac{\partial}{\partial \tau} A^{-1}(\tau) = -A^{-1}(\tau) \hat{A}(\tau) A^{-1}(\tau), \quad (9.26)$$

which is obtained by differentiating $A(\tau) A^{-1}(\tau) \equiv 1$ and applying the product rule for matrix-valued functions from Section 4.2 of [122]. The spot FX rate between economies $i$ and $j$ is given by

$$C_{ij}(t) = C_{ij}(0) \exp \left\{ - (\alpha^{(j)} - \alpha^{(i)}) t + \frac{1}{2} \left( X_t - c^{(j)} \right)' Q^{(j)} \left( X_t - c^{(j)} \right) - \frac{1}{2} \left( X_t - c^{(i)} \right)' Q^{(i)} \left( X_t - c^{(i)} \right) - \frac{1}{2} \left( x - c^{(j)} \right)' Q^{(j)} \left( x - c^{(j)} \right) + \frac{1}{2} \left( x - c^{(i)} \right)' Q^{(i)} \left( x - c^{(i)} \right) \right\}.$$  

(9.27)

We observe that the exponential-quadratic approach leads to rather cumbersome forms for the different prices yields and exchange rates. This makes further analytical evaluations rather tedious or even impossible. Numerically, if the task is just to compute the above expressions from simulated values of the driver process $X$, there are no particular problems
(for instance using Matlab). However, the – generally non-linear – constraints necessary to make the model relevant may be hard to implement in a calibration procedure. It is, therefore, worthwhile to think about simplifications:

**Remark 9.5** As is well known from linear algebra, for every symmetric bilinear form on $\mathbb{R}^d$, there exists an orthonormal basis with respect to which this bilinear form is represented by a diagonal matrix whose entries are the eigenvalues. So, we have for our bilinear form:

$$(x - c)'Q(x - c) = (x - c)'B'\Lambda B(x - c),$$

where $B$ is an orthogonal matrix and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)$. We note, moreover, that $\lambda_i$ is positive for all $i$. As long as $X$ is a purely abstract state process, we do not lose generality with respect to the original setup if we assume $X$ to be rotated by $B$, and let then $Q$ be a diagonal matrix with positive entries. This leads to more parsimonious models, and the requirement for positive definiteness of $Q$ takes a very simple form.

### 9.4 Properties of the Model

#### 9.4.1 Preliminaries

Following Remark 9.5, let us assume that the matrix $Q$ is diagonal with positive entries (i.e. positive definite), and that the matrix $\Gamma$ is according to Definition 9.1, such that the stationary law of the driver process $X$ exists; see Proposition 9.2. The structure of the model implies that we can express $R(t, T)$ as $R(X_t, \tau)$ for $\tau = T - t \geq 0$. Now, we can rewrite (9.19) as

$$R(X_t, \tau) = \alpha + \frac{1}{2\tau} \log |1 - QV(\tau)| - \frac{1}{2\tau} q_\tau(X_t),$$

(9.28)

where

$$q_\tau(x) = x' A(\tau)x + b(\tau)'x + c(\tau)$$
9.4. Properties of the Model

with

\[ A(\tau) = e^{-\Gamma''\tau} (1 - QV(\tau))^{-1} Q e^{-\Gamma\tau} - Q, \]
\[ b(\tau) = 2 \left[ Q - (1 - QV(\tau))^{-1} Q e^{-\Gamma\tau} \right] c, \]
\[ c(\tau) = c' \left[(1 - QV(\tau))^{-1} Q - Q\right] c. \]

We note that the matrix \( A(\tau) \) is symmetric. For any fixed time to maturity \( \tau \), the term \( q(\tau)(X_t) \) is a quadratic form of a normally distributed vector. In order to obtain a similar representation for the exchange rate, we rely on the log-return or depreciation rate \( c_{ij}(t, T) \) rather than on the spot exchange rate:

\[ c_{ij}(t, T) = \log C_{ij}(T) - \log C_{ij}(t), \quad T \geq t. \quad (9.29) \]

Combining (9.27) and (9.29), we obtain the representation

\[ c_{ij}(t, T) = - \left( \alpha^{(j)} - \alpha^{(i)} \right) (T - t) \]
\[ + \frac{1}{2} \left( q^{(j)}(X_T) - q^{(j)}(X_t) - q^{(i)}(X_T) + q^{(i)}(X_t) \right), \quad (9.30) \]

where

\[ q^{(i)}(x) = x'A_i x + b'_ix + c_i \quad \text{with} \quad A_i = Q^{(i)}, \]
\[ b_i = -2Q^{(i)}c_i, \quad c_i = c^{(i)}Q^{(i)}c^{(i)}; \]
\[ q^{(j)}(x) = x'A_j x + b'_jx + c_j \quad \text{with} \quad A_j = Q^{(j)}, \]
\[ b_j = -2Q^{(j)}c_j, \quad c_j = c^{(j)}Q^{(j)}c^{(j)}. \]

These representations in terms of quadratic forms of normal random vectors are the key to the further investigation of the properties of the model. Before proceeding further, we state some related general results which are due to Mathai and Provost; see [96].

**Proposition 9.6** Let \( X \sim \mathcal{N}_d(\mu, V) \) and let \( q(x) = x'Ax + b'x + c \) be a quadratic form with \( A \in \mathbb{R}^{d \times d} \) symmetric, \( b \in \mathbb{R}^d \) and \( c \in \mathbb{R} \). Then:

\[ \mathbb{E}[q(X)] = tr[AV] + \mu' A\mu + b'\mu + c, \]
\[ \mathbb{V}ar[q(X)] = 2tr[(AV)^2] + 4\mu'AVA\mu + 4b'AVA\mu + b'Vb. \]
If we have a second quadratic form \( q(x) = x'Ax + b'x + c \) of the same kind as \( q(x) \), then we have:

\[
\text{Cov}[q(X), q(X)] = 2 \text{tr}[AVÄV] + 2\mu'AVÄ\mu
+ 2\mu'AVÄb + 2b'VÄ\mu + b'VÄb.
\]

Proof. See [96], Section 3.3, in particular Theorem 3.3.2 and 3.3.4. \( \Box \)

**Proposition 9.7** Let \( X = (X_t)_{t \geq 0} \) be a multivariate Ornstein-Uhlenbeck process as given by Definition 9.1, and let \( q(x) \) and \( q(x) \) be quadratic forms as in Proposition 9.6. Then, for \( h \geq 0 \), we have

\[
\text{Cov}[q(X_t), q(X_{t+h})] = 2 \text{tr} \left[ e^{-\Gamma'h}Ae^{-\Gamma'h}V(\infty)ÄV(\infty) \right]
+ b'e^{-\Gamma'h}V(\infty)b.
\]

Proof. See [88]. \( \Box \)

### 9.4.2 Interest Rates

Now we can investigate the unconditional moments of \( R(X_t, \tau) \) under the stationary law of \( X_t \) which is normal with zero mean and covariance matrix \( V(\infty) \) according to Proposition 9.3. By applying Proposition 9.6 to (9.28), we can compute the unconditional expectation

\[
\mathbb{E}[R(X_t, \tau)] = \alpha + \frac{1}{2\tau} \log|1 - QV(\tau)| - \frac{1}{2\tau} \left( \text{tr}[A(\tau)V(\infty)] - c(\tau) \right). \tag{9.31}
\]

For the unconditional variance of \( R(X_t, \tau) \), we obtain

\[
\text{Var}[R(X_t, \tau)] = \frac{1}{4\tau^2} \left( 2\text{tr} [(A(\tau)V(\infty))^2] + b(\tau)'V(\infty)b(\tau) \right). \tag{9.32}
\]

The unconditional contemporaneous covariance can be investigated between different maturities \( \tau_1 \) and \( \tau_2 \) as well as between different economies \( i \) and \( j \):

\[
\text{Cov}[R_i(X_t, \tau_1), R_j(X_t, \tau_2)] = \frac{1}{4\tau_1\tau_2} \left( 2\text{tr} [A_i(\tau_1)V(\infty)A_j(\tau_2)V(\infty)]
+ b_i(\tau_1)'V(\infty)b_j(\tau_2) \right). \tag{9.33}
\]
9.4. Properties of the Model

Using Proposition 9.7, we can also investigate the autocovariance (the extension to multi-economy setting is straightforward):

\[
\text{Cov} [R(X_{t+h}, \tau), R(X_t, \tau)] = \frac{1}{4 \tau^2} \left( 2 \text{tr} \left[ e^{-\Gamma h} A(\tau) e^{-\Gamma h} V(\infty) A(\tau) V(\infty) \right] \right.

\left. + b(\tau)' e^{-\Gamma h} V(\infty) b(\tau) \right).
\]  

(9.34)

An investigation of the conditional dynamics of \( R(X_t, \tau) \) would most naturally also cover the conditional variance of the levels, that is

\[
\text{Var}[R(X_{t+h}, \tau)|R(X_t, \tau)].
\]

But this is not tractable due to the structure of the model, and also difficult to estimate empirically. However, we can easily investigate the unconditional variance of the differences, i.e. \( \text{Var}[R(X_{t+h}, \tau) - R(X_t, \tau)] \). As shown in [88], the latter is a good proxy for the former in situations where the levels are highly persistent, which is the case with interest rates as we see in Section 6.3. We have:

\[
\text{Var} [R(X_{t+h}, \tau) - R(X_t, \tau)] = 2 \left( \text{Var} [R(X_t, \tau)]

- \text{Cov} [R(X_{t+h}, \tau), R(X_t, \tau)] \right).
\]  

(9.35)

where the first-term on the right-hand side is given by (9.32) and the second by (9.34). This expression also has an easily tractable empirical counterpart. If this variance is considered for several time lags \( h \) simultaneously, then it is advisable to divide by \( h \) so as to have everything annualized. The question arises if – or under what circumstances – the expression \( h^{-1} \text{Var}[R(X_{t+h}, \tau) - R(X_t, \tau)] \) can reproduce the hump-shaped volatility pattern of the data shown in Figure 6.3. To investigate this, let us assume that the mean reversion matrix \( \Gamma \) is diagonal. We note that the diagonal entries \( \Gamma_{ii} \) are then positive according to Definition 9.1. Since the matrix \( Q \) is assumed to be diagonal anyway, we also have that the matrices \( A(\tau) \) and \( V(\infty) \) in (9.32) and (9.34) are diagonal. Given this, we can write

\[
f(h) = \frac{1}{h} \text{Var}[R(X_{t+h}, \tau) - R(X_t, \tau)]

= \frac{1}{2 \tau^2} \sum_{i=1}^{d} a_i f_i(h) + b_i g_i(h),
\]  

(9.36)
where

\[ f_i(h) = \frac{2}{h} (1 - e^{-2\Gamma_{ii}h}), \quad a_i = A_{ii}(\tau)^2 V_{ii}(\infty)^2 > 0, \]

\[ g_i(h) = \frac{1}{h} (1 - e^{-\Gamma_{ii}h}), \quad b_i = b_i(\tau)^2 V_{ii}(\infty) > 0, \]

by plugging (9.32) and (9.34) into (9.35) and rearranging. It is straightforward to show that both \( f_i(h) \) and \( g_i'(h) \) are negative and cannot have a zero for \( h > 0 \). Given the positivity of \( a_i \) and \( b_i \), the function \( f(h) \) in (9.36) is, therefore, downward-sloping. Hence, if the model has to produce hump-shaped dynamics as in Figure 6.3, we have to allow non-zero off-diagonal elements in the mean reversion matrix \( \Gamma \).

In this context, we also note that, due to the time-homogeneity of the model, we have

\[ E[R(X_{t+h}, \tau) - R(X_t, \tau)] = 0. \tag{9.37} \]

Desirable though it would be, it is not possible to make statements on the exact distribution of \( R(X_t, \tau) \). To get an idea, we note that the quadratic form \( q_\tau(x) \) in (9.28) can be re-written as

\[ q_\tau(x) = \left(x - \frac{1}{2} A(\tau)^{-1} b(\tau)\right)' A(\tau) \left(x - \frac{1}{2} A(\tau)^{-1} b(\tau)\right) - \frac{1}{4} b(\tau)' A(\tau)^{-1} b(\tau) + c(\tau) \]

provided \( A(\tau) \) is invertible. That is, up to some scaling and offset, \( R(X_t, \tau) \) has the same distribution as

\[ Y' A(\tau) Y, \quad \text{where} \quad Y \sim N_d \left( \mu(X_0, t) + A(\tau)^{-1} b(\tau)/2, V(t) \right), \]

with \( \mu(x, t) \) and \( V(t) \) as in Proposition 9.2. Under very restrictive conditions, which do not apply here, this would have a non-central \( \chi^2 \)-distribution. However, in the more general case relevant here, we are only left with a number of rather complicated infinite series representations; see Chapter 4 of [96]. Rather than studying suitable finite approximations of these representations, we will do an empirical evaluation of simulated data in Chapter 10 to investigate the features of the distribution of \( R(X_t, \tau) \).
As already mentioned, the model structure precludes the exact formulation of conditional moments given the interest rates at some time $t$. This is because the current value of the driver process is unobservable and can only be estimated. Doing this for the initial value $X_0 = x$ is necessary, but doing it for a large number of time steps during calibration would seriously hamper the accuracy of the latter and is, therefore, not useful. Nevertheless, the most important moments conditional on $x$ are stated here, in order to provide some a-priori understanding of the behaviour of a calibrated model. The conditional expectation is given by

$$\begin{align*}
\mathbb{E}^x[R(X_t, \tau)] &= \alpha + \frac{1}{2\tau} \log |1 - QV(t)| - \frac{1}{2\tau} \left( tr[A(\tau)V(t)] \\
&+ \mu(x, t)'A(\tau)\mu(x, t) + b(\tau)'\mu(x, t) + c(\tau) \right). \quad (9.38)
\end{align*}$$

This result is obtained by applying Proposition 9.6 to the representation in (9.28) using the result on the conditional distribution of the driver $X$ from Proposition 9.2. The terms $A(\tau)$, $b(\tau)$ and $c(\tau)$ are as in (9.28), and $\mu(x, t)$ and $V(t)$ are as in Proposition 9.2. Using the same machinery, one obtains

$$\begin{align*}
\operatorname{Var}^x[R(X_t, \tau)] &= \frac{1}{4\tau^2} \left( tr \left[ (A(\tau)V(t))^2 \right] \\
&+ 4\mu(x, t)'A(\tau)V(t)A(\tau)\mu(x, t) \\
&+ 4b(\tau)'V(t)A(\tau)\mu(x, t) + b(\tau)'V(t)b(\tau) \right). \quad (9.39)
\end{align*}$$

Finally, by using Proposition 9.7 instead of Proposition 9.7, one obtains a result on the conditional dependence structure:

$$\begin{align*}
\operatorname{Cov}^x \left[ R^{(i)}(X_t, \tau_i), R^{(j)}(X_t, \tau_j) \right] &= \frac{1}{4\tau_i \tau_j} \left( 2tr \left[ A^{(i)}(\tau_i)V(t)A^{(j)}(\tau_j)V(t) \right] \\
&+ 4\mu(x, t)'A^{(i)}(\tau_i)V(t)A^{(j)}(\tau_j)\mu(x, t) + 2\mu(x, t)'A^{(i)}(\tau_i)V(t)b^{(j)}(\tau_j) \\
&+ 2b^{(j)}(\tau_j)V(t)A^{(j)}(\tau_j)\mu(x, t) + b^{(i)}(\tau_i)'V(t)b^{(j)}(\tau_j) \right). \quad (9.40)
\end{align*}$$

### 9.4.3 Exchange Rates

In the same way as for the interest rates, we can now compute the unconditional moments of the log-returns $c_{ij}(t, T)$ of the the exchange rate. Recall
that the stationary distribution of the driver process is $\mathcal{N}_d(0, V(\infty))$ according to Proposition 9.3. Applying this result and Proposition 9.6 to the representation of $c_{ij}(t, T)$ in (9.30), we obtain

$$E[c_{ij}(t, T)] = -\left(\alpha^{(j)} - \alpha^{(i)}\right)(T - t), \quad T \geq t. \quad (9.41)$$

The computation of $\text{Var}[c_{ij}(t, T)]$ is, in principle, also straightforward, but rather tedious. Consider again (9.30) and define a random vector $Z := (q_j(X_T), -q_j(X_t), -q_i(X_T), -q_i(X_t))^\top$. Then we have

$$\text{Var}[c_{ij}(t, T)] = \frac{1}{4}e'\text{Cov}[Z]e, \quad (9.42)$$

where $e = (1, 1, 1, 1)^\top$. Let us assume the following notation for the structure of the matrix $\text{Cov}[Z]$

$$\text{Cov}[Z] = \begin{pmatrix} V_1 & -C_{12} & -C_{13} & C_{14} \\ -C_{12} & V_2 & C_{23} & -C_{24} \\ -C_{13} & C_{23} & V_3 & -C_{34} \\ C_{14} & -C_{24} & -C_{34} & V_4 \end{pmatrix}.$$  

Then we can use Proposition 9.6 to compute

$$V_1 = \text{Var}[q_j(X_T)], \quad V_2 = \text{Var}[q_j(X_t)],$$
$$V_3 = \text{Var}[q_i(X_T)], \quad V_4 = \text{Var}[q_i(X_t)].$$

and Proposition 9.7 to compute

$$C_{12} = \text{Cov}[q_j(X_T), q_j(X_t)], \quad C_{13} = \text{Cov}[q_j(X_T), q_i(X_T)],$$
$$C_{14} = \text{Cov}[q_j(X_T), q_i(X_t)], \quad C_{23} = \text{Cov}[q_j(X_t), q_i(X_T)],$$
$$C_{24} = \text{Cov}[q_j(X_t), q_i(X_t)], \quad C_{34} = \text{Cov}[q_i(X_T), q_i(X_t)].$$

It would be possible to further simplify this expression (e.g. $V_1 = V_2$), but we do not pursue this further as the computation of these expressions does not pose any fundamental problems. Along the same lines, one could as well compute further expressions like the autocovariance of the log returns, or the covariance of the log returns with interest rates.

### 9.5 The Context of the Model

We have already mentioned in Section 9.1 that the exponential-linear form of the price kernel according to (9.4) gives rise to models belonging to the
well-known class of Affine Term Structure Models (ATSM). But also the model that we have constructed in Section 9.3 based on the exponential-quadratic function from (9.3) is embedded in a broader class, namely the Quadratic Term Structure Models (QTSM).

**Definition 9.8** In a Quadratic Term Structure Model, the zero-coupon bond prices \( P(t, t + \tau) \) are quadratic functions

\[
P(t, t + \tau) = P(X_t, \tau) = \exp\{-X'_t A(\tau) X_t + b'(\tau) X_t + c(\tau)\},
\]

where \( X = (X_t)_{t \geq 0} \) is a \( d \)-dimensional Markov process solving

\[
dX_t = \mu(X_t) dt + \sigma(X_t) d\mathbf{B}_t
\]
as in Definition 8.1, \( A(\tau) \in \mathbb{R}^{d \times d} \) non-singular, \( b(\tau) \in \mathbb{R}^d \) and \( c(\tau) \in \mathbb{R} \).

Considering the expression for the ZCB prices in (9.18), we immediately see that our model fits into the definition. We can also see that Definition 9.8 extends the class of ATSM (9.7) by adding a quadratic term. As for the ATSM (see [49]), there are also characterisation results for QTSM.

**Proposition 9.9** The following conditions are necessary and sufficient for a model to be a QTSM:

1. The short rate \( r(t) \) is a quadratic function of the state process \( X \).
2. The drift of the state process \( X \) under the risk-neutral measure \( Q \) is an affine function.
3. The diffusion matrix of the state process \( X \) is constant.

**Proof.** See [87]. □

This proposition can be used as an alternative definition of a QTSM, as done in [3]. Plugging the market price of risk from (9.16) into the state process \( X \) according to Definition 9.1, we can see that the \( Q \)-dynamics of the latter are given by

\[
dX_t = (Qc - (\Gamma + Q)X_t) dt + d\mathbf{B}_t.
\]
Together with the expression for the short rate in (9.15), this shows that our model also fits into the alternative equivalent characterization provided by Proposition 9.9. Moreover, Proposition 9.9 shows that if we used a CIR-type process according to (9.2), we would obtain a model outside the QTSM class. Yet another characterization, used by [57] is that the forward rates must be a quadratic function of the state process \( X \), that is:

\[
f(t, t + \tau) = X'_t A_f(\tau) X_t + b_f(\tau)' X_t + c(\tau).
\]

Since we have

\[
P(t, t + \tau) = \exp \left\{ \int_0^\tau f(t, t + s) ds \right\} \quad \text{and} \quad f(t, t + \tau) = -\frac{\partial}{\partial \tau} \log P(t, t + \tau),
\]

the specification in terms of forward rates is equivalent to the one in Definition 9.8. (9.20) shows that our model is also compatible with this last specification of QTSM.

As in the case of ATSM, the functions \( A(\tau) \), \( b(\tau) \) and \( c(\tau) \) in Definition 9.8 are generally only defined implicitly as the solutions of a system of differential equations, see [88]. However, there exist explicit solutions in a much wider range of cases than for ATSM. In particular, we have managed to define a fairly general class of QTSM where all important quantities have explicit expressions.

Unlike for ATSM or HJM-type models, the literature on QTSM is very limited. The term QTSM itself was probably coined by Ahn, Dittmar and Gallant [3] when they elaborated a unified representation for a few older models (including [36]). Moreover, they showed in an empirical study that QTSM are much better capable of reproducing the empirical properties of interest rates than ATSM. This is an important advantage in view of the requirements posed by DFA.

Thorough investigations were also carried out by Leippold and Wu. In [86], they review and formulate models by using the price kernel notation. The case study [88] confirms the good tractability of the models — also in terms of calibration — and the superior empirical characteristics. In [87], they show that QTSM also have favourable properties if it comes to the
9.5. The Context of the Model

pricing of a wide range of fairly general options.

Finally, one can even argue that QTSM are the most general sensible class of models in the following sense: In the specification of the ZCB prices from Definition 9.8, one could add terms of degrees higher than two, leading to *Polynomial Term Structure Models*. However, Filipović shows in [57] that any polynomial model with degree higher than two violates essential consistency conditions. Finally, [30] also provides a very rigorous mathematical treatment of the QTSM class in the language of general Markov process theory. Moreover, they suggest an extension of the QTSM class to include defaultable securities.
Seite Leer / Blank leaf
Chapter 10

Calibration of Models

10.1 Problem Statement and Basics

We start by describing the general estimation problem and introducing a few fundamental notions. A comprehensive reference on modelling and estimation theory is [62]. A stochastic model is a pair $(\mathcal{Y}, \mathcal{P})$, where $\mathcal{Y}$ is the set of possible observations, and $\mathcal{P}$ is a set of probability distributions. Calibration is the task of selecting a suitable distribution $P \in \mathcal{P}$ for a given observation $Y \in \mathcal{Y}$. In the sequel, we will only deal with parametric models, i.e. with models where $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^K$ is the parameter space. In general, we will estimate some function $g(\theta)$ rather than $\theta$ itself. An estimator is then a mapping of the form:

$$
\delta : \mathcal{Y} \rightarrow g(\Theta)
$$

$$
Y \rightarrow g(\theta).
$$

A desirable property of any estimator is unbiasedness, meaning that, on average, the estimator produces the true value of the parameter:

Definition 10.1 An estimator $\delta$ is an unbiased estimator of $g(\theta)$ if

$$
\mathbb{E}_\theta[\delta(Y)] = g(\theta), \quad \theta \in \Theta.
$$
An observation $Y \in \mathcal{Y}$ is assumed to be of the form $Y = \{Y_1, \ldots, Y_n\}$, where each $Y_i$ is in $\mathbb{R}^p$ for some $p \geq 1$. Typically, one $Y_i$ consists of a cross-section of yields for different maturities and FX rates at some point in time, thus making $Y$ a multivariate time series. In general, the length $n$ of the observation $Y$ — or, rather, the fact that $n$ is typically smaller than desirable — plays an important role in the calibration of interest rate models. Given this, we can consider sequences of models $(\mathcal{Y}_n, \mathcal{P}_n) = \{P^n_\theta : \theta \in \Theta\}$ that depend on $n$. Notice, however, that the parameter $\theta$ is assumed independent of $n$, otherwise we run into problems. In turn, we also consider sequences of estimators $\delta_n : \mathcal{Y}_n \rightarrow g(\Theta)$. This naturally leads to

**Definition 10.2** A sequence of estimators $\{\delta_n\}$ is asymptotically unbiased if

$$E_\theta[\delta_n(Y)] \xrightarrow{n \to \infty} g(\theta), \quad \theta \in \Theta.$$  

A more useful criterion is the one of consistency:

**Definition 10.3** A sequence of estimators $\{\delta_n\}$ is called

1. Weakly consistent if $\delta_n(Y)$ converges in $P_\theta$-probability to $g(\theta)$.

2. Strongly consistent if $\delta_n(Y)$ converges $P_\theta$-almost surely to $g(\theta)$.

Another important property of parametric models is its identification. In the absence of the latter, it may not be possible to obtain sensible parameter estimates, even under perfect information. In particular, if $g(\theta)$ is a non-identified parameter function, then there exists no unbiased estimator of $g(\theta)$; see [62], p. 129.

