VAKUUTUSVALVONTA

EQUITY AND INTEREST RATE MODELS IN LONG-TERM INSURANCE SIMULATIONS

REPORTS 2008:1

KUVAILULEHTI/ BESKRIVNING/DESCRIPTION

Julkaisija/Utgivare/ Publisher
Vakuutusvalvontavirasto, Försäkringsinspektionen, Insurance Supervisory Authority

Tekijä/Redaktör/Editor
Laura Koskela
Vesa Ronkainen
Anne Puusteli

Julkaisun nimi/Titel/Title
Equity and interest rate models in long-term insurance simulations

Sisältö/Innehåll/Contents
Text

Tiivistelmä/Referat/Summary
Raportissa tutustumme eräisiin yleisiin stokastisiin osake- ja korkomalleihin päähuomion ollessa pitkäntähtäimen simulaatioissa, jotka ovat tyypillistä mm. henki- ja eläkevakuuttamiselle. Pohdimme käytännön mallintamisen eri vaiheita Solvenssi II -projektin sisäisten mallien näkökulmasta.

In this report we review and implement some commonly used stochastic models for equity prices and interest rates with a view to long-term simulations such as in life and pension insurance. We give practical details of the various modelling steps involved, and relate our discussion to the Solvency II project's internal models.

I denna rapport granskar och verkställer vi några allmänt tillämpade stokastiska modeller för aktiekapitalpriser och räntepriser med tanke på långfristig simuleringar, t.ex. inom liv- och pensionsförsäkringsbranschen. Vi framlägger praktiska detaljer av de olika stegen som ingår i modellskapandet och refererar i vår diskussion till de interna processerna inom Solvency II-projektet.

Avainsanat/Nyckelord/Keywords
Equity price models; interest rate models; internal models; risk management; Solvency II

Sarja/nimi ja numero Serie/namn och nummer Series/name and number	ISSN	ISBN
Reports 2008:1	1457-201X	978-952-5350-52-4

Sivumäärä/ Antal sidor/ Number of pages	Kieli Språk/Language	Hinta Pris/Price
101	English	

Jakaja/Distributör/Distributor
Vakuutusvalvontavirasto/
Försäkringsinspektionen
Insurance Supervisory Authority
Kirjaamo ☏ +358 9 4155 9542
e-mail: kirjaamo@vakuutusvalvonta.fi

Kustantaja/Förläggare/Publisher
Vakuutusvalvontavirasto
Försäkringsinspektionen
Insurance Supervisory Authority

Equity and interest rate models in long-term insurance simulations

Laura Koskela, Vesa Ronkainen and Anne Puustelli[1]

Abstract

In this report we review and implement some commonly used stochastic models for equity prices and interest rates with a view to long-term simulations such as in life and pension insurance. We give practical details of the various modelling steps involved, and relate our discussion to the Solvency II project's internal models.

[1]Anne Puustelli is a researcher in the Department of Mathematics and Statistics in the University of Tampere.

Contents

Abstract **1**

1 Introduction **4**

2 Internal models and solvency supervision **5**

2.1 Internal modelling . . . 5
2.2 Solvency II . . . 6
2.2.1 Validation process . . . 6
2.2.2 Modelling areas and partial models . . . 6
2.2.3 QIS 4 and internal models . . . 7
2.2.4 Internal models in insurance companies' risk and capital management . . . 7
2.3 International developments . . . 7

3 Modelling frameworks for internal models **8**

4 Equity market modelling **13**

4.1 Introduction . . . 13
4.2 Common equity models . . . 14
4.3 Data . . . 14
4.4 Choosing the model . . . 16
4.5 The lognormal model . . . 18
4.6 ARIMA models . . . 19
4.7 GARCH models . . . 22
4.8 Regime-switching models . . . 26
4.9 Other nonlinear models . . . 29
4.10 Forecasting . . . 30
4.11 On option pricing . . . 31

5 Interest rate modelling **33**

5.1 Introduction . . . 33
5.2 Notations and relationships . . . 36
5.3 The four types of an interest rate model . . . 37
5.3.1 Arbitrage-free vs. equilibrium model . . . 37
5.3.2 Risk neutral vs. real world . . . 38
5.3.3 When to use each of the model types . . . 39
5.4 Desirable features . . . 40
5.5 On one-factor short rate models . . . 40
5.6 Interest rate models in the market consistent valuation of life insurance policies . . . 45
5.6.1 The model . . . 45

5.6.2 Bayesian estimation . . . 46
5.6.3 Insurance contract . . . 47
5.7 On multifactor models . . . 48
5.7.1 Are one-factor models sufficient for practical purposes? 48
5.7.2 Introduction to multifactor models . . . 49
5.8 Heath, Jarrow & Morton (HJM) model . . . 50
5.8.1 Introduction to the HJM model . . . 50
5.8.2 Initial curves and the availability of market data . . . 53
5.8.3 Volatility functions in HJM . . . 54
5.9 On market models . . . 56

References **57**

A Descriptive analysis of the equity index data **62**
A.1 Data transformations . . . 62
A.2 Typical features of financial time series . . . 64
A.3 Data-analysis . . . 64
A.3.1 Summary statistics . . . 64
A.3.2 Data description . . . 65

B ARIMA modelling **71**
B.1 S&P 500 yearly Total Return Index 1925–2006 . . . 71
B.2 S&P 500 yearly Total Return Index 1955–2006 . . . 76
B.3 S&P 500 monthly Total Return Index 1955–2006 . . . 76

C GARCH modelling **77**
C.1 S&P 500 yearly Total Return Index 1925-2006 . . . 77
C.2 S&P 500 yearly Total Return Index 1955-2006 . . . 81
C.3 S&P 500 monthly Total Return Index 1955-2006 . . . 82

D Regime-switch modelling **83**
D.1 S&P 500 monthly Total Return Index 1955-2006 . . . 83

E Stochastic interest rate and equity modelling **87**

F HJM modelling **90**
F.1 Interest rate data . . . 90
F.1.1 Some data-analysis . . . 90
F.2 Forward rate curve estimation . . . 95
F.3 Factors and volatilities . . . 96
F.4 Pricing derivatives . . . 99

1 Introduction

This report grew out of the documentation of our current research project that deals with simulation models and their use in insurance company modelling and solvency supervision. An early example of this type of approach is given in Pentikäinen et al. (1989). The project is carried out and coordinated by the Research Unit of the Insurance Supervisory Authority, Finland. It is closely related to a broader international context, and these international developments, notably the use of internal models in Solvency II, are reviewed in Chapters 2 and 3. A comprehensive simulation model that takes into account all relevant aspects of an insurance company is a very resource intensive and challenging project. Therefore it is quite natural that our scope is here limited. We cover a number of models for equity returns in Chapter 4 and briefly discuss interest rate models in Chapter 5. We regard the material in Appendices relevant especially for those interested in implementation aspects. We recognize that those working on practical implementation will be faced with many issues that are not covered in this report. For instance interest rate modelling is a vast and challenging area where the current literature does have gaps when it comes to implementation details. Despite the limitations of our scope we hope that this report can serve as a useful introduction and reference for model builders in the insurance field. At a later stage we also aim to cover e.g. stochastic mortality forecasting.

The authors would like to express their appreciation to Research Director Dr. Lasse Koskinen (ISA) and Dr. Arto Luoma (Department of Mathematics and Statistics, University of Tampere) for their assistance during different phases of this work.

---

2 Internal models and solvency supervision

2.1 Internal modelling

Stochastic modelling and forecasting of the various cash-flows that an insurance company must deal with is very topical both in the EU and internationally. It is an issue of importance for both insurance companies and their various stakeholders, i.e. regulators, supervisors, rating agencies, financial analysts, policyholders etc. These so called internal models have a broad field of potential applications in insurance business as the following examples will show.

Economic capital models, typically based on a VaR (or percentile) risk measure, are used widely to determine the desired level and allocation of capital. ALM-models take an integrated view to assets and liabilities and their dependencies, and they allow various strategies to be tested regarding asset allocation, profit-sharing policy etc. Portfolio optimization for the asset allocation strategy is also related to ALM-modelling. Another increasingly important modelling application is the market consistent valuation of embedded guarantees and options, which can also be effectively used for the profit testing of various life insurance products. Embedded value and its market consistent variant is a traditional and commonly used tool to measure the value of a life insurance company. For further discussion on these issues we refer to Section 10.1.1 of the QIS 4 report (for QIS 4 see a separate section below).

We can already observe that a number of problems relevant for the effective management of an insurance company can be analysed with the help of simulation models. 'One size fits one' is a quite common phenomenon in certain modelling areas, for instance in interest rate modelling. So, it may not be practicable to attempt to solve all the problems with only one universal model. However, it is of course neither practical nor economical to develop more models than is absolutely necessary. Based on this point of view, insurance regulators are currently working towards European (Solvency II) and international (IAIS) solvency standards which will allow insurance companies to use a single model for their capital assessment both for the internal economic capital purposes and for the regulatory capital requirements. However, before giving their approval for such a model, supervisors will require that the aspects they consider important are adequately reflected in the model and its use environment. We now turn to these issues in more detail.

2.2 Solvency II

Solvency II is a fundamental and far-reaching project that will raise the current out-dated insurance regulations in the EU to a new, international best practice level. The focus is on better risk management and therefore the methods to quantify various risks are fundamentally important. It is here that the internal models of insurance firms become valuable tools. Below we outline the validation process for these models that supervisors will use.

2.2.1 Validation process

The general validation process for internal models is given in articles 109-122 in the Solvency II Directive proposal (COM(2008)119 final, available at `ec.europa.eu/internal_market/insurance`). The main steps in the process are:

1. Use test (art. 117)
2. Statistical quality standards (art. 118)
3. Calibration standards (art. 119)
4. Validation standards (art. 121)

Additionally there are requirements for documentation standards (art. 122), profit and loss attribution (art. 120), governance and management (art. 113) etc. The articles in the directive are necessarily quite general and thus have to be detailed by implementing measures and supervisory guidance. These are currently under development in CEIOPS. Some of the challenging issues relate to insufficient data, finding a right balance between too little and too much harmonisation, and getting rid of any 'black boxes' in the modelling process.

2.2.2 Modelling areas and partial models

The Solvency Capital Requirement (SCR) is based on the idea that an insurance firm should have the amount of capital that is sufficient with a 99,5 percent confidence-level to guarantee that the firm will have enough assets to cover its liabilities at the end of the one year period. This idea and the risks that need to be modelled are described in articles 102-106 of the Solvency II directive (Operational risk; Non-life underwriting risk; Life Underwriting risk; Market risk including interest rate, equity, property, spread, currency and concentration risk; Default risk). The directive proposal also gives an option to use a partial internal model, i.e. to model only certain risk modules or business units and to use the standard formula for the remaining parts (art.109).

In this report we will only deal with market risk and in particular equity and interest rate risk. Our approach is tailored for long-term forecasting and risk management purposes and it is therefore statistical, not focused on current market values of various derivative etc. instruments and their arbitrage-free modelling. However, we will also discuss the latter approach in the context of interest rate modelling.

2.2.3 QIS 4 and internal models

Quantitative Impact Study 4 (QIS 4), the most recent field test of Solvency II, included internal models as one key area. The summary report, CEIOPS SEC-82-08, is available at `www.ceiops.eu`. It gives up to date information on how insurance companies are currently using, and planning to use in the future, internal models as a part of their risk management and capital assessment frameworks. It also compares the Solvency Capital Requirement (SCR) and its various components based both on the standard formula and internal models.

2.2.4 Internal models in insurance companies' risk and capital management

Korhonen & Koskinen (2008) explored several critical aspects of risk and capital management of an insurance company in the internal model context and provided examples that illustrated the potential importance of management science tools for internal model users and developers. The problem was formulated as a multiple criteria decision making task with a hierarchical structure. Korhonen & Koskinen used Analytical Hierarchy Process as a planning tool to analyze management criteria, causal and risk factors. The evaluation was carried out (in October 2007) by a panel consisting of senior managers of major Finnish insurance companies. "Investment Management" was found to be the most important management criterion. Other important criteria were "Cycle Management" and "Risk Management". As a final result, panel's main concern was that the new regulatory regime may create a potential for new sources of systemic risk and supervisory over-control of the internal models. Model related important risk factors are (pure) model risk and dependence assumptions (considered separately).

2.3 International developments

We have discussed above the coming EU rules and on-going work relating to internal models. However, we should also note that there are international projects under way both in the supervisory community (IAIS, `www.iaisweb.org`) and the actuarial profession (IAA, `www.actuaries.org`) that deal with the same subject. One important example and reference is Wüthrich et al. (2008). It is the intention of the EC and CEIOPS to take into account these other developments also in the design of the EU rules.

3 Modelling frameworks for internal models

There are several conceptual frameworks that can be useful for the modelers. One useful way of thinking is the so called actuarial control cycle. The idea is to take into account the links and loops between various profit and loss and balance sheet items as well as actuarial, risk management and business planning goals (see page 31 of the IAA report 'Global Framework for Insurer Solvency Assessment', available at `www.actuaries.org`).

Statistical model fitting is a valuable area to gain modelling insights. For instance the Box-Jenkins method for time series modelling consists of the following steps (Box et al. (1994)):

1. Specification
2. Estimation
3. Evaluation

In practice we would first choose the general linear ARIMA class of models and analyse the data in order to choose a tentative ARIMA(p,d,q) model, and then we would estimate the parameters. The evaluation phase may use such tools as overfitting, analysis of error diagnostics, testing for in-sample and out-of-sample fit using some suitable criteria. As a result we should be able to identify the best fitting model that can be used to solve our practical problems of forecasting or control. We will illustrate these steps later in the context of equity price modelling.

The following list of key steps and things to consider in statistical modelling is based on Chatfield (1995), and we find it relevant for internal models also.

1. The objectives of the model have to be clearly specified and continuously kept in mind.
2. The management and responsibilities relating to data collection, modelling and analysis should be arranged in such a way that the whole process can be executed and supervised appropriately.
3. Data collection principles should define the important variables required by the model, the way they are measured, and the procedual steps for carrying out the actual sampling of the data.

4. Data scrutiny involves meticulous assessment of the structure and quality of the data, and the results of this analysis and the corrective actions taken should be documented. In particular credibility, consistency and completeness of the data should be examined. This step also includes the processing of the data in a suitable input form for the model (databases etc). Issues to be checked include the following: a) the coding of the data is accurately done and suitable to be used as input for the model, b) the sampled data is representative and of a suitable size, c) analysis and appropriate actions regarding errors, missing observations and outliers, d) modifying the data if necessary by transformations, corrections, using virtual data to complement the lack of observed extreme values etc, e) consistency checks if data from several sources have been merged, f) summarizing the data in suitable and clear numerical and graphical forms.

5. The processing of the data begins with an initial data analysis (IDA) and proceeds with the model building steps: formulation, fitting and validation. All these steps should be documented in order to justify the choices made at various stages of the model development.

6. Initial data analysis uses various descriptive methods (summary statistics, tables, graphs) to analyse the data. It gives useful background information for the modelling decisions and should therefore always be carried out and documented.

7. Model formulation/specification process may vary from case to case and as such cannot be standardised. At this stage one should carefully consider the fundamental modelling questions such as the following: a) the context and objectives, b) background theory and literature review, c) earlier experiences, d) market best practice and expert knowledge, which give valuable input for the modeller when choosing and testing which types of statistical analyses and inference[1] would best suit the research question and data at hand, e) the assumptions made and their validity have to be documented and justified (e.g. what is known or assumed or approximated and why).

8. Model fitting/estimation focuses on finding point and interval estimates for the model parameters using appropriate statistical methods.

9. Model checking/evaluation/validation concentrates on the performance assessment of the model. This includes diagnostic/residual analysis, benchmarking with other models, data sets and prior knowledge. Model robustness should be assessed with respect to chosen assumptions and data. These analyses will also give indications of the model uncertainty involved, which should be considered and accounted for. When the performance is not satisfactory in some respect, a new round of model formulation, fitting and checking should be carried out iteratively until an acceptable solution has been found.

[1]Frequentist, Bayesian, decision theoretic

10. Appropriate software and hardware are crucial when implementing numerical stochastic methods. As a rule the better the quality of these components, the more precise and reliable are the results. Therefore the choices made should be justified and documented. Sometimes it may be necessary or preferable to use a number software packages during the modelling process. The packages used should be thoroughly tested (e.g. by running the same data sets through competing packages) and well documented and supported. They as well as all the other software used in the modelling should be based on sound, published and tested algorithms. All the numerical results should be replicable (double-checking should be done during the validation, auditing and approval processes).

11. Literature references should be a key part of the model documentation at every stage.

12. Effective communication is essential during the model development and documentation, and when the results are presented. This requires that the papers are accurately and illustratively written, and that plenty of open discussions between the model builders, experts and managers have been carried out. In Solvency II context this relates also to the pillar 3 requirements that deal with disclosure and market discipline.

There exists a risk that the constructed model is wrong or does not adequately perform the tasks designed to it. The inescapable consequence of model use is called model risk. A nice review on model risk involved in using models to value financial securities is provided by Derman (1996). A broad typology for a risk model's model risk is given e.g. in Down (2002). He classifies:

- Misspecified model: Stochastic process might be misspecified, missing risk factors, misspecified relationships, transaction costs and liquidity factors;
- Incorrect model application;
- Implementation risk;
- Other sources: Incorrect calibration, programming problems and data problems.

Further, Down (2002) argues that there is no single strategy for avoiding model risk, but to combat model risk we can:

- Be aware of model risk;
- Identify, evaluate and check the key assumptions;
- Test models against known problems;
- Choose the simplest reasonable model;
- Backtest and stress-test the model;

- Estimate model risk quantitatively;
- Do not ignore small problems;
- Plot results and use non-parametric statistics;
- Re-evaluate models periodically.

Hence, in order to assess model risk an intimate knowledge of the modelling process is required. The model risk is discussed in the Solvency II framework e.g. in Ronkainen et al. (2008).

Financial derivative modelling is an area which is rapidly gaining ground in insurance due to market consistent valuation of assets and liabilities (required by insurance and accounting regulators). Many life and pension insurance contracts have features that can be described using derivatives terminolgy. The basic ideas of derivative modelling approach are summarised in Section 4.11.

General forecasting principles (available at `www.forecastingprinciples.com`) outline the best practice in the forecasting field and are very useful for insurance modelling too.

As a final illustration we present the iterative steps of model development cycle as follows:

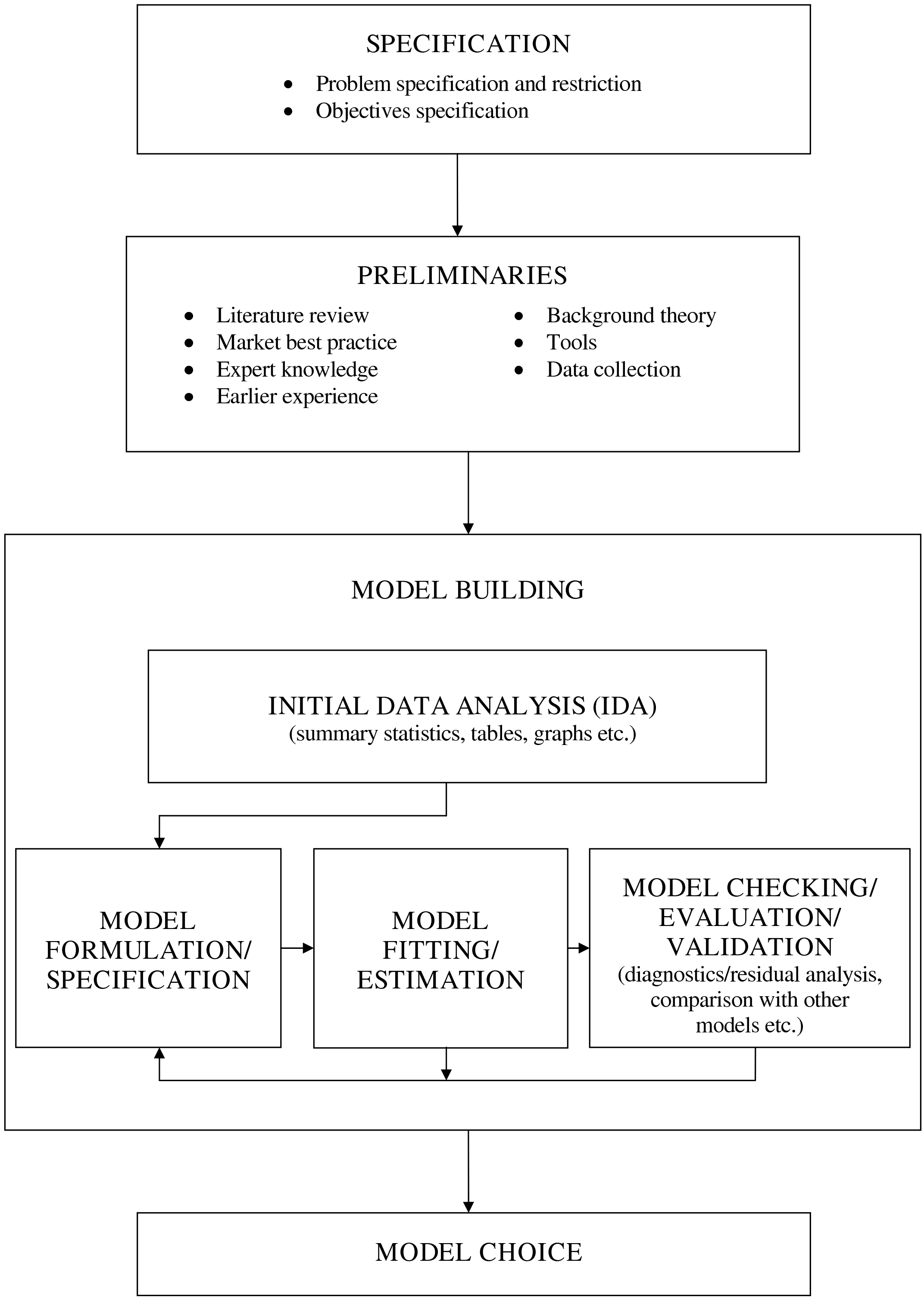


Figure 3.1: Iterative steps of model development cycle.

4 Equity market modelling

4.1 Introduction

Some part of life or pension insurer's assets is typically invested in equity (i.e. stocks or shares). The reason is that in the long run they are expected to provide a higher, although more volatile, return than what interest rate securities can provide. However, as we are here dealing with a complex stochastic process, the exact form of which is ultimately unknown, sound risk and return analysis for shorter and longer horizons is a necessary part of the risk management process (typically both nominal and real returns have to be considered).

The models for equity or other asset prices or returns, may be expressed in continuous or discrete time. Amongst actuaries, most notable is the pioneering work of Wilkie (1986, 1995). The Wilkie model is based on Box–Jenkins methodology. According to Cochrane (2001) the choice between the discrete- and continuous-time representations of the same idea is one of modelling language. He further states that the modeller should be familiar enough with both discrete- and continuous-time representations to be able to choose the one that is most convenient for a particular application. Some of the advantages of continuous-time models are that such models are analytically tractable and they provide a convenient framework for pricing the derivatives under the so-called complete market hypothesis. The problem faced with the continuous-time models is, however, that the price data are essentially always recorded at discrete points in time (e.g annually, monthly, weekly or daily), whereas the continuous-time models assume that the price is monitored continuously in time. Due to the unavailability of a continuous sample of observations, the estimation has usually been performed by first appropriately discretizing the model and then applying various estimation methods. Since discrete-time models are particularly useful for numerical computations, they may be a better starting point for many practical situations.

Below we will review and implement several models that are commonly applied for equities. We will mainly resort to a discrete-time framework and will discuss the models as financial time series models. Financial time series analysis provides us with statistical tools useful for analysing various types of models. Using the real observed data sets we aim at studing how well a specific model describes the stylized facts of the equity market data.

However, before closing this section, it is essential to understand what is here meant by volatility. In words volatility is the degree to which the value of a financial instrument tends to fluctuate, i.e variability. More precisely, volatility refers to the standard deviation of the change in value of a financial instrument with a specific time horizon. Volatility is often used to quantify the risk of the instrument over that time period. The causes of volatility are discussed e.g. in Hull (1999). These causes mentioned are the random arrival of new information about the future returns from the stock and trading.

4.2 Common equity models

A number of models can be applied to equity price or return[1] modelling as is explained e.g. in Hardy (2003), Tsay (2005), and Franses & Dijk (2000). Here are some common examples:

1. The lognormal model (see Section 4.5)
2. ARIMA time series models (see Section 4.6)
3. GARCH time series models (see Section 4.7)
4. Regime-switching models (see Section 4.8)
5. Jump-diffusion models (see Section 4.9)
6. Stochastic volatility models (see e.g. Tsay, 2005)
7. State-space model (see e.g. Shumway & Stoffer, 2006, ch. 6)

As will be illustrated below, these models differ with respect to their complexity, which in turn is related to their ability to take into account certain stylized facts observed in equity markets. These may include a trend, seasonality, atypical observations, clusters of outliers and nonlinearity. State-space modelling uses structural approach where trend, volatility, seasonal effects etc. components can appear explicitly in the model equations. It thus allows more transparent modelling than e.g. Box-Jenkins ARIMA approach. Typically the so called Kalman filter recursive methodology is used for parameter estimation, although it is limited to linear and Gaussian models in its basic form. However, we will not discuss this modelling approach further in this report.

4.3 Data

Stock market indices track the performance of a specific ”basket” of stocks considered to represent a particular market or sector of a particular stock market or the economy. They can be used as measures of asset price evolutions, benchmark for evaluating the performance of portfolio management,

[1]Direct statistical analysis of prices is difficult, because consecutive prices are correlated and the variance of prices often increases with time. It is usually more convenient to build up the model for returns, i.e. changes in prices. This is because return series commonly have more attractive statistical properties. The results for returns can be used to give appropriate results for prices.

support of derivatives, and economic indicators (Gourieroux & Jasiak, 2001, p. 413).

Diversification is the key tool to manage idiosyncratic investment risk (also called unsystematic or diversifiable risk), and this principle is generally applied in the asset management of insurance companies and pension funds[2]. Consequently, in many cases an appropriately chosen basket of market indices can be expected to provide a good approximation — in terms of return and risk — to the actual asset portfolio that an insurance company holds[3]. This is mentioned for instance in Morningstar (2007) on page 126: ”It is safe to say that, on average, the pension funds and balanced mutual funds are not adding value above their asset allocation policy due to their combination of timing, security selection, management fees, and expenses. Thus, about 100 percent of the total return is explained by asset allocation policy.”

In the most recent quantitative impact studies QIS 3 and QIS 4 the potential economic implications of Solvency II were tested. In this exercise the benchmark indices according to which the parameters of market risk module for equities were calibrated were Global and European MSCI Developed Market Indices (starting from 1970). Another example is the following basket of equity indices:

- Standard and Poor’s 500 Index
- Dow Jones STOXX 600 Price Index EUR
- OMX Helsinki Cap Index
- NASDAQ Composite Index
- Nikkei 225 Index

A serious problem for the modeller is the fact that for many market indices the data series is too short when long-term predictions are in focus. In particular the Euro-zone has not yet long enough history to provide appropriate data for long-term forecasting purposes. One rule of thumb in statistical modelling says that there should be at least the same amount of historical time series data available for estimation as is the length of forecasting period. However, as emphasized by Alexander (2001), in choosing how far to go back with the data, one has to take a view on whether or not current forecasts should be influenced by events that occured many years ago. Reliance on market values is the foundation of most financial models, and it also underlies the current international projects in the field of insurance, e.g. Solvency II.

In this report our approach is mainly statistical (econometrical) and we wish to use long data series as the basis of our modelling. The main reasons for this decision are the following:

[2]Diversification is one requirement in the prudent person rule which is an investment standard for pension funds.

[3]When not taking into account the effect of hedging etc. financial engineering may of course significantly change the probability distribution of returns compared to the index.

1. Forecasting horizon for pensions have to extend beyond what is typically available in the security markets for equity derivatives or government bonds[4].

2. We are mainly concerned with risk management and forecasting, which have to be done under real world probabilities (see Chapter 5).

3. In the time series of equity prices/returns we can observe bubbles and subsequent crashes from time to time. These overreactions can, to a large extent, be explained by the theory of behavioral finance (for an overview and references see e.g. Kaliva et al., 2007). As the current market volatility confirms, there seems to be no reason to believe that human behaviour in the market (motivated by hope, greed, panic, herd behaviour etc human traits) has changed essentially from its historically implied levels. The same applies to economic recessions, violent conflicts etc., which also create atypical observations.

The equity market data analyzed and used in this report to study the characteristics of various models are the S&P 500 yearly Total Return Index 1925-2006 and the S&P 500 monthly Total Return Index 1955-2006 from Morningstar (2007). These indices are nominal and include also the effect of reinvested dividends. The S&P 500 Total Return Index is a widely recognized index of common stock prices of U.S. companies. It consist of 500 large stocks, which are weighted by market capitalisation[5]. The S&P is a common benchmark for institutional investors.[6]

Very often return series are investigated instead of the inital series of prices, and it is common to define the returns as the logarithmic returns (log returns). The log return at time t, i.e. y_t, is computed as

$$y_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1} = p_t - p_{t-1}, \tag{4.1}$$

where P_t is the equity price at time t and $p_t = \ln P_t$. In case of a stock market index series, P_t is the index value at time t. The transformation is discussed in more detail in Appendix A. Appendix A also gives a descriptive analysis of the data series used in this study and examines how well these series reflect the stylized facts of financial time series.

4.4 Choosing the model

When suitable data series have been identified, a number of appropriate models should be estimated in order to find the best fit for the application at hand.

[4]Among the longest bonds in the Euro area is Germany's 30 year paper and France's 50 year bond. The longest actively traded equity derivatives are of much shorter duration, typically they tend to expire in two or three years.

[5]If a stock has 10 million shares outstanding and sells for \$20 per share, its market capitalisation is 200 million dollars.

[6]The Dow Jones Industrial Average is less diversified (only 30 large stocks) and uses average prices without weighting by market capitalisation. These two indices generally move in the same direction, but the magnitude of the changes can differ because of the differences in their definitions.

Both theoretical and applied literature should prove helpful when choosing the models to be studied further. The fit of various models can be assessed by optimizing some suitable criterion for the in-sample fit or by assessing the forecasting performance of various models[7].

For equity modelling both linear and nonlinear time series models might be used, but in many applications — in particular for high-frequency data — nonlinear models perform significantly better. When comparing models based on their in-sample fit, one can use for instance the AICC statistic (a bias-corrected AIC-criterion) that is applicable both for univariate and multivariate modelling. Another common criterion to assess the in-sample fit is the BIC, that is the Bayesian (or Schwarz-Rissanen) information criterion.

There is a difference between the AIC and BIC in their basic assumptions (see e.g. McQuarrie & Tsai, 1998). If one believes that the true model is of infinite dimension or that it is not included in the set of candidate models, the goal in model selection is to choose the finite-dimensional candidate model that best approximates the true model. In this approach, one needs to fix how to measure the distance between models. A widely used distance measure is the so-called Kullback-Leibler distance, on which the AIC and AICC are based. These criteria are also asymptotically efficient in the sense that they choose the model with minimum mean squared error distribution, when the sample size tends to infinity.

On the other hand, if one believes that the true model is finite-dimensional and that it is included in the set of candidate models, the goal in model selection might be to choose the model which has the highest probability to be the correct one. The BIC is based on this idea, and it is consistent in the sense that it chooses the correct model with probability one when the sample size tends to infinity. In the statistical approach these assumptions are almost always too strong. Especially in our econometric forecasting and case study purposes, the AIC and AICC are the most appropriate but in most cases we have also computed the BIC values for comparison.

