Universität Ulm
Fakultät für Mathematik und Wirtschaftswissenschaften



Aspects of Estimation Uncertainty in Risk Management

Dissertation

zur Erlangung des Doktorgrades Dr. rer. pol.
der Fakultät für Mathematik und Wirtschaftswissenschaften
der Universität Ulm

vorgelegt von
Paul Manuel Körbitz
aus Bielefeld

Mai 2013

Amtierender Dekan: Prof. Dr. Dieter Rautenbach
1. Gutachter: Prof. Dr. Gunter Löffler
2. Gutachter: Prof. Dr. Rüdiger Kiesel
Tag der Promotion: 16. Juli 2013

Contents

1. Introduction **1**
- 1.1. Motivation and problem formulation . . . 1
- 1.2. Outline and contributions . . . 5

2. A Bayesian approach to risk management **11**
- 2.1. Introduction . . . 11
  - 2.1.1. Why be Bayesian? . . . 11
  - 2.1.2. Related literature . . . 12
  - 2.1.3. Overview . . . 19
- 2.2. Bayesian statistical models and estimation . . . 19
  - 2.2.1. Bayesian models . . . 19
  - 2.2.2. Point estimates . . . 24
  - 2.2.3. Bayes factors . . . 25
  - 2.2.4. Posterior marginal distributions and Bayesian model averaging . . . 28
- 2.3. Markov chain Monte Carlo . . . 30
  - 2.3.1. Markov chains . . . 31
  - 2.3.2. Simulating and sampling from Markov chains . . . 40
  - 2.3.3. Monitoring and convergence diagnostics . . . 45
- 2.4. Summary and discussion . . . 54

3. Bayesian parameter estimation of short-rate models **55**
- 3.1. Introduction . . . 55
  - 3.1.1. Literature . . . 55
  - 3.1.2. Motivation and contribution . . . 59
- 3.2. Methodology . . . 60
  - 3.2.1. Approximating the transition density of the CKLS model 61
  - 3.2.2. Maximum likelihood estimation and the likelihood ratio test . . . 65
  - 3.2.3. Maximum a-posteriori estimates and Bayes factors . . . 66
  - 3.2.4. Re-parameterizing the CKLS model for MCMC simulations 69
  - 3.2.5. Testing model predictions . . . 71
- 3.3. Data sets and estimation results . . . 73
  - 3.3.1. Data sets . . . 73
  - 3.3.2. Estimation results . . . 77

3.3.3. Model comparison . . . 88
3.3.4. Model predictions test . . . 91
3.4. Conclusion . . . 94

4. Lattice methods for pricing options under stochastic interest rates 97
4.1. Introduction . . . 97
4.1.1. Prior work . . . 97
4.1.2. Motivation and contribution . . . 100
4.2. Trinomial trees for the CEV-CKLS model . . . 100
4.2.1. Constructing the grid . . . 102
4.2.2. Transition probabilities . . . 103
4.3. The QUAD algorithm for models with stochastic interest rates . 107
4.3.1. The QUAD method for one dimensional asset models . . 109
4.4. Computational examples . . . 118
4.4.1. European call . . . 118
4.4.2. Traffic light option . . . 119
4.4.3. Pricing European options with trees and the QUAD algorithm . . . 121
4.5. Conclusion . . . 124

5. Risk management of implicit options under uncertainty 129
5.1. Introduction . . . 129
5.2. A simple term-fix insurance contract . . . 131
5.3. Pricing the contract under stochastic interest rates . . . 135
5.3.1. The asset model . . . 135
5.3.2. Numerical pricing . . . 136
5.4. Estimating parameters and uncertainty . . . 139
5.4.1. Maximum likelihood estimation . . . 140
5.4.2. Bayesian estimation . . . 140
5.4.3. Risk-neutral dynamics . . . 142
5.4.4. Data set and estimation results . . . 143
5.5. Risk management techniques . . . 144
5.5.1. Static reserving . . . 145
5.5.2. Dynamic hedging . . . 152
5.6. Conclusion . . . 158

6. Conclusion 161
6.1. Summary and conclusions . . . 161
6.2. Outlook . . . 163

A. Additional results for Chapter 3 165
A.1. Additional tables . . . 165

B. Additional results for Chapter 4 **173**
B.1. The drift term for the transformed interest rate process 173
B.2. The drift term for the transformed stock process 174

Bibliography **184**

Zusammenfassung **185**

Acknowledgements **191**

Curriculum Vitæ **192**

List of Figures

2.1. Trajectories of two Markov chains 47
2.2. Autocorrelation functions of two Markov chains 48
2.3. Distance of cumulative sums from the mean 51

3.1. Convergence of transition density approximation methods 63
3.2. Likelihood function before and after reparametrization 70
3.3. Short-term interest rates for the US, UK, Japan, France, Germany, Italy, Switzerland, Australia, and the Euro zone 74

4.1. Numerical integrals as branches 110
4.2. Adaptable grid of the QUAD algorithm 112

5.1. Parameter estimates 144
5.2. Loss distributions under static reserving 146
5.3. Loss distributions under dynamic hedging 153

List of Algorithms

1. Gibbs sampling 41
2. Metropolis-Hastings algorithm 43
3. Random walk Metropolis-Hastings for CKLS model with special proposal distribution. 71
4. Random walk Metropolis-Hastings for CEV-CKLS model with special proposal distribution. 143

List of Tables

3.1. Parameter restrictions for models nested within the CKLS model. 61
3.2. Interpretation of Bayes factors. 67
3.3. Descriptive statistics of the interest rate time series 76
3.4. CKLS model parameter estimates for the ten year observation window 79
3.5. CKLS model parameter estimates for the 25 year observation window 80
3.6. Comparison of parameter estimates to results from literature . . 81
3.7. Likelihood ratio and Bayes factor test statistics for the ten year observation window 85
3.8. Likelihood ratio and Bayes factor test statistics for the 25 year observation window 86
3.9. Comparison of model test statistics to results from literature . . 87
3.10. Posterior probabilities of the restricted models. 91
3.11. Model prediction test statistics 93

4.1. European call option pricing results 125
4.2. European traffic light option pricing results 126

5.1. Simplified balance sheet 132
5.2. Maximum likelihood estimates for the base and extended model . 141
5.3. Risk-neutral values of the European contract and its embedded options with parameters based on maximum likelihood estimates. 145
5.4. Descriptive statistics of the loss resulting from static reserving . . 147
5.5. Comparison of effects of uncertainty in different parameters upon losses when reserving statically 148
5.6. Comparison of effects of uncertainty for different time series and guaranteed interest rates when reserving statically 149
5.7. Descriptive statistics of the loss resulting from dynamic hedging . 152
5.8. Comparison of effects of uncertainty in different parameters upon losses when hedging dynamically 154
5.9. Comparison of effects of uncertainty for different time series and guaranteed interest rates when hedging dynamically 155

A.1. P-values of parameter estimates for the ten year observation window 166

A.2. P-values of parameter estimates for the 25 year observation window 167
A.3. Descriptive statistics for MCMC chains for the ten year observation window 168
A.4. Descriptive statistics for MCMC chains for the 25 year observation window 169
A.5. Convergence diagnostics for MCMC chains for the ten year observation window 170
A.6. Convergence diagnostics for MCMC chains for the 25 year observation window 171

1. Introduction

1.1. Motivation and problem formulation

The measurement and management of financial risks has a long history. As McNeil, Frey, and Embrechts (2005) note, some authors from the popular literature have interpreted part of the Code of Hammurabi from Babylon of 1750BC as the first evidence of the use of options for insurance purposes. Clear references of options appear as early as 1688 in Vega's Confusión de Confusiones, taking place in Amsterdam at the end of the 17th century. Some date insurance policies also to the Code of Hammurabi. Clearly they were quite established by 1762, when the first insurance company *"Society for Equitable Assurances on Lives and Survivorship"* was founded in London. Although these developments show the desire of controlling and managing financial risk, the development of probabilistic models to measure and manage financial risk has seen its most dramatic growth during the last four decades.

Before the 1950s, the attractiveness of an investment opportunity was mostly judged by the expected return of that investment. Markowitz's 1952 groundbreaking publication first established that the risk of an investment must be another factor in deciding its desirability and laid the foundation for modern portfolio selection theory. This paradigm shift was soon followed by an explosive growth in risk management methodology, with the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT), including the Fundamental Theorem of Asset Pricing, standing out as major advances. During this innovative push, the famous Black-Scholes-Merton formula appeared (see Black and Scholes (1973) and Merton (1973)). After all these centuries in which financial options were used, finally their rational price could be established and a methodology for mitigating their risks was developed.

The power of these new evaluation models lead to a dramatic growth in the options and derivative markets. At the same time, valuation and risk management started to rely more and more on mathematical models: RiskMetrics was introduced in 1989 by JP Morgan and sought to measure the Value-at-Risk (VaR) of the daily risk exposure of the bank to market risk. On the credit risk side, JP Morgan introduced CreditMetrics eight years later to analytically model the VaR due to credit risk in bond portfolios. During the mid 1990s,

many banks also introduced more sophisticated credit scoring models. The goal of these models was to better measure and differentiate between the credit risks of loans to retail and enterprise clients. These developments were strengthened by regulation introduced during the late 1990s. The regulatory framework Basel II[1] requires banks to use models to compute their exposure to credit and market risks. These computations can either be performed with a simple, regulator-prescribed model, or a more sophisticated, regulator-approved, internally developed model, which estimates these risks more precisely. These internally developed models give banks the opportunity to arrive at lower risk estimates, which encouraged them to adapt sophisticated models for their credit risk portfolios to reduce capital requirements. Due to these incentives, virtually all large banks have adapted internal statistical models to estimate risk in their credit portfolios.

The application of mathematical theory to develop quantitative risk management techniques to describe the evolution of market assets as random processes is thus a relatively new phenomenon. One of its characteristics is that its descriptions of "future states of the world", upon which any subsequent calculations are built, are driven by probabilistic models. Typically, these models are *parametric*, i.e. they specify a certain structure and the exact nature of the outcomes they predict depends on a number of parameters. To actually use one of these models for risk measurement, it is necessary to choose a specific model and then select the parameters to describe future states of the world.

This model based approach has a strong reductive nature. It seeks to describe future states of the world by a parametric model which is usually rigid in the type of distribution it can predict and only specifies this distribution by a limited number of parameters. While this reductive nature is often required to make practical predictions, its presence must make a risk manager cautious: it is clear that any of the predictions made with such a model can only be as good as the selected model, and even more so, only as good as the particular model combined with the particular parameters chosen. As Robert Merton noted in his Nobel lecture on December 9, 1997: *"At times we can lose sight of the ultimate purpose of the models when their mathematics become too interesting. The mathematics of financial models can be applied precisely, but the models are not at all precise in their application to the complex real world. Their accuracy as a useful approximation to that world varies significantly across time and place. The models should be applied in practice only tentatively, with careful assessment of their limitations in each application."* It is thus clear that by introducing parametric models to describe future states of the world, two new uncertainties are introduced into the prediction process: the uncertainty about which model adequately describes the possible states, and the uncertainty in

[1] Basel Committee (1999): *A new capital adequacy framework*

the choice of parameters that control the behavior of the chosen model.

While these concerns are not new, the financial crisis starting in 2007 has made the problem of an overly optimistic view on the output of risk assessment models obvious. In an ever stronger quest for profitability, many financial institutions took the predictions of their risk models at face value and used them to optimize their balance sheets. A large part of the blame for the financial crisis sits with the risk estimation of structured credit products. In retrospect, it is clear that the risks of these products were dramatically underestimated. A number of studies has since pointed out that the models used to estimate the risk of structured credit products were very sensitive to the input parameters, specifically to the hard-to-estimate default correlation. At least for these products, one can conclude that ignoring model and estimation risks has had dramatic consequences.

The regulators had to respond to the shortcomings of the currently used risk models and regulatory requirements. The advent of Basel III[2] (and to some degree Solvency 2) is an attempt to account for previously ignored or underestimated risks. These new regulatory frameworks account for previously ignored risks and require larger capital buffers and thus provide a larger margin for error. At the same time, it is imperative to study financial models more thoroughly for their weaknesses and behavior when their assumptions are violated. Had risk managers been more aware of the fragility of some of the risk assessment models used, perhaps some of the recent financial disasters could have been avoided.

It is thus clear that the narrow focus on a single model with a certain parametrization can have strong adverse effects in risk management. This view ignores the wisdom that *all models are wrong, but some are useful*, and confuses the necessarily imperfect assessment of a model with an accurate description of all possible future states of the world. These uncertainties, typically referred to as model risk and estimation uncertainty, have been studied in a number of contexts, for example when computing market risk, credit risk, or in option valuation. Typically, two approaches are considered: A worst case approach, where the value of the risk measure is computed for a number of different models, and a Bayesian approach, which incorporates the uncertainty in model and parameter choice into a full posterior profit and loss distribution. This distribution integrates all uncertainties and allows to compute risk measures including these uncertainties.

Nevertheless, the problems introduced by model and parameter uncertainty have only seldom been studied quantitatively in the context of interest rate

[2]Basel Committee (2011): *Basel III: A global regulatory framework for more resilient banks and banking systems*

models. This is interesting as there is a large number of interest rate models to choose from, and some of their parameters can be quite difficult to estimate. At the same time, these models are used extensively for pricing and risk management purposes.

For these reasons, this thesis studies the impact of model and parameter uncertainty in interest rate models. Specifically, I focus on one factor short-term interest rate models as they provide a relatively easy starting point, yet are widely used, and therefore relevant. The problem that this thesis seeks to explore can be summarized by the following research questions:

- *How large is estimation uncertainty in short-term interest rate models?*
- *Does the predictive power of short-term interest rate models improve when estimation uncertainty is considered?*
- *How can the impact of estimation uncertainty upon risk measures be estimated and how large is it?*

To answer these questions, I turn to the aforementioned Bayesian approach. There are multiple reasons why this approach is a good choice in this situation. The Bayesian paradigm gives an easy and coherent way of incorporating model and parameter uncertainty. As the Bayesian approach already describes initial uncertainty in parameters by a probability distribution, it also does so after updating this knowledge with insights gained through observations. This approach to modeling uncertainty extends easily to incorporate several models. The Bayesian approach summarizes the available models and the respective parameter spaces as a large hierarchical model. Another nice property of this approach is that it can easily be applied to much of the practical risk management machinery in place today. Frequently, the result of a statistical model is used in risk management to draw paths, where each path describes the evolution of the risk drivers of interest in one particular state of the world. These paths are then used to compute the value of the assets under study. The model resulting from the Bayesian approach is structurally similar, and thus paths can be drawn in the same way as before. However, these paths incorporate both parameter and model uncertainty and may thus be much more variable than the ones ignoring these additional risks. Nevertheless, the application of the usual risk management machinery remains much the same once these paths have been computed. This fact makes this approach interesting to many financial institutions which have already invested great sums of money into this machinery.

1.2. Outline and contributions

Apart from the introduction, this thesis is made up of four main chapters and two appendices. Following the introduction, Chapter 2 summarizes the most important results of the Bayesian approach. These results will be used in the following chapters. It also details computational aspects which are crucial to the application of the Bayesian approach. Chapter 3 uses these results to study the impact of estimation uncertainty in the context of short-rate models. Chapter 4 digresses slightly and describes the extension of two pricing methods, a bivariate trinomial tree and a bivariate version of the QUAD algorithm[3] in a model with stochastic interest rates. These methods are crucial for the development of the results found in Chapter 5, which studies the impact of model and parameter uncertainty on different hedging strategies for implicit options in participating life insurance contracts. Below, I provide a brief overview of the content and the main contributions of each chapter. Each chapter ends in a synopsis summarizing its main findings.

Chapter 2 introduces the Bayesian methodology used in Chapters 3 and 5 and summarizes the most important results used in the later chapters. The approach is motivated and the main reasons for its use are presented. Apart from the inherent consistency of the Bayesian approach, the motivations for its use include that it can be easily extended to the problem at hand, works with methods which are currently in use in risk management, and is based on a well developed theory. Additionally, computational advances have made the application of the Bayesian approach easier, which is one of the main reasons that it has become increasingly popular in recent years. This chapter also reviews the literature pertaining to my use of the Bayesian approach, as well as that of computational aspects such as Markov chain Monte Carlo (MCMC) methods and the important theory of Markov chains underlying this approach.

In the next part of this chapter, the Bayesian methodology is summarized, presenting the definition of a Bayesian model along with prior and posterior distributions. This section also discusses the Bayesian methods of point estimation, in particular the maximum a-posteriori estimator (MAP), which is used in Chapter 3. Additionally, Bayes factors, an objective Bayesian method, marginal distributions, and the technique of Bayesian model averaging are discussed.

Section 2.3 addresses computational aspects of the Bayesian method, which are an important concern in any application. At the heart of these computational methods are MCMC algorithms, which in turn rely on Markov chains. To understand why MCMC algorithms work, the theory of Markov chains is essential.

[3]The QUAD algorithm is a numerical pricing algorithm based on quadrature methods and was first presented by Andricopoulos et al. (2003).

For this reason, the most important results of this theory are briefly summarized. Once these results have been presented, this section presents the two most important MCMC algorithms, namely the Gibbs sampling algorithm and the Metropolis-Hastings (MH) algorithm. The delicate topic of convergence assessment is treated in the next part of Section 2.3. Assessing convergence of MCMC algorithms is much more difficult than it is for regular Monte Carlo methods, therefore these techniques are discussed in some length. The treatment first focuses on informal methods and then turns to more formal methods, which are based on adaptive non-parametric tests, tests of stationarity, or comparing the variance between chains and within chains.

Chapter 3 applies Bayesian methods to the estimation of short-rate models, specifically to the estimation of models which are nested within the Chan-Karolyi-Longstaff-Sanders (CKLS) model (Chan et al. (1992)). One motivation for this chapter are the disparate estimation results for a number of time series resulting from the studies by Chan et al. (1992), Brailsford and Maheswaran (1998), Treepongkaruna and Gray (2003), Gray (2005), and Sanford and Martin (2006). As the volatility and level-effect parameter exhibit a strong dependency, this chapter seeks to explore if the disparate estimation results in literature can be explained by estimation uncertainty in these two parameters. After exploring the previous results in literature, this chapter makes a number of methodological contributions. I apply the methods presented in Aït-Sahalia (2002) for approximating transition densities of the CKLS model and use these approximations for the maximum likelihood and Bayesian estimates. Another contribution is the reparametrization I introduce for the CKLS model, which greatly simplifies the simulation of posterior distributions with MCMC algorithms. In the traditional parametrization, the two parameters controlling the volatility term have a strong non-linear dependency and are extremely difficult to sample. The reparametrized representation eliminates this dependency and makes the construction of well-mixing Markov chains much easier. To explore how well the different models under consideration describe the observed data, I compute Bayes factors, where I combine an approximation method presented by Kass and Raftery (1995) with a Box-Cox transform (Box and Cox (1964)). These approaches were suggested by Verdinelli and Wasserman (1995), but have not seen applications in the realm of short-rate interest model estimation. This chapter also tests whether incorporating model and parameter uncertainty into estimates improves the predictive performance of the models. To this end, an out-of-sample test for testing value-at-risk models is adapted.

Once the methodology has been presented, I apply it to time series in nine different currencies, most specifically United States dollar, British pound, Euro, Swiss franc, Japanese yen, Australian dollar, Deutschmark, French franc and Italian lira. Two different time horizons, ten and 25 years, are studied, and

each time series is estimated using daily, weekly, and monthly estimates.

The estimation results are consistent with previous studies. There is only very weak evidence of mean reversion, and the estimated level-effect parameter varies widely among time series. Furthermore, while Bayesian confidence intervals and test results suggest more uncertainty in the level effect parameter than classical statistical approaches do, they cannot fully explain the different estimates produced. Therefore, I cannot conclude that the different parameter estimates seen in literature are driven by estimation uncertainty in the level effect parameter. The tests regarding the predictive ability of the model indicate that while incorporating uncertainty in the predictions improves the performance, the overall predictive performance of the model remains unsatisfactory.

Chapter 4 presents two new extensions to existing pricing methods, namely a bivariate trionmial tree for a CEV-CKLS model[4] and an extension of the QUAD algorithm (Andricopoulos et al. (2003)) for the Black-Scholes-Vasicek model. These extensions are motivated by Chapter 5 where they are used to accurately and quickly price implicit options using these models.

The first part of Chapter 4 summarizes the literature on tree methods and the QUAD algorithm. The second part describes the development of a bivariate trinomial tree for the CEV-CKLS model. There are several challenges that need to be addressed in order to construct this tree. First of all, the two processes comprising the model are heteroskedastic, which complicates the construction of a recombining tree. Additionally, depending on the parameter constellation, the two asset processes may hit the zero boundary but must not surpass it. It is thus important to construct a tree that hits this boundary exactly. Furthermore, the correlation between the two processes should be preserved when the transition probabilities for the tree are constructed. This can be difficult in certain situations.

The third section of Chapter 4 describes the extension of the QUAD algorithm. To this end, I first describe the mechanics of the regular QUAD algorithm as it was presented by Andricopoulos et al. (2003). This algorithm is then extended to the Black-Scholes-Vasicek model, which consists of a geometric Brownian motion whose drift is given by a Vasicek short-rate process. This part shows how the two-dimensional transition density must be computed in order to ensure the convergence of the method. The important topic of grid construction and which quadrature methods can be applied on non-regular grids is also covered.

The last section of Chapter 4 applies the two methods to two different options, a European call option and a so-called traffic light option. Closed-form solutions are available for both options. Rabinovitch (1989) derived a pricing formula for

[4]The constant elasticity of variance-Chan-Karolyi-Longstaff-Sanders is a generalization of the Black-Scholes-Vasicek model.

the European call under the Black-Scholes-Vasicek model and Løchte (2007) found a closed form solution for pricing the traffic light option. Both options are priced under the Black-Scholes-Vasicek model. In this setting, the two-dimensional QUAD-algorithm fares much better than the bivariate trinomial tree, outperforming the latter both in terms of accuracy and computational performance. While this gives a clear indication that the QUAD algorithm should be the preferred method in this particular setting, it does not render the bivariate trinomial tree useless. Indeed, the bivariate trinomial tree can be applied to price options in the more general CEV-CKLS model, and may fare much better in other situations, for example, when applied to American-style options.

Chapter 5 studies the impact of parameter and model uncertainty on risk measures of implicit options in participating life-insurance contracts. It is the second application of Bayesian methods to estimation and risk management aspects. The application in this chapter is motivated by the work of Zaglauer and Bauer (2008), who showed that the options in this particular contract are sensitive to stochastic interest rates. Additionally, the parameters of interest rates are hard to estimate reliably. Furthermore, there is some interest in the effects of model and parameter uncertainty in the insurance literature, as the works by Cairns (2000) and Hardy (2002) show.

The first part of Chapter 5 motivates the problem and reviews the literature with a focus on studies treating the pricing and risk management of participating life-insurance contracts. The next section presents the contract under study and shows how it can be decomposed into five different options. The third section presents the asset model, which is the CEV-CKLS model studied in Chapter 4. Next, I discuss how numerical prices can be derived using the two-dimensional QUAD model. The algorithm applied is an extension of the QUAD algorithm from Chapter 4. It needs to keep track of an additional state variable and requires a few tricks to stay on the grid in spite of dividend payments and other complications. This algorithm performs significantly better than other methods, for example the PDE approach presented by Zaglauer and Bauer (2008) and is therefore an interesting alternative method for pricing such contracts. The following section in Chapter 5 presents how parameters are estimated with and without considering parameter and model uncertainty. Finally, the chapter considers two risk management methods. The first method, called static reserving, does not try to hedge the implicit options and instead buys an investment portfolio with any capital obtained from the sale of the contract. It is assumed that any possible payout from the contract will be made from this portfolio. The second method hedges the options, but only rebalances the hedging portfolio on a monthly basis. The results from this chapter show that both approaches suffer significantly under parameter and model uncertainty. Loss

distributions under both risk management methods are significantly worsened when parameter and model uncertainty are considered.

2. A Bayesian approach to risk management

2.1. Introduction

This chapter presents the approach I take to describe parameter and model uncertainty and their effects upon risk measures. I choose a Bayesian approach as it provides a natural and consistent theory to model these additional uncertainties. In the following chapter, I discuss the motivation for choosing this approach, cover the necessary foundations, review the related literature, and describe computational methods needed for applying the theory in practice.

2.1.1. Why be Bayesian?

As described in the introduction, this dissertation seeks to better understand the implications of parameter and model uncertainty and the effect they have on risk measures. To accomplish this goal, a systematic way to measure parameter and model uncertainty and to incorporate it into risk estimates is needed. The Bayesian approach offers a natural way to consider these additional sources of uncertainty in estimates and risk measures. Based on the simple principles of sufficiency and likelihood, it provides a general framework in which additional uncertainties can be systematically incorporated and their effects on estimates can be studied. The attractiveness of this approach lies in the fact that these additional sources of uncertainty can be studied under the same framework, making this approach consistent. Comparisons to the 'standard' case, where estimation uncertainties are ignored, become easy.

The distinguishing feature of the Bayesian approach is that it describes uncertainty in the input parameters of a model with a probability distribution, the so-called *prior distribution*. Using a prior distribution on the input parameters is directly suggested by Bayes theorem: it allows the inversion of the conditionality in probability distributions, making it natural to describe the information on parameters after making an observation as the combination of prior knowledge and the likelihood of an observation given certain parameters. In a way, this way of interpreting the information on the parameters is a direct

consequence of the likelihood principle. Nevertheless, considering a probability distribution on the input parameters is a conceptual leap, as it lifts the *cause* (the parameters) on the same level as the *effects* (the realizations), both being modeled by probability distributions. However, if one regards statistics as providing an *interpretation* instead of a description of a phenomenon, this approach can be justified. In this case, randomness can be interpreted as a lack of information, or uncertainty, and using probability distributions to describe this is a natural solution.

Indeed, this description of uncertainty with a probability distribution is what makes the Bayesian approach so attractive. Describing the uncertainty in the parameters as a probability distribution directly allows this uncertainty to be incorporated in further calculations of interest. For example, to compute the probability distribution of a quantity of interest under parameter uncertainty, one can simply compute the marginal distribution of the quantity of interest, integrating over the probability distribution (the uncertainty) of the parameters.

The Bayesian approach is sometimes criticized for its use of subjective prior distributions. As researchers usually strive for objectivity, the specification of a subjective distribution resulting in the modeling process can indeed seem bothersome. However, it is important to note that the prior distribution is only one – and, if there are sufficient observations in an estimate, a small – factor which influences the results. Indeed, a statistician will always have some preconceived notion of likely outcomes and realistic parameter values, even if they are not explicitly stated. This notion will influence the decision, be it by selecting a certain parametric model, running an estimation procedure again if it yields unexpected results, or by other means. Therefore, the Bayesian approach can also be seen as forcing a degree of honesty upon the modeler, as it forces her to make prior knowledge and assumptions explicit. If one wants to express total ignorance about the parameter or model space, this can always be achieved by choosing prior distributions that represent this absence of prior knowledge.

2.1.2. Related literature

A large body of literature has been written on Bayesian statistics, starting with the early seminal works of Bayes (1764) and Laplace (1774)[1]. After these early works, the book *Theory of Probability* by Jeffreys is often considered the first modern treatment of Bayesian theory. Jeffreys (1961) made many important contributions, for example introducing the use of Bayes factors and the scale against which they are evaluated, among many others. *Theory of Probability* is

[1] See also Stigler (1986) who discusses Laplace's work.

also the topic of a recent article by Robert, Chopin, and Rousseau (2009), who discuss it in the context of modern statistical theory and point out which parts are still relevant today.

Since the work by Jeffreys, many other good books on Bayesian theory have been written. Examples include Berger (1985), Bernardo and Smith (2001), and Robert (2007). All these books base the Bayesian approach on decision theoretic foundations and justify the use of specific methods by loss functions. The most recent work by Robert gives an excellent and thorough treatment of Bayesian statistics, covering decision theoretic foundations, prior elicitation, point estimation, testing, model fit analysis, and computational aspects. The book is rigorous in its treatment and yet accessible. Apart from giving a great summary of the current state of Bayesian statistics, the author passionately defends the Bayesian approach and stresses its inherent soundness and consistency. My summary of Bayesian statistics in the first part of this chapter is mostly based on this book.

Computational aspects play a large role in Bayesian statistics, and it is no coincidence that Bayesian methods have gained in popularity as increased computational power has made previously hard problems tractable. The center piece of these computational methods are Markov chain Monte Carlo (MCMC) algorithms, which are typically used to simulate posterior distributions. Markov chains, the cornerstone of these algorithms, have themselves a deep theory which has been thoroughly explored. The book *Monte Carlo Statistical Methods* by Robert and Casella (2004) gives a rigorous and thorough review of these computational aspects and includes a chapter on Markov chains with the requirements of MCMC algorithms in sight. Meyn and Tweedie (1993) give an in-depth treatment of Markov chains, which ignores computational aspects and is not focused on MCMC, but rigorously presents the theory behind Markov chains. At the other end of the spectrum, Gamerman and Lopes (2006) give an application-oriented introduction to Bayesian statistics and MCMC. These authors cover Bayesian theory and Markov chain theory with less depth, but give valuable insights into the practical aspects of MCMC applications. As such, this book is a valuable resource for any practitioner who wants to apply MCMC algorithms to solve a problem at hand.

The use of Bayesian methods in finance and economics is widespread, especially for estimating time series models and stochastic processes and in the application of portfolio choice problems. For example, for estimating models of the short-term interest rate, Elerian, Chib, and Shephard (2001), Eraker (2001), and Jones (1998) introduced the idea of using Bayesian methods and MCMC algorithms with simulated latent data points to reduce the discretization bias when estimating parameters of continuous time stochastic models. Gray (2005) develops Gibbs samplers for the Vasicek and CKLS models. Sanford and Mar-

tin (2006) use the Bayesian methodology in a comparative study of several continuous time models of the Australian short-rate. Bayesian methods and MCMC algorithms are also popular for estimating complex stochastic processes such as stochastic volatility models. Furthermore, Bayesian methods are also important in the field of portfolio optimization. The fact that model and parameter choices heavily affect what portfolio is considered optimal, lead to the recognition that point estimates are never optimal from a decision theoretic standpoint. This point was made early on, for example by Zellner and Chetty (1965) and Brown (1978). Another reason that the use of Bayesian methodology is appealing in portfolio optimization is that prior distributions can be used to incorporate an investor's views or proprietary knowledge not reflected in market models. This fact was exploited by Black and Litterman (1991) and Pastor (2000). *Risk and Asset Allocation* by Meucci (2005) discusses both classical and Bayesian approaches to portfolio optimization extensively. Another recent book on Bayesian methods in finance has been written by Rachev et al. (2008), who give a practical view on the applications of these methods. The book discusses both the estimation of stochastic processes with Bayesian methods as well as their application to portfolio optimization.

The use of Bayesian methods to describe model and parameter uncertainty became popular in statistics in the mid-1990s. A widespread method is to use *Bayesian model averaging*, i.e. the use of the marginal distribution across models which incorporates both model and parameter uncertainty. This approach was advocated by Madigan and Raftery (1994), Draper (1995), and Raftery, Madigan, and Hoeting (1997) and has seen applications in many fields such as economics (Fernandez, Ley, and Steel (2001)), biology (Yeung, Bumgarner, and Raftery (2005)), ecology (Wintle et al. (2003)), public health (Morales et al. (2006)), and toxicology (Koop and Tole (2004)). Hoeting et al. (1999) provide an in-depth discussion of how to use Bayesian techniques to make estimates using multiple models in the face of model uncertainty. George and Clyde (2004) give an excellent review of the developments in Bayesian methodology, with a focus on model averaging and variable selection. George and Clyde advocate the use of Bayesian methods in face of model and parameter uncertainty and discuss the advantages of Bayesian model averaging over selecting a model with the highest posterior probability.

There is also a significant body of literature studying the risk introduced by parameter and model uncertainty. From a methodological standpoint, most papers can be associated with one of four groups: first, there are works that discuss the qualitative aspects of model risk, describing what factors lead to model risk and how these factors may be mitigated. Second, there are empirical studies that analyze the impact of model and parameter uncertainty using historical or simulated data. Third, there are methodological works, which fo-

cus on worst-case risk measures to describe model and estimation risk. Finally, there are papers that apply Bayesian methods to measure parameter and model risk and their impact. These methods and views are then applied to different investment products, such as financial options, credit products, and investment portfolios.

The paper by Derman (1996) discusses the risks associated with valuation models qualitatively. The author identifies a number of risks when modeling financial securities. The first risk he states is that modeling may simply not apply, i.e. the underlying dynamics may be so complicated and poorly understood that any attempt to capture them in a mathematical model would be doomed to fail. As an example, Derman states that it is possible that forecasting stock price movements is more like forecasting political occurrences than like projecting spacecraft trajectories. Second, the modeler may fail to take into account all the factors that affect the value of a security. Third, he may incorrectly assume that certain stochastic variables can be approximated as deterministic or apply incorrect dynamics to one or more of the variables. Fourth, the modeler may make incorrect assumptions about relationships between these variables. Fifth, a previously valid assumption may have been invalidated by the time the model is used because the underlying dynamics have changed over time. Sixth, some assumptions may apply in an idealized world (e.g. no trading costs), but may invalidate the model in the real world. Seventh, a model may be correct, but the data driving it may be incorrectly estimated. Even when a modeler has carefully avoided all of the above pitfalls, there are still a number of problems that may occur in practice even though the model is theoretically sound. Users of a model may not fully understand its limitations, requirements, and implications, leading to incorrect conclusions from models that are correct in principle. Furthermore, numerical methods such as numerical optimization or finite differences may be unstable or converge slowly and may thus not be sufficiently reliable. Finally, models often require intricate implementation as computer programs, which creates ample opportunities for programming errors.

The paper by Gibson et al. (1998) also discusses the qualitative aspects of model risk. Specifically, the authors give an overview over model risk issues in interest rate models. They distinguish between three main sources of uncertainty, namely uncertainty about the model structure, uncertainty about model estimates and estimation methods as well as uncertainty about the applicability of a model in a certain context.

Green and Figlewski (2002) investigate the model risk of option writing institutions empirically. They study the position of a bank who writes European call and put options and prices them using a Black-Scholes model and the best-estimate volatility available to the bank at each point in time. Their simulation study assumes that the bank writes options of 1-month, 3-month, and 1-year

maturities daily and longer running options with 2-years and 5-years maturities monthly. Furthermore, the bank writes at-the-money options with a strike equal to the current asset price and out-of-the-money options that have, under the assumption of the model, a chance of about 1/3 of turning into the money. This approach is analyzed for four different underlyings: the S&P 500 index, the three-month LIBOR interest rate, the yield on 10-year US Treasury bonds, and the Deutschmark / US dollar exchange rate. Appropriate variants of the Black-Scholes formula are used to price assets in each market. Green and Figlewski conclude that there are considerable risks involved with writing options. First of all, not hedging options at all involves large risks with standard deviations of the final profit and loss distribution often exceeding the initial option value. Cash flow matching, on the other hand, works well and essentially eliminates model risk, leaving the party pursuing this approach only exposed to credit risk (in the case that the provider of the hedge defaults and leaves the hedger exposed). However, since banks are in total short options, this is not an approach that can be followed in general. This leaves delta hedging, which also entails considerable risk exposure. The authors also discuss the possibility of marking options up, i.e. selling them above their fair price implied by the model. While this practice reduces risk, the worst losses are still large in comparison to the initial values. Furthermore, in competitive option markets, the amount of markup that a bank can demand is likely to be quite limited.

Hull and Suo (2002) also present a simulation study to understand the impact of model risk when pricing exotic options. In this study, data is generated by a complicated multi-factor model which is calibrated to market data. Then the performance of a continuously recalibrated implied volatility function model is measured. The authors argue that the models should be considered in the way that they are actually used by traders and therefore study how the model performs when it is continuously recalibrated. When applied to pricing and hedging compound and barrier options, the authors find that the implied volatility function model works well for compound but not for barrier options.

Another paper that analyzes the performance of models empirically has been written by Berkowitz and O'Brien (2002). These authors analyze the historical accuracy of Value-at-Risk (VaR) models of six large US commercial banks. The performance of the VaR models for market risks was observed between January 1998 and March 2000. While the authors find that overall the models are conservative in estimating the true VaR, at times the values can be breached substantially and such events appear in clusters. The results from the banks models are also compared against a GARCH model. While in theory the internal model should be superior, the GARCH model generally provides lower VaR estimates, yet predicts changes in volatility better. This finding calls into question how effective internally used models are, since they are build on much

more detailed information than a simple GARCH model built with historical data.