**Definition 10.4** Let $(\mathcal{Y}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ be a parametric model.

1. A parameter value $\theta_1 \in \Theta$ is identified if there exists no other parameter value $\theta_2 \in \Theta$, $\theta_2 \neq \theta_1$ such that $P_{\theta_1} = P_{\theta_2}$.

2. The parametric model $\{P_\theta : \theta \in \Theta\}$ is identified if each $\theta \in \Theta$ is identified.
10.1. Problem Statement and Basics

3. A function $g(\theta)$ of the parameter is identified if

$$g(\theta_1) \neq g(\theta_2) \Rightarrow P_{\theta_1} \neq P_{\theta_2}, \quad \theta_1, \theta_2 \in \Theta.$$ 

Up until now, we have focused on ways in which the estimate $\delta(Y)$ converges to the true value $g(\theta)$. In the sequel, we turn our attention to the dispersion of the estimate $\delta(Y)$ around $g(\theta)$. We assume that the parametric model $(\mathcal{Y}, \mathcal{P} = \{P_\theta : \theta \in \Theta\})$ has densities $\ell(y; \theta)$ with respect to some reference measure $\mu$ and for all $\theta \in \Theta$.

**Definition 10.5** A parametric model with densities $\ell(y, \theta)$, $\theta \in \Theta$, is regular if:

1. $\Theta$ is an open subset of $\mathbb{R}^p$.
2. $\ell(y, \theta)$ is differentiable with respect to $\theta$.
3. $\int_{\mathcal{Y}} \ell(y, \theta) d\mu(y)$ is a differentiable function of $\theta$ with

$$\frac{\partial}{\partial \theta} \int_{\mathcal{Y}} \ell(y, \theta) d\mu(y) = \int_{\mathcal{Y}} \frac{\partial}{\partial \theta} \ell(y, \theta) d\mu(y).$$

4. The Fisher information matrix

$$\mathcal{I}(\theta) = \mathbb{E}_\theta \left[ \frac{\partial}{\partial \theta} \log \ell(y, \theta) \frac{\partial}{\partial \theta} \log \ell(y, \theta)' \right]$$

exists and is positive definite for all $\theta \in \Theta$.

**Theorem 10.6** Assume that we are given a regular model in the sense of Definition 10.5. Every unbiased estimator $\delta(Y)$ for $g(\theta) \in \mathbb{R}^q$ that satisfies the additional regularity condition

$$\mathbb{E}_\theta \left[ \|\delta(Y)\|^2 \right] < \infty, \quad \theta \in \Theta,$$

has a covariance matrix that satisfies

$$\text{Cov}_\theta [\delta(Y)] \succeq \frac{\partial g(\theta)'}{\partial \theta} \mathcal{I}(\theta)^{-1} \frac{\partial g(\theta)}{\partial \theta},$$

where $A \succeq B$ is to be understood in the sense that $A - B$ is positive semidefinite. The right-hand side of the inequality is called the Cramér-Rao bound.
Proof. See [62], p. 132-133.

As a consequence, we can identify an optimality criterion for unbiased estimators:

**Definition 10.7** Under a parametric model which is regular in the sense of Definition 10.5, a square-integrable and unbiased estimator of $g(\theta)$ is called efficient if its covariance matrix attains the Cramér-Rao bound.

This definition naturally extends to the asymptotic case. An efficient estimator is an estimator whose value is – on average – equal to the true value to be estimated and has the lowest possible dispersion around the true value.

### 10.2 Generalized Method of Moments

In this Section, we introduce the Generalized Method of Moments (GMM) as the tool for calibrating the interest rate and FX model as introduced in Chapter 9 to historical data. The presentation here takes into account the fact that our model has a structure that is somewhat different from most of the standard interest rate models. A survey of alternative methods, that also serves a motivation of the choice of GMM, will be given in Section 10.3.

Our interest rate and FX model has the following generic structure: We have a stationary process $(Z_t)_{t \geq 0}$ of observables, where $Z_t \in \mathbb{R}^n$ for all $t \geq 0$ and some $n \in \mathbb{N}$. These observables are the yields for several times to maturity and several economies plus the exchange rates between the involved economies. Instead of yields, one can also use ZCB prices, forward rates or yield spreads, and instead of plain values, one can also use first differences or log returns for numerical or analytical convenience. In all cases, the observables are a function of an unobservable $d$-dimensional driver process $(X_t)_{t \geq 0}$, specifically:

$$Z_t = z(X_t, W_t, \rho_1), \quad t \geq 0.$$  (10.1)
Here, \((W_t)_{t \geq 0}\) is a noise process. In the model class considered here, this noise process is absent, but it may prove useful to introduce it artificially for estimation purposes. \(\rho_1 \in \mathbb{R}^{P_1}\) is a vector of parameters pertaining specifically to the function \(z(.)\). The state process is the well known driver process as introduced in Section 8.1; it follows a time-homogeneous Itô diffusion

\[
dX_t = \mu(X_t, \rho_2)\,dt + \sigma(X_t, \rho_2)\,dB_t
\]

according to Definition 8.1. This process is also Markovian. We will restrict our attention to the specific choice introduced in Section 9.2. \(\rho_2 \in \mathbb{R}^{P_2}\) is a vector of parameters pertaining specifically to the driver process. The parameter vector of the full model is \(\rho = (\rho_1', \rho_2')' \in \mathbb{R}^P\), where \(P = P_1 + P_2\). In general – and in all models considered here – \(\rho\) is subject to a number of constraints that assure certain desirable properties of the model like numerical stability, positive interest rates or absence of arbitrage. In general, we can formulate this as

\[
c(\rho) \in \mathcal{R} \subset \mathbb{R}^Q,
\]

where \(c : \mathbb{R}^P \rightarrow \mathbb{R}^Q\) is some – not necessarily linear – function. Note that the calibration literature, including the references cited further below, either completely disregards the issue of constraints, or only gives it a marginal role.

On the empirical side, we have observations \(\tilde{Y}_t \in \mathbb{R}^N\) at discrete equi-spaced time points \(t\Delta t\) for \(t = 1, \ldots, T\) and \(\Delta t > 0\). For the sake of simplicity, we index the observations by \(t\) only. The task is then to find a suitable value for \(\rho\) given the observations \((\tilde{Y}_t)_{t=1}^T\) and respecting possible constraints.

If our model is stationary for some \(\rho \in \mathcal{R}\), then we can analyze the stationary, time-invariant law of the observables. We assume that this law has a density

\[p(Z_0, Z_{-1}, \ldots, Z_{-L}; \rho)\]

for any set of observables from different lags. For simplicity of notation, we stack the observables from the different lags in one vector \(\mathbf{Y} = \)
Chapter 10. Calibration of Models

$(Z_0, Z_{-1}, \ldots, Z_{-L})' \in \mathbb{R}^N$ where $N = (L+1)n$; with the empirical counterpart $\bar{Y}_t$ being defined equivalently. Suppose now that we can derive some moment relation

$$E_\rho [f(Y; \rho)] = g(\rho) \quad (10.4)$$

for suitable functions $f: \mathbb{R}^N \times \mathbb{R}^P \rightarrow \mathbb{R}^M$ and $g: \mathbb{R}^P \rightarrow \mathbb{R}^M$. Equivalently, we can look for some function $m(\rho; Y) \in \mathbb{R}^M$ such that

$$E_\rho [m(\rho; Y)] = 0, \quad (10.5)$$

i.e. (10.4) turns into $m(\rho; Y) = f(Y, \rho) - g(\rho)$. If our model is also ergodic for $\rho$, then the empirical counterpart

$$m(\rho; \bar{Y}) = \frac{1}{T} \sum_{t=1}^T m(\rho; \bar{Y}_t) \quad (10.6)$$

provides a consistent estimate of the theoretical expectation in (10.5). If $M = P$ then we can simply solve $m(\rho; \bar{Y}) = 0$ to obtain the estimate $\hat{\rho}$. This is known as the Classical Method of Moments (CMM). However, in order to bring more information into the calibration, we generally work with an over-specified model, i.e. $M > P$. Then we can solve

$$\hat{\rho}_T = \arg\min_{\rho \in \mathbb{R}^P} m(\rho; \bar{Y})' W m(\rho; \bar{Y}) \quad \text{s.t.} \quad c(\rho) \in \mathcal{R} \quad (10.7)$$

for some symmetric and positive definite weighting matrix $W \in \mathbb{R}^{M \times M}$. This is then called the Generalized Method of Moments (GMM). Let $\rho_0$ denote the true parameter of the model. Then, for the asymptotic distribution of the GMM estimate $\hat{\rho}_T$, we have

$$\sqrt{T} (\hat{\rho}_T - \rho_0) \overset{\text{as}}{\sim} \mathcal{N}(0, V), \quad (10.8)$$

meaning that GMM is asymptotically unbiased in the sense of Definition 10.2. The covariance matrix $V$ is given by

$$V = (D_0' WD_0)^{-1} D_0' WSWD_0 (D_0' WD_0)^{-1}, \quad (10.9)$$

where $D_0$ is related to the Jacobian of the moment condition functions, namely

$$D_0 = E_{\rho_0} \left[ \frac{\partial}{\partial \rho_0} m(\rho_0; Y) \right], \quad (10.10)$$
and $S$ is given by

$$S = \lim_{T \to \infty} \operatorname{Var} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(\rho_0; Y_t) \right]. \tag{10.11}$$

From (10.9) it becomes clear that the GMM estimator attains minimum variance if the weighting matrix $W$ is equal to $S^{-1}$ or any scalar multiple thereof. In this case, the optimal value of $V$ is $V^* = (D_0^\prime S^{-1} D_0)^{-1}$. Moreover, if $W = S^{-1}$, then $T$ times the minimized objective function from (10.7) is asymptotically $\chi^2$-distributed with $M - P$ degrees of freedom. Therefore, this type of estimation is often also called Minimum $\chi^2$ Estimation.

The problem with the optimal weighting matrix $W = S^{-1}$ is that $S$ actually depends on the parameter $\rho$ to be estimated; see (10.11). However, even with a non-optimal weighting matrix $W$, the resulting parameter estimate is still asymptotically unbiased, albeit not with minimum variance. This leads to the following practical procedure:

1. Compute a preliminary estimate $\hat{\rho}_T^0$ by running GMM optimization according to (10.7) with some initial weighting matrix $W^0$, e.g. $W^0 = 1$.

2. Use the preliminary parameter estimate $\hat{\rho}_T^0$ to obtain an estimate $\hat{S}(\hat{\rho}_T^0)$ of $S$ according to (10.11), plug this into $W = S^{-1}$ and re-run the GMM optimization in (10.7) to obtain a refined parameter estimate $\hat{\rho}_T$.

This procedure is called two-stage GMM. It could be extended to $n$-stage GMM by repeating Step 2 above. However, practical experience shows that this does not generally lead to substantial improvements. Having done this staged optimization, the asymptotic covariance matrix $V$ from (10.8) can be estimated as

$$\hat{V}_T = \left( \hat{D}_T(\hat{\rho}_T)^\prime \hat{S}_T(\hat{\rho}_T)^{-1} \hat{D}_T(\hat{\rho}_T) \right)^{-1}. \tag{10.12}$$

where $\hat{D}_T(.)$ is the empirical counterpart of (10.10). We also note that $\hat{V}_T$ is an estimate for the inverse of the Hessian of the GMM objective
function at the optimum. This is useful since optimization routines like Matlab's \texttt{fmincon} also return a numerical estimate of the Hessian with the results, which saves us some calculations.

In order to make GMM according to (10.7) operational, one has to address a number of issues:

1. Number and type of moment conditions; see Section 10.4.
2. Estimation of weighting matrix; see Section 10.5.
3. Numerical optimization method; see Appendix B.2.
4. Diagnostics to assess goodness of fit; see Section 10.6.

After having treated all these issues in depth, it will turn out that putting GMM to work is all but a simple task, involving the solution of a number of difficult theoretical and numerical problems.

There is a vast literature on GMM and its refinements and applications. The introduction in this section was based on [28] and [62], which are both good theoretical presentations and include also pointers to the relevant original articles. The latter reference also treats constraints on the parameters. A comprehensive account from a more applied viewpoint is provided by [74]. References dealing with specific topics and alternatives are given further below.

### 10.3 Survey of Alternative Methods

In Section 10.2, we have presented GMM as the calibration method of choice without any look at possible alternatives. Indeed, there is a wide range of alternatives, of which Part IV of [74] provides a comprehensive overview, with emphasis on the application to more classical models like Vasicek or CIR. In this section, we only give a brief overview which takes into account the somewhat special structure of our model, that is, the fact
that there is an unobservable stochastic process $X$ underlying the model for the observable quantities. An interesting survey in this context is [47].

The first calibration method usually cited in the literature is *Maximum Likelihood Estimation (MLE)*. Letting $p(y; \rho)$ again denote the stationary density of the observables $Y$, this can be formalized as

$$
\hat{\rho} = \arg \max_{\rho \in \mathcal{R}} \frac{1}{T} \sum_{t=1}^{T} \log p(\tilde{Y}_t; \rho).
$$

(10.13)

However, MLE is not an option in the present context for several reasons. First of all, we do not have an analytically tractable likelihood function for our model. Indeed, unless for very specific parameters, the densities of the interest rates as stated in Section 9.3 are only given as infinite series representations; see [96]. Even finite approximations thereof would be hard to handle numerically. Another reason is that MLE is very sensitive to even slight mis-specifications, and we generally must face the possibility that such mis-specifications are present in our models. Finally, MLE does not allow any freedom in setting optimality conditions or in integrating external judgement, which GMM provides quite naturally through the flexibility in selecting number and kind of moment conditions. The first problem could, at least in principle, be circumvented by quasi-MLE methods or by simulated MLE (see Section 17.5 of [74] for both). However, this does not address the other two problems, and one can also expect considerable numerical problems.

Though we disregard MLE in general for the reasons stated above, it is worthwhile here to make a link between GMM and MLE. Under some regularity conditions for the stationary density $p(y; \rho)$ we have

$$
E_{\rho} \left[ \frac{\partial}{\partial \rho} \log p(Y; \rho) \right] = 0.
$$

(10.14)

The expression in the square brackets is known as the *score vector*, and – as $M = P$ in this case – it can be plugged into classical method of moments estimation, leading to the condition

$$
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \rho} \log p(\tilde{Y}_t; \rho) = 0.
$$
This is exactly the first-order condition for MLE as in (10.13). Moreover, one can show that, in this situation, the matrix $S$ according to (10.11) is the Fisher information according to Definition 10.5. Hence, MLE (or GMM with the score vector) attains the Cramér-Rao bound and is asymptotically efficient according to Definition 10.7. More details are given in [28].

Depending on the underlying model, it may be impossible or overly complicated to compute the theoretical moment functions $g(\rho) = \mathbf{E}_\rho[f(\rho; \mathbf{Y})]$ analytically. In this case, for some proposed value of $\rho$, one can simulate a long trajectory $(\mathbf{Y}_u)_{u=1}^U$ and estimate $g(\rho)$ from this simulated sample. Depending on the complexity of the model, the length $U$ of this simulated sample is in the thousands or even tens of thousands; see Figure 11.1 to get an idea. Moreover, it has to be assured that no transient effects enter the sample. In principle, this brute-force approach, known as *Simulated Method of Moments* (SMM) is always doable, but it is computationally very expensive and severe accuracy problems are to be expected.

A special case of SMM that has become fairly popular is the so-called *Efficient Method of Moments* (EMM). Starting point is GMM with the score vector as above. In many practical situations – including the model presented in this thesis – the score functions may be impossible or overly complicated to calculate. The idea is then to approximate the original law $p(\mathbf{y}; \rho)$ with an auxiliary model $q(\mathbf{y}; \theta)$ that is more tractable. Note that we must have $\dim(\theta) > \dim(\rho)$ for the approximation to work properly. The auxiliary model is fitted to the original observations $(\mathbf{Y}_t)_{t=1}^T$, yielding an auxiliary parameter estimate $\hat{\theta}$. On the other hand, one simulates a long sample $(\hat{\mathbf{Y}}_s)_{s=1}^S$ from the original model and uses this to compute an accurate estimate of the scores of the original model through the scores of the auxiliary model:

$$m(\rho, \hat{\theta}) = \frac{1}{S} \sum_{s=1}^S \frac{\partial}{\partial \theta} \log q(\mathbf{Y}_s; \hat{\theta}).$$

The estimate of the original parameter is then obtained as in normal GMM by

$$\hat{\rho} = \arg\min_{\rho} m(\rho, \hat{\theta})' W m(\rho, \hat{\theta}).$$
Note that the long sample \((Y_s)_{s=1}^S\) has to be re-simulated for each investigated value of \(\rho\), making the method computationally highly expensive. Moreover, EMM estimates are only as good as the auxiliary model. As mentioned, this auxiliary model must have more parameters than the original model, thus exposing the method to the curse of dimensionality. Common practice in EMM is to use a relatively universal auxiliary model, most often the \textit{Semi-Non-Parametric (SNP) model} proposed in [61], the original inventors of EMM. SNP is a Hermite expansion of multivariate Normals, with AR or ARCH structures for the residuals if necessary. SNP requires at least \(n(n+1)(3/2)\) parameters to fit an \(n\)-dimensional time series, which is definitely too much in the situation of interest here. Another caveat is that such an auxiliary model generally introduces properties like arbitrage or negative interest rates through the backdoor. Given that the models considered here were carefully constructed to avoid such features, this is definitely undesirable. For the reasons stated here, we will not use EMM. A sound description of EMM and SNP is given in [61], a critical discussion providing some of the arguments above comes from [47].

\textit{Robust GMM} as described in [108] works, in principle, like the ordinary GMM outlined in Section 10.2. The only difference is that the moment estimators are based on methods from the realm of robust statistics which are insensitive against outliers. The latter are not too much of an issue for interest rates on monthly aggregation; see Section 6.3. For other asset classes (equity, real estate) and lower time aggregation (daily), this can be different. However, there is a fundamental problem to be borne in mind: in historical financial data, there are generally no severe measurement errors. So, each outlier must be considered as an authentic phenomenon. Therefore, if a model driven by Gaussian noise is calibrated to data with enough outliers so that robust calibration is necessary, then this model must be complemented with a generator for such outliers. Otherwise, the scenarios will underestimate the true risk.

A further estimation approach is \textit{Kalman Filtering}. In principle, it fits particularly well into the context of our model as it explicitly assumes a setup with an unobservable state process \(X\) and an observable process \(Y\).
The generic setup in discrete time looks like

\[ Y_t = f(X_t, \epsilon_t; \rho) \quad \text{measurement equation,} \]
\[ X_t = g(X_{t-1}, \eta_t; \rho) \quad \text{system equation,} \]

where \( \epsilon \) and \( \eta \) are Gaussian noise processes and \( \rho \) is the parameter vector of interest. The algorithm works recursively: For time \( t - 1 \), we assume that we have estimates of \( X_{t-1} \), its covariance \( P_{t-1} \) and the parameters \( \hat{\rho}_{t-1} \). Then, there are three steps:

1. Prediction step: find forecasts \( X_{t|t-1} \) and \( P_{t|t-1} \).

2. Update step: plug in the new observation \( \tilde{Y}_t \) to compute refined estimates of \( X_t \) and \( P_t \).

3. Estimation step: use \( X_t \) and \( P_t \) to obtain an updated estimate \( \hat{\rho}_t \).

In the pure Kalman filter, the functions \( f \) and \( g \) must be linear, so it does not apply to the exponential-quadratic model as set up in Chapter 9. However, there exist Extended Kalman Filters, where non-linear functions are approximated by their first-order Taylor expansions, and this can be applied to our model.

Details on Kalman Filtering are given in Sections 18.2 and 18.3 of [74] and further discussions are provided by [47]. An application of Extended Kalman Filtering in a setup quite close to the one in this thesis is presented in [107].

Extended Kalman Filtering can be a viable alternative to GMM, and extensions of [107] to settings relevant in the realm of this thesis are an interesting topic for future research. Preference was given to GMM mainly because of some points of concern: underlying the estimation step is quasi-MLE with all the associated problems as discussed above, moreover, one has to incorporate constraints on the parameters (see Section 9.3), which may have drawbacks.
10.4 Moment Conditions

The problem of the selection of moment conditions actually encompasses two questions, for none of which there is a generally agreed-upon answer:

1. Which moment conditions to use?
2. How many moment conditions to use?

In practice, the most popular and promising approach to selecting moment conditions is pragmatic. One selects the types of moment conditions to be used based on different types of reasonings on the model, the data and the application area:

**Mathematical reasoning** The mathematical structure of the model often induces choices of moment conditions – because certain moments are either particularly simple or impossible to calculate. Also, the model structure may make certain moment conditions particularly important. For instance, in the equilibrium models that we use here, calibrating the unconditional (= equilibrium) expectations is particularly important.

**Economic reasoning** In the realm of interest rates and FX rates, certain economically motivated conditions play a role. There are hard requirements like the absence of arbitrage or positive nominal rates. These generally enter the calibration through constraints, but they can also influence moment conditions. On the other hand, there are various constraints, like the Purchasing Power Parity (PPP) or the expectation hypothesis, that can be translated into moment conditions. In this case, however, it is important to verify whether such hypotheses are actually supported by the data.

**Statistical reasoning** As the purpose of DFA scenario generation is the faithful reproduction of the real-world behavior of the data, this is probably the most important kind of reasoning. Careful statistical analysis identifies the salient features of the data that are then translated into moment conditions. Economic reasoning or the study goals may provide guidance for the analysis, and the mathematical
structure of the model may constrain the feasibility of certain moment conditions.

Focus and goals of study Depending on these issues, certain moment conditions may prove more important than others. For instance, in a study with a short time horizon, the long-term equilibrium of the model is of minor importance, and when tail-based risk measures are used, then it becomes more important to get the volatilities right.

Given that we are usually interested in the entire yield curve, it is important to look at moment conditions for different times to maturity and not just for the short rate. This does not mean that all moment conditions have to be evaluated at all times to maturity, but if statistical evaluation suggests that a feature heavily depends on time to maturity, as e.g. the equilibrium rate, then this moment condition must be evaluated at a representative set of times to maturity.

Concerning the number of moment conditions to use, the conclusion from practical work is that one should not try to determine it explicitly. The important issue is to find a set of sensible moment conditions evaluated at representative times to maturity. This leads implicitly to a reasonable number of moment conditions.

There are many case studies describing such pragmatic approaches to finding moment conditions in GMM. An important one that provided valuable guidance for the models considered here is [88]. More case studies and further references are provided in [74]. Besides this application-oriented approach, there also exist more generic methods for obtaining moment conditions:

One generic way to obtain moment conditions is to use the scores as introduced in (10.14). Using the scores only amounts to doing MLE with all its practical disadvantages. One could complement the scores with other moment conditions to come to some kind of "MLE++", but there remain disadvantages: For the models considered here, the scores are overly complicated or impossible to compute. Moreover, the scores do not have an
intuitive interpretation relative to the data. For reasons outlined there, we also refrain from using EMM described in Section 10.2.

Another, very elegant, way to obtain generic moment conditions is suggested by [64]. Let $\mathbf{X} = (\mathbf{X}_t)_{t \geq 0}$ be a strictly stationary, time-continuous, $d$-dimensional Markov process – as the one defined in Section 9.2 and used in our model – with infinitesimal generator $A$. Then, for every function $f$ for which $(Af)(\mathbf{X}_t)$ is well-defined, we have $\mathbb{E}[(Af)(\mathbf{X}_t)] \equiv 0$ under the stationary distribution of $\mathbf{X}$. Moreover, a similar result holds for the reverse-time version of $\mathbf{X}$. The problem with this approach is in the present setting is that we actually have to compute the infinitesimal generator of the observables $R(t, \tau) = r_\tau(\mathbf{X}_t)$ rather than of the process $\mathbf{X}$ itself, which is difficult.