In the Bayesian approach to modelling, one can use as a model selection criterion the average discrepancy, defined as $D_{\text{avg}}(y) = \mathrm{E}(D(y,\theta)|y)$. In Bayesian terminology, it is the posterior mean of the deviance $D(y,\theta) = -2\log p(y|\theta)$ (see e.g. Gelman et al., 2004) and can be estimated as $\widehat{D}_{\text{avg}}(y) = \sum_{l=1}^{L} D(y,\theta_l)/L$, where the vectors θ_l are posterior simulations of the parameters. One can also use the deviance information criterion, defined as $\mathrm{DIC} = 2\widehat{D}_{\text{avg}}(y) - D_{\hat{\theta}}(y)$, where $D_{\hat{\theta}}(y) = D(y,\hat{\theta}(y))$ is the minimum value of the deviance (see e.g. Spiegelhalter et al., 2002). The minimum deviance can be approximately obtained, in addition to direct minimization, by using extensive posterior simulations or by replacing $\hat{\theta}$ with any point estimate, such as the posterior mean or median. The DIC can also be written in the form $\mathrm{DIC} = D_{\hat{\theta}}(y) + 2p_D$, where $p_D = \widehat{D}_{\text{avg}}(y) - D_{\hat{\theta}}(y)$ is called the effective number of parameters. It represents the expected improvement in the fit, improvement which is expected from estimating the parameters of the model. In simple cases, p_D is asymptotically equal to the true number of parameters.

[7]In this case a part of the sample data has to be reserved for the out-of-sample forecasting purposes which may not be possible in many long-term insurance applications.

It can also be computed in models where the number of parameters is not clearly defined, such as in hierarchical or state-space models. Therefore, the DIC is more general than AIC, AICC or BIC.

The goal of both DIC and AIC (or AICC) is to choose the model which minimizes the Kullback-Leibler distance between the distribution of the true model and the predictive distribution of the fitted model. This is equivalent to minimizing $E[D(y^{rep}, \hat{\theta}(y))]$, where y^{rep} is the replicated data set and the expectation is taken under the assumed true model. Therefore, the DIC and AIC (or AICC) values are comparable.

There are a number of criteria to assess the out-of-sample forecasting performance of a model, such as mean absolute error (MAE) and root mean square error (RMSE). Kennedy (2003) points out that no single criterion is always "best" — the "best" criterion depends on the particular problem being analysed. There is, however, some agreement in the literature that the "best" forecasting method overall is a "combined forecast" that is formed as a weighted average of various different forecasts. This issue is discussed from the insurance point of view in Kaliva et al. (2007).

Modelling error is a very important question to be considered when internal models are being used for regulatory purposes. In general the longer the forecasting horizon, the greater the role of subjectivity and expert opinions (due to lack of data and because of the possibility that the data-generating process will change during the forecasting period). In this report the equity model comparisons are based on the results for the in-sample using the above described model selection criteria.

4.5 The lognormal model

The lognormal model is a discrete version of the geometric Brownian motion (GBM)

$$\frac{dP_t}{P_t} = \mu\, dt + \sigma\, dW_t, \tag{4.2}$$

where P_t is the equity price at time t, μ and $\sigma \geq 0$ are constants and W_t is the standard Wiener or Brownian motion process[8]. The term dP_t/P_t represents the percentage change or return in the asset price, μ represents the expected rate of return of the equity price (drift) and σ is the volatility. The independent lognormal model, obtained by discretizing the GBM, can be expressed as

$$y_t = \mu + \sigma z_t, \tag{4.3}$$

where y_t is the log return, and z_t are independent and identically distributed standard normal for all t. It follows from (4.3) that $P_t/P_{t-1} \sim \text{LogN}(\mu, \sigma)$, where LogN denotes the lognormal distribution. Given last period's price

[8]Wiener process can be defined by two components: (1) dW_t is normally distributed with mean zero and variance dt, (2) the values of dW_t over two different, non-overlapping increments of time are independent (see e.g. Clewlow & Strickland, 1998).

P_{t-1}, also P_t follows a lognormal distribution. We may express the price at time t as

$$P_t = P_{t-1} \exp(\mu + \sigma z_t). \tag{4.4}$$

This shows that since both P_{t-1} and $\exp(\mu + \sigma z_t)$ are non-negative, also P_t will be non-negative.

The lognormal model, or geometric Brownian motion, is most widely used to describe the equity price behaviour and it is one of the original assumptions of the Black-Scholes framework (Black & Scholes, 1973). Since the model states that all subsequent price changes represent random departures from previous prices, it is consistent with the efficient market hypothesis[9]. The lognormal model provides a useful approximation for short and medium maturity. However, over longer term, the model does not usually provide a satisfactory fit to the data. Empirical studies indicate that the model fails to capture more extreme price movements, does not allow for autocorrelation in the data and also fails to capture volatility clustering (Hardy, 2003). Note that the lognormal model is a special case of ARMA models applied to log returns, or of ARIMA models applied to log prices. In the ARIMA context (Section 4.6) the lognormal model is referred to as a random walk model with a drift.

4.6 ARIMA models

It is in particular the correlations between the variable of interest and its past values (i.e. autocorrelation or serial correlation) which are the focus of linear time series analysis. ARMA models constitute a very useful class of linear time series models for a wide range of applications. These discrete stochastic processes depend linearly on their past values y_{t-i} and random shocks of white noise ϵ_{t-i}.[10]

The foundation of time series analysis is stationarity. It is therefore common to apply ARMA model to return series instead of prices or market indices. In this report we are mainly concerned with weakly stationary time series[11], which show the tendency of fluctuating with constant variation around a fixed level. For the definition of weak and strict stationarity we refer to Tsay (2005). For the stationary time series the ARMA(p, q) model can be expressed as

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}, \tag{4.5}$$

where ϵ_t is white noise with mean zero and variance $\sigma_\epsilon^2 > 0$, $\phi \neq 0$ and $\theta \neq 0$ (see e.g. Shumway & Stoffer, 2006). The parameters p and q are called the

[9] According to the efficient market hypothesis, the market accurately reflects all knowledge available about all stocks at all times. In other words, past stock prices do not provide information that permits an investor to outperform the market.

[10] The white noise is a collection of uncorrelated random variables ϵ_t, with mean 0 and finite variance σ_ϵ^2. It shall be sometimes denoted as $\epsilon_t \sim wn(0, \sigma_\epsilon^2)$.

[11] In the finance literature, it is common to assume that an asset return series is weakly stationary. This assumption can be checked empirically provided that a sufficient number of historical returns are available (Tsay, 2005).

autoregressive and the moving average orders, respectively. If y_t has a nonzero mean μ, we may replace y_t by $y_t - \mu$ in (4.5), i.e.

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + ... + \phi_p(y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + ... + \theta_q\epsilon_{t-q}, \tag{4.6}$$

or write

$$y_t = \alpha + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + ... + \theta_q\epsilon_{t-q}, \tag{4.7}$$

where $\alpha = \mu(1 - \phi_1 - \ldots - \phi_p)$. It is important to understand the meaning of a constant term in a time series model. As explained by Tsay (2005), for MA(q), i.e. ARMA(0,q), model the constant term is simply the mean of the series. For a stationary AR(p), i.e. ARMA(p,0), model or ARMA(p, q) model, the constant term is related to the mean via the relationship given above, that is $\alpha = \mu(1 - \phi_1 - \ldots - \phi_p)$. For a random walk model (see ARIMA(0,1,0) model below), the constant term becomes the time slope.

ARIMA model (Integrated ARMA) is a broadening of the class of ARMA models to include differencing. If the time series can be modelled with ARMA(p, q) after it has been differenced d times, then the original series is modelled with ARIMA(p, d, q). For example, a process is said to be ARIMA($p, 1, q$), if $\nabla y_t = y_t - y_{t-1}$ is ARMA(p, q). Similarly, a process is said to be ARIMA($p, 2, q$), if $\nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}$ is ARMA(p, q). Note also, that ARIMA($p, 0, q$) is equivalent to ARMA(p, q).

Univariate ARIMA models are in effect a sophisticated extrapolation method, using only past values of the variable being forecast to generate forecasts. Kennedy (2003) considers that the potential sources of errors in econometric forecasting relate to specification, conditioning, sampling, and randomness. Then (on page 361) he goes on to compare traditional causal econometric models with ARIMA models as follows: "It is generally acknowledged that whenever specification or conditioning errors render econometric models impractical (which some claim is most of the time), the Box-Jenkins approach has considerable merit for forecasting. It is also recognized that if an econometric model is outperformed by an ARIMA model, this is evidence that the econometric model is misspecified".

We have fitted an ARIMA model to S&P 500 Total Return Index data using the approach explained in Brockwell & Davis (2002). This method, also known as the Box-Jenkins approach, consists of the following, possibly iterative, steps:

1. Make data transformations, when necessary, to get a stationary series (i.e. there should be no trend or cyclic components or non-constant level and variability with time). In our case this step consist of taking a logarithm and differencing at lag 1 (see Appendix A). This is a practical and common procedure when analysing financial data. Still, one has to appreciate the theoretical issues involved, e.g. regarding unit roots (cf. Franses & Dijk, 2000; Kennedy, 2003)

2. Calculate key statistics from the data such as autocorrelations (ACF and PACF) and compare the values or sizes of these statistics with the

theoretical ones that would hold true if a certain model is adequate. As the attention is restricted to linear ARIMA models, the main objective in this step is to determine the preliminary orders p and q.

3. Fit various models and compare the results. Choose the one which gives the smallest AICC.

4. Use diagnostic tests for the residuals of the fitted model. This step uses e.g. a histogram, a Q-Q plot as well as some specific tests of randomness such as Ljung-Box, McLeod-Li, turning points, Jarque-Bera (see Brockwell & Davis, 2002).

Below we are content with only briefly introducing the fitted models. Results that led to the choice of these models are left to Appendix A and B. Appendix B focuses on Steps 2–4 while Appendix A presents and justifies the data transformation in Step 1.

For the logarithmic index values p_t of S&P 500 yearly data 1925-2006, the procedure chose the random walk model (i.e. ARIMA$(0,1,0)$) with a constant α, which is usually called a drift. The drift term here represents the time trend of the logarithmic index. The model can be written as

$$p_t = \alpha + p_{t-1} + \epsilon_t, \tag{4.8}$$

or, alternatively, in terms of the log return as

$$y_t = \alpha + \epsilon_t, \tag{4.9}$$

where $t = 1, \ldots, 82$. As can be seen in (4.9), for the log returns y_t the model is just white noise with a drift[12], i.e. $y_t - \alpha \sim wn(0, \sigma_\epsilon^2)$, which indicates that returns cannot be predicted from past changes in a time series of historical returns. The random walk model reflects the efficient market hypothesis, which states that it is not possible to consistently outperform the market by using any information that the market already knows, except through luck. The estimated model for the log return is

$$y_t = 0.0992 + \epsilon_t, \tag{4.10}$$

with $\hat{\sigma}_\epsilon^2 = 0.0364$. According to (4.10), the expected yearly log return (growth rate) of S&P 500 Total Return Index is about 9.9%, and its 95% confidence interval spreads from about -27% to 47%.

The random walk model with a drift was also suggested for the logaritmic index of our shorter historical index series (S&P 500 yearly Total Return Index 1955-2006). The estimated model is

$$y_t = 0.1002 + \epsilon_t, \tag{4.11}$$

with $\hat{\sigma}_\epsilon^2 = 0.0234$. According to (4.11), the expected yearly log return (growth rate) of S&P 500 Total Return Index is about 10.0%.

[12] ARMA$(0,0)$ with a drift

For the monthly log index we also fitted the random walk model with a drift. The estimated model can be written as

$$y_t = 0.0086 + \epsilon_t, \tag{4.12}$$

with $\hat{\sigma}_\epsilon^2 = 0.0017$. According to (4.12), the expected yearly log return (growth rate) of S&P 500 Total Return Index is about 10.3%, and the volatility is significantly lower than for the longer series.

Our results are in accordance with the existing literature. As stated by Tsay (2005), although daily returns of a market index often show some minor serial correlation[13], monthly return of the index may not contain any significant serial correlation. Hence, for most asset return series, building a mean equation amounts to removing the sample mean from the data if the sample mean is significantly different from zero. This is exactly what has been done above.

As discussed in Appendix A, the stylized facts of equity returns imply that linear models have in many instances only limited use in financial modelling. Particularly, linear models such as ARIMA, assume a constant variance and are hence not capable of describing the time-varying volatility typical for many financial time series. The asset return typically has a time-varying variance and volatility comes in clusters where tranquil periods of small returns are interspersed with volatile periods of large returns. In Section 4.7 we will introduce a special class of nonlinear time series models (GARCH class of models) which are capable of describing some of the stylized facts of financial time series not captured by linear models. In distinction from the models that are nonlinear in mean, the GARCH models are nonlinear in variance, because their conditional variances[14] evolve over time[15]. ARIMA models are linear in both sense.

4.7 GARCH models

In time series analysis appropriate modeling of the volatility may improve the efficiency in parameter estimation and the accuracy in interval forecast (Tsay, 2005). It is therefore essential to not only model the mean equation but also to find a volatility model which adequately reflects the stylized facts of asset return series.

The volatility of a time series is not directly observable, but it has some characteristics that are commonly seen in asset returns (especially in high-frequency returns). According to Tsay (2005) these characteristics are: (1)

[13] A possible cause of autocorrelation in equity indices is the news arrival process, where new information affects only trading in some stocks before others (Alexander, 2001, p. 385).

[14] A conditional distribution, in this context, is a distribution that governs a return at a particular instant in time. In more general terms, a conditional distribution is any distribution that is conditioned on a set of known values for some of the variables. In time series that often means conditioning on all the past values that were realized in the process. The conditional variance at time t is the variance of the conditional distribution at time t. (Alexander, 2001, pp. 12–13).

[15] In fact, the conditional variance is assumed to be an autoregressive process.

There exist volatility clusters (i.e., volatility may be high for certain time periods and low for other periods). (2) Volatility evolves over time in a continuous manner (i.e., volatility jumps are rare). (3) Volatility does not diverge to infinity (i.e, volatility is often stationary). (4) Volatility seems to react differently to a big price increase or a big price drop, referred to as the leverage effect. The characteristics listed above reflect the fact that ARIMA models, which assume a constant variance, turn out to be insufficent for many financial applications. To capture changes in volatility, models such as the autoregressive conditionally heteroscedastic (ARCH) model and its extension called generalized ARCH (GARCH) were introduced by Engel (1982) and Bollerslev (1986), respectively. The focus in this section will be on standard ARCH and GARCH models which are able to not only describe the feature of volatility clustering but also capture the excess kurtosis and fat-tailedness of time series of asset reuturns.

A time series model can generally be expressed as the sum of predictable and unpredictable part, where the former is an expectation conditional on the information set available at time $t-1$, denoted by I_{t-1}. Typically, I_{t-1} consists of all linear functions of the past returns. Using mathematical notations, this can be expressed as

$$y_t = \mathrm{E}[y_t|I_{t-1}] + \epsilon_t. \tag{4.13}$$

In the previous section we studied ARIMA models and assumed that the shock ϵ_t is white noise, and that in particular $\mathrm{E}[\epsilon_t^2] = \mathrm{E}[\epsilon_t^2|I_{t-1}] = \sigma_\epsilon^2$ for all t, i.e. ϵ_t was assumed both conditionally and unconditionally homoscedastic[16]. A convenient way to introduce more realism is to assume that the conditional variance of the shock is not constant over time but instead some time-dependent and nonnegative function $h_t = h(I_{t-1})$. In other words, $\mathrm{E}[\epsilon_t^2|I_{t-1}] = h_t$. This can also be expressed as $\epsilon_t = z_t\sqrt{h_t}$, where z_t is independent and indentically standard normal or t-distributed.

The basic idea of ARCH models is that (1) the shock ϵ_t of an asset is serially uncorrelated but dependent and (2) the dependence of ϵ_t can be described by a simple quadratic function of its lagged values. To be more specific, in the basic ARCH model, the conditional variance of the shock that occur at time t is a linear function of the squares of past shocks. The ARCH(s) model can be expressed as

$$h_t = \omega + \alpha_1\epsilon_{t-1}^2 + ... + \alpha_s\epsilon_{t-s}^2, \tag{4.14}$$

where $\omega > 0$ and $\alpha_i \geq 0$ for all $i = 1, \ldots, s$. The coefficients α_i must satisfy some regularity conditions to ensure that the unconditional variance of ϵ_t is finite. It can be seen from the structure of the model that a major market movement that occured yesterday or up to s time units ago will have an effect to increase today's conditional variance. It makes no difference wheather the movement is positive or negative. So, good and bad news have the same (symmetric) effect on the volatility in this model.

[16]The term homoscedasticity means that the variance is the same throughout the process. The term heteroscedasticity means that the variance changes over time.

The ARCH models often require many parameters to adequately describe the volatility of asset returns making the estimation of the model parameters more difficult. To reduce the computational problems Bollerslev (1986) suggested adding lagged conditional variances to the ARCH model. The GARCH(r, s) model can be expressed as

$$h_t = \omega + \alpha_1 \epsilon_{t-1}^2 + ... + \alpha_s \epsilon_{t-s}^2 + \beta_1 h_{t-1} + ... + \beta_r h_{t-r}, \tag{4.15}$$

where $\omega > 0$, $\alpha_i > 0$, and $\beta_i \geq 0$ to guarantee that $h_t \geq 0$ and the model is identifiable. The coefficients α_i and β_i must satisfy $\sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) < 1$ to ensure that the unconditional variance of ϵ_t is finite. The GARCH model assumes conditional heteroscedasticity, while the unconditional variance is homoscedastic. It is also worth noting that ARCH models are special cases of GARCH models, namely GARCH(0,s) = ARCH(s). To summary, the variance in GARCH model is a weighted average of three components: a constant or unconditional variance, yesterday's forecast and yesterday's news.

Higher-order GARCH models often tend to be overcomplicated and are rarely used in practice. Instead a GARCH$(1, 1)$ model, which has just one lagged error square and one lagged conditional variance, will most often be sufficient to capture the volatility clustering in the data. It can be argued that the GARCH$(1, 1)$ is equivalent to an ARCH(∞) model with exponentially declining weights (see e.g. Alexander, 2001, p. 72). The GARCH$(1, 1)$ model is the simplest and most robust of the family of volatility models, and its parameters have nice interpretations: A large coefficient of the lagged error square in the GARCH$(1, 1)$ model indicates that volatility responds quite intensively to market movements while a large coefficient of the lagged conditional variance term means that shocks to conditional variance take a long time to die out. The parameter estimates of the GARCH$(1, 1)$ model are sensitive to the data used. In particular, long-term volatility forecasts will be affected by the stress events included in the historic data. For more detailed discussion of the choice of the historic data for model estimation we refer to Alexander (2001).

As already pointed out, the standard GARCH(r, s)-models presented in this section are able to take into account the stylized fact of volatility clustering because they tend to generate a more volatile regime after one such atypical observation has been observed. In addition these models generate certain other desirable features such as fatter tails and excess kurtosis of the return distributions. However, the basic GARCH formulation is rather simple and restrictive and cannot address all the peculiarities that are observed in the financial data. Most notable exemptions are the asymmetric effect of positive and negative shocks on volatility, and possible correlation between the return and volatility. The GARCH models also tend to overpredict volatility because they respond slowly to large isolated returns. As a result numerous improvements have been suggested in the literature (see e.g. Franses & Dijk, 2000; Tsay, 2005). However, in order to avoid adding unnecessary complexity to the model, we have not implemented any of these variants of the GARCH model, although a better in-sample fit might be possible with these more complex models.

The practical modelling steps for GARCH model can be done in a similar fashion as for ARIMA models above. According to Tsay (2005), the model building procedure can be devided into the following four steps:

1. Specify a mean equation by testing for serial dependence in the data and, if necessary, build a model (e.g. an ARIMA model) for the time series to remove any linear dependence.

2. Use the residuals of the mean equation to test for ARCH/GARCH effects.

3. Specify a volatility model if ARCH/GARCH effects are statistically significant and perform a joint estimation of the mean and volatility equations.

4. Check the fitted model carefully and refine it if necessary.

For most asset return series serial correlations are weak, if any. Thus, building a mean equation amounts to removing the sample mean from the data if the sample mean is significantly different from zero. In financial application typically only lower order models are considered (e.g. GARCH(1,1), GARCH(1,2) or GARCH(2,1)). As in the case of ARIMA modelling, the ranking of the GARCH model candidates can, for example, be based on the AICC criteria.

The GARCH model is fitted to our index series in Appendix C. For the logarithmic index p_t of S&P500 yearly data 1925-2006, we chose the ARIMA(0,1,0)-ARCH(1) model, i.e. the mean equation is modelled by ARIMA(0,1,0), i.e. a random walk with a drift, and the volatility by the ARCH(1). The model can be written as

$$p_t = \alpha + p_{t-1} + z_t\sqrt{h_t} \quad \text{and} \quad h_t = \omega + \alpha_1\epsilon_{t-1}^2, \tag{4.16}$$

or, alternatively, in terms of the log return as

$$y_t = \alpha + z_t\sqrt{h_t} \quad \text{and} \quad h_t = \omega + \alpha_1\epsilon_{t-1}^2, \tag{4.17}$$

where $\omega > 0$, $\alpha_1 \geq 0$ and $t = 1, \ldots, 82$. The estimated model is

$$y_t = 0.1163 + z_t\sqrt{h_t} \quad \text{and} \quad h_t = 0.0183 + 0.5829\epsilon_{t-1}^2. \tag{4.18}$$

According to the model the expected yearly log return is 11.6%, which is somewhat higher than that of the pure ARIMA model. The relatively large α_1 coefficient indicates that volatility responds quite intensively to market movements. Since there are no lagged conditional variances included in the model the shocks to conditional variance die out quickly. By taking a look at the plotted time series in Figure A.1 in Appendix A, one may surely notice that the estimated model indeed reflects the features of our yearly data series quite well. Large shocks can be easily identified in the data (e.g. the Great Depression of the 30's, World War II, 1973-1974 and 2000-2002) while volatilty clustering is more difficult to observe. This is also consistent with the common view that GARCH effects may not be so apparent at lower-frequency data series.

For the log index of the shorter yearly index series, namely S&P500 yearly index 1925-2006, the random walk with a drift seems to be appropriate. That is, log returns follow the ARMA(0,0) model with a drift. According to our investigations, the ARCH/GARCH terms do not improve the mean fit. This is consistent with our visual observation that a more calm yearly return series is obtained if the starting date of the series is set to year 1954 instead of 1925. The exclusion of the Great Depression of the 30's and World War II leaves a yearly return series that does not show much deviation from stationarity.

For the monthly log index we chose the ARIMA(0,1,0)-GARCH(1,1) model. For the log return, the model can be written as

$$y_t = \alpha + z_t\sqrt{h_t} \quad \text{and} \quad h_t = \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{4.19}$$

where $\omega > 0$, $\alpha_1 > 0$, $\beta_1 \geq 0$ and $t = 1, \ldots, 624$. The estimated model becomes

$$y_t = 0.00863 + z_t\sqrt{h_t} \quad \text{and} \quad h_t = 0.000075 + 0.0967\epsilon_{t-1}^2 + 0.8645h_{t-1}. \tag{4.20}$$

As expected for the higher-frequency data series, the model now reflects the feature of volatility clustering. The relatively high value of β_1 indicates that shocks to conditional variance take a long time to die out, so volatility is persistent. According to the model the expected yearly log return is 10.4%.

A more detailed description of the fitting procedure, the estimated models and the model diagnostics with conclusions are given in Appendix C.

4.8 Regime-switching models

We have discussed above how the GARCH models can be used to address the issue of changing regimes with respect to volatility. But there can also be different regimes for investment returns, i.e. we observe periods of significantly lower or higher returns than on the average. For example, from Table 2-7 of Morningstar (2007) we can observe that the compound annual returns for 10-year holding periods of S&P 500 total return index vary between -0.89 and 20.06 percents during 1925-2006. In the following, we will see how the changing regimes can be taken into account in equity return modelling. Note that this section is mathematically more advanved than the previous ones.

There are two general approaches to model these so-called regime-switches. The first class of models assumes that the regimes can be determined by an observable variable, while the models in the second class assume that the regime is determined by an underlying unobservable stochastic process (see e.g. Franses & Dijk, 2000). In the following, we will present two models belonging to the latter class. One of them is the Hamilton (1989) model which we will discuss in the end of this section. The other model is given by

$$y_t = \alpha_0 + s_t\alpha_1 + (1 - s_t)\epsilon_t^0 + s_t\epsilon_t^1, \tag{4.21}$$

where y_t is the log return and the state variable s_t determines if the expected stock return is high ($s_t = 0$) or low ($s_t = 1$). The parameters α_0 and α_1 are unobservable and must be estimated. We set α_1 to be negative. The error processes $\epsilon_t^0 \sim \mathrm{N}(0, \sigma_0^2)$ and $\epsilon_t^1 \sim \mathrm{N}(0, \sigma_1^2)$ are i.i.d. Gaussian processes. Equation 4.21 can be rewritten in the form

$$y_t = \begin{cases} \alpha_0 + \epsilon_t^0, & \text{where } s_t = 0, \\ \alpha_0 + \alpha_1 + \epsilon_t^1, & \text{where } s_t = 1. \end{cases}$$

The transitions between the states are controlled by the first-order Markov process with transition probabilities

$$\begin{aligned} \mathrm{P}(s_{t+1} = 0 | s_t = 0) &= p, \\ \mathrm{P}(s_{t+1} = 1 | s_t = 0) &= 1 - p, \\ \mathrm{P}(s_{t+1} = 0 | s_t = 1) &= 1 - q, \\ \mathrm{P}(s_{t+1} = 1 | s_t = 1) &= q. \end{aligned}$$

Thus, the transition matrix is given by

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}.$$

The stationary probabilities $\boldsymbol{\pi} = (\pi_0, \pi)'$ of the Markov chain satisfy the equations $\boldsymbol{\pi}'\mathbf{P} = \boldsymbol{\pi}'$ and $\boldsymbol{\pi}'\mathbf{1} = 1$, where $\mathbf{1} = (1, 1)'$.

This model can be estimated, for example, by using the Gibbs sampler, introduced by Geman & Geman (1984) in the context of image restoration. Examples of Gibbs sampling can be found in Gelfand et al. (1990) and Gelman et al. (2004). The advantage of using the Bayesian approach is that we need not rely on asymptotic inference and that the inference on the state variables is not conditional on the parameter estimates. Carlin et al. (1992) provide a general approach to the use of the Gibbs sampler in nonlinear state-space modelling.

Gibbs sampling, also called alternating conditional sampling, is a useful algorithm for simulating multivariate distributions for which the full conditional distributions are known. Let us assume that we wish to simulate the random vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p)$ whose subvectors $\boldsymbol{\theta}_i$ have known conditional distributions $p(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{(-i)})$, where $\boldsymbol{\theta}_{(-i)} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, ..., \boldsymbol{\theta}_p)$. In each iteration the Gibbs sampler goes through $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p$ and draws values from their conditional distributions $p(\boldsymbol{\theta}_i | \boldsymbol{\theta}_{(-i)})$ where the conditioning subvectors have been set at their most recently simulated values. It can be shown that this algorithm produces an ergodic Markov chain whose stationary distribution is the desired target distribution of $\boldsymbol{\theta}$. In Bayesian inference one can use the Gibbs sampler to simulate the posterior distribution if one is able to generate random numbers or vectors from all the full conditional posterior distributions.

To simplify some of the expressions we will use the following notations: $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$ and $\mathbf{s} = (s_1, s_2, \ldots, s_T)'$. The vector of all parameters is denoted by $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \sigma_0^2, \sigma_1^2, p, q)'$ and the states at the other time points than t are denoted by $\mathbf{s}_{(-t)}$.

In order to facilitate computations, we use the following prior distributions:

$$
\begin{aligned}
&p(\alpha_0) \propto 1,\\
&p(\alpha_1) \propto \mathrm{N}(m, v^2) \times \mathrm{I}(\alpha_1 < -0.01),\\
&\sigma_0^2 \sim \text{Inv-}\chi^2\left(v_{\sigma_0^2}, s^2_{\sigma_0^2}\right),\\
&\sigma_1^2 \sim \text{Inv-}\chi^2\left(v_{\sigma_1^2}, s^2_{\sigma_1^2}\right),\\
&p \sim \mathrm{Beta}(\alpha_p, \beta_p),\\
&q \sim \mathrm{Beta}(\alpha_q, \beta_q),
\end{aligned}
$$

where $\mathrm{I}(\alpha_1 < -0.01)$ denotes the indicator function obtaining the value 1, if $\alpha_1 < -0.01$, and 0, otherwise. In order to implement the Gibbs sampler the full conditional posterior distributions of the parameters are needed:

$$
\begin{aligned}
&\{\alpha_0|\mathbf{y}, \mathbf{s}, \alpha_1, \sigma_0, \sigma_1\} \sim \mathrm{N}\left(\mu_{\alpha_0}, \sigma^2_{\alpha_0}\right),\\
&p(\alpha_1|\mathbf{y}, \mathbf{s}, \alpha_0, \sigma_0, \sigma_1) \propto \mathrm{N}(\mu_{\alpha_1}, \sigma^2_{\alpha_1}) \times \mathrm{I}(\alpha_1 < -0.01),\\
&\{\sigma_0^2|\mathbf{y}, \mathbf{s}, \alpha_0, \alpha_1\}\\
&\quad \sim \text{Inv-}\chi^2\left(v_{\sigma_0^2} + S_0, \frac{v_{\sigma_0^2} s^2_{\sigma_0^2} + \sum_{t=1}^T (1-s_t)(y_t - \alpha_0)^2}{v_{\sigma_0^2} + S_0}\right),\\
&\{\sigma_1^2|\mathbf{y}, \mathbf{s}, \alpha_0, \alpha_1\}\\
&\quad \sim \text{Inv-}\chi^2\left(v_{\sigma_1^2} + S_1, \frac{v_{\sigma_1^2} s^2_{\sigma_1^2} + \sum_{t=1}^T s_t(y_t - \alpha_0 - \alpha_1)^2}{v_{\sigma_1^2} + S_1}\right),\\
&\{p|\mathbf{s}\} \sim \mathrm{Beta}\left(\sum_{t=1}^T (1-s_t)(1-s_{t-1}) + \alpha_p, \sum_{t=1}^T s_t(1-s_{t-1}) + \beta_p\right),\\
&\{q|\mathbf{s}\} \sim \mathrm{Beta}\left(\sum_{t=1}^T s_t s_{t-1} + \alpha_q, \sum_{t=1}^T s_{t-1}(1-s_t) + \beta_q\right),\\
&\{s_t|\mathbf{s}_{(-t)}, \mathbf{y}, \boldsymbol{\eta}\}\\
&\quad \sim \mathrm{Bernoulli}\left(\frac{\mathrm{P}(s_t = 1|\mathbf{s}_{(-t)}, \boldsymbol{\eta}) p(\mathbf{y}|s_t = 1, \mathbf{s}_{(-t)}, \boldsymbol{\eta})}{\sum_{j=0}^1 \mathrm{P}(s_t = j|\mathbf{s}_{(-t)}, \boldsymbol{\eta}) p(\mathbf{y}|s_t = j, \mathbf{s}_{(-t)}, \boldsymbol{\eta})}\right), t = 1, \ldots, T,
\end{aligned}
$$

where we have denoted

$$
\mu_{\alpha_0} = \frac{\frac{\sum_{t=1}^T (1-s_t)(y_t - s_t\alpha_1)}{\sigma_0^2} + \frac{\sum_{t=1}^T s_t(y_t - s_t\alpha_1)}{\sigma_1^2}}{\frac{S_0}{\sigma_0^2} + \frac{S_1}{\sigma_1^2}}, \qquad \sigma^2_{\alpha_0} = \left(\frac{S_0}{\sigma_0^2} + \frac{S_1}{\sigma_1^2}\right)^{-1},
$$

$$
\mu_{\alpha_1} = \frac{\frac{m}{v^2} + \frac{\sum_{t=1}^T s_t(y_t - \alpha_0)}{\sigma_1^2}}{\frac{1}{v^2} + \frac{S_1}{\sigma_1^2}}, \qquad \sigma^2_{\alpha_1} = \left(\frac{1}{v^2} + \frac{S_1}{\sigma_1^2}\right)^{-1},
$$

and

$$
S_0 = \sum_{t=1}^T (1 - s_t), \qquad S_1 = \sum_{t=1}^T s_t.
$$

The notation Inv-$\chi^2(\nu, s^2)$ means the scaled inverse-chi-square distribution, defined as $\nu s^2/\chi^2_\nu$, where χ^2_ν is a chi-square distributed random variable with ν degrees of freedom.