Kerkhof (2003) explores the impact of model risk upon risk measures in his dissertation. He develops a worst case risk measure which is based on computing the desired risk measure for a number of models and comparing these values to the one obtained by the chosen model. Model risk is then calculated as the difference between the highest value for the risk measure computed by the different models and the value of the risk measure obtained by the chosen model. Kerkhof empirically investigates model risk when writing plain vanilla straddle options on the S&P 500 equity index as well as on three foreign currency exchange rates (\$/£, \$/¥, £/¥). He finds that estimation risk is significant for all products and markets studied, while model risk affects options of longer maturities. The approach to model risk laid out in Kerkhof's dissertation is further explored in Kerkhof, Melenberg, and Schumacher (2010), who focus on the same examples. This paper uses a backtesting approach to examine whether estimation risk alone can explain the observed outcomes in the time series or whether model risk needs to be considered as well. The results suggest that model risk cannot be ignored.

Cont (2006) analyzes model risk in the context of derivatives pricing. He discusses two approaches of dealing with model uncertainty: a Bayesian approach and a worst case approach. Cont distinguishes the difference between *statistical uncertainty* about the objective measure $\mathbb{P}$ and *uncertainty* in the pricing measure $\mathbb{Q}$. Even if there is no uncertainty about $\mathbb{P}$, there may still be uncertainty about $\mathbb{Q}$ when the market is incomplete. As he explores model risk when pricing derivatives, he focuses on uncertainty about $\mathbb{Q}$. Alas, it is not clear how a Bayesian approach would apply in this setting, and therefore Cont uses a worst-case approach. He postulates four requirements that a measure of model risk in the context of derivatives pricing should satisfy. These requirements are as follows: first, the price of liquidly traded options is determined by the bid-ask spread and there is no model risk for such options. Second, if an instrument can be hedged in a model-free way, then this instrument should also not have model risk. If it can be partially hedged in such a way, then this partial hedge should reduce model risk. Third, model risk should be measured in monetary units to be comparable to other risks. Fourth, model risk should decrease as the set of liquid instruments becomes larger, since more instruments can be hedged model-free. Cont then proceeds to construct a coherent measure of model risk satisfying the above properties. This measure is the difference between the upper and lower price bounds of an instrument X over all pricing models $\mathbb{Q}$ over the set of models $\mathcal{Q}$. Cont shows that this model fulfills the axioms given above. The use of this measure is then illustrated for a number of models.

The impact of parameter and model uncertainty on risk measures has also been a concern when estimating credit risk. Often, Bayesian methods have been applied in this analysis. Löffler (2003) studies the impact of noisy input parameters on measures of credit risk produced by the CreditMetrics model. Using Monte Carlo simulations to estimate VaR for credit portfolios with and without considering parameter uncertainty, he finds the that losses that are estimated to occur with a probability of only 0.3% when parameter uncertainty is ignored may actually occur with a probability of 1.0%. Furthermore, he shows that parameter uncertainty can lead to biased estimates of the VaR, even if the original parameter estimates are unbiased. Similarly to Löffler, Gössl (2005) argues that estimating default probabilities and correlations within credit risky portfolios is plagued by the scarce availability of data which results in uncertain parameter estimates. He introduces a Bayesian credit portfolio approach and derives the joint posterior distribution of default probabilities and correlation parameters. Coval, Jurek, and Stafford (2009) discuss the impact of parameter uncertainty upon structured finance products in general and collateral debt obligations (CDO) in particular. The authors stress the point that the high ratings obtained by such products are only possible in the virtual absence of uncertainty in probability of default and default correlation parameters. They further illustrate that such uncertainties can have devastating effects on the estimated riskiness of structured products and argue that if these products should make a comeback, these uncertainties will need to be accounted for, for example by Bayesian methods. Heitfield (2009) takes a similar view on the importance of parameter uncertainty in computing the risk of structured finance products. He analyzes the effects with a more quantitative focus and concludes that estimation error is a large problem when estimating the riskiness of CDOs. Interestingly, the CDOs with the safest collateral may be the most difficult to estimate, as the very rare default events of the underlyings make estimating default correlation between obligors very difficult. Tarashev and Zhu (2008) study the impact of several estimation and misspecification errors on capital requirements estimated by the asymptotic single-risk factor model and find that estimation errors along with errors in modeling the tail dependence of credit losses have the strongest influence on estimated capital requirements. In a recent paper, Tarashev (2010) uses Bayesian methods to discuss the effect of parameter uncertainty upon credit risk measures. Specifically, he extends the popular asymptotic single risk factor (ASRF) model to incorporate parameter uncertainty. He finds that the correct VaR, which considers parameter uncertainty in the probability of default of each obligor as well as the asset-return correlation, is up to 27% higher than the naive VaR, which takes parameter estimates at face value.

In the insurance literature, Bayesian methods have also found application in the study of estimation uncertainty. Cairns (2000) suggested the use of the

Bayesian approach when computing risk estimates in the insurance context and criticizes that both parameter and model uncertainty are typically ignored in such calculations. Hardy (2002) illustrates an application in such a setting: she derives posterior parameter distributions and studies the impact of parameter uncertainty on risk measures for participating life insurance contracts. I discuss both papers in more depth in Chapter 5, where I use a similar methodology to study the impact of parameter and model uncertainty upon risk measures for implicit options.

2.1.3. Overview

The remainder of this chapter is structured as follows: in the following section, I present estimation and testing procedures based on the Bayesian paradigm. Section 2.3 focuses on computational aspects and challenges connected to the method. As the posterior distributions which result from the estimation methods are usually not known, Markov chain Monte Carlo (MCMC) methods are used to obtain results computationally. Markov chains lie at the heart of this method, and their most important results are summarized. In the next part, two sampling algorithms are presented, and convergence assessment methods used in later chapters are reviewed.

2.2. Bayesian statistical models and estimation

This section describes the foundations of the Bayesian methodology. First, I describe Bayes' theorem, the foundation of the following work. Later, I describe the connection between prior and posterior distributions, the maximum a posteriori point estimator (MAP), Bayes factors, and Bayesian model averaging.

2.2.1. Bayesian models

Statistical analysis is at its heart an *inversion problem*. It aims at recovering the *causes*, summarized by the parameters of a model, from the *effects*, which are embodied in the observations. When observing a random experiment governed by the parameters θ, statistical methods allow an inference about θ, while probabilistic techniques describe how future realizations are controlled by θ.

A general description of the inversion of probabilities is given by *Bayes' theorem*: If A and E are events such that $P(E) \neq 0$, then $P(A|E)$ and $P(E|A)$ are related

by

$$\begin{aligned} P(A|E) &= \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^C)P(A^C)} \\ &= \frac{P(E|A)P(A)}{P(E)}. \end{aligned} \tag{2.1}$$

In light of today's probability theory, Bayes' theorem seems trivial. However, it appears to have been a major breakthrough as it is the first inversion of probabilities. Thomas Bayes also proved a continuous version of (2.1). For two random variables X and Y with conditional distribution $f(x|y)$ and marginal distribution $g(y)$, the conditional distribution of y given x is

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}. \tag{2.2}$$

While this inversion theorem is quite natural from the point of view of modern probability theory, Bayes and Laplace went further and described the uncertainty of the parameters θ to be modeled through a probability distribution π on Θ, called the *prior distribution*. After making observations x, inference is then based on the distribution of θ, *conditional on* x, i.e $\pi(\theta|x)$. This distribution is called the *posterior distribution* and is defined by

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}. \tag{2.3}$$

The main addition brought by a Bayesian statistical model is to describe the uncertainty in the parameters with a probability distribution. Thus a *Bayesian statistical model* is made up of a *parametric statistical model*, $f(x|\theta)$ and a *prior distribution* on the parameters $\pi(\theta)$.

Bayes' Theorem updates the information on θ obtained from the observation x. It does so by lifting causes (the parameters) on the same level as the effects (the observations). Both of them are described by probability distributions, and, from a modeling standpoint, there is little difference between the two. Nevertheless, there is quite a conceptual difference between an *unknown* parameter and a *random* parameter. This is a major paradigm shift introduced by the Bayesian approach. The Bayesian approach sees this as an *interpretation* of reality, rather than a *description*. In this light, a random parameter is the interpretation of uncertainty we have about this parameter, and the update from the prior to the posterior distribution describes how this uncertainty is reduced by incorporating the information contained in an observation.

There are thus a number of distributions that are available and of interest in a Bayesian setting. As described, the *prior* distribution $\pi(\theta)$ and the *sample* distribution $f(x|\theta)$ are the basic building blocks of a Bayesian model. Furthermore, the following distributions are of interest:

1. the joint distribution of (θ, x):

$$\varphi(\theta, x) = f(x|\theta)\pi(\theta),$$

2. the marginal distribution of x

$$m(x) = \int f(x|\theta)\pi(\theta)d\theta,$$

3. the posterior distribution of θ, as obtained by Bayes' formula:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}$$

4. the *predictive distribution* of y, when $y \sim g(y|\theta, x)$, obtained by

$$g(y|x) = \int g(y|\theta, x)\pi(\theta|x)d\theta.$$

Bayesian models provide a coherent description of an inference problem. All inference and testing techniques adhere to two simple principles, the *sufficiency principle* and the *likelihood principle*. The sufficiency principle states that inference about a distribution should be based only on sufficient statistics and that two observations leading to the same statistics must lead to identical inference. The likelihood principle states that inference should only be based on the likelihood function. Both principles can be derived from fundamental axioms, see Robert (2007, p. 25). These principles give the Bayesian approach a fundamental coherence.

Example 2.1
Suppose I want to estimate the Value-at-Risk (VaR) of an investment in a stock index, say the German DAX. My initial model is simply a geometric random walk, i.e. I assume that each new value is related to the last value by

$$S_{t+1} = S_t \cdot (1 + \epsilon_t), \quad \epsilon \sim \mathcal{N}(0, \sigma^2). \tag{2.4}$$

I may have some preconceived notion of likely values of the volatility parameter σ, for example I may think that it is on average 25%, but that it could vary about this value in either direction. One way to describe this knowledge analytically is to choose a prior distribution. A convenient choice for the prior distribution of the variance σ^2 is the scaled inverse χ^2 distribution. This is the conjugate prior for the variance in the normal distribution, and this choice simplifies computations considerably. A conjugate prior to a parametertic model is a

distribution with the property that the posterior distribution belongs to the same family as the prior distribution.

I therefore formulate my prior knowledge as

$$\pi(\sigma^2) = I\chi^2(\sigma^2; \nu, \tau^2) = \frac{(\tau^2\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{\exp\{-\frac{1}{2}\sigma^{-2}\nu\tau^2\}}{(\sigma^2)^{1+\nu/2}}. \tag{2.5}$$

In the scaled inverse χ^2 distribution ν denotes the degrees of freedom and τ^2 is a scale parameter. If I deem that σ^2 should have a mean of 25% and the same standard deviation, then I would choose $\nu = 5$ and $\tau^2 = 0.00521$ to achieve that.

Equation (2.4) implies that $\epsilon_t = {}^{(S_{t+1}-S_t)}/_{S_t}$. I denote by x_t the realizations of ϵ_t, i.e. that observations $S_0, \ldots, S_n$ lead to $x_1, \ldots, x_n$, where $x_i = {}^{(S_{i+1}-S_i)}/_{S_i}$ for all i. The parametric model is thus

$$f_0(\mathbf{x}|\sigma^2) = \prod_{i=1}^{n} \phi(x_i|0, \sigma^2), \tag{2.6}$$

where $\phi(.|0, \sigma^2)$ is the normal probability density with mean 0 and variance σ^2.

Given this formulation, I can now compute the joint, marginal, posterior, and predictive distributions for this model.

The joint distribution is given by

$$\begin{aligned} \varphi(\sigma^2, x) = f(x|\sigma^2)\pi(\sigma^2) &= \prod_{i=1}^{n} \phi(x_i|0, \sigma^2) I\chi^2(\sigma^2; \nu, \tau^2) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2}\right\} \cdot \frac{(\tau^2\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{\exp\{-\frac{1}{2}\sigma^{-2}\nu\tau^2\}}{(\sigma^2)^{1+\nu/2}}, \end{aligned} \tag{2.7}$$

which is, as expected, the product of a multivariate normal and a scaled inverse χ^2 distribution.

To obtain the posterior distribution, I use a trick, namely the fact that the posterior of σ^2 must be a scaled inverse χ^2 distribution. Using this knowledge, I omit computing the integral in the denominator and simply scale the resulting

term so that it becomes a proper probability density:

$$\begin{aligned}
\pi(\sigma^2|\mathbf{x}) &= \frac{\varphi(\mathbf{x},\sigma^2)}{\int \varphi(\mathbf{x},\sigma^2)d\sigma^2} \\
&\propto \varphi(\mathbf{x},\sigma^2) \\
&\propto \sigma^{-2-\nu-n}\exp\left\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} x_i^2+\nu\tau^2)\right\} \\
&= (\sigma^2)^{-(1+\nu+n/2)}\exp\left\{-\frac{1}{2\sigma^2}(\nu+n)\frac{\sum_{i=1}^{n} x_i^2+\nu\tau^2}{\nu+n}\right\} \\
&\propto \frac{(\tilde{\tau}\frac{\nu+n}{2})^{\nu+n/2}}{\Gamma(\nu+n/2)}(\sigma^2)^{-(1+\nu+n/2)}\exp\left\{-\frac{1}{2\sigma^2}(\nu+n)\tilde{\tau}^2\right\} \\
&= I\chi^2(\sigma^2;\nu+n,\tilde{\tau}^2),
\end{aligned} \tag{2.8}$$

where $\tilde{\tau}^2 = \sum_{i=1}^{n} x_i^2+\nu\tau^2/\nu+n$. *Thus, the* σ^2*'s posterior is indeed a scaled inverse* χ^2 *distribution with* $\nu+n$ *degrees of freedom and scale parameter* $\tilde{\tau}$.

The marginal probability $m(\mathbf{x})$ *can be obtained by similar compuations, except that this time, I need to keep track of the constants and separate* σ^2*'s density to integrate it out. With similar compuations to those in (2.8) I obtain:*

$$\begin{aligned}
m(\mathbf{x}) &= \int f(\mathbf{x}|0,\sigma^2)\pi(\sigma^2)d\sigma^2 \\
&= \int \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right\} \\
&\qquad \frac{(\tau^2\nu/2)^{\nu/2}}{\Gamma(\nu/2)}\cdot(\sigma^2)^{-(1+\nu/2)}\exp\left\{-\frac{\nu\tau^2}{2\sigma^2}\right\}d\sigma^2 \\
&= \frac{1}{(2\pi)^{n/2}}\frac{\Gamma(\frac{\nu+n}{2})(\tau^2\frac{\nu}{2})^{\nu/2}}{\Gamma(\frac{\nu}{2})(\tilde{\tau}^2\frac{\nu+n}{2})^{\nu+n/2}}\cdot \\
&\qquad \underbrace{\int \frac{(\tilde{\tau}^2\frac{\nu+n}{2})^{\nu+n/2}}{\Gamma(\frac{\nu+n}{2})(\sigma^2)^{(1+\frac{\nu+n}{2})}}\exp\left\{-\frac{(\nu+n)\tilde{\tau}^2}{2\sigma^2}\right\}d\sigma^2}_{=1} \\
&= \frac{1}{(2\pi)^{n/2}}\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})}\frac{(\tau^2\nu)^{\nu/2}}{(\frac{\nu\tau^2+\sum_{i=1}^{n} x_i^2}{\nu+n}\frac{\nu+n}{2})^{\nu+n/2}} \\
&= \frac{1}{\pi^{n/2}}\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})}\frac{1}{(\nu\tau^2)^{n/2}}\frac{1}{(1+\frac{1}{\nu}\mathbf{x}^T\frac{1}{\tau^2}\mathbf{x})^{\frac{\nu+n}{2}}} \\
&= t_\nu(\mathbf{x};0,\tau^2),
\end{aligned} \tag{2.9}$$

where $\tilde{\tau}$ *is defined as before. The marginal density of* $\mathbf{x}$ *is thus the* n*-dimensional Student's* t*-distribution with* ν *degrees of freedom, mean* $\mathbf{0}$ *and covariance matrix* $\tau^2 I$.

Finally, I can compute the predictive distribution of a new realization y. Since the ϵ_i are i.i.d., the new realization only depends on σ^2, therefore I may write $g(y|\sigma^2, \mathbf{x}) = g(y|\sigma^2)$. Observing that g is the same as f in (2.9) with $n = 1$ and that $\pi(\sigma^2|\mathbf{x})$ has the same form as $\pi(\sigma^2)$, only with different parameters ν and τ^2, the computations follow immediately from those in (2.9):

$$\begin{aligned} g(y|\mathbf{x}) &= \int_0^\infty g(y|\sigma^2)\pi(\sigma^2|\mathbf{x})d\sigma^2 \\ &= \int_0^\infty \phi(y; 0, \sigma^2) I\chi^2(\sigma^2; \nu + n, \tilde{\tau}^2)d\sigma^2 \\ &= t_{\nu+n}(y; 0, \tilde{\tau}^2). \end{aligned} \tag{2.10}$$

The predictive distribution for an observation y is thus the one-dimensional Student's t-distribution with $\nu + n$ degrees of freedom and variance $\tilde{\tau}^2$. It is interesting to observe how parameter uncertainty has changed the structure of the predictive distrituion. With a more classical approach, one might have estimated σ^2 using maximum likelihood estimators and then used that estimate to predict y with a normal distribution. Considering the uncertainty in σ^2 has led to Student's t-distribution, which has fatter tails than the normal distribution, indicating higher uncertainty. However, as the number of observations and thus the degrees of freedom in the t-distribution increases, the difference between the two approaches decreases.

With this predictive distribution, I can now compute an estimate for the VaR of the stock index. Using monthly observations over 2011, I find that $\sum_{i=1}^{12} x_i^2 \approx 0.0625$ and with $\nu = 5$ and $\tau^2 = 0.00521$ I obtain $\tilde{\tau}^2 \approx 0.00521$.

$$g(y|\mathbf{x}) = t_{5+12}(y; 0, 0.00521). \tag{2.11}$$

Estimates for the 95%, 99%, and 99.9% VaR over one month are then 12.55%, 18.52%, and 26.31% of the value of the portfolio. In contrast, using the aforementioned maximum likelihood approach, I estimate $\hat{\sigma^2} \approx 0.00521$ which yields VaR of 11.86%, 16.79% and 22.30%. This shows that the Bayesian approach yields slightly larger estimates.

2.2.2. Point estimates

When a prior distribution $\pi(\theta)$ is available, the posterior distribution $\pi(\theta|x)$ can be formally derived from the observation x with distribution $f(x|\theta)$. This updated distribution is the full summary of the information available on the parameter θ. It integrates the prior information and the information contained in the observation x. The likelihood principle implies that inference on θ should rely entirely on $\pi(\theta|x)$. Although one might not really think of θ as a random

variable, it is still possible to think of $\pi(\theta|x)$ as a probability distribution. Thus summaries of $\pi(\theta|x)$ such as the mode, mean, and variance can be computed and used as point estimates.

The MAP estimator

It is often desirable to have a narrower summary, such as a point estimate, of the full posterior distribution. To derive such a summary in a decision theoretic setting, a loss criterion must be introduced. One possible estimator of θ is then the *maximum a posteriori estimator (MAP)*, which is the mode of the posterior distribution. In a discrete setting, it is associated with 0–1 loss. In a continuous setting, 0–1 loss must be replaced by a series of loss functions $L_\epsilon(d,\theta) = \mathbb{1}_{|\theta-d|>\epsilon}$, and the MAP estimator is the limit of the Bayesian estimators associated with L_ϵ as ϵ goes to 0. The MAP can also be described as a *penalized maximum likelihood*. As intuition suggests, the MAP converges to the maximum likelihood estimator when the number of observations grows to infinity. The influence of the prior distribution vanishes as more and more observations are made. The MAP has the additional advantage of being available for finite sample sizes, while the maximum likelihood estimator is mainly justified on asymptotic grounds.

Example 2.2
Since an analytical representation of the posterior distribution $\pi(\sigma^2|\mathbf{x})$ was derived in Example 2.1, computing the MAP is now simple: it is the mode of the posterior distribution. For the scaled inverse χ^2-distribution, this is $\hat{\sigma}^2_{MAP} = \frac{(\nu+n)\tilde{\tau}^2}{\nu+n+2} \approx 0.00466$.

2.2.3. Bayes factors

The *Bayes factor* for two hypotheses is the ratio of the posterior probabilities multiplied by the inverse ratio of the prior probabilities:

$$B^\pi_{01} = \frac{P(\theta \in \Theta_0|x)}{P(\theta \in \Theta_1|x)} \cdot \frac{\pi(\theta \in \Theta_1|x)}{\pi(\theta \in \Theta_0|x)}. \tag{2.12}$$

The Bayes factor describes the modification of the odds of Θ_0 against Θ_1 due to the observed data x. It can be easily compared to 1, although for an exact comparison a loss function is needed. The Bayes factor depends on prior information, but is considered an *objective* Bayesian measure, since it eliminates part of the influence of the prior distribution. It can be considered the Bayesian pendant to the likelihood ratio. If π_0 and π_1 denote the prior distributions under

H_0 and H_1, respectively, then B_{01}^{π} can be written as

$$B_{01}^{\pi} = \frac{\int_{\Theta_0} f(x|\theta_0)\pi_0(\theta)d\theta}{\int_{\Theta_1} f(x|\theta_1)\pi_1(\theta)d\theta} = \frac{m_0(x)}{m_1(x)}. \tag{2.13}$$

Equation (2.13) replaces the likelihoods with marginals when compared to the classical likelihood ratio. The likelihood ratio

$$R(x) = \frac{f(x|\hat{\theta}_0)}{f(x|\hat{\theta}_1)} = \frac{\sup_{\Theta_0} f(x|\theta)}{\sup_{\Theta_1} f(x|\theta)} \tag{2.14}$$

is a peculiar case of (2.13), where π_0 and π_1 are Dirac measures at $\hat{\theta}_0$ and $\hat{\theta}_1$. These are not proper prior distributions, since they both depend on the observation x, and are therefore not available before making the observation.

Many Bayesians now consider the Bayes factor on its own ground, (see Kass and Raftery (1995), for example). Jeffreys (1961) proposed the following scale to judge the evidence for or against a hypothesis H_0 brought by the observation:

- For $0 \leq \log_{10}(B_{10}^{\pi}) \leq 0.5$ the evidence against H_0 is *poor*,
- for $0.5 \leq \log_{10}(B_{10}^{\pi}) \leq 1.0$ it is *substantial*,
- for $1 \leq \log_{10}(B_{10}^{\pi}) \leq 2$ it is *strong*,
- for $\log_{10}(B_{10}^{\pi}) > 2$ it is *decisive*.

This scale is indicative of the strength of the evidence for or against H_0, but the choice of the levels is somewhat arbitrary. This is a consequence of the lack of decision-theoretic derivation via a loss function. However, this criticism applies equally to the choice of conventional significance levels such as 0.05 or 0.01, which are common in classical statistics. Depending on the model setting, Bayes factors can be challenging to compute. I revisit this topic in Chapter 3, where I compute Bayes factors for several models of the short-term interest rate.

Example 2.3
In the initial example, I assumed that the returns of the stock index were zero on average. That was a simple assumption to get started, and I would now like to relax this assumption. Fortunately, incorporating non-zero returns requires only a small change to the previous formulation.

In the second model, I assume that the returns are i.i.d. normally distributed with mean μ and variance σ^2, i.e.

$$S_{t+1} = S_t \cdot (1 + \epsilon_t), \quad \epsilon \sim \mathcal{N}(\mu, \sigma^2). \tag{2.15}$$

The parameteric model is then

$$f_1(\mathbf{x}|\sigma^2) = \prod_{i=1}^{n} \phi(x_i|\mu, \sigma^2). \tag{2.16}$$

I also need a prior distribution for both μ *and* σ^2*. The conjugate prior to the parameteric model* f_1 *is the normal inverse* χ^2 *distribution, i.e.*

$$\pi(\mu, \sigma^2) \sim NI\chi^2(\mu_0, \kappa_0, \nu_0, \tau_0^2). \tag{2.17}$$

Here, ν_0 *and* τ_0^2 *are the same hyperparameters as before.* μ_0 *is the mean of* μ *and* σ^2/κ_0 *is* μ*'s variance. Therefore,* κ_0 *can be interpreted as the confidence in the prior knowledge of* μ*.*

To compute the Bayes Factor for the models f_0 *and* f_1*, one can divide the marginal likelihood of the two models as shown in equation (2.13). Computing the marginal* $m_1(\mathbf{x})$ *is quite similar to the computations in equation (2.9), which yields*

$$\begin{aligned} m_1(\mathbf{x}) &= \int\int f_1(\mathbf{x}|\mu, \sigma^2)\pi(\mu, \sigma^2)d\mu d\sigma^2 \\ &= \frac{1}{\pi^{n/2}} \frac{\Gamma(\nu_n/2)}{\Gamma(\nu_0/2)} \sqrt{\frac{\kappa_0}{\kappa_n}} \frac{(\nu_0\tau_0^2)^{\nu_0/2}}{(\nu_n\tau_n^2)^{\nu_n/2}}, \end{aligned} \tag{2.18}$$

where $\kappa_n = \kappa_0 + n$*,* $\nu_n = \nu_0 + n$*, and* $\tau_n^2 = \frac{1}{\kappa_n}(v_0\tau_0^2 + \sum_{i=1}^{n} x_i^2 - \frac{(n\bar{x}+\kappa_0\mu_0)^2}{\kappa_n} + \kappa_0\mu_0^2)$*.*

To compute the Bayes factor between models 0 *and* 1*, I still need to specify the hyperparameters of model* 1*. Since I set the average return to* 0 *in the first model, I will use this as the mean* μ_0*.* κ_0 *is a measure for the confidence in the prior estimate* μ_0*. Suppose I am not very confident in this value and set it to* 1*.*

Using these values and the observations from Example 2.2, I obtain:

$$\begin{aligned} m_0(\mathbf{x}) &= 1071472 \\ m_1(\mathbf{x}) &= 34159.96 \\ B_{01}^{\pi}(\mathbf{x}) &= 31.36 \\ \log(B_{01}^{\pi}(\mathbf{x})) &= 3.445 \end{aligned} \tag{2.19}$$

The scale of evidence indicates that there is substantive evidence in favor of model 0 *and against model* 1*. This result may seem surprising, given that model* 1 *has more parameters than model* 0*. However, contrary to maximum likelihood methods, the Bayes factor takes each additional parameter into account and automatically penalizes models with additional parameters. Techniques such as the Akaike information criterion or the Bayes information criterion are used to achieve a similar effect with maximum likelihood estimators.*

2.2.4. Posterior marginal distributions and Bayesian model averaging

Bayesian model averaging is the technique of using the posterior marginal distribution of a number of models as the result of the estimation process. Any predictions made with the models are hence made with this posterior marginal distribution. To make this notion precise, suppose that there are q models, $\mathcal{M}_1, \ldots, \mathcal{M}_q$,[2] where each model $\mathcal{M}_k$ consists of a family of distributions parametrized by θ_k: $\{P(Y|\theta_k, \mathcal{M}_k)\}$. I denote by $\pi(\mathcal{M}_k)$ the prior distribution over the model space. The prior distributions on the parameters θ_k now depend on the chosen model:[3]

$$\theta_k|\mathcal{M}_k \sim \pi(\theta_k|\mathcal{M}_k).$$

This prior formulation induces a joint distribution including data, parameters, and models:

$$P(Y, \theta_k, \mathcal{M}_k) = P(Y|\theta_k, \mathcal{M}_k)\pi(\theta_k|\mathcal{M}_k)\pi(\mathcal{M}_k).$$

The definition of these prior distributions creates a large hierarchical mixture model embedding the separate models $\mathcal{M}_1, \ldots, \mathcal{M}_q$. This model describes the generation of Y in three stages: First, a model $\mathcal{M}_k$ is chosen, second, the parameter vector θ_k is drawn according to $\pi(\theta_k|\mathcal{M}_k)$, and finally the data is generated from $P(Y|\theta_k, \mathcal{M}_k)$. The joint distribution $P(Y, \theta_k, \mathcal{M}_k)$ can be conditioned and marginalized to generate desired summaries.

For example, by integrating out the parameters θ_k

$$P(Y|\mathcal{M}_k) = \int_{\Theta_k} P(Y|\theta_k, \mathcal{M}_k)\pi(\theta_k|\mathcal{M}_k)d\theta_k, \tag{2.20}$$

I obtain the marginal probability distribution of Y given model $\mathcal{M}_k$. An application of Bayes' theorem yields the posterior probability of model $\mathcal{M}_k$ given data Y:

$$P(\mathcal{M}_k|Y) = \frac{P(Y|\mathcal{M}_k)\pi(\mathcal{M}_k)}{\sum_{k=1}^{q} P(Y|\mathcal{M}_k)\pi(\mathcal{M}_k)}. \tag{2.21}$$

Under the full three-stage hierarchical model, $P(\mathcal{M}_k|Y)$ can be interpreted as the probability that $\mathcal{M}_k$ generated the observed data Y. This probability can be used for Bayesian model selection, for example, by choosing the mode of the posterior model distribution, i.e. choosing the model with the highest posterior probability. An alternative is Bayesian model averaging, where both the available models and their parameters are marginalized to obtain a distribution that

[2] I focus on a finite number of models here for illustrative purposes - the method works just as well for a countable or even uncountable infinite number of models.

[3] Note that the parameters can have a different form per model.

fully incorporates all posterior probabilities. For example, suppose that I would like to make a prediction about X given the observed data Y. The conditional distribution of X given Y using Bayesian model averaging is then:

$$\begin{aligned} P(X|Y) &= \sum_{k=1}^{q} P(Y|\mathcal{M}_k, Y)P(\mathcal{M}_k|Y) \\ &= \sum_{k=1}^{q} \int_{\Theta_k} P(Y|\theta_k, \mathcal{M}_k)P(\theta_k|\mathcal{M}_k, Y)d\theta_k P(\mathcal{M}_k|Y). \end{aligned}$$

By marginalizing models and parameters, $P(X|Y)$ incorporates both model and parameter uncertainty. The *Bayesian model average* is thus the marginal probability distribution of the quantity of interest when the models and parameters are integrated out. This probability distribution is a natural way of including both parameter and model uncertainty in the target distribution. This approach extends easily to an infinite number of models.[4]

Example 2.4
To compute the predictive distribution of future returns under both models, I first need to specify a prior distribution over both models. To specify a prior distribution, I assume that I don't have any prior knowledge which model is correct, and thus assign each model a prior probability of 0.5*. Given that there are two models, I can compute the posterior probability via equation (2.21) as*

$$\begin{aligned} P(\mathcal{M}_k|Y) &= \frac{P(Y|\mathcal{M}_k)\pi(\mathcal{M}_k)}{P(Y|\mathcal{M}_0)\pi(\mathcal{M}_0) + P(Y|\mathcal{M}_1)\pi(\mathcal{M}_1)} \\ &= \frac{P(Y|\mathcal{M}_k)}{P(Y|\mathcal{M}_0) + P(Y|\mathcal{M}_1)} \\ &= \frac{1}{1 + BF_k}, \end{aligned} \tag{2.22}$$

where $BF_0 = {}^{P(Y|\mathcal{M}_1)}/_{P(Y|\mathcal{M}_0)} = BF_{10}^{\pi}$ *and* $BF_1 = BF_{01}^{\pi} = {}^{1}/_{BF_0}$*. Therefore, given the Bayes factor of* $BF_1 \approx 31.3663$ *from Example 2.3, I obtain the following posterior model probabilities:*

$$\begin{aligned} P(\mathcal{M}_0|Y) &\approx \frac{1}{1 + {}^{1}/_{31.3663}} \approx 0.9691, \\ P(\mathcal{M}_1|Y) &\approx \frac{1}{1 + 31.3663} \approx 0.0309. \end{aligned} \tag{2.23}$$

As the Bayes factor already indicated, there is little support for model $\mathcal{M}_1$*.*

To compute the full hierarchical model, I still need to compute the posterior

[4] From a modeling perspective, the distinction between parameters and models is somewhat arbitrary since the model index can be seen as another parameter. Nevertheless, this distinction is benefital for practical reasons, as the models may be known under different names and may be already established.

distribution and the predictive distribution produced by model 1. *As the normal inverse* χ^2 *distribution is the conjugate prior to the normal distribution with unknown mean* μ *and variance* σ^2*, the posterior distribution over* μ *and* σ^2 *is again a normal inverse* χ^2 *distribution. Similarly to the computations in (2.8) I obtain*

$$\begin{aligned}\pi(\mu,\sigma^2|\mathbf{x}) &= \frac{\phi(\mathbf{x},\mu,\sigma^2)}{\iint \phi(\mathbf{x},\mu,\sigma^2)d\mu d\sigma^2} \\ &\propto \iint f_1(\mathbf{x}|\mu,\sigma^2)\phi(\mu;\mu_0,\sigma^2/\kappa_0)\chi^{-2}(\sigma^2;\nu_0,\tau_0^2)d\mu d\sigma^2 \\ &\propto N\chi^{-2}(\mu,\sigma^2;\mu_n,\kappa_n,\nu_n,\tau_n^2),\end{aligned} \tag{2.24}$$

where $\mu_n = \frac{n\bar{x}-\kappa_0\mu_0}{\kappa_n}$*,* $\kappa_n = \kappa_0 + n$*,* $\nu_n = \nu_0 + n$*, and* $\tau_n^2 = \frac{1}{\kappa_n}(v_0\tau_0^2 + \sum_{i=1}^n x_i^2 - \kappa_n\mu_n^2 + \kappa_0\mu_0^2)$*.*

For the predictive distribution, I observe that $g(y|\mathbf{x},\mu,\sigma^2) = g(y|\mu,\sigma^2)$*, then computations similar to those in (2.10) and (2.18) yield*

$$\begin{aligned}g_1(y|\mathbf{x}) &= \iint g(y|\mu,\sigma^2)\pi(\mu,\sigma^2|\mathbf{x})d\mu d\sigma^2 \\ &= \iint \phi(y;\mu,\sigma^2)\phi(\mu;\mu_n,\sigma^2/\kappa_n)\chi^{-2}(\sigma^2;\nu_n,\tau_n^2)d\mu d\sigma^2 \\ &= t_{\nu_n}(y;\mu_n,\frac{1+\kappa_n}{\kappa_n}\tau_n^2).\end{aligned} \tag{2.25}$$

The full hierarchical model thus takes the form

$$\begin{aligned}g(y|\mathbf{x}) &= 0.9691\cdot t_{\nu_n}(y;0,\tau_n^2) + 0.0309\cdot t_{\nu_n}(y;\mu_n,\frac{1+\kappa_n}{\kappa_n}\tau_n^2) \\ &= 0.9691\cdot t_{17}(y;0,0.00521) + 0.0309\cdot t_{17}(y;-0.0105,0.00723)\end{aligned} \tag{2.26}$$

Again, I compute the 95%*,* 99%*, and* 99.9% *VaR of this distribution which gives values of* 12.67%*,* 18.71%*, and* 26.63%*. These values are marginally larger than those seen in Example 2.1. Model uncertainty does increase these values slightly. However, since the impact of the second model is small, so is its effect on the results.*

2.3. Markov chain Monte Carlo

One major problem for Bayesian statistics is that it is often challenging to compute results. Contrary to classical statistics, many problem descriptions do not give rise to analytical or quasi-analytical solutions. Therefore, computational techniques to approximate the desired calculations are of major importance.

Markov chain Monte Carlo (MCMC) is a very popular technique, which allows simulating posterior distributions.

As the name implies, Markov chain Monte Carlo methods rely on simulating a distribution by drawing from a Markov chain. The objective is to construct a Markov chain whose equilibrium distribution is the desired distribution one wants to sample from. This method relies on the fact that many statistical properties established for independent samples also hold for samples that are the path of Markov chains, provided that these chains fulfill certain conditions.

Given that Markov chains are the centerpiece of this methodology, I review the most important definitions and results for Markov chains in the following section. These results are the foundations of the computations presented in subsequent chapters.