10.5 The Weighting Matrix

As we have seen in Section 10.2, using the right weighting matrix $W$ is a key to proper GMM estimation. The optimal choice is $W = S^{-1}$, where $S$ is the asymptotic covariance matrix of the moment conditions as specified in (10.9). This matrix depends on the observations as well as on the parameter and must be estimated. To this end, we rewrite (10.9) as

$$S = \lim_{T \to \infty} \mathbb{E} \left[ \frac{1}{T} \left( \sum_{t=1}^{T} \mathbf{m}(\rho_0; \mathbf{Y}_t) \right) \left( \sum_{t=1}^{T} \mathbf{m}(\rho_0; \mathbf{Y}_t)' \right) \right]$$

$$= S_0(\rho_0) + \sum_{j=1}^{\infty} (S_j(\rho_0) + S_j(\rho_0)') , \quad (10.15)$$

where

$$S_j(\rho_0) = \mathbb{E} [\mathbf{m}(\rho_0; \mathbf{Y}_t) \mathbf{m}(\rho_0; \mathbf{Y}_{t+1-j})']$$

is the $j$-th autocovariance matrix of $\mathbf{m}(\rho_0; \mathbf{Y}_t)$. There are various ways for estimating $S$, depending on whether serial correlation or heteroskedasticity have to be dealt with. A generic form of the estimator is given by

$$\hat{S} = \hat{S}_0 + \sum_{j=1}^{J} w(j) \left( \hat{S}_j + \hat{S}_j' \right) , \quad (10.16)$$
where
\[ \hat{S}_j = \frac{T}{T - P} \frac{1}{T} \sum_{t=j+1}^{T} m(\hat{\rho}_t; \tilde{Y}_t) m(\hat{\rho}_t; \tilde{Y}_{t-j})' . \]

The term \( T/(T - P) \) is a small sample correction, where \( P \) is again the number of estimated parameters. The estimator requires some previous estimate \( \hat{\rho} \) of the parameters of interest. This estimate usually comes from the first stage of the two-stage estimation procedure described in Section 10.2. \( J \) is a lag truncation parameter that excludes estimates from higher lags where the autocovariance is close to zero, \( w(j) \) is a kernel weight optimizing the estimator.

If the values \( m(\hat{\rho}, \tilde{Y}_t) \) of the moment function evaluated at the different observations are uncorrelated and homoskedastic, then one can set \( J = 0 \), and (10.16) boils down to the usual estimator for the contemporaneous covariance matrix. In general, however, we must be aware of autocorrelation and heteroskedasticity. For this case, following the suggestions of [28], we use the Newey-West kernel given by
\[ w(j) = \frac{J - j}{J} \quad \text{for} \quad 0 \leq j \leq J \quad \text{and} \quad w(j) \equiv 0 \quad \text{otherwise}, \quad (10.17) \]
and a lag truncation factor of \( J = \lceil T^{-1/3} \rceil \). Alternative weighting schemes are suggested e.g. by [61]. A more refined estimation of the lag truncation factor \( J \) could be based on fitting a VAR model and using the resulting diagnostic tests.

### 10.6 Diagnostics

Besides providing estimates for the parameters themselves, the GMM estimation procedure is also supposed to provide information on how well the calibrated model fits the data. As already observed in Section 10.2, we have
\[ T m(\hat{\rho}_T, \tilde{Y})' W(\hat{\rho}_T, \tilde{Y}) m(\hat{\rho}_T, \tilde{Y}) \overset{\text{as}}{\sim} \chi^2_{M-P}, \quad (10.18) \]
where \( \hat{\rho}_T \) is the GMM estimate for the parameters according to (10.7), and \( W(\hat{\rho}_T, \tilde{Y}) \) is the optimal weighting matrix according to (10.11), estimated with the GMM estimate for the parameters. \( M \) is the number of
10.7. Fitting the Initial Yield Curve

moment conditions, and \( P \) is the number of parameters, where we must have \( M > P \). The test induced by Formula 10.18 is commonly known as the GMM Omnibus Test and it provides information on whether the calibrated model as a whole fits the data. In practical work, it turned out not to be very selective.

In order to obtain inference for each single parameter, we can use the fact that the parameter estimate is asymptotically normally distributed according to Formula 10.8 with covariance matrix according to (10.9):

\[
\sqrt{T} (\hat{\rho}_T - \rho_0) \overset{\text{as}}{\sim} N(0, V),
\]

where \( V = (D_0'WD_0)^{-1}D_0'WSWD_0(D_0'WD_0)^{-1} \).

If the optimal weighting matrix \( W = S^{-1} \) according to (10.11) is used, then (10.12) provides an estimate \( \hat{V}_T(\hat{\rho}_T) \) of \( V \), namely

\[
\hat{V}_T(\hat{\rho}_T) = (\hat{D}_T(\hat{\rho}_T)'\hat{S}_T(\hat{\rho}_T)^{-1}\hat{D}_T(\hat{\rho}_T))^{-1}.
\]

Form this, for each single parameter \( \rho^i \), one can obtain an asymptotic \( t \)-test for the null hypothesis that the true value \( \rho_0^i \) is actually zero. More importantly, one can also obtain a \( \alpha \)-confidence interval:

\[
\rho_0^i \in \hat{\rho}^i_T \pm t_{1-\frac{\alpha}{2}};T-1 \sqrt{(\hat{V}_T(\hat{\rho}_T)))ii}/T, \tag{10.19}
\]

where \( t_{q,T-1} \) denotes the \( q \)-quantile of Student's \( t \)-distribution with \( T - 1 \) degrees of freedom. Given the usual values for \( T \) (> 100), the \( t \)-distribution can usually be replaced by the normal one. See [28] or [74] for more information.

10.7 Fitting the Initial Yield Curve

The GMM method as presented above can fit a model to a stationary stretch of historical observations. In other words: the estimated parameter vector \( \hat{\rho}_T \) obtained from GMM determines the behavior of the model in equilibrium. However, in DFA scenario generation as well as in other applications as e.g. derivative valuation, we need to start from a distinct
initial state at some distinct time point $t_0$, i.e. the model must also be calibrated to this initial state. We also note that the initial value $x_0$ of the driver process can never be determined using GMM, as it does not belong to the specification of the equilibrium of the model.

The approach here is that we maintain the parameter estimate $\hat{\rho}_T$ from GMM without any changes – so as not to lessen the content of historical information in it – and only optimize over the initial value $x_0$ of the driver process to adapt the model to the initial conditions. Formally, let $Y(x; \hat{\rho}_T) \in \mathbb{R}^D$ be the vector of observable yields and FX rates as defined by the model as a function of the value $x$ of the driver process, and let $\bar{Y}_0 \in \mathbb{R}^D$ be their observed values at the initial time point. Then we can simply determine $x_0$ by non-linear least squares, i.e.

$$x_0 = \min_{x \in \mathbb{R}^d} \frac{1}{2} \left\| W \left( Y(x; \hat{\rho}_T) - \bar{Y}_0 \right) \right\|_2^2,$$

where $W$ is a diagonal weighting matrix allowing to put more or less emphasis on the fit of certain initial conditions. If the driver process is of the Ornstein-Uhlenbeck type according to Definition 9.1, then no constraints need to be imposed on $x_0$. The implementation uses the Matlab function \textit{fmincon} as described in Section B.2.

10.8 Case Studies

10.8.1 Single-Currency Case – USD

In this section, we calibrate a two-factor version of the exponential-quadratic model as introduced in Section 9.3 to a subset of the data introduced in Section 6.3, specifically the USD yield curve data from May 1991 up to December 2003, corresponding to 152 monthly observations.

With the empirical facts from Section 6.3 and their theoretical counterparts for the exponential-quadratic model as given in Section 9.4, we can now set up a GMM calibration. The terminology follows the one given in Section 9.3. Motivated by the PCA mentioned above, we opt for a
two-dimensional driver process, which is at the lower end of the sensible range. In the mean reversion matrix \( \Gamma \), we explicitly allow for non-zero off-diagonal elements. The dependence matrix \( Q \), however, is restricted to diagonal form following Remark 9.5. The discount factor \( \alpha \) is fully specified through the other parameters by (9.14) and, therefore, not an element of the calibration. This results in a parameter vector \( \rho \) with \( P = 8 \) elements and assignments

\[
\begin{align*}
\rho_1 &= \Gamma_{11}, & \rho_2 &= \Gamma_{12}, & \rho_3 &= \Gamma_{21}, & \rho_4 &= \Gamma_{22}, \\
\rho_5 &= Q_{11}, & \rho_6 &= Q_{22}, & \rho_7 &= c_1, & \rho_8 &= c_2.
\end{align*}
\]

In order for the calibrated model to have the desirable properties of stationarity, positive rates, absence of arbitrage and complete markets, we have to impose a number of constraints on the above parameters, irrespective of the moment conditions. For the mean reversion matrix \( \Gamma \), this is

\[
0 < \Gamma_{11} \leq 1, \quad \Gamma_{12} \leq 0, \quad \Gamma_{22} \leq \Gamma_{11},
\]

\[
0 < \Gamma_{22} \leq 1, \quad \Gamma_{21} \leq 0, \quad \text{Re}(\lambda_i) > 0, \quad i = 1, 2,
\]

where \( \lambda_i, i = 1, 2, \) are the eigenvalues of \( \Gamma \). The constraint \( \Gamma_{22} \leq \Gamma_{11} \) is for identification purposes only. The other ones assure that the driver process has the desired stationarity properties; see Section 9.2, in particular Proposition 9.2. It is easily verified that the eigenvalues of \( \Gamma \) are always real if the other constraints are fulfilled. The constraints for the matrix \( Q \) are

\[
Q_1 > 0, \quad Q_2 > 0, \quad \text{Re}(\nu_i) > 0, \quad i = 1, 2,
\]

where \( S \) is the symmetric matrix defined by (9.13) and \( \nu_i, i = 1, 2, \) are its eigenvalues. The eigenvalue constraint thus assures positive definiteness. If \( \Gamma \) is diagonal, then the eigenvalue constraint simplifies to the form \( Q_i \leq 2\Gamma_{ii} \) for \( i = 1, 2 \). The positivity constraint on \( Q_1 \) and \( Q_2 \) assures positive definiteness of \( Q \) and a complete market. These constraints can easily be mapped to the arguments of the \textit{fmincon} function of Matlab; see Section B.2. The non-linear eigenvalue constraints are somewhat annoying as they hamper the numerical stability of the optimization, but practical experience showed that the parameter values go completely astray when these constraints are dumped.
Moment conditions are selected based on the intuition obtained from the statistical evaluations in Chapter 6. First of all, we have to get the mean level of the yield curve right. To this end, we fit the unconditional expectation of the yield as given by (9.31) to its empirical counterpart for a representative cross-section of maturities, i.e.

\[10 \cdot E_{\rho}[R(X_t, \tau_i)], \quad \text{where } \tau_i = 3m, 1y, 3y, 5y, 10y, 15y \text{ and } 20y.\]

Contemporaneous dependence is accounted for by fitting the theoretical correlation coefficient to its empirical counterpart, i.e.

\[
\frac{\text{Cov}_\rho[R(X_t, \tau_0), R(X_t, \tau_i)]}{\sqrt{\text{Var}_\rho[R(X_t, \tau_0)] \cdot \text{Var}_\rho[R(X_t, \tau_i)]}},
\]

where \(\tau_0 = 3m\) and \(\tau_i = 1y, 3y, 5y, 10y, 15y \text{ and } 20y.\). Theoretical covariance and variances are computed according to Equations 9.33 and 9.32, respectively. Given the very high persistence of the interest rate time series (see Section 6.3), we can reflect the conditional volatility by the unconditional variance of lagged differences, i.e.

\[
\frac{\text{Var}_\rho[R(X_{t+h_i}, \tau) - R(X_t, \tau)]}{h_i \cdot \text{Var}_\rho[R(X_t, \tau)]},
\]

where the theoretical variances are given by (9.35) and 9.32. Figure 6.3 suggests that letting \(\tau = 5y\) and \(h_i = 1m, 6m, 1y \text{ and } 2y\) captures well the hump-shaped volatility without having to sacrifice too much of the scarce data.

The scaling with the constant factor and with the variances is for a practical purpose: It turned out that the optimization is only numerically stable if all the moment conditions are approximately equal in location and scale, and the scale should, moreover, be sufficiently higher than the accuracy of the optimizer. It should not be left unmentioned that a good deal of tedious fine tuning was necessary to get the optimization running smoothly.

With the settings just introduced and the Matlab implementation, constrained GMM estimations as described in Section 10.2 were carried out. Specifically, two-step GMM was used: the first step with weighting matrix
Figure 10.1: Comparison of theoretical moment properties and their empirical counterparts for the three parameter sets given in Table 10.1 obtained from the GMM calibration of the single-currency model according to Section 10.8.1. The stars represent the empirical values, whereas the lines represent the theoretical values of the model under each one of the parameter sets given in Table 10.1: solid = diagonal model, dashed = symmetric model, dotted = general model. One sees, in particular, that the calibrated model does not reproduce the empirically observed hump-shaped volatility structure (SW panel), and the contemporaneous correlation structure is not very well reproduced either (NE). Mean (NW) and initial (SE) yield curves, on the other hand, fit fairly well.
Chapter 10. Calibration of Models

Table 10.1: Results of the GMM calibration of the single-currency model according to Section 10.8.1 for three configurations of the mean reversion matrix \( \Gamma \): diagonal, symmetric and general. The upper panel shows the estimates together with a 95% confidence band obtained from the t-statistic in (10.19). The middle panel shows the estimates for \( x_0 \) obtained according to (10.20). The lower panel shows the results of the GMM Omnibus Test derived from (10.18).

\[ W = 1 \] and the second step with the optimal weighting matrix according to the Newey-West estimator given in Section 10.5, and diagnostics according to Section 10.6 were computed. The estimation was done for three settings of the mean reversion matrix \( \Gamma \):

**Diagonal** \( \Gamma_{12} = \Gamma_{21} = 0 \). Corresponds to a driver process with two completely decoupled components. This is interesting statistically as the base case. Moreover, the matrices \( \Gamma, Q \) and \( S \) according to Section 9.3 are all diagonal, so that the related eigenvalue constraints can be translated into linear inequality constraints which leads to improved speed and accuracy of the constrained numerical optimization.

**Symmetric** \( \Gamma_{12} = \Gamma_{21} \). This is mainly interesting for numerical reasons, as the covariance matrix \( V(\tau) \) can be computed in closed form instead of numerically by using Proposition 9.3.
General $\Gamma_{12}$ and $\Gamma_{21}$ can take general (non-positive) values. This is the most flexible version of the model, and the computationally most expensive, as $V(\tau)$ must be computed numerically according to Proposition 9.2.

The results are summarized in Table 10.1. At first glance, they look fairly plausible; in particular, they are all in the interior of the constrained parameter space. The moderate to slow mean reversion of the driver process (expressed by $\Gamma_{11}$ and $\Gamma_{22}$) corresponds to what is found in many other models as well; see [74] for a survey. In terms of the $\chi^2$-statistic, all versions of the model can be reconciled with the data.

Some of the $t$-confidence intervals for the symmetric and general cases do not look very credible. This is due to the fact that the $t$-statistics are computed from a purely numerical estimate of the Hessian matrix of the target function; see the discussion around (10.12). In the mentioned cases – due to the presence of the non-linear constraints – this estimate tends to be numerically unstable; see [97]. A conclusion of this is that better diagnostic statistics are needed, for instance the GMM Bootstrap as suggested in [72].

Finally, using the GMM parameter estimates, the initial value for the driver process was estimated based on (10.20), using observed yields for 1m, 3m, 6m and 1y, ..., 20y. The results are also reported in Table 10.1.

Figure 10.1 shows a comparison of the theoretical features of the three calibrated model settings (diagonal, symmetric, general) with their empirical counterparts. Features include the ones used for the GMM optimization as well as the fit to the initial yield curve. The fits are generally not very good and, interestingly enough, the simplest model, i.e. the diagonal one, provides the best fit. In particular, the hump in the conditional dynamics is not very well captured. The fits shown here are not as good as the ones in [88]. It should, however, be pointed out that the study here fits a model with fewer parameters to a much wider range of maturities and a shorter data set.
Figure 10.2: Summary statistics over time at several maturities of 5000 US interest rate scenarios produced by the exponential-quadratic model calibrated with the "diagonal" parameter set from Table 10.1, but without the time-inhomogeneous overlay according to Section 8.4. Solid line = scenario mean, dashed lines = 25% and 75% quantiles, dotted lines = 5% and 95% quantiles. The stars indicate the empirical long-term average. One can clearly see the asymmetric distribution of the interest rates and the convergence towards the long-term equilibrium.
Figure 10.3: The same summary statistics as in Figure 10.2, but this time with the time-inhomogeneous overlay according to Section 8.4. The effect of the latter on the mean interest rate is clearly visible.
Figure 10.4: Comparison of one-step-ahead point forecasts for several US rates with the actual realizations for the 120 months from January 1994 to December 2003. Forecasts were generated according to the "diagonal" parametrisation of the model as given in Table 10.1. The solid diagonal line represents equality of forecast and realization; the dash-dotted line is a least-squares regression line through the data points. There is at least a small bias for all times to maturity, but the 1-year rate is particularly problematic since the bias is highest for the range in which values are situated most frequently.
Figure 10.2 shows some Summary statistics over time at several maturities of 5000 interest rate scenarios produced by the exponential-quadratic model calibrated with the "diagonal" parameter set from Table 10.1. One can see there that the scenarios deterministically start at the initial yield curve, and that their average approaches a long-term equilibrium within a few years. Also the dispersion around the mean stabilizes after some time, indicating the reversion of the model towards its stationary distribution. Other salient features are the marked asymmetry of the rates for all times and maturities and the decreased dispersion with increasing maturity. Negative rates entirely absent in the scenario set.

Figure 10.3 was produced in the same way as Figure 10.2, but this time with the time-inhomogeneous extension according to Section 8.4 superimposed on the scenarios. One can see that the stubborn reversion towards the equilibrium is not present anymore. The inhomogeneous correction has quite a heavy impact mainly on the shorter-term rates, diminishing for longer-term rates and for the farther future – as one would expect.

A further evaluation is shown in Figure 10.4. The diagonal model according to Table 10.1 was fitted to each historical yield curve and then used to generate one-month-ahead point forecasts. These forecasts are then compared to the actual realizations in order to assess the forecasting accuracy. The pictures for some key rates show that the model does not produce completely unbiased forecasts, but the deviations are not dramatic either, particularly in view of the intended purpose.

In Figure 10.5, we explore the range of possible yield curve shapes under the "diagonal" version of the model according to Table 10.1. The figure shows that even this simplest version of the model is capable of producing upward-sloping and downward-sloping as well as other (humped) shapes of the yield curve, as it is observed in real-world data; see Section 6.3. The related probabilities of the model (75.5% upward, 4.4% downward and 20.1% others) correspond approximately with the observed frequencies in the data (81.6%, 3.5% and 14.9%, respectively).
Figure 10.5: Survey of the attainable yield curve shapes as a function of the value of the driver process $X = (X_1, X_2)'$ for the "diagonal" parameterization of the two-factor model according to Table 10.1. The hyperbolic lines delimit the areas giving rise to upward-sloping, downward-sloping and other shapes, respectively; the ellipses represent the density of the stationary distribution. The probabilities stated on top are with respect to the latter distribution. For further discussions, see p. 153.
The evaluations shown here represent first attempts of backtests of the calibrated model. They evaluate how well the model fits the data and how well it can forecast the future. Of course, these few punctual evaluations are not sufficient for a full validation of the model for practical use. Therefore, the topic of validation will be dealt with in more detail in Chapter 11.

### 10.8.2 Double-Currency Case – CHF and USD

Based on the experience with the single-currency calibration as described in Section 10.8.1, we are now going to calibrate a double-currency version of the exponential-quadratic model as introduced in Chapter 9 to the data presented in Section 6.3. The terminology follows the one set forth in Section 9.3. We select the USD as the pivot currency and denote it by the superscript (1), whereas the CHF is the secondary currency denoted by (2). The general procedure is the same as in Section 10.8.1.

Given the results of the Principal Component Analysis in Section 6.3, we select a four-factor version of the model. This should suffice to explain 95% of the variability of the combined data sample. Following Remark 9.5, the dependence matrices $Q^{(k)}$, $k = 1, 2$, are restricted to diagonal. Moreover, the parameters $\alpha^{(k)}$ are fully determined by the other parameters; see (9.14). This results in the following parameter vector $\rho$:

\begin{align*}
&\rho_1 = \Gamma_{11}, \quad \rho_9 = \Gamma_{31}, \quad \rho_{17} = Q_{11}^{(1)}, \quad \rho_{25} = Q_{11}^{(2)}, \\
&\rho_2 = \Gamma_{12}, \quad \rho_{10} = \Gamma_{32}, \quad \rho_{18} = Q_{22}^{(1)}, \quad \rho_{26} = Q_{22}^{(2)}, \\
&\rho_3 = \Gamma_{13}, \quad \rho_{11} = \Gamma_{33}, \quad \rho_{19} = Q_{33}^{(1)}, \quad \rho_{27} = Q_{33}^{(2)}, \\
&\rho_4 = \Gamma_{14}, \quad \rho_{12} = \Gamma_{34}, \quad \rho_{20} = Q_{44}^{(1)}, \quad \rho_{28} = Q_{44}^{(2)}, \\
&\rho_5 = \Gamma_{21}, \quad \rho_{13} = \Gamma_{41}, \quad \rho_{21} = c_1^{(1)}, \quad \rho_{29} = c_1^{(2)}, \\
&\rho_6 = \Gamma_{22}, \quad \rho_{14} = \Gamma_{42}, \quad \rho_{22} = c_2^{(1)}, \quad \rho_{30} = c_2^{(2)}, \\
&\rho_7 = \Gamma_{23}, \quad \rho_{15} = \Gamma_{43}, \quad \rho_{23} = c_3^{(1)}, \quad \rho_{31} = c_3^{(2)}, \\
&\rho_8 = \Gamma_{24}, \quad \rho_{16} = \Gamma_{44}, \quad \rho_{24} = c_4^{(1)}, \quad \rho_{32} = c_4^{(2)}.
\end{align*}

If we allow a fully unstructured mean reversion matrix $\Gamma$, we have $P = 32$
parameters. However, in order to take advantage of the better numerical properties, we will mainly deal with submodels where the mean reversion matrix \( \Gamma \) is symmetric. Specifically, we explore a diagonal, a tri-diagonal and a fully symmetric version. In order for the models to have the desired properties, we have to impose the constraints specified in Section 9.3. For the mean reversion matrix \( \Gamma \), they read as

\[
0 < \Gamma_{ii} \leq 1, \quad i = 1, \ldots, 4, \\
\Gamma_{ij} \leq 0, \quad i \neq j, \\
\text{Re}(\lambda_i) > 0, \quad i = 1, \ldots, 4,
\]

where \( \lambda_i, i = 1, \ldots 4 \), are the eigenvalue of \( \Gamma \). The constraints on the dependence matrices \( Q^{(k)} \) are

\[
Q_{ii}^{(k)} > 0, \quad k = 1, 2, i = 1, \ldots, 4, \\
\text{Re}(\nu_i^{(k)}) > 0, \quad k = 1, 2, i = 1, \ldots, 4,
\]

where \( \nu_i^{(k)}, i = 1, \ldots, 4 \), are the eigenvalues of \( S^{(k)}, k = 1, 2 \), respectively, the symmetric matrices from (9.13).

Moment conditions are selected based on the experience with the single-currency model in Section 10.8.1 and on the statistical evaluations in Section 6.3. The first batch of conditions is for the unconditional expectation:

\[
10 \cdot \mathbb{E}_\rho \left[ R^{(k)}(X_t, \tau_i) \right],
\]

where \( k = 1, 2 \) and \( \tau_i = 3m, 1y, 3y, 5y, 10y, 15y \) and 20y. The theoretical expectation is according to (9.31). The contemporaneous dependence within each economy is accounted for by

\[
\frac{\text{Cov}_\rho \left[ R^{(k)}(X_t, \tau_0), R^{(k)}(X_t, \tau_i) \right]}{\sqrt{\text{Var}_\rho \left[ R^{(k)}(X_t, \tau_0) \right] \text{Var}_\rho \left[ R^{(k)}(X_t, \tau_i) \right]}},
\]

where \( k = 1, 2, \tau_0 = 3m \) and \( \tau_i = 1y, 3y, 5y, 10y, 15y \) and 20y. Theoretical variances and covariances follow from Equations 9.33 and 9.32. The conditional dynamics are again given by

\[
\frac{\text{Var}_\rho \left[ R^{(k)}(X_{t+h_i}, \tau) - R^{(k)}(X_t, \tau) \right]}{h_i \text{Var}_\rho \left[ R^{(k)}(X_t, \tau) \right]},
\]
where $k = 1, 2, \tau_j = 5y$ and $h = 1m, 6m, 1y, 2y$. The conditions up to now correspond to those for the single-currency case. In order to account for cross-economy dependence, we add the conditions

$$\frac{\text{Cov}_\rho \left[ R^{(1)}(X_t, \tau_i), R^{(2)}(X_t, \tau_i) \right]}{\sqrt{\text{Var}_\rho \left[ R^{(1)}(X_t, \tau_i) \right] \text{Var}_\rho \left[ R^{(2)}(X_t, \tau_i) \right]}}$$

where $\tau_i = 3m, 1y, 3y, 5y, 10y, 15y \text{ and } 20y$. The covariance term is given in (9.33). The FX rate is explicitly accounted for by

$$10 \cdot \text{E}_\rho [c_{12}(t, t + \Delta t)] \text{ and } \text{Var}_\rho [c_{12}(t, t + \Delta t)]$$

according to Equations 9.41 and 9.42, where $\Delta t = 1m$. This makes a total of 43 moment conditions for $P \leq 32$ parameters. Parameter estimation was carried out by the GMM method as described in the previous sections of this chapter, and the results are shown in Table 10.2. The results of the classical optimization procedure (fmincon of Matlab) turned out to depend heavily on the chosen initial values; see Appendix B for an illustration. In order to overcome this problem, a global search for suitable initial values was carried out by using the genetic algorithm as described in Appendix B. Table 10.2 only presents results for simplified setups with symmetric mean reversion matrix $\Gamma$, so that the covariance matrix $V(\tau)$, according to Proposition 9.2, can be computed in the closed form according to Proposition 9.3. In these cases, the execution of fmincon took less than two minutes, and the previous global search took about 30 minutes. For general mean reversion matrices, where $V(\tau)$ must be computed iteratively, the standard optimization only took more than an hour on a 2.8 GHz Mobile Pentium machine with 512 MB of RAM and was plagued with heavy numerical problems.