Note that the probability of state 1 at time t, given the states at the other time points, $\mathbf{s}_{(-t)}$, is easily calculated as

$$\mathrm{P}(s_t = 1|\mathbf{s}_{(-t)}, \boldsymbol{\eta}) = \frac{\mathrm{P}(s_t = 1|s_{t-1}, p, q)\mathrm{P}(s_{t+1}|s_t = 1, p, q)}{\mathrm{P}(s_{t+1}|s_{t-1}, p, q)}, \quad 0 < t < T.$$

When we assume that $\alpha_1 = 0$, the model (4.21) simplifies to the form

$$y_t = \alpha_0 + (1 - s_t)\epsilon^0_t + s_t\epsilon^1_t, \tag{4.22}$$

where error processes $\epsilon^0_t \sim \mathrm{N}(0, \sigma^2_0)$ and $\epsilon^1_t \sim \mathrm{N}(0, \sigma^2_1)$ are i.i.d. Gaussian processes. Equation 4.22 can be rewritten in the form

$$y_t = \begin{cases} \alpha_0 + \epsilon^0_t, & \text{where } s_t = 0, \\ \alpha_0 + \epsilon^1_t, & \text{where } s_t = 1. \end{cases}$$

The conditional posterior distributions of the parameters are the same as earlier if α_1 is set at 0 in all the equations.

Next we move to our second model, that is, the Hamilton model. The Hamilton model may be expressed as $y_t = \alpha_0 + \alpha_1 s_t + z_t$, where y_t denotes the log return at time t, s_t the state (low or high) of the return and z_t a zero-mean stationary random process, independent of s_t. The parameters α_0 and α_1 and the state s_t are unobservable and must be estimated. We will assume that z_t is an autoregressive process of order r, denoted by $z_t \sim \mathrm{AR}(r)$. It is defined by the equation $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_r z_{t-r} + \epsilon_t$, where $\epsilon_t \sim \mathrm{N}(0, \sigma^2_\epsilon)$ is an i.i.d. Gaussian error process.

The Hamilton model was originally estimated by maximizing the marginal likelihood of the data series y_t. Then the probabilities of the states were calculated conditional on these maximum likelihood estimates. The numerical evaluation was done by a kind of nonlinear version of the Kalman filter. By contrast, we will use the Gibbs sampler. The technical details of the method can be found in Puustelli et al. (2008).

The estimation results for both models are given in Appendix D. For the monthly data a better fit than that of the GARCH model was found in both cases. For the yearly data there appeared to be convergence problems possibly due to insufficient amount of data.

4.9 Other nonlinear models

Finally, we mention two more nonlinear approaches that are based on empirically observed findings from equity market data. A jump-diffusion process is capable to generate atypical observations such as the one experienced on Black Monday in October 1987. This class of models adds for instance lognormally distributed random shocks to the Brownian motion e.g. according to the Poisson process. A model that gives a good fit to our long yearly data can be formulated by adding Gamma-distributed negative jumps to the random walk process according to Geometric-distributed waiting times. This model will be described in detail in a separate research report (to be published by Ronkainen and Alho).

Instead of modelling a single shock, the whole bubble formation phase can be modelled (which often precedes the downward shock). For an example of this approach see Kaliva & Koskinen (2008a,b).

4.10 Forecasting

Once the time series model which adequately describes the data is chosen, the next step is to use the model for predictive purposes. The goal is to predict l future values of a time series based on the data collected to the present time point t. That is, at time point t one wants to obtain values $\hat{y}_{t+1|t}, \hat{y}_{t+2|t}, \ldots, \hat{y}_{t+l|t}$, where $\hat{y}_{t+l|t}$ denotes a forecast of a future value y_{t+l} made at time t. There are many different ways to forecast the future value y_{t+l}. In time series analysis the forecast is commonly considered optimal, if it minimizes the mean square error (MSE)

$$\mathrm{E}\left[y_{t+l} - \hat{y}_{t+l|t}\right]^2. \tag{4.23}$$

The forecast that minimizes (4.23) is the conditional expectation of y_{t+l} at time t, that is $\hat{y}_{t+l|t} = \mathrm{E}\left[y_{t+l}|I_t\right]$.

Forecasting a stationary time series using ARMA models is quite straightforward in practice. To illustrate this, we first take a look at the forecasting with an AR(p) model. The forecasting procedure with an AR(p) model starts with generating the one-step-ahead forecast which is then used for a two-step-ahead forecast and so on. The optimal one-step-ahead prediction at time t is

$$\hat{y}_{t+1|t} = \hat{\alpha} + \hat{\phi}_1(y_t - \hat{\alpha}) + \hat{\phi}_2(y_{t-1} - \hat{\alpha}) + \ldots + \hat{\phi}_p(y_{t-p+1} - \hat{\alpha}),$$

and the two-step-ahead prediction is

$$\hat{y}_{t+2|t} = \hat{\alpha} + \hat{\phi}_1(\hat{y}_{t+1|t} - \hat{\alpha}) + \hat{\phi}_2(y_t - \hat{\alpha}) + \ldots + \hat{\phi}_p(y_{t-p+2} - \hat{\alpha}),$$

and the subsequent forecasts are obtained consequently. For an MA(q) model and $l \leq q$, the l-step-ahead prediction is

$$\hat{y}_{t+l|t} = \hat{\alpha} + \hat{\theta}_l \epsilon_t + \hat{\theta}_{l+1}\epsilon_{t-1} + \ldots + \hat{\theta}_q \epsilon_{t-q+l}.$$

For $l > q$ we simply have $\hat{y}_{t+l|t} = 0$. This is because the unknown errors $\epsilon_{t+1}, \epsilon_{t+2}, \ldots$ are set to zero. For an ARMA(p, q) model the l-step-ahead predictions are

$$\begin{aligned}\hat{y}_{t+l|t} = \hat{\alpha} + \hat{\phi}_1(\hat{y}_{t+l-1|t} - \hat{\alpha}) + \hat{\phi}_2(\hat{y}_{t+l-2|t} - \hat{\alpha}) + \ldots +\\ + \hat{\phi}_p(y_{t+l-s} - \hat{\alpha}) + \hat{\theta}_l\epsilon_t + \hat{\theta}_{l+1}\epsilon_{t-1} + \ldots + \hat{\theta}_q\epsilon_{t-q-l},\end{aligned}$$

for $l \leq q$. For $l > q$ only the AR part determines the forecasts.

Accurate predictions of volatilities are critical e.g. for option pricing, risk management and portfolio management. Volatility forecasts of an ARCH model can be obtained recursively as those of an AR model. Also in many GARCH models the forecasts are relatively simple to construct. This is because they take a simple analytic form and need no approximations or complicated simulations.

In the case of GARCH(1,1) model, the optimal h-step-ahead forecast of the conditional variance can be computed recursively from

$$\hat{h}_{t+l|t} = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1)\hat{h}_{t+l-1|t}, \tag{4.24}$$

which is obtained by noticing that the unexpected return at time $t + j$ is unknown for $j > 0$, but $\mathrm{E}(\epsilon^2_{t+j}) = h_{t+j}$. Note, that volatility term structure forecasts that are constructed from GARCH models mean-revert to the long-term level of volatility. If $\hat{\alpha} + \hat{\beta} < 1$, this long-term volatility level is

$$\hat{h} = \hat{\omega}/(1 - \hat{\alpha} - \hat{\beta}),$$

which is obtained by replacing $\hat{h}_{i+j}$ by $\hat{h}$ for all j in (4.24). For more detailed discussion of the forecasting with both linear and nonlinear models, we refer to the books of Franses & Dijk (2000) and Lai & Xing (2008). For practical issues specific to long-term forecasting, e.g. the role of expert judgement, we refer to the introduction given in Chapters 9–10 in Makridakis et al. (1998).

4.11 On option pricing

The fundamental assumption in option pricing is that there are no arbitrage opportunities in an efficient financial market. Therefore if one can find a (dynamic) trading strategy on the underlying security and the risk-free bond that exactly replicates the cash-flows of the derivative, this replicating portfolio must have the same price as the derivative. If the market is complete, every security can so be replicated, i.e. hedged, and a unique arbitrage-free price can be found. However, in insurance applications we need to keep in mind and adjust for the limitations of the basic derivative models (see Kaliva et al., 2007, and later chapters of this report for further discussion).

Another concept that needs special care is risk neutral valuation. As Wilmott (2001) states, it is a very important and very confusing topic in financial option valuation. Wilmott summarises the key issues as follows: "Real and risk neutral, this idea is probably more confusing than anything else in quantitative finance, but it is extremely important. ... But remember also that such risk-neutral valuation is only valid when hedging can be used to eliminate all risk. If hedging is impossible, risk-neutral valuation is meaningless." He also points out that if one uses simulations to get an idea what may happen to unhedged positions in the future, then the real world probabilities should be used for the underlying asset. This is very important to keep in mind when modelling non-hedgeable risks and carrying out risk management. Wilmott however underlines that for pricing of derivatives one has to use the risk neutral probabilities.

When pricing options or any derivatives one thus should use risk-neutral valuation. In risk neutral valuation one assumes a risk-neutral world which is a world where all investors are risk-neutral. In other words, investors are insensitive to risk and the value of a derivative does not depend upon the risk preference of the investor. So, to find the value of an option, one need to move to a risk-neutral world and then calculate the expected payoff at the expiration time and then discount it using risk-free rate. Typically, a model which is estimated from the historical data in a way we have done above (i.e. in the real world) differs from a model that should be used for pricing options (a risk-neutral model). A person doing the pricing should know how the two models differ [17]. We will discuss the risk-neutral world again in the context of interest rate modelling in Section 5.3.2. For a theoretical and practical pricing of non-life insurance contracts within a financial option pricing context see e.g. Holtan (2007).

[17] A risk-neutral model requires a change of probability measure, which is based on the Girsanov's theorem of stochastic calculus. A risk-neutral model is often achieved by setting the drift term of the real world process equal to the risk-free rate. This is because in the risk-neutral world the return on any traded investment is simply the risk-free rate.

5 Interest rate modelling

5.1 Introduction

In equity option pricing the interest rates are often assumed constant[1]. However, as options and derivatives are written either on interest rates, or on securities whose values are dependent on interest rates (e.g. bonds and swaps), a description of the stochastic behaviour of interest rates becomes essential. The assumption on the constant interest rate may be acceptable when the life of the option is only a few months. However, as longer time periods are considered, such an assumption becomes far too unrealistic. Recall that it is exactly the fluctuation of interest rates that the option buyer seeks to hedge.

An interest rate model is a probabilistic description of the future evolution of interest rates. It characterizes the uncertainty involved in interest rates. Interest rate modelling is a very important part of any life and pension insurance simulation exercise. This is because the valuation of the liabilities of life insurance companies and pension funds depends crucially on interest rates. In Solvency II there are two different approaches to the valuation of life insurance liabilities, namely: (1) hedgeable and (2) non-hedgeable approach. The first approach is applied to those cash-flows that can be hedged. It states that the cash-flows should be valued based on the respective hedging cost in the financial market (arbitrage-free market value). The second approach is used for those cash-flows that cannot be hedged. The valuation of the cash-flows should then be based on the so called best estimate (expected value) plus risk margin[2]. The future cash-flows are discounted by the relevant risk-free interest rates of the financial markets. Thus, the interest rates affect the valuation of the liabilities both through certain asset classes as well as through the discounting of the cash-flows. The proper modelling of interest rates is crucial for the appropriate valuation of the liabilities in Solvency II.

The behaviour of interest rates is complex providing one of the most challenging modelling areas. A convenient way to approach interest rate modelling is to first introduce bonds and bond yields.[3] A bond's yield, or yield-to-maturity, can be thought of as the interest earned per year from buying and holding the bond until its maturity. More specifically, it is the interest rate at which the present value of the stream of payments is exactly equal to

[1]This is the case, for example, when the Black-Scholes formula is used for short-dated options.

[2]The risk margin should address the uncertainty in the valuation of the best estimate in terms of regulatory capital costs. It should cover the risk linked to the future liability cash-flows over their whole time horizon.

[3]A bond is an obligation by the bond issuer to pay money to the bond holder according to rules specified at the time the bond is issued.

the current price (Luenberger, 1998). The (annual) yield-to-maturity, λ, of a coupon-bearing bond can be solved from the equation

$$P = \frac{F}{[1 + (\lambda/m)]^n} + \sum_{k=1}^{n} \frac{C/m}{[1 + (\lambda/m)]^k},$$

where P is the current price of the bond, F is the face value[4] of the bond, C is the yearly coupon payment[5], m is the number of coupon payments per year and n is the number of coupon periods remaining to maturity. A zero-coupon bond is a bond paying no coupons. The formula above of course implies that a rise in interest rates reduces bond prices while a fall has the opposite influence.

The interest rate that the issuer of a bond must pay is influenced by a variety of factors such as the bond's term to maturity and the creditworthiness of the issuer. The relationship between yield, or interest rate, and maturity for similar bonds at a particular point in time is called a yield curve, a zero-curve[6] or a term structure of interest rates. A yield curve can be constructed by plotting the yields of various available bonds that differ in maturity but are otherwise identical at a particular point in time. Usually, government bonds are used for derivation, since they are assumed default free, they exist in a wide range of maturities and are freely traded on the secondary markets. A common yield curve that investors consider is the U.S. Treasury yield curve. In this report we will concentrate on government bonds that have no probability of default.

The yield curve may take different shapes. The curve typically rises gradually with increasing maturity reflecting the fact that longer term interest rates are higher than shorter term interest rates. However, occasionally it may take on an inverted shape, where the yields decrease as the time to maturity increases. A flat yield curve is almost never observed. This suggests that investors require different rates of return depending on the maturity of the bond they are holding. A number of different theories have been proposed for the shape of the yield curve, the simplest of them being that the yield curve reflects the market's expectations of future interest rates. For example, according to this expectations theory, the upward rising yield curve can be explained by the market believing that the interest rate will rise. For a good review on the main theories proposed to explain the shape of the yield curve we refer to the article of Choudhry in Fabozzi (2002, Chapter 4).

As indirectly indicated above, the yield curve's shape and height changes through time. The fluctuation of the shape of the yield curve is called the evolution of the term structure of interest rates. The main challenge in the modelling of interest rates is to capture the random fluctuation of the yield curve, which is much more complex than the movements of a single stock or

[4]The face value is a specific amount paid by the bond at the date of maturity (expiration date).

[5]Coupon payment is the interest payment made to the bondholder. The coupon amount is described as a percentage of the face value. Here the coupon payments sum to C within a year.

[6]The yield curve is often represented in terms of a zero-coupon bond, i.e. a bond that provide no coupons.

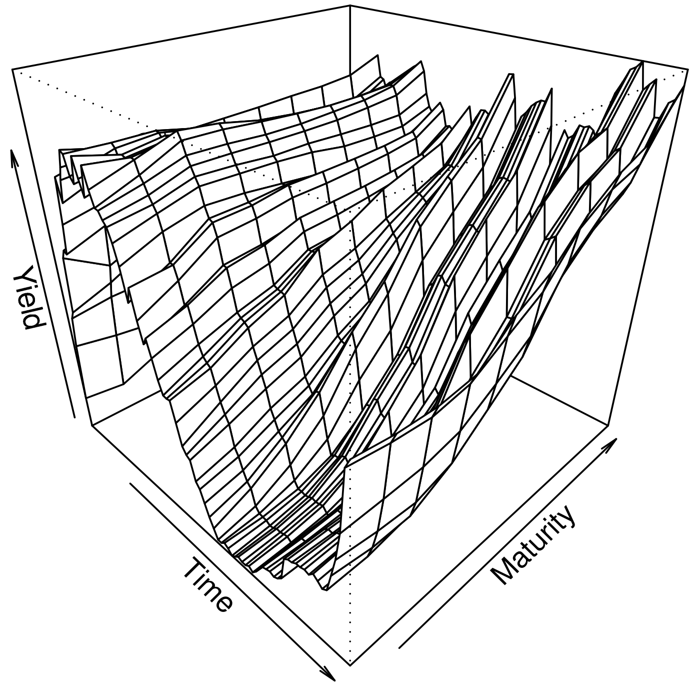


Figure 5.1: The evolution of the term structure of interest rates based on monthly observations on US Treasuries from July 2001 to March 2008.

index price. One can intuitively relate it to the difference in the dynamics of a scalar variable (e.g. stock index) and a vector (the yield curve). Figure 5.1 illustrates the evolution of the term structure of interest rates. The figure is based on monthly observations (US Treasuries) from July 2001 to March 2008.

Modeling the term structure of interest rates has a long tradition in finance and there are always many competing models available for each application at hand. There is no single ideal term structure model useful for all purposes, but the model needs to be chosen according to what is approporiate for the particular problem and data at hand. In the following we will take an insight into the interest rate term structure modelling. After introducing the basic notations used in the chapter, we will first discuss the four faces of an interest rate model (Section 5.3) and then list some desirable features of a term-structure model (Section 5.4). A few of the best-known one-factor short rate models are briefly introduced in Section 5.5. For the convenience, we will resort to the continous-time framework. In Section 5.6 one-factor short rate models are applied in the context of life insurance. The section is mathematically more advanced and is based on the paper of Luoma et al. (2008) which analyzes the role of the underlying asset and interest rate model in the market consistent valuation of life insurance policies. We then widen the perspective by moving from one-factor models to multifactor models. The HJM framework is introduced in Section 5.8 and the modelling in that context is demonstrated by a simulation study in Appendix F. Before closing the appendix we briefly discuss some issues on pricing. In Section 5.9 we give a brief introduction to the market models which have become widely popular in the past years.

5.2 Notations and relationships

Below we introduce the notations used in this chapter and also show some relationships:

$P(t,T)$ The price at time t of a zero-coupon bond that matures at time T, with $t \leq T$. Note that $P(t,t) = 1$ for all t.

$R(t,T)$ Spot rate. The spot rate $R(t,T)$ is the continuously compounded interest rate at time t implied by the price $P(t,T)$, that is

$$R(t,T) = -\frac{1}{T-t} \ln P(t,T), \quad t < T.$$

It follows that

$$P(t,T) = e^{-R(t,T)(T-t)}.$$

Interpretation: If we invest one euro at time t in a zero-coupon bond maturing at time T for $T-t$ years, then this will accumulate at an average rate of $R(t,T)$ over the whole period.

$F(t,T,S)$ Forward rate. The (continuously compounded annualized) forward rate at time t which applies between times T and S, is defined as

$$F(t,T,S) = \frac{1}{S-T} \ln \frac{P(t,T)}{P(t,S)}, \quad t \leq T < S.$$

Interpretation: Forward rate is the interest rate agreed at time t for an investment of one euro made at time T in a zero-coupon bond maturing at time S for $S-T$ years. Note that $F(t,t,S) = R(t,S)$.

$f(t,T)$ The (continuously compounded annualized) instantaneous forward rate at time t, is defined as

$$f(t,T) = \lim_{S \to T} F(t,T,S) = -\frac{\partial}{\partial T} \ln P(t,T), \quad t < T.$$

It follows that

$$P(t,T) = e^{-\int_t^T f(t,u)\ du}$$

and

$$R(t,T) = \frac{1}{T-t} \left(\int_t^T f(t,u)\ du \right).$$

r_t Short rate. The (continuously compounded annualized) instantaneous interest rate at time t is defined as

$$r_t = \lim_{T \to t} R(t,T) = R(t,t) = f(t,t), \quad t < T.$$

In practice one should take the short rate to be the yield on a liquid finite-maturity bond, say one of one month (Wilmott, 2001, pp. 286–287).

Note that $R(t,T)$, $f(t,T)$ and $P(t,T)$ all carry the same information. They are all functions of two variables: intiation time t and maturity time T. They provide three different but equivalent ways to represent the yield curve. The short rate $r(t)$ is a function of only one variable and hence contains less information. However, the short rate plays an important role in interest rate modelling.

5.3 The four types of an interest rate model

In this section the concepts arbitrage-free and equilibrium models, and risk neutral and realistic (real world) probabilities are explained. We also briefly discuss the choice between an arbitrage-free or equilibrium model, and the choice between risk neutral or realistic parameterizations of a model. These two dimension define four classes of model forms, each of which has its own proper use. For a proper discussion on the topic, we refer to the excellent article of Fitton & McNatt in Fabozzi (2002, Chapter 2). The following section is closely based on this particular article.

5.3.1 Arbitrage-free vs. equilibrium model

It is a common belief that market participants quickly take advantage of any opportunities for arbitrage[7] among financial assets, so that these opportunities do not exist for long. Thus, in many cases the term structure of the interest rates for different maturities is aimed to be modelled in an arbitrage-free way[8], that is, such that the interest rates implied by the model are consistent with the observed actual interest rates. "Arbitrage-free" is known as the law of one price. If one values the same cash flows in two different ways, one should get the same result for both (see e.g. Cheyette in Fabozzi, 2002, Chapter 1).

Arbitrage-free models assume some computationally convenient, but essentially arbitrary, random process underlying the yield curve. They take certain market prices as given, and adjust model parameters so that the models fit the prices exactly[9]. It should be realized, that an arbitrage-free model may achieve an exact fit to prices of assets in particular classes without regard to the reality. Examples of arbitrage-free models are the Ho & Lee, the Hull & White, the BDT and the Black & Karasinski models discussed in Section 5.5.

Equilibrium models attempt to capture the behaviours of the term structure over time. Rather than interpolating among prices at one particular point in time, they employ a statistical approach assuming that market prices are observed with some statistical error. Typically, equilibrium models put structure on the evolution of interest rates and then try to match the data as

[7]Arbitrage, also known as riskless profit and, more popularly, as free lunch, is the simultaneous purchase and sale of assets to make a profit from the difference in pricing.

[8]This is typically done under the risk neutral probability measure (see Section 5.3.2 below).

[9]In an arbitrage-free model, the drift term is, in general, dependent on time.

closely as the model will allow. Current term structure of interest rates hence is an output rather than an input in the model[10] (Hull, 1999).

Equilibrium models can be made to provide approximate fit to many of the term structures encountered in practice. However, they do not usually provide an exact fit. In certain cases, there may even be significant differences between the model fit and the observed actual rates, leading to arbitrage. This clearly is a disadvantage of the model. On the other hand, equilibrium models can have an advantage over arbitrage-free models in that they can be estimated from historical data when current market prices are unreliable or unavailable. The first short rate models being proposed in the financial literature were one-factor equilibrium models. Examples of such models are the Merton, the Rendleman & Bartter, the Vasicek and the CIR models introduced in Section 5.5. Some equilibrium models can be converted into arbitrage-free models by including a time-dependent drift term. Examples of such models are the Ho & Lee and the Hull & White models generalized from the equilibrium models of Merton and Vasicek, respectively.

5.3.2 Risk neutral vs. real world

When pricing interest rate derivatives, the task is to specify a random process for the instantaneous, risk-free interest rate called the short rate[11], which is the rate payable on an investment in default-free government bonds for a very short time period (cf. Section 5.2). The short rate is generally considered as the only truly riskless interest rate in financial markets. It is also believed to be the most important state variable driving the dynamics of the term structure of interest rates.

An investor in bonds (or any risky investment) subject to market risk expects to earn a risk-free return plus a (time-varying) risk premium, whose purpose is to take into account the aggregate risk preference of market participants. The spot rate for a particular term is then composed of the return expected under the random process for the short rate up to the end of that term, plus a term premium, an additional return to compensate the investor for the interest rate risk of the investment. Thus, in order to value a zero-coupon bond, it may seem necessary to not only know the random process for the short rate but also the term premium for every possible term.

In the valuation principle called risk neutral valuation, it is not necessary to separately identify the term premium embedded in each spot rate. The idea is to identify a set of spot rates such that investors' risk preferences do not affect the valuation of bonds or other interest rate derivatives. As described by Fitton & McNatt, this is eventually done by risk-adjusting the term structure model, that is, by changing probability distribution of the short rate so that the spot rate of every term is, under the new model, equal to the expected return from investing at the short rate over the same term.

[10] In arbitrage-free models the current term structure is an input. This means that while-constructing the model we take the observed actual rates and estimate the unobserved rates.

[11] The short rate is an abbreviation for the "short term interest rate". The short rate can be illustrated as being the short-term maturity edge of the yield surface in Figure 5.1.

This is accomplished by redefining the model so that, instead of being a random process for the short rate, it is a random process for the short rate plus a function of the term premium (for a mathematical formulation and more detailed description see the full article in Chapter 2 of Fabozzi, 2002).

As further described by Fitton & McNatt, the resulting risk neutral model can be thought of as a model for the true behaviour of the short rate in an imaginary world, where the investors do not require compensation for the extra risk in bonds of longer maturity. The important aspect of the risk neutral model is that the term premia, whatever their values, that exist in the marketplace are embedded in the interest rate process itself, so that the expected discounted value of a cash flow at the risk adjusted short rate is equal to the discounted value of the cash flow at the spot rate. Valuing assets without risk adjusting the model would require a more complicated discounting procedure. Under the parametrization of the risk neutral probability measure, however, an interest-sensitive instrument's price can be estimated by averaging the present values of its cash flows, discounted at the short term interest rates along each path of the short rate under which those cash flows occur. Thus, the specific change of variables that produces a risk neutral model simply makes the algebra easier than the others, because one can ignore risk preferences.

It is important to distinguish the risk neutral term structure model (specified under the risk neutral probability measure) from the realistic risk averse term structure process (specified under the real world probability measure). In the real world, term premia commonly are different from zero, while in scenarios generated by a risk neutral process all term premia are zero. As discussed above, risk neutral interest rate scenarios are preferred for pricing bonds. However, such scenarios that are lacking realism are not appropriate for all purposes. Realistic simulations, which can be provided only if the test environment is like the real environment, are desired especially for many risk management purposes e.g. for stress testing cases.

5.3.3 When to use each of the model types

The two dimensions, risk neutral versus real world and arbitrage-free versus equilibrium, define four classes of modelling approached, each of which has its appropriate use. The risk neutral and arbitrage-free type of model is the most familiar interest rate model form for most analysts. However, as shown in Table 5.1, it is not the only valid kind of a term structure model. From Solvency II point of view the risk neutral approach is related to mark-to-market valuation of liabilities while the real world approach is needed for the calculation of Solvency Capital Requirement. For a comprehensive discussion on the use of each of the modelling approaches we refer to Fabozzi (2002, Chapter 2).

Table 5.1: When to use each of the model types (Fabozzi, 2002, Table 2.1)

Model classification	**Risk neutral**	**Real world**
Arbitrage-free	• Current pricing, where input data (market prices) are reliable	• Unusable, since term premium cannot be reliably estimated
Equilibrium	• Current pricing, where inputs (market prices) are unreliable or unavailable • Horizon pricing	• Stress testing • Reserve and asset adequacy testing

5.4 Desirable features

Cairns (2004) discusses basic characteristics which, in varying degree, are desirable but not essential for the development of a term-structure model. We list some of these charateristics below. For the comprehensive list we refer to Cairns (2004, pp. 53–55).

1. Models should not allow negative interest rates.

2. Models should incorporate mean reversion. By mean reversion (or autoregression) is meant a tendency of a stochastic process to remain near or return over time to a long-run average value. Interest rates typically are mean reverting while stock prices do not exhibit this tendency.

3. Model formulaes obtained for bond and derivative prices should be simple. However, a more important property of a model is, of course, that a proposed model gives a good approximation to what we observe in reality.

4. Bond and derivative prices are simple to calculate numerically. This relaxes the requirement that prices be available using analytical formulae and is a reflection of the existence of increasing computing power.

5. Models are flexible enought to cope with new and more complex derivative products.

6. Models produce dynamics that are realistic.

7. Models fit historical data well (in the statistical sense) or at least adequately.

5.5 On one-factor short rate models

One-factor models provide a solid foundation upon which we can build more complex models. These models have only a single source of randomness and they are often built for the short rate, which is commonly considered as the only truly riskless interest rate in financial markets. The risk neutral valuation requires that one knows the sequence of short rates for each scenario. Since an interest rate model must provide this information, many interest rate models are simply models of the stochastic evolution of the short rate. The desirability of short rate models also follows from that they have the Markov property (”absence of memory”) meaning that the evolution of the short rate at each instant depends only on its current value and not on how it got there.

A short rate model is a mathematical model that describes the future evolution of interest rates by describing the future evolution of the short rate. The one-factor short rate models often use a stochastic differential equation (SDE) to represent the short rate r_t. Commonly, r_t is assumed to be governed by an equation of the Ito type,

$$dr_t = \mu(r_t, t)\ dt + \sigma(r_t, t)\ dW_t,$$

where $\mu(.)$ and $\sigma(.)$ are the instantaneous drift and standard deviation (volatility) of the process, respectively, and W_t is the standard Wiener or Brownian motion process. The left-hand side of the equation is the change in the short rate over the next instant. Given an initial condition r_0, the equation defines a stochastic process r_t. Many such models have been proposed as being good approximations to actual interest rate processes. We should note that the manipulation of SDEs requires some special stochastic calculus rules (see e.g. James & Webber, 2000). Intuitively SDEs can be described as continuous-time counterparts of discrete time processes that we discussed in the context of equity modelling.

Below we list a few of the best-known one-factor short rate models. All processes for the short rate will be presented in the risk neutral world. Before we go to the short rate models it is worth noting that a change of measure from real world to the risk neutral will only affect the drift term. The volatility term is the same under both the measures. Using historical data we can estimate[12] the parameters of the models under the real world measure. To move from one probability measure to another we need know (or approximate) the market price of risk. For a comprehensive introduction to estimation and calibration techniques we refer to James & Webber (2000). A nicely written paper by Zeytun & Gupta (2007) may help the reader to better understand the model specification and parameter estimation under the risk neutral respective real world probability measure.

1. **Merton model**:
   The model was proposed by Merton in 1973. It assumes that the short rate follows a Brownian motion with a drift, that is,

   $$dr_t = \alpha\ dt + \sigma\ dW_t,$$

   where α and σ are constants. The model assumes that r_t is normally distributed. This implies that interest rates can become negative with

[12]The process of fitting an interest rate model to historical and current data is known as estimation. The process is often called a model calibration if the parameters of the models are being estimated from current market data (James & Webber, 2000).

a positive probability, which clearly is an undesirable property of the model at least for the nominal interest rates. The parameter σ in the model represents the local volatility of short rates, i.e. volatility per unit of time dt.

2. **Rendleman & Bartter model**:
   In 1980 Rendleman and Bartter suggested a model

   $$dr_t = \alpha r_t \, dt + \sigma r_t \, dW_t,$$

   where α and σ are constants. The model is the lognormal model suggested for the equity price and presented in Section 4.5. It assumes that the short rate follows a GBM. The GBM leads to lognormal distribution of short rates. Hence, unlike the Merton model, the Rendleman & Bartter model does not permit negative interest rates. The Rendleman & Bartter model is nowadays rarely advocated as a realistic model of the short rate process. One reason is that the model does not attempt to model a mean reversion property often exhibited by the interest rates in practice. Interest rates do not usually exhibit the long-term exponential growth seen in the equity markets.