Once these results are established, I turn my attention to sampling methods. The theoretical results hold for Markov chains that are at their equilibrium distribution, therefore a major problem is how to create Markov chains that have a given target distribution as their equilibrium distribution. In the subsequent section, I discuss two popular methods that allow this, the Gibbs sampler and the Metropolis-Hastings algorithm.

Finally, for any of the desired properties to hold, it is important that the Markov chain fulfills a number of convergence criteria. As it is often not easy to detect the convergence of a given Markov chain, a number of methods that detect convergence, or the absence thereof, are covered towards the end of this chapter.

This section is based on the books by Meyn and Tweedie (1993), Robert and Casella (2004), and Gamerman and Lopes (2006). The first work is the most rigorous and provides an excellent exhibition of the theory on discrete and general state-space Markov chains. Robert and Casella (2004) gives a less rigorous but broader review, which also covers the theory of sampling algorithms such as the Gibbs sampler and the Metropolis-Hastings algorithm. Finally, Gamerman and Lopes (2006) is the most accessible text, which gives an introduction more oriented towards application rather than theory.

2.3.1. Markov chains

A Markov chain is a special kind of stochastic process over a discrete index set T, usually thought of as time. For each index $n \in T$, there is a random variable, taking values in a state space S. Markov chains may be defined over countable or general state spaces S.

Markov chains considered in later chapters generally have $\mathbb{R}^d$ (or a subset thereof) as their state space. Thus, these Markov chains clearly have non-countable state spaces. Most of this section will nevertheless focus on Markov chains with countable state spaces. The reason for this choice is that the definitions and results are much more intuitive for chains with countable state spaces. As this chapter's focus is on intuition, rather than mathematical rigor, a presentation of the results for Markov chains with countable state spaces seems more appropriate. One nice property of Markov chains is that the results available for countable state spaces carry over without major problems to more general state spaces. Meyn and Tweedie (1993, p. 66) state that *"The countable space approach guides the development of the theory we shall present in this book for a much broader class of Markov chains, on quite general state spaces: it is one of the more remarkable features of this seemingly sweeping generalization that the great majority of the countable state space results carry over virtually unchanged, without assuming any details structure on the space."* Therefore, much of this section focuses on Markov chains with discrete state spaces and will only mention how the results presented here transfer to chains with continuous state spaces. Meyn and Tweedie (1993) give a thorough treatment of Markov chains with both discrete and continuous state spaces.

A *Markov chain* is thus a stochastic process $\{X_n, n \in T\}$, where each X_n is a random variable on a state space S over an index set T. Both S and T are assumed to be countable. The defining property of a Markov chain is that it has no memory, or, given the present state of the chain, the past and future are independent:

$$P(X_{n+1} \in A | X_n = x_n, \ldots, X_0 = x_0) = P(X_{n+1} \in A | X_n \in x_n). \tag{2.27}$$

This property is also called the *Markov property*. When P and A do not depend on n, the chain is called *homogeneous*. The remainder of this chapter will focus on such chains. For homogeneous chains I denote the transition probability from x to $\{y\}$ by $P(x, y) := P(X_{n+1} \in \{y\} \mid X_n = x)$ for all $n \in T$.

Many interesting properties of Markov chains require the consideration of multistep transition probabilities. Given equation (2.27) the probability to move from state x to state y in two steps must be the sum of the probabilities of all paths that move from x to some state z and then from z to y, i.e.

$$P^2(x, y) = \sum_{z \in S} P(x, z) P(z, y). \tag{2.28}$$

Inductively, one can easily obtain the n-step transition probability from x to y:

$$P^n(x, y) = \sum_{z_1 \in S} \cdots \sum_{z_{n-1} \in S} P(x, z_1) \cdot \ldots \cdot P(z_{n-1}, y). \tag{2.29}$$

This representation of the n-step transition probability gives directly rise to the *Chapman-Kolmogorov equations*:

$$P^{n+m}(x,y) = \sum_{z \in S} P^n(x,z) P^m(z,y). \tag{2.30}$$

Decomposing the state space

To obtain further results for Markov chains, for example to arrive at stationary and limiting distributions, I need to decompose the state space. In particular, I would like to know which states are visited repeatedly by the chain and which states are only visited a few times. To this end, I need to define a few additional terms, such as hitting times, hitting probabilities, and the number of visits. For each element of the state space $y \in S$ a *hitting time* T_y can be defined as:

$$T_y = \min\{n \in T : X_n = y\}, \tag{2.31}$$

which is the first index n where the chain visits y. The *hitting probability*, i.e. the probability that the chain at state x will hit state y at any point in the future is given by

$$\rho_{xy} = P_x(T_y < \infty). \tag{2.32}$$

The total *number of visits* to state y is

$$N(y) = \#\{n > 0 : X_n = y\} = \sum_{n=1}^{\infty} \mathbb{1}(X_n = y). \tag{2.33}$$

For the first hitting time and the expected number of visits it can be shown that $\mathbb{E}(T_y | X_0 = x) = \sum_n^{\infty} P_x(T_y > n)$ and that $\mathbb{E}(N(y)|X_0 = x) = \sum_{n=1}^{\infty} P^n(x,y)$.

Given these definitions, I can now define what I mean by states that are visited repeatedly by the chain. A state $x \in S$ is said to be *recurrent* if a chain starting at x will return to x almost surely, that is $\rho_{xx} = P_x(T_x < \infty) = 1$. If $\rho_{xx} < 1$ the state is said to be *transient*. A state may be recurrent, yet the expected hitting time can be infinite. In this case, the state is said to be *null recurrent*, if the expected hitting time is finite, it is said to be *positive recurrent*.

If a state y is transient, then the number of visits to y is finite with probability 1. Furthermore, $\mathbb{E}(N(y)|X_0 = x) = \frac{\rho_{xy}}{1-\rho_{yy}} < \infty$ for any x. On the other hand, if y is recurrent then $P_y(N(y) = \infty) = 1$ and $\mathbb{E}(N(y)|X_0 = x) = \infty$.

I will now consider how recurrent and transient states make up the state space S. For two states $x, y \in S$ I say that x *hits* y ($x \to y$) if the hitting probability of y starting in x is positive ($\rho_{xy} > 0$). A set $C \subset S$ is called *closed* if the probability of $\rho_{xy} = 0$ for any $x \in C$ and $y \notin C$. This definition formalizes the intuition of a closed set which the chain does not leave once it has entered it.

C is called *irreducible* if $x \to y$ for any $x, y \in C$. The notion of irreducibility means that one cannot remove any states from an irreducible set without also modifying the transition probabilities of the Markov chain. A Markov chain is *irreducible* if its state space S is irreducible. It is easy to show that $\rho_{xy} > 0$ if and only if $P^m(x, y) > 0$ for some $m \in \mathbb{N}$. With this, one can show that if $x \in S$ is recurrent and $\rho_{xy} > 0$, then y is also recurrent. In this case $y \to x$ and (since $x \to y$) one can also write $x \leftrightarrow y$. Furthermore, it can be shown that recurrence defines an equivalence class with respect to $\leftrightarrow$. If $C \subset S$ is closed, finite, and irreducible, then all states of C are recurrent. If the set of recurrent steps S_R is not empty, it can be written as a union of disjoint, closed, and irreducible sets. When a chain hits a closed, irreducible set C of recurrent states, it stays in C forever and visits all elements of C infinitely often (i.o.) with probability 1.

Stationary distributions

A stationary distribution π is a distribution over the state space S. In a sense, it is an equilibrium distribution that once attained is maintained through further transitions. In short, π is a stationary distribution for a Markov chain $\{X_n\}$ if, for all $y \in S$

$$\sum_{x \in S} \pi(x) P(x, y) = \pi(y). \tag{2.34}$$

I would like to know under what circumstances a chain $\{X_n\}$ possesses a stationary distribution and what form this distribution takes. I will denote by $N_n(y)$ the number of visits to state y in n steps and by $G_n(x, y) = \mathbb{E}_x(N_n(y))$ the average number of visits to y when starting in x after n steps. Furthermore, let $m_x = \mathbb{E}_x(T_x)$ denote the average return time to a state x, given that the chain is currently in state x.

These definitions help solidify an intuition one may have about a stationary or equilibrium distribution π. One would expect that in such a distribution, states which are visited infinitely often have a probability equal to the relative frequency with which the chain visits them. Similarly, states which are only visited a finite number of times should have probability 0 in an equilibrium distribution. This is indeed the case. It can be shown that if $y \in S$ is transient then

$$P\left(\lim_{n \to \infty} \frac{N_n(y)}{n} = 0\right) = 1, \tag{2.35}$$

i.e. the relative frequency of transient states approaches zero with probability 1 as the chain is run infinitely long. Consequently, the expected relative frequency with which a chain visits a transient state y is also zero:

$$\lim_{n \to \infty} \frac{G_n(x, y)}{n} = 0 \quad \text{for all } x \in S. \tag{2.36}$$

On the other hand, if $y \in S$ is recurrent, then

$$P\left(\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{I(T_y < \infty)}{m_y}\right) = 1 \tag{2.37}$$

and

$$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \frac{\rho_{xy}}{m_y}. \tag{2.38}$$

These results are in line with the intuition about the relative frequency of recurrent states. If $y \in S$ is positive recurrent, then the relative frequency approaches the inverse of the average return time with probability 1 and consequently this also holds for the average time spent in state x. These observations imply that for a stationary distribution π, the following two conditions hold: first, if $x \in S$ is transient or null-recurrent, then $\pi(x) = 0$. Second, if $x \in S$ is positive recurrent, then $\pi(x) = \frac{1}{m_x}$. This result, which is a restatement of equations (2.36) and (2.38), is implied by *Kac's theorem*.[5]

Finally, an irreducible Markov chain is called *positive recurrent* if and only if it has a stationary distribution π such that

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} P^k(x,y)}{n} = \lim_{n\to\infty} \frac{G_n(x,y)}{n} = \pi(y). \tag{2.39}$$

Otherwise the chain is called *null recurrent*. Null recurrent chains may still have a measure π which fulfills (2.34), but then π is not a probability distribution.

Limiting theorems

As is the case when working with i.i.d. samples, one is often interested in the properties of statistics computed on a set of such samples. In the i.i.d. case, the strong law of large numbers and the central limit theorem are two important results providing support to the use of the empirical mean. For Markov chains, analogs to these important theorems exist. Naturally, some conditions for the underlying Markov chain need to be established.

Under some circumstances, a Markov chain may have a stationary distribution, yet a limiting distribution $\lim_{n\to\infty} P^n(x,y)$ does not exist. However, the existence of a limiting distribution is crucial for the development of limiting theorems, such as the ergodic theorem and versions of the central limit theorem. Conceptually, the problem with these chains is that they are periodic, i.e. that they deterministically jump between states and thus never arrive at a limiting

[5] See Theorem 10.2.2 in Meyn and Tweedie (1993, p. 240) and Theorem 6.37 in Robert and Casella (2004, p. 224).

distribution for an arbitrary starting distribution.[6] To answer the question of when a Markov chain has a limiting distribution, one must first understand the notion of periodicity. The *period* of a state $x \in S$ is the greatest common divisor

$$d_x = \gcd\{n \geq 1 : P^n(x, x) > 0\}, \tag{2.40}$$

which is the maximum number of steps that a chain cannot return to a state x when it is currently at x. It is clear that if a chain has a positive probability of returning to state x in the next step, i.e. $P(x, x) > 0$, then $d_x = 1$. Furthermore, if x and y are recurrent states and $\rho_{xy} = \rho_{yx} = 1$ then $d_x = d_y$. This can be understood intuitively since if one can get from x to x in d_x steps and pass through y then one can also get from y to y in d_x steps, and thus both x and y must have the same period. Therefore, the states of an irreducible chain all have the same period. A state with a period $d_x = 1$ is called *aperiodic*. A positive recurrent and aperiodic state is called *ergodic*.

Aperiodicity is a necessary condition for the existence of a limiting distribution. Let $\{X_n\}$ be an irreducible and positive recurrent chain with stationary distribution π and let $||\xi_1 - \xi_2|| = \sup_{A \subset S} |\xi_1(A) - \xi_2(A)|$ be the total variation distance between two distributions ξ_1 and ξ_2. If the chain is aperiodic, then $\lim_{n\to\infty} P^n(x, y) = \pi(y)$ for all $x, y \in S$. In fact, irreducibility and ergodicity are equivalent to $\lim_{n\to\infty} ||P^n(x, \cdot) - \pi(\cdot)|| = 0$ for all $x \in S$.[7] Furthermore, if the chain is periodic with period d, then for every $x, y \in S$, there is an integer r, $0 \leq r \leq d$, such that $P^n(x, y) = 0$ unless $n = md + r$ for some $m \in \mathbb{N}$ and $\lim_{n\to\infty} P^{md+r}(x, y) = d\pi(y)$.

The preceding results lead to the *ergodic theorem*,[8] an important result for Markov chains which is the equivalent of the strong law of large numbers for i.i.d. samples. The ergodic average is the average of a real valued function t over the path of a Markov chain: $\bar{t}_n = \frac{1}{n}\sum_{i=1}^{n} t(x_i)$. The ergodic theorem states that if the chain is ergodic and $\mathbb{E}_\pi[t(X)] < \infty$ for the unique limiting distribution π, then

$$\bar{t}_n \xrightarrow{a.s.} \mathbb{E}_\pi(t(X)) \quad (n \to \infty). \tag{2.41}$$

Thus, the ergodic theorem states that the average over a function's value along the path of an ergodic Markov chain converges against the expected value of that function under the limiting distribution. This statement is similar to that of the strong law of large numbers, which states that the average of a function on i.i.d. samples converges against the expected value of the function under the distribution of those samples. The ergodic theorem is an important justification

[6] A simple example is a two-state chain which alternates between both states. This chain has the stationary distribution $(\frac{1}{2}, \frac{1}{2})$, but has no limiting distribution for any other initial distribution.

[7] See Theorem 6.50, Robert and Casella (2004, p. 234)

[8] Theorem 6.63, Robert and Casella (2004, p. 241)

for much of the subsequent calculations based on Markov chains.

Just as there is a version of the strong law of large numbers for Markov chains, there is also a pendant for the central limit theorem. To establish this, two additional definitions are needed. First, a chain is said to be *geometrically ergodic* if there is a $0 \leq \lambda < 1$ and a real, integrable function $M(x)$ such that

$$||P^n(x, \cdot) - \pi(\cdot)|| \leq M(x)\lambda^n, \ \forall x \in S. \tag{2.42}$$

If M is independent of x, the ergodicity is said to be *uniform*. Second, the *autocovariance* with lag k for a function t defined on the state space of the Markov chain is $\gamma_k = \mathrm{Cov}_\pi(t(X_n), t(X_{n+k}))$. The variance of $t(X_n)$ is then $\sigma^2 = \gamma_0$ and the autocorrelation is $\rho_k = \frac{\gamma_n}{\sigma^2}$.

Additionally, define $\frac{\tau^2}{n} = \mathrm{Var}_\pi(\bar{t}_n)$. It can then be shown that

$$\tau_n^2 = \sigma^2(1 + 2\sum_{k=1}^{n} \frac{n-k}{n}\rho_k). \tag{2.43}$$

Furthermore, if $\sum_{k=1}^{\infty} \rho_k$ converges then $\tau_n^2 \to \tau^2$ as $n \to \infty$, where

$$\tau^2 = \sigma^2(1 + 2\sum_{k=1}^{\infty} \rho_k). \tag{2.44}$$

The term $1 + 2\sum_{k=1}^{\infty} \rho_k$ is sometimes called the *inefficiency factor*, as it describes how much $\bar{t}_n$'s variance increases due to the fact that it is not computed on independent random samples. Furthermore, it can be used to obtain the effective sample size

$$n_{\text{eff}} = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho_k}, \tag{2.45}$$

that is, the size of an i.i.d. sample leading to the same variance.

The preceding definitions can be used to establish a version of the central limit theorem for Markov chains. Let $\{X_n\}$ be a positive recurrent and geometrically ergodic Markov chain with stationary distribution π, and let $\mathbb{E}_\pi[|t(X)|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$, then[9]

$$\sqrt{n}\frac{\bar{t}_n - \mathbb{E}_\pi[t(X)]}{\tau} = \sqrt{n_{\text{eff}}}\frac{\bar{t}_n - \mathbb{E}[t(X)]}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1). \tag{2.46}$$

This statement is akin to the central limit theorem and provides support for the construction of confidence intervals based on Markov chain samples.

[9] Theorem 6.64, Robert and Casella (2004, p. 243)

Reversible chains

Let $\{X_n\}$ be a time homogeneous Markov chain with transition probability $P(x, y)$ and stationary distribution π. It can be shown that the reversed chain fulfills

$$P(X_n = y|X_{n+1} = x_{n+1}, X_{n+2} = x_{n+2}, \ldots) = P(X_n = y|X_{n+1} = x_{n+1}). \tag{2.47}$$

Therefore, the reversed chain is also Markov. The transition probabilities can be computed as

$$\begin{aligned}\tilde{P}_n(x, y) &= P(X_n = y|X_{n+1} = x) \\ &= \frac{P(X_{n+1} = x|X_n = y)P(X_n = y)}{P(X_{n+1} = x)} \\ &= \frac{\pi^{(n)}(y)P(y, x)}{\pi^{(n+1)}(x)},\end{aligned} \tag{2.48}$$

which shows that they are in general not homogeneous. However, if π is a stationary distribution and thus $\pi^{(n)} = \pi^{(n+1)} = \pi$, then the reversed Markov chain is homogeneous. If additionally $P(X_{n+1}|X_{n+2} = x) = P(X_{n+1}|X_n = x)$, i.e. the distribution of X_{n+1} conditional on X_{n+2} being at x is the same as when conditioned on X_n being at x, then the chain is said to be *reversible*. This definition leads to the following equation: A Markov chain with transition probability P is said to satisfy the *detailed balance equation* if there exists a function f such that

$$P(x, y)f(y) = P(y, x)f(x). \tag{2.49}$$

If a Markov chain fulfills the detailed balance equation with a probability density π, then π is the invariant density of the chain and the chain is reversible.[10] The detailed balance equation expresses an equilibrium in the flow of the Markov chain, as it states that being in x and moving to y is equally probable as being in y and moving to x. While it is not a necessary condition for a stationary distribution to exist, it is sufficient and can often be easily verified for Markov chains.

A nice property of reversible chains is that they make it easy to verify conditions which imply that the Central Limit Theorem holds for a Markov chain: if $\{X_n\}$ is aperiodic, irreducible, and reversible with invariant distribution π, and if $0 < \tau < \infty$, then the Central Limit Theorem (2.46) applies.[11] These conditions are more restrictive than they need be, but they are often easy to verify for Markov chains in practice.

[10] Theorem 6.46, Robert and Casella (2004, p. 230)
[11] Theorem 6.65, Robert and Casella (2004, p. 244)

Markov chains with general state spaces

As previously mentioned, the results obtained for Markov chains with discrete state spaces can be mostly transferred to chains with continuous state spaces, although in some situations additional conditions must be satisfied by the chains. These conditions are mostly of technical nature to exclude pathological cases that can occur in a more general setting.

In the general case, one cannot work with transition probabilities defined for atoms, as they have zero measure in continuous sets. Instead, one must use a σ-field $\mathcal{B}(S)$ on the state space S. On this σ-field one can then define a transition kernel $P : S \times \mathcal{B}(S) \to [0, 1]$, where $P(x, \cdot)$ must be a probability measure on $\mathcal{B}(S)$ and $P(\cdot, A)$ must be measurable for all $A \in \mathcal{B}(S)$. To obtain a notation similar to that of the discrete setting, define $P(x, y) := P(x, \{\omega : X(\omega) \leq y\})$. Transition kernels can then also be defined by the transition density $p(x, y) = \frac{\partial P(x,y)}{\partial y}$, provided that such a density exists. This will generally be the case in all situations that are of interest in this thesis.

The Markov property, multi-step transition probabilities, and the Chapman-Kolmogorov equations have a similar form as in the discrete case, except that now integrals replace the sums. The marginal distribution after $n+1$ steps with density $\pi^{(n+1)}$ may be obtained from the marginal distribution of the previous step as

$$\pi^{(n+1)}(y) = \int_{-\infty}^{\infty} p(x, y)\pi^{(n)}(x)dx. \tag{2.50}$$

The stability and limiting results obtained for the discrete case transfer to the general case with minor technical complications. Hitting times and the number of visits must now be defined for sets instead of atoms. The definition of irreducibility is slightly more complicated: A chain is said to be ν-irreducible if, for a measure ν and all $A \in \mathcal{B}(S)$ with $\nu(A) > 0$, the hitting probability $P_x(T_A < \infty)$ is greater than zero for all $x \in S$. Nevertheless, irreducibility is a property of the Markov chain and not of the irreducibility measure. If a chain is ν-irreducible there exists a *maximally irreducible* measure ψ, which dominates all other measures for which the chain is irreducible.[12] A ψ-irreducible chain with a σ-finite stationary measure π is said to be *positive*. Recurrent chains that do not possess such an invariant measure are called *null recurrent*.

In some cases the notion of *Harris recurrence*, which is stronger than regular recurrence, is needed. A set A is Harris recurrent if for all start points $x \in A$ the number of visits to A is infinite with probability 1, i.e. $P_x(N_A = \infty) = 1$. Regular recurrence only requires that the expectation of the number of visits is infinite. A ψ-irreducible chain $\{X_n\}$ is Harris recurrent if every set A with

[12] Theorem 6.16, Robert and Casella (2004, p. 214)

$\psi(A) > 0$ is Harris recurrent. To define aperidocity in the continuous case, the notion of small sets is needed. A set C is *small* if there exists some m such that the m-step transition kernel from any $x \in C$ is bounded below by a non-zero measure ν. Given a small set C, the period of a chain is defined as the greatest common divisor over the set of integers m for which the m-step transition kernel is bounded below in the definition of the small set. A chain is called *ergodic* if it converges in the total variation norm to its stationary distribution, regardless of the starting distribution.

The most important stability and limiting results obtained in the discrete case also follow in the general case. Similar to the discrete case, Harris positive and aperiodic chains are *ergodic*, and their limiting distribution π is unique in the total variation sense.[13] Ergodic averages of real valued functions $t(x)$ converge almost surely to their limiting expectations, provided that both expectations exist.[14] The central limit theorem with τ^2 as in (2.44) applies to ergodic averages in the continuous case for an aperiodic, irreducible, and positive Harris recurrent chain which is geometrically ergodic, and if the real valued function t is $2 + \epsilon$ integrable with some $\epsilon > 0$.[15]

Just like in the discrete case, a continuous state-space Markov chain whose transition kernel K fulfills the detailed balance equation with a probability density π is reversible, and π is the invariant density of the chain. If the chain is also aperiodic and irreducible and if the variance fulfills $0 < \tau < \infty$, then the Central Limit Theorem applies.

2.3.2. Simulating and sampling from Markov chains

The Gibbs sampler

The Gibbs sampler was first devised by Geman and Geman (1984) in the context of image processing, where it was used to create samples of a Gibbs distribution. Gelfand and Smith (1990) first pointed out that this method could in fact be used to sample from any distribution, provided that the full conditional distributions are known and easy to sample from.

The Gibbs sampler works by subsequently sampling from the full conditional distributions. If $\pi(x)$ denotes the target distribution and $x = (x_1, \ldots, x_d)'$, then the full conditional distributions

$$\pi_i(x_i) = \pi(x_i|x_{-i}) = \pi(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) \tag{2.51}$$

[13] Theorem 6.51, Robert and Casella (2004, p. 234)
[14] Theorem 6.63, Robert and Casella (2004, p. 241)
[15] Theorem 6.67, Robert and Casella (2004, p. 245)

Algorithm 1 Gibbs sampling

1. Set $i = 1$ and choose a start value $x^{(0)} = (x_1^{(0)}, \ldots, x_d^{(0)})'$.

2. Get the next value $x^{(i)} = (x_1^{(i)}, \ldots, x_d^{(i)})'$ from $x^{(i-1)}$ by successively sampling from the full conditional distributions:

$$\begin{aligned} x_1^{(i)} &\sim \pi(x_1 | x_2^{(i-1)}, \ldots, x_d^{(i-1)}), \\ x_2^{(i)} &\sim \pi(x_2 | x_1^{(i-1)}, x_3^{(i-1)}, \ldots, x_d^{(i-1)}), \\ &\vdots \\ x_d^{(i)} &\sim \pi(x_d | x_2^{(i-1)}, \ldots, x_{d-1}^{(i-1)}). \end{aligned}$$

3. Increment the counter from i to $i+1$ and return to step 2 until convergence is reached.

must be available and one must be able to sample from them. The Gibbs sampler is a means to approximately sample from the target distribution $\pi(x)$ when in fact one can only sample from the full conditionals. Algorithm 1 is a formal description of the Gibbs sampler.

It is apparent that this sampling scheme produces a Markov chain, as each new sample x_{n+1} only depends on x_n. Furthermore, this chain is homogeneous since the full conditionals do not depend on n. The transition kernel is easily obtained as

$$p(x, y) = \prod_{i=1}^{d} \pi(y_i | y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_d). \tag{2.52}$$

Even though the update of each component is reversible, this chain is not reversible. It is possible to obtain a reversible chain by sweeping through the components in step 2 first forwards and then backwards, but this makes the chain less efficient. An alternative is to make randomized sweeps, which also produces reversible chains.

It is perhaps surprising that knowledge of the full (univariate) conditionals alone is sufficient to construct a chain that converges against the target distribution. The *Hammersley-Clifford* theorem[16] states that the full density can be recovered from the marginal densities up to a normalizing constant. This is enough to produce a sampling algorithm.

The Gibbs sampler has the desired property that it produces Markov chains with equilibrium distribution π under mild conditions. If the Markov chain defined by a Gibbs sampler is ergodic, then its stationary distribution is the

[16] Theorem 10.5, Robert and Casella (2004, p. 377)

target distribution π.[17] Furthermore, if the kernel of the Gibbs chain $\{X_t\}$ is absolutely continuous with respect to a starting measure μ and it is aperiodic, then its limiting distribution converges against the target distribution π in the total variation sense.[18]

Example 2.5
To sample from the posterior distribution $\pi(\mu, \sigma^2|\mathbf{x})$ presented in Example 2.3, the Gibbs algorithm is a possible solution. The posterior distribution is a normal inverse χ^2 distribution and its density is the product of a normal and an inverse χ^2 density:

$$\pi(\mu, \sigma^2|\mathbf{x}) = \phi(\mu; \mu_n, \sigma^2/\kappa_n)\chi^{-2}(\sigma^2; \nu_n, \tau_n^2). \tag{2.53}$$

This implies that the conditional densities, namely $\pi(\mu|\sigma^2)$ and $\pi(\sigma^2|\mu)$, are the normal and inverse χ^2 densities:

$$\begin{aligned} \pi(\mu|\sigma^2) &= \phi(\mu; \mu_n, \sigma^2/\kappa_n) \\ \pi(\sigma^2|\mu) &= \chi^{-2}(\sigma^2; \nu_n, \tau_n^2). \end{aligned} \tag{2.54}$$

To sample from the χ^{-2} distribution, I draw from the χ^2 distribution and then invert the sample. Therefore, both distributions have well-known sample algorithms, which are implemented by any major statistical software package. The Gibbs algorithm could then be used to draw a sample form the normal-inverse χ^2 distribution:

1. *Set $i = 1$ and choose $(\mu_{(0)}, \sigma^2_{(0)}) = (0.0, 0.005)$.*
2. *a) Draw $\mu_{(1)} \sim \mathcal{N}(\mu_{(0)}, \sigma^2_{(0)}) = \mathcal{N}(0.0, 0.005)$.*
   b) Draw $\frac{1}{\sigma^2_{(1)}} \sim \chi^2(\nu_n, \tau_n^2)$.
3. *Set $i = i + 1$ and go to step 2.*

In this case, only the distribution of μ depends on σ^2, not vice-versa, but this is not a problem for the Gibbs algorithm.

The Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm was first described by Metropolis et al. (1953) and then generalized by Hastings (1970). The algorithm is very general and can be used to produce samples from many target distributions. In fact, other MCMC algorithms, such as the Gibbs sampler, can be considered a special case of the Metropolis-Hastings algorithm. The main idea of the algorithm is to create a Markov chain by repeatedly sampling from a proposal

[17] Theorem 10.6 Robert and Casella (2004, p. 378)
[18] Theorem 10.10 Robert and Casella (2004, p. 380)

Algorithm 2 Metropolis-Hastings algorithm

1. Set $i = 1$ and choose a start value $x^{(0)} = (x_1^{(0)}, \ldots, x_d^{(0)})'$.
2. Move the chain to the next value:
   a) Draw $y = (y_1, \ldots, y_d)'$ from the proposal distribution with density $q(\cdot|x^{(i-1)})$.
   b) Accept y as new value of the chain with probability $\alpha(x, y) = \min\left\{\frac{\pi(y)q(x|y)}{\pi(x)q(y|x)}, 1\right\}$. In the case of acceptance set $x^{(i)} = y$, otherwise set $x^{(i)} = x^{(i-1)}$.
3. Increment the counter from i to $i+1$ and return to step 2 until convergence is reached.

distribution and then accepting or rejecting the proposed sample so that the equilibrium distribution of the chain is the desired distribution. Let $\pi(x)$ be the density of the target distribution with values $x = (x_1, \ldots, x_d)'$ and let $q(\cdot|x)$ be the density of the proposal distribution, conditioned on x. To sample from π the Metropolis-Hastings algorithm repeatedly draws from $q(\cdot|x)$ and accepts or rejects a proposed sample with an *acceptance probability* α. When a sample is accepted, the proposal distribution is conditioned on this new value in the next step. Algorithm 2 describes the procedure formally.

This algorithm works well if $q(\cdot|x)$ is easy to sample from and is either available in closed form up to a normalizing constant or if it is symmetric, i.e. $q(x|y) = q(y|x)$. The density π must be available to some extent, it suffices if $\pi(y)/q(y|x)$ is available up to a normalizing constant.

It is clear that this algorithm defines a Markov chain, as each value only depends on the previous value of the chain. However, it is not as clear that the chain converges and that its equilibrium distribution is the target distribution π. Nevertheless, convergence can be shown under conditions which are easy to satisfy in practice. Let $\mathcal{E} = \text{supp}(\pi)$ denote the support of π. A sufficient condition on the conditional density is *positivity*, i.e. that

$$q(y|x) > 0 \quad \text{for every } (x, y) \in \mathcal{E} \times \mathcal{E}. \tag{2.55}$$

This means that for any $x \in \mathcal{E}$ the support of π is a subset of the support of $q(\cdot|x)$. In this case the chain $\{X_n\}$ produced by the Metropolis-Hastings algorithm with π and q fulfills the detailed balance equation (2.49):[19] the transition kernel of the chain is

$$P(x, y) = \alpha(x, y)q(y|x) + (1 - r(x))\delta_x(y), \tag{2.56}$$

[19] Theorem 7.2, Robert and Casella (2004, p. 272)

where δ_x is the Dirac-measure in point x and

$$r(x) = \int \alpha(x,y)q(y|x)dy. \tag{2.57}$$

It follows that

$$\begin{aligned}\alpha(x,y)q(y|x)\pi(x) &= \alpha(y,x)q(x|y)\pi(y)\\ (1-r(x))\delta_x(y)\pi(x) &= (1-r(y))\delta_y(x)\pi(y)\end{aligned}$$

and therefore $\{X_n\}$ satisfies the detailed balance equation:

$$P(x,y)\pi(y) = P(y,x)\pi(x). \tag{2.58}$$

Since $\{X_n\}$ fulfills the detailed balance equation with π, π is the stationary distribution of $\{X_n\}$. The positivity of the conditional density is also sufficient to show that $\{X_n\}$ is π-irreducible and, furthermore, that the chain is Harris-recurrent.[20] This suffices for the ergodic theorem and it follows that the ergodic average of an integrable function $t \in \mathcal{L}^1(\pi)$ converges against the expected value and that the limiting distribution of the chain is π for any starting distribution μ in the total variation sense.[21] It can be shown that these properties also hold under weaker conditions, but this is not of concern here as the positivity of the conditional transition density is easy to establish in all cases considered in this thesis. Since Markov chains produced by the Metropolis-Hastings algorithm are reversible by design, if they are also aperiodic and irreducible and the variance fulfills $0 < \tau < \infty$, then the Central Limit Theorem applies. Aperiodicity and irreducibility follow from the positivity condition on the conditional proposal density, thus the only requirement is that the variance be finite.

Example 2.6
To sample from the posterior distribution $\pi(\mu, \sigma^2|\mathbf{x})$ from Example 2.3 with the Metropolis-Hastings algorithm, I first need to choose a proposal distribution. One popular choice is to use a so-called random walk Metropolis-Hastings algorithm which works by perturbing the last sample. To choose this method, I set the proposal distribution to

$$q(\mu_{(i)}, \sigma_{(i)}|\mu_{(i-1)}, \sigma_{(i-1)}) \sim \mathcal{N}((\mu_{(i-1)}, \sigma_{(i-1)})^T, \Sigma).$$

Thus the proposal distribution is a normal distribution centered around the previous value of the chain and with covariance matrix Σ. Choosing a good covariance matrix is important to get a well mixing chain, the negative inverse of the log-likelihood function's Hessian evaluated at the maximum of the likelihood function is often a good choice. For the normal inverse χ^2 distribution with

[20] Lemma 7.3, Robert and Casella (2004, p. 273)
[21] Theorem 7.4, Robert and Casella (2004, p. 274)

parameters given in Example 2.3, this matrix is given by

$$\Sigma = \begin{bmatrix} 3.94\times10^{-4} & -2.26\times10^{-13} \\ -2.26\times10^{-13} & 3.28\times10^{-6} \end{bmatrix}.$$

A first run through the Metropolis-Hastings algorithm could then take the following form:

1. *Set* $i = 1$ *and choose* $\theta_{(0)} = (\mu_{(0)}, \sigma^2_{(0)}) = (0.0, 0.005)$.
2. *a) A draw from*

$$\tilde{\theta} \sim \mathcal{N}((\mu_{(0)}, \sigma^2_{(0)})^T, \Sigma) = \mathcal{N}\left(\begin{bmatrix} 0.0 \\ 0.005 \end{bmatrix}, \begin{bmatrix} 3.94\times10^{-4} & -2.26\times10^{-13} \\ -2.26\times10^{-13} & 3.28\times10^{-6} \end{bmatrix}\right)$$

gives $\tilde{\theta} = (0.0192, 0.00683)^T$.

b) $\alpha(\theta_{(0)}, \tilde{\theta}) = \min\left\{\frac{\pi(\tilde{\theta}|\mathbf{x})q(\theta_{(0)}|\tilde{\theta})}{\pi(\theta_{(0)}|\mathbf{x})q(\tilde{\theta}|\theta_{(0)})}, 1\right\} = 0.9018$. *Drawing* $u \sim U[0,1]$ *results in* $u = 0.5391$ *and since* $u < \alpha$ *I accept* $\tilde{\theta}$ *and set* $\theta_{(1)} = \tilde{\theta} = (0.0192, 0.00683)^T$.

3. *To continue the algorithm, I would now increment* i *to* $i + 1$ *and go to step 2.*

The first two values in the chain produced by the Metropolis-Hastings algorithm are thus $(0.0, 0.005)^T$ *and* $(0.0192, 0.00683)^T$.

2.3.3. Monitoring and convergence diagnostics

One difficulty in the application of MCMC methods is to determine when to stop the chain, i.e. to know at what point one has generated enough samples to perform the desired calculations with sufficient accuracy. While for regular Monte Carlo methods the Monte Carlo variance provides an easily applicable tool, this question is much more difficult to answer for MCMC methods.

In the previous section, I have discussed the Gibbs sampler and the Metropolis-Hastings algorithm and cited the theoretical justifications which show that under weak conditions, these methods produce chains which are guaranteed to converge to a target distribution and produce the right ergodic averages. However, while this provides the necessary theoretical justification for the applicability of MCMC methods, these results are not sufficient in practice: even though one might know that a chain must converge *eventually*, this could well take much longer than one is prepared to wait. Therefore, these results must be accompanied by more practical methods which detect that a chain has converged and that it gives a sufficiently good estimate of the desired distribution.