The results for the diagonal version of the model in Table 10.2 are easiest to interpret intuitively. From the four independent Ornstein-Uhlenbeck processes, two have a very fast mean reversion, and the other two have a rather slow one. This corresponds to the conventional wisdom that interest rates are governed by factors with both slow and fast mean reversion. For the USD (economy 1), the interest rates are mainly governed by two factors with only weak dependence on the other ones. For CHF (economy 2), the interest rates mainly depend on three factors, rather than only two.
### Table 10.2: Results (analogous to Table 10.1) of the GMM calibration of the double-currency model according to Section 10.8.2 for three configurations of the mean-reversion matrix: diagonal, tri-diagonal symmetric and general symmetric.
This asymmetric dependence between the two economies makes sense intuitively: while the Swiss economy clearly depends on the US economy, the converse is certainly only the case to a very limited extent. For the tri-diagonal case, we also see the mixture between slow and fast mean reversion. Moreover, at least some of the admitted dependencies between the components of the driver process do, indeed, obtain non-zero values. The patterns of dependence of the interest rates on the driver process are not significantly different from the diagonal case. The same is true for the general symmetric setup. The $t$-confidence intervals according to Section 10.6 are rather wide in some cases, which has numerical reasons and emphasizes, once more, the need for better diagnostics.

Figure 10.6 shows the comparison of the theoretical features of the calibrated models according to Table 10.2 with their empirical counterparts. The goodness of fit is approximately equal to the one for the single-currency model shown in Figure 10.1. Figure 10.7 shows summary statistics over time for 5000 scenarios generated with the parameters for the diagonal case in Table 10.2. As in the single-currency case, we can clearly see the features of mean reversion, asymmetric distribution and non-negative rates. The impact of the time-inhomogeneous correction is also visible, but not as strong as in the single-currency case (Figure 10.3). Figure 10.8 provides further insight into the distribution of the interest rates. The asymmetry of the distribution is clearly confirmed. The scenarios are not only concentrated on the positive half-axis, but they also have a lower bound greater than zero, which is a disadvantage of the model. The GPD estimates of the tail coefficient $\xi$ (see [99]) are somewhat intriguing. While the plot for the one-year rate (top right) clearly indicates no heavy tails, the situation for the five-year and ten-year rates is less clear, in that some of the possible estimates are positive. However, the most stable part of the shape plots is always around zero, so that we can also assume absence of heavy tails here, as we would expect it for a quadratic form of normal random variables.

As the time-inhomogeneous correction function according to Section 8.4 has such a considerable impact, it is worthwhile to check whether it is actually admissible, that is, whether it preserves the supermartingale prop-
Figure 10.6: Comparison of theoretical moment properties and their empirical counterparts for the three parameter sets given in Table 10.2 obtained from the GMM calibration of the double currency model according to Section 10.8.2. The stars represent the empirical values, whereas the lines represent the theoretical values of the model under each one of the parameter sets given in Table 10.2: solid = diagonal model, dashed = tri-diagonal model, dotted = general symmetric model. As in the single-currency case, we notice that the calibrated model does not reproduce the hump-shaped volatility structure; see p. 159 for further discussions.
Figure 10.7: Summary statistics over time (as in Figure 10.2) for 5000 interest rate scenarios generated with the "diagonal" parameter set according to Table 10.2, including the time-inhomogeneous overlay according to Section 8.4. Asymmetric distribution of interest rates, reversion towards the equilibrium and the impact of the time-inhomogeneous overlay on the mean rates are clearly visible.
Figure 10.8: Evaluations of the distribution of 10000 scenarios of CHF interest rates generated with the "diagonal" parameter set from Table 10.2. The left panels show the empirical densities of rates for different times to maturity after one year (dotted), five years (dashed) and 10 years (solid). The right panels show the GPD estimates of the tail coefficient $\xi$ as a function of the threshold (solid) with the 95% confidence intervals (dotted) according to [99]. For further discussions see p. 159.
Figure 10.9: Evaluation of the time-inhomogeneous correction function by using condition (8.8). The white surfaces represent $h(t + \tau)/h(t)$ and the gray surfaces represent the minimum attainable values of $1/P(t, t + \tau)$, where $t$ is time and $\tau$ is time to maturity, both in years. Since the white surfaces do not intersect with the gray one, condition (8.8) is satisfied and the model with the time-inhomogeneous overlay still produces positive yield curves without arbitrage.

Figure 10.10 shows a comparison of point forecasts against actual realisations analogous to Figure 10.4, but with overlapping three-month-ahead forecasts instead of one-month-ahead forecasts. Again, we can see some bias in the forecasts, but this bias is not very high, and it attains its minimum in those regions where most of the actual values are located.
Figure 10.10: Comparison of overlapping three-month ahead forecasts for several rates with the actual realizations for the 200 months from January 1988 to September 2004. Forecasts were generated according to the "diagonal" parameter set in Table 10.2 The solid diagonal line represents equality; the dash-dotted line is the least squares regression line through the data points. The $R^2$ statistics of these regressions are stated in the title lines of the panels. A discussion is provided on p. 163.
All in all, the goodness of fit of the model to the data is far from being perfect, but it should be sufficient for the purpose of DFA scenario generation where the emphasis is mainly on exploring the range of possible outcomes. Further tests of the calibrated model will be presented in Section 11.5.2.
Seite Leer / Blank leaf
Chapter 11

Validation of Models

11.1 Problem Statement

This chapter provides methods for backtesting a calibrated model. Backtesting is the ultimate assessment whether a calibrated model is able to produce scenarios that are representative for the possible future behaviour of the risk factors modelled. Therefore, backtesting is the mandatory last step before a calibrated model can be used for DFA scenario generation or related tasks.

Let us recall the situation. We consider a discrete time scale $t \in \mathbb{Z}$ and some random variables of interest $\{Y_t\}_{t=1}^T$ for which we have a stochastic model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$. For instance, $Y_t$ can be a cross-section of interest and FX rates at time $t$, and $\mathcal{P}$ might be the class of exponential quadratic models as introduced in Section 9.3. Assuming that $t$ is the current time, we use a calibration sample $Y = \{Y_t, Y_{t-1}, ..., Y_{t-s}\}$ and a calibration method $\delta : Y \mapsto g(\theta)$ to select one particular model $P_\theta$ from the model class $\mathcal{P}$. For instance, the calibration method $\delta$ might be the constrained GMM as presented in Section 10.2.
The calibration as such already provides us with some backtesting. First, the model is an optimal choice with respect to some carefully chosen target function. Moreover, calibration methods usually allow for the computation of certain diagnostic statistics to assess whether the entire model or certain parameters can be reconciled with the data. For the case of GMM, some diagnostics of this kind are presented in Section 10.6. Others, like AIC or entropy, could be added. On a less formal note, one can compare theoretical features of the model $P_\theta$ with empirical features of the calibration sample $Y$. Some examples of this kind are shown in Section 10.8.1. The common feature of all these evaluations is that they assess how well the model fits the data used for its calibration. This is also called in-sample backtesting, and it measures the goodness-of-fit or the explanatory power of the model.

Whenever working with a stochastic model, a decent goodness-of-fit is necessary. But in the context of DFA, it is not sufficient because DFA is about forecasting the future, i.e. exploring the range of possible future outcomes of $Y_{t+k}$ of the modelled variables for $k > 0$. Hence, the model $P_\theta$ must also prove its suitability under this point of view. The validation of the model regarding its capability to capture outcomes beyond the data to which it was calibrated is called out-of-sample backtesting.

Before proceeding further, we introduce a few notions. The selected stochastic model $P_\theta$ and the set $\mathcal{F}_t$ of information available up to and including current time $t$ are used to make forecasts about the behaviour of $Y_{t+k}$ at the future time $t + k$, where $k > 0$ is the forecast horizon. If $k = 1$, this is called a single-step forecast, if $k > 1$, we have a multi-step forecast. If simultaneous forecasts for several $k_i > 0$ are made at $t$, then we have a multi-horizon forecast. The more important distinction is, however, between the different types of forecasted quantities:

1. **Point forecasts** project a single numerical quantity representing the behaviour of $Y_{t+k}$, usually the conditional expectation $E^\theta[Y_{t+k}|\mathcal{F}_t]$.

2. **Interval forecasts** project an interval $[l^\theta(t + k, \mathcal{F}_t), u^\theta(t + k, \mathcal{F}_t)]$ in which the value of $Y_{t+k}$ is expected to lie with some probability $p$. The most popular interval forecast is the Value-at-Risk, with the
11.1. Problem Statement

forecast interval \((-\infty, q_p(Y_{t+k})]\), where \(q_p(Y_{t+k})\) is the projected \(p\)-quantile of \(Y_{t+k}\).

3. **Distribution forecasts** project the full distribution function \(F_{t+k}^\theta(y|\mathcal{F}_t)\) of \(Y_{t+k}\), or the related density \(f_{t+k}^\theta(y|\mathcal{F}_t)\) if applicable. A special case is if only a part of the distribution is forecasted, usually the tail beyond some threshold. This is relevant if projections of the Expected Shortfall are to be made, for instance by using the Generalised Pareto Distribution (GPD) above a high threshold; see [52].

Point forecasts are the most classical and most frequent application of forecasting and there is a huge related literature in the realm of econometrics, and there exist well-established methods for out-of-sample backtesting. A standard reference is [35]. Interval forecasts have gained a lot of importance in recent years due to the emergence of Value-at-Risk in the financial industry. Though less well researched than for point forecasts, there is also established methodology for backtesting interval forecasts, with [33] being a good reference. The case applying to DFA scenario generation, however, is the forecasting of full (conditional) distributions, as is easily seen from stylized DFA setups as given in [14]. The problem of backtesting distribution or density forecasts is much less researched, and there is essentially one viable approach, namely the **Probability Integral Transform (PIT)** explored in Section 11.2.

The **Probability Integral Transform (PIT)** is – to my knowledge – the only viable generic method for the out-of-sample backtesting of distribution forecasts. The PIT states that if we plug the actual realizations \(\tilde{y}_{t+h}\) into the correct conditional distributions \(F_{t+h}(.|\mathcal{F}_t)\), then the resulting sequence \(\{F_{t+h}(\tilde{y}_{t+h}|\mathcal{F}_t)\}_{t=1}^{T}\) is iid \(U(0,1)\). This is a very classical result, probably first stated by Paul Lévy in 1937 for discrete probability spaces [89]. In 1952, Murray Rosenblatt extended Lévy's result to more general settings; see [109]. Therefore, the PIT is also referred to as the **Rosenblatt Transform** in a part of the existing literature. Afterwards, the PIT was used occasionally in pure mathematical statistics and in miscellaneous applications; see [112] for a thorough formal treatment and pointers to early applications. The use of the PIT for backtesting financial models is more recent. The foundations were laid by Francis Diebold and coauthors: the
basic univariate case is introduced in [45], while [46] extends the PIT to a multivariate setting and provides a method for improving density forecasts by exploiting remaining irregularities in the PIT residuals. Since backtesting in general, and backtesting of distribution forecasts in particular, is not a very popular theme in finance, the fundamental ground-laying papers by Diebold and coauthors are only followed by few others, and most of these are only of limited interest to the pragmatic practitioner. This is because Diebold's more academic followers generally take the fact that the $F_{t+h}(\bar{y}_{t+h}|F_t)$'s must be iid uniform under the null hypothesis as a starting point and then develop ever more sophisticated tests for this null hypothesis, with power against ever more fancy alternatives. In practice, however, one is not so much interested in abstract global statements, but rather in information about specific aspects that the model is not able to cover. Also interesting are statements on the PIT in situations with overlapping forecast intervals and multiple forecast horizons. These issues will be dealt with in the sequel.

We note that there is quite some interest in the development of methods for the out-of-sample validation of distribution forecasts. With the advent of Solvency II, there will also be a need for internal models in non-life insurance, and if these models are to be used for the computations relevant to the regulator, then they must be validated. On the other hand, usable internal models in the DFA context are likely to be of the DFA type, i.e. their underlying scenario generators must produce forecasts for entire distributions, and good distribution forecasts become even more important if tail-based risk measures like expected shortfall are used.

### 11.2 The PIT: Probability Integral Transform

Let us now introduce the PIT formally, in a univariate setting first. Let $\{Y_t\}_{t \in \mathbb{Z}}$ be a univariate time series, and let $\{\bar{y}_t\}_{t=-S}^T$ be a sample of observed realizations of this time series. Observations from $t = -S$ up to $t = 0$ are used as calibration sample, $t = 1$ up to $t = T$ is the out-of-sample period. Let $F_t$ denote the $\sigma$-algebra of information available up to and
11.2. The PIT: Probability Integral Transform

including time $t$, so that $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ is a filtration. Though it goes unmentioned in some situations, we assume all distributions to be conditional on $\mathcal{F}_0$. The variable of interest is $Y = (Y_1, \ldots, Y_T)'$, and we assume that its joint CDF $F(y_1, \ldots, y_T) := F(y_1, \ldots, y_T|\mathcal{F}_0)$ is absolutely continuous with respect to the Lebesgue measure. We let $F_t(y_t) := P[Y_t \leq y_t|\mathcal{F}_0]$ denote the marginal CDFs of the $Y_t$'s. The conditional laws of the $Y_t$'s are given by

$$F_t(y_t|\mathcal{F}_{t-1}) := F(y_t|y_{t-1}, \ldots, y_1, \mathcal{F}_0)$$

$$:= P[Y_t \leq y_t|Y_{t-1} = y_{t-1}, \ldots, Y_1 = y_1, \mathcal{F}_0].$$

Now we can state the theorem that is the foundation of the backtesting methods presented in the sequel:

**Theorem 11.1 (Lévy Transform, Rosenblatt Transform or Probability Integral Transform)**

Assume the setup introduced above and define a transformation $R: \mathbb{R}^T \rightarrow \mathbb{R}^T$ componentwise by

$$R_1(y) = F_1(y_1), R_2(y) = F_2(y_2|\mathcal{F}_1), \ldots, R_T(y) = F_T(y_T|\mathcal{F}_{T-1}).$$

Let $Z = R(Y)$. Then we have $Z_1, \ldots, Z_T \sim \text{iid } U(0,1)$.

**Proof.** From [89] or [109]

This theorem is the basis for the following proposition from [45] that covers the situation of single-step distribution or density forecasts:

**Proposition 11.2** Assume that the sequence of observations $\{\tilde{y}_t\}_{t=1}^T$ is generated from the sequence of unknown conditional laws $\{G_t(y|\mathcal{F}_{t-1})\}_{t=1}^T$. If a sequence of distribution forecasts $\{F_t(y|\mathcal{F}_{t-1})\}_{t=1}^T$ coincides with $\{G_t(y|\mathcal{F}_{t-1})\}_{t=1}^T$ then the sequence of probability integral transforms

$$\{z_t := F_t(\tilde{y}_t|\mathcal{F}_{t-1})\}_{t=1}^T$$

is iid $U(0,1)$.

**Proof.** See [45].
Hence, we only need to plug the observations $\tilde{y}_t$ into the distribution forecasts $\{F_t(\cdot | \mathcal{F}_{t-1})\}_{t=1}^{T}$ to obtain the sequence $\{z_t\}_{t=1}^{T}$ and the test the joint null hypothesis that the latter is independent and identically $U(0,1)$-distributed. This issue will be dealt with in Section 11.4.

Proposition 11.2 extends quite naturally to the situation of $k$-step-ahead forecasts, i.e. to the situation where we make forecasts $F_t(y|\mathcal{F}_{t-k})$ of the distribution of $Y_t$ given information only up to time $t-k$ for some fixed $k \in \mathbb{N}$. Let again $\{\tilde{y}_t\}_{t=1}^{T}$ be the set of out-of-sample observations and assume for simplicity that $T = k \cdot S$. In the same way as for the one-step-ahead forecasts, we can compute the PIT values $z_t := F_t(\tilde{y}_t|\mathcal{F}_{t-k})$ for each $t$. Now, we group the PIT values into $k$ subsets as follows:

$$
\begin{align*}
\{z_1, z_{1+k}, z_{1+2k}, \ldots, z_{1+(S-1)k}\}, \\
\{z_2, z_{2+k}, z_{2+2k}, \ldots, z_{2+(S-1)k}\}, \\
\vdots \\
\{z_k, z_{2k}, z_{3k}, \ldots, z_{Sk}\}.
\end{align*}
$$

Each one of these subsets contains PIT values from $k$-step-ahead forecasts on non-overlapping forecast intervals, so that Proposition 11.2 applies. Hence, the $z_t$'s within each one of the subsets above are iid $U(0,1)$, and the combined sample contains only dependence up to lag $k-1$. Formally:

**Proposition 11.3** Let $k \in \mathbb{N}$ and assume that the sequence of observations $\{\tilde{y}_t\}_{t=1}^{T}$, where $T = kS$, is generated from the sequence of conditional laws $\{G_t(y|\mathcal{F}_{t-1})\}_{t=1}^{T}$. If a sequence of $k$-step-ahead distribution forecasts $\{F_t(y|\mathcal{F}_{t-k})\}_{t=1}^{T}$ coincides with the related true conditional laws $\{G_t(y|\mathcal{F}_{t-k})\}_{t=1}^{T}$, then

$$
\{z_{i+s_k} = F_{i+s_k}(\tilde{y}_{i+s_k}|\mathcal{F}_{i+(s-1)k})\}_{s=1}^{S} \sim iid U(0,1), \quad i = 0, \ldots, k-1,
$$

and the combined sample $\{z_t\}_{t=1}^{T}$ only contains dependencies up to and including at most lag $k-1$.

The joint null hypothesis that the $z_t$'s are identically $U(0,1)$-distributed and that there is no dependence beyond lag $k-1$ can be tested by using a variety of methods; see Section 11.4 below.
11.2. The PIT: Probability Integral Transform

Using the same techniques as above, the PIT can also be extended to multivariate time series. Let now \( \{Y_t\}_{t \in \mathbb{Z}} \) be a multivariate time series with \( Y_t \in \mathbb{R}^N \) and let \( \{\tilde{y}_t\}_{t=-S}^T \) be the related sample of observations. We use the analogues of the different distribution functions given for the univariate case. In particular,

\[
F_t(y|F_{t-1}) = P\left( Y_t^1 \leq y^1, \ldots, Y_t^N \leq y^N \mid F_{t-1} \right)
\]

is the joint conditional CDF of \( Y_t \) given information up until time \( t-1 \). We assume that \( F_t(y|F_{t-1}) \) is absolutely continuous with respect to the Lebesgue measure. Using a copula argument to be made precise further below, we can show that \( F_t(y|F_{t-1}) \) has a representation

\[
F_t(y|F_{t-1}) = F_t\left( y^1 \mid F_{t-1} \right) \prod_{i=2}^{N} F_t\left( y^i \mid y^{i-1}, \ldots, y^1; F_{t-1} \right). \tag{11.1}
\]

The terms on the right-hand side can be identified with the components of the PIT given in Theorem 11.1, but in this case along the dimensions of \( Y_t \) for a fixed \( t \) rather than along the time axis, i.e.

\[
F_t(y|F_{t-1}) = R_1(y)R_2(y) \cdots R_N(y),
\]

so that, for each fixed \( t \), \( Z_t = R(Y_t) \) is uniformly distributed on \([0,1]^N\) with independent components. Together with the result for the univariate case along the time axis, we obtain

**Proposition 11.4** Assume that the sequence of observations \( \{\tilde{y}_t\}_{t=1}^T \) is generated from the sequence of unknown joint conditional laws \( \{G_t(y|F_{t-1})\}_{t=1}^T \). If a sequence of joint distribution forecasts \( \{F_t(y|F_{t-1})\}_{t=1}^T \) coincides with \( \{G_t(y|F_{t-1})\}_{t=1}^T \), then each one of the sequences \( \{z_t^i\}_{t=1}^T \), where

\[
z_t^i = F_t\left( y^i \mid y^{i-1}, \ldots, y^1; F_{t-1} \right),
\]

of componentwise probability integral transforms is iid \( U([0,1]) \), and the different univariate series \( \{z_t^i\}_{t=1}^T \) are mutually independent.

Methods to test for the joint null hypothesis and to extract specific information from deviations from the latter will be given in Section 11.4.
11.3 PIT: The Non-Parametric Case

The PIT as introduced in Section 11.2 implicitly assumes that the forecast distribution $F_t(y|\mathcal{F}_{t-1})$ is given analytically. This is often not the case, either because the underlying model is too complicated, or because it is non-parametric. However, the Monte Carlo approach provides a solution to this problem, thus making the PIT procedure universally applicable. We can generate a large sample $\{Y_i\}_{i=1}^N$ of iid replicates of $Y \sim F$, where we let $Y := Y_t$ and $F(y) := F_t(y|\mathcal{F}_{t-1})$ for simplicity. Then we can approximate the analytical distribution $F$ by the related empirical distribution function (EDF)

$$\hat{F}_N(y) = \frac{1}{N} \sum_{i=1}^N 1\{Y_i \leq y\}, \quad y \in \mathbb{R}. \quad (11.2)$$

As long as we can assume that $F$ is continuous, the theorem of Glivenko-Cantelli assures that the ECDF converges uniformly to the analytical one:

$$\sup_{y \in \mathbb{R}} |\hat{F}_N(y) - F(y)| \to 0 \quad \text{P-a.s.} \quad (N \to \infty). \quad (11.3)$$

This is reassuring, but of less practical value. Of interest is rather the number $N$ of Monte Carlo replicates that we have to generate such that $\hat{F}_N(y)$ deviates from $F(y)$ by at most $\varepsilon$ for some confidence level $\alpha$:

$$\mathbb{P} \left[ \sup_{y \in \mathbb{R}} |\hat{F}_N(y) - F(y)| > \varepsilon \right] \leq \alpha. \quad (11.4)$$

In the sequel, we provide a simple approach to tackle this problem: Let $\hat{F}_N$ again denote the EDF obtained from $N$ iid Monte Carlo replicates $\{\hat{Y}_i\}_{i=1}^N$ as in (11.2) above. Moreover, let $\hat{U}_N$ denote the EDF of an iid sequence $\{\hat{U}_i\}_{i=1}^N$ of $U(0,1)$-distributed random observations. Then, as shown in Section 6.10 of [105], we have

$$D_N := \sup_{y \in \mathbb{R}} |\hat{F}_N(y) - F(y)| \overset{d}{=} \sup_{y \in \mathbb{R}} |\hat{U}_N(y) - U(y)|, \quad (11.5)$$

where the second equality denotes equality in distribution. This means that the distribution of $D_N$ does not depend on the underlying distribution function $F$, as long as the latter is continuous. (11.5) is the basis for the well-known Kolmogorov-Smirnov test for the goodness-of-fit. Here,
however, we continue on a different path. In fact, $D_N$ has the limiting property

$$\sqrt{N}D_N \xrightarrow{d} D \quad (N \to \infty),$$

where

$$P[D > d] = 2 \sum_{k=1}^{\infty} (-1)^{k-1}e^{-2k^2d^2}, \quad d \geq 0. \quad (11.6)$$

This approximation is fairly accurate for $N \geq 100$. Now let $0 < \alpha < 1$ and $\varepsilon > 0$. Then we must have as a boundary case

$$P[D_N > \varepsilon] = \alpha,$$

which is equivalent to

$$P[\sqrt{N}D_N > \sqrt{N}\varepsilon] = \alpha.$$

Given the asymptotic result above, we have

$$P[D > \sqrt{N}\varepsilon] \approx \alpha$$

with fairly high accuracy for sufficiently large $N$ (see above).