3. **Vasicek model**:
   One of the most widely used models was developed by Vasicek in 1977. It assumes that the short rate follows an Ornstein-Uhlenbeck process[13]

   $$dr_t = \alpha(\beta - r_t) \, dt + \sigma \, dW_t,$$

   where α, β and σ are constants. The model incorporates mean reversion through the instantaneous drift term $\alpha(\beta - r)$. It is incorporated in the Vasicek model so that if the interest rate is above the long-run mean, that is, $r_t > \beta$, the drift term will be negative and the short rate is pulled down. Likewise, if the short rate is less than the long-run mean, that is, $r_t < \beta$, the drift will be positive and the rate is pulled upward. The coefficient α is the speed of adjustment of the interest rate towards its long-run normal level. A drawback of the Vasicek model is that the volatility of the short rate is constant. Another undesirable feature of the model is that it permits negative interest rates.

4. **Cox, Ingersoll & Ross (CIR) model**:
   Cox, Ingersoll and Ross proposed in 1985 an alternative to the Vasicek model where rates are always non-negative. The model is derived from the equilibrium conditions of the economy and it considers an interest rate process of the type

   $$dr_t = \alpha(\beta - r_t) \, dt + \sigma\sqrt{r_t} \, dW_t,$$

   where α, β and σ are constants. The model has the same mean reverting drift as the Vasicek model. The volatility part of the model is, however,

[13]The Ornstein-Uhlenbeck process is the continuous-time analogue of the discrete-time AR(1) process.

different. In the CIR model the volatility of the short rate is assumed to be proportional to the level of interest rates through $\sqrt{r_t}$. The model hence allows more variability at times of high interest rates and less variability when rates are low. Cairns (2004) regards the CIR model as the first tractable model that keeps rates of interest positive. However, despite the fact that the CIR model is nowadays widely used, it is quite incapable to fit different types of shapes for the yield curve. The CIR model presents a good fit only for upward term structures of interest rates.

5. **Ho & Lee model**:
   Ho and Lee proposed their model in 1986. The model extended the Merton model to fit a given initial yield curve perfectly in a discrete-time framework. It actually was the first arbitrage-free model of the term structure of interest rates. The Ho & Lee model is one of the simplest models that can be calibrated to market data. The continuous-time limit of the model is

$$dr_t = \theta(t)\ dt + \sigma\ dW_t,$$

   where σ is a constant repsenting the instantaneous volatility of the short rate and $\theta(t)$ is a time-dependent drift that defines the average direction that the short rate moves at time t. The drift $\theta(t)$, which is independent of the short rate, is chosen so that the model generates the observed yield curve. Like in the Merton model, the volatility of the short rate is assumed to be constant. The Ho & Lee model is easy to apply and provides an exact fit to the current term structure of interest rates. The main drawback of the model is that it has no mean reversion property. The model also has a disadvantage of permitting negative interest rates.

6. **Hull & White model**:
   Hull and White proposed in 1990 a model that extended the Vasicek model to provide an exact fit to an initial yield curve. The Hull & White model is hence sometimes also referred to as an extended Vasicek model. One version of it is

$$dr_t = (\theta(t) - \beta r_t)\ dt + \sigma\ dW_t, \tag{5.1}$$

   where β and σ are constants, and $\theta(t)$ is a function of time. The model is also often expressed as

$$dr_t = \alpha(\mu(t) - r_t)\ dt + \sigma\ dW_t,$$

   where α and σ are constants, and $\mu(t)$ is a function of time chosen to ensure that the model fits the initial yield curve. Now $\mu(t)$ has a straightforward interpretation of a local mean reversion level, i.e. at time t the short rate reverts to $\mu(t)$ at rate α. The model can also be characterized as the Ho & Lee model with the mean reversion property. The main difference to the Vasicek model is that in the Vasicek model $\mu(t) = \beta$, i.e. a constant.

7. **Black, Derman & Toy (BDT) model**:
   Black, Derman and Toy suggested in 1990 a generalization of the Hull & White model, which matches both the observed initial yield curve and market volatility data. The continuous-time limit of the model is

$$d \ln r_t = \left(\theta(t) - \frac{\sigma'(t)}{\sigma(t)} \ln r_t \right) \, dt + \sigma(t) \, dW_t,$$

   where $\theta(t)$ and $\sigma(t)$ are independent functions of time, and $\sigma'(t)$ is the partial derivative of $\sigma(t)$ with respect to t. The functions $\theta(t)$ and $\sigma(t)$ are chosen so that the model fits the existing term structure of interest rates and volatilities, respectively. Since the changes in the short rate are lognormally distributed, the problem of negative interest rates is avoided. A drawback of the BDT model is that in certain cases the model may be mean fleeing rather than mean reverting. A version often implemented in practice holds the future short rate volatility constant. The convergence limit then reduces to

$$d \ln r_t = \theta(t) \, dt + \sigma \, dW_t,$$

   which is virtually indentical to the Ho & Lee model, except that the underlying variable is $\ln r_t$ rather than r_t. The BDT model is also sometimes referred to as the exponential Vasicek model.

8. **Black & Karasinski model**:
   In 1991, Black and Karasinki suggested a generalization of the BDT model, in which the reversion rate was explicitly decoupled from the volatility. The continuous-time limit representation of the model is

$$d \ln r_t = (\theta(t) - \beta(t) \ln r_t) \, dt + \sigma(t) \, dW_t,$$

   where $\theta(t)$, $\beta(t)$ and $\sigma(t)$ are three independent functions of time, or

$$d \ln r_t = \alpha(t)(\ln \mu(t) - \ln r_t) \, dt + \sigma(t) \, dW_t,$$

   where $\mu(t)$, $\alpha(t)$ and $\sigma(t)$ have the interpretations of the target rate, the mean reversion and the local volatility in the expression for the local change in $\ln r_t$, respectively. The three time-dependent functions are chosen to match three features of the world (see Black & Karasinski, 1991). The Black & Karasinski model avoids the problem of negative interest rates and it has become quite popular amongst practitioners mainly due to its good fitting quality to market data.

A more detailed description on the above one-factor models can be found in many books dealing with interest rate models such as Cairns (2004), Hull (1999) and Brigo & Mercurio (2001). For an introduction on the parameter estimation and model calibration we refer to the same books and Lai & Xing (2008). Chan et al. (1992) has conducted an empirical comparison of various short-term interest rate models. Lai & Xing (2008) briefly summarize the results.

One-factor models are simple and tractable. However, as discussed in Cairns (2004), these models generally fail many of the desirable characteristics listed in Section 5.4. This can be explained by one-factor models being dependent on only a single factor which makes the models too inflexible and unrealistic. The practical sufficiency of one-factor model is discussed in Section 5.7.1. Multifactor models which incorporate more than one factor and are hence more flexible are considered in Section 5.7.2. However, before going into the discussion of these two topics, we will in the next section take a look at the work of Luoma et al. (2008) in which one-factor short rate models are applied in the context of life insurance.

5.6 Interest rate models in the market consistent valuation of life insurance policies

In this section we introduce a bivariate modelling of stochastic interest rate and equity index. These models are used by Luoma et al. (2008) in a market consistent valuation of a participating life insurance contract. The contract is, in this setup, an American-style path-dependent derivative. A Bayesian approach is utilized in the estimation of the underlying processes. Specifically, the processes are estimated using the Markov Chain Monte Carlo method, and their simulation is based on their posterior predictive distribution, which is, however, adjusted to give risk-neutral dynamics. The contract prices are estimated using the regression method.

The focus is on a novel application of advanced theoretical and computational methods, which enable us to deal with a fairly realistic valuation framework and to address model and parameter error issues. Our empirical results support the use of elaborated instead of stylized models for asset dynamics in practical applications.

5.6.1 The model

The short-term interest rate model we use is a generalization of the Vasicek and CIR models. It was introduced by Chan et al. (1992), who provide a useful summary of short-term interest rate models in their paper.

We assumed that the dynamics of riskless short-term rate r_t and stock index S_t are described by the following system of SDEs:

$$\mathrm{d}r_t = \kappa(\xi - r_t)\mathrm{d}t + \sigma r_t^{\gamma}\mathrm{d}W_t^{(1)}, \tag{5.2a}$$

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \nu S_t^{1-\alpha}\mathrm{d}W_t^{(2)}, \tag{5.2b}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are two standard Brownian motions, correlated through $W_t^{(2)} = \rho W_t^{(1)} + \sqrt{1-\rho^2}W_t^{(3)}$, where $W_t^{(1)}$ and $W_t^{(3)}$ are independent standard Brownian motions under the real-world probability measure. Thus the correlation of $W_t^{(1)}$ and $W_t^{(2)}$ is ρ.

By parameter restriction the short-term interest rate model becomes the following: If $\gamma = 0$, the model becomes the Vasicek model, and, if $\gamma = \frac{1}{2}$, it becomes CIR model. The stock index model becomes a geometric Brownian motion if $\alpha = 0$.

Substituting $Z_t^{(1)} = W_t^{(1)}$ and $Z_t^{(3)} = W_t^{(3)} + (\mu - r_t)\nu^{-1}(1-\rho^2)^{-1/2}S_t^{\alpha}dt$, the system of SDEs (5.2a) and (5.2b) can be written in the form

$$\text{(5.3a)} \qquad \mathrm{d}r_t = \kappa(\xi - r_t)\mathrm{d}t + \sigma r_t^{\gamma}\mathrm{d}Z_t^{(1)},$$

$$\text{(5.3b)} \qquad \mathrm{d}S_t = r_t S_t \mathrm{d}t \;+\; \nu S_t^{1-\alpha}\mathrm{d}Z_t^{(2)},$$

where $Z_t^{(2)} = \rho Z_t^{(1)} + \sqrt{1-\rho^2}Z_t^{(3)}$. Now a risk-neutral probability measure $\mathbb{Q}$ may be introduced by assuming that $Z_t^{(1)}$ and $Z_t^{(3)}$ are two independent standard Brownian motions under this measure. It can then be shown that the discounted price $\tilde{S}_t = S_t\exp(-\int_0^t r_s ds)$ is a martingale under $\mathbb{Q}$.

To our knowledge, the transition densities of the bivariate process described by (5.2a) and (5.2b) do not have a closed form solution, and its Euler discretization is used to estimate the unknown parameters κ, ξ, σ, γ, μ, ν and α. Accordingly, we simulate the risk-neutral process using the Euler discretization of (5.3a) and (5.3b).

In order to obtain numerical stability in estimation, we reparametrize the model (5.2a) as

$$\mathrm{d}x_t = (\beta - \kappa x_t)\mathrm{d}t + \tau x_t^{\gamma}\mathrm{d}W_t^{(1)},$$

where $x_t = 100\ r_t$ (the interest rate given in percentages), $\beta = 100\ \kappa\xi$ and $\tau = (100)^{1-\gamma}\sigma$. Assuming that the bivariate process has been observed at equally-spaced time points $0, \delta, ...N\delta$, the likelihood function can be written in the form

$$\text{(5.4)} \quad p(y|\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\tau^2 x_{(i-1)\delta}^{2\gamma}\delta}} \exp\left(-\frac{\left(\Delta x_{i\delta} - (\beta - \kappa x_{(i-1)\delta})\delta\right)^2}{2\tau^2 x_{(i-1)\delta}^{2\gamma}\delta}\right)$$
$$\times \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\nu^2 S_{(i-1)\delta}^{2(1-\alpha)}(1-\rho^2)\delta}}$$
$$\exp\left(-\frac{\left(\Delta S_{i\delta} - \mu S_{(i-1)\delta}\delta - \nu S_{(i-1)\delta}^{1-\alpha}\rho\Delta W_{i\delta}^{(1)}\right)^2}{2\nu^2 S_{(i-1)\delta}^{2(1-\alpha)}(1-\rho^2)\delta}\right),$$

where y is data, $\theta = (\mu, \nu, \alpha, \beta, \kappa, \tau, \gamma, \rho)$, $\Delta x_{i\delta} = x_{i\delta} - x_{(i-1)\delta}$, $\Delta S_{i\delta} = S_{i\delta} - S_{(i-1)\delta}$ and

$$\Delta W_{i\delta}^{(1)} = \frac{x_{i\delta} - x_{(i-1)\delta} - (\beta - \kappa x_{(i-1)\delta})\delta}{\tau x_{(i-1)\delta}^{\gamma}}.$$

5.6.2 Bayesian estimation

We use Bayesian methods to estimate the unknown parameters of the stock index and interest rate models. This makes it possible to take parameter uncertainty into account when evaluating the fair prices of derivatives. We take the model uncertainty into account by using a sufficiently general, continuously parametrized family of distributions (see Gelman et al., 2004). The

Metropolis algorithm introduced by Metropolis et al. (1953) is used to simulate the joint posterior distribution of unknown parameters. The posterior density is proportional to the product of the prior density and the likelihood,

$$p(\theta|y) \propto p(\theta)p(y|\theta).$$

We use an improper uniform prior distribution

$$p(\theta) \propto \begin{cases} 1 & \text{when } |\rho| < 1 \text{ and } \min(\kappa, \xi, \sigma, \nu, \alpha) > 0 \\ 0 & \text{otherwise} \end{cases}$$

and the posterior function is thus proportional to the likelihood (5.4) in a feasible region of parameters. The estimation results are summarized in Appendix E.

5.6.3 Insurance contract

The goal is to price a participating life insurance contract, and it is done from insurance company's perspective. The contract may be viewed as an option whose seller is an insurance company and buyer its client. The contract consists of two parts. The first part is a guaranteed interest and the second part a bonus depending on the yield of some total return equity index. The amount of savings in the insurance contract at time t_i is denoted by $Y(t_i)$. Then its growth during a time interval of length $\delta = t_{i+1} - t_i$ is given by

$$\log \frac{Y(t_{i+1})}{Y(t_i)} = g\,\delta + b \max\left(0, \log \frac{X(t_{i+1})}{X(t_i)} - g\,\delta\right), \tag{5.5}$$

where $X(t_i) = \sum_{j=0}^{q} S(t_{i-j})/(q+1)$ is a moving average of the total return equity index $S(t_i)$. One can see from (5.5) that the accumulated capital is guaranteed to the customer. The guarantee rate g is fixed for one year at a time. It is set annually at kr_t, where r_t is the riskless short-term interest rate at time t and $k < 1$. The bonus rate b is the proportion of the excessive equity index yield that is returned to the customer. We use the time interval $\delta = 1/255$, where 255 is approximately the number of the days in a year on which the index is quoted. The model also incorporates a surrender (early exercise) option and the possibility for a penalty p which occurs if the customer reclaims the contract before the final expiration date. The parameters k, g, b and p are predefined by the insurance company. A further condition is that there will be a 1 % penalty if the contract is reclaimed during the first 10 working days. This condition essentially improves the estimation of the fair bonus rate, which is the main goal in Luoma et al. (2008).

This particular participating life insurance contract is in practice an American option with a path-dependent moving average feature. An American option gives the holder the right to exercise the option at any time up to the expiry date. The pricing of an American option is based on an optimal exercising strategy. The idea is to compare the dicounted immediate exercise value with the corresponding discounted continuation value. In pricing a simple but powerful least squares method introduced by Longstaff & Schwartz (2001) is adopted.

One should note that the methods used in determining prices of American options are approximative. In addition to Monte Carlo simulation errors, there is a modelling error related to the choice of regressors in the least squares method. These sources of error are taken into account in the confidence intervals which are also provided in Luoma et al. (2008).

5.7 On multifactor models

5.7.1 Are one-factor models sufficient for practical purposes?

The short rate is a key interest rate in all the one-factor models constructed above, even though this rate cannot be directly observed[14]. The short rate may constitute the fundamental coordinate with which the whole yield curve can be characterized. Knowledge of the short rate and of its distributional properties leads to knowledge of bond prices, from which one can then construct the whole zero-coupon interest rate curve. The evolution of the whole curve is hence characterized by the evolution of the single quantity namely the short rate (Brigo & Mercurio, 2001).

A one-factor model does not often capture the subtleties of the yield curve that are important for particular contracts/products. A look at the historical data in Figure 5.1, for example, shows that changes in interest rates with different maturities are not perfectly correlated. However, this is indeed what one-factor models assume.[15] The Vasicek model, for example, assumes that the thirty-year interest rate at a given instant is perfectly correlated with, say, the three-month rate at the same instant. This means that a shock to the interest rate curve at time t is transmitted equally through all maturities, and the curve, when its initial point (the short rate r_t) is shocked, moves almost rigidly in the same direction (Brigo & Mercurio, 2001). Since in reality interest rates are known to exhibit in a different manner, a more satisfactory model of curve evolution is often needed.

According to Brigo & Mercurio (2001), many one-factor short rate models may prove useful when the product to be priced does not depend on the correlations of different rates but depends at every instant on a single rate of the whole interest rate curve, say for example the six-month rate. One-factor models may also be acceptable in the cases where two or more rates jointly influence the payoff at every instant. However, in such cases the real correlation between the rates need to be high enough so that the perfect correlation induced by the one-factor model provides an acceptable approximation. Hence,

[14]The one- or three-month interest rate is often taken to be the best available proxy for the short rate. It is not ideal, however, as the short rate is defined to have an instantaneous holding period, compared with which three months is a long way down the yield curve (James & Webber, 2000, p. 77).

[15]According to Hull (1999), the assumption on a single factor is not as restrictive as it might appear. Although one-factor model implies that all rates move in the same direction over any short time interval, they do not require that all move by the same amount. Hull also adds that the term structure need not always have the same shape, but a fairly rich pattern of term structures can occur under a one-factor model.

such rates usually need to be close, say, for example, six-month and one-year rates.

Since the obligations of pension funds or life insurers stretch far into the future, the model used should not only fit the short end of the yield curve well but also the long end. It is relevant that such a model is able to capture the nature of correlations among rates of different maturities. In general, whenever correlation plays a more relevant role or when higher precision is required, there is a need for a model providing a more realistic yield curve evolution. For example, an option may be defined in terms of the difference between the one- and five-year rates. As noted by Cairns (2004), in such a case one-factor model would possibly overprice this contract because of its assumption that the underlying rates are perfectly, non-linearly correlated. The nature of correlations among rates of different maturities, including the way that those correlations are influenced by the shape of the term structure, are better captured by multifactor models discussed in the next section.

5.7.2 Introduction to multifactor models

Multifactor models have the potential to explain the lack of perfect correlation. They have more than one source of randomness allowing for increased variety of yield curves. However, as a result of their increased flexibility, multifactor models require more computation time than the one-factor models introduced above.

There have been a number of attempts to extend one-factor short rate models so that they involve two or more factors. Brennan & Schwartz (1982), for example, chose the two stochastic factors to be the short rate and the long-term interest rate. Fong & Vasicek (1991) chose the variance of the short term rate to be the second state variable to more accurately describe the term structure of interest rates. Two-factor models can be further extended to three or more factors. The choice of the number of factors then involves a compromise between numerically-efficient implementation and capability of the model to represent realistic correlation patterns (and covariance structures in general) and to fit satisfactorily enough market data in most concrete situations (Brigo & Mercurio, 2001).

When modelling interest rate risk, insurance companies seem to commonly use models that involve more than one factor. Apparently models such as Black & Karasinski, CIR or Hull & White models with two or three factors are often used. Unfortunately, there seems to be no single answer to the question of how many factors should appear in the model. Research on term structure has established that much of the variablity in government bond returns can be summarized by movements in a few, usually two or three, underlying factors. In Section 5.8.3 we will discuss a common method used to extract the driving factors, namely principal component analysis (PCA).

5.8 Heath, Jarrow & Morton (HJM) model

5.8.1 Introduction to the HJM model

An important alternative to the short rate models is the Heath, Jarrow & Morton (HJM) model framework[16] which was proposed in a groundbreaking paper by Heath et al. (1992)[17]. Instead of modelling a short rate and deriving the forward rates, or equivalently the yield curve, from that model, the HJM model starts with directly modelling the whole forward rate curve. The advantage of modelling forward rates by the HJM model is that the current term structure of rates is, by construction, an input of the selected model.

The HJM model is based on the instantaneous forward rates $f(t,T)$, $0 \leq t < T$.[18] The HJM model simultaneously considers infinitely many processes $f(t,T)$, namely one process for each T. However, if a model depends on, say, three sources of randomness, then we are fortunately able to consider the problem as three dimensional rather than infinite dimensional. The n-factor HJM model assumes that the evolution of the instantaneous forward rates, under the real world measure, is governed by the stochastic differential equation

$$df(t,T) = \alpha(t,T)\ dt + \sum_{i=1}^{n} \sigma_i(t,T)\ dW_{t,i}, \quad t \leq T, \tag{5.6}$$

where $\alpha(t,T)$ is the drift and $\sigma_i(t,T)$ the volatility function of factor i at maturity T, and $W_{t,i}$ are independent standard Wiener processes. The volatilities and drifts can depend on the history of the Wiener processes $W_{t,i}$ and on the rates themselves up to time t. Equation 5.6 is the most general formulation of the HJM approach with n sources of randomness.

If there is only one source of randomness, that is $n = 1$, the forward rate for any fixed maturity T evolves according to its own volatility $\sigma(t,T)$ and its own drift $\alpha(t,T)$. Since the forward rate processes (for different maturities) in such a setting are dependent upon the same one-dimensional source of uncertainty W_t, the changes of all forward rates, and hence all yields and all bond prices, are perfectly but non-linearly correlated (cf. one-factor models). As already discussed above, for many applications this kind of an assumption is too coarse.

In a multifactor version of the HJM model (with n sources of randomness) the various processes are driven by a collection of independent Wiener process $W_{t,1}, W_{t,2}, \ldots, W_{t,n}$. The forward rate for any fixed maturity T then has a

[16] It is important to observe that the HJM approach to interest rates is not a proposal of a specific model, such as e.g. Vasicek model. It is instead a framework to be used for analyzing interest rate models. Every short rate model can be equivalently formulated in forward rate terms (Björk, 1998).

[17] The work of Heath, Jarrow and Morton was motivated by the earlier work of Ho & Lee (1986). The Ho-Lee model is a special case of the HJM model.

[18] Although the HJM model is often described as a model of forward evolution, it can be re-expressed so that the evolution of a spot rate curve, or indeed of a bond price curve, is fundamental. Empirical work, for instance, is often performed on spot rate curves, not on forward rate curve (James & Webber, 2000).

volatility $\sigma_i(t,T)$ for each Wiener process term $W_{t,i}$. This allows different bonds to depend on external 'shocks' in different ways, and to have strong correlation with some bonds and weaker correlation with others.

Given a non-random initially observed forward rate curve $f(0,t)$, Equation 5.6 can be integrated as

$$f(t,T) = f(0,T) + \int_0^t \alpha(s,T)\ ds + \sum_{i=1}^{n} \int_0^t \sigma_i(s,T)\ dW_{s,i}. \tag{5.7}$$

Equation 5.7 says that the forward rate process starts with initial value $f(0,T)$ and is driven by various Wiener process terms and a drift. The instantaneous short rate $r_t = f(t,t)$ can be written

$$r_t = f(0,t) + \int_0^t \alpha(s,t)\ ds + \sum_{i=1}^{n} \int_0^t \sigma_i(s,t)\ dW_{s,i}. \tag{5.8}$$

It is worth noting that the short rate process r_t is not necessarily Markov as the evolution of the term structure could depend on the entire path taken by the term structure since it was initialized at time 0. This property may lead to considerable increase in computation times when implementing the model since one may be forced to use Monte Carlo simulation or non-combining trees to value derivative securities (for the numerical techniques involved with the HJM model see e.g. Clewlow & Strickland, 1998, Section 10).

Suppose now that we have specified the coefficient functions $\alpha(t,T)$ and $\sigma_i(t,T)$, $i = 1,\ldots,n$, and the initial forward rate curve $f(0,T)$. Then we have specified the entire forward rate structure and thus, by the relation

$$P(t,T) = e^{-\int_t^T f(t,s)\ ds}, \tag{5.9}$$

we have also specified the entire term structure of zero-coupon bonds, $t \leq T$. Since in the dynamics (5.6) we have n sources of randomness and an infinite number of traded assets (one bond for each maturity T), we may have introduced arbitrage possibilities into the bond market. Contrary to the short rate modelling case, to rule out arbitrage we now cannot choose the drift rate $\alpha(t,T)$ independently of the volatility structure. We need to impose conditions on the forward-rate dynamics, so that they are consistent with absence of arbitrage opportunities. For simplicity, we will first discuss these conditions in one-factor case ($n = 1$). The conditions are then given in a more general case of n factors.

In one-factor case the stochastic differential equation for $P(t,T)$ under the real world probability measure ($\mathbb{P}$-measure) is

$$dP(t,T) = \mu_P(t,T)P(t,T)dt + \sigma_P(t,T)P(t,T)dW_t, \tag{5.10}$$

where

$$(5.11) \qquad \sigma_P(t,T) = -\int_t^T \sigma(t,s)\ ds \quad \text{and}$$

$$(5.12) \qquad \mu_P(t,T) = r_t - \int_t^T \alpha(t,s)\ ds + \frac{1}{2}\ \sigma_P(t,T)^2.$$

All bonds are now driven by the same Wiener process, so to rule out the arbitrage we now have to impose the restriction

$$(5.13) \qquad \mu_P(t,T) = r_t - \lambda(t)\sigma_P(t,T), \quad \text{for all } T,$$

where $\lambda(t)$ is the market price of risk at time t. Using (5.11) and (5.12), the arbitrage-free condition can be written as

$$(5.14) \qquad \int_t^T \alpha(t,s)\ ds = -\lambda(t)\int_t^T \sigma(t,s)\ ds + \frac{1}{2}\left(\int_t^T \sigma(t,s)\ ds\right)^2,$$

which, after differentiating with respect to T on both sides of the equation and rearranging the terms, becomes

$$(5.15) \qquad \alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,s)\ ds - \lambda(t)\sigma(t,T).$$

It follows from (5.15), that the SDE for instantaneous forward rates in one-factor case (under the measure $\mathbb{P}$) has the expression

$$(5.16) \qquad df(t,T) = \left[\sigma(t,T)\int_t^T \sigma(t,s)\ ds - \lambda(t)\sigma(t,T)\right] dt + \sigma(t,T)dW_t.$$

For the n-factor case, (5.15) is written as

$$(5.17) \qquad \alpha(t,s) = \sum_{i=1}^{n} \sigma_i(t,T)\int_t^T \sigma_i(t,s)\ ds - \sum_{i=1}^{n} \lambda_i(t)\sigma_i(t,T),$$

where $\lambda_i(t)$ is the market prices of risk associated with $W_{t,i}$. Equation 5.17 is called the "HJM Rate Drift Condition". It tells us that to rule out the arbitrage, we cannot choose the forward rate drift $\alpha(t,T)$ independently of the volatility structure, but the drift must be a function of the volatility structure and the market price of risk, $\lambda(t)$.

However, as discussed earlier, in order to price fixed-income derivatives, we need the distribution of $f(t,T)$ and hence bond prices under the risk-neutral measure ($\mathbb{Q}$-measure). So, what we are still missing is the model formulation under the measure $\mathbb{Q}$. What we also would like to see, is that the market price of risk $\lambda(t)$ in the drift term of (5.17) would disappear. Fortunately, for HJM models setting the price of risk $\lambda(t)$ to zero takes us from the drift under the real world measure $\mathbb{P}$ to the drift under $\mathbb{Q}$. Thus, the "HJM Drift Condition" under the risk-neutral world (under $\mathbb{Q}$) has the expression

$$\alpha(t,s) = \sum_{i=1}^{n} \sigma_i(t,T) \int_t^T \sigma_i(t,s)\, ds. \tag{5.18}$$

As a consequence, as stated by James & Webber (2000), in the HJM framework one does not have to separately model a price of risk (unless one is calibrating to time series data). It is now also easy to see that the forward rate is in fact completely specified by the volatility functions. In other words, no drift estimation is needed.

An "algorithm" for the use of an HJM model can be written schematically as follows (see Björk, 1998, p. 269):

1. Specify, by your own choice, the volatilities $\sigma_i(t,T)$, $i = 1,\ldots,n$ (see Section 5.8.3).
2. The drift parameter of the forward rates is now given by (5.18).
3. Go to the market and observe today's forward rate structure $f(0,T)$, $T \geq 0$ (see Section 5.8.2).
4. Integrate in order to get the forward rates as in (5.7).
5. Compute bond prices using (5.9).
6. Compute prices for derivatives.

5.8.2 Initial curves and the availability of market data

The HJM model requires as an input any sufficiently smooth initial yield curve. However, fitting HJM models empirically requires estimating a continuous forward rate curve from the discrete set of bond prices observed in the market. The main problem occuring is that only rates with few maturities can be derived directly from the market[19]. This is typically quite enough for short terms but only with one-year lag for maturities longer than one or two years. The problem is usually solved by interpolating between the given points with splines or other parameterized families (see e.g. Lai & Xing, 2008, Chapters 7 and 10.2). However, some analyses show that the choice of the interpolation method, and thereby the choice of the initial curve, should not be completely independent of the particular model used. Instead, one should use an initial curve that is consistent with the model (see e.g. Angelini & Herzel, 2002).

[19]A common procedure for calculating the zero-coupon yield curve from market data is called a bootstrap method. Bootstrapping starts with the shortest term security and steps through them all in ascending order of maturity. At every step, zero rates from the preceding securities are used to determine the zero rates for the current one. The curve generation process strips each security into its individual cashflows and then prices it using zero coupon pricing. See e.g. Hull (1999).

The lack of appropriate and liquid securities of different maturities is not only a problem for the estimation of the initial curve. Market incompleteness, i.e. there are no deep and liquid markets of appropriate securities, is a serious practical problem especially for the long term hedging of life and pension products. The CRO Forum in Appendix A of their paper Market value of liabilities for insurance firms (see `www.croforum.org/publications.ecp`) use a 60-year cash flow or interest rate option and a 30-year equity option as examples of non-hedgeable risks. Other examples of this type include mortality and lapse (policyholder behaviour) risk. On the other hand many life insurance financial modelling practitioners (see e.g. Life and Pension, April 2007, pages 15-18) hold the view that at the moment about 10 years seems to be the limit after which illiquidity causes the projections based on market-calibrated data being more of a guesswork (extrapolation). Another example is given in Fabozzi (2002, p. 24) where it is noted for the US market that models with good statistical fit is needed for the valuation of caps and floors beyond 5-year tenor. In the Euro area the longest benchmarks of government bonds are the 30-year German bond and the 50-year French bond. In swap markets one may find even longer maturities but it is still questionable how liquid and deep the market is, and if it can be used as a good proxy for risk-free interest rate[20].

5.8.3 Volatility functions in HJM

The HJM models require only the specification of the form of the volatility structure of forward interest rates along with the initial term structure of interest rates as inputs. The choices of the volatility functions $\sigma_i(t,T)$, $i = 1,\dots,n$, are essential for the valuation of derivatives. For example, volatility functions that give Gaussian forward rate processes may lead to explicit formulae for simpler options. In practice it is convenient to choose the volatility structure to be Markov, since such specifications avoid unnecessary complexity and are likely to result in valuations using approximate trees. Non-Markov specifications may require difficult simulations or non-recombining trees even for the valuation of simple options.