There are several convergence criteria that are important to the user of MCMC methods. First and foremost, I am interested in the *convergence to the stationary distribution*. Clearly, this was the original motivation for using MCMC methods, and must therefore be a minimum criterion for accepting samples from a Markov chain. However, this stationarity is not sufficient, and is often not the major problem when applying MCMC algorithms: given an initial draw from the target distribution, one can often construct Markov chains which are at the stationary distribution right from the start. In practice, Markov chains often suffer from two related problems: first, the samples can exhibit *strong autocorrelation*, i.e. each sample depends strongly on previous samples. The second problem are chains that have difficulty exploring the entire support of the target distribution. Chains that exhibit this behavior are said to *mix slow* while chains that do not have this problem are said to *mix fast*. Additional convergence criteria which can diagnose problems of this kind focus on *convergence of the ergodic averages* and *convergence to i.i.d. samples*. Stationarity tests also provide some assurance that a chain has converged to its equilibrium distribution. Another difference between convergence diagnostics is whether they use a single chain or multiple chains. The use of multiple chains can be advantageous in certain situations, as these methods do not suffer as much from the *'you've only seen where you've been'* problem. However, using multiple chains reduces the efficiency of samples and may not yield the desired answers when applied naively.

This section will present a number of diagnostic methods which can diagnose problems with the previously mentioned types of convergence. The first part will focus on informal methods, which give a quick visual indication of potential problems. The second part focuses on more formal methods. A problem that all methods share is that while they can often indicate that the chain has not reached its equilibrium distribution, they can never guarantee convergence or give general criteria that allow for automatic stopping rules. Therefore, the right combination of methods is important and it must further be augmented by the judgment of the statistician.

Informal methods

The simplest, yet effective, method to diagnose convergence problems is to plot the path of a Markov chain in each dimension. Examining the paths of the Markov chain can quickly show problems. For example, a chain can exhibit high autocorrelation or have difficulties traversing the entire state space. Several authors have proposed methods that are based on the observation of the paths. For example, Gamerman and Lopes (2006) suggests to plot the paths from the beginning and the end of a region where convergence is suspected. If these

paths are noticeably different, this is an indicator against convergence. As an illustration, Figure 2.1 shows two paths from Markov chains based on Example 2.3 and described in more detail in Example 2.7.

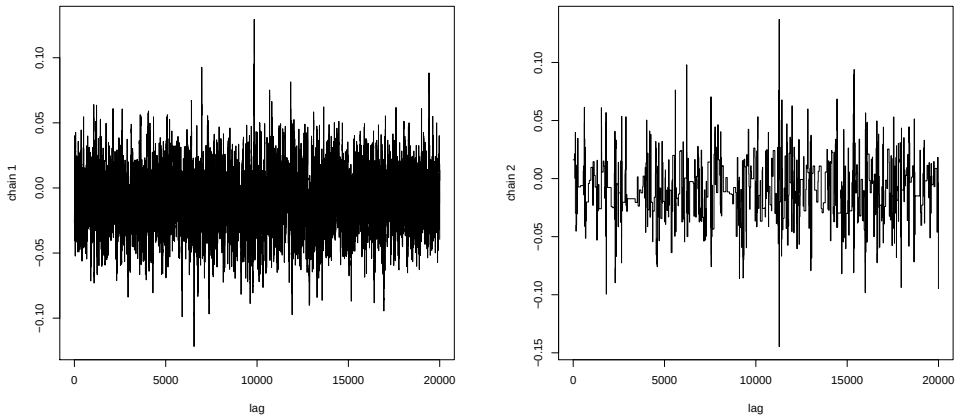


Figure 2.1.: Two Markov chains estimating the parameter μ of the second model from Example 2.3. The procedure used to obtain these chains is described in Example 2.7. The first path traverses the parameter space quickly and visits all values frequently. It is said to *mix fast*. The second chain is often stuck at a single value for many steps, it *mixes slowly*.

Closely related is the method of examining histograms of the observed chains. Again, one can compare histograms from the start and end of a chain and accept convergence if these are indistinguishable and reject otherwise. An alternative suggested by Gelfand and Smith (1990) is to run several chains in parallel and compare histograms produced by these chains. One approach is to run n chains in parallel for m steps and to plot a histogram of a function applied to the n values of the m-th state. Then these chains can be iterated for another k steps and the plotting procedure is repeated. If the histograms appear indistinguishable convergence is accepted, otherwise it is rejected.

Another simple, yet effective, method is to examine the autocorrelation function for each component of the chain. The Central Limit Theorem for Markov chains in equation (2.46) gives a motivation for this approach: autocorrelations drive the inefficiency factor and reduce the effective sample size, thereby increasing the required sample size and making the MCMC approach less efficient. Ideally, one would like the autocorrelation to be as low as possible, to approximate i.i.d. samples. Furthermore, high autocorrelation, especially when maintained over many lags, is usually an indication that the chain mixes slowly. Figure 2.2 shows autocorrelation functions for the Markov chain presented in Figure 2.1.

Example 2.7

To estimate the posterior distribution of model 1 *from Example 2.3, I construct two random walk Metropolis-Hastings Markov chains. The first chain is the one described in Example 2.6, the second uses* $100 \cdot \Sigma$ *as the covariance matrix of the proposal distribution. I then run these chains for* 25 000 *iterations and discard the first* 5000 *draws as burn-in samples. Figure 2.1 shows the path of* μ *from both of these chains while Figure 2.2 shows the autocorrelation function of the two paths.*

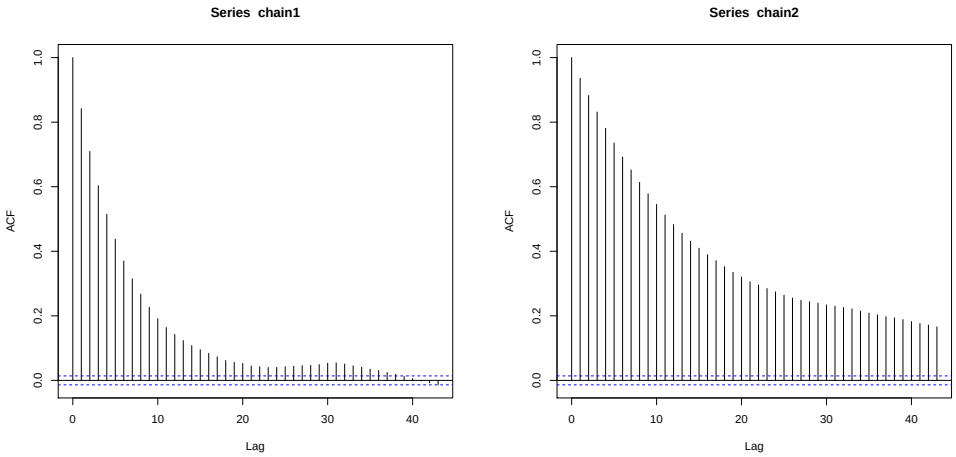


Figure 2.2.: The autocorrelation function for the two chains from Figure 2.1 and Example 2.7. The autocorrelations of the first chain drop off quickly and are close to zero from about the 20th lag on. Autocorrelations for the second chain remain fairly high even after 50 lags. This is another indication that the second chain does not mix well.

Formal methods

Several authors have proposed convergence detection methods which are more formal than the rules-of-the-thumb provided in the previous section. These methods are often designed to test some property that holds if the Markov chain has converged to its stationary distribution. Often, these methods are theoretically motivated, but lack complete rigor as they usually do not specify a completely accurate test of convergence. Furthermore, none of the tests can say with complete certainty that a Markov chain has indeed converged. Nevertheless, these methods test other properties than the previous ones and are therefore a useful addition to the convergence-detection toolbox.

Nonparametric tests

Nonparametric tests, such as the Kolmogorov-Smirnov test, can be applied to test stationarity of a single chain $\{X_n\}$. When the chain is stationary, then X_{n_1} and X_{n_2} have the same distribution for any n_1 and n_2. Given a path of the chain, one can thus split this in half and compare the two halves $\{X_1, \ldots, X_{n/2}\}$ and $\{X_{n/2+1}, \ldots, X_n\}$. Non-parametric tests are usually developed for i.i.d. samples, thus there needs to be some kind of correction for the dependence between the samples. An easy, although somewhat imprecise, way to obtain quasi independence is to thin the chain by a factor k, using only every k-th sample, such that $\{X_k, X_{2k}, \ldots\}$ are nearly independent. Then, the *Kolmogorov-Smirnov* statistic is

$$K = \frac{1}{M} \sup_h \left| \sum_{l=1}^{M} \mathbb{1}_{(-\infty,h)}(t(X_{lk})) - \sum_{l=M+1}^{2M} \mathbb{1}_{(-\infty,h)}(t(X_{lk})) \right|. \tag{2.59}$$

For one-dimensional chains, t can be the identity function, for multi-dimensional chains it could select one component of the vector. Under the assumption that $\{X_t\}$ is stationary, the limiting distribution of $\sqrt{M}K$ has the cumulative distribution function

$$\mathbb{P}(\sqrt{M}K \leq x) = 1 - \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2}. \tag{2.60}$$

A derivation of the exact statistic is quite difficult. Robert and Casella (2004) also suggest to plot $\sqrt{M}K_n$ against n and to check visually for a stable distribution around small values.

Example 2.8
To compute the KS-statistic for the chains from Example 2.7, I must first select the thinning factor k. For the first chain I select $k = 50$, as autocorrelations drop quickly and are around zero already for a few lags below 50. Starting from 20 000 samples, this leaves me with 400 observations, which I split into two chains $(X_{50}, X_{100}, \ldots, X_{10000})$ and $(X_{10050}, \ldots, X_{20000})$, each containing 200 observations. The KS-statistic for $t_\mu((\mu, \sigma^2)^T) = \mu$ is then attained for $h = -0.009774$ and is

$$\begin{aligned} K &= \frac{1}{200} \sup_h \left| \sum_{l=1}^{200} \mathbb{1}_{(-\infty,h)}(\mu_{50l}) - \sum_{l=201}^{400} \mathbb{1}_{(-\infty,h)}(\mu_{50l}) \right| \\ &= \frac{1}{200} \left| \sum_{l=1}^{200} \mathbb{1}_{(-\infty,-0.009774)}(\mu_{50l}) - \sum_{l=201}^{400} \mathbb{1}_{(-\infty,-0.009774)}(\mu_{50l}) \right| \\ &= \frac{1}{200} |106 - 87| = 0.095 \end{aligned} \tag{2.61}$$

The p-value of this statistic is

$$\mathbb{P}(\sqrt{200} \cdot K \leq 0.095) \approx 0.5. \tag{2.62}$$

Therefore, using the Kolmogorov-Smirnov test, I cannot reject the stationarity of the Markov chain.

Monitoring the convergence of averages

The Ergodic theorem states that the empirical (ergodic) average of a function t along the path of an ergodic Markov chain must converge to the expected value $\mathbb{E}_\pi[t(X)]$. Under certain conditions the central limit theorem applies and specifies the rate of convergence, although it can be much slower than in the i.i.d. case. One way to use this to monitor convergence is to consider the empirical average and check if it converges against a fixed value. As always, plotting the average $S_N = \frac{1}{N}\sum_{n=1}^{N} t(X_n)$ against N provides a quick visual check. Yu (1994) and Yu and Mykland (1998) take this one step further and propose to use cumulative sums, graphing the distances

$$D_N^i = \sum_{n=1}^{i} (t(X_n) - S_N), \tag{2.63}$$

where

$$S_N = \frac{1}{N}\sum_{n=1}^{N} t(X_n) \tag{2.64}$$

is the sample average up to N. According to Yu (1994) and Yu and Mykland (1998), if the mixing speed of the chain is high, then the graph of D_N^i is very irregular and oscillates around 0. When mixing speed is low, then the graph of D_N^i is more regular with long excursions away from 0.

Example 2.9
Figure 2.3 shows D_N^i for the two paths of μ produced by the chains from Example 2.7. As one may have expected, the second chain takes much longer excursions from zero and also reaches larger values. In comparison, it is thus clear that the first chain mixes better than the second.

Within and in-between variances

Gelman and Rubin (1992) provide a method based on the idea that the paths of multiple chains should have the same distribution after convergence. Their method tests if the dispersion between different chains is greater than the dispersion within chains. This method is based on the same idea as the comparison of

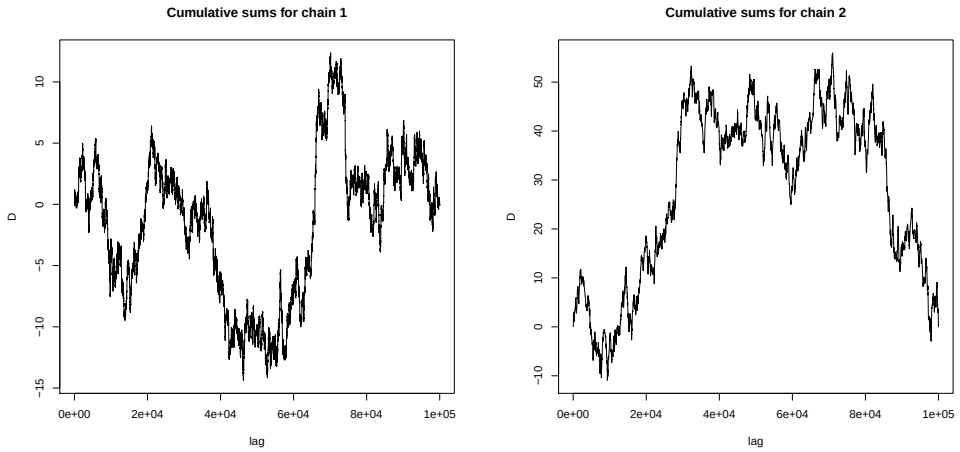


Figure 2.3.: This plot shows the evolution of the distance of the cumulative sum from the mean when the two chains from Example 2.7 are run for 100 000 iterations. The second path takes much longer excursions away from 0 which also reach higher values. This is another indication that the second chain does not mix well.

histograms of different chains and single chains. The method uses several chains started from points which are overdispersed with respect to the target distribution. For example, if the components of $\{X_t\}$ are restricted to an interval, one might start two chains at opposite end points of the interval.

Considering m parallel chains and a function $\varphi = t(X)$, there are m paths $\{\varphi_1^{(i)}, \ldots, \varphi_n^{(i)}\}$, $i = 1, \ldots, m$. The variances between chains B and within chains W are given by

$$B = \frac{n}{m-1} \sum_{i=1}^{m} (\bar{\varphi}^{(i)} - \bar{\varphi})^2, \tag{2.65}$$

$$W = \frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (\varphi_j^{(i)} - \bar{\varphi}^{(i)})^2. \tag{2.66}$$

Here, $\bar{\varphi}^{(i)}$ is the average over chain i and $\bar{\varphi}$ is the average of all averages. If the chain has converged, then all samples are drawn from the stationary distribution and the variance of φ, σ_φ^2, can be consistently estimated by W, B and $\hat{\sigma}_\varphi^2 = (1 - 1/n)W + (1/n)B$.

If the chains have not yet converged, then the variance between chains will be larger than the variance within chains, and $\hat{\sigma}_\varphi^2$ will tend to overestimate σ_φ^2, while W will tend to underestimate it. An indicator of convergence can thus be constructed by comparing the two estimates. Gelman and Rubin (1992) suggest $\hat{R} = \sqrt{\hat{\sigma}_\varphi^2 / W}$, which is always larger than 1. As $n \to \infty$ both estimators will

approach σ_φ^2 and thus $\hat{R} \to 1$. Gelman (1996) suggest accepting convergence when the value of $\hat{R}$ is less than 1.2.

Example 2.10
Returning to the chain from Example 2.7, I would like to apply the Gelman-Rubin diagnostic to see if I can accept convergence with this diagnostic. To do so, I need to create additional Markov chains from overdispersed starting points. Apart from the first chain, which starts at $(0.0, 0.005)^T$*, I choose* $(-0.3, 0.00001)^T$*,* $(0.3, 0.00001)^T$*,* $(-0.3, 0.1)^T$*, and* $(0.3, 0.1)^T$ *as starting points. Running each chain for* 20 000 *iterations, I obtain*

$$\begin{aligned} W &\approx 0.0005555 \\ B &\approx 0.0003636 \\ \hat{R} &\approx 1.0 \end{aligned} \tag{2.67}$$

The Gelman-Rubin diagnostic thus indicates that despite starting from overdispersed points, all chains have converged to the same stationary distribution.

Variance estimates and effective sample size

Even when one is reasonably certain that a Markov chain has converged and is at its stationary distribution π, there is still an important difference between the ergodic average for $h(X_t)$ as an estimator of the expected value $\mathbb{E}_\pi(h(X))$ and one that is based on regular Monte Carlo samples. While regular Monte Carlo samples are i.i.d. and the variance of the estimate can be approximated by

$$\hat{v}_N = \frac{1}{N^2} \sum_{n=1}^{N} \left(t(X_n) - S_N\right)^2, \tag{2.68}$$

this is not the case for the ergodic average obtained via MCMC. One possibility to get a better estimator for the variance is to use the *effective sample size* in lieu of N. The effective sample size $\hat{N}^S$ is given by

$$\hat{N}^S = N/\kappa(t), \tag{2.69}$$

where κ is the *autocorrelation time* associated with the sequence $\{t(X_n)\}$,

$$\kappa(t) = 1 + 2 \sum_{n=1}^{\infty} \operatorname{corr}\left(t(X_0), t(X_n)\right). \tag{2.70}$$

Replacing N with $\hat{N}^S$ in (2.68) then gives the more reliable estimate of the variance

$$\hat{v}_N = \frac{1}{N \cdot \hat{N}^S} \sum_{n=1}^{N} \left(t(X_n) - S_N\right)^2. \tag{2.71}$$

Of course, this raises the difficulty of estimating κ. One possibility is to use the spectral density at frequency zero. This method is implemented in the CODA package.[22]

Example 2.11
For the two Markov chains from Example 2.7 the effective sample size and the variance based on this effective sample size can be computed as follows. As before, I focus on $t_\mu((\mu, \sigma^2)^T) = \mu$ for illustrative purposes. For the first chain, I obtain an effective sample size of

$$\hat{N}^S = 1645.96$$

and a sample variance of

$$\hat{v}_N \approx 3.4 \times 10^{-7}.$$

By comparison, for the second chain I obtain an approximate effective sample size of

$$\hat{N}^S = 473.34$$

and thus a sample variance of

$$\hat{v}_N \approx 1.1367 \times 10^{-6}.$$

These results show that the second chain, which mixes poorly, leads to a considerably lower effective sample size and a higher sample variance.

Other methods

Beyond the methods presented here, there are a large number of other methods which have been proposed. I have focused on the previous methods, because they are applicable in the cases I will consider in the subsequent chapters, are relatively easy to implement, and they address the different convergence criteria and typical problems of MCMC methods. They should provide a good overview and allow me to detect most convergence problems. As Robert and Casella (2004) state, choosing a wide array of different methods is important to asses all possible convergence difficulties. Given that the methods presented focus on different difficulties, they should give sufficient confidence that a chain which does not show problems with respect to these indicators has indeed converged and can be used as a basis for further analysis.

[22] See Plummer et al. (2006).

2.4. Summary and discussion

This chapter has presented many important aspects of Bayesian theory and the computational methods that are required for any practical implementation. Initially, I discussed the motivation for choosing a Bayesian approach in estimation and risk management and referenced some of the literature that has previously taken this approach.

Next, the key concepts of Bayesian theory were presented, such as prior distributions and Bayes' theorem. Additionally, I have presented how point estimates such as the MAP can be computed under a Bayesian method. Bayes factors, an objective Bayesian method which measures the impact of the observed data on the posterior distribution, were then discussed. It was illustrated how Bayes factors can be used for testing and how they are related to the likelihood ratio known from classical statistics. A testing scale developed by Jeffreys (1961) was also presented. Finally, I mentioned how the Bayesian approach allows to combine several models to a large hierarchical model via prior distributions. Marginalizing the parameters for each model and applying Bayes' theorem leads to posterior model probabilities – which are closely related to Bayes factors – while marginalizing both models and their parameters leads to a full posterior distribution. This technique is known as Bayesian model averaging.

The last part of this chapter focused on computational aspects, which play an important role in Bayesian methodology. As it is often difficult to obtain analytical results for the posterior distributions obtained via Bayesian analysis, the usual approach is to apply Markov chain Monte Carlo algorithms to obtain approximations to the posterior density. Markov chains are the centerpiece of this methodology, and I have summarized the most important results as they pertain to MCMC algorithms. In particular, I reviewed the Ergodic theorem and Central Limit Theorem for Markov chains and how reversibility is an easily verified sufficient condition. Next, I described the two most important MCMC algorithms – the Gibbs sampler and the Metropolis-Hastings algorithm – and indicated how these algorithms fulfill the conditions postulated in the earlier part of this section. While theoretical convergence results are reassuring and motivate the use of MCMC algorithms, they are not sufficient in practice. Therefore, I also covered several convergence diagnostic methods. A number of informal and formal convergence diagnostics were covered and while none of the methods discussed can provide a guarantee for convergence, the full range of convergence diagnostics should provide sufficient assurance that a particular MCMC algorithm has worked. With the discussion of these methods, I have prepared a toolbox of Bayesian methods which will find its application in the following chapters.

3. Bayesian parameter estimation of short-rate models

3.1. Introduction

The goal of this chapter is to study how much parameter estimates of short-rate models are affected by estimation uncertainty and whether this uncertainty can explain some of the different estimation results in literature. In this chapter, I estimate the parameters of the Chan-Karolyi-Longstaff-Sanders (CKLS) model introduced in Chan et al. (1992). In contrast to previous studies, I use the approximation technique introduced by Aït-Sahalia (2002) to obtain very accurate approximations of the transition densities of the CKLS model. With these transition densities, I compute both maximum likelihood estimates as well as posterior parameter distributions for short-term interest rate time series in United States dollar, British pound, Euro, Swiss franc, Japanese yen, Australian dollar, Deutschmark, French franc, and Italian lira. The CKLS and a number of nested models are compared using likelihood ratio and Bayes factor tests. Furthermore, I study how estimation uncertainty improves the predictive performance of the model. As an additional contribution, I introduce an alternative parametrization of the CKLS model for the Bayesian computations. This alternative parameterization leads naturally to a Metropolis-Hastings algorithm that produces well-mixing chains, which is difficult to achieve with the classical parameterization.

3.1.1. Literature

The short-term interest rate is a key variable in any economy. As it represents the *price of money*, it affects businesses and expectations about its development result in longer-term interest rates. For central banks, it is the most important instrument for their monetary policy and an important economic indicator. Therefore, it is unsurprising that the economic and finance literature has produced a wide range of models to describe this important variable. The models can be broadly classified into *short-rate models* and *no-arbitrage term structure models*. The former describe the short-term interest rate as a stochastic process which is typically observable. As a consequence, the parameters governing such

processes are usually estimated directly from a proxy of the short-rate. In contrast, no-arbitrage term structure models describe the entire term structure of interest rates by latent processes. These models provide a relationship between risk-free bond prices and the latent factors based on no-arbitrage arguments. The parameters governing these models can be obtained by fitting the term structure produced by such a model to an observed term structure.

The advantage of no-arbitrage term structure models is that more complex models can be fitted and therefore better fits to observed term structures can be obtained. The disadvantage is that the latent factors are often difficult to interpret as they typically do not correspond to any economically meaningful variables. In comparison, short-rate models can be directly estimated from a proxy of the short-rate and are immediately interpretable. This intuitive nature makes short-rate models popular even though more complex models are available.

Given the economic importance of the short-rate, it is not surprising that the finance literature has produced a large number of models describing this process. The early work by Cox (1975) and Vasicek (1977) lead the way and was shortly followed by Dothan (1978), Brennan and Schwartz (1980), and Marsh and Rosenfeld (1983). The square root-diffusion process introduced by Cox, Ingersoll, and Ross (1985) remains one of the most popular short-rate models to this day.

The seminal study by Chan et al. (1992) generalized seven previously known short-rate models by introducing an additional exponent parameter in the volatility term. This parameter is usually called the level effect or elasticity parameter. The level effect parameter controls the heteroskedasticity of the short-rate. When estimated on 1-month United States Treasury Bill yields with monthly observations between June 1964 and December 1989, Chan et al. found that a level effect parameter close to 1.5 fit the data best. This finding lead to some controversy, as it suggests a level of heteroskedasticity much higher than what is assumed by the popular Vasicek and CIR models.

After its publication, the CKLS model has been the main focus of studies that deal with the estimation of short-rate models. For example, Nowman (1997) introduces a Gaussian estimation technique and estimates the CKLS model on British short-rates[1] with monthly observations between March 1975 and March 1995, as well as on the U.S. Treasury Bill data used by Chan et al. (1992). He finds that, in contrast to the U.S. T-Bill yields, the British short-rate volatility does not strongly depend on the level of the interest rate.

Bliss and Smith (1998) reexamine and criticize the results of Chan et al. (1992). They argue for a structural shift during October 1979 and Fall 1982, known

[1] Nowman (1997) uses one-month sterling interbank rate (middle rate).

as the Federal Reserve Experiment period. Once this regime shift is taken into account and the model is estimated separately for the periods during and outside of the shift, Bliss and Smith find evidence for an elasticity parameter around or below 1.0. In particular, the CIR model with an elasticity parameter of 0.5 and the Brennan-Schwartz model with an elasticity of 1.0 are not rejected, while a structurally similar model, which forces an elasticity parameter of 1.5, is.

Brailsford and Maheswaran (1998) estimate the CKLS model along with a time-varying volatility model and a combination of both models for the Australian short-term interest rate. For the CKLS model, the authors find estimates of the elasticity parameter above 1.0 with both the generalized method of moments and quasi maximum likelihood as estimation methods. The models are estimated on 30-day bank accepted bills data with weekly observations between July 1976 and December 1995.

Treepongkaruna and Gray (2003) provide an extensive empirical study of the CKLS model by applying it to interest rate time series in eight currencies, namely the United States dollar, British pound, Japanese yen, Deutschmark, Australian dollar, Italian lira, French franc, and the Swiss franc. Using both the generalized method of moments as in Chan et al. (1992), as well as quasi maximum likelihood estimators, they come to the conclusion that estimates of the interest rate processes are sensitive to the estimation method used and differ between different time series. The authors conclude that short-term interest rate models should be used with caution.

Aït-Sahalia has made several important contributions to the field of estimating diffusion processes. In Aït-Sahalia (1996), he develops a methodology for testing diffusion models non-parametrically and applies this to 7-day Eurodollar deposit spot rates with daily observations between June 1, 1973 and February 25, 1995. In the non-parametric tests, Aït-Sahalia rejects all but a very sophisticated model whose diffusion and volatility functions consist of four and three terms, respectively. This model allows for a non-linear drift which is the principal source of rejection of the other models. In Aït-Sahalia (2002), he develops an approximation technique to compute approximations of transition densities with high precision for processes where no closed-form solution of the transition density exists. The technique allows approximating the transition density of diffusion processes with arbitrary precision and is shown to give very accurate approximations with very few approximation terms. Approximations with even only one term are vastly more accurate than, for example, those obtained with an Euler discretization of the process. With two or three terms the remaining errors are negligible. In the companion work Aït-Sahalia (1999), the approach is applied to many stochastic processes popular in interest rate modelling, for

example the Vasicek, CIR, and CKLS processes. Aït-Sahalia (2008) further generalizes the approach to multivariate diffusions.

Chib and Greenberg (1996) were among the early adapters of Bayesian methods in general and Markov chain Monte Carlo algorithms in particular. These authors showed how to apply these techniques to problems in finance and made them accessible to a larger audience. Subsequently, MCMC algorithms became ever more popular and several authors have used these techniques and applied them to the estimation of short-rate models.

Elerian, Chib, and Shephard (2001), Eraker (2001), and Jones (1998) introduced the idea of using MCMC algorithms with simulated latent data points to reduce the discretization bias when estimating parameters of continuous time stochastic models. Elerian, Chib, and Shephard (2001) use a Metropolis-Hastings algorithm with latent data points and present in-depth information on the derivation of posterior distributions, convergence characteristics, marginal distributions, and the reduction of bias achieved by introducing latent data points. They use these approaches to estimate the Vasicek and CIR interest rate models. Eraker (2001) develops Gibbs samplers for the constant elasticity of variance model (CEV) and a stochastic volatility model. Similar to Elerian, Chib, and Shephard (2001), he uses unobserved latent interest rate data points which are integrated out during the Monte Carlo simulation. The method is applied to United States Treasury Bill data, which is better fitted by the stochastic volatility model. In contrast, Jones (1998) uses the augmentation approach to estimate the parameters of stochastic models describing the evolution of stock prices. He analyzes the differences of posterior parameter distributions from the analytically available solutions in a geometric Brownian motion model and then applies the technique to estimate the parameters of a stochastic volatility model on S&P return data. He also extends his method to jump-diffusion processes and analyzes the U.S. Dollar / Deutschmark exchange rate.

Gray (2005) develops Gibbs samplers for the Vasicek and CKLS models. For the CKLS model, these samplers rely on an Euler-discretization of the process to obtain the log-likelihood function. To reduce the bias introduced by this discretization, Gray introduces latent data points. He concludes that data augmentation is not essential for point estimates, but helps reduce the bias in posterior densities. With this methodology, he estimates the parameters of both models on 30, 90, and 180-day Australian bank accepted bill data with weekly observations between July 1976 and February 2002, as well as on yields of 13-week and 26-week Treasury notes with observations between June 1986 and February 2002. For all of these data sets, the estimated level effect parameters are consistently above 1.0, ranging from 1.2561 for the 26-week Treasury notes to 1.6606 for the 30-day Australian bank accepted bills.

Sanford and Martin (2006) use the Bayesian methodology in a comparative study of several continuous time models of the Australian short-rate. Specifically, they take the CKLS model and restrict the level effect parameter to the four values $0.0, 0.5, 1.0$, and 1.5. Using an Euler discretization with data augmentation to reduce discretization bias, they derive posterior distributions of the remaining parameters with a MCMC algorithm. Both Bayes factors and posterior probabilities for these models are then computed. For the Australian 90-day bank acceptance bill rate with weekly observations between January 1990 and July 2000, the posterior probability of the model with the level parameter at 0.5 is the largest. The Bayes factors indicate little or no support for the other models.

3.1.2. Motivation and contribution

The varying estimation results for the level effect parameter found in literature call into question if a single short-rate model with fixed level effect parameter can adequately describe the general dynamics of short-rates. The study by Chan et al. (1992) suggested that a more complex model such as the CKLS model may be required. Some of the subsequent studies, in particular that by Bliss and Smith (1998), shed doubt upon this finding and have lent support to the popular CIR model. On the other hand, the studies by Brailsford and Maheswaran (1998), Treepongkaruna and Gray (2003), and Gray (2005) show that estimates of the level effect parameter above 1.0 are not a one-off event.

To estimate the level effect parameter, previous studies have used the General Method of Moments (GMM), (Chan et al. (1992), Bliss and Smith (1998)), a Gaussian estimation technique (Nowman (1997)), or, most frequently, quasi-maximum likelihood methods (e.g. Treepongkaruna and Gray (2003), Gray (2005)). Accordingly, either χ^2-tests (Chan et al. (1992), Nowman (1997)), Newey-West (Bliss and Smith (1998)), or likelihood ratio tests (Treepongkaruna and Gray (2003)), are used to examine the support for the competing models. Observing that the level effect parameter and the volatility parameter exhibit a strong non-linear dependence, and that a range of combinations of the two parameters give values of similar likelihood, one might question if the uncertainty in these parameters is sufficiently taken into account by these methodologies. In fact, the likelihood ratio test is sometimes criticised for ignoring parameter uncertainty by solely focussing on the likelihood values at the maximum likelihood estimates. An alternative is the Bayes factor, which is similar to the likelihood ratio test, but takes parameter uncertainty into account. Furthermore, given that the parameters governing the short-rate model are potentially hard to estimate, I would like to study how estimation uncertainty affects the predictive ability of the CKLS model.

To assess whether parameter uncertainty can explain some of the discrepancies in the estimation results found in literature, I compute both likelihood ratios and Bayes factors for four short-rate models nested within the CKLS model and for nine currencies over two different estimation lengths (ten and 25 years) and three estimation frequencies. This wide range of data sets allows me to assess if, and to what extent, uncertainty in the level parameter can explain different estimation results. Furthermore, I apply an approximation technique introduced by Aït-Sahalia (2002) to obtain very accurate approximations of the transition densities where these are not available in closed form. This technique has not found widespread use in the empirical literature, even though it is superior in accuracy and computational efficiency compared to other approximation methods. As an additional contribution, I derive a re-parametrization of the CKLS model. This new parameterization facilitates the derivation of posterior parameter distributions by Metropolis-Hastings algorithms. Due to the strong dependence between the volatility and the level effect parameter, it is difficult to obtain a well-mixing chain when applying a Metropolis-Hastings algorithm to the original model. Finally, I apply an out-of-sample validation test to study if parameter uncertainty can enhance the predictive ability of the CKLS model.

The remainder of this chapter is structured as follows. Section 3.2 introduces the methodology used in this study, in particular the derivation of the likelihood functions based on Aït-Sahalia's approximation method, the re-parametrization of the CKLS model, the techniques used to obtain the Bayes factors, and the out-of-sample test for the predictive ability of the CKLS model. In Section 3.3, I introduce the data sets and present and discuss the estimation results and the insights gained from applying the likelihood ratio tests and the Bayes factor. Section 3.4 concludes with a discussion of the differences between the Bayes factor and the likelihood ratio tests and describes the findings of the chapter.

3.2. Methodology

This section discusses the methods I use to estimate and compare the different short-rate models under consideration. I focus on the CKLS model (H_0) defined in equation (3.1) and on four models nested within the CKLS model: the Vasicek model (H_1), the CIR model (H_2), the Brennan-Schwartz model (H_3) and a model that restricts the level effect parameter to 1.5, which I call VR (H_4). Table 3.1 shows the parameter restrictions imposed by the models H_0-H_4.

Model	Hypothesis	α	β	σ	ξ
CKLS	H_0	–	≤ 0	≥ 0	≥ 0
Vasicek	H_1	–	≤ 0	≥ 0	0.0
CIR	H_2	–	≤ 0	≥ 0	0.5
Brennan-Schwartz	H_3	–	≤ 0	≥ 0	1.0
VR	H_4	–	≤ 0	≥ 0	1.5

Table 3.1.: Parameter restrictions for models nested within the CKLS model.

3.2.1. Approximating the transition density of the CKLS model

One of the difficulties in working with the CKLS model is that the transition density is not available in closed form. For two points in time t_0 and t_1 it is impossible to describe the density $p(r, t_1 - t_0 \mid r_{t_0})$ such that

$$\mathbb{P}(r_{t_1} \leq x \mid r_{t_0}) = \int_{-\infty}^{x} p(r, t_1 - t_0 \mid r_{t_0}) dr.$$

Studies that seek to estimate the CKLS model must deal with this issue. The general approach is to apply an Euler discretization, meaning that one assumes that the transition density between two points in time is in fact normal. For the CKLS model, this means that the stochastic differential equation

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^{\xi} dW_t \tag{3.1}$$

is replaced with the discretized version

$$r_{t+\triangle} = r_t + (\alpha + \beta r_t)\triangle + \sigma r_t^{\xi} \sqrt{\triangle}\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1). \tag{3.2}$$

This approach is popular due to its simplicity and has been used by Chan et al. (1992), Nowman (1997), Bliss and Smith (1998), and Treepongkaruna and Gray (2003). One problem with this relatively simple approximation is that it introduces a discretization bias which depends on the size of the time step $\triangle$. Several approaches have been devised to address the issue of discretization bias. One popular method is the introduction of unobserved values of the interest rate r between observed time points t_0 and t_1. These unobserved values shorten the time step $\triangle$ and can be integrated out of the likelihood function using simulation techniques. This approach has been independently developed by Elerian, Chib, and Shephard (2001), Eraker (2001), and Jones (1998) and was also applied by Gray (2005). While this method successfully reduces the discretization bias introduced by the Euler approximation, it has the disadvantage of complicating the computation of the transition density and likelihood function. With the Euler approximation, both functions are available in closed form while the improved versions using unobserved time points now require

simulation for each evaluation.

Instead of relying on Euler approximations or simulation techniques, I apply a technique which was developed by Aït-Sahalia (2002). In this technique, Aït-Sahalia takes a diffusion process, such as the one used in the CKLS model, and transforms it into a process whose transition density can be approximated by a normal density plus a series of Hermite polynomials. As the transformation is known, it can then be reverted to obtain an approximating series that converges to the original transition density. This means that one can obtain arbitrarily good approximations of the true transition densities by introducing additional terms. In practice, the use of the approximation up to an order of $K = 1$ already yields a vast improvement over other available methods, and using two or three approximation steps should eliminate any error on a scale relevant for practical applications. Importantly, Aït-Sahalia also shows that a maximum likelihood estimator based on the approximated transition densities converges to the true maximum likelihood estimator as more approximation terms are added. Aït-Sahalia (2002) compares the different approximation methods in a chart that I have reproduced in Figure 3.1. This chart shows the vast superiority of this method, both in terms of accuracy as well as in computation time.