Note that, by (11.6), $\alpha$ decreases as $N$ increases. Hence, for given $\alpha$ and $\varepsilon$, we have to find

$$N_{\alpha,\varepsilon} = \min \left\{ n \in \mathbb{N} \, \bigg| \, 2 \sum_{k=1}^{\infty} (-1)^{k-1}e^{-2k^2n\varepsilon^2} \leq \alpha \right\}. \quad (11.7)$$

Then, for $N \geq N_{\alpha,\varepsilon}$, the Monte Carlo EDF $\hat{F}_N$ satisfies the accuracy property in (11.4). For the sake of completeness, we note that

$$2 \sum_{k=1}^{\infty} (-1)^{k-1}e^{-2k^2d^2} = 1 - \vartheta_4 \left(0, e^{-2d^2}\right),$$

where $\vartheta_4(z, q)$ is the Jacobi Theta function of the fourth type according to (16.27.4) of [1]. Within the present study, the evaluation of (11.6) is done by brute force, i.e. by summing over a finite number of terms, which is justified given the quick decay of the absolute value of the terms.
Figure 11.1: Required number $N$ of Monte Carlo simulations according to (11.7) versus allowed discrepancy $\varepsilon$ between theoretical and empirical distribution for different confidence levels $\alpha$ in the case of continuous univariate distributions (on double-logarithmic scale).

Figure 11.1 shows the required number $N$ of Monte Carlo simulations for different allowed discrepancies $\varepsilon$ and for different confidence levels $\alpha$ according to (11.7). We see that, even in the simple case of continuous univariate distributions, the minimum number of simulations for high accuracy levels easily reaches the thousands. In particular, the number of simulations increases very quickly for higher required accuracy.

This consideration is just for the univariate case only. An equivalent statement for the multivariate case with general dependence structure is desirable though it may not be so straightforward. There exists generalization of the Kolmogorov-Smirnov statistic for higher dimensions; see [22]. However, it would require an amount of work that is momentarily beyond the scope of what is possible in the realm of this thesis to make
such a generalization operational. Therefore, for the time being, we give here a more ad-hoc approach.

Suppose now that we have to simulate a multivariate random variable \( Y \in \mathbb{R}^d \), and let \( \varepsilon > 0 \) and \( \alpha \in (0, 1) \) again be the maximum discrepancy and the confidence level, respectively. In analogy to (11.5), we consider the marginal Kolmogorov-Smirnov discrepancy for each dimension,

\[
D_N^i := \sup_{y^i \in \mathbb{R}} |\hat{F}_N^i(y^i) - F_i(y^i)|, \quad i \in \{1, \ldots, d\},
\]

(11.8)

where \( \hat{F}_N^i \) and \( F_i \) are the empirical and the theoretical marginal distribution functions, respectively. We are interested in the event that none of the discrepancies \( D_N^i \) exceeds the tolerance \( \varepsilon \) at a given confidence probability \( \alpha \),

\[
P \left[ \exists i \in \{1, \ldots, d\}: D_N^i > \varepsilon \right] \leq \alpha.
\]

(11.9)

This is equivalent to

\[
P \left[ \forall i \in \{1, \ldots, d\}: D_N^i \leq \varepsilon \right] \geq 1 - \alpha,
\]

(11.10)

which can be reformulated as

\[
P \left[ \forall i \in \{1, \ldots, d\}: D_N^i \leq \varepsilon \right] \geq 1 - \sum_{i=1}^{d} P \left[ D_N^i > \varepsilon \right] = 1 - d \cdot P \left[ D_N > \varepsilon \right].
\]

(11.11)

The first part is by Bonferroni’s inequality, and the second part is by (11.5) which states that all \( D_N^i \)'s have the same distribution. Combining Equations 11.10 and 11.10, we obtain the condition

\[
P \left[ D_N > \varepsilon \right] \leq \frac{\alpha}{d},
\]

(11.12)

where we can again use the approximation from (11.7). This method actually neglects dependencies between components of \( Y \) and, therefore, is a very crude upper bound for the required number \( N \) of simulations. In the case of strong dependence of the components of \( Y \), the actually required number may be considerably lower.
This generalization to the multivariate case is better than nothing, but otherwise clearly unsatisfactory, as it only specifies a very crude upper bound for the required number of simulations. A more sophisticated evaluation, e.g. along the lines of [22], would be desirable.

11.4 PIT: Evaluation

Up until now, we have explored the PIT in different settings and we have shown that, if the forecasted distributions \( \{F_t(\cdot|\mathcal{F}_{t-1})\}_{t=1}^T \) coincide with the true conditional distributions \( \{G_t(\cdot|\mathcal{F}_{t-1})\}_{t=1}^T \), then the PIT transforms \( \{z_t\}_{t=1}^T \) of the observations \( \{y_t\}_{t=1}^T \) are iid \( U(0,1) \). The tasks are now

- To test whether the joint hypothesis of \( \{z_t\}_{t=1}^T \) being iid and \( U(0,1) \) holds.
- To extract as much information as possible from possible deviations from the null hypothesis.

There are two basic approaches for these tasks, which could be labeled as parametric and non-parametric. In the parametric approach, one assumes that a parametric model is underlying \( \{z_t\}_{t=1}^T \). This model must correspond to the iid \( U(0,1) \) case for certain parameter values, and the test is then made for the parameters to have these specific values. The parametric approach in the real of PIT evaluation was pioneered by [11]. Here, we consider the more general approach by [32], which actually contains [11] as a special case. The advantage of the parametric approach is that it requires relatively low amounts of data to come to significant rejections of the null hypothesis under realistic conditions, which is important in the setting investigated in this thesis. The disadvantage is that parametric approach tests do not allow to extract much information beyond rejection or non-rejection. The non-parametric approach basically consists of evaluating histograms or empirical densities (for the \( U(0,1) \) distributional assumption) and correlograms or periodograms (for the independence) calculated from \( \{z_t\}_{t=1}^T \). This approach is advocated in [45] and [46] and presented here with a number of refinements. The advantage
of the non-parametric approach is that it allows to extract specific information on the mis-specification and improvement potential of the model. The disadvantage is that it generally requires relatively large amounts of data, and it is often not so simple to come to a significant overall rejection.

### 11.4.1 Parametric Evaluation

The task at hand here is to test for the joint null hypothesis that the sequence \( \{z_t\}_{t=1}^T \) of PIT values is iid and \( U(0,1) \)-distributed. Note that testing for the \( U(0,1) \) hypothesis alone is quite simple; it suffices to use one of the well-known tests such as \( \chi^2 \) or Kolmogorov-Smirnov; see [115]. Testing the joint hypothesis, or testing for the absence of dependence irrespective of the marginal distribution, is considerably more difficult. For these issues, we introduce and implement here a copula-based approach suggested by [32].

The idea is strikingly simple: We embed the null hypothesis that \( \{z_t\}_{t=1}^T \) is iid \( U(0,1) \) in a broad class of copula-based Markov Chain models and then test for the iid \( U(0,1) \) case within this class. This allows us to obtain power against a broad range of alternative hypotheses, which is not the case for more simplistic evaluation frameworks; see below. More specifically:

**Assumption 11.5** We assume that the sequence of PIT values \( \{z_t\}_{t=1}^T \) is the realization of a Markov Chain with true stationary distribution \( G^*(\cdot) \) and copula \( C^*(\cdot,\cdot) \) for the dependence structure of \( (z_{t-1}, z_t) \). Moreover, we assume that both \( G^*(\cdot) \) and \( C^*(\cdot,\cdot) \) are absolutely continuous with respect to the Lebesgue measure on their respective domains of definition.

For our test framework, we select a parametric family \( \{G(\cdot;\beta) : \beta \in B\} \) of marginal distributions and a parametric family of copulae \( \{C(\cdot,\cdot;\alpha) : \alpha \in A\} \).

**Assumption 11.6** For the selected parametric families, we assume

1. \( G(\cdot;\beta) \) and \( C(\cdot,\cdot;\alpha) \) have densities for all \( \alpha \in A \) and \( \beta \in B \).
2. \( \exists \beta^* \in B : G(u; \beta^*) = G^*(u), \quad u \in [0, 1]. \)

3. \( \exists \alpha^* \in A : C(u, v; \alpha^*) = C^*(u, v), \quad (u, v) \in [0, 1]^2. \)

4. \( \exists \beta_0 \in B : G(u; \beta_0) = u, \quad u \in [0, 1]. \)

5. \( \exists \alpha_0 \in A : C(u, v; \alpha_0) = uv, \quad (u, v) \in [0, 1]^2. \)

Assumption 11.6 means that the model must encompass both the true model and the null hypothesis. The log-likelihood function of \( \{z_t\}^T_{t=1} \) is then given by

\[
\ell(\alpha, \beta) = \frac{1}{T} \sum_{t=1}^{T} \log g(z_t; \beta) + \frac{1}{T} \sum_{t=2}^{T} \log c(G(z_{t-1}; \beta), G(z_t; \beta); \alpha).
\]

We note that Assumption 11.6 implies that \( \ell(\alpha_0, \beta_0) = 0. \) The null hypothesis now reads as

\[
H_0 : \quad \alpha^* = \alpha_0 \quad \wedge \quad \beta^* = \beta_0,
\]

which can be tested for by using the likelihood ratio framework as described, for instance, in [103]: If \((\hat{\alpha}, \hat{\beta})\) is the unrestricted maximum likelihood estimate of \((\alpha, \beta)\), then, under \(H_0\), we have

\[
S_{LR} := 2T \left( \ell(\hat{\alpha}, \hat{\beta}) - \ell(\alpha_0, \beta_0) \right) \sim \chi_k^2,
\]

where \(k\) is the difference in the number of parameters to be estimated, that is, \(k = \dim((\alpha^*, \beta^*)') - \dim((\alpha'_0, \beta'_0)').\) This is the general framework, and we can now define specific models for both the stationary distribution and the copula.

For \(G(\cdot; \beta)\), we need a parametric family of distributions that can approximate a wide range of distributions for variables taking values only in \([0, 1]\), as \(z_t\) will never assume any value outside this interval. We loosely follow a suggestion given in [32] and define

\[
g(z; \beta) = C(\beta) \exp \left\{ \sum_{n=1}^{N} \beta_n P_n \left( \frac{2z - 1}{2} \right) \right\}, \quad 0 \leq z \leq 1,
\]

(11.16)
where \( k \in \mathbb{N}, \beta = (\beta_1, \ldots, \beta_N)' \in \mathbb{R}^N \), \( C(\beta) \) is a constant ensuring that \( \int_0^1 g(z; \beta) \, dz = 1 \), and \( P_n(x) \) denotes the Legendre polynomial of degree \( n \). These polynomials are widely used in numerical mathematics for the approximation of functions with compact support; see e.g. [111] or [101] for details. If necessary, the Legendre polynomial of degree \( n \) can be expressed explicitly as

\[
P_n(x) = \frac{1}{2^n} \sum_{m=0}^{\left\lfloor \frac{n}{2} \right\rfloor} (-1)^m \binom{n}{m} x^{n-2m} \left( \frac{2n-2m}{n} \right), \quad -1 \leq x \leq 1.
\]

The Legendre polynomials are defined on \([-1, 1]\). Therefore, the variable transformation \( x = \frac{2z - 1}{2} \) is done to transform \( P_n(x) \) in such a way that it covers the domain \([0, 1]\) of the \( z_t \)'s. The case of \( \theta = 0 \) corresponds to the \( \mathcal{U}(0,1) \) distribution. The advantage of this model is that it can approximate a wide range of different distributions on the unit interval. The downside is that calibration is quite cumbersome, since the DF needed for the full log-likelihood function of (11.13) cannot be computed in closed form, and Matlab does not offer an explicit function. Moreover, the most suitable order \( N \) must also be determined, which means iteration over different values of \( N \) and comparison based on AIC. A somewhat simpler alternative to the model of (11.16) is the \( \beta \)-distribution, which has density

\[
g(z; \beta) = \frac{\Gamma(\beta_1 + \beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} z^{\beta_1-1}(1-z)^{\beta_2-1}, \quad 0 \leq z \leq 1,
\]

where \( \Gamma(\cdot) \) is the Gamma function and \( \beta = (\beta_1, \beta_2)' \); see [77] for details. The case \( \beta = (1, 1)' \) corresponds to the \( \mathcal{U}(0,1) \)-distribution. Although less versatile than the generic model in (11.16), the beta distribution also covers a wide range of different shapes of distributions on the unit interval. Moreover, computational support in Matlab is readily available through the functions betacdf and betapdf.

Now, we turn to possible models for the copula \( C(u, v; \alpha_0) \) in Assumption 11.6. The use of copulae in risk management is fairly well publicized so that we do not go further into details here (see [54], [53] and [31]). The crucial technical condition for a copula to be easily usable in the present context is that its density is continuous and has an explicit representation. The list of copulae presented hereafter is not exhaustive; other
The first copula considered is the Gaussian copula, which is a representative of the class of Elliptical copulas and which allows the approach studied here to encompass some more classical setups. Its density is given by

\[
c(u_1, u_2; \alpha) = \frac{\phi_{\alpha}(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}{\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))}, \quad -1 \leq \alpha \leq 1, \tag{11.18}
\]

where \(\Phi(\cdot)\) and \(\phi(\cdot)\) are the CDF and PDF respectively of the univariate standard Normal distribution and \(\phi_{\alpha}(\cdot, \cdot)\) is the joint PDF of a bivariate standard Normal distribution with correlation \(\alpha\). The case of \(\alpha = 0\) corresponds to independence. The Gaussian copula is symmetric and does not exhibit tail dependence, but allows to produce both positive and negative dependence.

A radical alternative is the Gumbel copula. Its CDF is given by

\[
C(u_1, u_2; \alpha) = \exp \left\{ -\left(\tilde{u}_1^{\alpha} + \tilde{u}_2^{\alpha}\right)^{1/\alpha} \right\}, \quad 1 \leq \alpha < \infty,
\]

where \(\tilde{u}_i = -\log u_i\). The related PDF is given by

\[
c(u_1, u_2; \alpha) = C(u_1, u_2; \alpha)(u_1u_2)^{-1} \times
\frac{(\tilde{u}_1 \tilde{u}_2)^{\alpha-1}}{(\tilde{u}_1^{\alpha} + \tilde{u}_2^{\alpha})^{2-1/\alpha}} \left[(\tilde{u}_1^{\alpha} + \tilde{u}_2^{\alpha})^{1/\alpha} + \alpha - 1 \right]. \tag{11.19}
\]

Independence occurs at \(\alpha = 1\), which may be problematic as this is on the boundary of the parameter space. This copula is not symmetric and allows for tail dependence. It cannot produce negative dependence.

The Farlie-Gumbel-Morgenstern (FGM) copula is given by

\[
C(u_1, u_2; \alpha) = u_1 u_2 \left[1 + \alpha(1 - u_1)(1 - u_2)\right], \quad -1 \leq \alpha \leq 1.
\]

It has density

\[
c(u_1, u_2; \alpha) = 1 + \alpha(1 - 2u_1)(1 - 2u_2), \quad -1 \leq \alpha \leq 1. \tag{11.20}
\]

Independence corresponds to \(\alpha = 0\). This copula is symmetric and can produce both positive and negative dependence. However, it does not
11.4. PIT: Evaluation

exhibit tail dependence.

The use of this type of tests in finance was actually pioneered in [11], but in a much simpler setting where \( \{z_t\}_{t=1}^T \) is embedded into the model

\[
z_t - \mu = \alpha (z_{t-1} - \mu) + \sigma \varepsilon_t, \quad \text{where } \varepsilon_t \sim \text{iid } \mathcal{N}(0, 1).
\]  

(11.21)

It can be shown – see [32] – that this is actually a special case of the setup presented in this section: If one takes the Gaussian copula from (11.18), and if one selects

\[
G(z) = \Phi \left( \frac{\Phi^{-1}(z) - \mu}{\sigma/\sqrt{1 - \alpha^2}} \right)
\]

(11.22)

for the marginal distribution, then this is equivalent to the time series model in (11.21).

Let us now consider a variant of the above tests that verifies the dependence structure only, assuming that the marginal law \( G^*(\cdot) \) is completely unknown. The original null hypothesis from (11.14) simplifies to

\[
H_0 : \alpha^* = \alpha_0,
\]

(11.23)

irrespective to the choice of a model for \( G^*(\cdot) \). This test is based on a Pseudo-MLE estimate of the copula parameter \( \alpha \), specifically

\[
\hat{\alpha} = \arg\max_{\alpha \in \mathcal{A}} \hat{\ell}(\alpha),
\]

where

\[
\hat{\ell}(\alpha) = \frac{1}{T} \sum_{t=2}^T \log c(G_T(z_{t-1}), G_T(z_t); \alpha).
\]

(11.24)

For the marginal distribution, we plug in the rescaled empirical distribution function

\[
G_T(z) = \frac{1}{T+1} \sum_{t=1}^T 1\{z_t \leq z\}, \quad 0 \leq z \leq 1.
\]

(11.25)

In order for this modified estimation to work, some additional conditions on top of assumptions stated above must be put in place:
Proposition 11.7 Let Assumptions 11.5 and 11.6 as well as the conditions of Proposition 4.3 of [31] hold. Then, under the null hypothesis of (11.23), the estimate \( \hat{\alpha} \) from (11.24) is consistent and asymptotically normal, that is

\[
\sqrt{T} (\hat{\alpha} - \alpha^*) \to \mathcal{N}(0, B^{-1}) \quad \text{as} \quad n \to \infty,
\]

where \( B \) is the Fisher information

\[
B = \mathbb{E} \left[ \frac{\partial}{\partial \alpha} \log c(G^*(z_{t-1}), G^*(z_t); \alpha^*) \frac{\partial}{\partial \alpha} \log c(G^*(z_{t-1}), G^*(z_t); \alpha^*)' \right].
\]

Proof. See [31] and [32]. \( \square \)

Remark 11.8 The additional conditions mentioned in Proposition 11.7 are highly technical, and the details are omitted here for the moment. However, they should be added later on. We only mention here that \( \alpha^* \) should be in the interior of the parameter space \( \mathcal{A} \subseteq \mathbb{R}^d \), which precludes the use of the Gumbel copula (see (11.19) in this context).

The parametric likelihood ratio test from (11.15) can now be replaced by a pseudo-likelihood analogue. If the assumptions of Proposition 11.7 hold, then, under the modified null hypothesis from (11.23), we have

\[
S_{PLR} = 2T \left( \tilde{\ell}(\tilde{\alpha}) - \ell(\alpha^*) \right) \sim \chi_k^2,
\]

where \( \tilde{\ell}(\cdot) \) is the modified likelihood function of (11.24) and \( k = \text{dim}(\alpha) \). Alternatively, [32] also suggest a Wald-type test using the asymptotic property established in Proposition 11.7. To this end, they show that the Fisher information \( B \) can be estimated as

\[
\hat{B} = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \frac{\partial}{\partial \alpha} \log c(U_i, U_j; \alpha_0) \frac{\partial}{\partial \alpha} \log c(U_i, U_j; \alpha_0)',
\]

where \( \{U_i\}_{i=1}^N \) is a large, simulated sample of iid \( U(0,1) \) random variables. Guidelines on how large this sample has to be are provided in Section 11.3.

Proposition 11.9 Let the same conditions hold as in Proposition 11.7, and let

\[
\tilde{W} = (\hat{\alpha} - \alpha_0)' \hat{B} (\hat{\alpha} - \alpha_0).
\]
11.4. PIT: Evaluation

<table>
<thead>
<tr>
<th>Marginal</th>
<th>Copula</th>
<th>Statistic</th>
<th>1000 Pts.</th>
<th>500 Pts.</th>
<th>250 Pts.</th>
<th>125 Pts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal Only</td>
<td></td>
<td>$\chi^2$</td>
<td>0.6921</td>
<td>0.8767</td>
<td>0.2375</td>
<td>0.1522</td>
</tr>
<tr>
<td>Dependence Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>PLR</td>
<td>0.8598</td>
<td>0.6304</td>
<td>0.1179</td>
<td>0.4628</td>
</tr>
<tr>
<td></td>
<td>Wald</td>
<td></td>
<td>0.8537</td>
<td>0.6316</td>
<td>0.0873</td>
<td>0.4451</td>
</tr>
<tr>
<td></td>
<td>FGM</td>
<td>PLR</td>
<td>0.8821</td>
<td>0.5795</td>
<td>0.1042</td>
<td>0.2174</td>
</tr>
<tr>
<td></td>
<td>Wald</td>
<td></td>
<td>0.8774</td>
<td>0.5745</td>
<td>0.0894</td>
<td>0.1910</td>
</tr>
<tr>
<td>Joint Marginal and Dependence</td>
<td></td>
<td>LR</td>
<td>0.7407</td>
<td>0.6196</td>
<td>0.2052</td>
<td>0.1703</td>
</tr>
<tr>
<td></td>
<td>Beta</td>
<td>Gumbel</td>
<td>0.9572</td>
<td>0.6800</td>
<td>0.2211</td>
<td>0.2008</td>
</tr>
<tr>
<td></td>
<td>Gauss</td>
<td>LR</td>
<td>0.7864</td>
<td>0.5842</td>
<td>0.1983</td>
<td>0.1419</td>
</tr>
</tbody>
</table>

Table 11.1: Sample table of results for the various parametric evaluations described in Section 11.4.2. Data used here consists of different numbers of artificial iid $U(0,1)$ replicates. All numbers are $p$-values of the respective tests.

Then, under the null hypothesis from (11.23), we have

$$TW^d \xrightarrow{d} \chi^2_k$$

as $T \to \infty$,

where $k = \text{dim}(\alpha)$, and $d$ denotes convergence in distribution.

PROOF. See [31] and [32].

The downside of the Wald approach is that it relies only on asymptotic properties. Hence it may be less suitable for the typical situation where only small amounts of data are available.

Table 11.1 provides a sample evaluation table containing the results of all the tests presented in this section. The first column specifies the assumption made for the marginal distribution $G(\cdot;\beta)$ and the second column specifies the assumption for the copula $C(\cdot, \cdot; \alpha)$ of the Markov chain according to Assumption 11.5. The third column specifies the type of test used. Each further column gives the $p$-values of the different tests for a specific series of PIT values. The upper part of the table contains an evaluation of the marginal distribution only. The test is the $\chi^2$-test as given by (11.28). The middle part contains tests for the correct dependence structure. For each choice of copula, two tests are carried out,
11.4.2 Non-parametric Evaluation

For a non-parametric evaluation of \( \{z_t\}_{t=1}^T \), which does not impose any particular model assumptions, we basically follow the informal approach of [45] and [46], but we add some further evaluations in order to extract more information from the sample. All evaluations are collected in a standardized figure which contains six panels. An example of such an evaluation, based on artificial data, is shown in Figure 11.2. Note that all tests are for a confidence level of \( \alpha = 5\% \) unless stated otherwise.

The first panel on the upper left contains a simple bar plot of the \( z_t \)'s in their chronological order. It allows to get a first superficial idea as to whether the \( z_t \)'s are uniformly distributed on the unit interval, and whether there are intertemporal dependencies.

The second panel on the upper right provides a formal assessment of the hypothesis that the \( z_t \)'s are unconditionally uniformly distributed on the unit interval. To this end, a histogram of \( \{z_t\}_{t=1}^T \) with \( H = \lfloor \sqrt{T} \rfloor \) equispaced bins is plotted. Let \( n_h \) denote the number of observed \( z_t \)'s falling into a particular bin \( h = 1, \ldots, H \). Under the null hypothesis, the probability \( p_h \) of any observation \( z_t \) falling into bin \( h \) is \( 1/H \) for each \( h \). Hence, for each bin \( h \), the number \( n_h \) of observations falling into it has a Binomial distribution with parameters \( T \) and \( 1/H \). The dash-
Figure 11.2: Sample picture for the non-parametric evaluation described in Section 11.4.2 based on 500 simulated \( U(0,1) \) replicates. Even though the data is iid \( U(0,1) \), one can see that the relatively small sample leads to some partial evaluations that are in contradiction to the null hypothesis, particularly the correlograms for the second and fourth power.
dotted horizontal lines in the panel specify the related confidence interval. Moreover, a $\chi^2$-test for uniform distribution is carried out, see [115], the $p$-value of which is indicated above the panel. Specifically, under the null hypothesis we have:

$$S_{\chi^2} := \sum_{h=1}^{H} \frac{(n_h - T/H)^2}{T/H} \sim \chi^2_{H-1}. \quad (11.28)$$

The remaining four panels investigate the intertemporal independence of $\{z_t\}_{t=1}^T$. In particular, they show the autocorrelation functions of $\{(z_t - \mu(z))^k\}_{t=1}^T$, where $\mu(z) = (1/T) \sum_{t=1}^T z_t$ and $k = 1, 2, 3, 4$. These evaluations allow us to assess whether the model is able to capture the conditional dynamics of the data in terms of mean, variance, skewness and kurtosis. The specific plots are Autocorrelation Diagnostic Plots as described in Appendix A.2.

A reading guide on how to extract information from the evaluation figure is provided in [45] for some basic mis-specifications and in [46] for some more sophisticated ones. The additional test statistics present in the evaluations here allow to make better-founded decisions on the significance of a phenomenon if the plot itself is not very clear.