One of the volatility functions often used in practice is $\sigma(t,T) = \sigma$, that is a constant. This is a Ho & Lee type volatility, which is tractable but unrealistic. Another common function is that of a Vasicek type volatility, namely $\sigma(t,T) = \sigma e^{-a(T-t)}$, with σ and a constants. This specification is consistent with the Hull & White model or the Vasick model with time-varying drift. Although this volatility performs better than a constant volatility, it does not provide a very realistic volatility structure. In practical implementations one often uses several volatility functions. For instance, a two-factor model where the first function is given by the Ho & Lee structure and the second by the Hull & White structure was introduced by Heath et al. (1992).

PCA provides a recommended way of modelling the volatility structure of interest rates in the HJM framework. By applying PCA to a time series of

[20]In QIS 3 and QIS 4 field-tests for Solvency II the risk-free interest rates up to 50 years were obtained from the swap market data. However, the swap rates do include a risk premium.

historical term structure, it is possible to not only determine the number of factors that explain the data in a satisfactory way, but also to identify the shape of the volatility functions. PCA is a multivariate prodedure whose main use is to reduce the dimensionality of a data set while retaining as much information as is possible. PCA transforms two or more correlated variables into a smaller number of uncorrelated variables ordered by reducing variability. These new variables are called principal components (sometimes also referred to as factors). Essentially, principal components are linear combinations of the original (correlated) variables. The first principal component accounts for as much of the variability in the data as possible, the second principal component accounts for as much of the remaining variability as possible, and so on. The last of these variables can be removed with minimum loss of real data. For a more detailed and theoretical description of PCA see e.g. Johnson & Wichern (1998) and Lai & Xing (2008).

Empirical studies based on PCA reveal that three factors (principal components) capture about 90%–95% of variations in the yield curve (see e.g. Litterman & Scheinkman, 1991; Jamshidian & Zhu, 1997). The three components, or factors, are often called ”level”, ”slope” and ”curvature” (Litterman & Scheinkman, 1991). The names describe how the yield curve shifts or changes shape in response to a shock. A ”level” shock changes the interest rates of all maturities by almost identical amounts, inducing a parallel shift that changes the level of the whole yield curve. The influence of the ”slope” factor on yield curve increases short-term interest rates by much larger amounts than the long-term interest rates, so that the yield curve becomes less steep and its slope decreases. The ”curvature” factor affects medium-term interest rates, and consequently the yield curve becomes more ”humpshaped” than before. According to James & Webber (2000), the ”level” factor will often explain 80%–90% or more of the variance, depending on the data set, confirming that parallel shifts are the most usual term structure movements. As a consequence, as stated by Cheyette in Fabozzi (2002, Chapter 1), valuation of securities can be reduced to a one-factor problem in many instances with little loss of accuracy. A two- or three-dimensional process is, however, needed to provide a realistic evolution of the yield curve. [21] [22]

The use of PCA to estimate HJM volatility functions was already proposed in the original paper of Heath, Jarrow and Morton. Later literature applying PCA to determine the volatility functions specifically in the context of an HJM model contain e.g. the paper of Bühler et al. (1999) and Driessen et al. (2003). As discussed above, two or three factors (principal components) are usually needed to provide a realistic evolution of the yield curve. Once the driving principal components are extracted from the time series of historical term structure, the volatility functions are determined by the volatilities of

[21] Note that much of the empirical studies has been done under the real world measure. However, since instantaneous-covariance structure of the same process when moving from the real world probability measure to the risk neutral probability measure does not change, one may guess that also in the risk-neutral world a two- or three-dimensional process may be needed in order to obtain satisfactory results (Brigo & Mercurio, 2001).

[22] One must be aware that these conclusions on the number of components are in reference to the historical term structure data only.

those principal components and the corresponding factor loadings. We will demonstrate the volatility function estimation using PCA in Appendix F. Appendix F will also contain a cursory discussion of how to price derivatives when the forward rate process follows a HJM model.

5.9 On market models

Since the market models are of increasing importance in interest rate modelling, we briefly discuss the framework here. For more detailed introduction on the topic we refer to e.g. Brigo & Mercurio (2001).

A practically appealing alternative to the HJM framework, called market models, was proposed by Brace et al. (1997), Jamshidian (1997) and Miltersen et al. (1997). Market models aimed at correcting some problems encountered by the HJM model, such as, that (1) the HJM model is based on the instantaneous forward rates, which cannot be directly observed in the market, and that (2) the model is not always easy to calibrate to prices of actively traded instruments such as caps (for calibrating interest rate models in the financial industry see e.g. Lai & Xing, 2008, Chapter 10.6).

Market models shifted from a concentration on unobservable, instantaneous rates of interest such as the short rate or forward rates, to rates (e.g. LIBOR[23]) which are directly observable in the market. Accompanied with the assumption that relevant market interest rates are lognormal, which results in analytical formulae for some commonly traded derivatives, market models create an environment which makes calibration of a model relatively straighforward compared with models arising from alternative frameworks. The analytical tractability has made the market models popular amongst practitioners. The HJM model, on the other hand, has become popular amongst the academics due to its attractive theoretical properties. The most widely used version of the market models is the LIBOR market model (LMM), which is expressed in terms of successive LIBOR forward rates. For the calibration of the LMM and the model plausibility for pricing swaption based products see e.g. Salminen (2007).

[23]LIBOR = London Inter Bank Offer Rate

References

Alexander, C. (2001), *Market models: A Guide to financial data analysis*, John Wiley & Sons Inc.

Angelini, F. & Herzel, S. (2002), ”Consistent initial curves for interest rate models”, *Journal of Derivatives*, 9, 8–17.

Bera, A. K. & Jarque, C. M. (1980), ”Efficient tests for normality, homoscedasticity and serial independence of regression residuals”, *Economics Letters*, 6, 255–259.

Björk, t. (1998), *Arbitrage theory in continuous time*, Oxford University Press.

Black, F. & Karasinski, P. (1991), ”Bond and option pricing when short rates are lognormal”, *Financial Analysts Journal*, 47, 52–59.

Black, F. & Scholes, M. (1973), ”The pricing of options and corporate liabilities”, *Journal of Political Economy*, 81, 637–654.

Bollerslev, T. (1986), ”Generalized autoregressive conditional heteroscedasticity”, *Journal of Econometrics*, 31, 307–327.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994), *Time series analysis: forecasting and control* (3rd ed.), Prentice-Hall.

Brace, A., Gatarek, D. & Musiela, M. (1997), ”The market model of interest rate dynamics”, *Mathematical Finance*, 7, 127–155.

Brennan, M. & Schwartz, E. S. (1982), ”An equilibrium model of bond pricing and a test of market efficiency”, *Journal of Financial and Quantitative Analysis*, 17, 301–329.

Brigo, D. & Mercurio, F. (2001), *Interest rate models theory and practice*, Springer-Verlag.

Brockwell, P. J. & Davis, R. A. (2002), *Introduction to time series and forecasting* (2nd ed.), New York: Springer.

Bühler, W., Uhrig-Homburg, M., Walter, U. & Weber, T. (1999), ”An empirical comparison of forward-rate and spot-rate models for valuing interest-rate options”, *The Journal of Finance*, 54, 269–305.

Cairns, A. (2004), *Interest rate models: an introduction*, Princeton, New Jersey: Princeton University Press.

Carlin, B. P., Polson, N. G. & Stoffer, D. S. (1992), ”A Monte Carlo approach to nonnormal and nonlinear state-space modeling”, *Journal of the American Statistical Association*, 87, 493–500.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. & Sanders, A. B. (1992), ”An empirical comparison of alternative models of the short-term interest rate”, *The Journal of Finance*, 47, 1209–1227.

Chatfield, C. (1995), *Problem solving, a statistician's guide* (2nd ed.), Florida: Chapman & Hall/CRC.

Clewlow, L. & Strickland, C. (1998), *Implementing derivative models*, John Wiley & Sons Ltd.

Cochrane, J. H. (2001), *Asset pricing*, Princetone University Press.

Derman, E. (1996), ”Model risk”, Quantitative Strategies Research Notes, April 1996, Goldman Sachs.

Down, K. (2002), *Measuring market risk*, John Wiley & Sons.

Driessen, J., Klaassen, P. & Melenberg, B. (2003), ”The performance of multifactor term structure models for pricing and hedging caps and swaptions”, *The Journal of Financial and Quantitative Analysis*, 38, 635–672.

Engel, R. F. (1982), ”Auroregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations”, *Econometrica*, 50, 987–1007.

Fabozzi, F. J. (ed.) (2002), *Interest rate, term structure, and valuation modelling*, New Jersey, USA: John Wiley & Sons.

Fong, H. G. & Vasicek, O. A. (1991), ”Fixed-income volatility management”, *Journal of Portfolio Management*, 17, 41–46.

Francq, C., Roy, R. & Zakoian, J.-M. (2005), ”Diagnostic checking in ARMA models with uncorrelated errors”, *Journal of the American Statistical Association*, 100, 532–544.

Franke, J., Härdle, W. & Hafner, C. M. (2004), *Statistics of financial markets*, Germany: Springer-Verlag.

Franses, P. H. & Dijk, D. (2000), *Non-linear time series models in empirical finance*, Cambridge, UK: Cambridge University Press.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. & Smith, A. F. M. (1990), ”Illustration of bayesian inference in normal data models using Gibbs sampling”, *Journal of the American Statistical Association*, 85, 972–985.

Gelman, A., Carlin, J. B., Stern, H. S. & Rubin, D. B. (2004), *Bayesian data analysis* (2nd ed.), Boca Raton, Florida: Chapman & HallCRC.

Geman, S. & Geman, D. (1984), ”Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gourieroux, C. & Jasiak, J. (2001), *Financial econometrics, problems, models and methods*, New Jersey: Princeton University Press.

Hamilton, J. (1989), ”A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica*, 57, 357–384.

Hardy, M. (2003), *Investment guarantees, modelling and risk management for equity-linked life insurance*, Hoboken, New Jersey: John Wiley & Sons.

Heath, D., Jarrow, R. & Morton, A. (1992), ”Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation”, *Econometrica*, 60, 77–105.

Ho, T. & Lee, S. (1986), ”Term structure movements and pricing interest rate contingent claims”, *Journal of Finance*, 41, 1011–1029.

Holtan, J. (2007), ”Pragmatic insurance option pricing”, *Scandinavian Actuarial Journal*, 1, 53–70.

Hull, J. C. (1999), *Options, futures & other derivatives* (4th ed.), Prentice Hall.

James, J. & Webber, N. (2000), *Interest rate modelling*, Chichester, England: John Wiley & Sons.

Jamshidian, F. (1997), ”LIBOR and swap market models and measures”, *Finance and Stochastics*, 1, 293–330.

Jamshidian, F. & Zhu, Y. (1997), ”Scenario simulation: theory and methodology”, *Finance and Stochastics*, 1, 43–67.

Jarrow, R. A. (2002), *Modelling fixed-income securities and interest rate derivatives* (2nd ed.), Stanford, USA: Stanford University Press.

Johnson, R. & Wichern, D. W. (1998), *Applied multivariate statistical analysis*, Prentice Hall.

Jones, C. (1998), ”Bayesian estimation of continuous-time finance models”, Working paper, Simon School of Business, University of Rochester, New York.

Kaliva, K. & Koskinen, L. (2008a), ”The long-term risk caused by stock market bubble”, to appear in *Journal of Risk*, 9, 1–13.

——— (2008b), ”Stock market bubbles, inflation and investment risk”, *International Review of Financial Analysis*, 17, 592–603.

Kaliva, K., Koskinen, L. & Ronkainen, V. (2007), ”Internal models and arbitrage-free calibration”, in *AFIR Colloquium 2007*, Stockholm.

Kennedy, P. (2003), *A guide to econometrics* (5th ed.), Cambridge, Massachusetts: The MIT Press.

Korhonen, P. & Koskinen, L. (2008), ”Searching guidelines for the use of internal models in insurance company’s risk and capital management”, Submitted (under revision).

Kwon, H.-C. (1992), ”The time variant term premium of interest rates”, Ph.D. thesis, Ohio State University Economics Department.

Lai, L. L. & Xing, H. (2008), *Statistical models and methods for financial markets*, Springer.

Litterman, R. & Scheinkman, J. (1991), ”Common factors affecting bond returns”, *The Journal of Fixed Income*, 1, 54–61.

Longstaff, F. A. & Schwartz, E. S. (2001), ”Valuing American options by simulation: a simple least-squares approach”, *Review of Financial Studies*, 14, 113–148.

Luenberger, D. G. (1998), *Investment science*, Oxford University Press.

Luoma, A., Puustelli, A. & Koskinen, L. (2008), ”Bayesian analysis of participating life insurance contracts with American-style options”, Submitted. (A shorter version in *AFIR Colloquium 2008*, Rome).

Makridakis, S., Wheelwright, S. C. & Hyndman, R. J. (1998), *Forecasting: methods and applications*, John Wiley & Sons, Inc.

McCulloch, J. H. (1975), ”The tax-adjusted yield curve”, *Journal of Finance*, pp. 811–830.

——— (1990), ”U.S. term structure data, 1946-87”, in *Handbook of Monetary Economics* (Vol. 1), eds. B. M. Friedman & F. H. Hahn, Elsevier Science, pp. 672–715.

McLeod, A. I. & Li, W. K. (1983), ”Diagnostic checking ARMA time series models using squared-residual autocorrelations”, *Journal of Time Series Analysis*, 4, 269–273.

McQuarrie, A. D. R. & Tsai, C. (1998), *Regression and time series model selection*, Singapore: World Scientific.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), ”Equations of state calculations by fast computing machines”, *Journal of Chemical Physics*, 21, 1087–1092.

Miltersen, K., Sandmann, K. & Sondermann, D. (1997), ”Closed form solutions for term structure derivatives with log-normal interest rates”, *Journal of Finance*, 52, 409–430.

Morningstar (2007), *Stocks, bonds, bills, and inflation 2007 yearbook*, Chicago, US: Morningstar, Inc.

Pentikäinen, T., Bonsdorff, H., Pesonen, M., Rantala, J. & Ruhonen, M. (1989), *Insurance solvency and financial strength*, Helsinki, Finland: Finnish Insurance Training and Publishing Company Ltd.

Puustelli, A., Koskinen, L. & Luoma, A. (2008), "Bayesian modelling of financial guarantee insurance", *Insurance: Mathematics and Economics*, 43, 245–254.

Romano, J. L. & Thombs, L. A. (1996), "Inference for autocorrelations under weak assumptions", *Journal of the American Statistical Association*, 91, 590–600.

Ronkainen, V., Koskinen, L. & Koskela, L. (2008), "Challenges in developing internal models for Solvency II", *Scandinavian Insurance Quarterly*, pp. 154–158.

Royston, P. (1980), "An extension of Shapiro and Wilk's W test for normality to large samples", *Applied Statistics*, 31, 115–124.

Salminen, T. (2007), "Calibration of a forward rate market model", Master's thesis, Helsinki University of Technology, Department of Engineering Physics and Mathematics, Finland.

Shumway, H. S. & Stoffer, D. D. (2006), *Time series analysis and its applications: with R examples* (2nd ed.), Springer.

Spiegelhalter, D., Best, N., Carlin, B. & van der Linde, A. (2002), "Bayesian measures of model complexity and fit", *Journal of the Royal Statistical Society B*, 64, 583–639.

Tsay, R. S. (2005), *Analysis of financial time series* (2nd ed.), Hoboken, New Jersey: John Wiley & Sons.

Wilkie, A. D. (1986), "A stochastic investment model for actuarial use", *Transactions of the Faculty of Actuaries*, 39, 341–403.

——— (1995), "More on a stochastic investment model for actuarial use", *British Actuarial Journal*, 1, 774–964.

Wilmott, P. (2001), *Paul Wilmott introduces quantitative finance*, Chichester, England: John Wiley & Sons.

Wüthrich, M. V., Bühlmann, H. & Furrer, H. (2008), *Market-consistent actuarial valuation* (EAA Lecture Notes), Springer Verlag.

Zeytun, S. & Gupta, A. (2007), "A comparative study of the Vasicek and the CIR model of the short rate", Fraunhofer ITWM, Bericht 124 (2007), Kaiserslautern, Germany, 26 pages. Available from Internet: `http://www.itwm.fraunhofer.de/zentral/download/berichte/bericht124.pdf`.

Appendix A

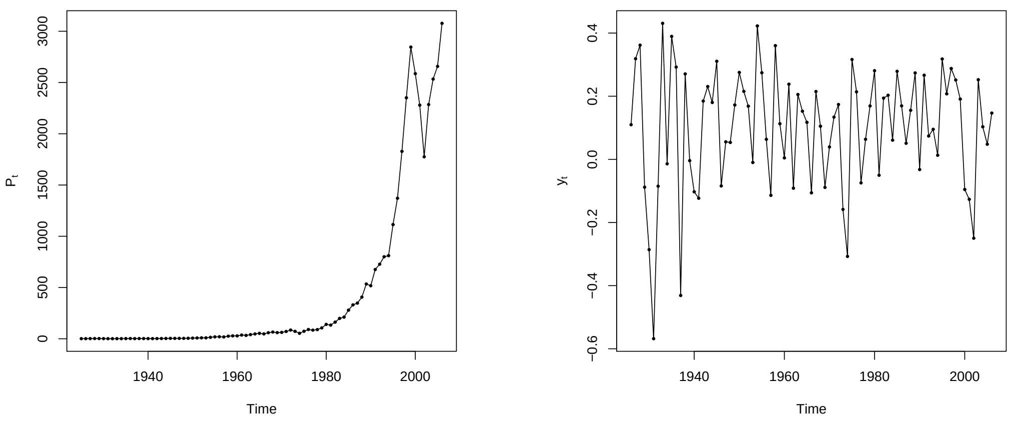
Descriptive analysis of the equity index data

We here describe our data series in order to obtain insight into their patterns. The main purpose of this initial data analysis is to describe some characteristic features of the time series at hand.

A.1 Data transformations

To justify and illustrate the transformation applied to our equity index series, we will first take a closer look at one of the original data series, namely S&P 500 yearly Total Return Index (SP500TRIy) from years 1925-2006 (Morningstar, 2007). Figure A.1(a) displays the index series. It can be easily seen that the series is nonstationary and can be characterized by the irregular exponential growth and heteroscedasticity (i.e. variance increases as the level of the original series rised over time).

In financial studies, returns are very often investigated instead of the inital series of prices or stock market index values. This is because returns in general display more regular patterns, i.e. have more attractive statistical properties, and are hence easier to handle. There are, however, several definitions of an asset return (see e.g. Tsay, 2005, pp. 2–6). Academics often use in their research logarithmic returns (log returns), also called continuously compounded



(a) Index values P_t. (b) Log returns y_t.

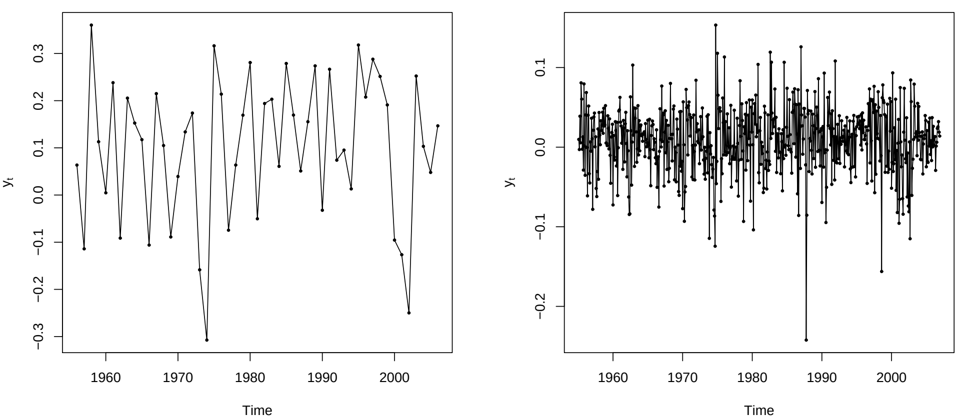
Figure A.1: S&P 500 yearly Total Return Index from 1925-2006.

returns, since their statistical properties are more tractable. The log return y_t at time t is defined as

$$y_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1}, \tag{A.1}$$

where P_t is the equity price or stock market index value at time t. Note that the purpose of the transformation in (A.1) is to make the data series stationary. The log return approximately represents the relative (percentage) change in equity price from period to period, i.e. growth rate. The approximation is almost exact if the percentage change is small[1]. The series y_t, displayed in Figure A.1(b), shows more or less stationary and random appearance. The stationarity of the time series can be tested e.g. by augmented Dickey-Fuller (ADF), Phillips-Perron (PP) or Kwiatkowski-Phillip-Schmidt-Shin (KPSS) test (see unit-root tests e.g. in Franses & Dijk, 2000).

The transformation in (A.1) is applied to all our index series. Two other series considered in this report are S&P 500 yearly and monthly Total Return Index series covering the years 1955-2006 (Morningstar, 2007). These data series are denoted by SP500TRIs and SP500TRIm, respectively. The former time series corresponds to the last 52 observations of SP500TRIy. Figures A.2(a) and A.2(b) display the log returns (y_t) of the corresponding data series. In the sequel, as we refer to any of our time series, we will actually refer to the log returns y_t instead of the index values P_t.



(a) Log returns of SP500TRIs. (b) Log returns of SP500TRIm.

Figure A.2: S&P 500 yearly (SP500TRIs) and monthly (SP500TRIm) Total Return Index series from 1955-2006.

[1]According to Franke et al. (2004), the log return approximates the relative change well with returns under 10%. The authors also state that this is usually above all the case when one is studying financial time series with high frequency, as, for example, with daily values.

A.2 Typical features of financial time series

Many models that are commonly used in empirical finance to describe asset returns and volatility are linear. There are, however, several indications that nonlinear models may be more appropriate for time series of asset returns. The typical features suggesting the necessity of considering nonlinear models are (Franses & Dijk, 2000):

1. Large returns (in absolute terms) occur more frequently than one might expect under the assumption that the data are normally distributed.[2]

2. Large absolute returns tend to appear in clusters.[3]

3. Large negative returns appear more often than large positive ones in stock markets, while it may be the opposite for exchange rates.

4. Volatile periods are often preceded by large negative returns.

A.3 Data-analysis

In this section we will take a closer look at our data series to discuss the above features. The analysis follows closely that of Franses & Dijk (2000) and starts by introducing a set of summary statistics. After that we report the values of these summary statistics as they are computed from our datasets. We also use some graphical tools and statistical tests to analyze the data.

A.3.1 Summary statistics

The set of summary statistics include the number of observations (N) in the data series, the arithmetic mean and median (measures of location), the minimum, maximum and standard deviation (measures of statistical dispersion) and the skewness and kurtosis (measures of the shape of the distribution).

Skewness (Skew) and kurtosis (Kurt) characterize the location and variability of a data set. Skewness is a measure of symmetry of the distribution and kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Data sets with negative values for the skewness indicate data that are skewed left, while positive values indicate data that are skewed right. By skewed left, we mean that the left tail is long relative to the right tail. Similarly, skewed right means that the right tail is long relative to the left tail. Data sets with high kurtosis tend to have a distinct and sharper peak near the mean, decline rather rapidly and have heavy tails, while data sets with low kurtosis, on the other hand, tend to have a flat and more rounded top near the mean and wider "shoulders". The kurtosis (K_y) and skewness (S_y) are defined as

$$K_y = \mathrm{E}\left[\frac{(y_t - \mu)^4}{\sigma^4}\right] \quad \text{and} \quad S_y = \mathrm{E}\left[\frac{(y_t - \mu)^3}{\sigma^3}\right],$$

[2]Normality is often closely related to the use of linear models.

[3]Clusters indicate the possible presence of time-varying risk or volatility.

where μ and σ^2 are the mean and the variance of y_t, respectively. For an observed time series $y_1, \ldots, y_n$ these statistics can be estimated by the sample analogues

$$\hat{K}_y = \frac{1}{n}\sum_{t=1}^{n}\frac{(y_t-\hat{\mu})^4}{\hat{\sigma}^4} \quad \text{and} \quad \hat{S}_y = \frac{1}{n}\sum_{t=1}^{n}\frac{(y_t-\hat{\mu})^3}{\hat{\sigma}^3},$$

where $\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} y_t$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{\mu})^2$ are the sample mean and variance, respectively.

A usual assumption in the finance literature is that the logarithmic returns y_t are normally distributed as

$$y_t \sim N(\mu, \sigma^2).$$

For normal distribution the kurtosis is equal to three and the skewness is equal to zero (the latter is true for all other symmetric distributions as well). If the kurtosis is greater than three, the distribution is said to be leptokurtic. This kind of distribution is said to have heavy tails, implying that the distribution puts more mass on the tails than a normal distribution does. In other words, a random sample from a leptokurtic distribution tends to contain more extreme values. If the kurtosis is less than three, the distribution is said to be platykurtic.

A.3.2 Data description

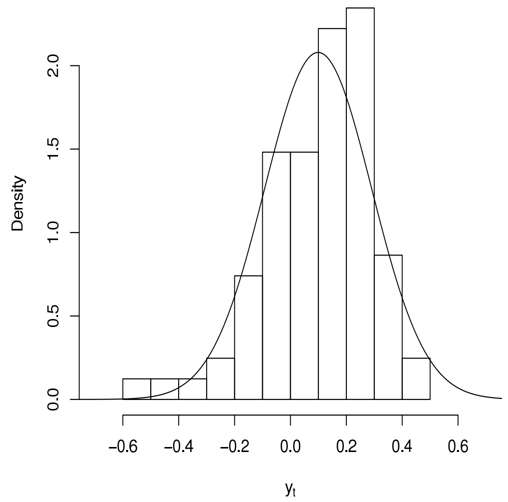
We will now examine if and how strongly our data sets reflect the features listed above. We will start by reporting the summary statistics computed for log returns y_t in Table A.1. The table will serve as a basis for the discussion in this section.

Table A.1: Summary statistics of the data series.

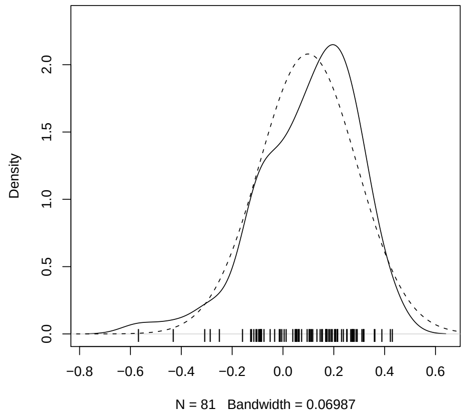
Data series	N	Mean	Median	Min	Max	Sd	Skew	Kurt
SP500TRIy	81	0.099	0.134	-0.568	0.431	0.192	-0.853	3.893
SP500TRIs	51	0.100	0.117	-0.307	0.360	0.154	-0.573	2.621
SP500TRIm	623	0.0086	0.0110	-0.242	0.153	0.0416	-0.599	5.545
		(0.103)	(0.132)	.	.	(0.144)	.	.

The data series are plotted in Figures A.1(b), A.2(a) and A.2(b). It can be seen that all three paths are characterized by subsequent ups and downs. In addition, both positive and negative outlying observations occur during the time periods, although extreme negative returns seem to be more present than positive extremes. The feature of volatility clustering is more difficult to observe in the data series. Some volatilty clusters may, however, be distinquished in the higher-frequency monthly series in Figure A.2(b).

The summary statistics of all our index series are shown in Table A.1. The table shows that the log returns in the data series SP500TRIy vary between -0.568 and 0.431. However, the mean return is 0.099 which implies that on average the yearly growth rate of the index is about 9.9% during the sample



(a) Histogram and normal distribution.



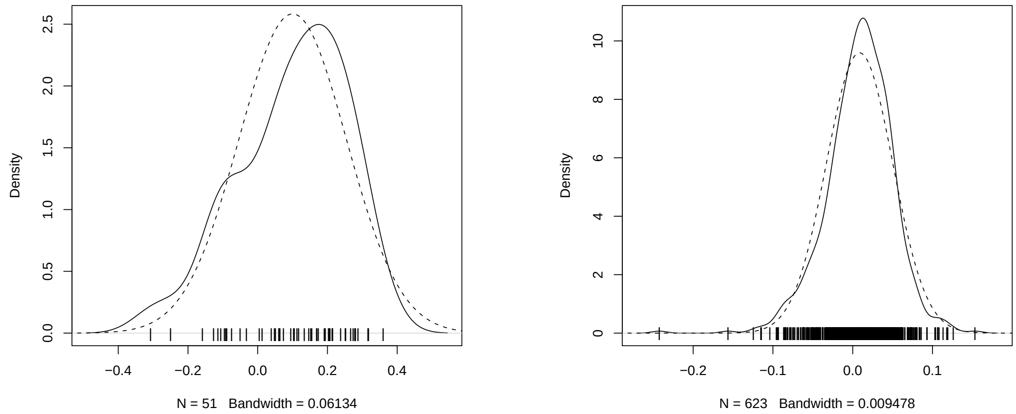
(b) Kernel estimate (solid) and normal distribution (dashed).

Figure A.3: The sample distribution of y_t of SP500TRIy.

period. The standard deviation (volatility) of the corresponding data set is 0.192, that is about 19%. The mean log return of the shorter yearly series SP500TRIs indicates a slightly higher growth rate, namely 10.0%. According to the monthly data (SP500TRIm) the yearly growth rate is 10.3%. The yearly mean, median and standard deviation of the monthly returns are given in parentheses on the last row of the table.

We then take a look at the distributional patterns of the data series. Figures A.3(a) and A.3(b) graphically illustrate the sample distribution of log returns of SP500TRIy. In Figure A.3(a) the sample distribution of y_t is described by a histogram while in Figure A.3(b) the corresponding distribution is estimated by a kernel density estimator. For comparison we also show the normal distribution which has the distribution parameters equal to the sample mean and variance of the data. The whiskers on the horizontal axis in Figure A.3(b) represent individual observations. Figures A.3(a) and A.3(b) reflect the fact that the distribution of log returns of SP500TRIy is more peaked and has fatter tail probabilities than the normal distribution, i.e. large absolute returns occur more often than would be expected if the data were normally distributed. The value of the kurtosis ($K > 3$) in Table A.1 also supports the conclusion. The skewness in Table A.1 is negative for SP500TRIy indicating non-symmetry and long left tail relative to the right. Hence, large negative returns appear more often than large positive ones. The yearly log returns of SP500TRIy well reflect the distributional patterns commonly associated with high-frequency financial time series.

The kernel density based estimates of the sample distribution of y_t of SP500TRIs and SP500TRIm are displayed in Figures A.4(a) and A.4(b), respectively. The kurtosis and skewness values can be found in Table A.1. The conclusions of the distributional pattern of monthly returns are similar to those of SP500TRIy, i.e. the distribution of log returns is skewed left and has fat tail probabilities. Also the distribution of yearly log returns SP500TRIs is skewed left. However, unlike the yearly log returns of the longer series



(a) SP500TRIs.

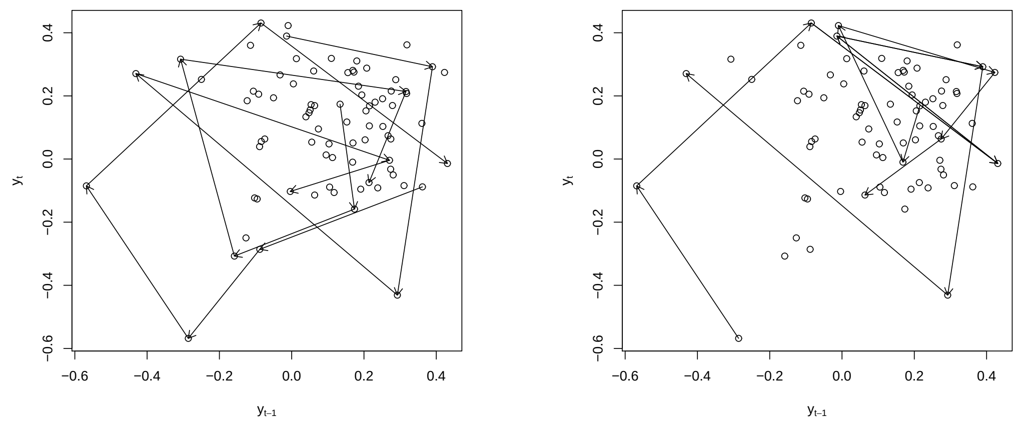
(b) SP500TRIm.