The only difficulty in applying this method lies in the actual computation of the closed form approximation which may lead to quite complex expressions. In the following, I show how this method can be used to obtain an approximation up to order K. I use an approximation of order $K = 2$, which should ensure sufficient accuracy for the purpose of this chapter.

To maintain the notation introduced by Aït-Sahalia (2002), let the random variable X denote the value of $r_{t+\triangle}$ and x_0 the observed value r_t. The true, unknown transition density from x_0 to x over time $\triangle$ and with parameters θ is denoted by $p_X(x, \triangle \mid x_0; \theta)$. Aït-Sahalia (2002) provides the approximation formulas for the transition densities in terms of Y. Y is the random variable obtained by applying the first transformation to X. This transformation is denoted by γ and takes the form

$$Y = \gamma(X; \theta) = \int^{X} \frac{du}{\sigma(u; \theta)} = \int^{r_{t+\triangle}} \frac{du}{\sigma u^{\xi}}, \tag{3.3}$$

where $\sigma(.; \theta)$ denotes the volatility term of the model. The approximated transition density for Y is then given by

$$p_Y^{(K)}(y, \triangle \mid y_0; \theta) = \\ \triangle^{-1/2} \phi\left(\frac{y - y_0}{\triangle^{1/2}}\right) \exp\left\{\int_{y_0}^{y} \mu_Y(\omega; \theta) d\omega\right\} \sum_{k=0}^{K} c_k(y \mid y_0; \theta) \frac{\triangle^k}{k!}, \tag{3.4}$$

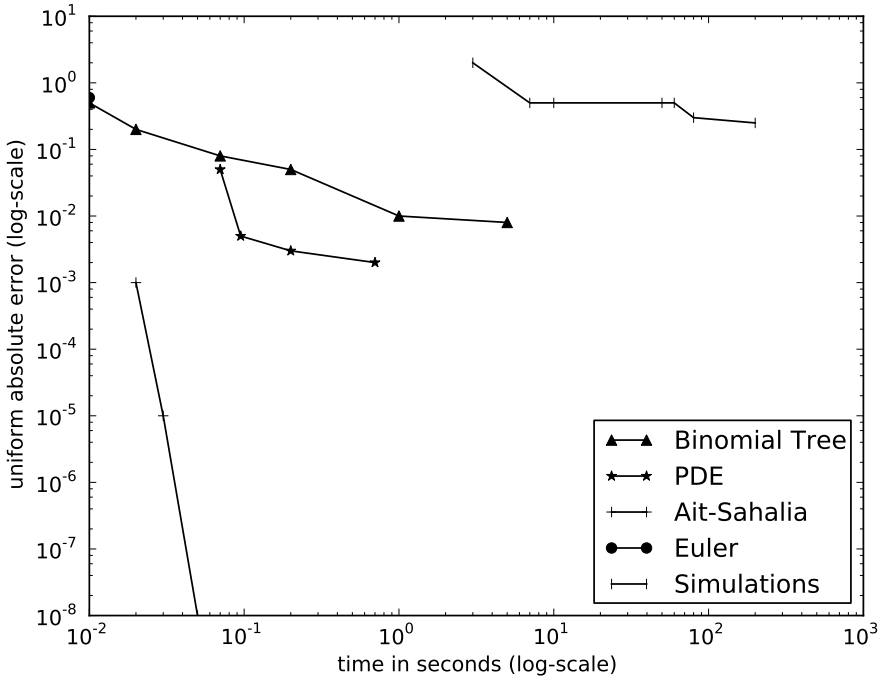


Figure 3.1.: This chart is reproduced from Aït-Sahalia (2002). It shows the average uniform absolute error of various density approximation techniques applied to the Vasicek, Cox-Ingersoll-Ross, and Black-Scholes models. *Euler* is the first-order Gaussian approximation scheme; *Binomial Tree* is a discrete-time, discrete-state approximation; *Simulations* refers to an implementation of Pedersen (1995)'s simulated-likelihood method; *PDE* refers to the numerical solution of the Fokker-Planck-Kolmogorov partial differential equation.

where K denotes the number of approximating terms and ϕ denotes the density of the standard normal distribution. The coefficients c_j can be calculated by the following equations

$$c_0(y \mid y_0;\theta) = 1 \tag{3.5}$$

$$c_j(y \mid y_0;\theta) = j(y-y_0)^{-j} \int_{y_0}^{y} (w-y_0)^{j-1} \times \left\{ \lambda_Y(\omega;\theta) c_{j-1}(\omega \mid y_0;\theta) + \frac{1}{2} \frac{\partial^2 c_{j-1}(\omega \mid y_0;\theta)}{\partial \omega^2} \right\} d\omega \tag{3.6}$$

$$\lambda_Y(y;\theta) = -\frac{1}{2} \left(\mu_Y^2(y;\theta) + \frac{\mu_Y(y;\theta)}{\partial y} \right). \tag{3.7}$$

To obtain the transition density of interest, $p_X^{(K)}(x, \triangle \mid x_0;\theta)$, one can revert the transformation form X to Y:

$$p_X^{(K)}(x, \triangle \mid x_0;\theta) = \sigma(x;\theta)^{-1} p_Y^{(K)}(\gamma(x;\theta), \triangle \mid \gamma(x_0;\theta), \theta). \tag{3.8}$$

In particular, for the CKLS model the first transformation function γ is given by

$$\gamma(X_t;\theta) = \begin{cases} \sigma^{-1}(1-\xi) X^{1-\xi} & \xi \in (0,1) \cup (1,\infty) \\ \sigma^{-1} \log X & \xi = 1 \end{cases}. \tag{3.9}$$

Denoting by μ and σ the drift and volatility terms of the original variable X, the drift of Y required by equations (3.4) and (3.7) is given by

$$\mu_Y(y;\theta) = \frac{\mu(\gamma^{-1}(y;\theta);\theta)}{\sigma(\gamma^{-1}(y;\theta);\theta)} - \frac{1}{2} \frac{\partial \sigma}{\partial x}(\gamma^{-1}(y;\theta);\theta). \tag{3.10}$$

Computing μ_Y for the CKLS model yields

$$\mu_Y(y;\theta) = \begin{cases} \frac{\xi}{2(\xi-1)y} + \beta(\xi-1)y - \alpha\beta^{1/(\xi-1)}(\xi-1)^{\xi/(\xi-1)} y^{\xi/(\xi-1)} & \xi \neq 1 \\ \frac{1}{\sigma}\left(\frac{\alpha}{\exp(\sigma y)} - \beta \right) - \frac{\sigma}{2} & \xi = 1 \end{cases}. \tag{3.11}$$

Equations (3.9) and (3.11) can be substituted into (3.8) to obtain the desired approximation formula for the transition density p_X and, by equation (3.6), the desired number of coefficients can be computed.

3.2.2. Maximum likelihood estimation and the likelihood ratio test

With the transition densities derived in the previous section, it is clear how to define the log-likelihood function. Since diffusion processes are Markov, the likelihood function for observations $r_{t_0}, \ldots, r_{t_n}$ is simply the product of the conditional transition densities:

$$L(\theta \mid \mathbf{r}) = \prod_{j=1}^{n} p(r_{t_j}, t_j - t_{j-1} \mid r_{t_{j-1}}; \theta). \tag{3.12}$$

Taking the logarithm on both sides and substituting the approximated density $p^{(K)}$ for p, one obtains the approximate log-likelihood function, which I denote by ℓ:

$$\ell(\theta \mid \mathbf{r}) = \sum_{j=1}^{n} \log p^{(K)}(r_{t_j}, \triangle \mid r_{t_{j-1}}; \theta). \tag{3.13}$$

To obtain the maximum likelihood estimator, I maximize (3.13) numerically. This numerical optimization is realized in two implementations, one in Haskell[2] and one in R[3]. Both implementations rely on the CMA-ES algorithm, which was developed in Hansen and Ostermeier (1996). This optimization algorithm proved to be more stable and reliable than the standard L-BFGS algorithm.[4] Standard errors of the maximum likelihood estimates are typically computed as the negative inverse of the log-likelihood's Hesse matrix. However, the numerical approximation of the second order derivatives needed for the Hessian matrix is sometimes unstable, therefore, I run MCMC algorithms to obtain samples from the log-likelihood and use these samples to compute standard errors.

The maximum likelihood estimator $\hat{\theta}_i$ for the models H_i, $i \in 0, \ldots, 4$ can be computed by restricting ξ where necessary. Once maximum likelihood estimates for all models have been computed, the likelihood ratio test statistic is readily available:

$$D = 2\ell(\theta_0) - 2\ell(\theta_i). \tag{3.14}$$

As mentioned previously, 0 indexes the most general model and subsequent models $i = 1, \ldots, 4$ are nested under model 0. In this case, Wilks (1938) shows that D is asymptotically χ^2_d distributed with d degrees of freedom. The degrees of freedom d are equal to the number of parameters fixed under model i, therefore $d \equiv 1$. Model i can be rejected if the test statistics exceeds a certain threshold.

[2] `http://www.haskell.org`
[3] `http://www.r-project.org`
[4] See Nocedal (1980) for a description of the algorithm.

3.2.3. Maximum a-posteriori estimates and Bayes factors

To obtain Bayesian estimates, such as maximum a-posteriori estimates (MAP) or Bayes factors, prior distributions are required and must first be specified. I specify two prior distributions, one informative prior and one flat, non-informative prior. As informative prior, I choose a normal distribution which uses the sample mean and sample standard deviation of the estimates made by previous studies. The motivation for this prior distribution is to model the prior knowledge of the statistician who has seen a number of studies on the models and thus a preconceived notion of what likely parameter values will be. The first prior distribution is thus given by

$$\pi_{\text{inf}}(\theta) = \phi\left(\theta; \mu_{\text{inf}}, \Sigma_{\text{inf}}\right), \tag{3.15}$$

where ϕ is the four-dimensional normal density with mean $\mu_{\text{inf}} = (0.0451, -0.5891, 2.6783, 1.2600)$ and standard deviations

$$\Sigma_{\text{inf}}^{1/2} = \begin{bmatrix} 0.1337 & 0 & 0 & 0 \\ 0 & 1.9427 & 0 & 0 \\ 0 & 0 & 6.9906 & 0 \\ 0 & 0 & 0 & 0.7326 \end{bmatrix}. \tag{3.16}$$

The second prior distribution is a flat prior which encompasses all values which are economically sensible for θ by a large margin. This prior is given by

$$\pi_{\text{flat}}(\theta) = \frac{1}{\lambda(\Theta)} \mathbb{1}_{\Theta}(\theta), \tag{3.17}$$

where $\Theta = \{(\alpha, \beta, \sigma, \xi) \,|\, \alpha \in [-50, 50], \beta \in [-100, 0], \sigma \in [0, 100], \xi \in [0, 100]\}$ and λ denotes the Lebesgue measure.

Given the prior densities for θ, the posterior density is proportional to

$$\pi_{\text{inf|flat}}(\theta|\mathbf{r}) \propto \prod_{j=1}^{n} p^{(K)}(r_{t_j}, \triangle \mid r_{t_{j-1}}; \theta) \cdot \pi_{\text{inf|flat}}(\theta). \tag{3.18}$$

Just as in the case of the maximum likelihood estimator, I take the logarithm of the posterior density and maximize it numerically to obtain the maximum a-posteriori estimates.

Obtaining Bayes factors from posterior distributions

From a Bayesian perspective, model selection is performed by determining the posterior probability of a model given prior probabilities and observed data. Let

$2\log(B)$	B	Evidence against H_0
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	>150	Very Strong

Table 3.2.: Interpretation of Bayes factors.

$\mathbf{r}$ denote the observed data and H_i the competing hypothesis with corresponding a priori probability $p(H_i)$. Observing the data $\mathbf{r}$ allows me to compute the probability of the data under the model, namely $p(\mathbf{r} \mid H_i)$. Using these inputs, Bayes theorem gives the posterior probability of model i:

$$p(H_i \mid \mathbf{r}) = \frac{p(\mathbf{r} \mid H_i)p(H_i)}{\sum_{i=0}^{n} p(\mathbf{r} \mid H_i)p(H_i)}. \tag{3.19}$$

To assess the evidence in favor of model i versus model j, I compute the quotient of their posterior distributions:

$$\frac{p(H_i \mid \mathbf{r})}{p(H_j \mid \mathbf{r})} = \frac{p(\mathbf{r} \mid H_i)}{p(\mathbf{r} \mid H_j)} \frac{p(H_i)}{p(H_j)}. \tag{3.20}$$

The term $\frac{p(\mathbf{r}|H_i)}{p(\mathbf{r}|H_j)}$ is the Bayes factor B_{ij} as described in Section 2.2.3 on page 25. Kass and Raftery (1995) propose a scale on which values of the Bayes factor can be interpreted in the same scale as the values obtained by likelihood-ratio tests. This scale is reproduced from their paper in Table 3.2.

The main motiviation for using the Bayes factor lies in the possibility that parameter uncertainty can be incorporated. In this case, the probability of the data $\mathbf{r}$ appearing under model i is

$$p(\mathbf{r} \mid H_i) = \int p(\mathbf{r} \mid \theta_i, H_i)p(\theta_i \mid H_i)d\theta_i, \tag{3.21}$$

where $p(\theta_i \mid H_i)$ represents the prior parameter density under model i.

The representation of $p(\mathbf{r} \mid H_i)$ in equation (3.21) raises the question of how the integral can be computed. Kass and Raftery (1995) discuss a number of approaches, such as approximations based on Laplace methods and simulations employing either Monte Carlo or Markov chain Monte Carlo techniques. Han and Carlin (2001) give a more comprehensive review of possible approximation techniques. I focus on a technique introduced by Verdinelli and Wasserman (1995) based on MCMC simulations. A description of this techniques follows.

Let $\theta = (\alpha, \beta, \sigma, \xi) \in \Theta$ be the parameter vector of interest and denote by $H_0 : \xi = \xi_0$, with alternative $H_1 : \xi \neq \xi_0$. For notational convenience, define $\psi = (\alpha, \beta, \sigma)$. Dickey (1971) noted that if $p(\psi \mid \xi_0) = p_0(\psi)$ then $B = \frac{p(\xi_0|\mathbf{r})}{p(\xi_0)}$. This requirement applies, for example, when flat prior of the form $p(\theta) = \frac{1}{\lambda(\Theta)} \mathbb{1}_{\{\theta \in \Theta\}}$ is used. When it does not apply, Verdinelli and Wasserman (1995) show that the Bayes factor may be computed as $B = p(\xi_0 \mid \mathbf{r}) \cdot \mathbb{E}[p_0(\psi)/p(\psi, \xi_0)]$, where p_0 and p are the parameter prior densities under H_0 and H_1, respectively.

The problem is thus reduced to the computation of $p(\xi_0 \mid \mathbf{r})$. Verdinelli and Wasserman (1995) propose a number of methods based on samples from the posterior distribution. I discuss how such samples are obtained in the next section. For now assume that $(\psi_1, \xi_1), \ldots, (\psi_n, \xi_n)$ represent a sequence of such samples. A brute force method to estimate $p(\xi_0 \mid \mathbf{r})$ is to calculate

$$\hat{p}(\xi_0 \mid \mathbf{r}) = \frac{1}{n} \sum_{i=1}^{n} \frac{L(\xi_0, \psi_i) p(\xi_0, \psi_i)}{\int L(\xi, \psi_i) p(\xi, \psi_i) d\xi}.$$

The integral in the denominator needs to be approximated, but it is only one-dimensional. Nevertheless, it needs to be computed for every sample which makes this method computationally burdensome if many samples are used. Furthermore, for the many data sets I employ, I have found it difficult to devise a general method to compute the integral accurately and reliably over all data sets in a computationally efficient way. Therefore, I rely on a different method, also presented by Verdinelli and Wasserman (1995).

Raftery (1996) notes that only an approximation of the Bayes factor is required and argues for using a normal approximation to the posterior distribution of ξ. This assumption is much less restrictive than to assume that the entire posterior distribution over all parameters follows a normal distribution. Nevertheless, this approximation may be inaccurate. Therefore, I follow the suggestion by Verdinelli and Wasserman and employ a Box-Cox transformation on the posterior samples before applying the normal approximation.

Box and Cox (1964) describe a family of transformations parameterized by λ. The topic of transformation functions has since been discussed very actively, a good summary is given in Sakia (1992). The goal of the transformation in each case is to transform the input variables in such a way that the transformed variables are normal or very close to normal. Since the values of ξ are non-negative, I employ the first transformation discussed by Box and Cox (1964) which takes the form

$$g(y; \lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \log y & (\lambda = 0) \end{cases}. \tag{3.22}$$

Box and Cox (1964) show that when the transformed variable is normal, the original variable y has density $f(y) = g'(y)\phi(g(y); \mu, \sigma)$, where μ and σ are the mean and standard deviation of the transformed variable and ϕ is the density of a standard normal distribution. To estimate the parameter λ, one approach proposed by Box and Cox (1964) is to use maximum likelihood estimation, i.e. to maximize the value of the product of the previous density at all observations y with respect to λ. Since an arbitrary number of samples can be obtained from the posterior distribution, this approach should accurately produce the λ that gives the best fit.

Once the parameter λ has been computed, the sample mean $\hat{\mu}$ and sample standard deviation $\hat{\sigma}$ can be estimated from the transformed samples of $g(\xi_1), \ldots, g(\xi_n)$. The target density $p(\xi_0 \mid \mathbf{r})$ can be approximated by

$$\hat{p}(\xi_0 \mid \mathbf{r}) = \frac{|g'(\xi_0)|}{\sqrt{2\pi}\hat{\sigma}} \exp\left\{-\frac{(g(\xi_0) - \hat{\mu})^2}{2\hat{\sigma}^2}\right\}. \tag{3.23}$$

3.2.4. Re-parameterizing the CKLS model for MCMC simulations

This section describes how I compute samples from the posterior distribution to calculate the Bayes factors. I employ a MCMC algorithm, namely a random walk Metropolis-Hastings algorithm to obtain these samples. However, it is difficult to obtain a well-mixing Markov chain when using either standard random-walk or independence Metropolis-Hastings algorithms with the standard parametrization of the CKLS model. Therefore, I introduce a random-walk Metropolis-Hastings algorithm which uses a carefully chosen proposal distribution.

The standard parameterization of the CKLS model as defined in equation (3.1) is

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^{\xi} dW_t. \tag{3.24}$$

With this parametrization and a current value of the short-rate r_0, an infinite combination of values σ and ξ give rise to the same instantaneous volatility $c = \sigma r_0^{\xi}$. In fact, given an instantaneous volatility c and a value of the short-rate r_0, it is possible to define one in term of the other:

$$\sigma = c r_0^{-\xi}. \tag{3.25}$$

Given that short-rates often tend around a center value r_0 (this gives rise to the idea of mean-reversion in short-rate models, after all), it is unsurprising that the likelihood function takes similar values over a range of combinations of σ and ξ.

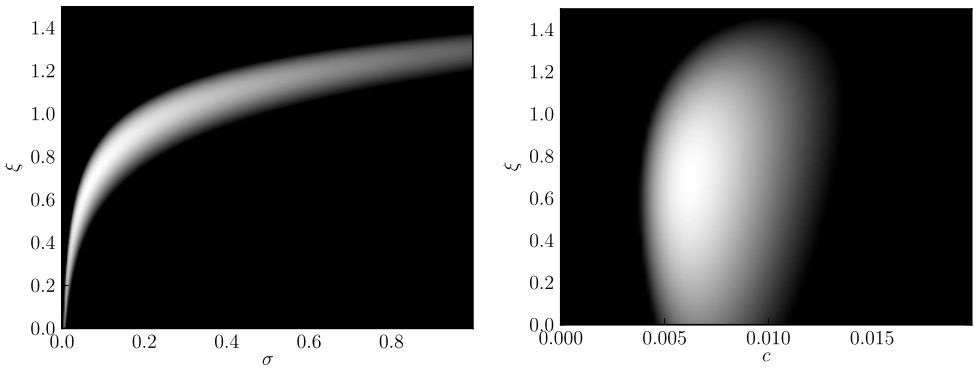


Figure 3.2.: This figure shows the values of the log-likelihood function of the CKLS model for 1-month EUR LIBOR rates from Dec 30, 1998 to June 30, 2011 with monthly observations. White indicates the maximum of the function, while black indicates a drop by 100 from the maximum. The log-likelihood function is plotted in terms of σ and ξ (left-hand side) and in terms of c and ξ (right-hand side). On the left hand side the non-linear relationship between σ and ξ becomes apparent. No such relationship is found between c and ξ.

The area where likelihood values are similar can be very narrow and has a non-linear shape. This is a problem for standard Metropolis-Hastings algorithms. I have found that both random walk and independence Metropolis-Hastings algorithms have difficulties in exploring posterior parameter distributions with these features.

To illustrate this issue, the left-hand side of Figure 3.2 shows the likelihood function of the CKLS model for values of σ and ξ while the parameters α and β are held at their maximum likelihood values. The non-linear relationship between σ and ξ becomes apparent as the likelihood function takes similar values along a narrow ridge. Observing that these combinations lead to very similar instantaneous volatility values c for a short-rate value of $\bar{r}$, I have plotted the likelihood function in terms of c and ξ on the right hand side of Figure 3.2. While the likelihood values in this plot are also not perfectly symmetric, there is not a non-linear dependence as there is on the left-hand side.

Motivated by this observation, I introduce a proposal distribution for a random walk Metropolis-Hastings algorithm that takes this dependence into account. The two drift parameters α and β exhibit little dependence on the volatility parameters σ and ξ. Therefore, candidates $\tilde{\alpha}$ and $\tilde{\beta}$ are drawn independently from a bivariate normal distribution centered at the previous values with some fixed covariance matrix: $(\tilde{\alpha}, \tilde{\beta}) \sim \mathcal{N}((\alpha, \beta), \Sigma)$. The proposal density is thus normal:

$$q_{\alpha,\beta}(\tilde{\alpha}, \tilde{\beta} \mid \alpha, \beta) = \phi((\tilde{\alpha}, \tilde{\beta}) \,;\, (\alpha, \beta), \Sigma). \tag{3.26}$$

Algorithm 3 Random walk Metropolis-Hastings for CKLS model with special proposal distribution.

1. Choose a start value $\theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \sigma^{(0)}, \xi^{(0)})$
2. For $i = 0, \ldots$
   a) Draw a new candidate $\tilde{\theta}$ with the following procedure:
      - Draw $(\tilde{\alpha}, \tilde{\beta}) \sim \mathcal{N}((\alpha^{(i)}, \beta^{(i)})\,;\, \Sigma)$.
      - Draw $\tilde{\xi} \sim \mathcal{N}(\xi^{(i)}\,;\, \delta_\xi)$.
      - Draw $\tilde{\sigma} \sim \mathcal{N}(\sigma^{(i)} r_0^{-\tilde{\xi}}\,;\, \delta_c r_0^{-\tilde{\xi}})$.
   b) If $\min\left\{\frac{p_X(\tilde{\theta}) q_\theta(\theta^{(i)}|\tilde{\theta})}{p_X(\theta^{(i)}) q_\theta(\tilde{\theta}|\theta^{(i)})}, 1\right\} > u$, with $u \sim U[0,1]$, set $\theta^{(i+1)} = \tilde{\theta}$, otherwise set $\theta^{(i+1)} = \theta^{(i)}$.
   c) If $i =$ MAXSTEPS stop, otherwise go to step 2(a).

For the volatility parameters σ and ξ, I perform a Gibbs-within-Metropolis step in the proposal distribution. First, draw $\tilde{\xi}$ from a normal distribution centered at its previous value ξ: $\tilde{\xi} \sim \mathcal{N}(\xi; \delta_\xi)$, where δ_ξ denotes the standard deviation of the proposal distribution. To draw $\tilde{\sigma}$, I first compute the previous instantaneous volatility based on r_0: $c = \sigma r_0^{-\xi}$. I then draw a disturbed $\tilde{c} \sim \mathcal{N}(c, \delta_c)$ and compute the candidate $\tilde{\sigma}$ based on this candidate for the instantaneous volatility: $\tilde{\sigma} = \tilde{c} r_0^{-\tilde{\xi}}$. Given $\tilde{\xi}$, $\tilde{\sigma}$ is conditionally normal with mean $c r_0^{-\tilde{\xi}}$ and standard deviation $\delta_c r_0^{-\tilde{\xi}}$. The joint proposal density for σ and ξ is thus

$$q_{\sigma,\xi}(\tilde{\sigma}, \tilde{\xi} \mid \sigma, \xi) = \phi(\tilde{\xi}\,;\, \xi, \delta_\xi) \cdot \phi(\tilde{\sigma}\,;\, \sigma r_0^{-\tilde{\xi}}, \delta_c r_0^{-\tilde{\xi}}). \tag{3.27}$$

Denoting by $\theta = (\alpha, \beta, \sigma, \xi)$ the full parameter vector, the complete proposal density is obtained by multiplying the proposal densities (3.26) and (3.27):

$$\begin{aligned} q_\theta(\tilde{\theta} \mid \theta) &= q_{(\alpha,\beta,\sigma,\xi)}(\tilde{\alpha}, \tilde{\beta}, \tilde{\sigma}, \tilde{\xi} \mid \alpha, \beta, \sigma, \xi) \\ &= q_{\alpha,\beta}(\tilde{\alpha}, \tilde{\beta} \mid \alpha, \beta) \cdot q_{\sigma,\xi}(\tilde{\sigma}, \tilde{\xi} \mid \sigma, \xi). \end{aligned} \tag{3.28}$$

Algorithm 3 details the full Markov chain Monte Carlo algorithm using the previously discussed proposal density.

3.2.5. Testing model predictions

One way in which short-rate models may be used is to predict the future distribution of the value of the short-rate or of values dependent upon it. In addition to comparing how well the different models fit the data, it is also interesting to see how well they fare in out-of-sample predictions. In particular, I would

like to test whether the presented short-rate interest rate models are sufficiently cautious in their prediction of possible short-rate movements and to what degree parameter uncertainty can affect these results. To evaluate the soundness of the model predictions, I use a backtesting approach which was developed by Christoffersen (1998) to test the fitness of VaR models.

The idea behind this backtest is to repeatedly compute quantiles of the predictive distribution of the short-rate and to count the number of times that realizations fall outside of these quantiles. To formalize this method, assume that the process X_t describes the evolution of the short-rate and that, given a current realization x_t and a time delta $\triangle$, the transition density $f(x_{t+\triangle}|\triangle, x_t)$ is known. Denoting by $F(x) := F(x|\triangle, x_t)$ the cumulative density function and by $F^{-1}(y) = \inf_{x\in\mathbb{R}}\{F(x) \geq y\}$ its inverse, the α-quantile is given by $q_\alpha = F^{-1}(\alpha)$. Let

$$I_t(\alpha) = \mathbb{1}(X_{t+\triangle} < q_\alpha) \tag{3.29}$$

denote the random variable that is 1 whenever the realization at time point $t+\triangle$ falls below the α-quantile and 0 otherwise. Since the short-rate process is assumed to be Markov, the $I_t(\alpha)$ are independent identically Bernoulli-distributed with success probability α. This implies that the sum over all I_t has a Binomial distribution:

$$I_t(\alpha) \overset{iid}{\sim} B(1, \alpha), \tag{3.30}$$

$$\sum_{t=0}^{n-1} I_t(\alpha) \sim B(n, \alpha). \tag{3.31}$$

Under the hypothesis that the assumed model accurately describes the transition density of the short-rate, the probability of observing i or more realizations below the α-quantile in a total of n realizations can then be computed as:

$$P\left(\sum_t I_t(\alpha) \geq i\right) = \sum_{k=i}^{n} \binom{n}{k} \alpha^k (1-\alpha)^{n-k}. \tag{3.32}$$

If this probability is small, then the hypothesis that the given model correctly describes the transition densities can be rejected. The same test can be constructed to count the number of occurences that exceed, rather than fall below, a given quantile.

To apply this test, it is thus necessary to obtain the transition density of the model. I apply this test to the CKLS model both in the case where parameters are affected by uncertainty and where they are not. In both cases, parameter estimates are obtained by using a *rolling window* approach, i.e. by estimating the model's parameters over a fixed time frame which immediately precedes the next observation.

In the case where parameter uncertainty is ignored, parameters are estimated using the maximum likelihood estimator on a 5-year window. Since parameters have to be estimated many times over the length of the path, I use Euler approximations in this case to ease the computational burden. This also facilitates computing the quantiles of the predictive distribution as the predictive distribution under the Euler approximation is normal.

In the case that considers parameter uncertainty, a posterior parameter distribution is estimated using flat priors and the MCMC algorithm described in Section 3.2.4. Again, I use an Euler approximation for the transition density to ease the computational burden. Once an estimate of the posterior distribution has been obtained, I compute the normal distribution that results from using an Euler approximation of the model with this particular parameter set. I then integrate over these distributions to obtain the predictive distribution.

3.3. Data sets and estimation results

3.3.1. Data sets

I estimate the parameters of the CKLS short-rate model and the four models nested within it for nine major currencies: the US dollar (USD), British pound (GBP), Japanese yen (JPY), French franc (FRF), Deutschmark (DEM), Italian lira (ITL), Swiss franc (CHF), Australian dollar (AUD) and the Euro (EUR). All models are estimated on two observation windows, the first spanning ten years (from July 14, 2001 to July 14, 2011) and the second spanning 25 years (from July 14, 1986 to July 14, 2011). For the ten year observation window, the European currencies French franc, Deutschmark, and Italian lira are considered from the beginning of the observation window up to April 19, 2000. This approach mirrors that of Treepongkaruna and Gray (2003). For the 25 year window, interest rate data for the Deutschmark is used in lieu of Euro data prior to December 30, 1998. All estimations were conducted with three observation frequencies, namely with monthly, weekly, and daily observations.

For all currencies, except the Australian dollar, I use one-month LIBOR rates provided by the British Bankers Association. For the Australian dollar, I use the 90-day dealer bill rate, which is the most liquid instrument in that market. Daily observations of this data were obtained from Datastream beginning on January 2, 1986 and ending on July 14, 2011. For the French franc, the Italian lira, and the Japanese yen, data is available only starting after January 2, 1986, and were retrieved from the earliest available time point onward. The time series for the Euro starts on December 30, 1998. As in Treepongkaruna and Gray (2003), weekends and holidays are not treated specially, since French

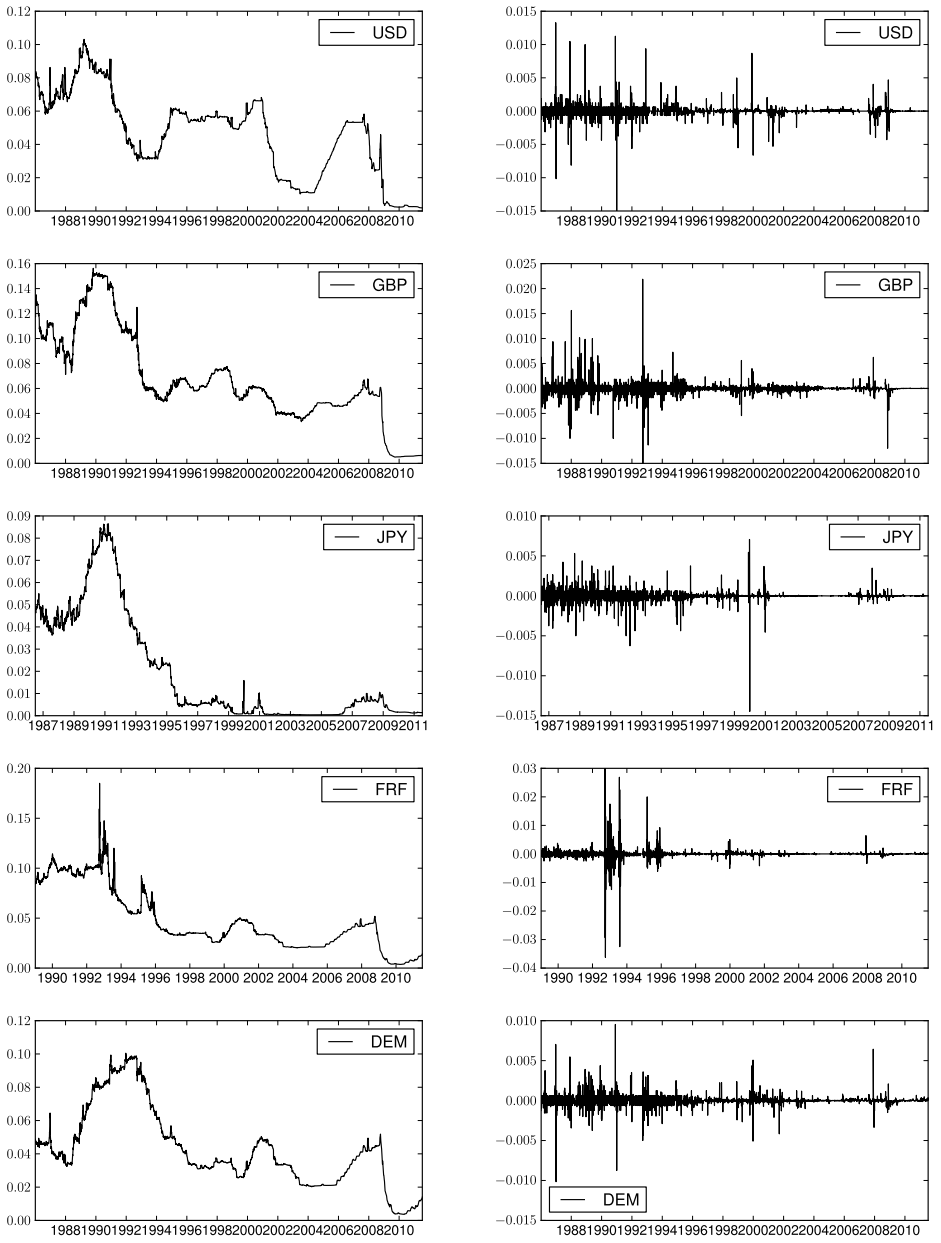


Figure 3.3.: Short-term interest rates for the US, UK, Japan, France, Germany, Italy, Switzerland, Australia, and the Euro zone. The left column shows the levels of the short-term interest rates while the right column shows first differences. Australian rates are 90-day bank accepted bill rates, all others are one-month LIBOR rates as quoted by the British Bankers Association. Data is collected from January 2, 1986 or from the first time point available in Datastream. The Euro interest rate time series starts on December 30, 1998 (continued on next page).