A word of warning: The measurement of (auto)correlations suffers from the notorious problems described in [54]. On the one hand, independence always implies zero correlation, so that rejecting the null hypothesis in the case of significant non-zero correlation is no problem. On the other hand, absence of significant correlation is absolutely no guarantee for independence in general situations. In statistical language: The evaluation here has no power at all against a very wide range of specific alternatives. There is, however, no need to worry about possible bounds for the correlation other than $-1$ and $1$, since we can reasonably assume that the distributions of $(z_t - \mu(z))$ and $(z_{t-1} - \mu(z))$ are of the same type, so that Theorem 4 of [54] applies.

The example in Figure 11.2 also shows a serious problem of non-parametric evaluation: Although it was generated with 500 iid $U(0,1)$ replicates,
11.5. Backtesting: Case Studies

there are a few significant autocorrelation estimates, and the histogram does not look very uniform as well. In practical settings, backtest samples will usually be shorter, so that this type of evaluation suffers somewhat from small-sample uncertainty.

11.5 Backtesting: Case Studies

11.5.1 Single-Currency Case – USD

In this section, we apply the backtesting methodology introduced above to the single-currency, two-factor exponential-quadratic model calibrated according to Section 10.8.1. Specifically, we only use the "diagonal" parameter set according to Table 10.1. Given the general wisdom that a yield curve model in a DFA setting should have three factors (see Section 6.3) and given the GMM diagnostics from Section 10.8.1, we do not expect a good performance in the backtests. Therefore, we do not lose a lot by restricting the backtest to an in-sample period, namely the 120 months from January 1994 to December 2003. Moreover, we only carry out univariate backtests for the 1-, 5-, 10-, and 20-year rates.

The non-parametric evaluations for the one-year and 20-year rates are shown in Figure 11.3 and Figure 11.4, respectively. The results are rather sobering: the unconditional distributions are significantly non-uniform, and there are significant autocorrelations for all cases investigated. The histogram in Figure 11.3 suggests that the unconditional level and variance of the one-year rate are systematically overstated by the model, and none of the conditional moments is properly reflected either. Figure 11.4 suggests that the unconditional distribution of the 20-year rate is systematically under-estimated; the rejection with respect to the conditional moments is not as blunt as for the one-year rate. Given this, it makes no sense to carry out further evaluations of this specific model here. A conclusion is that we should go for a model with three rather than two factors to ensure a sufficient richness of possible structures.
Figure 11.3: Non-parametric evaluation plot according to Section 11.4.2 for the one-year rate in the backtest setup of Section 10.8.1. The histogram as well as the correlograms suggest a clear rejection of the model; see p. 189 for a detailed discussion.
Figure 11.4: Non-parametric evaluation plot according to Section 11.4.2 for the 20-year rate in the backtest setup of Section 10.8.1. The histogram suggests a clear rejection w.r.t. the marginal distribution, whereas the rejection of the intertemporal dependence structure suggested by the correlograms is not very significant; see p. 189 for a detailed discussion.
Table 11.2: Result table for the various parametric evaluations from Section 11.4.1 applied to the backtest setup of Section 11.5.1. All numbers are p-values. Numbers in rows labeled with PLR are results of the parametric likelihood ratio test according to (11.15), Wald refers to the Wald test statistic in Proposition 11.9 and PLR refers to the semi-parametric likelihood ratio statistic from (11.26). Beta refers to the use of the beta-distribution (11.17) as the model for the marginal distribution, whereas Gauss refers to the Gaussian one from (11.22). For the dependence structures, we have FGM according to (11.20), Gumbel according to (11.19) and FGM according to (11.20). For further discussions see p. 189.

<table>
<thead>
<tr>
<th>Marginal</th>
<th>Copula</th>
<th>Statistic</th>
<th>1y-rate</th>
<th>5y-rate</th>
<th>10y-rate</th>
<th>20y-rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n/a</td>
<td>$\chi^2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependence Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n/a</td>
<td>Gauss</td>
<td>PLR</td>
<td>0.1487</td>
<td>0.0268</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.1131</td>
<td>0.0153</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>n/a</td>
<td>FGM</td>
<td>PLR</td>
<td>0.5044</td>
<td>0.1534</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wald</td>
<td>0.4848</td>
<td>0.1478</td>
<td>0.0001</td>
<td>0.0003</td>
</tr>
<tr>
<td>Joint Marginal and Dependence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta</td>
<td>FGM</td>
<td>LR</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Beta</td>
<td>Gumbel</td>
<td>LR</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Gauss</td>
<td>Gauss</td>
<td>LR</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
11.5. Backtesting: Case Studies

Table 11.2 shows the results of the parametric evaluation described in Section 11.4.1. As one would expect from the histograms in Figures 11.3 and 11.4, the null hypothesis that the marginal distribution is $U(0,1)$ is grossly rejected, and so is the joint hypothesis. However, we can also see that the partial hypothesis regarding the conditional dynamics is clearly not rejected for the 1-year rate and only weakly rejected for the 5-year rate. This goes clearly against the intuition obtained from the correlograms in Figure 11.3. The background may be the well-known fact that MLE is very sensitive against even slight violations of the underlying model assumptions. So, if the model assumptions are as clearly violated as is suggested by Figure 11.3, then the output of the MLE should not be trusted. This suggests to follow a pragmatic procedure:

1. Do the non-parametric evaluation as described in Section 11.4.2. If this evaluation suggests a clear violation of the assumptions (marginal, dependence, or both) then reject the model or at least the clearly violated assumptions.

2. In case there is no clear rejection (of marginal, dependence, or both), carry out the parametric tests as described in Section 11.4.1 to obtain additional insight or further confirmation.

11.5.2 Double-Currency Case – CHF and USD

In this section, we apply the backtesting methodology introduced above to the double-currency, four-factor exponential-quadratic model calibrated according to Section 10.8.2. Again, we only use the "diagonal" parameter set according to Table 10.2, and we consider the rates for 1y, 5y, 10y and 20y time-to-maturity. We also do the evaluation in-sample rather than out-of-sample as it should be. This is mainly in order to obtain a backtest sample of sufficient length with the limited amount of historical data available. However, given the wide variety of behavioral patterns in the historical data (see Section 6.3), the results of such an in-sample evaluation can be considered as strong indicators for the behaviour of the model out-of-sample.
Figure 11.5: Non-parametric evaluation plot according to Section 11.4.2 for the CHF 10-year rate in the backtest setup of Section 10.8.2. The histogram clearly suggests no rejection w.r.t. the marginal distribution. The rejection of the intertemporal dependence structure is rather clear for the even moments, but little to not significant for the odd ones; see p. 193 for a more detailed discussion.
11.5. Backtesting: Case Studies

<table>
<thead>
<tr>
<th></th>
<th>Marginal $\chi^2$</th>
<th>Dependence F-G-M</th>
<th>Marginal &amp; Dependence F-G-M + $\beta$</th>
<th>Gauss + Gauss</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHF 1-year</td>
<td>0.0007</td>
<td>0.8712</td>
<td>0.0001</td>
<td>0.0000</td>
</tr>
<tr>
<td>CHF 5-year</td>
<td>0.1104</td>
<td>0.9436</td>
<td>0.0204</td>
<td>0.0677</td>
</tr>
<tr>
<td>CHF 10-year</td>
<td>0.2268</td>
<td>0.2551</td>
<td>0.0057</td>
<td>0.0124</td>
</tr>
<tr>
<td>CHF 20-year</td>
<td>0.0000</td>
<td>0.0131</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>USD 1-year</td>
<td>0.0000</td>
<td>0.1107</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>USD 5-year</td>
<td>0.0000</td>
<td>0.0235</td>
<td>0.3395</td>
<td>0.3182</td>
</tr>
<tr>
<td>USD 10-year</td>
<td>0.0001</td>
<td>0.0447</td>
<td>0.5176</td>
<td>0.4535</td>
</tr>
<tr>
<td>USD 20-year</td>
<td>0.0142</td>
<td>0.0909</td>
<td>0.2061</td>
<td>0.0712</td>
</tr>
</tbody>
</table>

**Table 11.3**: Result of the formal tests according to Section 11.4.1 applied to the backtest setup of Section 11.5.2. All numbers are p-values. Those in the first column were obtained from the $\chi^2$ statistic (11.28), those in the second column from the semi-parametric likelihood ratio statistic (11.26) with the FGM copula (11.20), and those in the two last columns were obtained from the LR statistic according to (11.15) with beta marginal (11.17) and FGM copula (11.20) and Gaussian marginal (11.22) with Gauss copula (11.18). For further discussions see p. 197.

Figure 11.5 and Figure 11.6 show the non-parametric evaluation plots according to Section 11.4.2 for the 10-year rates of CHF and USD, respectively. They can be considered as representative for the other rates within the respective economy. The results are not as devastating as the ones in Section 11.5.1: the values of the PIT transforms $z_i$ cover the full interval $[0, 1]$, and for CHF, the null hypothesis of uniformity can clearly not be rejected. Moreover, autocorrelations for the $z_i$'s are only weakly significant. However, there are significant autocorrelations for the second and fourth powers of the $z_i$'s, indicating that the model does not properly reflect the conditional variance of the sample. This is not very astonishing given the rather modest fit of the related moment conditions according to Figure 10.6.

The results of the (semi-)parametric evaluations according to Section 11.4.1 are presented in Table 11.3. Based on the experience from Section 11.5.1, we only consider the Farlie-Gumbel-Morgenstern (F-G-M) copula for the dependence, the $\beta$-distribution for the marginals and the likelihood ratio statistic for inference. The results of the Berkowitz test (see [11]; equivalent to a parametric setup with a Gaussian copula and transformed
Figure 11.6: Non-parametric evaluation plot according to Section 11.4.2 for the USD 10-year rate in the backtest setup of Section 10.8.2. The histogram suggests rejection w.r.t. the marginal distribution. The rejection of the intertemporal dependence structure is rather clear for the even moments, but less so for the odd ones; see p. 193 for a detailed discussion.
Gaussian marginals) is added for the sake of completeness and comparison to this special case.

The general observation from Table 11.3 is that the part pertaining to CHF obtains better results than the one pertaining to USD, which goes in line with the intuition obtained from Section 10.8.2. The tests for the dependence structure produce more or less affirmative results, whereas the picture for the joint hypothesis is more mixed. In particular, the high $p$-values for some the USD rates are somewhat counter-intuitive. This may be due to the general problems with MLE discussed in Section 11.5.2. In very general, the evaluation neither provides a strong corroboration of the calibrated model, nor does it produce such a blunt rejection as in Section 11.5.2.
Seite Leer / Blank leaf
Chapter 12

Case Study – Funds Withheld Asset

12.1 Introduction

In order to illustrate the use of the models introduced in this thesis, we present here an application to a fictitious reinsurance deal. Any resemblance between this example and authentic reinsurance deals is neither intentional nor accidental, but inevitable. This type of deal is called Funds Withheld, and its use is quite commonplace in the reinsurance markets. The deal has two involved parties, namely the Cedent and the Reinsurer. The Cedent has some loss reserves that must be removed from its balance sheet. To this end, these reserves are ceded to the Reinsurer on a 100%-quota share basis. This is generally known as a Loss Portfolio Transfer (LPT) in ART, see e.g. [113]. The premium that the reinsurer receives for accepting the liabilities corresponds to the NPV of the losses plus expenses plus risk premia depending on what risks are included.

There are, however, two problems in place. On the one hand, the Cedent is in dire straits – the reason behind the need for balance sheet relief – and does not have the cash to pay the premium up front. The assets are there,
but they cannot instantly be liquidated at reasonable conditions. Hence, the Cedent is not insolvent, but temporarily illiquid. This no real problem, since the liabilities are long-tailed and the actual payments are far in the future when liquidity is – hopefully – restored. On the other hand, the Reinsurer has a less-than-optimal credit rating so that this LPT – which we assume to be of quite substantial size – represents a non-negligible credit risk for the Cedent which is even more accentuated by the fact that rating agencies also penalize concentration risk. The solution to both of these problems is that the parties agree to set up a Funds Withheld Asset (FWA) structure, which works as outlined hereafter.

Everything takes place at discrete time intervals of length $\Delta t > 0$. For simplicity, we let $t \in \mathbb{N}_0$ denote the number of time intervals, i.e. the actual time on the continuous scale is $t\Delta t$. The outset of the deal is at $t = 0$. We also assume that there is a cutoff time, denoted $T$, at which the deal is redeemed.

There is an ultimate loss amount of $X_u$ assumed to be net of expenses. The payout pattern of the losses is denoted by $(p(t))_{t=1}^{T}$, where each $p(t)$ denotes the percentage of the initial reserve that is paid in the $t$-th time interval $[(t - 1)\Delta t, t\Delta t)$. Hence, the cash outflow in the interval is $p(t)X_u$, and the cash outflow up to time $t$ is $X_u \sum_{s=1}^{t} p(s)$.

At the outset, the Cedent sets up a notional account to the benefit of the Reinsurer, with initial balance $B(0)$. Then, in each time period $t$, the Cedent makes the claims payments due, $p(t)X_u$. Now, instead of receiving the reinsurance recoverables from the 100%-quota share, the paid amount is deducted from the FWA account balance – as long as the latter is non-negative. Hence, the account balance $B(t)$ behaves like

$$B(t) = \max (B(t-1) - p(t)X_u, 0) = \max \left( B(0) - X_u \sum_{s=1}^{t} p(t), 0 \right).$$

If, for some reason, there are still claims payments due when the account balance $B(t)$ is already at zero, then the Reinsurer has to inject these claims payments. These cash injections are denoted by $I(t)$, and they
12.1. Introduction

amount to

\[
I(t) = \begin{cases} 
0 & \text{if } B(t-1) \geq p(t)X_u, \\
p(t)X_u - B(t-1) & \text{if } p(t)X_u > B(t-1) > 0, \\
p(t)X_u & \text{if } B(t-1) = 0.
\end{cases}
\] (12.2)

At the end of each time period \( t \), the Cedent pays to the Reinsurer a **coupon** at a fixed coupon rate \( r \) on the current balance \( B(t) \) of the account. Moreover, at the cutoff date \( T \), the final balance \( B(T) \) is paid out to the Reinsurer. The cash flows \( A(t) \) to the reinsurer therefore amount to

\[
A(t) = \begin{cases} 
rb(t) & \text{if } t < T, \\
(1 + r)B(t) & \text{if } t = T.
\end{cases}
\] (12.3)

As long as there is no uncertainty in the loss amount \( X_u \) or in the loss development \((p(t))_{t=1}^T\), this setup is fully deterministic. We will, however, deviate from this assumption.

Things get further complicated by the fact that there are \( N \) currency zones, denoted by superscripts \((1), \ldots, (N)\). The setup introduced above is in place for each one of these currency zones. So, there is a set of variables \( B(t)^{(i)}, I(t)^{(i)}, p(t)^{(i)}, r^{(i)}, \) and \( A(t)^{(i)} \) for each currency zone \((i)\). Currency \((1)\) is the reporting currency of the Reinsurer.

Specifically, we have two currency zones, namely the USD \((1)\), which is also the reporting currency, and the CHF \((2)\), which is the foreign currency. The time horizon \( T \) is 10 years, and the unit time interval \( \Delta t \) is one year. The detailed settings are presented in Table 12.1, where one can also find a number of summary statistics, in particular the **Mean Time to Payment (MTP)**, defined as

\[
\text{MTP} = \frac{\sum_{t=1}^{T} t p(t) X_u}{\sum_{t=1}^{T} p(t) X_u},
\] (12.4)

the **MacAulay Duration** \( D_M \), defined as

\[
D_M = \frac{\sum_{t=1}^{T} t p(t) X_u e^{-tR(0,t)}}{\sum_{t=1}^{T} p(t) X_u e^{-tR(0,t)}},
\] (12.5)
The Net Present Value in Local Currency \(NPV_{LC}\), defined as

\[
NPV_{LC} = \sum_{t=1}^{T} p(t) X_u e^{-tR(0,t)},
\]

and the Net Present Value translated into the reporting currency, which is simply \(NPV_{LC} \cdot C_{12}(0)\). One can see in Table 12.1 that the payout pattern for the US, \(\{p^{(1)}(t)\}_{t=1}^{T}\), develops relatively quickly. The MTP and duration are relatively short, and the expected NPV is not far below the expected ultimate loss. On the other hand, the Swiss payout pattern, \(\{p^{(2)}(t)\}_{t=1}^{T}\), is relatively slow, resulting in long MTP and duration and an expected NPV well below the expected ultimate loss. The summary statistics suggest that acceptable coupon rates would be around \(r^{(1)} = 2.81\%\) for USD and \(r^{(2)} = 2.24\%\) for CHF. Initial account balances should be above the NPV’s, i.e. \(B^{(1)}(0) \geq 92\) and \(B^{(1)}(0) \geq 86\).
12.2. Some Projections

We will also allow for the loss amounts $X_u^{(i)}$ to be random. We will, however, always assume that $X_u^{(1)}$ and $X_u^{(2)}$ are independent of each other and of interest rates and foreign exchange rates. Specifically, we will use two models for $X_u^{(i)}$, namely the Pareto distribution and the log-normal distribution. The Pareto distribution has CDF

$$F(x) = 1 - \left(\frac{D + \beta}{x + \beta}\right)^\alpha, \quad x \geq D, \alpha > 0, \beta > -D. \quad (12.7)$$

The mean (provided $\alpha > 1$) and the variance (provided $\alpha > 2$) are given by

$$\mathbb{E}[X] = \frac{\alpha D + \beta}{\alpha - 1} \quad \text{and} \quad \text{Var}[X] = \frac{\alpha(D + \beta)^2}{(\alpha - 1)^2(\alpha - 2)^2}. \quad (12.8)$$

More details can be found in [41] and [52]. In our specific setting, we assume for both $X_u^{(i)}$ that $B = 0$ (no threshold), $\alpha = 4$ (moderately heavy tails), and $\beta = 300$ (such that $\mathbb{E}[X_u^{(i)}] = 100$). Note that Matlab does not offer a function for sampling from a Pareto distribution. However, this was easily implemented by using the well-known quantile transform: If $U$ has a uniform distribution on the unit interval, then $F_1(U)$, where $F(.)$ is the CDF from (12.7) has a Pareto distribution.

On the other hand, a random variable $X$ has a translated log-normal distribution if it can be expressed in the form $X = d + e^Y$, where $Y \sim \mathcal{N}(\mu, \sigma^2)$. For the moments, we have

$$\mathbb{E}[X] = e^\mu e^{\sigma^2/2} + d \quad \text{and} \quad \text{Var}[X] = e^{2\mu} e^{\sigma^2} \left(e^{\sigma^2} - 1\right); \quad (12.9)$$

see [41]. For our study, we let $d = 0$, $\mu = 4.2586$ and $\sigma = 0.8326$, so as to have the same mean and variance as in the Pareto case. We note, however, that the log-normal distribution is thin-tailed; so we have less extreme expected shortfall values.

12.2 Some Projections

During the entire case study, we take the point of view of the Reinsurer. From this point of view, the deal is represented by the stream of cash
Chapter 12. Case Study – Funds Withheld Asset

flows \((K(t))_{t=1}^T\), where a cash flow \(K(t)\) is given by \(K(t) = A(t) - I(t)\) with \(A(t)\) according to (12.3) and \(I(t)\) according to (12.2). The deal is acceptable to the reinsurer, if the value – where the exact definition of this term remains to be specified – of the cash flow sequence \((K(t))_{t=1}^T\) is positive. The adjustable parameters of the deal are the coupon rate \(r\) and the initial account balance \(B(0)\). We assume, however, that the coupon rate is set equal to the market interest rate at the duration of the payout pattern, as described in Section 12.1. So, the question is: *What is the minimal amount for the initial balance \(B(0)\) such that the deal is acceptable for the reinsurer?* Note that we always assume that the initial account balance is set up in local currency.

We start by simulating and evaluating the setup as given in Section 12.1 in a typical DFA manner. We consider the cash flow \(K(t)\) for each single time point \(t \in \{1, \ldots, T\}\). The distribution of \(K(t)\) is characterized by the mean as a measure for gain and by the Expected Shortfall at the 1%-level as a measure for risk. For CHF, we distinguish two cases:

1. All amounts translated into USD at the fixed rate \(C_{ij}(0)\). Economically, this corresponds to a full currency hedge (and is compatible with the valuation principle in (7.27)).

2. Cash flows are translated into USD at simulated current spot rates. This corresponds to the assumption that the exchange rate risk is completely unhedged and enters the overall risk of the reinsurer.

On top of the period-wise evaluations, we also consider the NPV of the cash flow stream \((K(t))_{t=1}^T\), that is

\[
\text{NPV} = \sum_{t=1}^{T} K(t)P(0,t),
\]

(12.10)

where \(P(0,t)\) is the ZCB price corresponding to the given initial term structure. This NPV is itself a random variable and evaluated through mean and Expected Shortfall. Note that, for the case of random FX rates, this NPV computation is incompatible with the valuation principle from (7.27) and corresponds to the implicit assumption that the FX risk is completely unhedged and carried by the reinsurer. In insurance settings, this
12.2. Some Projections

Figure 12.1: Results of the cash flow projections for the three settings: USD (top), CHF translated into USD at fixed rate (middle) and CHF translated into USD (bottom). The panels on the left show the mean (solid) and Expected Shortfall (dotted) of the cash flows $K(t)$ for each single year. The panels on the right show the density of the NPV distribution. The vertical lines indicate mean (solid), VaR (dashed) and Expected Shortfall (dotted). For further discussions see p. 205.

may not be too unrealistic an assumption; see [15] for further explanations.

The results of these projections, for which we have assumed initial ac-
count balances $B(0)(1) = 150$ USD and $B(0)(2) = 150$ CHF, are shown in Figure 12.1. The expected values of the cash flows $K(t)$ for $t < T$ are so low that they are barely visible on the selected scale. Only at $t = T$, there is a substantial expected positive cashflow, since the initial balance was chosen higher that the expected loss. On the other hand, the Expected Shortfall values are considerable due to the heavy-tailed nature of the loss distribution. The NPV distributions have positive means due to the sufficiently high initial balances, but rather long left tails as indicated by the high Expected shortfall values. The stochastic FX rates have a visible impact in that the NPV distribution is is less peaked and has a higher expected shortfall.

Risk for the reinsurer emanates from the cash injections $I(t)$ according to (12.2) that can result in rather hefty shortfalls in the annual cash flows $K(t)$ as can be seen from Figure 12.1. The reinsurer has to set up risk capital in order to cover such cash injections if they arise. We account for this risk here by using a somewhat simplified procedure: The risk capital level is set corresponding to the 1%-Expected Shortfall of the NPV distribution. However, since we are in a multi-period setup, we also have to account for the average holding period of the risk capital, which is represented by the time factor

$$\tau = \mathbb{E} \left[ \sum_{t=1}^{T} t \cdot I(t) \right] \left/ \sum_{t=1}^{T} I(t) \right].$$

(12.11)

Then, we can define the Risk Adjusted Value (RAV) of the deal as

$$RAV = \mathbb{E} \left[ NPV \right] - \tau r_C \mathbb{E} \left[ ES_\alpha \left[ NPV \right] \right],$$

(12.12)

where $r_C$ is the cost of capital of the reinsurer, assumed to be 15% here. This RAV expresses the actuarial value of the deal in a single number.

In Figure 12.2, we can see the RAV for all three settings as a function of the initial account balances $B(1)(0)$ in order to explore the profitable range of the deal. Note that the initial CHF balance $B(2)(0)$ was translated to the USD scale at the initial FX rate $C_{12}(0)$. For USD, the deal is profitable if the initial account balance $B(1)(0)$ is above approximately

...
190 USD. For CHF, the threshold is around $B^{(1)}(0) = 200$ CHF, corresponding to approximately 158 USD at $C_{12}(0)$ for both fixed and random FX rates. This approximate equality is by coincidence; one can also see that for higher initial account balances, the profitability of the deal with random FX rates is clearly higher. We can see that the cost of the risk capital has a quite considerable impact: while the average NPV of the USD deal at $B^{(1)}(0) = 150$ USD is positive (see Figure 12.1), the RAV is still negative.

**Figure 12.2:** Risk-Adjusted Values (RAV) according to (12.12) as a function of the initial balances $B^{(1)}(0)$ (stated in USD) for the three setups: USD (marked by *), CHF at fixed FX rate (marked by +), and CHF at random FX rate (marked by ×). For a discussion see p. 206.
12.3 Valuation and Embedded Options

After the actuarial evaluations in Section 12.2, we are now going to take a strictly financial point of view. The differences between these two paradigms are explained in [50], and this section together with Section 12.2 provides a related illustration.

The intention here is to assign a value to the FWA deal that is consistent with the financial market as modelled by the interest rate and FX rate model. To this end, we have to re-express the FWA as set forth in Section 12.1 in terms of a portfolio of securities from the complete market spanned by the interest rate and FX rate model.