Figure A.4: Kernel estimate (solid) of the sample distribution of y_t of data series SP500TRIs and SP500TRIm and the normal distribution (dashed).

SP500TRIy, the returns of SP500TRIs do not have a fat-tailed sample distribution. This is indicated in Table A.1 by the kurtosis which is slightly less than three, i.e. less than that of the normal distribution.

The analysis of the behaviour of the volatile periods will be based on the scatterplots or lagplots (see Franses & Dijk, 2000). The scatterplots of SP500TRIy are displayed in Figures A.5(a) and A.5(b) below. In Figure A.5(a) the SP500TRIy values of year t (y_t) are plotted against the values of year $t-1$ (y_{t-1}) and the observations for the three smallest negative values of y_t are connected with the two preceding and the two following observations by arrows that point in the direction in which the data series evolves. The corresponding plot for the three largest values of y_t is shown in Figure A.5(b).

It commonly appears for the higher-frequency financial series that large absolute returns occur in clusters. In such a case relative volatile periods (i.e.



(a) Three smallest observations.

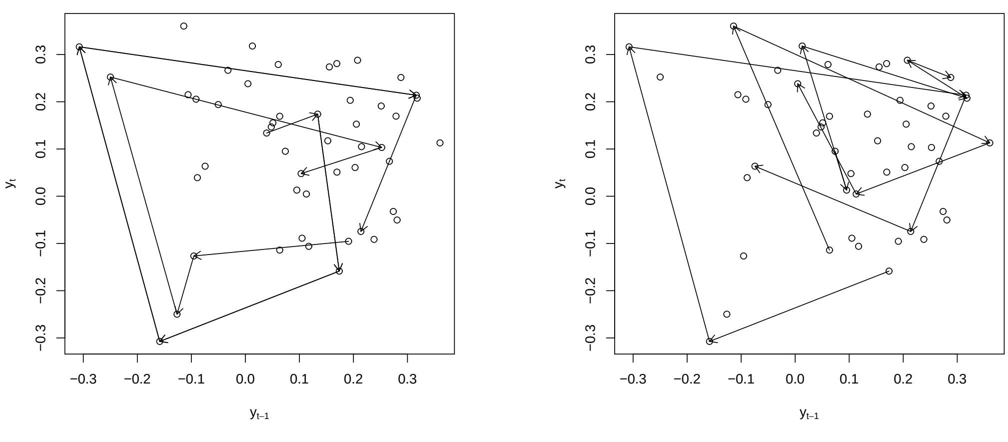
(b) Three largest observations.

Figure A.5: Scatterplot of the return of SP500TRIy on year t (y_t) against the return on year $t-1$ (y_{t-1}).

periods with large price changes and hence large absolute returns) alternate with those in which prices are rather stable and absolute returns remain small. In scatterplots this feature is reflected by the routes formed by the arrows showing tendency of travelling around the main cloud of observations for an extended period of time. A single large return would, on the other hand, constitute a three-cycle in the figure. This can be justified by noting that a large value y_t (in absolute terms) is needed to leave the main cloud of observations and that the next observation necessarily also appears outside the main cloud, since then y_{t-1} is large. See Franses & Dijk (2000) for more discussion and illustration.

In Figures A.5(a) and A.5(b) the routes formed by the arrows show some tendency of travelling around the main cloud of observations for an extended period of time. If the routes in Figures A.5(a) and A.5(b) were combined, the arrows would only comprise three stretches. As noted above, this indicates that large absolute returns occur in clusters. The first stretch starts at $(y_t, y_{t-1}) =$ (0.36,-0.09) which correspond to years 1928 and 1929, and ends at (-0.003,-0.10) where the latter coordinate corresponds to year 1940. The second stretch starts at (0.22,0.17) and ends at (0.06,-0.11). The first coordinate of the starting point and the latter coordinate of the ending point correspond to years 1951 and 1957, respectively. Finally, the third stretch covers the period from year 1971 to 1977. This closer examination of the arrow paths reveals some more volatile periods in our time series, one of which being the Great Depression of the 30's.

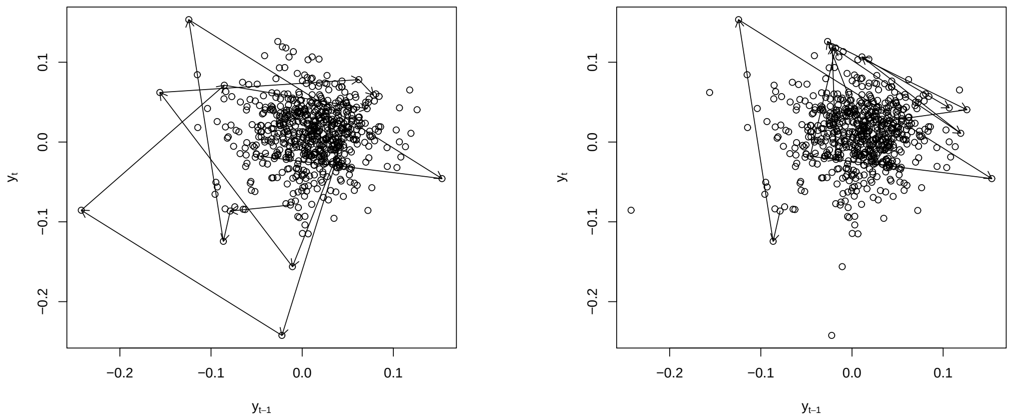
The scatterplots of the shorter yearly series SP500TRIs are displayed in Figures A.6(a) and A.6(b), respectively. At the first glance, the stretches may seem to travel around the main cloud of observations. However, it should be noted that, after all, the observations form a relatively compact cloud, so the clustering effect is not as obvious as in the case of the longer yearly data. If the figures are combined, the arrow paths form four stretches with the longest



(a) Three smallest observations. (b) Three largest observations.

Figure A.6: Scatterplot of the return of SP500TRIs on year t (y_t) against the return on year $t-1$ (y_{t-1}).

one covering the years from 1970 to 1978. The second path covers the years from 1956 to 1961 and corresponds to only one outlying observation. The last two paths correspond to the years both sides of the millennium (1992-1998 and 1999-2005). Looking at Figure A.1(b), some more extreme events are observed at those years. However, compared to the years in the 30's, for example, the overall volatility remains smaller.



(a) Three smallest observations.

(b) Three largest observations.

Figure A.7: Scatterplot of the return of SP500TRIm on month t (y_t) against the return on month $t-1$ (y_{t-1}).

Figures A.7(a) and A.7(b) display the scatterplots of the monthly series. Based on the arrow paths in Figure A.7(b), especially, it may not be that easy to say whether the paths reflect clustering or not. However, if we combine the stretches in both figures we may found that they correspond to the time periods from June 1974 to January 1975, from May 1982 to November 1982, from October 1986 to January 1988 and from May 1998 to November 1998. As these periods are traced in the time plot of the series shown in Figure A.2(b), some more volatile clusters can be observed.

It was also claimed earlier that volatile periods are often predeced by large negative returns in financial time series. If this were the case, the routes formed by the arrows in the scatterplots should almost invariably leave the main cloud in a southern direction (i.e. today's return is large and negative). This feature is quite well reflected in Figures A.5(a) A.5(b). The negative sample correlation (-0.259) between y_t^2 and y_{t-1} also supports the conclusion. Also the two other series indicate the same behaviour. The sample correlations between y_t^2 and y_{t-1} of data series SP500TRIs and SP500TRIm are -0.425 and -0.172, respectively.

It is interesting to note, that although the features studied above (fatter-tailed distribution and time varying volatility) are more commonly associated with the higher-frequency financial time series, they do seem to be present also in our yearly based data series SP500TRIy. However, as discussed above, not all features seem to be that clearly presented in the shorter yearly data series SP500TRIs which excludes the Great Depression in the 30's as well as the World War II. Commonly, nonlinear time series models are suggested to describe the features studied above. However, we will first take a look at the linear models in Section 4 and Appendix B and leave the more complex models to the subsequent Appendices.

Appendix B

ARIMA modelling

In this chapter we report and discuss the most essential findings and results of the iterative steps 2–4 of the ARIMA modelling procedure described in Section 4.6. The R software (version 2.5.1) is here used to analyse the data. For the ARIMA modelling we have applied the function `arima()` contained in package tseries. Alternatively, for example, `armaFit()` function in package fseries could have been used. We aim at giving a comprehensive description only on the ARIMA modelling of the S&P yearly Total Return Index 1925–2006. For the other two models we content with only briefly commenting the chosen models. The modelling procedure for all three data sets is exactly the same.

B.1 S&P 500 yearly Total Return Index 1925–2006

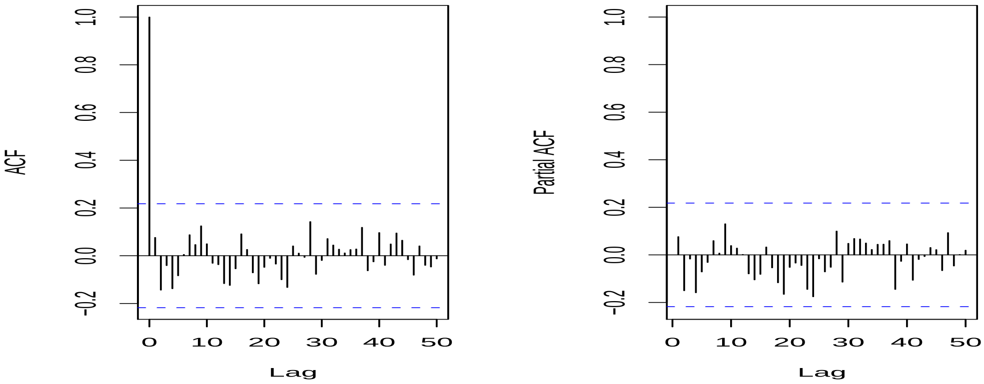
Step 1: Make data transformations

The data transformation was discussed in Section A.1 in Appendix A. The abbreviation SP500TRIy is used for the log returns y_t of the particular data set.

Step 2: Calculate key statistic and determine preliminary values of the orders p and q

After suitably transforming the data (i.e., the series indicates no apparent deviations from stationarity nor apparent trend or seasonality), the next step is to identify, if necessary, preliminary values of the autoregressive order p and the moving average order q. The sample autocorrelation function (ACF) and the partial autocorrelation function (PACF) are the most relevant statistics to be used in this model specification step (see e.g. Brockwell & Davis, 2002). For example, a sample ACF that is 1 at lag zero and close to zero elsewhere suggests that a white noise model might be appropriate for the data. The sample ACF and PACF of the transformed data series are displayed in Figure B.1. The figure shows that neither the sample ACF nor the sample PACF show significant peaks[1] and hence a white noise model (i.e. ARMA(0,0)) might be adequate for the data.

[1]For independent and indentically distributed (IID) random variables with mean zero and constant variance the sample autocorrelations are approximately IID $N(0, 1/n)$ for large n. If the underlying noise process is IID, approximately 95% of the sample autocorrelations for nonzero lags should fall between the bounds $\pm 1.96/\sqrt{n}$.



(a) The sample ACF. (b) The sample PACF.

Figure B.1: The sample ACF and PACF of SP500TRIy.

Step 3: Fit various models and compare the results

Although the sample ACF and PACF suggest the ARMA(0,0) model, we will fit various models with low orders p and q. Conditional-sum-of-squares method is used to find starting values. Maximum likelihood (ML) method is used for the final estimation of the model parameters. The models are compared with the AIC, BIC and AICC criteria functions. Table B.1 reports the criteria values of the fitted models. It can be seen that ARMA(0,0) gives the smallest value for all three criteria functions. A closer look at the estimated ARMA(0,0) model shows that the constant parameter is significant ($\hat{\alpha} = 0.0992$, $\hat{\sigma}_\alpha = 0.0212$) and should be included in the model. We choose this model for further analysis.

Table B.1: Criteria values of the various ARMA(p, q) models fitted to SP500TRIy.

p	q	m	AIC	BIC	AICC
0	0	2	-34.621	-29.832	-34.467
0	1	3	-33.281	-26.097	-32.969
0	2	4	-33.456	-23.879	-32.930
1	0	3	-33.091	-25.907	-32.779
1	1	4	-32.827	-23.249	-32.300
1	2	5	-31.973	-20.000	-31.173
2	0	4	-32.938	-23.360	-32.412
2	1	5	-31.778	-19.806	-30.978
2	2	6	-35.316	-20.950	-34.181

AIC $= -2\log(L)+2m$, BIC $= -2\log(L)+m\log(n)$, AICC $= -2\log(L)+2mn/(n-m-1)$, L = the likelihood of the model, m = the number of estimated parameters in the model, n = the number of observations

Step 4: Use diagnostic tests for the residuals of the fitted model

This step includes the analysis of residuals. We first plot the standardized residuals of the model fit in Figure B.2. If the model fits well, these residuals should behave as a white noise sequence with mean zero and variance one. The plot should be inspected for any obvious departures from this assumption. One should look for trends, cycles and nonconstant variance, any of which suggest that the fitted model is inappropriate. It can be noted that Figure B.2 shows no obvious patterns and only one residual exceeds three standard deviations in magnitude.

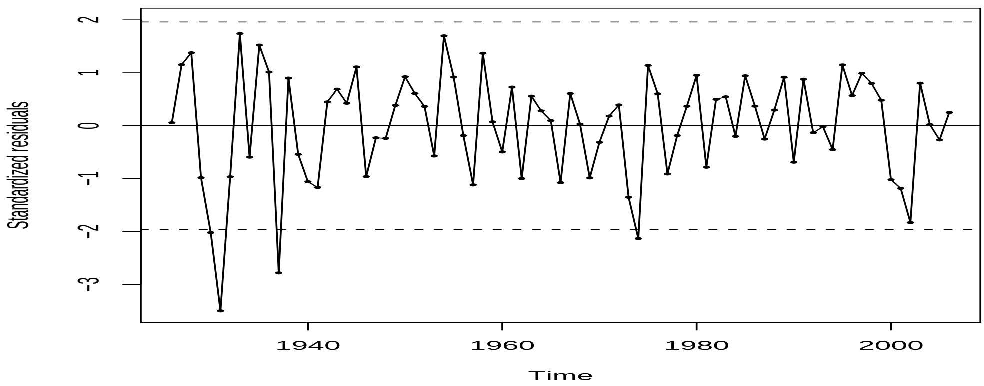
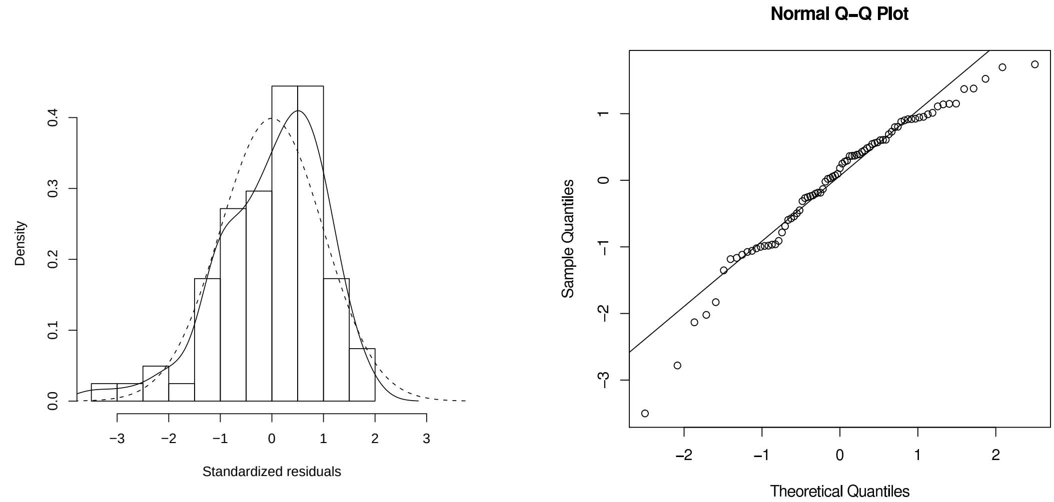


Figure B.2: Standardized residuals of the ARMA(0,0) model fitted to SP500TRIy. Also the ± 1.96 bounds (dashed) are drawn in the figure.

To further examine if the residuals behave as a white noise, we may also inspect the sample ACF of the residuals of the model fit and test whether the autocorrelations are equal to zero for lags greater than zero. If autocorrelation is found, there is a need to modify the model by incresing the value of p and/or q. In the case of the ARMA(0,0) model the individual elements of the sample ACF are identical to those calculated and drawn for the data in Figure B.1(a). Clearly, no significant peaks are seen in the figure.

The residuals of the model fit can be tested for residual autocorrelation by general tests that take into consideration the magnitudes of the single elements of the sample ACF as a group. We use the Box-Pierce test (or the Portmanteau test) and Ljung-Box test (or modified Box-Pierce test) to examine jointly if several residual autocorrelations are zero (see e.g. Shumway & Stoffer, 2006; Brockwell & Davis, 2002; Tsay, 2005). Both tests show no significant autocorrelation (i.e. p-value > 0.05) for the lags 1–20.

Note, that it is a common practice that the 5% significance limits of the sample autocorrelations and the goodness-of-fit tests above are based on the asymptotic results obtained for an IID process. For example, in ARMA models with non-independent innovations, the standard Box-Pierce and Ljung Box tests can perform poorly (Romano & Thombs, 1996). For those more interested in the behaviour of the residual autocorrelations in the framework of ARMA models with non-independent error terms, we refer to Francq et al.



(a) Histogram of the residuals. (b) Normal Q-Q plot of the residuals.

Figure B.3: Normality check of the standardized residuals of the ARMA(0,0) model fitted to SP500TRIy time series.

(2005). The turning point test, the difference-sign test and the rank test can be used to check the hypothesis that the residuals are observed values of IID random variables (see Brockwell & Davis, 2002).

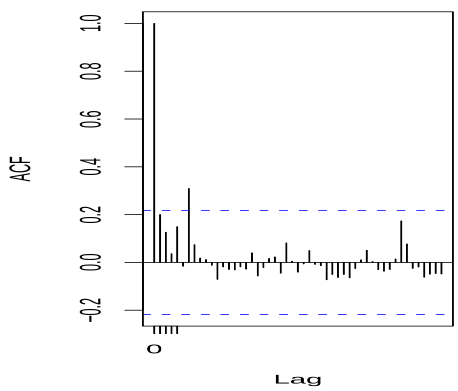
A usual more strict assumption for the white noise series ϵ_t is that its realizations are independent and indentically distributed according to a normal distribution with mean zero and variance σ_ϵ, i.e $\epsilon_t \sim$ IID $\mathrm{N}(0, \sigma_\epsilon)$. In Step 4 also this normality assumption should be checked. According to Franses & Dijk (2000), the rejection of normality may indicate (1) the existence of outlying observations, (2) heteroscedastic error process, and/or (3) that the data should better be described by a nonlinear time series model.

As a first check of marginal normality one can draw a histogram and a Q-Q plot (or a normal probability plot) from the estimated strandardized residuals. These are shown in Figures B.3(a) and B.3(b), respectively. In Figure B.3(a) also the kernel estimate (solid) with bandwith = 0.3665441 and normal distribution (dashed) with the parameters equal to the sample mean and variance of the residuals are drawn. In the Q-Q plot, if the residuals follow a normal distribution, the points should fall approximately along the reference line. Fat tails show up as deviations below the reference line at the lower quantiles and as deviations above the line at the upper quantiles. Both the figures indicate slight deviation from the normal distribution, especially at the tail ends. The normality assumption can also be inspected in Figure B.2. If substantially more than 5% of the estimated standardized residuals lie outside the bounds ± 1.96 (dashed lines) or if there are rescaled residuals far outside these bounds, then the normality assumption should be rejected. In our case four (4.9%) of the estimated standardized residuals are greater in magnitude than 1.96 and a few fall far outside the bounds. Also this reflects a slight deviation from normality.

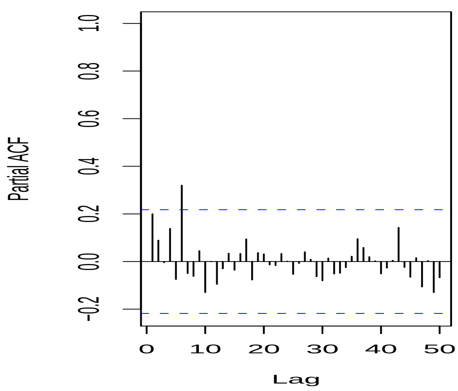
In addition to the visual inspection of the marginal normality, we may also use various general tests for examining the normality of the residuals. We use here the Jarque-Bera (Bera & Jarque, 1980) and the Shapiro-Wilk

(Royston, 1980) tests, which both give us highly significant p-values leading to the rejection of the null hypothesis, i.e. normality.

As mentioned earlier, one indication of non-normality may be that the error process is heteroscedastic. According to Franses & Dijk (2000) neglecting the heteroscedasticity may have quite severe consequences. One way to visually inspect the heteroscedasticity is to draw the sample ACF and PACF of the squared residuals shown in Figures B.4(a) and B.4(b), respectively. Positive autocorrelation in squared terms would, for example, indicate that volatility comes in clusters. Both the sample ACF and PACF indicate that there might be some sort of autocorrelation in the squared residuals. One may also use the sample ACF of the absolute residuals for the visual inspection of heteroscedasticity. In our case the sample ACF of the absolute residuals shows a significant peak for the lag 1. Note, that the focus is in the absolute and the squared residuals since our main interest, the variability (volatility) of the residuals, cannot be observed. The observable quantities $|R_t|$ and R_t^2 are often considered as surrogates or estimators of σ_t and σ_t^2, respectively.



(a) The sample ACF.



(b) The sample PACF.

Figure B.4: The sample ACF and PACF of the squared residuals of the ARMA(0,0) fitted to SP500TRIy.

A general test for testing the null hypothesis of constant residual variance against an unspecified alternative was developed by McLeod & Li (1983). The statistic is actually computed in exactly the same way as the Box-Ljung test except that it tests for autocorrelation in the squared residuals (see Franses & Dijk, 2000). We calculated the test statistic for the lags 1–20 and found significant p-values (p-value < 0.05) for the lags 6, 7 and 8. We conclude that despite no autocorrelation in the residuals (cf. the Box-Pierce and Ljung-Box tests above), there may exist some other type of serial dependence in the error process.

To summarize the results of the whole procedure, among all the fitted linear time series models the ARMA(0,0) model seems to be most appropriate for the log returns y_t. This means that ARIMA(0,1,0) model is suggested for the logarithmic index values $\ln P_t$. For the model formula and the estimated model we refer to Equations 4.9 and 4.10 in Section 4.6, respectively. However, according to the diagnostics in Step 4 there are strong arguments for using nonlinear time series models instead of linear models. This is consistent with the numerous authors who have obtained better agreement for their financial data using nonlinear models.

B.2 S&P 500 yearly Total Return Index 1955–2006

Among the ARMA models, the ARMA(0,0) model seems to be most appropriate for the log returns, i.e. the ARIMA(0,1,0) model is suggested for the logarithmic index series. For the estimated model we refer to Equation 4.11 in Section 4.6, respectively. The diagnostics support our model choice quite nicely. However, the squared residuals indicate that there still might be some serial dependence left in the residuals and hence nonlinear models should also be considered.

B.3 S&P 500 monthly Total Return Index 1955–2006

Among the fitted linear time series the ARMA(0,0), model seems to be most appropriate for the logarithmic index series, i.e. the ARIMA(0,1,0) model is suggested for the log prices. For the estimated model we refer to Equation 4.12 in Section 4.6. However, the diagnostics reveal that linear time series model might be insufficient for the monthly data and nonlinear models should be used instead.

Appendix C

GARCH modelling

In this chapter we report and discuss the most essential findings and results of the GARCH modelling described in Section 4.7. We will base the modelling on the log returns y_t. The R software is used for the model estimation. Similar to Appendix B, also here we aim at giving a comprehensive description only on the modelling of the S&P yearly Total Return Index 1925–2006. For the other two data series we content with only brifely commenting the chosen models. The modelling procedure for all three data sets is exactly the same.

C.1 S&P 500 yearly Total Return Index 1925-2006

Step 1: Specify a mean equation by testing for serial dependence in the data and, if necessary, build a model for the time series to remove any linear dependence

This step was accomplished in Appendix B, where the ARMA(0,0) model with a drift was chosen.

Step 2: Use the residuals of the mean equation to test for ARCH/GARCH effects

The form of the mean equation was identified in Appendix B. In the sequel, the residuals will be denoted by r_t. We now proceed by studying if the residuals of the fitted mean equation model show evidence of ARCH/GARCH effects. When a time series is said to have the ARCH effect or GARCH effect, it exhibits autoregressive conditionally heteroskedasticity, that is volatility clustering. Volatility clustering implies a strong autocorrelation in squared returns. Hence, a simple method for detecting ARCH/GARCH effects is to calculate the first-order autocorrelation coefficient in squared residuals. The ARCH/GARCH effetcs were actually tested already in Step 4 of the ARIMA modelling procedure in Appendix B. For SP500TRIy, the sample ACF and PACF as well as the McLeod & Li test indicated autocorrelation in squared residuals. Hence, ARCH/GARCH effects, i.e. volatility clustering, seem to be present in the data. More on testing ARCH/GARCH effects can be found e.g. in Franses & Dijk (2000).

Before moving to the next step, we may take a look at the standardized

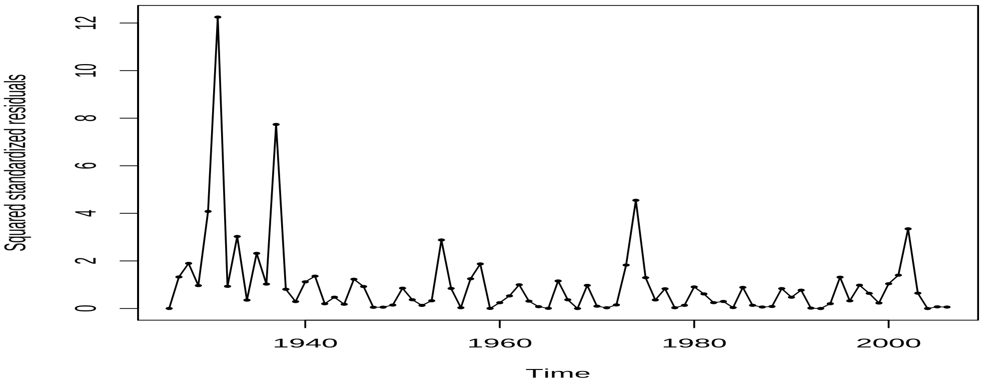


Figure C.1: The squared residuals of the ARMA(0,0) fitted to SP500TRIy.

residuals of the fitted mean model in Figure B.2 and also at their squared terms in Figure C.1. The purpose is to see wheather the series exhibit volatility clustering or varying volatility which are typical, especially, for high-frequency financial time series. However, in a yearly data like this the clusters may not be as clearly observed as they usually are in more frequent financial time series. Figure C.1 shows certain higher peaks and longer calm periods reflecting possible changes in volatility during the study period.

Step 3: Specify a volatility model if ARCH/GARCH effects are statistically significant and perform a joint estimation of the mean and volatility equations

The PACF of r_t^2 can be used to determine the ARCH order but it may not be effective if the sample size is small (Tsay, 2005). Also more generally, the orders obtained this way are not very accurate. Figure B.4 in Appendix B suggests setting the ARCH order s equal to 6. The specification of the order of the GARCH model is more difficult and we will not suggest any tools for that. Since overparameterization is not recommended, we confine ourselves with only lower order models such as e.g. GARCH(1,1), GARCH(1,2) and GARCH(2,1). Usually, GARCH(1,1) is already good enough to capture the variance's performance. We also consider some pure ARCH models.

We use R to find the maximum likelihood estimates of the conditionally normal model, i.e. assuming that ϵ_t are i.i.d. standard normal. The estimation of the ARMA(p, q)-GARCH(r, s) model can be done in two phases: Phase 1 consists of fitting a pure ARMA (ARIMA) model to the data, i.e. modelling the mean equation, and Phase 2 consists of modelling residuals, i.e. volatility, with an ARCH/GARCH model. In R the model estimation in Phase 2 can be done, for example, by function `garch()` contained in tseries package. With this function the only option for the conditional distribution is normal. We will here proceed by fitting the model with simultaneous estimation procedure using the R function `armagarch()`. By simultaneous estimation we mean that both Phase 1 and Phase 2 are accomplished at once. The function is written by Dr. Arto Luoma and the code is available at `http://mtl.uta.fi/`

`codes/tser/armagarch.R`. This function has two options for the conditional distribution: the normal and the Student t-distribution. The simultaneous estimation can also be done in R by the function `garchOxFit()`in fSeries package. This function offers several choices for the conditional distribution.

The AIC, BIC and AICC values of our model candidates are displayed in Table C.1. It can be seen that the ARMA(0,0)-ARCH(1) and ARMA(0,0)-GARCH(2,1) models provide the smallest criteria values. The AIC and AICC values are only slighty better (smaller) for the ARMA(0,0)-GARCH(2,1) reflecting that the more parsimonious ARMA(0,0)-ARCH(1) model might be adequate. A closer look at the estimated models shows that for this particular model all the parameters are significant ($\hat{\alpha} = 0.1163$, $\hat{\sigma}_\alpha = 0.0160$, $\hat{\omega} = 0.0183$, $\hat{\sigma}_\omega = 0.0051$, $\hat{\alpha}_1 = 0.5829$, $\hat{\sigma}_{\alpha_1} = 0.2493$) and should hence be included in the model. This is not the case for the other model candidate. We choose ARMA(0,0)-ARCH(1) model for further analysis.

Table C.1: Criteria values of the various GARCH(r, s) models when mean equation is ARMA(0,0) and the original data series is SP500TRIy. Note that GARCH(0,s) = ARCH(s).

r	s	m	AIC	BIC	AICC	n_{eff}
0	1	3	-41.759	-34.576	-41.448	80
0	2	4	-40.724	-31.146	-40.197	79
1	1	4	-40.338	-30.760	-39.812	80
2	1	5	-42.259	-30.287	-41.459	79
1	2	5	-38.840	-26.867	-38.040	79
2	2	6	-36.923	-22.556	-35.788	79

AIC = $-2C\log(L)+2m$, BIC = $-2C\log(L)+m\log(n)$, AICC = $-2C\log(L) + 2mn/(n-m-1)$, L = the likelihood of the model, m = the number of estimated parameters in the model, n = the number of observations, $C = n/n_{eff}$ is a correction term, where n_{eff} is the effective number of observations used in the computation (see, Brockwell & Davis, 2002, p. 355)

Step 4: Check the fitted model carefully and refine it if necessary

The GARCH models assume that the innovations (shocks) are independent and indentically distributed. Hence, if the model is correctly specified, the standardized residuals (residuals devided by their estimated conditional volatility) should possess a constant variance, lack of serial correlation etc. In particular, the adequacy of a fitted mean model can be checked by examining standardized residuals while squared standardized residuals can be used to test the validity of the volatility equation. To verify that the previous model is appropriate we use the diagnostics introduced in Appendix B (in Step 4).