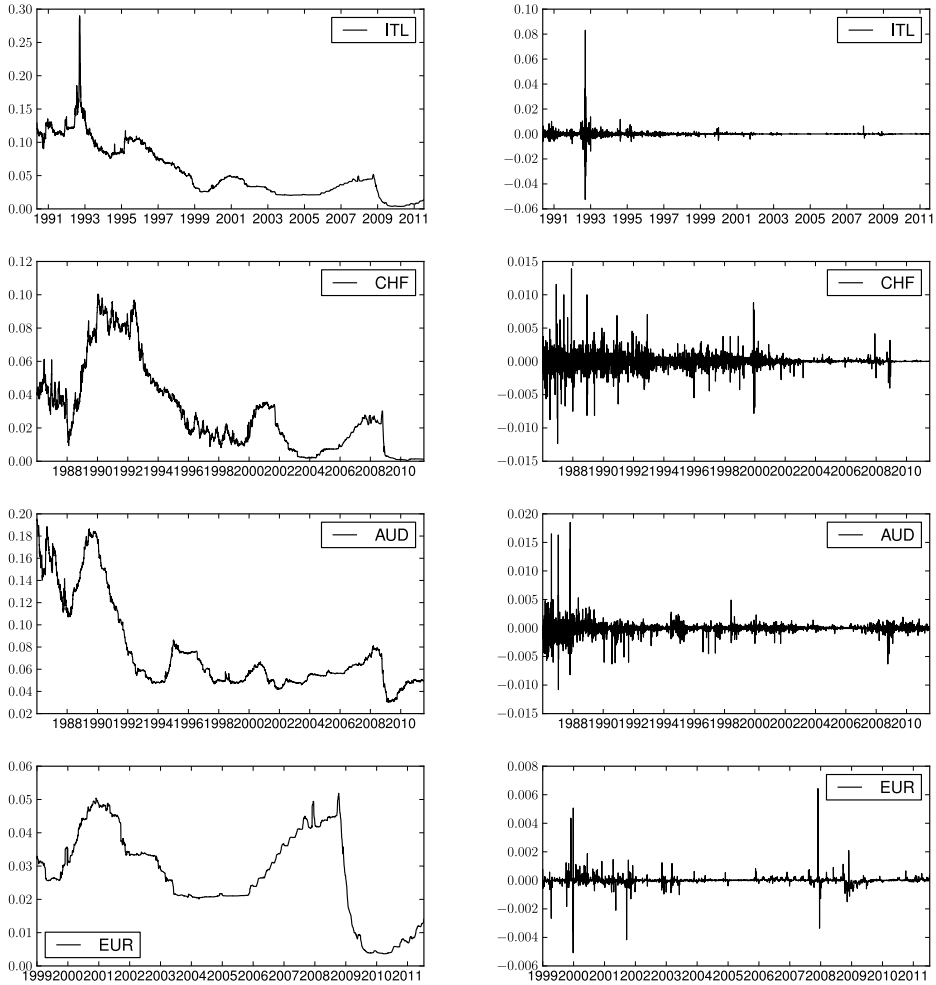


Figure 3.3.: (continued)

Currency	USD	GBP	JPY	FRF	DEM	ITL	CHF	AUD	EUR
Start date	Jan 2, 86	Jan 2, 86	Jul 1, 86	Jan 3, 89	Jan 2, 86	May 17, 90	Jan 2, 86	Jan 2, 86	Dec 30, 98
End date	Jul 14, 11	Jul 14, 11	Jul 14, 11	Apr 19, 00	Apr 19, 00	Apr 19, 00	Jul 14, 11	Jul 14, 11	Jul 14, 11
Number of observations	6660	6660	6532	5877	6660	5520	6660	6660	3271
Spot rate (%)									
Mean	4.56	6.66	1.89	4.70	4.26	5.56	2.99	7.64	2.76
Standard dev.	2.49	3.61	2.41	3.05	2.36	3.99	2.60	4.01	1.32
Skewness	-0.14	0.58	1.22	0.80	0.74	0.90	1.01	1.40	-0.20
Kurtosis	-0.77	-0.01	0.24	-0.34	-0.03	0.56	0.06	0.71	-0.90
Minimum	0.19	0.50	0.04	0.36	0.36	0.36	0.03	3.00	0.36
Maximum	10.31	15.62	8.64	18.50	10.05	29.00	10.05	19.53	5.19
$\rho(1)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho(10)$	0.99	1.00	1.00	0.99	1.00	0.99	0.99	0.99	1.00
$\rho(20)$	0.99	0.99	1.00	0.98	1.00	0.98	0.99	0.98	0.99
First difference daily (%)									
Mean	-0.001 19	-0.001 67	-0.000 72	-0.001 23	-0.000 52	-0.002 08	-0.000 59	-0.002 19	-0.000 55
Standard dev.	0.06	0.09	0.05	0.15	0.05	0.21	0.09	0.08	0.03
Skewness	0.70	2.48	-4.22	0.12	-0.13	8.66	0.46	3.52	2.23
Kurtosis	143.76	129.66	172.16	217.29	86.68	587.68	41.76	111.04	172.00
Minimum	-1.50	-1.50	-1.45	-3.62	-1.02	-5.25	-1.23	-1.08	-0.51
Maximum	1.33	2.19	0.71	3.00	0.95	8.31	1.39	1.85	0.64
$\rho(1)$	0.09	0.02	0.07	0.00	0.09	-0.12	0.05	0.03	0.08
$\rho(10)$	0.00	0.04	-0.03	-0.05	0.04	0.01	0.05	-0.03	0.07
$\rho(20)$	0.01	0.03	0.00	-0.02	0.03	0.01	-0.01	0.03	0.13

Table 3.3.: This table presents descriptive statistics of the various time series. Data for the US, UK, Japan, France, Germany, Italy, and Switzerland and the Euro are inter-bank one-month rates quoted by the British Bankers Association. Data for Australia are Dealer Bill 90-day rates. All data was obtained from Datastream.

and Roll (1986) document an inconclusive weekend effect for money market instruments. Descriptive statistics for the time series are presented in Table 3.3.

The minimum observed interest rate is 0.035 percent for the Swiss franc occurring on June 2, 2010. The maximum observed rate is 29.00 percent for the Italian lira on September 16, 1992, during the European currency crisis. The unconditional mean of the short-rate ranges from 1.89 percent for the Japanese yen to 8.96 percent for the Australian dollar. The mean of the first differences is negative across all time series, indicating falling interest rates over the observation window. The largest changes in the interest rates of 8.31 and -5.25 percent occur in the data for the Italian lira on September 16, 1992 and September 14, 1992. These dramatic jumps are also associated with the European currency crisis. As could be expected, all short-rates exhibit very high autocorrelation, but the autocorrelation in the first differences is negligible.

Plots of the different time series along with their first differences are provided in Figure 3.3. The graphs show that all short-term interest rates exhibit a falling trend over the observation horizon. This finding is in line with the negative mean of their first differences shown in Table 3.3. From 2009 on, all interest rates show a significant drop as central bankers lowered the target rate to help the ailing economy and to ease the stress on banks after the financial crisis of 2008. European rates have risen slightly since then, but the US and UK rates remain firmly close to zero. The Japanese short-rate has tended around zero since the second half of 2001.

3.3.2. Estimation results

Tables 3.4 and 3.5 present the estimation results for the ten year and 25 year observation windows. The maximum likelihood and maximum a-posteriori estimates are computed by maximizing the log-likelihood function or the posterior density, respectively. Stars on the estimates indicate their significance, where a single star indicates that an estimate is different from 0 at the 10% level, and two and three stars indicate significance at the 5% and 1% level, respectively. In case of the maximum likelihood estimator, the asymptotic normality is used to compute approximate z-scores. Standard errors for the estimator are obtained by computing the empirical covariance matrix from samples from the log-likelihood function.

For the Bayesian estimates, a similar method is used to provide a point of comparison. Instead of using samples from the log-likelihood function, I use samples produced by the respective MCMC chains, using either the informative or the flat prior. As such, these values are comparable to the p-values of the

maximum likelihood estimator. These p-values are presented in Tables A.1 and A.2 in Appendix A.1.

Descriptive statistic and convergence diagnostics for the Markov chains are also presented in Appendix A.1. In general, these results show the good convergence characteristics of the MCMC algorithm presented in Section 3.2.4. The descriptive statistics of the chains shown in Tables A.3 and A.4 indicate that the acceptance ratio of all chains is in the desirable range of 0.3 to 0.4. Furthermore, the convergence diagnostics in Tables A.5 and A.6 show that the autocorrelation with a lag size of 20 is fairly low. The p-value of the Kolmogorov-Smirnov statistic is generally well above 0.1 and the test statistic $\hat{R}$ described by Gelman and Rubin (1992) is 1.0 in all cases. This indicates that all chains have converged and are well mixing. The effective sample size for these chains, which consist of 40 000 draws after discarding 10 000 burn-in samples, is between 2200 and 3200 samples.

Drift parameters

The parameters α and β govern the drift function of the CKLS model. Their estimates over the ten and 25 year observation windows can be found in Tables 3.4 and 3.5. As I use the same notation as Treepongkaruna and Gray (2003) and other previous studies, the speed of mean reversion is given by $-\beta$, while $-\frac{\alpha}{\beta}$ is the mean reversion level.

Overall, there is limited evidence of mean reversion. The mean reversion parameter β is highly significant for daily observations of the British pound over both the ten and the 25 year observation horizon. It is furthermore significant at the 5%-level for weekly observations of the British pound, daily and weekly observations of the Swiss franc and the Euro over the ten year observation horizon, as well as monthly observations of the Australian dollar over the 25 year horizon. There is only weak evidence for a non-zero β for weekly observations of the Deutschmark and monthly observations of the Swiss franc over the ten year window. For all other currencies and observation frequencies, β is insignificant. The p-values for the estimates are presented in Tables A.1 and A.2 in appendix A.1.

Table 3.4.: This table shows parameter estimates for the CKLS model over the ten year observation window. The maximum likelihood estimates are obtained by maximizing the log-likelihood function numerically. The maximum a-posteriori estimates are obtained by maximizing the posterior probability distribution, obtained under the informative or flat prior, respectively. No stars indicate that the estimate is not significant, while *, **, and *** stars indicate significance at the 10%, 5%, and 1% level, respectively.

Timeseries	MLE				MAP, informative prior				MAP, flat prior			
	α	β	σ	ξ	α	β	σ	ξ	α	β	σ	ξ
USD, 10y, d	-0.0002	-0.0766	0.1245***	0.8629***	-0.0005	-0.0749	0.1149***	0.8440***	-0.0002	-0.0766	0.1245***	0.8629***
USD, 10y, w	-0.0002	-0.0609	0.2411***	0.8577***	0.0001	-0.1227	0.2704***	0.8876***	-0.0002	-0.0609	0.2411***	0.8577***
USD, 10y, m	0.0004	-0.0867	0.5654***	0.9636***	0.0007	-0.1622	0.6061***	0.9799***	0.0004	-0.0867	0.5654***	0.9636***
GBP, 10y, d	0.0003	-0.1655***	0.3629***	1.3265***	0.0004	-0.1840***	0.3583***	1.3286***	0.0003	-0.1952***	0.3587***	1.3256***
GBP, 10y, w	0.0002	-0.2148**	0.2966***	1.1296***	-0.0006	-0.0944**	0.3245***	1.1553***	0.0002	-0.2148**	0.2966***	1.1296***
GBP, 10y, m	0.0005	-0.0737	0.3332***	1.0012***	-0.0010	-0.0695	0.3444***	1.0126***	0.0005	-0.0737	0.3332***	1.0012***
JPY, 10y, d	0.0001	-0.0363	0.3760***	0.9861***	0.0002	-0.0938	0.3847***	0.9894***	0.0001	-0.0745	0.3926***	0.9909***
JPY, 10y, w	0.0001	-0.0093	1.1934***	1.1252***	0.0001	-0.0954	1.5330***	1.1632***	0.0001	-0.0093	1.1934***	1.1252***
JPY, 10y, m	0.0002*	-0.1048	7.2424***	1.3889***	0.0001*	-0.0109	6.5019***	1.3790***	0.0002*	-0.1048	7.2424***	1.3889***
FRF, 10y, d	0.0087*	-0.2118	3.0195***	1.8886***	0.0039*	-0.0688	2.9907***	1.8836***	0.0087*	-0.2118	3.0195***	1.8886***
FRF, 10y, w	0.0072	-0.2234	3.5037***	1.9290***	0.0076	-0.1026	3.4625***	1.9155***	0.0072	-0.2234	3.5037***	1.9290***
FRF, 10y, m	0.0065	-0.1629	1.7850***	1.6552***	0.0024	-0.0273	1.7577***	1.6191***	0.0065	-0.1629	1.7850***	1.6552***
DEM, 10y, d	0.0030	-0.1006	0.0663***	0.7875***	0.0014	-0.0376	0.0671***	0.7853***	0.0030	-0.1006	0.0663***	0.7875***
DEM, 10y, w	0.0067	-0.1890*	0.1679***	1.0472***	-0.0002	-0.0395*	0.1497***	1.0046***	0.0067	-0.1890*	0.1679***	1.0472***
DEM, 10y, m	0.0043	-0.1437	0.0958***	0.8159***	-0.0014	-0.0247	0.0965***	0.8120***	0.0043	-0.1437	0.0958***	0.8159***
ITL, 10y, d	0.0005	-0.0151	0.6573***	1.4266***	0.0012	-0.0090	0.6718***	1.4345***	0.0005	-0.0117	0.6717***	1.4367***
ITL, 10y, w	-0.0007	-0.0633	0.7091***	1.4393***	-0.0026	-0.0322	0.6975***	1.4214***	-0.0007	-0.0633	0.7091***	1.4393***
ITL, 10y, m	0.0040	-0.0030	1.1227***	1.4873***	-0.0004	-0.1035	1.1836***	1.5178***	0.0040	-0.0030	1.1227***	1.4873***
CHF, 10y, d	0.0002	-0.2658**	0.1379***	0.8044***	0.0000	-0.1259**	0.1437***	0.8113***	0.0002	-0.2658**	0.1379***	0.8044***
CHF, 10y, w	0.0004	-0.3866**	0.2604***	0.8576***	0.0004	-0.0738**	0.2712***	0.8632***	0.0004	-0.3866**	0.2604***	0.8576***
CHF, 10y, m	0.0012**	-0.3879*	0.9326***	1.0616***	0.0003**	-0.0461*	0.6819***	0.9989***	0.0012**	-0.3879*	0.9326***	1.0616***
AUD, 10y, d	0.0020	-0.0328	0.0076***	0.1693***	0.0006	-0.0011	0.0074***	0.1602***	0.0020	-0.0328	0.0076***	0.1693***
AUD, 10y, w	0.0008	-0.0014	0.0177***	0.3360***	0.0088	-0.1587	0.0250***	0.4520***	0.0008	-0.0014	0.0177***	0.3360***
AUD, 10y, m	0.0120	-0.1837	0.0268	0.3901**	0.0009	-0.0211	0.0415	0.5474**	0.0120	-0.1837	0.0268	0.3901**
EUR, 10y, d	0.0002	-0.0927**	0.0571***	0.8232***	0.0001	-0.0999**	0.0552***	0.8125***	0.0002	-0.0927**	0.0571***	0.8232***
EUR, 10y, w	0.0013	-0.2024**	0.0723***	0.7320***	0.0000	-0.0755**	0.0717***	0.7300***	0.0013	-0.2024**	0.0723***	0.7320***
EUR, 10y, m	0.0005	-0.1212	0.1588***	0.7984***	0.0026	-0.1371	0.1579***	0.7900***	0.0005	-0.1212	0.1588***	0.7984***

Table 3.5.: This table shows parameter estimates for the CKLS model over the 25 year observation window. The maximum likelihood estimates are obtained by maximizing the log-likelihood function numerically. The maximum a-posteriori estimates are obtained by maximizing the posterior probability distribution, obtained under the informative or flat prior, respectively. No stars indicate that the estimate is not significant, while *, **, and *** stars indicate significance at the 10%, 5%, and 1% level, respectively.

Timeseries	MLE				MAP, informative prior				MAP, flat prior			
	α	β	σ	ξ	α	β	σ	ξ	α	β	σ	ξ
USD, 25y, d	-0.0006	-0.0060	0.1012***	0.8195***	-0.0007	-0.0093	0.1004***	0.8184***	-0.0006	-0.0060	0.1012***	0.8195***
USD, 25y, w	-0.0011	-0.0432	0.0842***	0.6384***	0.0000	-0.0939	0.0892***	0.6543***	-0.0011	-0.0432	0.0842***	0.6384***
USD, 25y, m	-0.0002	-0.0691	0.0461***	0.4068***	0.0011	-0.0366	0.0399***	0.3664***	-0.0002	-0.0691	0.0461***	0.4068***
GBP, 25y, d	-0.0003	-0.0892***	0.3437***	1.3047***	-0.0004	-0.0376***	0.3417***	1.3018***	-0.0003	-0.0892***	0.3437***	1.3047***
GBP, 25y, w	-0.0016*	-0.0247	0.1612***	0.9616***	-0.0011*	-0.0571	0.1699***	0.9849***	-0.0016*	-0.0247	0.1612***	0.9616***
GBP, 25y, m	-0.0015	-0.0326	0.0517***	0.5075***	-0.0010	-0.0538	0.0501***	0.5036***	-0.0015	-0.0326	0.0517***	0.5075***
JPY, 25y, d	0.0009***	-0.0703	0.1205***	0.6680***	0.0007***	-0.0330	0.1249***	0.6740***	0.0009***	-0.0703	0.1205***	0.6680***
JPY, 25y, w	0.0014***	-0.0706	0.1533***	0.6724***	0.0013***	-0.0712	0.1615***	0.6804***	0.0014***	-0.0706	0.1533***	0.6724***
JPY, 25y, m	0.0011***	-0.0663	0.1351***	0.6830***	0.0010***	-0.0393	0.1231***	0.6663***	0.0011***	-0.0663	0.1351***	0.6830***
CHF, 25y, d	0.0008**	-0.0620	0.2393***	0.8504***	0.0003**	-0.0568	0.2375***	0.8474***	0.0001**	-0.0104	0.2398***	0.8504***
CHF, 25y, w	0.0003	-0.0108	0.2674***	0.8134***	0.0007	-0.0721	0.3030***	0.8392***	0.0003	-0.0108	0.2674***	0.8134***
CHF, 25y, m	0.0007	-0.0700	0.1829***	0.7376***	0.0021	-0.1564	0.1826***	0.7298***	0.0007	-0.0700	0.1829***	0.7376***
AUD, 25y, d	0.0026	-0.0568	0.1864***	1.2001***	0.0016	-0.0427	0.1849***	1.1952***	0.0026	-0.0568	0.1864***	1.2001***
AUD, 25y, w	0.0053	-0.1153	0.1651***	1.0624***	0.0021	-0.0518	0.1627***	1.0538***	0.0053	-0.1153	0.1651***	1.0624***
AUD, 25y, m	0.0099*	-0.1975**	0.1714***	1.0490***	0.0062*	-0.1314**	0.1528***	1.0080***	0.0099*	-0.1975**	0.1714***	1.0490***
DEM, 25y, d	0.0003	-0.0309	0.1346***	0.9895***	0.0001	-0.0260	0.1350***	0.9914***	0.0003	-0.0309	0.1346***	0.9895***
DEM, 25y, w	-0.0002	-0.0137	0.0934***	0.7792***	0.0011	-0.0732	0.0972***	0.7937***	-0.0002	-0.0137	0.0934***	0.7792***
DEM, 25y, m	0.0021	-0.0877	0.0391***	0.4539***	0.0017	-0.0306	0.0330***	0.4003***	0.0021	-0.0877	0.0391***	0.4539***

Table 3.6.: This table compares the estimation results from this chapter to the results of other studies. Values in parenthesis are z-scores and one, two, or three stars indicate significance at the 10%, 5%, or 1% level. Significance for the Bayesian MAP estimate is determined by the comparable method outlined in Section 3.3.2.

Study	Dataset	Method	α		β		σ		ξ	
My results	USD 1m LIBOR, weekly, Jan 2, 86-Jul 14, 11	MLE	-0.0011	-(0.92)	-0.0432	-(0.90)	0.0842	(12.16)***	0.6384	(26.67)***
My results	USD 1m LIBOR, weekly, Jan 2, 86-Jul 14, 11	MAP (inf)	0.0000	(0.02)	-0.0939	-(2.14)**	0.0892	(13.01)***	0.6543	(27.64)***
Treepongkaruna and Gray (2003)	USD 1m LIBOR, weekly, Jan 2, 86-Apr 19, 00	QMLE	0.0164	(1.44)	-0.2915	-(1.50)	0.1127	(7.06)***	0.7704	(32.24)***
Chan et al. (1992)	USD 1m T-Bills, monthly, Jun 64-Dec 89	GMM	0.4896	(1.85)*	-7.1052	-(1.55)	4.4771	(0.77)	1.4999	(5.95)***
Nowman (1998)	USD 1m Euro-Curr., monthly, Feb 81-Mar 95	Gauss	0.0180	(1.87)*	-0.3108	-(1.94)*	0.2960	(2.00)*	1.0519	(11.25)***
My results	GBP 1m LIBOR, weekly, Jan 2, 86-Jul 14, 11	MLE	-0.0016	-(1.89)*	-0.0247	-(0.66)	0.1612	(10.99)***	0.9616	(31.63)***
My results	GBP 1m LIBOR, weekly, Jan 2, 86-Jul 14, 11	MAP (inf)	-0.0011	-(1.31)	-0.0571	-(1.59)	0.1699	(11.49)***	0.9849	(32.07)***
Treepongkaruna and Gray (2003)	GBP 1m LIBOR, weekly, Jan 2, 86-Apr 19, 00	QMLE	0.0141	(1.23)	-0.2147	-(1.31)	0.5350	(6.21)***	1.4527	(41.33)***
My results	JPY 1m LIBOR, weekly, Jul 1, 86-Jul 14, 11	MLE	0.0014	(3.55)***	-0.0706	-(0.59)	0.1533	(9.37)***	0.6724	(34.18)***
My results	JPY 1m LIBOR, weekly, Jul 1, 86-Jul 14, 11	MAP (inf)	0.0013	(3.49)***	-0.0712	-(0.88)	0.1615	(9.87)***	0.6804	(34.63)***
Treepongkaruna and Gray (2003)	JPY 1m LIBOR, weekly, Jan 9, 86-Apr 19, 00	QMLE	0.0009	(0.23)	-0.1541	-(1.43)	0.0261	(13.60)***	0.2593	(38.99)***
Nowman (1998)	JPY 1m Euro-Curr., monthly, Feb 81-Mar 95	Gauss	0.0036	(0.41)	-0.1200	-(0.73)	0.2272	(1.15)	0.9838	(6.73)***
My results	FRF 1m LIBOR, weekly, Jan 3, 89-Apr 19, 00	MLE	0.0072	(1.09)	-0.2234	-(1.23)	3.5037	(8.14)***	1.9290	(37.70)***
My results	FRF 1m LIBOR, weekly, Jan 3, 89-Apr 19, 00	MAP (inf)	0.0076	(2.01)**	-0.1026	-(1.21)	3.4625	(7.90)***	1.9155	(36.54)***
Treepongkaruna and Gray (2003)	FRF 1m LIBOR, weekly, Jan 3, 89-Apr 19, 00	QMLE	0.0054	(0.51)	-0.1277	-(0.41)	2.7218	(7.09)***	1.7902	(72.92)***

Study	Dataset	Method	α		β		σ		ξ	
Nowman (1998)	FRF 1m Euro-Curr., monthly, Feb 81-Mar 95	Gauss	-0.0300	-(1.41)	0.5052	(1.65)*	25.6755	(1.27)	2.8259	(16.52)***
My results	DEM 1m LIBOR, weekly, Jan 2, 86-Apr 19, 00	MLE	0.0067	(1.36)	-0.1890	-(1.67)*	0.1679	(5.45)***	1.0472	(15.79)***
My results	DEM 1m LIBOR, weekly, Jan 2, 86-Apr 19, 00	MAP (inf)	-0.0002	-(0.06)	-0.0395	-(0.51)	0.1497	(4.88)***	1.0046	(15.14)***
Treepongkaruna and Gray (2003)	DEM 1m LIBOR, weekly, Jan 2, 86-Apr 19, 00	QMLE	0.0052	(0.74)	-0.1066	-(0.70)	0.1591	(5.16)***	0.9431	(28.71)***
My results	ITL 1m LIBOR, weekly, May 17, 90-Apr 19, 00	MLE	-0.0007	-(0.12)	-0.0633	-(0.54)	0.7091	(9.50)***	1.4393	(32.69)***
My results	ITL 1m LIBOR, weekly, May 17, 90-Apr 19, 00	MAP (inf)	-0.0026	-(0.54)	-0.0322	-(0.41)	0.6975	(9.44)***	1.4214	(32.57)***
Treepongkaruna and Gray (2003)	ITL 1m LIBOR, weekly, May 17, 90-Apr 19, 00	QMLE	0.0049	(0.71)	-0.1582	-(0.88)	0.8393	(14.06)***	1.4598	(106.17)***
Nowman (1998)	ITL 1m Euro-Curr., monthly, Feb 81-Mar 95	Gauss	0.0420	(1.07)	-0.2820	-(0.75)	3.8662	(1.50)	2.2020	(13.54)***
My results	CHF 1m LIBOR, weekly, Jan 2, 86-Jul 14, 11	MLE	0.0004	(0.69)	-0.3866	-(2.00)**	0.2604	(6.19)***	0.8576	(27.58)***
My results	CHF 1m LIBOR, weekly, Jan 2, 86-Jul 14, 11	MAP (inf)	0.0004	(0.90)	-0.0738	-(0.78)	0.2712	(6.39)***	0.8632	(28.02)***
Treepongkaruna and Gray (2003)	CHF 1m LIBOR, weekly, Jan 2, 86-Apr 19, 00	QMLE	0.0103	(1.21)	-0.2535	-(1.20)	0.0500	(5.00)***	0.3255	(11.00)***
My results	AUD 90d BAB, weekly, Jan 1, 86-Jul 14, 11	MLE	0.0053	(1.10)	-0.1153	-(1.40)	0.1651	(9.92)***	1.0624	(27.08)***
My results	AUD 90d BAB, weekly, Jan 1, 86-Jul 14, 11	MAP (inf)	0.0021	(0.51)	-0.0518	-(0.76)	0.1627	(9.71)***	1.0538	(26.66)***
Treepongkaruna and Gray (2003)	AUD 90d BAB, weekly, Jan 28, 76-Apr 19, 00	QMLE	0.0038	(0.42)	-0.0181	-(0.15)	0.7957	(8.21)***	1.5519	(60.50)***
Brailsford and Maheswaran (1998)	AUD 30d BAB, weekly, Jul 76-Dec 95	QMLE	0.0234	(1.07)	-0.1872	-(0.81)	0.3877	(1.21)	1.1148	(3.45)***
Sanford and Martin (2006)	AUD 90d BAB, weekly, Jan 1, 90-Jul 26, 00	Post. mean	0.0685	–	-0.0124	–	0.0055	–	0.6692	–

The ratio $-\alpha/\beta$ describes the mean reversion level. There is less evidence for a non-zero α than there is for a non-zero β. For the ten year observation window, α is only significant at the 5%-level for monthly observations of the Swiss franc and weakly significant at the 10%-level for monthly observations of the Japanese yen and daily observations of the French franc. For the 25 year observation window, there is more evidence for a non-zero α: it is highly significant at the 1% level for all observation frequencies of the Japanese yen and significant at the 5%-level for daily observations of the Swiss franc. Furthermore, it is also weakly significant at the 10%-level for weekly observations of the British pound and monthly observations of the Australian dollar.

The results from the Bayesian approach largely agree with those obtained with maximum likelihood. In fact, the maximum a-posteriori estimates obtained with the flat priors are equal to those obtained with maximum likelihood. Under the informative priors, the estimated values differ slightly, but the significance of the estimates also agrees in all cases.

A detailed comparison of the estimation results to those of previous studies is presented in Table 3.6. The estimated values of the drift parameters are in line with the results found in literature. Overall, there is little evidence of mean reversion. The estimates for α presented by Treepongkaruna and Gray (2003) are all insignificant, while Nowman (1998) and Chan et al. (1992) find α to be weakly significant for the US dollar. In comparison, my estimates of α are insignificant for almost all time series, save the Japanese yen. For the British pound and the French franc, one of the two presented estimates is significant. While the exact results differ due to the different time series under study, the generally weak evidence of a non-zero α is confirmed. Similarly to the results for α, the estimates of the mean reversion parameter β are mostly insignificant. Treepongkaruna and Gray (2003) find again no evidence of a non-zero β, while Nowman (1998) reports weakly significant β's for the US dollar. My results for β are mixed: for the US dollar the MAP estimate of β is significant at the 5% level while the maximum likelihood estimate is not significant. For the Swiss franc the maximum likelihood estimate of β is significant at the 5% level while the MAP estimate is not significant. Otherwise, there is no evidence of a non-zero β, which is in line with the results from other studies.

Volatility parameters

In contrast to the drift parameters, the volatility parameters σ and ξ are, with the single exception of monthly observations over ten years for the Australian dollar, highly significant across currencies, observation horizons, and observation frequencies. This result is expected and consistent with the results documented in previous studies.

I am particularly interested in the level effect parameter ξ, for which I estimate values between 0.1693 (daily observation of the Australian dollar over ten years) and 1.9290 (weekly observations of the French franc over ten years). For the 25 year observation window, the estimates for the level effect parameter lie between 0.4068 (monthly observations of the US dollar) and 1.3047 (daily observations of Pound Sterling). It follows that most estimated values for the level effect parameter lie between 0.5 and 1.5 and could therefore be approximated by one of the restricted models. These observations make it already apparent that support for the Vasicek model, which sets ξ to zero, is weak.

For many time series, the estimates of the level effect parameter are fairly consistent across observation frequencies, with the estimate of ξ on time series with daily, weekly, and monthly observations falling typically within a range of 0.2. There are, however, a few exceptions where this is not the case. The starkest example is that of the British pound over 25 years, where estimates of the level effect parameter vary between 0.5075 for monthly observations and 1.3047 for daily observations. The estimates made for the ten year and 25 year window also appear roughly consistent, but there is still some variation. The largest difference appears for the Australian dollar, where the estimated level effect parameter lies between 0.1693 and 0.3901 for the ten year observation window and between 1.0490 and 1.2001 for the 25 year window. The differences for the other time series are considerably smaller.

The results obtained from the maximum a-posterior estimates are again quite similar to those obtained with the maximum likelihood estimator. While the posterior distributions obtained from the MCMC simulations suggest more uncertainty in the volatility parameters, this is not expressed in the significance levels of the estimates. Furthermore, the maximum likelihood estimates are again identical to the MAP estimates obtained with the flat prior.

Similarly to the drift parameters, the estimation results for the volatility parameters are also compared to those from the literature in Table 3.6. In contrast to the drift parameters, the level effect parameter ξ is highly significant in all studies. The z-scores from my estimates are quite close to those presented by Treepongkaruna and Gray (2003). One difference between my results and those of Nowman (1998) and Chan et al. (1992) is that these authors often find σ to be insignificant. In contrast, my estimates of σ are almost always highly significant and have similar z-scores to those presented by Treepongkaruna and Gray (2003).

Table 3.7.: This table shows likelihood ratio and Bayes factor test statistics for the four models nested within the CKLS model and estimated on the ten year observation window. Likelihood ratio statistics are asymptotically χ^2_1 distributed. Stars either indicate significance at the 10%, 5%, or 1% level for one, two, or three stars, or the amount of evidence against a model according to Table 3.1.

Timeseries	Vasicek			CIR			BS			CIR-VR		
	LR	BF 1	BF 2	LR	BF 1	BF 2	LR	BF 1	BF 2	LR	BF 1	BF 2
USD, 10y, d	2298.2***	2846.5***	2840.3***	427.3***	459.4***	452.7***	103.0***	95.8***	87.5***	1792.8***	1755.5***	1744.7***
USD, 10y, w	317.1***	382.4***	378.9***	61.8***	60.5***	53.6***	11.3***	5.6*	-2.0	610.8***	217.6***	212.2***
USD, 10y, m	204.1***	130.6***	115.4***	209.1***	22.9***	13.3***	106.1***	-3.4	-11.1	572.0***	51.6***	42.5***
GBP, 10y, d	3246.3***	6034.5***	6070.8***	1542.3***	2328.2***	2341.0***	302.6***	350.7***	348.2***	113.1***	102.2***	92.6***
GBP, 10y, w	372.9***	587.6***	596.3***	140.9***	174.5***	173.0***	6.7***	0.6	-6.7	75.6***	65.9***	57.3***
GBP, 10y, m	51.8***	65.2***	56.3***	13.0***	10.0**	2.1*	0.4	-3.4	-11.1	31.6***	25.1***	16.5***
JPY, 10y, d	7591.5***	8455.4***	8501.8***	2165.4***	2013.8***	2017.4***	3.5*	-4.4	-12.1	1.1e10***	2400.9***	2413.4***
JPY, 10y, w	1221.4***	1274.9***	1209.7***	445.0***	399.4***	373.6***	21.2***	13.9***	5.2*	96.6***	118.3***	104.4***
JPY, 10y, m	316.2***	342.3***	334.6***	136.6***	127.8***	120.0***	20.4***	14.6***	6.7**	12.0***	2.7*	-5.3
FRF, 10y, d	6149.2***	7785.6***	7691.0***	3623.8***	4173.8***	4118.9***	1581.3***	1677.3***	1650.2***	296.9***	296.2***	284.9***
FRF, 10y, w	791.2***	1237.2***	1296.1***	463.8***	656.8***	684.3***	200.8***	257.5***	263.6***	37.0***	39.2***	33.8***
FRF, 10y, m	97.7***	131.9***	120.6***	49.5***	58.7***	49.0***	14.9***	13.6***	5.0*	0.2	-3.4	-11.3
DEM, 10y, d	935.9***	976.1***	975.4***	117.0***	112.5***	106.2***	84.3***	82.4***	74.8***	766.3***	885.6***	880.9***
DEM, 10y, w	208.7***	213.1***	209.0***	53.2***	47.8***	41.3***	0.0	-4.9	-12.8	55.7***	55.1***	46.8***
DEM, 10y, m	23.9***	20.0***	14.3***	2.0	-2.0	-9.2	4.5**	0.8	-6.8	29.3***	28.4***	21.5***
ITL, 10y, d	6046.5***	8490.5***	8649.0***	2929.0***	3555.4***	3617.1***	687.6***	732.0***	738.3***	27.1***	20.3***	12.9***
ITL, 10y, w	691.2***	976.7***	952.8***	320.4***	390.6***	375.7***	65.4***	66.1***	56.5***	8.4***	3.1*	-4.7
ITL, 10y, m	180.9***	268.2***	254.5***	90.9***	110.1***	99.5***	22.7***	19.6***	10.9***	0.7	-3.5	-11.4
CHF, 10y, d	3025.3***	3686.3***	3636.1***	475.3***	515.8***	502.1***	233.5***	216.4***	206.2***	3353.7***	2788.2***	2748.3***
CHF, 10y, w	539.2***	740.7***	717.9***	106.3***	117.1***	106.2***	26.7***	19.2***	11.7***	517.4***	447.2***	434.2***
CHF, 10y, m	162.4***	212.9***	199.6***	48.1***	51.0***	40.8***	0.0	-4.8	-12.7	51.5***	45.5***	39.1***
AUD, 10y, d	7.3***	-0.2	-6.1	71.5***	67.0***	61.2***	389.9***	397.1***	388.4***	960.4***	990.2***	975.5***
AUD, 10y, w	4.1**	-0.6	-6.4	3.3*	-1.9	-7.9	33.4***	29.4***	26.5***	95.6***	93.2***	96.5***
AUD, 10y, m	0.5	-2.3	-8.3	1.2	-3.2	-9.7	8.4***	5.4*	0.6	22.2***	23.5***	22.6***
EUR, 10y, d	2141.9***	3219.9***	3133.4***	387.7***	448.8***	429.5***	199.0***	189.3***	177.6***	2480.0***	2441.4***	2377.5***
EUR, 10y, w	212.0***	271.2***	267.9***	17.9***	12.4***	5.2*	66.0***	57.4***	51.1***	419.0***	406.3***	405.5***
EUR, 10y, m	47.5***	55.4***	48.8***	5.3**	0.5	-7.1	10.6***	5.3*	-1.4	90.8***	69.6***	65.9***

Table 3.8.: This table shows likelihood ratio and Bayes factor test statistics for the four models nested within the CKLS model and estimated on the 25 year observation window. Likelihood ratio statistics are asymptotically χ^2_1 distributed. Stars either indicate significance at the 10%, 5%, or 1% level for one, two, or three stars, or the amount of evidence against a model according to Table 3.1.