To avoid complications, we assume that the payout pattern is non-decreasing, that is, \( p(t) \geq 0 \) for all \( t \). We also assume that \( B(0) \) is given. Until further notice, we condition on \( X_u \). Figure 12.4 then shows a typical trajectory of the account balance, and it becomes clear that the latter can be replicated by a portfolio in which, for each \( t \in \{1, \ldots, T\} \), we hold

1. a long position of \( B(t - 1) - B(t) \) bonds with principal 1 and annual coupons at rate \( r \), as given in Section 12.1, maturing at time \( t \),

2. a long position of \( B(T) \) bonds with principal 1 and annual coupons at rate \( r \), maturing at time \( T \), and

3. a short position of \( p(t)X_u \) units of a ZCB maturing at time \( t \).

Note that the principals of the coupon bonds are offset by the short positions representing the payments. The reinsurer only benefits from the coupon payments and from an eventual positive final balance \( B(T) \). The value of the deal, given \( X_u \), for the reinsurer is

\[
\pi(0, \text{FWA}|X_u) = B(T)\pi_r(0,T) + \sum_{t=1}^{T} (B(t-1) - B(t)) \pi_r(0,t) - p(t)X_uP(0,t),
\]

(12.13)
where \( \pi_r(0,t) \) denotes the time-0 value of the bond with principal 1 and annual coupon at rate \( r \) maturing at time \( t \). So, since \( X_u \) is random, \( \pi(0,\text{FWA}|X_u) \) is still a random variable as well, but adjusted for all systematic risk emanating from the financial market. Therefore, \( \pi(0,\text{FWA}|X_u) \) can now be treated as an insurance risk without any conflict with financial valuation paradigms, e.g. along the lines of the risk capital-based principle in Section 12.2, or any other viable actuarial valuation principle; see [41]. The financial adjustment from (12.13) essentially consists of the valuation of a few fairly standardized securities.

As long as we only consider the basic setup from Section 12.1, the valuation according to (12.13) is equivalent to the crude discounting done in Section 12.2 as becomes clear from the valuation principles in Section 7.3. The real power of the method only becomes visible when we introduce some additional features into the deal:

**Case 1.** Instead of a fixed coupon rate \( r \), the parties agree on a *floating rate*, equal to the one-year rate, say. That is, the coupon bond in the replicating portfolio above is replaced by a floating-rate note that pays \( R(s, s + 1) \) for \( s = 1, \ldots, t - 1 \) and \( (1 + R(t, t + 1)) \) at maturity \( t \). The time-zero value of such a floating-rate note maturing at time \( t \) is denoted by \( \pi_R(0,t) \).

**Case 2.** The Cedent has the option to *call* the agreement fully or partly at any time \( t = 1, \ldots, 10 \). That is, the Cedent pays the account balance, or a part of it, to the Reinsurer. Note that this does not affect the Reinsurer’s obligation to pay the losses; it just means that the Cedent can avoid the coupon payments and leave the investment of the funds to the Reinsurer. The standard coupon bond in the replicating portfolio above is replaced by a *callable coupon bond* (American type), the value of which is denoted by \( \pi_C(0,t) \).

**Case 3.** The Reinsurer has the option to *put* the agreement fully or partly at any time \( t = 1, \ldots, 10 \). That is, the Reinsurer requires reimbursement of the account balance, or a part of it, from the Cedent. Note
that this does not affect the Reinsurer’s obligation to pay the losses; it just gives the reinsurer the opportunity to earn a higher interest than the coupon payments. The standard coupon bond in the replicating portfolio above is replaced by a puttable coupon bond (American type), the value of which is denoted by $\pi_p(0,t)$.

The structure of the valuation allows us to compute the values of the valuation units $\pi_R(0,t)$, $\pi_C(0,t)$ and $\pi_P(0,t)$ separately before putting them into (12.13). $\pi_R(0,t)$ can be obtained through the basic valuation principle in (7.24), whereas $\pi_C(0,t)$ and $\pi_P(0,t)$ can be obtained by using the algorithm from Section 7.3.2. The results are shown in Table 12.2. To understand the results, it is useful to consider the initial yield curves in Figure 10.6 and the summary statistics of the scenarios in Figure 10.7. The USD coupon rate is generally below market rates, so that the coupon bond trades at a discount for most maturities. Moreover, the Cedent has no incentive to call the debt, so that the callable bond is not very different from the coupon bond, with the exact opposite applying to the put option of the Reinsurer. The coupon rate for CHF better corresponds to market rates, leading to more balanced values. The spread between the puttable and the callable bonds increases with increasing time horizon due to the impact of volatility.

---

**Table 12.2:** Prices of the different valuation units for all time horizons. Prices are computed by using the diagonal version of the double-currency model according to Section 10.8.2.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$\pi_r(0,t)$</th>
<th>$\pi_R(0,t)$</th>
<th>$\pi_C(0,t)$</th>
<th>$\pi_P(0,t)$</th>
<th>$\pi_r(0,t)$</th>
<th>$\pi_R(0,t)$</th>
<th>$\pi_C(0,t)$</th>
<th>$\pi_P(0,t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0120</td>
<td>1.0057</td>
<td>1.0120</td>
<td>1.0120</td>
<td>1.0098</td>
<td>1.0128</td>
<td>1.0098</td>
<td>1.0098</td>
</tr>
<tr>
<td>2</td>
<td>1.0196</td>
<td>1.0113</td>
<td>1.0165</td>
<td>1.0207</td>
<td>1.0116</td>
<td>1.0237</td>
<td>1.0110</td>
<td>1.0174</td>
</tr>
<tr>
<td>3</td>
<td>1.0213</td>
<td>1.0142</td>
<td>1.0151</td>
<td>1.0259</td>
<td>1.0018</td>
<td>1.0258</td>
<td>1.0017</td>
<td>1.0177</td>
</tr>
<tr>
<td>4</td>
<td>1.0213</td>
<td>1.0180</td>
<td>1.0127</td>
<td>1.0319</td>
<td>0.9902</td>
<td>1.0280</td>
<td>0.9910</td>
<td>1.0191</td>
</tr>
<tr>
<td>5</td>
<td>1.0169</td>
<td>1.0194</td>
<td>1.0065</td>
<td>1.0351</td>
<td>0.9764</td>
<td>1.0298</td>
<td>0.9763</td>
<td>1.0206</td>
</tr>
<tr>
<td>6</td>
<td>1.0128</td>
<td>1.0232</td>
<td>1.0011</td>
<td>1.0404</td>
<td>0.9653</td>
<td>1.0369</td>
<td>0.9648</td>
<td>1.0275</td>
</tr>
<tr>
<td>7</td>
<td>1.0056</td>
<td>1.0253</td>
<td>0.9930</td>
<td>1.0438</td>
<td>0.9523</td>
<td>1.0434</td>
<td>0.9515</td>
<td>1.0340</td>
</tr>
<tr>
<td>8</td>
<td>0.9953</td>
<td>1.0252</td>
<td>0.9822</td>
<td>1.0449</td>
<td>0.9317</td>
<td>1.0433</td>
<td>0.9309</td>
<td>1.0338</td>
</tr>
<tr>
<td>9</td>
<td>0.9868</td>
<td>1.0280</td>
<td>0.9733</td>
<td>1.0487</td>
<td>0.9201</td>
<td>1.0541</td>
<td>0.9190</td>
<td>1.0445</td>
</tr>
<tr>
<td>10</td>
<td>0.9764</td>
<td>1.0294</td>
<td>0.9624</td>
<td>1.0509</td>
<td>0.8955</td>
<td>1.0510</td>
<td>0.8943</td>
<td>1.0413</td>
</tr>
</tbody>
</table>
12.3. Valuation and Embedded Options

Figure 12.3: Risk-adjusted values of the deal (for CHF only) under the different embedded interest rate options as specified on p. 209. For a discussion of the results see p. 212.

If one of the aforementioned optional features is present in the deal, then \( \pi_R(0,t) \), \( \pi_C(0,t) \) or \( \pi_P(0,t) \) can be plugged into (12.13) instead of \( \pi(t,t) \) so as to obtain the financially adjusted value given \( X_u \), that is, \( \pi(0,FWA|X_u) \). This is still a random variable, but adjusted for all risk emanating from the financial market. Hence, it can be treated by a purely actuarial, that is risk capital-based, valuation principle. We opt here for the one used in Section 12.2 which translates into

\[
\text{RAV} = E[\pi(0,FWA|X_u)] - \tau \text{ES}_\alpha[\pi(0,FWA|X_u)].
\]  

(12.14)

The evaluation was done in the same way as in Section 12.2, that is, the risk-adjusted value was computed for different values of the initial account balance \( B(0) \), and it was repeated for each one of the four investment regimes, that is, fixed coupon, floating coupon, put option of the Reinsurer and call option for the Cedent. Attention was restricted to CHF since the long-tailed payout pattern (see Table 12.1) provides more leeway for the investment component to have an influence. For the same
reason, the thin-tailed log-normal distribution was used for $X_u$ rather than the heavy-tailed Pareto distribution; see Section 12.2.

The results of the evaluation are presented in Figure 12.3. The fixed-coupon regime is the base case and corresponds to the evaluation in Figure 12.2, but with different values due to the thin-tailed loss distribution and the related lower risk capital costs. The break-even value for $B(0)$ under this regime is at 132.67 CHF. If the Reinsurer has the put option to take advantage of more favourable market interest rates, then the break-even value decreases to 129.13 CHF. On the other hand, if the Cedent has the put option to avoid above-market coupon payments, the break-even value increases to 133.63 CHF. Since, as discussed above, the coupon rate is slightly below market rates, the floating-rate arrangement is more favourable than the fixed-rate one, with a break-even value for $B(0)$ of 130.51 CHF. Moreover, the more the initial balance $B(0)$ is above break-even, the higher is the impact of the investment aspect. For instance, if $B(0) = 150$ CHF, then the risk-adjusted values are at 26.03 CHF for fixed coupons, 30.68 for floating-rate coupons, 24.20 CHF for the callable bond and 33.50 CHF for the puttable one, so the relative impact is quite considerable.

This example shows that, even in deals that are dominated by underwriting risk, embedded financial options can have a non-negligible impact on the total value. This would be even more accentuated in contracts that are dominated by the investment component, like participating life insurance policies. We have shown here a simple, tractable and flexible method for disentangling underwriting and investment risk and for valuating deals in a way compatible with both financial and actuarial principles.

### 12.4 Currency hedging

Recall that the Reinsurer has the USD as its reporting currency. However, one part of the FWA deal is denominated in CHF, thus exposing
12.4. Currency hedging

Figure 12.4: A typical trajectory of the account balance \( B(t) \) for \( t = 1 \) according to (12.1) of the FWA (horizontal lines). The respective lines beyond \( t = 5 \) are dashed because the effective balance according to (12.1) is zero. The downward arrows represent claims payments made.

the Reinsurer to FX rate risk. For the following considerations, we focus exclusively on the CHF part of the deal, and we take the point of view of the Reinsurer. Notice that the initial account balance \( B(0) \) is denominated in CHF, so that there is no FX rate risk for all payments made from this balance. There is, however, FX rate risk for the cash injections \( I(t) \) according to (12.2) (cash outflows) and for the coupons and principal repayments \( A(t) \) according to (12.3) (cash inflows). For the cash outflows, a rising FX rate is an adverse development, whereas for the cash inflows, a falling FX rate is an adverse development.

In order to get an idea of the impact of the FX rate \( C_{12}(t) \) on the NPV of the deal, we first perform a small drill-down study in the sense of Section 3.3. We simulate the CHF-related part of the deal with free-floating and unhedged FX rate and identify the 1% best and the 1% worst outcomes of the NPV. Then we identify the values of the ultimate loss \( X_u \) and the FX rate \( C_{12}(T) \) associated with these outcomes. Note that we use
the final value $C_{12}(T)$ of the FX rate as a simple one-dimensional proxy for the trajectory of the FX rate over time. The results are shown in Figure 12.5. We can see a very clear association between $X_u$ and the NPV for both ends of the range. For the FX rate, the situation is more subtle: the FX rate values associated with the 1% worst outcomes are scattered over the full range of outcomes, whereas the 1% best outcomes are quite clearly associated with high exchange rates, which have a favourable impact on the relatively high cash inflows $A(t)$ of the Reinsurer in cases of low ultimate losses.

In Section 12.2, we have already considered the cases where the FX rate
Figure 12.6: Mean (black bars) and Expected Shortfall (gray bars) of the NPV and RAV (white bars) as a function of the hedge ratio for different initial account balances $B(0)$. For a discussion see p. 215.

is considered as fixed at the initial rate $C_{12}(t) = 0.79$ USD/CHF, and the case where the exchange rate is freely floating according to the model. The former case corresponds to the situation where the Reinsurer holds funds in CHF and credits and debits all cash flows to these funds. In this way, the Reinsurer has a full currency hedge. The latter case corresponds to the situation where the Reinsurer has no funds in CHF and all CHF cash flows are converted to and from USD at current market FX rates. That is, there is no currency hedge at all.

Following [15], we assume now that the reinsurer converts $x\%$ of each cash flow at the fixed FX rate $C_{12}(0)$, and the remaining $(1 - x)\%$ at the market FX rate $C_{12}(t)$. This is a partial currency hedge, and $x$ is the hedge
Table 12.3: Values of European put and call options on one CHF for different strike prices and different times to maturity (in years).

ratio. The fixed-rate case above corresponds to \( x = 100\% \), whereas the floating-rate case corresponds to \( x = 0\% \). Figure 12.6 shows mean and expected shortfall of the NPV distribution and the risk-adjusted values as a function of the hedge ratio for some selected values of the initial account balance \( B(0) \). All computations are equivalent to the ones in Section 12.2. We can see that the expected shortfall always decreases with increasing hedge ratio due to the reduced influence of the exchange rate volatility. For low values of \( B(0) \), the expected shortfall and the risk-adjusted value tend to grow better with increasing hedge ratio, whereas for high values of \( B(0) \), the converse tendency prevails. This is for two reasons: on the one hand, the exchange rates produced by the model tend to increase over time reflecting market views as of late 2004; see Section 10.8.2. On the other hand, cash inflows start to dominate cash outflows for higher values of \( B(0) \), so that an increasing FX rate has a favourable impact. Therefore, a high hedge ratio is optimal for low values of \( B(0) \), whereas a low hedge ratio is favourable for high values of \( B(0) \).

Alternatively, the Reinsurer can also hedge the exchange rate risk by using derivatives. A cash outflow \( I(t) \) occurring at time \( t \) can be hedged against increasing FX rates by a long position of \( I(t) \) European call options maturing at \( t \) on one CHF with a strike price of \( C_c \) representing the maximum acceptable FX rate. Conversely, a cash inflow \( A(t) \) can be hedged against declining exchange rates by buying \( A(t) \) European put options maturing at \( t \) on one CHF with strike price \( C_p \) representing the
12.4. Currency hedging

Figure 12.7: Risk-adjusted values (RAV) as a function of the initial account balance $B(0)$ for different currency hedging regimes: no hedging ($\cdot$), hedging of cash inflow by put options ($+$), hedging of cash outflows by call options ($*$) and hedging of both cash inflows and outflows ($\times$). For a discussion see p. 217.

lowest acceptable exchange rate. Values of these options can easily be obtained by plugging the contingent cash flows $K_c = \max(C_{12}(t) - C_c, 0)$ and $K_p = \max(C_p - C_{12}(t), 0)$ into the valuation principle according to (7.23).

Option values for all maturities from one to ten years and different strike prices are shown in Table 12.3. Due to the increasing trend in the exchange rate, call option values tend to become fairly high, whereas put options are not assigned considerable values. Therefore, a hedge of the cash outflows will be rather expensive.
Assume now that the options for the currency hedge are embedded into the deal between the Cedent and the Reinsurer. That is, the Cedent grants the Reinsurer an exchange rate cap on all cash outflows and an exchange rate floor on all cash inflows. This modified deal can be evaluated along the same lines as in Section 12.3. Figure 12.7 shows the results of this evaluation for the case of $C_c = C_p = 0.79$ USD/CHF, that is, both cash inflows and outflows are hedged at the initial exchange rate. The base case is the variant without embedded options. All embedded options represent an improvement with respect to the base case in the sense of lower initial balances for break-even and higher risk-adjusted values for the same initial balance. As suggested by the values in Table 12.3, the impact of the put (floor for cash inflows) is not very strong, contrary to the impact of the call (cap for cash outflows).

This example emphasizes once more that embedded derivatives can have a considerable impact on the value of a deal, and that the price kernel method used within this thesis is an efficient and versatile tool for their valuation.
Chapter 13

Assessment and Outlook

13.1 Assessment of Present State

In Part II of this thesis, we have gone through an entire, engineering-like development process for an interest rate and FX rate model for DFA. We started with a set of requirements and we have covered all relevant theoretic and practical aspects so as to arrive at a fully operational and transparent model that is ready for use in industrial applications. None of the currently known interest rate and FX rate models for DFA is documented to such a degree of completeness. Therefore, apart from providing a new model, (a streamlined version of) this thesis may also serve as an engineering guide for other model development projects. Most of the elements used throughout the development had already been known before, but they were adapted and combined in new ways to achieve pre-specified goals. Where necessary, new features and components were added. The sequel of this section summarises the most important findings made along the development path, whereas the next section outlines some future perspectives.

From among the different interest rate modelling approaches, the price kernel approach was found to be best suited under the given requirements,
that is:

- It allows multi-economy modelling in a natural way.
- It can produce positive interest rates.
- It allows for the valuation of fairly general embedded derivatives under the physical measure, so that it integrates seamlessly into a wider simulation context.
- The equilibrium structure yields numerical stability.

Within the price kernel approach, there are two methods to develop specific models, namely the Flesaker and Hughston method and the potential method. The former method was fully worked out in the context of DFA in [23] and [24]. This thesis brings the latter method, i.e. the potential method, to the same maturity.

The price kernel was complemented by a new time-inhomogeneous extension that fully incorporates the information contained in the initial yield curve. Thus, the calibration here is different from the usual ones in that it incorporates both historical dynamics and current market expectations.

Within the potential approach, the attention is limited to exponential-quadratic models which generalize the better-known affine models by adding a quadratic term and represent the most general sensible step along this path. The exponential-quadratic model also proved to be very tractable. Under mild regularity conditions, there exist closed-form expressions for all relevant quantities, and even when these conditions are dropped, the numerical computations are limited to a narrow part of the model, contrary to models of the affine class.

For the calibration, the method of choice was GMM, because of its flexibility and tractability. Moreover, the exponential-quadratic class allowed to compute moment conditions in abundance. Some aspects that are different from the general GMM literature required special attention. First, since an unobservable state process was involved, moment conditions had
to be limited to unconditional ones, and suitable proxies had to be found for the conditional dynamics. Finally, an additional optimisation step for finding suitable initial values for the hidden state process had to be added. Moreover, optimisation had to be done under a number of constraints that assure the desirable properties of the model. In particular, the required positive definiteness of certain matrices proved to be a major challenge. The notorious dependence of the optimisation results on the initial values was circumvented by adding a global search for suitable initial values based on an adapted version of the genetic algorithm. The standard test statistics that GMM offers to assess the goodness-of-fit of the model were found to be of little use, and related improvements remain an open issue; see Section 13.2. Over all, however, the adapted GMM allowed to find parameter values that produce credible scenarios and that exhibit a reasonable performance in some restrictive backtests.

The forecasting capability of a calibrated model still needs to be validated before it can be used in practice. In the context of DFA, this means the evaluation of full distribution forecasts. One of the aims of this thesis was to provide generic validation methods that can be used for a wide range of models and variables. The method of choice is the Probability Integral Transform (PIT) that was introduced into finance by [45]. The approach was extended in such a way that it can also deal with Monte Carlo scenarios rather than closed-form forecasts. For the evaluation of PIT output, the non-parametric approach from [45] was refined, and a recent idea from [32] for parametric tests based on copula methods was implemented and evaluated in practice. While proving the general usefulness, the investigations also revealed some problems, mainly related to the lack of sufficient amounts of data and to the sensitivity of the maximum likelihood estimation against mis-specifications.

Finally, the calibrated and validated model was put at work in a setup that is representative for the valuation problems that insurers will face under fair value accounting and Solvency II, involving both financial and non-financial risk. Besides classical DFA-type projections, the model can also cope with the valuation of embedded options compatible with established financial principles. Because of the importance of market-consistent
valuation in future actuarial applications, the valuation capabilities of the price kernel were extensively reviewed, including its limitations. Moreover, an adaptation of the Longstaff-Schwartz algorithm for the valuation of American-style derivatives in the price kernel context was developed.

13.2 Outlook on Future Developments

In this section, we identify a number of modifications of the models and methods studied in this thesis that deserve further exploration. We start with a limitation: there is no flexibility for departing from Itô diffusions as the driver process. The resolvent construction in Chapter 8 would, in principle, allow for other, more general, Markov processes. However, absence of arbitrage as shown in Chapter 7 is intimately linked to Brownian motion. Any departure would lead into the realm of incomplete markets, where the full theory as in Chapter 7 has to be re-developed. Moreover, in incomplete markets, we have the problem of a non-unique price kernel, so that the valuation approach falls flat unless we find a striking argument for which price kernel to use. On the other hand, it would be worthwhile to explore more sophisticated forms of the Itô diffusion for the driver process, featuring stochastic volatility, for instance. The evaluations in Section 10.8 and Section 11.5 show that there is some related need.

If we consider double-economy versions of the model, as in Section 10.8.2, it is not written in stone that the two economies must represent two different currencies. One can also think of other settings, for instance: economy (1) is the USD with its non-defaultable term structure, but economy (2) is the US equity market as represented by some equity index. The exchange rate $C_{12}(t)$ according to Proposition 7.10 then represents the USD price of one unit of the equity index.

The exchange rate $C_{12}(t)$ according to (9.27) implied by the exponential-quadratic model is well capable of reproducing the most basic features of equity (index) prices. Due to its structure, it will never produce negative prices. Recalling from (9.41) that the expected value of the log-returns of
the exchange rate is given by

$$E[c_{12}(t, T)] = -(\alpha^{(2)} - \alpha^{(1)}) (T - t),$$

we can see that the model is also able to produce a drift as in the Black-Scholes model. The fact that $C_{12}(t)$ depends on the full driver process allows the modelling of dependence to the interest rates and a richer volatility structure than in the Black-Scholes model. As for the downside, it should be mentioned that heavy tails in equity returns are an issue even on monthly time aggregation, and the current model is not capable of reflecting this. For a basic GMM calibration, we can use moment conditions on the characteristics of equity log-returns and their dependence structure with interest rates as given in Section 9.4.

One can even go one step further: there exists also a term structure of equity returns to which one can fit the full yield curve model for economy (2). For the past, this term structure of equity returns can be constructed from historical returns over different periods if necessary. For the present, it can be inferred from the prices of traded derivatives (see [9]), and it bears information on the market consensus expectations for the future behaviour of the equity market. So, calibrating the model to the initial term structure of equity returns, using the method set forth in Section 10.7 and calculating the correction factors according to Section 8.4, can bring precious additional information into the model. Calibrating the model also to past equity term structures is mainly a means for obtaining more and simpler GMM moment conditions; the sheer size of (9.42) makes a point for this.

Another alternative use of two economies is more popular; see e.g. [23] or [69]: assume that economy (1) represents the ordinary nominal currency and economy (2) represents the real currency (inflation adjusted). Then, the yield curve of economy (1) represents nominal interest rates, the yield curve of economy (2) represents real interest rates, and the exchange rate $C_{12}(t)$ is the consumer price index (CPI). One drawback of this setup is that the model should allow real interest rates to become negative, unlike nominal ones. For some economies (mainly UK, also US), there is a market for inflation-linked bonds, so that we can have data for a full real
yield curve. However, for other economies (like CH or EMU) there is no such market, and we would have to base a calibration of the parameters of economy (2) on CPI data only, which may be problematic.

In calibration, although GMM is generally corroborated as a flexible and tractable method, some refinements appear worthwhile. First of all, better diagnostic tests are a must. An approach based on the bootstrap, as suggested in [72], appears most promising. Then, some more formal criteria for the selection of number and kind of moment conditions would be desirable. The useability of the intuitive approach, advocated in Section 10.4, turned out to be hampered by the sheer number of possible choices available. Finally, some more refined estimation methods of the weighting matrix could further improve the numerical stability of the estimation.

The use of the genetic algorithm was a success. Although it is not very precise, and final optimization by using standard methods cannot be abandoned, the genetic algorithm is a reliable means for finding sensible starting values in complex and non-linear optimization problems, thus solving the big problem of classical optimizers. Hence, there is a wide range of further applications for the genetic algorithm in the realm of DFA.

In the validation part, the PIT method proved to be a powerful generic approach, but there is some need for refinements. In particular, the likelihood-based tests proved to offer little robustness, and more robust test methods could greatly improve the reliability. Also, the investigation of further copulae would be worthwhile. In the specific context of DFA, the lack of sufficient amounts of data is also a problem. For regulatory use, it might prove necessary to develop simpler test criteria.