The mean and standard deviation calculated for the estimated standardized residuals are -0.0692 and 1.004, respectively. The residuals of the GARCH model fit in Figure C.2 seem to show better compatibility with the assumption of behaving as a white noise than those of the pure ARIMA model. This conclusion is supported by the sample ACF and PACF functions which do not indicate autocorrelation in the estimated residuals. As the presence of autocorrelation in the residuals is tested by the Box-Pierce and Ljung-Box

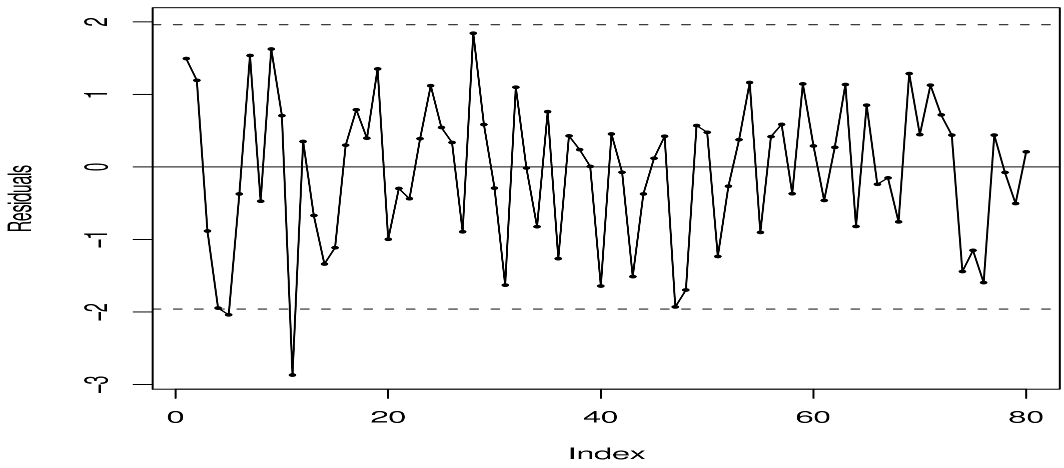
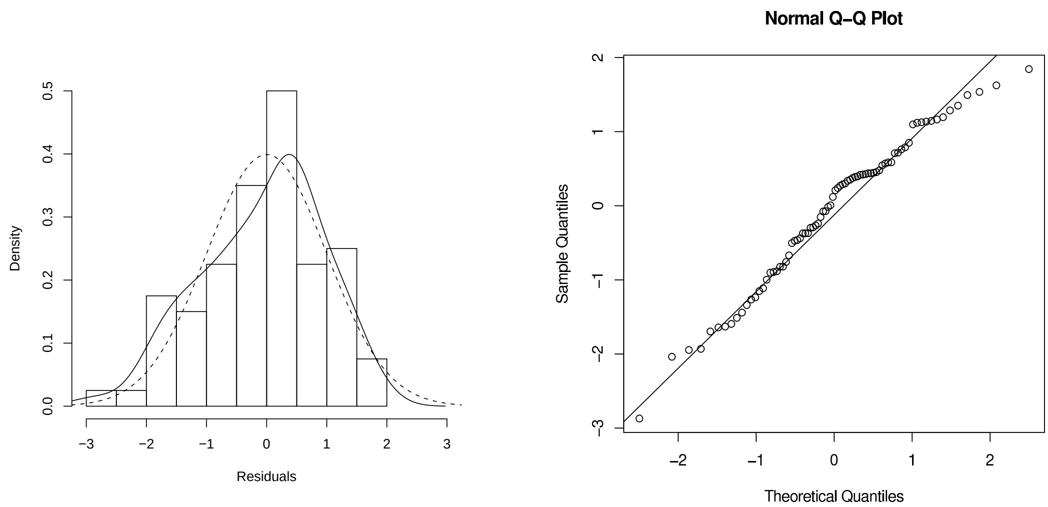


Figure C.2: The standardized residuals of the fitted ARMA(0,0)-ARCH(1) model.

tests, highly nonsignificant p-values (i.e. all p-values >> 0.05) are reported for the lags 1–20.

As the marginal normality[1] of the residuals is checked we first draw a histogram and a normal Q-Q plot of the estimated standardized residuals. The histrogram and the Q-Q plot are shown in Figures C.3(a) and C.3(b), respectively. The residuals show slightly better compatibility with normal distribution than those of of the pure ARMA model. The kurtosis and the skewness of the estimated standardized residuals are 2.540 and -0.4027, respectively. The negative skewness supports the conclusion of the distribution which is skewed left. As the marginal normality is tested by the Jarque-Bera and the Shapiro-Wilk tests, neither of the tests give significant p-values (i.e. p-values > 0.05). Hence, we accept the null hypothesis of marginal normality.

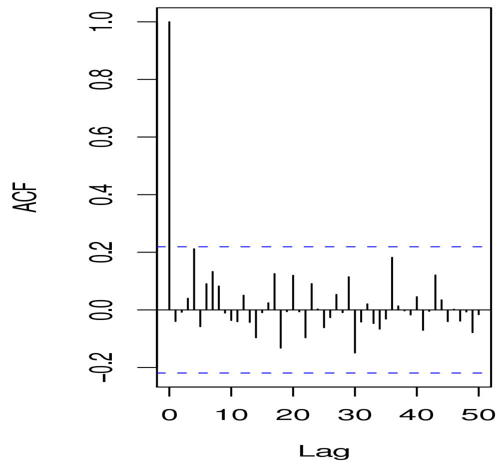


(a) Histogram of the residuals.

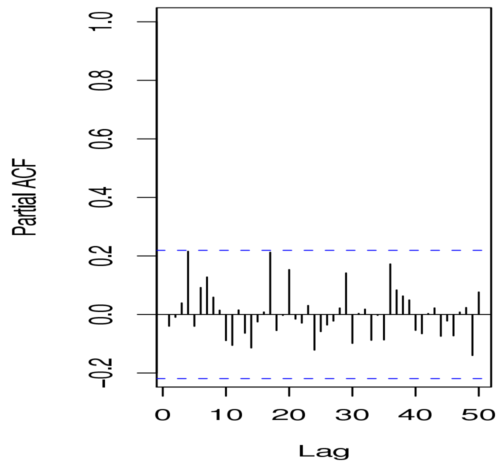
(b) Normal Q-Q plot of the residuals.

Figure C.3: Normality check of the standardized residuals of the ARMA(0,0)-ARCH(1) model fitted to SP500TRIy.

[1]Normal distribution was assumed for the conditional distribution.



(a) The sample ACF.



(b) The sample PACF.

Figure C.4: The sample ACF and PACF of the squared model residuals of the ARMA(0,0)-ARCH(1) fitted to SP500TRIy.

The sample ACF and PACF of the squared and absolute residuals can be used for visual inspection of heteroscedasticity, more specifically, of autoregressive conditional heteroscedasticity. The sample ACF and PACF of the squared standardized residuals are shown in Figures C.4(a) and C.4(b), respectively. Generally, the GARCH model is well specified, if there is no autocorrelation in the squared standardized returns (Alexander, 2001). It can be seen, that neither of the figures show any significant peaks. The McLeod & Li test supports the conclusion by giving highly non-significance p-values (i.e. all p-values >> 0.05) for the lags 1–20 indicating that the model is appropriate. However, if the sample ACF of absolute residuals were drawn, a significant peak could be seen for the lag 4.

To summarize the results, the ARMA(0,0)-ARCH(1) model seems to provide a good fit for the log return series. For the model formula and the estimated model we refer to Equations 4.17 and 4.18 in Section 4.7, respectively. The normality check of the model residuals shown slight negative skewness which indicates that large negative returns occur more often than large positive ones. This feature cannot be captured by the standard ARCH or GARCH models but some more advanced time series models should be applied if the asymmetric feature needs to be taken into account.

C.2 S&P 500 yearly Total Return Index 1955-2006

Our investigations indicate that no ARCH/GARCH terms are needed for SP500TRIs. Hence, we will confine ourselves in the ARMA(0,0) model which regards the conditional volatility as a constant.

C.3 S&P 500 monthly Total Return Index 1955-2006

For the monthly log returns we suggest the ARMA(0,0)-GARCH(1,1) model. For the model formula and the estimated model we refer to Equations 4.19 and 4.20 in Section 4.7, respectively. The model seems to capture well the conditional autoregressive heteroscedasticity observed in the original data series. However, criticism can be addressed to the choice of conditional distribution. The diagnostics clearly indicate that normal distribution is not appropriate but a more fat-tailed distribution might a better choice. The standardized residuals also reflect skewness which cannot be captured by the standard GARCH models.

Appendix D

Regime-switch modelling

In this chapter we report and discuss the most essential findings and results of the regime-switch modelling described in Section 4.8. We will base the modelling on the log returns y_t (cf. Appendix A).

D.1 S&P 500 monthly Total Return Index 1955-2006

We will denote the models described in Section 4.8 as follows. The versions of the Hamilton model will be denoted by Hamilton_0, Hamilton_1 and Hamilton_2, where the index represents the order of the autoregressive process. Var.model_1 and Var.model_2 represent the models described in Equations 4.21 and 4.22, respectively.

For all models we obtained noninformative prior distributions for p and q by specifying as prior parameters $\alpha_p = \beta_p = \alpha_q = \beta_q = 0.5$. These values correspond to the Jeffreys uninformative prior distribution in the standard Bernoulli model. In the Hamilton model we gave improper, noninformative prior distributions for α_0, $\boldsymbol{\phi}$ and σ_ϵ^2. Also in Var.model_1 and Var.model_2 we used improper, noninformative prior distributions for α_0, but for σ_0^2 and σ_1^2 we gave informative conjugate prior distributions. It is essential to give proper prior distributions for σ_0^2 and σ_1^2 so that the posterior distributions will be proper in the case when the entire data is estimated to be in one regime. As hyper-parameters we used the values $v_{\sigma_0^2} = 10$, $s^2_{\sigma_0^2} = s^2$, $v_{\sigma_1^2} = 10$ and $s^2_{\sigma_1^2} = s^2$, where s^2 corresponds to the variance of the observed data. (In principle, the prior distribution should not be dependent on the data but we used s^2 here to get a rough idea of the value range of σ_0^2 and σ_1^2.) The value 10 as a degree of freedom was chosen, since smaller values would produce extremely large values from the scaled inverse-chi-square distribution. The prior distribution of α_1 prevents it from getting a positive value (that is, the state can then be interpreted as a 'low price' state). In all models we specified the values of the prior parameters to be $m = -0.1$ and $v^2 = 0.1^2$, which results in a fairly noninformative prior distribution. In the Hamilton model we made the restriction $\alpha_1 < -0.03$ and in Var.model_1 the restriction $\alpha_1 < -0.01$. In the latter model the restriction was set to be milder because otherwise the estimation became too slow.

Table D.1: Criterion values of different models.

Model	DIC	D_{avg}	p
Var.model$_1$	-2049.78	-2234.631	184.851
Var.model$_2$	-2108.642	-2227.576	118.934
Hamilton$_0$	-2206.699	-2333.851	127.152
Hamilton$_1$	-2235.731	-2340.967	105.2360
Hamilton$_2$	-2231.287	-2341.459	110.172
Random Walk	-2190.972	-2192.983	2.011

In Table D.1 all the basic information criteria (see Section 4.4) for all our models are presented. Since the Hamilton model with AR(1) or AR(2) has less effective observations than the other models, we have done a bias correction by multiplying the likelihood function by $n/(n-r)$, where r is the order of the autoregressive process. According to the DIC, the best model is the Hamilton model with the AR(1) process. According to D_{avg} Hamilton$_1$ and Hamilton$_2$ are almost as good. Hamilton$_2$ has a slightly better value but the difference is very small. Hamilton$_1$ has also the lowest value in the number of effective parameters.

Below are some results from the best model, that is, Hamilton$_1$. In Figure D.1, one simulated chain, produced by the Gibbs sampler, is shown. The chain does not converge rapidly to its stationary distribution, but after achieving the convergency the component series of the chain mix well, that is, they are not too autocorrelated. With all the Hamilton models the convergency was sometimes slow and some chains did not converge at all. One can alleviate this problem by using good initial values, especially by setting the initial value of p to be close to 1. Figure D.2 shows the simulated marginal posterior distributions of the Hamilton model with the AR(1) process. Figure D.3 shows the growth rate of the original index series, i.e. P_t, and the probabilities of the low price state. The summary of the estimation results, based on ten simulated chains, as well as Gelman and Rubin's diagnostics (Gelman et al., 2004) are given below in Table D.1. The values of the diagnostic are close to 1 and thus indicate good convergence.

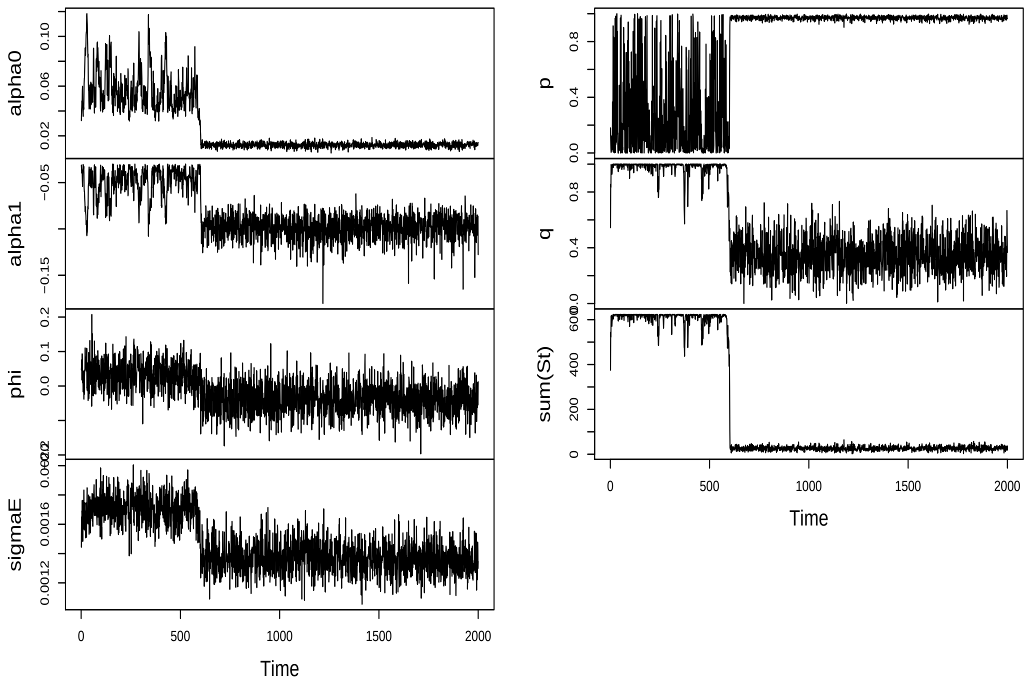


Figure D.1: Iterations of the Gibbs sampler.

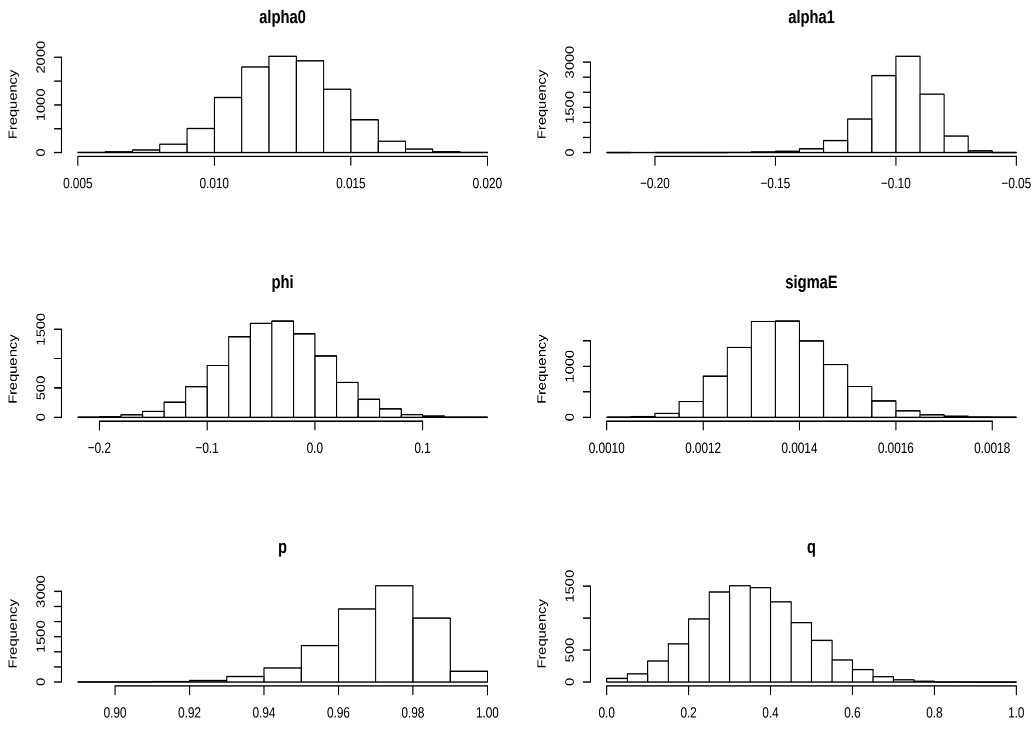


Figure D.2: Simulated posterior distributions of the parameters.

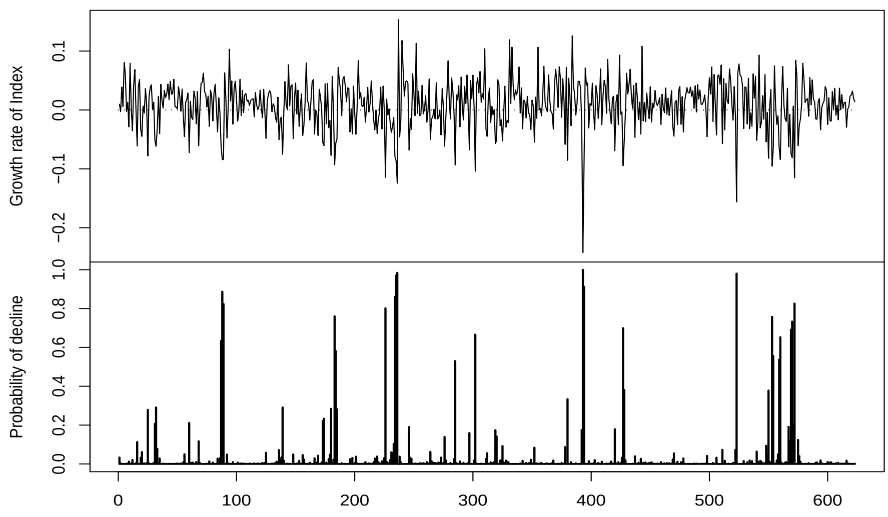


Figure D.3: The growth rate of Index and the probabilities of the low price state.

Table D.2: Estimation results of Hamilton_1 model.

```
Number of chains = 10
Sample size per chain = 1000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean        SD  Naive SE Time-series SE
alpha0    0.012640 0.0018619 1.862e-05      2.100e-05
alpha1   -0.098657 0.0133508 1.335e-04      1.544e-04
phi      -0.037307 0.0475509 4.755e-04      5.071e-04
sigmaE    0.001370 0.0001045 1.045e-06      1.193e-06
p         0.970413 0.0131783 1.318e-04      1.679e-04
q         0.354141 0.1283483 1.283e-03      1.269e-03
sum(St)  26.295600 9.2277048 9.228e-02      1.248e-01

2. Quantiles for each variable:

              2.5%       25%       50%       75%     97.5%
alpha0    0.009011  0.011380  0.012651  0.013922  0.016248
alpha1   -0.127994 -0.106189 -0.097589 -0.089878 -0.075753
phi      -0.130511 -0.069180 -0.037038 -0.005452  0.056659
sigmaE    0.001180  0.001298  0.001364  0.001436  0.001589
p         0.939601  0.962899  0.972043  0.979866  0.991318
q         0.113637  0.264646  0.349628  0.440199  0.615193
sum(St)   9.000000 20.000000 26.000000 32.000000 46.000000

Gelman and Rubin's diagnostics
(Potential scale reduction factors):

        Point est. 97.5% quantile
alpha0        1.00           1.00
alpha1        1.00           1.00
phi           1.00           1.00
sigmaE        1.00           1.00
p             1.00           1.00
q             1.00           1.00
sum(St)       1.00           1.00
```

Appendix E

Stochastic interest rate and equity modelling

In order to experiment with actual data and to estimate the unknown parameters of the models (5.2a) and (5.2b), we chose the following data sets: As an equity index we use the Total Return of Dow Jones EURO STOXX Total Market Index (TMI), which is a benchmark covering approximately 95 per cent of the free float market capitalization of Europe. The objective of the index is to provide a broad coverage of companies in the Euro zone including Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain. The index is constructed by aggregating the stocks traded on the major exchanges of Euro zone. Only common stocks and those with similar characteristics are included, and any stocks that have had more than 10 non-trading days during the past three months are removed. In estimation, the daily quotes from March 4th, 2002 until December 6th, 2007 are used.

As a proxy for riskless short-term interest rate, we use Eurepo, which is the benchmark rate of the large Euro repo market. Eurepo is the rate at which one prime bank offers funds in euro to another prime bank if in exchange the former receives from the latter Eurepo GC as collateral. It is a good benchmark for secured money market transactions in the Euro zone. In the estimation of the interest rate model we use the 3 month Eurepo rate, since it behaves more regularly than the rates with shorter maturities. Both the index and interest series are presented in Figure E.1.

We had no remarkable convergence problems when estimating the model parameters. We used three chains in MCMC simulation, and all chains converged rapidly to their stationary distributions. The summary of the estimation results, as well as Gelman and Rubin's diagnostics (see Gelman et al., 2004), are given in Table E.1. The values of the diagnostic are close to 1 and thus indicate good convergence. All computations were made and figures produced using the R computing environment. To speed up computations the most time consuming loops were coded in C++. The code and data needed to replicate the results can be found at `http://mtl.uta.fi/codes/savings`.

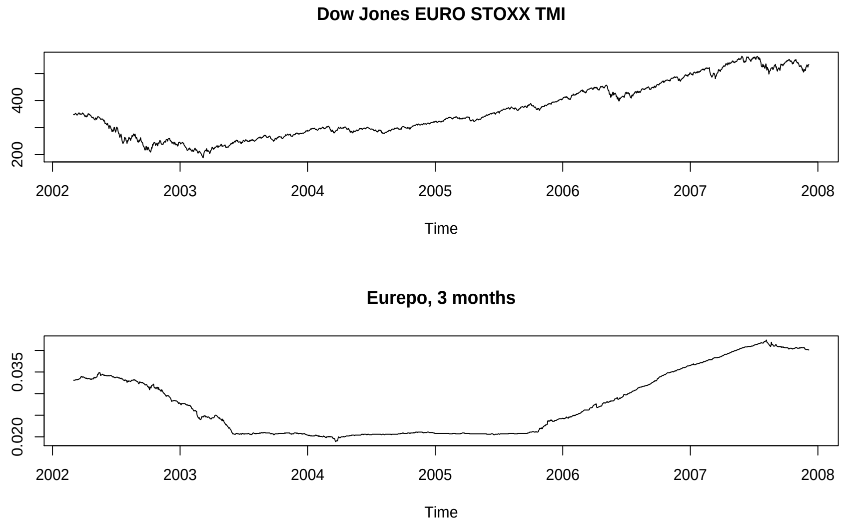


Figure E.1: The equity index and interest series.

Table E.1: Estimation results

```
Number of chains = 3
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean       SD  Naive SE Time-series SE
mu        0.079225 0.067939 5.547e-04      0.0042595
log nu    3.402534 0.327204 2.672e-03      0.0219129
alpha     0.880626 0.055834 4.559e-04      0.0037286
kappa     0.052439 0.045232 3.693e-04      0.0018596
beta      0.221869 0.132709 1.084e-03      0.0060462
tau^2     0.009487 0.001697 1.386e-05      0.0001046
gamma     0.683214 0.087154 7.116e-04      0.0051778
rho       0.091389 0.025618 2.092e-04      0.0016489

2. Quantiles for each variable:

               2.5%       25%       50%     75%   97.5%
mu        -0.048659 0.026263 0.079805 0.12508 0.21444
log nu     2.769493 3.176798 3.401768 3.61407 4.04341
alpha      0.772006 0.843013 0.881094 0.91714 0.98700
kappa      0.001606 0.018515 0.039505 0.07354 0.16534
beta       0.035210 0.126786 0.200871 0.29001 0.53552
tau^2      0.006695 0.008333 0.009257 0.01045 0.01355
gamma      0.504586 0.627500 0.687341 0.74042 0.84705
rho        0.041503 0.075450 0.090498 0.10661 0.14419

Gelman and Rubin's diagnostics
(Potential scale reduction factors):

           Point est. 97.5% quantile
mu               1.01           1.04
log nu           1.01           1.02
alpha            1.01           1.02
kappa            1.01           1.03
beta             1.01           1.03
tau^2            1.02           1.06
gamma            1.04           1.10
rho              1.01           1.03
```

The posterior distributions of the parameters α (Equation 5.2b) and γ (Equation 5.2a) are shown in Figure E.2. As already noted, the CEV model becomes the geometric Brownian motion (GBM) when $\alpha = 0$. The figure reveals clearly that the posterior probability of α being around zero is vanishingly small, which makes the GBM highly improbable. Both models have, under the risk-neutral probability measure, equal expected yields for the underlying index, but the volatility will be greater with the GBM, since in the CEV model the volatility decreases as the level of the process increases. But greater volatility increases the probability of great profits, while not increasing the probability of great losses, since the accumulated capital is guaranteed to the customer. Consequently, the price of an option is greater in the case of the GBM. This illustrates how the approach to use general models efficiently prevents the model error resulting from the use of a too simple model. On the other hand, we see that $\gamma = 1/2$ is not highly improbable in the interest rate model, so the model error would not be large if the CIR model were used instead of the more general model.

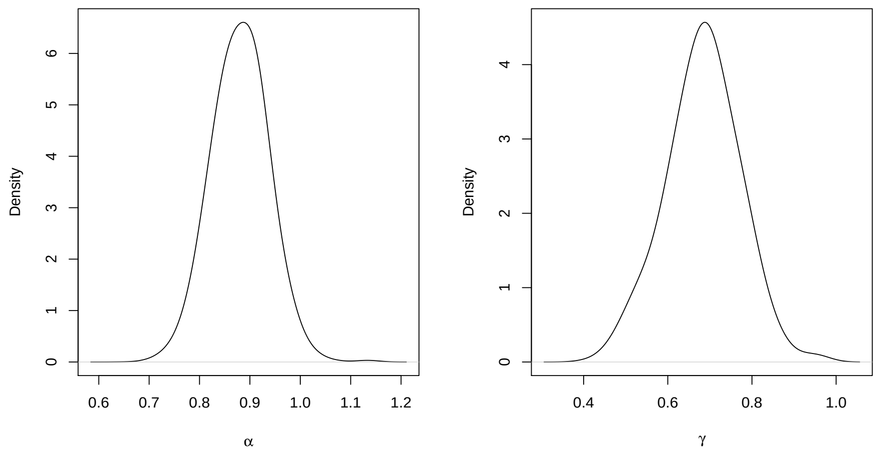


Figure E.2: Posterior distributions of the parameters α (index model) and γ (interest rate model).

There is an error related to the use of Euler discretization in estimation and simulation. However, the effect of discretization is vanishingly small, since our discretization interval is very short, one working day. If daily data were not available, one could use the high frequency augmentation technique described in Jones (1998) for estimation. On the other hand, it is important to select correct index and interest rate models. For example, a failure to choose a realistic model for the stock index might lead to over- or underestimation of volatility, which would make the price estimates biased.

Appendix F

HJM modelling

The aim in this Appendix is to present a HJM model where each of the factors has been determined by Principal Component Analysis. We first aim at obtaining the initial forward rate curve from the zero-coupon bond yields and estimating the volatility function(s) of the HJM model using PCA. In this part we closely follow Jarrow (2002, Chapter 16). In the latter part of this Appendix we briefly discuss how to price derivatives when the forward rate process follows a HJM model. This part of Appendix closely follows Wilmott (2001). The purpose of the Appendix is purely illustrative.

F.1 Interest rate data

For estimating the volatility function(s) we have chosen to use the U.S. monthly nominal interest rates which span from January 1952 to February 1991. The zero-coupon bond yields are obtained from the extended McCulloch dataset[1] (see McCulloch, 1975; 1990, and Kwon, 1992). Our dataset consists of a time series of monthly zero-coupon yields given as percentage per annum with eight different maturities (3 and 6 months, and 1, 2, 3, 5, 7 and 10 years). Rates are on a continuous-compounding basis and they represent the afternoon of the last business day of the month indicated.

The initial forward rate curve used for calibrating the HJM model is estimated based on U.S. monthly yields from January 1997. These yields are based on constant maturity yields taken from the Federal reserve H.15 Statistical Releases[2].

For various techniques for obtaining the zero-coupon bond prices implicit in observed coupon bond prices we refer to James & Webber (2000).

F.1.1 Some data-analysis

Unlike with equity index series, we omit any deeper data-analysis of our interest rate data. Data is, however, described in some detail. Various summary statistics used in this section were introduced in Appendix A.

[1]The dataset is available in `http://www.econ.ohio-state.edu/jhm/ts/mcckwon/mccull.htm` [cited November 11, 2008]. It consists of monthly zero-coupon rates from 1947 to 1991.

[2]The Federal reserve H.15 Statistical Releases are available in `http://www.federalreserve.gov/releases/h15/`.

F.1.1.1 Zero-coupon bond yields

In Table F.1 we display the summary statistics of the monthly zero-coupon bond yields for the period January 1952 to February 1991, for a total 471 months. It can be seen, that on average the yield increases as the maturity increases. The standard deviation is slighly smaller for the longer maturities than for the short ones. The larger variation in the long end can be guessed also by comparing the minimum and maximum values of different maturities. The kurtosis and skewness are also reported. The skewness indicates that the data is not symmetric for any maturity. All skewness values are positive indicating that the data is skewed right, i.e. the right tail is long relative to the left tail. This feature can be observed also by taking a look at the histograms in Figure F.1.

The evolution of the observed zero-coupon yields over the studied time period is graphed in Figure F.2. The yields for different maturities are seen to evolve in a similar kind of manner indicating strong correlation between different maturities. Correlation matrix given in table F.2 support the conclusion of strong correlation (close to one) between the maturities contained in the analysis.

Table F.1: Summary statistics of the monthly zero-coupon yields.

Maturity	N	Mean	Median	Min	Max	Sd	Skew	Kurt
3 months	471	0.0586	0.0528	0.0062	0.1600	0.0315	0.810	3.561
6 months	471	0.0593	0.0561	0.0069	0.1651	0.0319	0.756	3.431
1 year	471	0.0616	0.0592	0.0085	0.1634	0.0320	0.649	3.096
2 years	471	0.0632	0.0626	0.0115	0.1614	0.0314	0.612	2.964
3 years	471	0.0648	0.0638	0.0141	0.1583	0.0314	0.533	2.689
5 years	471	0.0663	0.0661	0.0177	0.1570	0.0312	0.519	2.557
7 years	471	0.0668	0.0655	0.0207	0.1528	0.0308	0.516	2.502
10 years	471	0.0677	0.0657	0.0234	0.1507	0.0307	0.459	2.313

Table F.2: Correlation matrix for the zero-coupon bond yields.