Timeseries	Vasicek			CIR			BS			CIR-VR		
	LR	BF 1	BF 2	LR	BF 1	BF 2	LR	BF 1	BF 2	LR	BF 1	BF 2
USD, 25y, d	5873.0***	9390.4***	9335.3***	1125.3***	1373.9***	1359.4***	537.5***	500.0***	489.7***	8159.6***	6768.6***	6725.9***
USD, 25y, w	502.9***	681.9***	660.0***	24.8***	18.9***	11.1***	281.8***	247.0***	234.6***	1628.3***	1366.2***	1330.5***
USD, 25y, m	24.2***	25.6***	17.1***	7.6***	1.3	-5.3	153.4***	121.3***	108.4***	524.8***	385.8***	358.1***
GBP, 25y, d	7899.0***	14 006.3***	14 006.8***	3518.8***	5286.0***	5281.3***	615.7***	728.3***	720.8***	386.1***	333.1***	325.2***
GBP, 25y, w	637.2***	942.5***	957.3***	170.0***	204.8***	202.5***	2.7*	-3.5	-11.4	417.3***	317.4***	315.7***
GBP, 25y, m	45.1***	46.4***	39.8***	0.5	-5.5	-12.3	85.1***	72.0***	64.2***	336.4***	278.9***	269.4***
JPY, 25y, d	9009.6***	8018.5***	7893.3***	509.2***	497.2***	483.5***	1743.1***	1976.2***	1935.8***	11 902.2***	12 455.5***	12 250.2***
JPY, 25y, w	1335.9***	1156.7***	1146.5***	78.9***	67.5***	59.9***	247.8***	272.9***	265.2***	2005.8***	1772.9***	1762.2***
JPY, 25y, m	408.3***	390.1***	385.3***	23.5***	14.6***	7.6**	96.6***	106.7***	99.1***	467.2***	666.3***	659.6***
CHF, 25y, d	7517.5***	10 296.6***	10 501.1***	1457.7***	1712.9***	1741.6***	371.8***	333.7***	332.3***	9602.7***	6158.8***	6273.1***
CHF, 25y, w	1051.7***	1496.7***	1490.1***	186.5***	214.9***	206.9***	87.6***	72.8***	65.8***	1680.8***	1070.2***	1066.8***
CHF, 25y, m	174.1***	238.9***	226.2***	17.5***	13.1***	5.1*	52.9***	41.1***	33.2***	512.1***	323.1***	310.5***
AUD, 25y, d	6486.1***	6485.8***	6214.3***	2208.9***	2158.4***	2062.9***	153.9***	148.5***	134.4***	416.5***	456.0***	428.7***
AUD, 25y, w	700.4***	658.7***	662.3***	181.5***	166.9***	161.9***	0.2	-5.7	-13.6	133.2***	140.9***	135.9***
AUD, 25y, m	111.3***	97.0***	92.5***	22.2***	16.8***	9.8**	0.8	-3.6	-11.4	35.6***	35.8***	28.8***
DEM, 25y, d	5781.6***	9709.8***	9734.0***	1730.9***	2329.0***	2329.8***	3.3*	-5.2	-13.1	3598.2***	2707.4***	2705.4***
DEM, 25y, w	550.6***	883.9***	828.7***	84.0***	97.7***	84.5***	91.6***	80.6***	68.7***	1051.2***	832.4***	781.3***
DEM, 25y, m	21.1***	17.7***	12.0***	5.0**	-1.2	-7.6	106.2***	81.7***	77.8***	374.7***	266.4***	268.0***

Table 3.9.: This table compares the model test statistics from this chapter to other studies.

Study	Dataset	Test statistic	Vasicek	CIR	BS	CIR-VR
My results	USD 1m LIBOR, weekly,	LR (χ^2_1)	502.91***	24.81***	281.77***	1628.29***
	Jan 2, 86-Jul 14, 11	Bayes Factor	681.87***	18.90***	247.03***	1366.25***
Chan et al. (1992)	USD 1m T-Bills, monthly,	GMM Min.	8.84***	6.15***	2.22	–
	Jun 64-Dec 89	Crit. (χ^2_1)				
Treepongkaruna and	USD 1m LIBOR, weekly,	LR (χ^2_1)	683.00***	153.86***	–	–
Gray (2003)	Jan 2, 86-Apr 19, 00					
My results	GBP 1m LIBOR, weekly,	LR (χ^2_1)	637.21***	169.99***	2.74*	417.28***
	Jan 2, 86-Jul 14, 11	Bayes Factor	942.47***	204.82***	-3.53	317.41***
Treepongkaruna and	GBP 1m LIBOR, weekly,	LR (χ^2_1)	614.84***	141.78***	–	–
Gray (2003)	Jan 2, 86-Apr 19, 00					
My results	JPY 1m LIBOR, weekly,	LR (χ^2_1)	1335.92***	78.88***	247.78***	2005.82***
	Jul 1, 86-Jul 14, 11	Bayes Factor	1156.74***	67.54***	272.93***	1772.91***
Treepongkaruna and	JPY 1m LIBOR, weekly,	LR (χ^2_1)	713.82***	30 331.91***	–	–
Gray (2003)	Jan 9, 86-Apr 19, 00					
My results	FRF 1m LIBOR, weekly,	LR (χ^2_1)	791.23***	463.81***	200.77***	37.04***
	Jan 3, 89-Apr 19, 00	Bayes Factor	1237.16***	656.82***	257.50***	39.19***
Treepongkaruna and	FRF 1m LIBOR, weekly,	LR (χ^2_1)	4505.67***	2689.23***	–	–
Gray (2003)	Jan 3, 89-Apr 19, 00					
My results	DEM 1m LIBOR, weekly,	LR (χ^2_1)	208.70***	53.23***	0.05	55.69***
	Jan 2, 86-Apr 19, 00	Bayes Factor	213.06***	47.76***	-4.88	55.14***
Treepongkaruna and	DEM 1m LIBOR, weekly,	LR (χ^2_1)	790.92***	157.89***	–	–
Gray (2003)	Jan 2, 86-Apr 19, 00					
My results	ITL 1m LIBOR, weekly,	LR (χ^2_1)	691.19***	320.35***	65.43***	8.37***
	May 17, 90-Apr 19, 00	Bayes Factor	976.70***	390.60***	66.08***	3.14*
Treepongkaruna and	ITL 1m LIBOR, weekly,	LR (χ^2_1)	4580.46***	2453.28***	–	–
Gray (2003)	May 17, 90-Apr 19, 00					
My results	CHF 1m LIBOR, weekly,	LR (χ^2_1)	539.17***	106.33***	26.73***	517.41***
	Jan 2, 86-Jul 14, 11	Bayes Factor	740.70***	117.08***	19.25***	447.19***
Treepongkaruna and	CHF 1m LIBOR, weekly,	LR (χ^2_1)	249.47***	92.93***	–	–
Gray (2003)	Jan 2, 86-Apr 19, 00					
My results	AUD 90d BAB, weekly,	LR (χ^2_1)	700.39***	181.52***	0.24	133.20***
	Jan 1, 86-Jul 14, 11	Bayes Factor	658.68***	166.92***	-5.68	140.88***
Treepongkaruna and	AUD 90d BAB, weekly,	LR (χ^2_1)	2511.82***	980.99***	–	–
Gray (2003)	Jan 28, 76-Apr 19, 00					
Sanford and Martin	AUD 90d BAB, weekly,	Bayes Factor	2.07×10^{13}***	0.49	941.62***	1.67×10^{22}***
(2006)	Jan 1, 90-Jul 26, 00					

3.3.3. Model comparison

To study the fit of the different models, I present likelihood ratio statistics, Bayes factors, and posterior model probabilities. These results should help answer which models best describe the evolution of the short-rate and if some of the discrepancies from previous studies can be explained by uncertainty in the level effect parameter.

Tables 3.7 and 3.8 present the values of the likelihood ratio and Bayes factor test statistics. The likelihood ratio statistic is computed according to equation (3.14) with the CKLS model as the unrestricted model and the Vasicek, CIR, Brennan-Schwartz, and VR model as restricted models. Since only one parameter is restricted, the test statistic is asymptotically χ^2-distributed with one degree of freedom. Stars on the estimates indicate the statistic's significance based on the p-value of the statistic. One, two, and three stars indicate significance at the 10%, 5%, and 1%-level, respectively. In addition to the likelihood ratio statistics, I present a statistic based on the Bayes factor. This statistic, labeled in the tables as BF, is twice the logarithm of the Bayes factor $B_{0j} = \frac{p(\mathbf{r}|H_0)}{p(\mathbf{r}|H_j)}$. Here, H_0 stands for the unrestricted CKLS model and the Bayes factor can be interpreted as evidence in favor of this model. The alternative interpretation is that it presents evidence against the alternative H_j which is one of the restricted models. By applying the logarithm and multiplying by two, this statistic is on the same scale as the likelihood ratio statistic.[5] This approach was suggested in Kass and Raftery (1995), who also give a scale for assessing the amount of evidence presented by this statistic. This scale is reproduced in Table 3.2 and the asterisks in Tables 3.7 and 3.8 visualize the amount of evidence presented by the Bayes factor statistic according to this scale. One asterisk stands for some evidence against a restricted model, while three asterisks stand for very strong evidence.

The likelihood ratio and Bayes factor test statistics presented in Tables 3.7 and 3.8 lead to similar conclusions in most cases. In particular, for large values of the likelihood ratio (≥ 25) the Bayes factor produces similar values as the likelihood ratio statistic and is sometimes considerably larger. This means that where models are very clearly rejected, the two methods do not differ significantly. For lower values of the likelihood ratio statistic, the corresponding value for the Bayes factor is usually smaller, sometimes by a considerable amount. This result is due to two factors: first, the Bayes factor incorporates parameter uncertainty, so that models which might be clearly rejected under the likelihood ratio statistic still appear acceptable under the Bayes factor. This effect is visible when comparing the test statistics of the Bayes factors from informative

[5] If the posterior parameter distribution for the Bayes factor is an atom at the maximum likelihood estimates, then the two measures coincide.

priors (BF 1) and flat priors (BF 2). The second statistic incorporates more parameter uncertainty as it starts from a more diffuse prior, and its test statistic in these cases is always lower than the one based on the informative prior. Second, the Bayes factor automatically penalizes models which have more free parameters. As the CKLS model has one additional parameter, it will be penalized compared to a model with fewer parameters that fits the data equally well. Another observation is that both the likelihood ratio as well as the Bayes factor increase with observation frequency. With more frequently observed time series the models are rejected more strongly than with less frequently observed time series. This finding is in accordance to the estimates of the level effect parameter, where the standard errors are also much lower for more frequently observed time series, leading to less uncertainty in these parameters and therefore stronger rejection of alternative models.

The test statistics presented in Tables 3.7 and 3.8 give a mixed picture of the restricted models' performance. For daily observations over ten years, most restricted models are rejected when tested against the CKLS model, even when considering the Bayes factor statistic that was obtained with a flat prior (BF 2). This statistic allows the most uncertainty in the parameters, as it does not meaningfully restrict the parameter space a-priori. For all but two time series, the Japanese yen and the Australian dollar, all restricted models are rejected. For the Japanese yen, the Brennan-Schwartz model cannot be rejected when using the most lenient BF 2 statistic. The same is true of the Vasicek model for the Australian dollar. The results for the time series with daily observations over 25 years are similar: there is only a single time series where not all restricted models are rejected in favor of the unrestricted CKLS model. For the combined Deutschmark/Euro time series, the Brennan-Schwartz model produces a good fit.

For weekly observations, the restricted models fare a little better, at least over the ten year estimation window. For the US dollar, the British pound, and the Deutschmark, the Brennan-Schwartz model is not rejected. This is also true of the VR model for the time series of the Italian lira and for both the CIR and Vasicek model for the Australian dollar. On the other hand, over the 25 year observation period, only the time series for the British pound and the Australian dollar are fitted adequately by the Brennan-Schwartz model, while all restricted models are rejected for all other time series.

For most time series with monthly observations there is a restricted model that cannot be rejected. Over the ten year observation window, the US dollar, the British pound, and the Swiss franc are best fitted by the Brennan-Schwartz model, while for the Japanese yen, the French franc, and the Italian lira the VR model cannot be rejected. For the Deutschmark and Euro time series, neither the CIR nor the Brennan-Schwartz model can be rejected, while for the

Australian dollar this is true of both the Vasicek and the CIR models. Over the 25 year observation window, more time series are well fitted by the CIR model: for the US dollar, the British pound, and the Deutschmark/Euro time series this model cannot be rejected and there is not very strong evidence against it for the Swiss franc. The Australian dollar over this horizon is fitted adequately by the Brennan-Schwartz model. The only time series where all restricted models are rejected in this case is that of the Japanese yen.

If one were to constrain oneself to the restricted models which fix the level effect parameter, one way to choose among these would be to examine the posterior probabilities of each model given the data. These probabilities are presented in Table 3.10. The results from this table are similar to the ones presented before, except that now one of the restricted models must be chosen, as there is no option of rejecting them all.

As before, little support is given to the Vasicek model, the only time series for which it is the most probable model are daily observations of the Australian dollar over ten years. The CIR model would be chosen for monthly observations of the Deutschmark, the Australian dollar, and the Euro, as well as for weekly observations of the Australian dollar and the Euro. The Brennan-Schwartz model has the strongest support, being the model of choice for all observation frequencies of the US dollar and the Swiss franc. Furthermore, it has the highest posterior probability for weekly and monthly observations of the British pound, daily and weekly observations of the Japanese yen and the Deutschmark, and daily observations of the Euro. The VR model is the most probable model for all observation frequencies of the French franc and the Italian lira and also for daily observations of the British pound. Over the 25 year observation period, the CIR model has more support. It describes the Japanese yen best across all observation frequencies and also fits weekly and monthly observations of the US dollar and monthly observations of the British pound, the Swiss franc and the Euro/Deutschmark the best. The Brennan-Schwartz model consequently has a bit less support over this window, but it is still the most probable model for all observation frequencies of the Australian dollar, daily and weekly observations of the Swiss franc and the Euro/Deutschmark, as well as daily observations of the US dollar and weekly observations of the British pound. The VR model would only be chosen based on daily observations of the British pound.

My model test results are compared to the results of other studies in Table 3.9. I find that my results are roughly comparable to the results from Treepongkaruna and Gray (2003). Unfortunately, these authors only present likelihood ratio test statistics for the Vasicek and CIR model. The results presented here agree with their results in that these two models are strongly rejected across all time series compared in Table 3.9. When comparing my results to those presented by Chan et al. (1992) for the US dollar, I find that the test statistics tend to be much

		10y observations				25y observations			
Currency	Freq	Vasicek	CIR	BS	VR	Vasicek	CIR	BS	VR
	d	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
USD	w	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	m	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	d	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000
GBP	w	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
	m	0.0000	0.0013	0.9987	0.0000	0.0000	1.0000	0.0000	0.0000
	d	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
JPY	w	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	m	0.0000	0.0000	0.0025	0.9975	0.0000	1.0000	0.0000	0.0000
	d	0.0000	0.0000	0.0000	1.0000				
FRF	w	0.0000	0.0000	0.0000	1.0000				
	m	0.0000	0.0000	0.0003	0.9997				
	d	0.0000	0.0000	1.0000	0.0000				
DEM	w	0.0000	0.0000	1.0000	0.0000				
	m	0.0000	0.7675	0.2325	0.0000				
	d	0.0000	0.0000	0.0000	1.0000				
ITL	w	0.0000	0.0000	0.0000	1.0000				
	m	0.0000	0.0000	0.0000	1.0000				
	d	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
CHF	w	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
	m	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	d	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
AUD	w	0.3305	0.6695	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
	m	0.3301	0.6660	0.0039	0.0000	0.0000	0.0000	1.0000	0.0000
	d	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
EUR	w	0.0000	1.0000	0.0000	0.0000	0.0000	0.0004	0.9996	0.0000
	m	0.0000	0.9459	0.0541	0.0000	0.0001	0.9999	0.0000	0.0000

Table 3.10.: Posterior probabilities of the restricted models.

higher in my case, meaning the models are rejected more strongly based on the test statistics I present. On the other hand, the Bayes factor statistics I present are much lower than the extreme values presented by Sanford and Martin (2006) in the case where the models are rejected.

3.3.4. Model predictions test

To test model predictions, I estimate parameters of the CKLS model on a five year rolling window using both maximum likelihood estimators and the random-walk Metropolis-Hastings MCMC algorithm. For each time series of interest rates, I use the longest available sequence. To obtain the initial estimate, I use the data points comprising the first five years of this time series and compute the maximum likelihood estimate or the posterior parameter distribution using the MCMC algorithm. Using these estimates, I compute the transition density and the 0.01, 0.05, 0.1, 0.9, 0.95, and 0.99-quantiles for the time point of the next observation. I then compare these quantiles against the actual observation. This procedure is repeated for the next observation by advancing the estimation window one step until the end of the observed time series is reached. Due to the computational intensity of the method, I only use the time series with weekly

and monthly observation frequencies.

The results from this test are presented in Table 3.11. The first row for each time series indicates how many realizations fall below the 0.01, 0.05, and 0.1-quantile or above the 0.9, 0.95, and 0.99-quantile, respectively. The second row indicates the p-value, i.e. the probability of observing this many or more realizations below or above the respective quantiles.

For many time series, there are many more realizations below the 0.01 and above the 0.99-quantile than expected. The p-value for the observed number of values below the 0.01-quantile are below 0.01 for eight and below 0.05 for twelve of the 16 time series when ignoring parameter uncertainty. For the 0.99-quantile the p-value for the number of crossings is below 0.01 for four and below 0.05 for ten of the 16 time series. When parameter uncertainty is considered, the results are similar. The p-value for the number of crossings is below 0.01 for eight and below 0.05 for eleven of the 16 time series. For the 0.99-quantile, the p-value for the number of crossings is below 0.01 for three and below 0.05 for seven out of the 16 time series. These results indicate that the CKLS model with or without parameter uncertainty would be rejected at the 0.01-level for about half of the time series when using the 0.01-quantile as reference. At the 0.05-level, the model would be rejected for approximately three out of four time series.

In contrast, when considering the 0.05 and 0.95-quantiles, the number of crossings is often less than expected. The CKLS model with or without parameter uncertainty would only be rejected at the 0.05-level for a single time series, namely for the Italian lira with weekly observations. Using the 0.1 or 0.9-quantiles as reference point, no rejection occurs.

While the predicted transition densities are wider when parameter uncertainty is considered, this does not lead to a qualitative difference in the results of the tests. As Table 3.11 shows, there are fewer crossings of the quantiles in a number of cases when parameter uncertainty is considered, however, the qualitative result of the tests is largely the same. I must therefore conclude that while parameter uncertainty does lead to additional uncertainty when predicting the transition density of the short-rate, this additional uncertainty is not sufficient to predict the range of the realizations accurately. This indicates that the CKLS model does not generate sufficiently cautious transition densities, even when parameter uncertainty is considered.

Table 3.11.: This table shows the model test statistics described in Section 3.3.4. The value before the slash indicates the number of realizations below (above) the respective quantile while the number after the slash are the total number of observed realizations. The decimal below indicates the p-value of that many or more crossings under the assumptions described in section 3.2.5.

	0.01		0.05		0.1		0.9		0.95		0.99	
Timeseries	noPU	PU	noPU	PU	noPU	PU	noPU	PU	noPU	PU	noPU	PU
USD, w	32 / 1044 0.0000	31 / 1044 0.0000	40 / 1044 0.9685	39 / 1044 0.9780	51 / 1044 1.0000	50 / 1044 1.0000	47 / 1044 1.0000	46 / 1044 1.0000	25 / 1044 1.0000	24 / 1044 1.0000	18 / 1044 0.0203	18 / 1044 0.0203
USD, m	9 / 240 0.0008	8 / 240 0.0032	13 / 240 0.4240	13 / 240 0.4240	21 / 240 0.7705	17 / 240 0.9525	14 / 240 0.9920	13 / 240 0.9962	12 / 240 0.5413	9 / 240 0.8516	7 / 240 0.0112	6 / 240 0.0349
GBP, w	22 / 1044 0.0011	21 / 1044 0.0025	36 / 1044 0.9936	35 / 1044 0.9960	58 / 1044 1.0000	58 / 1044 1.0000	50 / 1044 1.0000	47 / 1044 1.0000	27 / 1044 1.0000	25 / 1044 1.0000	11 / 1044 0.4721	11 / 1044 0.4721
GBP, m	8 / 240 0.0032	8 / 240 0.0032	14 / 240 0.3157	11 / 240 0.6582	20 / 240 0.8333	17 / 240 0.9525	20 / 240 0.8333	17 / 240 0.9525	12 / 240 0.5413	7 / 240 0.9580	3 / 240 0.4308	2 / 240 0.6931
JPY, w	14 / 1044 0.1686	15 / 1044 0.1073	30 / 1044 0.9997	29 / 1044 0.9999	40 / 1044 1.0000	40 / 1044 1.0000	30 / 1044 1.0000	30 / 1044 1.0000	21 / 1044 1.0000	21 / 1044 1.0000	15 / 1044 0.1073	14 / 1044 0.1686
JPY, m	4 / 240 0.2206	3 / 240 0.4308	13 / 240 0.4240	11 / 240 0.6582	18 / 240 0.9239	14 / 240 0.9920	6 / 240 1.0000	6 / 240 1.0000	6 / 240 0.9819	6 / 240 0.9819	6 / 240 0.0349	6 / 240 0.0349
FRF, w	6 / 260 0.0482	6 / 260 0.0482	9 / 260 0.9060	9 / 260 0.9060	14 / 260 0.9974	12 / 260 0.9995	19 / 260 0.9449	17 / 260 0.9804	11 / 260 0.7549	10 / 260 0.8409	6 / 260 0.0482	6 / 260 0.0482
FRF, m	2 / 60 0.1212	2 / 60 0.1212	4 / 60 0.3527	3 / 60 0.5826	6 / 60 0.5628	6 / 60 0.5628	5 / 60 0.7290	4 / 60 0.8626	3 / 60 0.5826	2 / 60 0.8084	2 / 60 0.1212	2 / 60 0.1212
DEM/EUR, w	28 / 1044 0.0000	28 / 1044 0.0000	51 / 1044 0.5875	48 / 1044 0.7439	80 / 1044 0.9960	75 / 1044 0.9994	67 / 1044 1.0000	66 / 1044 1.0000	50 / 1044 0.6425	45 / 1044 0.8640	24 / 1044 0.0002	23 / 1044 0.0005
DEM/EUR, m	7 / 240 0.0112	6 / 240 0.0349	16 / 240 0.1500	16 / 240 0.1500	22 / 240 0.6975	19 / 240 0.8845	19 / 240 0.8845	17 / 240 0.9525	13 / 240 0.4240	10 / 240 0.7643	6 / 240 0.0349	4 / 240 0.2206
ITL, w	10 / 260 0.0003	9 / 260 0.0014	21 / 260 0.0219	21 / 260 0.0219	24 / 260 0.6901	24 / 260 0.6901	16 / 260 0.9893	14 / 260 0.9974	8 / 260 0.9502	7 / 260 0.9767	2 / 260 0.7342	2 / 260 0.7342
ITL, m	1 / 60 0.4528	1 / 60 0.4528	2 / 60 0.8084	2 / 60 0.8084	4 / 60 0.8626	2 / 60 0.9862	3 / 60 0.9470	2 / 60 0.9862	2 / 60 0.8084	2 / 60 0.8084	0 / 60 1.0000	0 / 60 1.0000
CHF, w	21 / 1044 0.0025	21 / 1044 0.0025	42 / 1044 0.9396	41 / 1044 0.9558	70 / 1044 0.9999	66 / 1044 1.0000	69 / 1044 1.0000	68 / 1044 1.0000	50 / 1044 0.6425	47 / 1044 0.7888	27 / 1044 0.0000	22 / 1044 0.0011
CHF, m	7 / 240 0.0112	4 / 240 0.2206	14 / 240 0.3157	13 / 240 0.4240	23 / 240 0.6166	20 / 240 0.8333	20 / 240 0.8333	17 / 240 0.9525	14 / 240 0.3157	12 / 240 0.5413	8 / 240 0.0032	5 / 240 0.0949
AUD, w	27 / 1044 0.0000	26 / 1044 0.0000	49 / 1044 0.6949	48 / 1044 0.7439	68 / 1044 1.0000	66 / 1044 1.0000	80 / 1044 0.9960	78 / 1044 0.9980	46 / 1044 0.8289	42 / 1044 0.9396	23 / 1044 0.0005	22 / 1044 0.0011
AUD, m	7 / 240 0.0112	6 / 240 0.0349	13 / 240 0.4240	13 / 240 0.4240	17 / 240 0.9525	15 / 240 0.9846	22 / 240 0.6975	19 / 240 0.8845	14 / 240 0.3157	12 / 240 0.5413	7 / 240 0.0112	5 / 240 0.0949

3.4. Conclusion

This chapter has presented a comprehensive analysis of how well the CKLS model and four models nested within it fit short-term interest rate data in nine currencies over two different estimation windows with three different estimation frequencies. To capture the effect of parameter uncertainty, I applied Bayesian techniques in addition to the classical maximum likelihood approach. Furthermore, a new estimation technique introduced by Aït-Sahalia (2002) was used, which allows to accurately approximate transition densities not available in closed form. To estimate Bayes factors and posterior parameter distributions, I introduced a new parametrization of the CKLS model which produces well mixing and quickly converging random walk Metropolis-Hastings Markov chains. Additionally, I tested the CKLS model out-of-sample against the presented short-term interest rates to assess whether it is sufficiently cautious in predicting the distribution of the short-term interest rate.

One of the objectives of this chapter was to study whether parameter uncertainty can explain some of the conflicting estimation results found in literature. To this end, I applied the Bayesian estimation methodology. While Bayesian estimates suggest more uncertainty in the parameters than when using classical maximum likelihood theory, they cannot explain the different estimates of the level effect parameter for different time series. Bayes factors and likelihood ratio tests lead to the same conclusions with regard to the level effect parameter. Estimation uncertainty can also not fully explain the evolution of the short-rate for many time series. The CKLS model produces insufficiently wide transition densities both when parameter uncertainty is considered and when it is ignored.

One important observation of this chapter is that all nested models which restrict the level effect parameter are almost always rejected when tested against the unrestricted CKLS model using either the likelihood ratio or Bayes factor test on time series with daily observations. This is the case even though the level effect parameter falls in the range between 0.2 and 1.9, which are broadly covered by the restricted models. However, the CKLS model usually provides much higher likelihood values, leading to the rejection of the restricted models. Furthermore, if one of the restricted models would be chosen, a number of time series would require a different model for different observation frequencies or observation intervals. This is of course somewhat disappointing from a modeling perspective, as it is generally not very plausible that the dynamics of a short-rate should be different when considering daily in lieu of weekly data. While the Brennan-Schwartz model has the most support of the restricted models, this observation makes it unconvincing to choose any particular model that restricts the level effect parameter. The inability of the CKLS model to predict plausible

ranges for the next realization of the short-rate exacerbates this finding. Given that the predicted ranges of the short-rate are generally too narrow, even when parameter uncertainty is taken into account, I must question the model's ability to accurately predict the evolution of the short-rate.

4. Lattice methods for pricing options under stochastic interest rates

4.1. Introduction

In this chapter, I discuss two different methods which can be used to price options in equity models with stochastic interest rates on a lattice. The first method is a bivariate trinomial tree which I extend to the CEV-CKLS model. The second method is an extension of the QUAD algorithm which allows the pricing of options in a Black-Scholes-Vasicek model.

4.1.1. Prior work

Numerical pricing methods have been important since the advent of modern asset pricing theory in form of the Black-Scholes-Merton model by Black and Scholes (1973) and Merton (1973). As analytical solutions similar to those put forth by these authors are only available for a few options in certain models, numerical pricing methods soon came into view. As these developments coincided with a dramatic increase in the availability of computing power, interest in numerical pricing methods exploded.

In general, valuation methods are based either on the evaluation of the risk-neutral expectation or on finding an approximate solution to a pricing partial differential equation (PDE). Monte Carlo methods focus on approximating the risk-neutral expectations and are very popular in practice. This popularity can be attributed to their intuitive nature and to the relative ease with which they can be implemented. Glasserman (2003) gives an excellent review, detailing the techniques used in Monte Carlo simulations. On the other hand, lattice methods such as finite difference or finite element methods are deeply rooted in classical numerical mathematics and have also been applied extensively to solving pricing PDEs. While these methods often give superior performance for more complicated options, they are considerably more difficult to implement and understand. Binomial and trinomial trees, as introduced by Cox, Ross, and Rubinstein (1979), are an alternative which are both easier to understand and implement. Lattice methods based on trees are in fact a special case

of the explicit finite difference method as shown for example in Brennan and Schwartz (1978) and Rubinstein (2000). While their theoretical features are not as attractive as those of more sophisticated finite difference methods, they are immensely popular among practitioners due to their intuitive nature and the relative simplicity of their implementation.

Pricing options with trees

Ever since the seminal paper by Cox, Ross, and Rubinstein (1979), binomial and trinomial trees have been popular instruments to price derivatives. Their early success can be attributed to their ability to price American options, for which no closed form solutions exist, and which are difficult to price with Monte Carlo methods. Technical problems preclude the approach by Cox, Ross, and Rubinstein (1979) from being a mathematically valid derivation of the continuous time results of Black and Scholes (1973) and Merton (1973). However, the fact that Brennan and Schwartz (1978) showed that trinomial trees are a special case of the explicit finite difference method lends support to the use of trees as a numerical method. Similarly, Rubinstein (2000) showed that a suitably parametrized binomial tree can be interpreted as a special case of the explicit finite difference method. Given the intuitive nature of tree methods, and the relative ease by which they can be implemented in computer programs, it is unsurprising that these methods have enjoyed widespread success among practitioners.

Binomial and trinomial trees were soon applied to processes other than the geometric Brownian motion. For example, in two popular papers, Hull and White (1994a) and Hull and White (1994b) detail how to price interest rate derivatives in the Hull-White and other interest rate models. One development that is of particular interest to this chapter is the development of trees for the constant elasticity of variance (CEV) and the Chan-Karolyi-Longstaff-Sanders (CKLS) models. This work was pioneered by Nelson and Ramaswamy (1990). There are several difficulties in constructing trees for these processes. The first issue is that the process is heteroskedastic which makes it difficult to find a recombining tree, a prerequisite for constructing trees that are computationally feasible. The second problem is that under some parameter constellations the process can attain zero and exhibit strong drift in the next step. Nelson and Ramaswamy (1990) solved the first problem by transforming the heteroskedastic process into a homoskedastic process. The second problem was not solved in their paper, and the solution eluded the research community for several years. Finally, Nawalkha and Beliaeva (2007) proposed a solution in which the grid size is adjusted so that the process can attain the zero-boundary exactly. In the following step, multiple node jumps are permitted so that all path probabilities can remain positive.

Another topic of interest to this chapter is the development of multivariate trees. This approach was pioneered by Boyle (1988), who discusses the construction of a bivariate binomial tree for two underlying geometric Brownian motions. Similarly, Kamrad and Ritchken (1991) examine the construction of multivariate, multinomial trees and argue specifically for the use of trees with multiple (more than two) branches. Much of the subsequent discussion of bivariate trees has focused on constructing trees for multi-variate interest rate processes or for equity models with stochastic volatility. A popular discussion of the construction of a bivariate tree for interest rate models can be found in Hull and White (1994b). A number of other works discuss the construction of bivariate trees for interest rate models, for example Brigo and Mercurio (2007). A paper that is of particular interest in this chapter is that of Hilliard, Schwartz, and Tucker (1996), which discusses the construction of a bivariate tree for a stock process under stochastic interest rates. While these authors also cover a broad class of models, they assume that the two processes are separable and do not address the issue of hitting the zero boundary. Furthermore, they limit their approach to binomial trees, which has the advantage of only requiring four path probabilities in each step, but converges slower than an approach using trinomial trees.

The QUAD algorithm

The QUAD algorithm was introduced by Andricopoulos et al. (2003). It is in a sense an evolution of tree-based lattice algorithms. The algorithm was introduced as a means of valuing quickly and accurately path dependent options. The idea of the algorithm is to split the time domain at those points in time where the path dependence of the option comes into play. As an example, consider a barrier option where the barrier is checked only at discrete points in time. To value this option, one would split the time domain at the dates where the barrier is checked and the domain of the underlying at the value of the barrier. These critical points in the underlying's domain can be augmented by an arbitrary number of additional points to improve precision. To evaluate the option's value at each point, the risk-neutral expectation is computed. This is usually done by computing the expectations numerically with quadrature methods which give the algorithm its name. Andricopoulos et al. (2007) extend this algorithm to cover multi-asset options and options with more complex payoffs. The method is also analyzed in Fusai and Recchioni (2007), who study the convergence of the QUAD algorithm for the pricing of discrete barrier options and discuss the empirical performance under different asset models. Another paper using this method is Milev and Tagliani (2010), where the authors derive a similar method and apply it to barrier options written on multiple underlyings. Furthermore, some interest in the computational aspects of the method has

been stirred, as the works by Tse, Thomas, and Luk (2009a) and Tse, Thomas, and Luk (2009b) show.

The main advantage of the QUAD algorithm is that the computation of the integrals via quadrature methods is both very fast and very accurate. This means that the number of points needed along the time domain is typically much lower than with alternative methods. This effectively reduces the number of dimensions across which the grid needs to be made finer by one. As a result, the QUAD algorithm is usually both more accurate and faster than a comparable tree or finite difference method. The drawback of this method is that to apply the algorithm, one needs access to the transition density of the underlyings. If this density is not available in either analytical or approximated form, one cannot apply the algorithm or has to rely on cruder methods to approximate the density, possibly losing any advantage the algorithm would have provided.

4.1.2. Motivation and contribution

This chapter contributes to the existing literature in several ways. First, I develop a two-dimensional trinomial tree for the CEV-CKLS model. The approach is similar to that in Hilliard, Schwartz, and Tucker (1996), in that it focuses on the development of a two-dimensional tree for an asset model with stochastic interest rates. However, the tree I develop is truly two-dimensional in the sense that the two processes need not be separable as required in Hilliard, Schwartz, and Tucker (1996). Second, the critical issue of hitting the zero boundary is explicitly accounted for in my methodology. Third, I extend the QUAD pricing algorithm to an asset model with stochastic interest rates. This approach is based on Andricopoulos et al. (2007), which I extend by allowing two underlyings where the evolution of one may explicitly depend upon the other.

4.2. Trinomial trees for the CEV-CKLS model

In this section, I detail the construction of a bivariate trinomial tree for the CEV-CKLS model. The risk-neutral dynamics of the model are given by the following stochastic differential equations:

$$dr_t = \kappa(\theta - r_t)dt + \sigma_r r_t^{\xi} dW_t^r \tag{4.1}$$

$$dS_t = r_t S_t dt + \sigma_S S_t^{\alpha} dW_t^S, \tag{4.2}$$

where W_t^r and W_t^S are two Brownian motions defined over a suitable probability space with instantaneous correlation $dW_t^r \cdot dW_t^S = \rho dt$.

Since both processes are heteroskedastic, it is difficult to construct a recombining tree directly for the processes in the given form. Making the tree recombining is important because it is a prerequisite for obtaining trees that are computationally simple.[1] Nelson and Ramaswamy (1990) found a solution to this problem by applying a one-to-one transformation to the process so that the transformed process has unit volatility. I also apply this transformation here.

To transform a heteroskedastic process Z to a unit volatility process, I apply the following function:

$$\phi(Z,t) = \int^{Z} \frac{du}{\sigma_Z(u,t)}. \tag{4.3}$$

Here, $\sigma_Z(u,t)$ is the volatility function of the diffusion process Z. Since both processes (4.1) and (4.2) have a volatility process of the form $\sigma(u,t) = \sigma u^p$, I transform both processes in the same way. However, I do need two different transformations, depending on the value of the exponent parameter p in the volatility function. For $p = 1$, the transformation function is $\phi(Z,t) = \frac{1}{\sigma}\ln(Z)$, while for $p \neq 1$ it is $\phi(Z,t) = \frac{z^{(1-p)}}{\sigma(1-p)}$.

Applying the transformation to equations (4.1) and (4.2), I obtain two new processes X and Y:

$$X_t = \phi(r_t,t) = \int^{r_t} \frac{du}{\sigma_r u^{\xi}} = \int_0^t \mu_X(X_t,t)dt + \int_0^t dW_t^r, \tag{4.4}$$

$$Y_t = \phi(S_t,t) = \int^{S_t} \frac{du}{\sigma_S u^{\alpha}} = \int_0^t \mu_Y(X_t,Y_t,t)dt + \int_0^t dW_t^S. \tag{4.5}$$

The new drift terms μ_X and μ_Y can be obtained by an application of Itô's lemma. I provide the full form of these functions in appendix B. Equations (4.4) and (4.5) clearly show that the new processes have unit volatility, which makes it easy to construct a recombining tree.

I will now examine the construction of a recombining grid for the two processes. There are two difficulties that need to be solved before constructing the grid. First, the drift term can become dominant causing transition probabilities to become negative when the interest rate process approaches zero or becomes very large. To solve this problem, I follow Nelson and Ramaswamy (1990) and Nawalkha and Beliaeva (2007) in permitting multiple jumps between nodes. Allowing multiple jumps enables the correct modeling of large drift terms and ensures that all transition probabilities stay positive. The second difficulty is that the process can attain zero when the exponent in the volatility term is less than one. The problem in this scenario is that a grid constructed with standard

[1] Nelson and Ramaswamy (1990) define a tree as *computationally simple* if the number of nodes per time step grows at most linearly as the number of time steps increases.

practices will extend below zero, and this can lead to unnecessary errors when computing derivative prices. Nawalkha and Beliaeva (2007) show how the grid can be constructed so that it reaches zero exactly. In the following section, I adapt their approach to this chapter's setting.