Finally, driven by IFRS and Solvency II, the really big topic in insurance mathematics, DFA and scenario generation in the near future will be fair or market consistent valuation. Price kernel models like the one developed in this thesis, together with sound methods for calibration and validation, can be a crucial element for fair valuation. However, there
are still many open points as to what the fair value of some general cash flow is. Thorough exploration of these questions and also of the potential and limitations of price kernels in this context must be top priorities on the research agenda. In order to establish fair valuation methodology in the insurance world, and in order to assure that the actuaries can attain and maintain thought leadership in the area, it is particularly important to demonstrate the applicability of the methodology in as many practice-relevant case studies as possible.
Seite Leer / Blank leaf
Appendix A

Statistical Methods

A.1 Descriptive Statistics and Normality Tests

We assume that we are given a sample \( \{X(t)\}_{t=1}^{T} \) of observations that are iid with an unknown underlying distribution \( F \). The most basic tool to obtain an idea of the distribution underlying the observations if the empirical distribution function:

\[
\hat{F}_T(x) = \frac{1}{T} \sum_{t=1}^{T} 1\{X(t) \leq x\}.
\]  

The Glivenko-Cantelli lemma \[83\] assures that \( \hat{F}_T \) is a consistent estimator of the true underlying distribution \( F \). The properties of the stochastic law are generally summarised by looking at the moment of the distribution, namely

- **Mean**: \( \mu = E[X(t)] \) location,
- **Standard Deviation**: \( \sigma = E[(X(t) - \mu)^2]^{1/2} \) dispersion around mean,
- **Skewness**: \( S = E[(X(t) - \mu)^3]/\sigma^3 \) asymmetry around mean,
- **Kurtosis**: \( K = E[(X(t) - \mu)^4]/\sigma^4 \) heavy-tailedness.

As a convention, we are considering here the kurtosis \( K \) rather than the excess kurtosis \( K - 3 \). The empirical counterparts of the above moments
are given by
\[ \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} X(t), \quad \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (X(t) - \hat{\mu})^2, \]
\[ \hat{S} = \frac{1}{T\hat{\sigma}^3} \sum_{t=1}^{T} (X(t) - \hat{\mu})^3, \quad \hat{K} = \frac{1}{T\hat{\sigma}^4} \sum_{t=1}^{T} (X(t) - \hat{\mu})^4. \] (A.2)

By using the one-sample t-test [83] one can test whether the sample mean \( \hat{\mu} \) is significantly different from 0. Moreover, in practical applications, it is also interesting to know the minima and maxima of the samples to see the range of actual outcomes.

For model selection, it is of utmost importance to know whether the observations \( X(t) \) follow a normal distribution. The latter is, among other things, characterized by \( S = 0 \) (symmetry) and a Kurtosis \( K = 3 \). If the \( X(t)'s \) are actually normally distributed, then the estimators for skewness and kurtosis given in (A.2) are asymptotically normally distributed; see [28]:
\[ \hat{S} \sim \mathcal{N}(0, \sqrt{6/T}), \quad \hat{K} \sim \mathcal{N}(3, \sqrt{24/T}). \] (A.3)

So, if significant skewness or kurtosis is present in the data, then, by contraposition, we can also reject normality. In order to have a second source of inference, we also compute the Kolmogorov-Smirnov test (see [115]), which is not based on empirical moments but on the maximum absolute deviation between theoretical and empirical distribution function, and which is readily available through Matlab’s kstest function. A graphical tool for getting an idea of whether the empirical distribution of a sample can be reconciled with some theoretical distribution function \( F \) (e.g. the Gaussian one) is the QQ-plot, which consists of plotting the empirical quantiles against the theoretical ones, i.e.
\[ \left\{ \left( X(t), F^{-1}\left( \frac{T - t + 1}{T + 1} \right) \right); \ t = 1, \ldots, T \right\}, \] (A.4)

where \( X(t) \) is the \( t \)-th order statistic of \( \{X(t)\}_{t=1}^{T} \) in ascending order and \( F^{-1}(x) \) is the generalized inverse. In case \( F \) is continuous and strictly increasing, this is just the inverse \( F^{-1} \). Indications on how to read QQ-plots are given e.g. in [115]. In Matlab, the QQ-plot against the Gaussian
A.2. Intertemporal Dependence Diagnostics

This section introduces a simple toolkit for measuring and testing the significance of intertemporal dependence in some time series \( \{X(t)\}_{t=1}^{T} \) that we assume to be weakly stationary. The autocovariance at lag \( k \) is given by \( \gamma(k) := \text{Cov}[X(t), X(t+k)] \), where \( k = 0, 1, 2, \ldots \), and \( \gamma(0) = \text{Var}[X(t)] \). The real quantity of interest is, however, the autocorrelation at lag \( k \), defined as

\[
\rho(k) := \frac{\text{Cov}[X(t), X(t+k)]}{\sqrt{\text{Var}[X(t)]} \sqrt{\text{Var}[X(t+k)]}} = \frac{\gamma(k)}{\gamma(0)}. \tag{A.5}
\]

We will refer to \( k \to \rho(k) \) as the autocorrelation function (ACF) of the considered return series. Given (at least weak) stationarity, the following estimation of \( \gamma(k) \), and hence \( \rho(k) \) is possible:

\[
\hat{\gamma}(k) = \frac{1}{T} \sum_{t=1}^{T-k} (X(t) - \hat{\mu})(X(t+k) - \hat{\mu}), \tag{A.6}
\]

where \( \hat{\mu} \) is the sample mean. The autocorrelation can then be estimated as

\[
\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}. \tag{A.7}
\]

In order for \( \hat{\rho}(k) \) to remain meaningful, the lag \( k \) should not exceed \( T/4 \) according to [20]; other practitioners take \( \sqrt{T} \) as an upper bound for \( k \). Here, we will generally respect the more conservative of the two bounds. For further analysis, we would also like to have confidence intervals for the ACF estimates, or at least test statistics in order to assess whether \( \hat{\rho}(k) \) is significantly different from zero. The exact distribution of \( \hat{\rho}(k) \) is generally intractable, but there are some asymptotic relations (see [20]): letting \( \hat{\rho} := (\hat{\rho}(1), \ldots, \hat{\rho}(k))' \) we have

\[
\hat{\rho} \sim \mathcal{N}_k(\rho, T^{-1}W), \tag{A.8}
\]
where \( \rho := (\rho(1), \ldots, \rho(k))' \) is the vector of true autocorrelations, and the \((i,j)\)-th element of the matrix \( W \) is given by Bartlett's formula

\[
W_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)] \\
\times [\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)]. \tag{A.9}
\]

Bartlett's formula can only be used practically if we impose some model assumption on the \( X(t) \)'s, otherwise, the infinite series cannot be evaluated; see [20] for a number of examples. The only specific assumption that we pursue here is the random walk hypothesis, i.e. the null hypothesis the the observations \( X(t) \) are uncorrelated with finite variance. Then we have \( \rho(k) = 0 \) for \( k \neq 0 \) and, by A.9, \( w_{ii} = 1 \) and \( w_{ij} = 0 \) whenever \( i \neq j \). Hence, under the null hypothesis, we have

\[
\hat{\rho}(k) \sim N(0, T^{-1/2}), \quad k \neq 0, \tag{A.10}
\]

which is also known as the Kendall-Stuart test. This gives us a simple method for a first assessment of whether the ACF is significantly different from zero. Observing that, under the above null hypothesis, the off-diagonal elements of \( W \) in (A.8) are zero, we immediately obtain

\[
Q(m) = T^2 \sum_{k=1}^{m} \hat{\rho}(k)^2 \sim \chi_m^2, \tag{A.11}
\]

which is known as the Box-Pierce test. A close relative of the Box-Pierce test is the Ljung-Box test, whose test statistic incorporates a small sample correction (see [28]) and should, therefore, be preferred for the sample sizes present in this study.

\[
Q'(m) = T(T+2) \sum_{k=1}^{m} \frac{\hat{\rho}(k)^2}{T-k} \sim \chi_m^2. \tag{A.12}
\]

The problem with both tests is that the number of lags \( m \) must be chosen by the analyst and depends on his/her judgment. It is important to notice that the two tests test for a random walk hypothesis by using autocorrelations up to lag \( m \). A rejection for some \( m \) does not mean that the autocorrelation at this very lag \( m \) is significant, but that all autocorrelations up to this lag make for a significant deviation from the random walk
A.2. Intertemporal Dependence Diagnostics

hypothesis.

One problem with the ACF is that the autocorrelation $\rho_k$ for some lag $k > 1$ is often spurious in the sense that it is actually due to autocorrelations at shorter lags. Recall e.g. the well-known fact that for an AR(1) process $X(t) = \phi X(t - 1) + \varepsilon(t)$ with $|\phi| < 1$ we have $\rho(k) = \phi^k$, i.e. $\rho(k) \neq 0$ for all $k$, although there is only a causal dependence at lag 1. To adjust for this phenomenon, we complement the ACF by the Partial Autocorrelation Function (PACF) $\pi(k)$, where $\pi(0) := 1$, and for any lag $k \neq 0$, $\pi(k)$ is the coefficient $\phi_{kk}$ in the AR($k$) model

$$X(t) = \phi_{k1}X(t - 1) + \cdots + \phi_{kk}x(t - k) + \varepsilon(t). \quad (A.13)$$

The PACF $\pi(k)$ measures the additional correlation at lag $k$ after adjustments have been made for all intervening lags up to $k$. Estimating the PACF essentially amounts to fitting an AR($k$) model to the data for each lag $k$ for which $\pi(k)$ is wanted. The AR($k$)-fitting is most easily done by using Yule-Walker estimation, see e.g. [20]. It is easy to see that for the above-stated AR(1) example, we have $\pi(1) = \phi$ and $\pi(k) = 0$ for all $k > 1$.

ACF and PACF, complemented by some of the test statistics introduced above, represent a very powerful means for obtaining a quick picture of the intertemporal dependence structure of some observed time series. Therefore, a Matlab function was created that summarises all these diagnostic in one figure, the Autocorrelation Diagnostic Plot. The function call is

```
acDiagPlot(x,nlag,alpha).
```

Here, $x$ is a vector of length $T$ containing the observations, $nlag$ is the maximum lag up to which the evaluations are made (default value $\sqrt{T}$), and $alpha$ is the confidence level for the inference (default 5%).

An example of an Autocorrelation Diagnostic Plot is shown in Figure A.1. The dashed horizontal lines delimit the confidence band of the Kendall-Stuart test from (A.10). For each lag, there is a solid vertical line indicating the value of the ACF estimate, and a dotted line indicating the
value of the PACF estimate. If the estimates are significantly different from zero, then the related lines carry a dot at the end. The dots on the horizontal line indicate that the Ljung-Box test from (A.12) rejects its null hypothesis based on information up to this lag.

The Autocorrelation Diagnostic Plot can also be used for model identification: for a wide range of time series models (including ARMA(p,q)), the theoretical behaviour of the ACF and the PACF is known and can be compared with the observed one. [20] and [100] provide full details on theoretical ACF’s and PACF’s. The plot in Figure A.1, for instance, shows the typical pattern of autoregressive processes: exponentially decaying ACF and truncated PACF.
Appendix B

Numerical Methods

B.1 Simulating the Driver Process

The uncertainty in the specific model considered here is modelled by a driver process \( X = (X_t)_{0 \leq t \leq T} \), where \( T \) is some finite time horizon. In general, \( X \) is a \( d \)-dimensional Itô diffusion as specified in Definition 8.1:

\[
dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad X_0 = x.
\]

More specifically, we are interested in the special cases where \( X \) is a multivariate Ornstein-Uhlenbeck process as specified in Definition 9.1, i.e. \( b(x) = -\Gamma x \) and \( \sigma(x) = 1_{d \times d} \). Note that we have so-called diagonal noise, i.e. \( \sigma(x) \) is diagonal. In the given Ornstein-Uhlenbeck case, we even have additive noise, i.e. \( \sigma(x) \) does not at all depend on \( x \).

In a practical implementation, we have to simulate values \( Y_n := X_{t_n} \) for a sequence of valuation times

\[
0 = t_0 < t_1 < \ldots < t_n < \ldots < t_N = T.
\]

We will assume in the sequel that the valuation times are equispaced with time step \( \Delta t > 0 \), i.e.

\[
t_n - t_{n-1} \equiv \Delta t, \quad n = 1, \ldots, N.
\]
We have to distinguish between two cases: If the true conditional distribution of $X_{t_n}$ given $X_{t_{n-1}}$ is known and sufficiently tractable, then we can directly simulate $Y_n = X_{t_n}$ from this true distribution and there are no inaccuracies arising from the discretization as such. In the general case, the conditional distribution of $X_{t_n}$ given $X_{t_{n-1}}$ is not known explicitly; then we have to revert to numerical approximation schemes for SDE's as described in [81]. In this case, we have to deal with inaccuracies arising from the discretization scheme as such.

It turns out that we are faced with the more favorable situation of a known conditional distribution in the case where the driver process is a multivariate Ornstein-Uhlenbeck process. This cases will be dealt with in Section B.1.1, respectively. The general case will be dealt with in Section B.1.2.

### B.1.1 The Ornstein-Uhlenbeck case

If the driver process $X$ is a multivariate Ornstein-Uhlenbeck process as specified in Definition 9.1, then we have

$$X_{t_n} | \mathcal{F}_{t_{n-1}} = X_{t_n} | X_{t_{n-1}} \sim \mathcal{N}_d \left( \mu(X_{t_{n-1}}, \Delta t), V(\Delta t) \right), \quad n = 1, \ldots, N,$$

due to the Markov property and Proposition 9.2, where

$$\mu(X_{t_{n-1}}, \Delta t) = e^{-\Gamma \Delta t} X_{t_{n-1}} \quad \text{and} \quad V(\Delta t) = \int_0^{\Delta t} e^{-\Gamma s} \left( e^{-\Gamma s} \right)' ds.$$

Due to the setup with equispaced valuation times, it suffices to compute $e^{-\Gamma \Delta t}$ and $V(\Delta t)$ once for all time steps and for all simulated scenarios, which greatly increases the efficiency of the simulations. Fast and accurate computation of matrix exponentials is done by using the \texttt{expm} function of Matlab. The integral for $V(\Delta t)$ is computed numerically by using the trapezoid method as described in [111]. A simulated trajectory of $X$ is then computed by

$$X_{t_0} = x \quad \text{and} \quad X_{t_n} = e^{-\Gamma \Delta t} X_{t_{n-1}} + \varepsilon_{t_n}, \quad n = 1, \ldots, N, \quad \text{(B.1)}$$

where

$$\varepsilon_{t_1}, \ldots, \varepsilon_{t_N} \sim \text{iid } \mathcal{N}_d \left( 0, V(\Delta t) \right).$$
The innovations \( \varepsilon_t \) are simulated in the well-known manner: If \( \eta \sim \mathcal{N}_d(0, 1_{d \times d}) \) and \( R \in \mathbb{R}^{d \times d} \) such that \( RR' = V(\Delta t) \) then \( R\eta \sim \mathcal{N}_d(0, V(\Delta t)) \). One out of many possible choices for \( R \) is to let \( R \) be the Choleski factorization of \( V(\Delta t) \). Caveat: In Matlab, we have to set \( R \) equal to the transpose of the result of the \texttt{chol} function.

### B.1.2 The general case

Let now \( X \) be some general Itô diffusion according to Definition 8.1. We select a time step \( \Delta t > 0 \) such that \( T = N\Delta t \) for some \( N \in \mathbb{N} \). The aim is then to construct a sequence \( Y^{\Delta t} = (Y_n)_{n=0}^N \) of random variables such that, for each \( n \), \( Y_n \) is an approximation for \( X_{n\Delta t} \). Following the recommendations of [74], we use a relatively simple approximation, namely the explicit Milstein scheme which is described in detail in [81]. For the \( d \)-dimensional case with diagonal noise, its \( k \)-th component is given by

\[
Y_{n+1}^k = Y_n^k + b^k(Y_n)\Delta t + \sigma^{kk}(Y_n)\sqrt{\Delta t} \varepsilon_{n+1}^k + \frac{1}{2} \sigma^{kk}(Y_n) \frac{\partial \sigma^{kk}}{\partial x^k}(Y_n) \left\{ \Delta t (\varepsilon_{n+1}^k)^2 - \Delta t \right\} \tag{B.2}
\]

for \( n = 0, \ldots, N-1 \) and \( k = 1, \ldots, d \). The approximation always starts at \( Y_0 = X_0 \), and we let

\[ \varepsilon_1, \ldots, \varepsilon_N \sim \mathcal{N}_d(0, 1_{d \times d}). \]

For the case of the multivariate Ornstein-Uhlenbeck process, this boils down to

\[ Y_{n+1} = Y_n - \Gamma Y_n \Delta t + \sqrt{\Delta t} \varepsilon_{n+1}, \tag{B.3} \]

which is equivalent to the simpler Euler scheme. F

Under the regularity conditions prevailing here, the Milstein scheme as introduced in this section is, in the Terminology of [81], an order 1.0 strong Taylor scheme. It then follows (Theorem 10.6.3 and Corollary 10.6.4, op. cit.) that, for some constant \( K \) independent of \( \Delta t \) and \( p \geq 1 \), we have

\[
\left( \mathbb{E} \left[ \sup_{0 \leq n \leq N} \|X_{n\Delta t} - Y_n^{\Delta t}\|^p \right] \right)^{1/p} \leq K\Delta t. \tag{B.4}
\]
There is a fairly broad agreement, see e.g. [74], that the Milstein scheme generally provides a sufficiently accurate approximation for the diffusions in view, and that there is no point in using explicit schemes of higher order. There is, however, one caveat: As in the realm of deterministic differential equations, a stochastic differential equation can be stiff, and in this case, any explicit scheme provides very poor approximations; see Chapter 9.8 of [81]. In order to cope up with stiff SDE's, one has to use either implicit schemes or predictor-corrector schemes.

The real problem with a system of SDE's is to determine whether or not it is actually stiff. Stiffness is defined in terms of the Lyapunov exponents of the system, see Chapter 9.8 of [81], and these Lyapunov exponents are very difficult to compute or approximate in realistic situations (see Chapter 17.3, op. cit.). Therefore, we give here the implicit version of the Milstein scheme, but the question of when it is applicable is left open.

The following equation from Chapter 12.2 of [81] gives the implicit equivalent of (B.2) for the case of diagonal noise

\[ Y_{n+1}^k = Y_n^k + b^k(Y_{n+1}^1)\Delta t + \sigma_{kk}^k(Y_n)\sqrt{\Delta t} \varepsilon_{n+1}^k + \frac{1}{2} \sigma_{kk}^k(Y_n) \frac{\partial \sigma_{kk}^k}{\partial x^k}(Y_n) \left\{ \Delta t (\varepsilon_{n+1}^k)^2 - \Delta t \right\} , \quad (B.2*) \]

where the only difference to the explicit version is that the drift term \( b(x) \) is evaluated at \( Y_{n+1}^1 \) instead of \( Y_n \). For our special case, this boils down to

\[ Y_{n+1} = Y_n - \Gamma Y_{n+1} \Delta t + \sqrt{\Delta t} \varepsilon_{n+1} . \quad (B.3*) \]

Provided sufficient regularity, the fully implicit iteration step is then given by

\[ Y_{n+1} = (I_{d\times d} + \Delta t \Gamma)^{-1} \left( Y_n + \sqrt{\Delta t} \varepsilon_{n+1} \right) . \quad (B.3**) \]

This is fine as long as \( (I_{d\times d} + \Delta t \Gamma) \) is well conditioned. Otherwise, in order to circumvent the numerical problems with the inversion of badly conditioned matrices, one can apply a predictor-corrector scheme which consists of first computing a predictor value \( \tilde{Y}_{n+1} \) by using the explicit iteration step in (B.3), and then inserting \( \tilde{Y}_{n+1} \) into the right-hand side of (B.3*).
B.2. Numerical Optimization Issues

B.1.3 Random Number Generation

The simulation of the driver process involves quite some random number generation. For this task, the implementation entirely relies on the related built-in functions of Matlab which are described in [101]. At the basis is the rand function that produces uniformly distributed random numbers on the unit interval. The underlying algorithm by Marsaglia is specifically aimed at producing random real numbers rather than scaled integers. Internal tests confirmed that Matlab’s rand function is suitable for simulations of the size required here; see [12]. Independent standard Normal vectors are generated by using Matlab’s randn function, which is based on a numerical inversion of an approximation of the Normal CDF, called the Ziggurat algorithm. General Normal vectors are obtained via the usual affine transformation of standard Normal vectors, see Section B.1.1 for details. The Matlab function ncx2rnd is used for generating non-central \( \chi^2 \) numbers; the function is based on summing squared Normal numbers.

B.2 Numerical Optimization Issues

The GMM optimization problem that we have to solve is summarized in (10.7). From Section 10.2, it becomes clear that it is a non-linear optimization problem with constraints, and from Section 9.3 we can see that some of these constraints are even non-linear for the Exponential Quadratic model that is of main interest here. Therefore, we must use the Matlab function fmincon (for function minimize constrained) for implementing the parameter optimization. The function fmincon can address the following generic problem:

\[
\begin{align*}
\min_{x} & \quad f(x) \\
\text{s.t.} & \quad c(x) \leq 0; \quad c_{eq}(x) = 0; \quad A x \leq b; \quad A_{eq} x = b_{eq}; \quad b_{l} \leq x \leq b_{u}.
\end{align*}
\]  

(B.5)

where \( x, b, b_{eq}, b_{l} \) and \( b_{u} \) are vectors, \( A \) and \( A_{eq} \) are matrices, \( c(x) \) and \( c_{eq}(x) \) are vector-valued functions and \( f(x) \) is a scalar-valued function. The three functions can be non-linear. The user has to provide an initial value \( x_0 \) from where the search starts. In the general case the function fmincon uses an algorithm called Sequential Quadratic Programming. At
Appendix B. Numerical Methods

Figure B.1: Comparison of the performance of two non-linear programming methods. Both upper panels show the target function. The lower panels show the location of the minimum as a function of the starting value $x_0$ for two non-linear programming methods: Sequential Quadratic Programming ($fmincon$) on the left and Genetic Algorithm ($ga$) on the right.

Each optimization step, the non-linear problem is locally approximated by a Quadratic Programming subproblem based on gradients, Jacobians and Hessians of the three functions, which, therefore, must have some degree of smoothness. Details on the algorithm and the details of the function can be found in [97].

While the classical non-linear programming algorithms are quite fast and accurate, they suffer from a peculiar problem. Figure B.1 on provides an
B.2. Numerical Optimization Issues

illustration of the problem. We consider a simple test function

\[ f(x) = -\cos(\pi x) \exp(-|x|/3) \]

and we search for its minimum under the constraint that \(-5 \leq x \leq 5\). We repeat the search with different starting values \(x_0\) ranging also from \(-5\) to \(+5\) and plot the location of the minimum as found by \texttt{fmincon} against \(x_0\). We can clearly see that the solution depends quite heavily on the initial value \(x_0\), and generally only the local minimum next to \(x_0\) is found. Near local maxima or turning points, the solution can even go completely astray. This problem was also found to prevail in the real GMM optimization problem, and it is rather serious as, in general, it is not possible to give an initial guess that is close to a global minimum.

What we need is thus a global optimization method, i.e. one that finds a global minimum independently of the starting value. This method is provided by the Genetic Algorithm resp. its Matlab implementation \texttt{ga}, see [98] for details or [39] for other applications in statistics. The right-hand side of Figure B.1 documents the results of the same experiment as above, but by using the Genetic algorithm. In fact, in this case we do not use a starting point \(x_0\), but we start with a small interval of \(\pm 0.5\) around \(x_0\) in which the initial population is located. The few outliers could also be made disappear with slight adjustments on the optimizer settings, but they are left here to show that the good results of the Genetic Algorithm do not come automatically.
Bibliography


Curriculum Vitae

Personal Data
Name: Peter Blum
Date of Birth: May 16th, 1970
Nationality: Swiss, citizen of Buch & Beggingen, SH

Education
2000 – 2005: Graduate studies in Mathematics, ETH Zurich;
Supervisor: Prof. Dr. Paul Embrechts
1993 – 1999: Undergraduate studies in Mathematics, ETH Zurich;
Diploma awarded in 1999
1990 – 1993: Maturitätskurse für Erwachsene, Lucerne;
Matura Typus C
Federal Certificate of Capacity

Awards
2000: Walter Saxer - Versicherungs - Hochschulpreis
for the Diploma Thesis
1999: Diplom mit Auszeichnung

Employment
2005 – date: Senior Qualitative Analyst, Suva, Lucerne
2004 – 2005: Research Associate, RiskLab at ETH Zurich
2000 – 2004: Research Associate, Converium and ETH Zurich
1999 – 2000: Software Engineer, bbv Software Services, Lucerne
1997 – 1998: Teaching Assistant (part time), ETH Zurich
1996 – 1998: Software Engineer (part time),
bbv Software Services, Lucerne
1995 – 1995: Software Engineer (part time), Landis & Gyr, Zug
1990 – 1993: Software Engineer, Landis & Gyr, Zug

251