	3 m	6 m	1 y	2 y	3 y	5 y	7 y	10 y
3 m	1.000	0.997	0.990	0.975	0.962	0.943	0.934	0.923
6 m	0.997	1.000	0.995	0.983	0.971	0.953	0.945	0.933
1 y	0.990	0.995	1.000	0.994	0.986	0.971	0.964	0.954
2 y	0.975	0.983	0.994	1.000	0.996	0.989	0.984	0.976
3 y	0.962	0.971	0.986	0.996	1.000	0.996	0.993	0.988
5 y	0.943	0.953	0.971	0.989	0.996	1.000	0.998	0.996
7 y	0.934	0.945	0.964	0.984	0.993	0.998	1.000	0.998
10 y	0.923	0.933	0.954	0.976	0.988	0.996	0.998	1.000

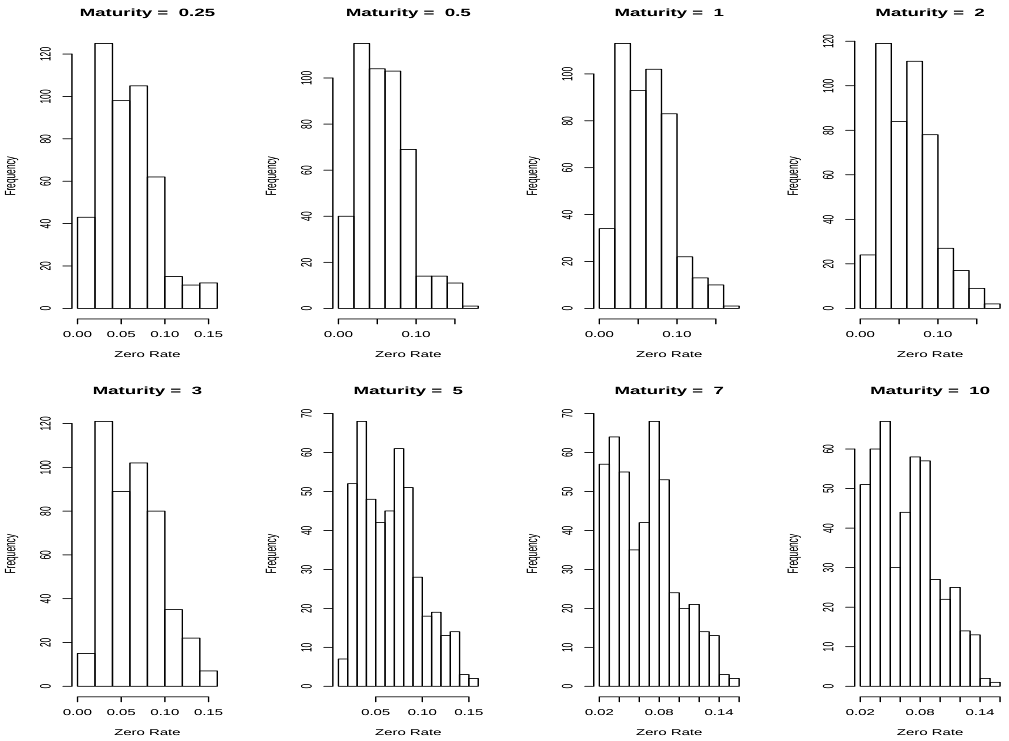


Figure F.1: Histograms of zero-coupon yields for various maturities.

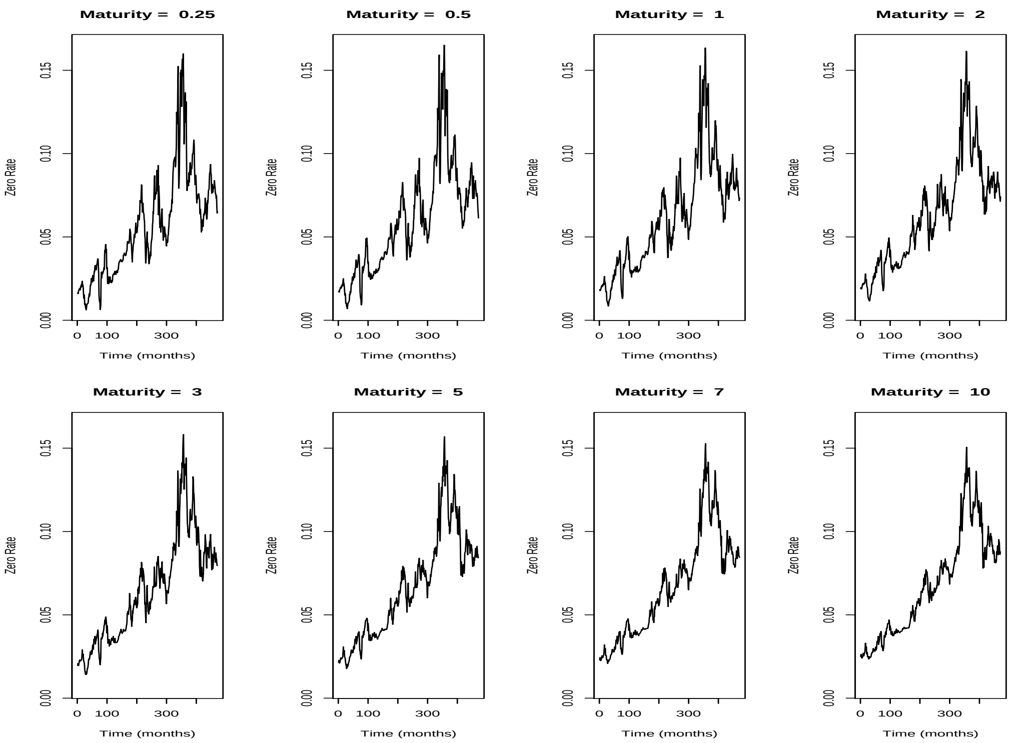


Figure F.2: Evolution of zero-coupon yields for various maturities.

F.1.1.2 Monthly forward rate changes

For the methods deriving the monthly forward rate changes see the subsequent sections of this appendix. In Figure F.3 we may see how the monthly changes in forward rates for various maturities evolve in time. The summary statistics calculated for the forward rate changes are displayed in Table F.3.

Table F.3: Summary statistics of the monthly changes in forward rates.

Maturity	N	Mean*	Median*	Min	Max	Sd*	Skew	Kurt
0–3 m	470	1.04	3.15	-0.046	0.024	5.79	-1.69	16.27
3–6 m	470	0.84	2.10	-0.052	0.027	6.36	-1.09	14.92
6 m–1 y	470	1.42	0.75	-0.038	0.040	6.21	0.11	10.49
1–2 y	470	1.17	3.10	-0.030	0.031	1.17	0.03	9.19
2–3 y	470	1.46	2.15	-0.034	0.039	5.64	-0.07	12.82
3–5 y	470	1.45	3.72	-0.018	0.025	4.14	-0.06	7.87
5–7 y	470	1.21	2.50	-0.039	0.028	5.21	-1.32	14.87
7–10 y	470	1.41	1.05	-0.020	0.029	3.96	0.25	10.57

*To get the actual mean, median and standard deviation, the first two values need to be multiplied by 0.0001 and the third one by 0.001.

Later in this Appendix we will apply PCA to estimate the number of factors that explain the data in a satisfactory way and to specify the volatility functions required to generate the forward rate evolutions using HJM model. Since PCA is employed for the monthly changes in forward rates, one should check if these rates are multivariate normal. Unfortunately, in practice, testing for multivariate normality is more diffucult than testing for univariate normality and relatively few formal methods are available in this context. We here omit the testing of multivariate normality. However, to get some sense on whether there exists at least marginal normality in monthly forward rate changes, we plot their histograms with a normal distribution superimposed on each of the figures (Figure F.4). The normality assumption can be examined also by taking a look at the summary statistics (e.g. skewness and kurtosis) in Table F.3. Clearly, all kurtosis values are greater than three indicating that the sample distibutions have sharper peak and heavier tails than the corresponding normal distributions.

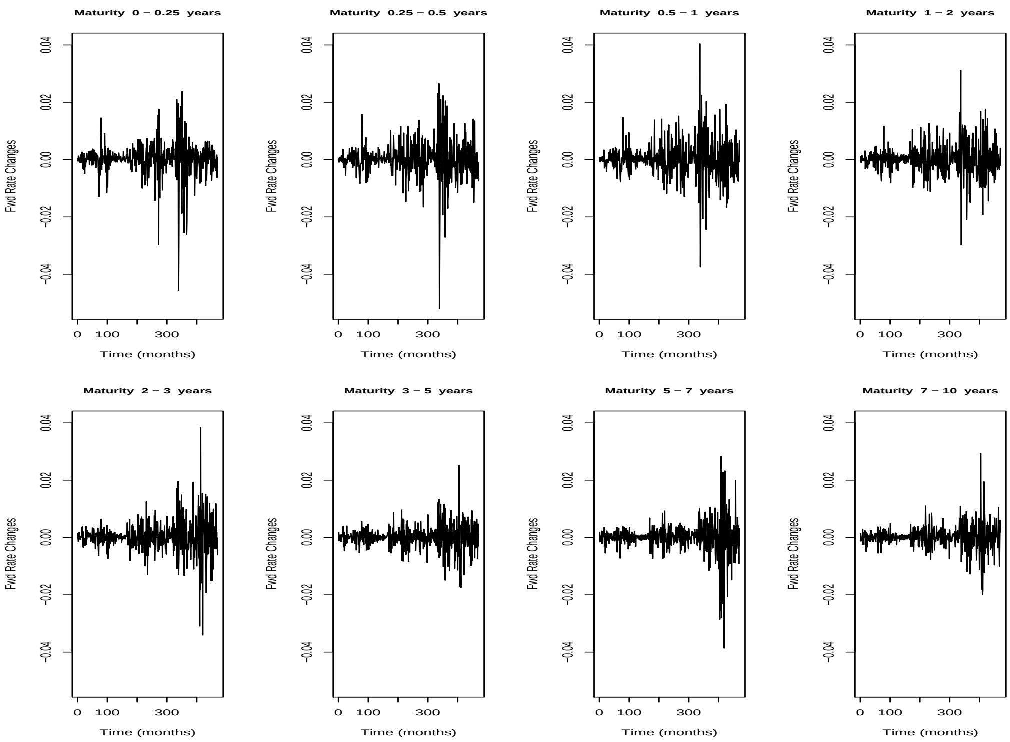


Figure F.3: Evolution of monthly changes in forward rates for various maturities.

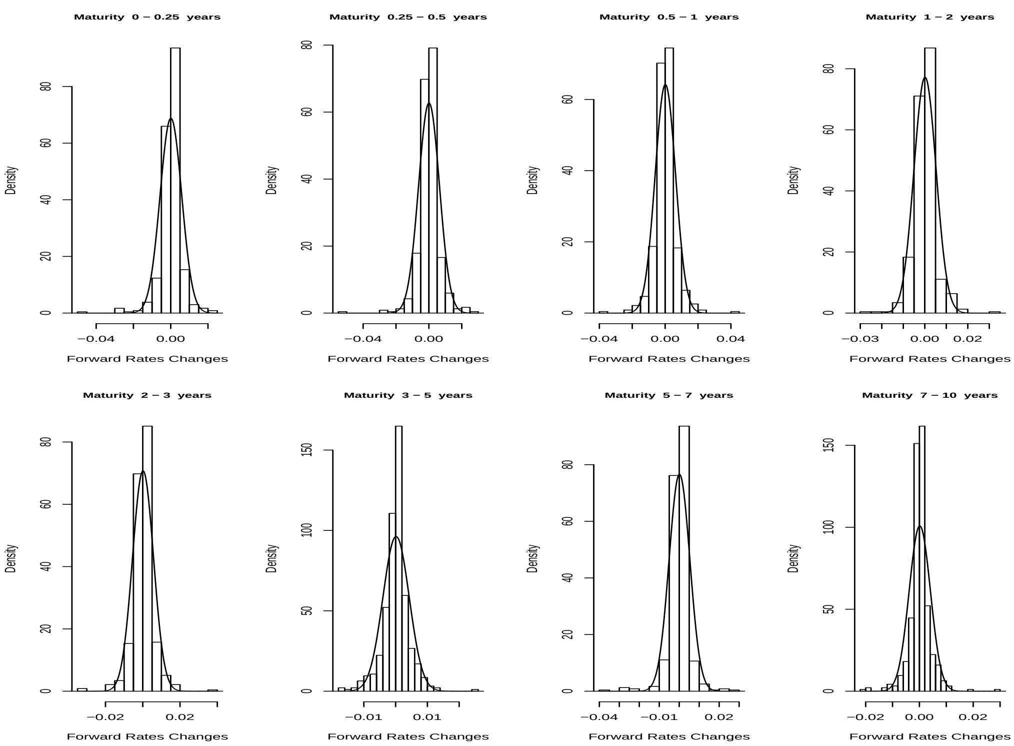


Figure F.4: Histograms of changes in forward rates of various maturities.

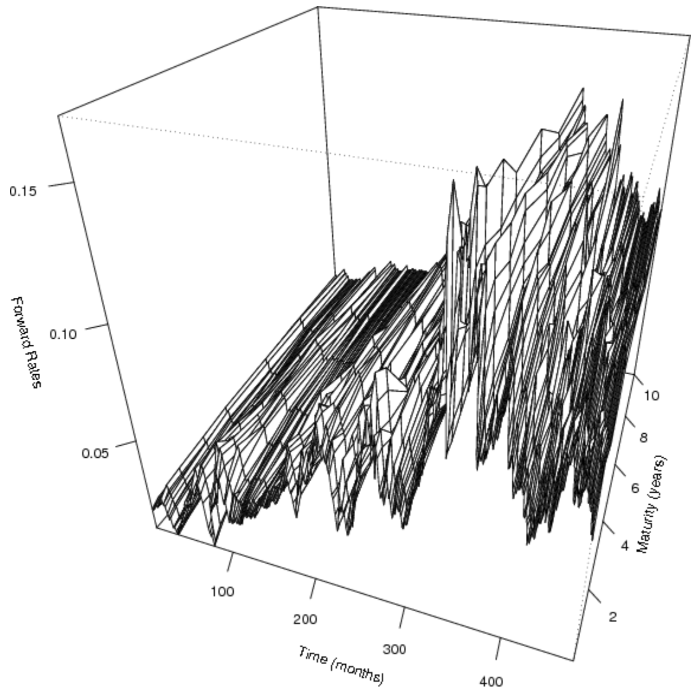


Figure F.5: Forward rate curve evolution over the period from January 1951 to February 1991.

F.2 Forward rate curve estimation

Instantaneous forward rates are not observable. As a proxy one often uses the continuously compounded forward rate from $t+x$ to $t+x+\tau$, with x and τ non-negative constants, that is

$$F(t, t+x, t+x+\tau) = -\frac{\log P(t, t+x+\tau) - \log P(t, t+x)}{\tau}. \tag{F.1}$$

It is worth noting, that (F.1) does not provide a satisfactory proxy when τ is large. This is because the approximation of the derivative may then be fairly inaccurate. Another method to obtain instantaneous forward rates would be to interpolate the term structure with a smooth curve (e.g. cubic spline) and then differentiate the curve.

We use (F.1) as a proxy for the instantaneous forward rate. The zero-coupon bond prices necessary for (F.1) we obtain from the zero-coupon bond yields of our dataset using the relationship

$$P(t, T) = e^{-R(t,T)(T-t)},$$

where $R(t, T)$ denotes the yield to maturity at time t of the zero-coupon bond maturing at time T.

Figure F.5 contains a time series graph of the estimated forward rates over the time period from January 1951 to February 1991, that is total 471 months. It can be seen that the forward rate curves evolve across time in a nonparallel fashion. These forward rates are used to estimate the volatility functions, i.e. to study the forward rate's stochastic evolution later in this appendix.

Table F.4: Forward rates and zero-coupon bond prices on January, 1997.

Maturity	Forward rate	Maturity	Zero-coupon bond price
0 – 3 months	0.05170	3 month	0.9871582
3 – 6 months	0.05450	6 month	0.9737994
6 months – 1 year	0.05910	1 year	0.9454446
1 – 2 years	0.06410	2 year	0.8867431
2 – 3 years	0.06460	3 year	0.8312705
3 – 5 years	0.06585	5 year	0.7286950
5 – 7 years	0.06820	7 year	0.6357817
7 – 10 years	0.06836	10 year	0.5178861

The difficulty encountered when estimating the forward rate curve is that we have observed zero-coupon bond yields/prices in discrete spacings, not continuously in time. The observed zero-coupon bond prices are hence insufficient to price interest rate derivatives which have cash flows occuring on days with no observed yield/price information. In Figure F.5, the forward rate curves are plotted piecewise linear. That is, linear interpolation is used to determine the missing forward rates. To get around the missing zero-coupon bond price observations, Jarrow (2002) uses the approach which assumes constant forward rates over the missing maturities. According to Jarrow, this approach which approximates the forward rate curve with a piecewise constant step function is the simplest but perhaps most robust.

As we estimate the initial forward rate curve needed for the calibration of the HJM model, we follow the approach of Jarrow and assume the forward rates to be constant over various maturity time intervals. Table F.4 contains the estimated forward rates and zero-coupon bond prices obtained for January, 1997. The forward rate for 0–3 month is 0.0517. It can be seen that the rates increase as we move to the sequential maturity points.

F.3 Factors and volatilities

We use historic volatility estimation to obtain volatility function(s), which means that we utilize time series observations of past forward rates. Another approach would be implicit volatility estimation which uses current market prices of various interest rate derivatives and inverts the computed price formulas to obtain the volatility functions such that the computed prices best match market prices. This approach is sometimes called curve-fitting (see Jarrow, 2002).

We concern with Gaussian models where the volatility function(s) are deterministic functions that depend only on time to maturity. Hence, $f(t, T)$ is a Gaussian process. An important feature of Gaussian models is analytical tractability, as they may lead, for simpler options, to explicit formulae. This follows from bond prices being log-normal. A drawback of these models is that forward rates may go negative.

As discussed in Section 5.8.3, two or three factors are usually needed to explain a high percentage of the variation in the yield curve. We may hence specify the volatility structure as

$$\sigma_k(t,T) = \sigma_k(T-t), \quad k = 1, 2, 3.$$

The one-factor model is obtained if $\sigma_2(T-t)$ and $\sigma_3(T-t)$ vanish, and the two-factor model is obtained if $\sigma_3(T-t)$ vanishes. The functions $\sigma_i(T-t)$ will be estimated using PCA.

We perform PCA for monthly changes of instantaneous forward rates. Using mathematical notations, monthly changes of instantaneous forward rates can be expressed as

$$\Delta f_i(t) = f(t+\delta, i) - f(t, i),$$

where $i = 1, \ldots, N$, is indexing a maturity point (we have assumed N different maturity points) and δ corresponds to the time period over which the differences are computed, that is one month. Then, $\Delta f(t) = (\Delta f_1(t), \ldots, \Delta f_N(t))'$ is a time-homogeneous N-dimensional normally distributed random process. Note that in our analysis $N = 8$.

We then compute the $N \times N$ sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ for $\Delta f(t)$ and decompose it as

$$\hat{\boldsymbol{\Sigma}} = \mathbf{ALA}', \tag{F.2}$$

where the $N \times N$ matrix $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_N)$ gives the N eigenvectors $(\mathbf{a}_i)$ of $\hat{\boldsymbol{\Sigma}}$, and the $N \times N$ diagonal matrix $\mathbf{L} = \text{diag}(l_1, \ldots, l_N)$ provides the N eigenvalues (l_i). Finding principal components reduces to finding all eigenvalues and eigenvectors of $\hat{\boldsymbol{\Sigma}}$. The eigenvectors are the principal components, and the eigenvalues indicate the amount of variance explained by each component. The eigenvector with the largest eigenvalue is hence the first principal component, the eigenvector with the second largest eigenvalue is the second principal component, and so on. The decomposition (F.2) gives the estimates of the N volatility functions as

$$\begin{bmatrix} \sigma_i(x_1) \\ \vdots \\ \sigma_i(x_N) \end{bmatrix} = \mathbf{a}_i \sqrt{l_i}, \quad \text{for } i = 1, \ldots, N. \tag{F.3}$$

See Jarrow (2002) for more details.

There is a normality assumption underlying the theory of PCA. By taking a look at Figure F.4, we may get sense of if the normality (at least marginally) holds for the monthly forward rate changes in our dataset. The normality assumption seems to appear reasonable. As the PCA analysis is done, one should also check if the data include jumps. This is because PCA assumes that the underlying process is a diffusion, i.e. jumps are not allowed. Since it is an empirical reality that the data do jump (e.g. do to interest rate setting by the monetary authorities), one may need to remove jumps from the data before doing the PCA (James & Webber, 2000, p. 462). The results on the analysis of PCA conducted to the monthly forward rate changes are presented in Table F.5

Table F.5: Results of PCA.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Eigenvalue*	1.300	0.284	0.232	0.187	0.112	0.076	0.066	0.054
Sd*	1.139	0.532	0.481	0.432	0.334	0.275	0.257	0.232
%	56.3	12.3	10.0	8.1	4.8	3.2	2.8	2.3
Cum-%	56.3	68.5	78.6	86.6	91.5	94.8	97.6	100
				Eigenvectors				
0 months	-0.414	0.341	-0.318	0.316	-0.313	0.250	0.520	-0.283
3 months	-0.512	0.178	-0.110	0.217	-0.283	-0.189	-0.527	0.504
6 months	-0.481	0.103	-0.020	-0.110	0.795	0.328	-0.064	0.021
1 year	-0.380	-0.119	-0.136	-0.343	0.030	-0.687	-0.013	-0.481
2 years	-0.299	-0.069	0.901	0.213	-0.115	0.011	0.057	-0.177
3 years	-0.222	-0.144	0.104	-0.577	-0.175	0.020	0.511	0.542
5 years	-0.148	-0.818	-0.194	0.461	0.107	-0.052	0.175	0.121
7 years	-0.171	-0.365	-0.088	-0.363	-0.365	0.564	-0.389	-0.308

*To get the correct eigenvalues and standard deviations (sd), the former values need to be multiplied by 0.0001 and the latter ones by 0.01.

Since the data is eight dimensions, a total of eight principal components are computed. All eight components are given in Table F.5 and the estimates of volatility functions, obtained from (F.3), are contained in Table F.6. As discussed in Section 5.8.3, various studies show that a very high proportion of the movement of the forward rate curve is explained by just three components. This suggests a three-factor model. In our analysis, the first component alone counts for 56.3%, and the second and third components together for an additional 22.3%. As a result the three first components explain about 79% of the total variance. We graph the first three (most significant) principal components against the relevant maturity in Figure F.6. The first component

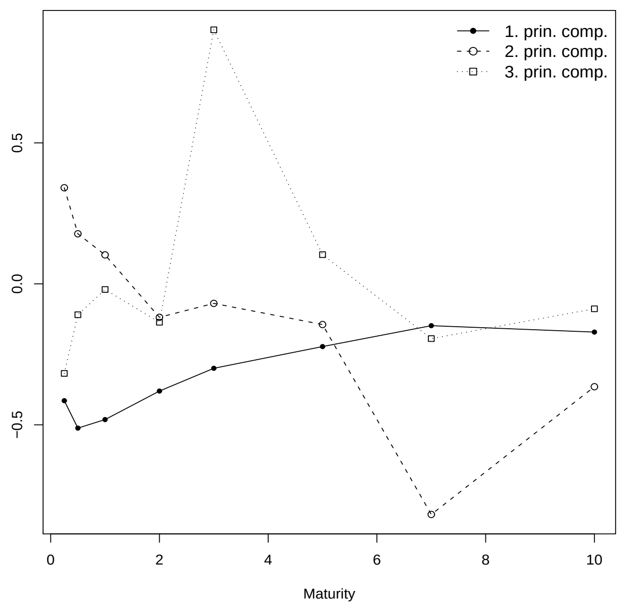


Figure F.6: The three most significant principal components.

is roughly flat, so it moves the forward rates up and down together. The second component is downward sloping causing the short and long end forward rates to move in opposite directions. The third component is hump-shaped and causes the term structure to flex.

The first three volatility functions are given in Table F.6. They are the inputs required in the next section to generate forward rate evolution.

Table F.6: The estimated volatility functions ($\times 10^{-3}$).

	VF1	VF2	VF3
0 months	-4.725	1.817	-1.528
3 months	-5.834	0.946	-0.527
6 months	-5.489	0.548	-0.096
1 year	-4.336	-0.632	-0.655
2 years	-3.414	-0.369	4.336
3 years	-2.536	-0.769	0.498
5 years	-1.691	-4.358	-0.934
7 years	-1.950	-1.944	-0.425

F.4 Pricing derivatives

Pricing derivatives is all about finding the expected present value of all cashflows under the risk-neutral probability measure $\mathbb{Q}$. So, when we come to pricing derivatives we must do so in the risk-neutral world.

Since the HJM model is very general and not necessarily Markov, it is usually not possible to do the calculations via a finite-dimensional partial differential equation. It thus often leaves us with only two alternatives for estimating the necessary expectations. The first alternative is to simulate the random evolution of the risk-neutral forward rates. The other alternative is to build up a tree structure. We here discuss the Monte Carlo simulation approach in some more detail. This discussion closely follows Wilmott (2001). For the tree structure approach we refer to Jarrow (2002) and James & Webber (2000). Matlab software (Financial Derivatives Toolbox) provides functions to generate HJM forward rate trees. Matlab also provides functions to calculate the price of any set of supported instruments, based on an HJM interest rate tree.

When using a Monte Carlo simulation method, we must first simulate the evolution of the whole forward rate curve, then calculate the value of all cashflows under each evolution and finally calculate the present value of these cashflows by discounting at the realized short rate r_t. As we perform a Monte Carlo simulation, we need to proceed by the following steps (see Wilmott, 2001, pp. 324–325):

1. Simulate a realized evolution of the whole risk-neutral forward rate curve for the necessary length of time T.

2. At the end of simulation we will have the realized prices of all maturity zero-coupon bonds at every time up to T.

3. Using this forward rate path calculate the value of all the cashflows that would have occured.

4. Using the realized path for the short rate r_t calculate the present value of these cashflows. Note that we discount at the continuously compounded risk-free rate, not at any other rate. In the risk-neutral world all assets have an expected return of r_t.

5. Return to Step 1 to perform another realization, and continue until you have a sufficiently large number of realizations to calculate the expected present value as accurately as required.

The Monte Carlo simulation from the HJM model may be very slow. The bond prices at all maturities are, however, trivial to find during the simulation.

We illustrate Step 1 of the above list in Figure F.7. The HJM model that serves as the basis for the simulation has been calibrated by the initial forward rate curve obtained in Section F.2 (January, 1997) and the volatility functions specified by PCA in Section F.3. We have chosen to use a three-factor HJM model. The simulation has been carried out by applying the spreadsheet formulae and Visual Basic programs provided by Wilmott (2001). The implementation is that provided by Wilmott, except for some modifications made to match the volatility structure with that estimated by PCA above. The timestep was chosen to be one month, i.e. 1/12 years. Simulation methods for HJM models are discussed e.g. in James & Webber (2000). For information on methods for HJM models one should also see Clewlow & Strickland (1998).

The simulation outcome is displayed in Figures F.7 and F.8. Figure F.7 shows how the simulated forward rates for various maturities evolve over a time period of four years. Figure F.8 shows how the whole forward rate curve evolves during the same time period in our simulation study. Note that this is only one realization of the random process. For the derivative pricing we should generate a large number of evolutions from the same model (cf. Step 5 above). The pricing of put and floor options and more complex exotic options is relevant for some life insurance products (see Section 5.6). We note also that Step 3 above in life insurance applications may not always be straightforward under the risk neutral measure (for instance customer and management actions are typically assessed under the real world measure).

Evaluation of of term structure models is quite often performed in terms of pricing accuracy. For the scope of this report, we omit any evaluation of our model calibrated above. Naturally, it would be important and highly essential to know how well the model serves its purposes.

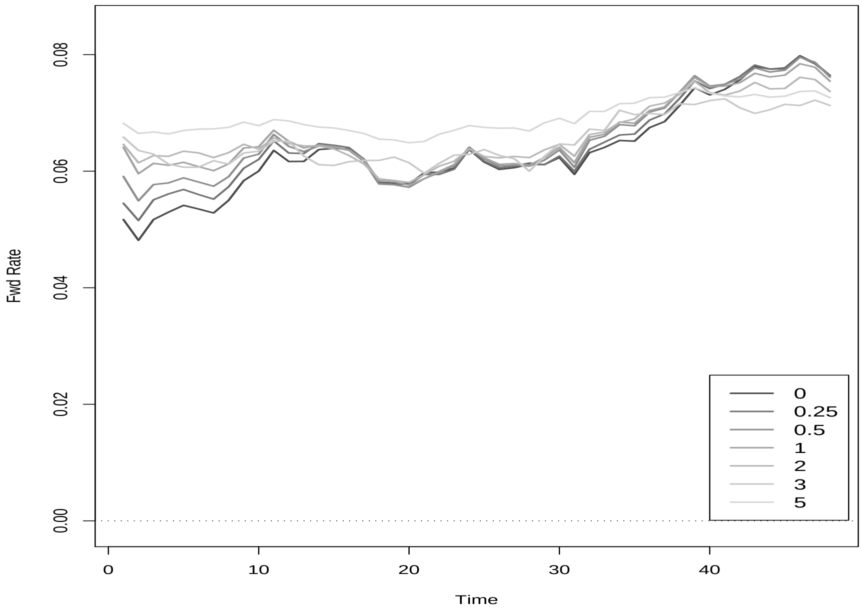


Figure F.7: Simulated paths of forward rates for various maturities.

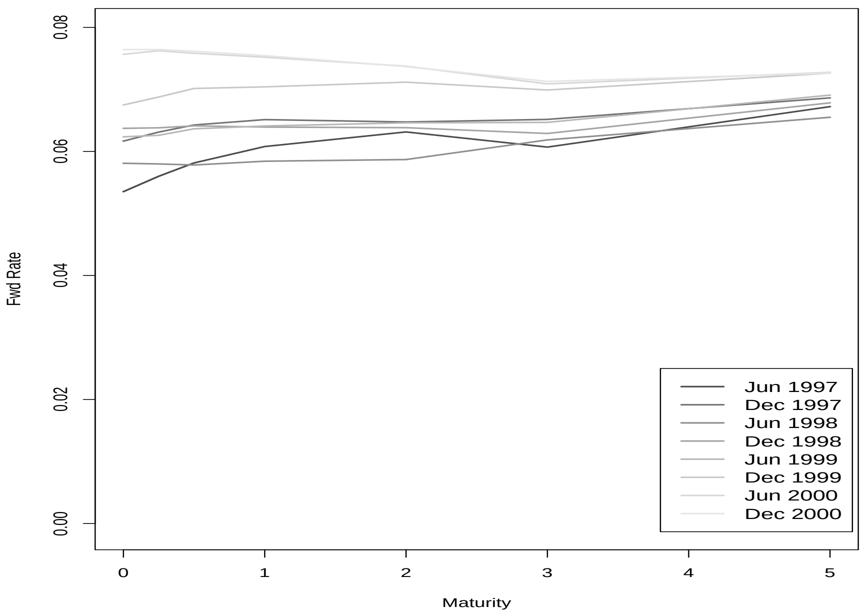


Figure F.8: Simulated forward rate curves at various times.

Vakuutusvalvontavirasto
Mikonkatu 8, PL 449
00101 Helsinki
+358-9-4155 950
+358-9-4155 9515
www.vakuutusvalvonta.fi

Försäkringsinspektionen
Mikaelsgatan 8, PB 449
FIN-00101 Helsingfors
+358-9-4155 950
+358-9-4155 9515
www.vakuutusvalvonta.fi

Insurance Supervisory Authority
Mikonkatu 8, P.O. Box 449
FIN-00101 Helsinki
+358-9-4155 950
+358-9-4155 9515
www.vakuutusvalvonta.fi

ISSN 1457-201X
ISBN 978-952-5350-52-4