4.2.1. Constructing the grid

The construction of the grid is very similar for both the X_t and Y_t process. I will denote by x_t and y_t their discretized versions which result from the approximation on the tree. Since the construction is similar, the discussion below will focus on the x_t process. Whenever the construction of the grid differs for the two processes, both will be discussed separately.

To construct the grid of the x_t process, I discretize the process X_t both in time and space. I first create a grid in t, where I assume for simplicity that all time points are equidistant, i.e. that $\triangle t = t_{i+1} - t_i$ is constant for all i. The main purpose of this assumption is to simplify notation, the generalization to the case of non-equidistant time points is straightforward. The next step is to discretize the space domain of the process. The goal for the discretization in the space domain is to create a recombining grid on which the discretized process can match the mean and variance of the continuous process.

To construct this homogeneous grid, I add or subtract the multiple of a step size to the initial value x_0. The kth node is therefore given by

$$x_k = x_0 + k\triangle x, \tag{4.6}$$

where $k \in \mathbb{Z}$ denotes the number of steps above (below) the initial value x_0 and $\triangle x$ denotes the step size. At all time steps the value of the node with $k = 0$ equals the starting value of the process. The step size $\triangle x$ is computed as

$$\triangle x = b\sqrt{\triangle t}. \tag{4.7}$$

The value b is important for the convergence of the tree and is discussed in detail in the next section.

Choosing the step size

Choosing the right factor b in the step size is important to ensure that the grid can reach zero where necessary, while obtaining positive values for all transition probabilities. For trinomial trees a value of $b = \sqrt{3}$ is popular, this value is chosen in Hull and White (1994a) and Hull and White (1994b). Indeed, I will choose $b = \sqrt{3}$ in case $\xi \geq 1$ ($\alpha \geq 1$ for the y_t process). When $\xi < 1$ ($\alpha < 1$), the

x_t (y_t) process can reach zero. The problem here is that while the process could reach zero, it should not surpass it and go below zero. The question is therefore how to construct a grid which hits zero exactly. Nelson and Ramaswamy (1990) allowed the grid of the transformed process to go below zero, however, this can lead to biases in the derived prices, as was pointed out by Nawalkha and Beliaeva (2007). Therefore, I adapt the approach by Nawalkha and Beliaeva (2007), who modify the step size b in this case.

To handle the case where the process can attain zero correctly, I adapt the step size. To obtain the correct step size, I modify b so that it is close to $\sqrt{3}$, but allows the process to hit the zero boundary exactly. The idea is to slightly expand or contract the step size, so that the last downward step reaches zero exactly. To this end, I define the following helper variables b_c (contraction) and b_e (expansion):

$$b_c = \frac{x_0}{\sqrt{\triangle t}\left\lfloor \frac{x_0}{\sqrt{3\triangle t}} + 1 \right\rfloor}, \tag{4.8}$$

$$b_e = \frac{x_0}{\sqrt{\triangle t}\left\lfloor \frac{x_0}{\sqrt{3\triangle t}} \right\rfloor}. \tag{4.9}$$

With these definitions, I define the step size b so that it is closest to $\sqrt{3}$ in absolute terms:

$$b = \begin{cases} b_c & |b_c - \sqrt{3}| < |b_e - \sqrt{3}| \\ b_e & \text{otherwise} \end{cases}. \tag{4.10}$$

Given this definition, b will tend to $\sqrt{3}$ as the grid is made finer and $\triangle t$ decreases.

4.2.2. Transition probabilities

Now that I have defined a grid for the discretized process, I will consider the transition probabilities that describe its dynamics on this grid. The goal in choosing the transition probabilities is to ensure that the discretized processes converges to the continuous process in distribution. This convergence is achieved by matching the first two moments of the discretized process to those of the Euler-discretization of the continuous process. Let x be the value of the process at time t and denote by E the expected value and by V^2 the variance of the Euler-discretization of the continuous process, i.e.

$$E = X_t + \mu(X_t)\triangle t, \tag{4.11}$$

$$V^2 = \triangle t. \tag{4.12}$$

Choose k such that

$$k = \text{round}\left(\frac{E - x_0}{b\sqrt{\triangle t}}\right). \tag{4.13}$$

Equation (4.13) implies that x_k is the grid point closest to the expected value of the process, and I will choose this node as the middle node in the transition. Now consider the mean and the variance of the discretized process:

$$\mathbb{E}[x_{t+\triangle t} \mid x_t = x] = p_d x_{k-1} + p_m x_k + p_u x_{k+1} \stackrel{!}{=} E, \tag{4.14}$$

$$\text{Var}[x_{t+\triangle t} \mid x_t = x] = p_d x_{k-1}^2 + p_m x_k^2 + p_u x_{k+1}^2 - E^2 \stackrel{!}{=} V^2. \tag{4.15}$$

Denote by $\epsilon = \text{E} - x_k$ the distance between the middle node and the expected value of the process. Since the path probabilities should add up to one, I can substitute $p_m = 1 - p_d - p_u$. With these substitutions equations (4.14) and (4.15) then yield the following solutions for the path probabilities:

$$p_d = \frac{V^2}{2b^2\triangle t} + \frac{\epsilon^2}{2b^2\triangle t} - \frac{\epsilon}{2b\sqrt{\triangle t}} = \frac{1}{2b^2} + \frac{\epsilon^2}{2b^2\triangle t} - \frac{\epsilon}{2b\sqrt{\triangle t}}, \tag{4.16}$$

$$p_m = 1 - \frac{V^2}{b^2\triangle t} - \frac{\epsilon^2}{b^2\triangle t} = 1 - \frac{1}{b^2} - \frac{\epsilon^2}{b^2\triangle t}, \tag{4.17}$$

$$p_u = \frac{V^2}{2b^2\triangle t} + \frac{\epsilon^2}{2b^2\triangle t} + \frac{\epsilon}{2b\sqrt{\triangle t}} = \frac{1}{2b^2} + \frac{\epsilon^2}{2b^2\triangle t} + \frac{\epsilon}{2b\sqrt{\triangle t}}. \tag{4.18}$$

On the one hand, there is some degree of freedom in choosing the step size multiplier b. On the other hand, I need to ensure that all probabilities are non-negative. Because of the way I have chosen k in equation (4.13), x_k is no more than half a step size away from the mean, i.e. $\epsilon \leq \frac{1}{2}b\sqrt{\triangle t}$. For equation (4.16) (and analogously for (4.18)) I obtain:

$$\begin{aligned} p_d = \quad & \frac{1}{2b^2} + \frac{\epsilon^2}{2b^2\triangle t} - \frac{\epsilon}{2b\sqrt{\triangle t}} = \frac{1}{2b^2} + \frac{1}{2}\left(\frac{\epsilon}{b\sqrt{\triangle t}} - \frac{1}{2}\right)^2 - \frac{1}{8} \\ & \geq \frac{1}{2b^2} - \frac{1}{8} \\ \Rightarrow \quad & \frac{1}{2b^2} - \frac{1}{8} \stackrel{!}{\geq} 0 \\ \Leftrightarrow \quad & b \leq 2. \end{aligned} \tag{4.19}$$

Therefore, for p_d and p_u to be positive, b needs to be less than or equal to 2.

On the other hand, equation (4.17) yields:

$$\begin{aligned} p_m = \quad & 1 - \frac{1}{b^2} - \frac{\epsilon^2}{b^2 \triangle t} \geq 1 - \frac{1}{b^2} - \frac{1}{4} \overset{!}{\geq} 0 \\ \Leftrightarrow \quad & \frac{2}{\sqrt{3}} \leq b. \end{aligned} \tag{4.20}$$

Therefore, b needs to be chosen such that $\frac{2}{\sqrt{3}} \leq b \leq 2$. My choice of taking b close to $\sqrt{3}$ fulfills this requirement.

Adjusting probabilities at the zero boundary

In the case where $\xi < 1$ ($\alpha < 1$), the zero boundary can be attained by the x_t (y_t) process. As discussed above, the processes should reach the zero boundary exactly to avoid biases. When the process x_t approaches the zero boundary, its drift becomes infinite. However, this is not the case for the r_t process, and I thus replace the drift of x_t with that of r_t at zero.

In case x_t is at zero, I would like to avoid that the tree goes below zero. Therefore, I modify the way the index of the middle node is chosen. Define $\tilde{k}$ as

$$\tilde{k} = \max \left\{ k, 1 - \frac{x_0}{b\sqrt{\triangle t}} \right\}, \tag{4.21}$$

where k is chosen according to (4.13). This choice of the middle node $x_{\tilde{k}}$ ensures that all tree nodes remain non-negative.

To compute the probabilities, only the drift of the short-rate needs to be matched. For $r_t = 0$ this is simply $\kappa \triangle t$. If $\tilde{k}$ is the lowest permissible index and $\kappa \triangle t \leq x_{\tilde{k}}$, I set $p_u = 0$, $p_m = \frac{\kappa \triangle t}{x_{\tilde{k}}}$ and $p_d = 1 - p_m$. Otherwise, define $\tilde{\epsilon} = \kappa \triangle t - x_{\tilde{k}}$. By substituting $\tilde{\epsilon}$ for ϵ in equations (4.16), (4.17), and (4.18), I can compute the transition probabilities in the same way as when x_t is above zero.

For the processes y_t and S_t, this procedure is different, as S_t exhibits no mean reversion. Therefore, both the drift and volatility are zero, and once the process reaches zero, it will stay at zero forever.

Incorporating correlation into the transition probabilities

Now that I have defined the transition probabilities for the x_t process (and in principle for an independent y_t process), I need to define the probabilities of the y_t process such that they reflect the correlation between the two processes.

As before, I will match only the dynamics of the Euler-approximation of the process. I will denote the Euler-approximation of Y_t by $\tilde{Y}_t$. Given $\tilde{Y}_t$, $\tilde{Y}_{t+\triangle t}$ is a normally distributed random variable whose distribution is fixed at time t. Furthermore, $\tilde{Y}_{t+\triangle t}|\tilde{X}_{t+\triangle t}$ is also normally distributed with the following parameters:

$$\begin{aligned}\tilde{Y}_{t+\triangle t} \mid \tilde{X}_{t+\triangle t} &= \mathcal{N}\left(\mu_{\tilde{Y}} + \frac{\sigma_{\tilde{Y}}}{\sigma_{\tilde{X}}}\rho(\tilde{X}_{t+\triangle t} - \mu_{\tilde{X}}), (1-\rho^2)\sigma^2_{\tilde{Y}}\right) \\ &= \mathcal{N}\left(\mu_{\tilde{Y}} + \rho(\tilde{X}_{t+\triangle t} - \mu_{\tilde{X}}), (1-\rho^2)\triangle t\right).\end{aligned} \tag{4.22}$$

With this insight, I can now find the transition probabilities in much the same way as I did for the x_t process. First, compute the index l of the middle node as in equation (4.13), where $E_{\tilde{Y}}$ is the unconditional expectation of $\tilde{Y}_{t+\triangle t}$. Then define $\tilde{E}_j$ and $\tilde{V}_j^2$ as

$$\tilde{E}_j = \mathbb{E}\left[\tilde{Y}_{t+\triangle t} \mid \tilde{X}_{t+\triangle t} = x_j\right] \tag{4.23}$$

$$\tilde{V}_j^2 = \mathrm{Var}\left[\tilde{Y}_{t+\triangle t} \mid \tilde{X}_{t+\triangle t} = x_j\right] \tag{4.24}$$

for $j \in \{k-1, k, k+1\}$. To make the notation more intuitive, I label $k-1$, k, and $k+1$ as d, m, and u, respectively. Next compute $\tilde{\epsilon}_j = \tilde{E}_j - y_l$ and substitute the respective terms into equations (4.16)-(4.18). This yields the following transition probabilities:

$$p^y_{d|j} = \frac{1}{2b^2} + \frac{\tilde{\epsilon}_j^2}{2b^2\triangle t} - \frac{\tilde{\epsilon}_j}{2b\sqrt{\triangle t}}, \tag{4.25}$$

$$p^y_{m|j} = 1 - \frac{1}{b^2} - \frac{\tilde{\epsilon}_j^2}{b^2\triangle t}, \tag{4.26}$$

$$p^y_{u|j} = \frac{1}{2b^2} + \frac{\tilde{\epsilon}_j^2}{2b^2\triangle t} + \frac{\tilde{\epsilon}_j}{2b\sqrt{\triangle t}}. \tag{4.27}$$

One problem is that I cannot guarantee that all probabilities are greater than zero in all cases. For choices of the model parameters that are common in practice, this is likely the case, but one should briefly consider the edge cases. When the correlation parameter ρ is close to -1 or 1, probabilities may become negative. There are two cases to consider:

In the first case, the conditional expectation $\tilde{E}_j$ could either be greater than y_u or less than y_d. There is no easy solution to match the drift and the volatility in this case, and I simply set either $p_{u|j}$ or $p_{d|j}$ to one and all other probabilities to zero. This means that neither the conditional expectation nor the conditional variance are matched, but it is the best I can do within this setup.

The second case occurs when either $p_{u|j}$ or $p_{d|j}$ is less than zero, but the other

two probabilities are still greater than or equal to zero. As the probabilities need to add up to one, there is only one degree of freedom, and it is in general impossible to match both the conditional expectation and the conditional variance. In this case, I match the conditional drift, and accept that the conditional variance will not be perfectly matched.

While this approach might seem crude, it is more precise than other methods used to incorporated correlations in transition probabilities. For example, Hull and White (1994b) suggest to replace the transition probabilities with transition probabilities reflecting independence should their method of correlating the state variables lead to negative probabilities. This suggestion is also found in Brigo and Mercurio (2007). In the approach presented here, the transition probabilities are approximated at least as accurately as possible. Furthermore, one should keep in mind that these cases occur only in situations with extreme parameter values, which are not of much relevance in practice.

Having solved the equations for the conditional probabilities, I can now specify the matrix of the joint probabilities for x_t and y_t:

$$P_0 = \begin{pmatrix} p_{ud} & p_{um} & p_{uu} \\ p_{md} & p_{mm} & p_{mu} \\ p_{dd} & p_{dm} & p_{du} \end{pmatrix} = \begin{pmatrix} p_u^x p_{d|u}^y & p_u^x p_{m|u}^y & p_u^x p_{u|u}^y \\ p_m^x p_{d|m}^y & p_m^x p_{m|m}^y & p_m^x p_{u|m}^y \\ p_d^x p_{d|d}^y & p_d^x p_{m|d}^y & p_d^x p_{u|d}^y \end{pmatrix}. \tag{4.28}$$

Nelson and Ramaswamy (1990) show that discretized processes converge in distribution against their continuous-time counterparts if the following conditions are met: Both processes must start at the same value, the jump size of the discrete process must decrease sufficiently quickly as the time step size decreases, and the local mean and volatility must converge against the respective values of the continuous time process. The discretized process on the bivariate trinomial tree adheres to these requirements,[2] and thus converges in distribution to its continuous-time counterpart.

4.3. The QUAD algorithm for models with stochastic interest rates

While Section 4.2 focuses on extending a bivariate trinomial tree to the CEV-CKLS model, this section discusses how the QUAD method can be extended to a model with stochastic interest rates. The QUAD method bears some similarities to trees, but also differs from them in important aspects. The QUAD method is similar in that it seeks to compute the price of a derivative by computing

[2] This applies except where the probabilities need to be modified to stay positive. As discussed above, these situations should not occur very often in practice.

the risk-neutral expectation of its discounted payoff. To this end, similarly to tree methods, the underlying asset process is discretized in both time and space. However, the two methods differ in how the transition probabilities between nodes are computed. Tree methods construct explicit paths with fixed transition probabilities and then use these probabilities to compute risk-neutral expectations. The QUAD method computes the risk-neutral expectation by numerically evaluating the required integrals. This means that rather than explicitly deciding on specific paths and computing path probabilities, this is done implicitly while rolling back.

This method has a few advantages. Perhaps most importantly, it completely avoids discretization error with respect to time, as using numerical integration to compute the expectations allows one to use the exact transition densities. In contrast, binomial or trinomal trees use discrete distributions that converge slowly against the desired transition distribution. This has the important consequence that the QUAD method does not need the interval between time steps to become small to achieve convergence. On the contrary, convergence is independent of the size of the time steps.[3] Avoiding the necessity of a very fine grid of time points reduces the dimensions of the problem by one and improves the speed and memory requirements of the method significantly. Furthermore, approximating the transition distributions with numeric integrals is much more accurate than using discrete transition probabilities, which also reduces the need for very fine grids in the space dimension. Another advantage of this method is that the grids in the space dimension may be chosen with relative freedom. This means that one can easily align the grid, which does not need to be homogeneous, with critical points in the option's payoff. Trees or finite difference methods make it difficult or impossible to achieve such an alignment, especially for options where multiple points in the space dimension must be hit. Additional strong points of the method are that it is fairly easy to implement, easy to understand, and it is readily extensible to models with multiple underlyings.

The single strongest disadvantage of the QUAD method is that one needs closed form solutions of the underlying's transition density. This rules out using this method for the CEV-CKLS model presented in equations (4.1) and (4.2). One alternative where a density's closed-form solution is not available would be to use high precision approximations, for example the ones put forth in Aït-Sahalia (1999), Aït-Sahalia (2002), and Aït-Sahalia (2008). Another disadvantage of the method is that it works best for discretely monitored options, such as discrete Barrier options or Bermudan options. While Andricopoulos et al. (2007) present extrapolation methods to deal with continuously monitored options, such as

[3]The method does require the time steps to be chosen in such a way that the option's payoff can be approximated well. However, once such a grid is chosen, the time steps do not need to become smaller.

American options, it is unclear what the convergence characteristics are in these cases.

4.3.1. The QUAD method for one dimensional asset models

Andricopoulos et al. (2003) introduce the QUAD method to price discretely monitored options in a Black-Scholes model with a single stock process. The algorithm is generalized in Andricopoulos et al. (2007) to options with multiple underlyings, each driven by a geometric Brownian motion. The idea of the QUAD algorithm is based on computing the risk-neutral expectation of an option by discretizing both the time and space domain and approximating the risk-neutral expectation by numerically evaluating integrals where closed-form solutions are not available.

To make this precise, consider an asset process that follows a geometric Brownian motion

$$dS_t = rS_t dt + \sigma S_t dW_t, \tag{4.29}$$

where W_t is a standard Brownian motion, defined over a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with canonical filtration $\mathcal{F}_t$. As an example, I would like to price a simple European call option V whose payoff $v(S_T) = [S_T - K]^+$ solely depends on the value of the underlying at time T. It is well known that the time t price of the option may be computed as the discounted risk-neutral expectation of the option's payoff:

$$V_t = \mathbb{E}\left[e^{-r(T-t)} v(S_T) \mid \mathcal{F}_t\right]. \tag{4.30}$$

For the simple European call option with payoff $v(S_T) = [S_T - K]^+$, the risk-neutral expectation in equation (4.30) can be computed analytically. Denoting by f S_T's log-normal transition density from S_t, equation (4.30) can be written as

$$V_t = \int_0^\infty e^{-r(T-t)} v(S_T) f(S_T) dS_T. \tag{4.31}$$

To solve (4.31), one can substitute $e^y = \frac{S_T}{K}$, which yields

$$\begin{aligned} V_t &= \int_{-\infty}^{\infty} e^{-r(T-t)} K[e^y - 1]^+ e^{-y} f(Ke^y) dy \\ &= \int_{-\infty}^{\infty} e^{-r(T-t)} K[1 - e^{-y}]^+ g(y) dy, \end{aligned} \tag{4.32}$$

where g is the density of the normal distribution with mean $y_0 + (r - \frac{1}{2}\sigma^2)(T-t)$ and standard deviation $\sigma\sqrt{T-t}$. Solving the integral in (4.32) yields the well-known Black-Scholes formula.

While for a simple option, such as the European call under the Black-Scholes model, an analytical formula is available, this is seldom the case for more compli-

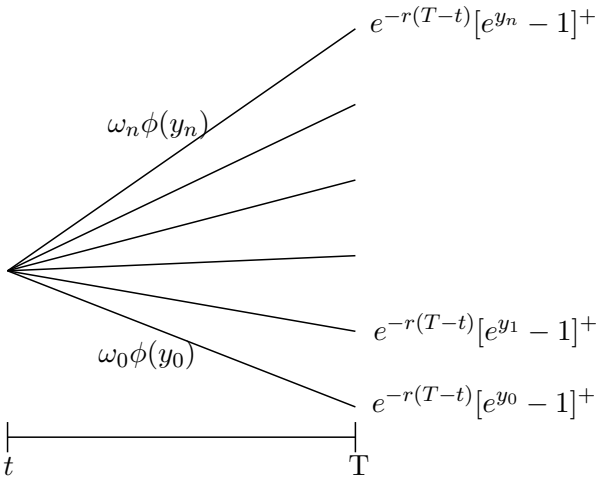


Figure 4.1.: Computing the discounted expectation via numerical integrals can be interpreted as a rollback step in a tree with $n+1$ branches. The path probabilities are given by the integration weight multiplied by the underlying's density.

cated options. In this case, the approach of the QUAD algorithm is to evaluate the integral in (4.32) numerically.

Using numeric integration to compute (4.32), one needs to construct a grid that covers the support of the integrand. Since the integrand vanishes for $y < 0$, a logical choice is to start the grid at zero. The integrand is unbounded above, so that the grid would theoretically have to reach out to infinity. One method to deal with such integrals is to transform the integrand to one with finite support. However, the normal probability density for the integrand in (4.32) decreases exponentially. This makes the integrand negligible for large values of y, and the grid can be cut off at a value where the integrand is deemed small enough to no longer contribute to the value of the integral.

To compute the value of the integral numerically, I split the interval into n subintervals $[y_0, y_1], [y_1, y_2], \ldots, [y_{n-1}, y_n]$ with $y_0 < y_1 < \ldots < y_n$. With a suitable quadrature rule one obtains integration weights $w_0, \ldots, w_n$ and can compute the approximate value of the integral as

$$V_T = \sum_{i=0}^{n} w_i \phi(y_i) e^{-r(T-t)} [e^{y_i} - 1]^+. \tag{4.33}$$

Figure 4.1 shows how this numerical integral can be interpreted as a tree with $n+1$ branches. However, in contrast to traditional tree methods, it is clear that the branches and their respective weights are chosen implicitly. This makes this approach more flexible than trees, where both the branches and their weights are fixed by the algorithm.

Discretely monitored options

The next step is to move from single period options to discretely monitored options. Discretely monitored options are options where the option's value depends on the value of the underlying at multiple points in time. Barrier options, where the option is *knocked out* or *knocked in* if the underlying crosses a certain threshold, are a good example of this.

To value such options, I first discretize the time domain at the time points where the option is monitored, and denote these points by $t = t_0 < t_1 < \ldots < t_m = T$. I will focus on options whose value can be computed on a tree without the introduction of new state variables. Such options include barrier and lookback options, but exclude Asian options. There is a large body of literature which discusses extensions of tree algorithms to price Asian options, see for example Klassen (2001), Dai, Huang, and Lyuu (2005), or Hsu and Lyuu (2007). While it is possible to adapt such methods to the QUAD algorithm, this would distract from the main point and will therefore not be covered here.[4] The value of such an option can be written as $v_0(S_{t_0}, V_{t_0})$, where V_{t_0} is the t_0 value of an option whose payoff only depends on the value of S_t at and after time t_1. V_{t_0}'s value can in turn be written as a function $v_1(S_{t_1}, V_{t_1})$, where again V_{t_1} is the time t_1 value of an option whose payoff is determined by S_t's values on or after time t_2. With repeated application of the tower law, one can write the risk-neutral value of such options as follows:

$$\begin{aligned} V_t &= \mathbb{E}[e^{-r(T-t)} V_T \mid \mathcal{F}_t] \\ &= \mathbb{E}[v_0(S_{t_0}, V_{t_0}) \mid \mathcal{F}_{t_0}] \\ &= \mathbb{E}[v_0(S_{t_0}, \mathbb{E}[v_1(S_{t_1}, \mathbb{E}[v_2(\ldots) \mid \mathcal{F}_{t_2}] \mid \mathcal{F}_{t_1}] \mid \mathcal{F}_{t_0}]. \end{aligned}$$

Writing the risk-neutral valuation formula in this way hints at how the option can be valued using the QUAD algorithm. First, the innermost expectation is evaluated by computing the respective integral using quadrature methods. Next, the second innermost expectation can be computed by building on the values of the innermost expectation, and so on, until finally the outermost expectation is evaluated, and the value of the option is computed. This method of valuing the option is similar to valuing an option on a tree, where the risk-neutral expectation in one step is also computed by building on the risk-neutral expectations computed in the previous steps. The important difference is that there is no need to make the time discretization arbitrarily fine, and that the choice of the position of the nodes is quite flexible.

Figure 4.2 illustrates how the grid is constructed, and demonstrates that the

[4] These methods are based on the introduction of a new state variable. I use this method in Chapter 5 to price implicit options of an insurance contract which depends on an additional state variable.

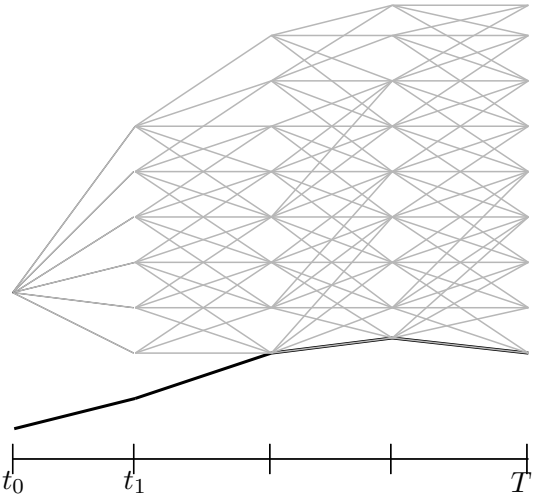


Figure 4.2.: The grid of the QUAD algorithm can be adapted to critical points on the options path. This Figure illustrates how the grid and be adapted for a discretely monitored barrier option. Adapting the grid so that nodes fall on the barrier improves the approximation's accuracy.

grid can be chosen quite flexibly, with grid points that do not need to be aligned between steps. This flexibility has the advantage that the location of grid points can be chosen to maximize the accuracy of the approximation. For example, for barrier options it is possible to align the start and end points of the grid at each time point exactly with the barrier values. This eradicates any non-linearity error introduced by a grid with points not exactly on the barrier. Avoiding this non-linearity error is difficult with trees or finite difference methods, especially when there are several different barrier points.

Obtaining a recombining grid

When constructing a grid for the QUAD algorithm, it is important to make this grid recombining for the same reason that it is important when constructing a grid for a tree: making the grid recombining is essential to make the computation efficient and to avoid an exponentially growing number of nodes. Depending on the quadrature rules used, there may be some additional flexibility in how the grid can be chosen, which can have several advantages. First of all, the number of nodes and the weights of each node are not fixed by the design of the algorithm or the grid. This means that the number of nodes can be chosen flexibly to obtain a desired accuracy, in particular the density of nodes can vary between time points or even in the space domain. Secondly, the nodes need not necessarily be distributed evenly in the space domain. While it is often desirable to transform the underlying processes so that they become homoskedastic, this is not required for the QUAD algorithm to work. Furthermore, it is often easier to reconcile a change in the underlying, such as a price

correction mandated by a dividend payment. While such a change typically leads to values off the grid, it is often possible to calculate with hypothetical values so that values after the adjustment are once again on the grid. This is possible because the transition probabilities are computed 'on the fly' and not fixed at construction time. The caveat with this approach is that many integration methods require a homogeneous grid. The following section discusses two rules which do not make this assumption.

Quadrature rules for fixed, non-homogeneous grids

In this section, I will briefly discuss two quadrature rules from the Newton-Cotes family. Both rules can work with fixed, non-homogeneous grids. Such rules are useful in the context of the QUAD algorithm, because each step's expected values need to be evaluated based on a grid from the previous step. Therefore, the choice of grid points is not entirely in the modeler's hands, and flexible quadrature rules to work with a given grid are needed.

The well known trapezoidal rule is the simplest quadrature rule in the Newton-Cotes family. An integral of the function f on the interval $[a, b]$ is simply approximated as:

$$\int_a^b f(x)dx = \frac{1}{2}(b-a)(f(a)+f(b)). \tag{4.34}$$

This formula lends itself easily to a composite integration rule. With nodes at $a = x_0 < x_1 < \ldots < x_n = b$ the above integral can be approximated by

$$\int_a^b f(x)dx = \frac{1}{2}f(a)(x_1-a) + \sum_{i=1}^{n-1}(x_i - x_{i-1})f(x_i) + \frac{1}{2}f(b)(b-x_{n-1}). \tag{4.35}$$

For each interval $[x_{i-1}, x_i]$, the error of approximating the integral by the formula in (4.34) is given by

$$\frac{1}{12}(x_i - x_{i-1})^3 f''(\xi), \tag{4.36}$$

where ξ is some intermediate value between x_{i-1} and x_i.

Next, I discuss a Newton-Cotes formula for a quadratic interpolation polynomial. This rule is an adaption of the popular Simpson rule to a non-homogeneous grid. The Simpson rule is the quadrature rule one obtains by using a quadratic polynomial in the Newton-Cotes formulas with equidistant supporting points:

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right). \tag{4.37}$$

Having equidistant supporting points is popular, because it allows the Simpson rule to give accurate integrals for polynomials up to fourth degree. However, this is not necessary and a similar formula can be derived for supporting points $a < c < b$:

$$\int_a^b f(x)dx \approx \frac{(a-b)(2a+b-3c)}{6(c-a)} f(a) + \frac{(a-b)^3}{6(a-c)(b-c)} f(c) + \frac{(a+2b-3c)(a-b)}{6(c-b)} f(b). \quad (4.38)$$

Similar to the trapezoidal rule, this formula can also be applied to several intervals, giving a composite non-homogeneous Simpson rule. Due to the extra factors, this rule is somewhat unwieldy, which is why I omit it here. As mentioned, the error estimate is slightly worse for the non-homogeneous Simpson rule when compared to the standard rule. The error can be computed as[5]

$$\text{error}_{\text{nSimp}} = K \cdot h^4 \left| f^{(3)}(\xi) \right|, \quad (4.39)$$

where h is the larger of the distances, $b - c$ and $c - a$ and K is a constant between $\frac{1}{12}$ and $\frac{2}{9}$. This error estimate is worse than the actual error

$$\text{error}_{\text{Simp}} = \frac{1}{90} \left(\frac{b-a}{2} \right)^5 \left| f^{(4)}(\xi) \right| \quad (4.40)$$

for the regular Simpson rule with equidistant points. However, for a sufficiently smooth integrand, this error is still better than that of the trapezoidal rule which may speed up the valuation of options.

Extending the QUAD Algorithm to a model with stochastic interest rates

After discussing the application of the QUAD algorithm to models with a single stochastic process, I want to now extend this algorithm to cover a model which includes stochastic interest rates. The extended model has the following form:

$$\begin{aligned} dr_t &= \kappa(\theta - r_t)dt + \sigma_r dW_t^r \\ dS_t &= r_t S_t dt + \sigma_S S_t dW_t^S, \end{aligned} \quad (4.41)$$

where again W_t^r and W_t^S are two Brownian motions with instantaneous correlation $dW_t^r \cdot dW_t^S = \rho dt$, defined over a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with canonical filtration $\mathcal{F}_t$.

[5] This is an application of the general error estimate for the Newton-Cotes formulas, see for example Schwarz and Köckler (2011, pp. 308-313) for a discussion.

I will now consider an option V with monitoring points $t_0 < \ldots < t_m$ that depends on r_t and S_t in a way that allows me to value it on a tree without the introduction of additional state variables. As in the one-dimensional case, the time t_0 value of such an option can be written as a function $v_0(r_{t_0}, S_{t_0}, V_{t_0})$, where V_{t_0} is the t_0 value of the future payoff of the option, which only depends on values of S_t and r_t at and after time t_1. Repeated application of the tower law allows me to write the option's time t_0 value as

$$\begin{aligned} V_t &= \mathbb{E}[e^{-\int_t^T r_u dr_u} V_T \mid \mathcal{F}_t] \\ &= \mathbb{E}[v_0(r_{t_0}, S_{t_0}, V_{t_0}) \mid \mathcal{F}_{t_0}] \\ &= \mathbb{E}[v_0(r_{t_0}, S_{t_0}, \mathbb{E}[v_1(r_{t_1}, S_{t_1}, \mathbb{E}[v_2(\ldots) \mid \mathcal{F}_{t_2}] \mid \mathcal{F}_{t_1}] \mid \mathcal{F}_{t_0}]. \end{aligned} \tag{4.42}$$

Equation (4.42) indicates that a valuation approach similar to the one in the one-dimensional case can be applied. The main difference is that the option's value now depends on both r_t and S_t. Therefore, a two-dimensional integral must be computed to evaluate the expectation in equation (4.42). Specifically, to evaluate $\mathbb{E}[v_i(r_{t_i}, S_{t_i}, v_{i+1}) \mid \mathcal{F}_{t_i}]$ I need to compute

$$\mathbb{E}[v_i(r_{t_i}, S_{t_i}, v_{i+1}) \mid \mathcal{F}_{t_i}] = \iint v_i(r_{t_i}, S_{t_i}, v_{i+1}(r, S)) f(r, S | r_{t_i}, S_{t_i}) d(r, S), \tag{4.43}$$

where f is the joint transition density of r_t and S_t from time t_i to t.

One way to evaluate the two-dimensional integral numerically is the iterated application of one-dimensional quadrature rules. The first step is to split the joint density of r and S into an inner and outer part, where the inner part is conditioned on the outer part. In an asset model with stochastic interest rates, it is natural to condition the density of the asset process on the interest rate process, as the drift of the asset process depends on the interest rate. For the Black-Scholes-Vasicek model in equation (4.41), the integral in (4.43) can then be written as

$$\begin{aligned} &\mathbb{E}[v_i(r_{t_i}, S_{t_i}, v_{i+1}) \mid \mathcal{F}_{t_i}] = \\ &\qquad \iint v_i(r_{t_i}, S_{t_i}, v_{i+1}(r, S)) f(S \mid S_{t_i}, r_{t_i}, r) f(r | r_{t_i}) dS dr, \end{aligned} \tag{4.44}$$

where $f(S|S_{t_{i+1}}, S_{t_i}, r)$ is S's transition density from S_{t_i} to S from time t_i to t_{i+1} conditional on r_t taking values r_{t_i} at time t_i and r at time t_{i+1}.

For the Black-Scholes-Vasicek model presented in equation (4.41), I proceed as follows. While it is not necessary to transform the processes to be homoskedastic, it is desirable to do so as it allows me to use a homogeneous grid. To further simplify the computations and reduce the book-keeping, I will scale the processes to have unit volatility. The interest rate process is already ho-

moskedastic so I only need to scale it by $\frac{1}{\sigma_r}$. To transform the asset process S_t, I apply the function given in equation (4.3), which for $p = 1$ is $\phi(Z_t) = \frac{1}{\sigma}\ln(Z_t)$. Applying this transformation and Itô's lemma to equation (4.29) I obtain

$$X_t = X_s + \int_s^t \kappa\left(\frac{\theta}{\sigma_r} - X_u\right)du + \int_s^r dW_u^r, \tag{4.45}$$

$$Y_t = \frac{1}{\sigma_S}\ln(S_t) = Y_s + \int_s^t \frac{\sigma_r}{\sigma_S}X_u - \frac{1}{2}\sigma_S du + \int_s^t dW_u^S. \tag{4.46}$$

To apply the QUAD algorithm, I need closed form solutions of both X_t's and Y_t's transition density. The transformed interest rate process X_t follows the Ornstein-Uhlenbeck process with known solution

$$X_t = X_s e^{-\kappa(t-s)} + \frac{\theta}{\sigma_r}(1 - e^{-\kappa(t-s)}) + \int_s^t e^{-\kappa(t-u)}dW_u^r. \tag{4.47}$$

To solve equation (4.46), I first need to find a solution for $\int_s^t X_u du$. Based on a convolution argument the integral over the Itô integral is once again an Itô integral and one obtains:

$$\int_s^t X_u du = \frac{\theta}{\sigma_r}(t-s) + \left(X_s - \frac{\theta}{\sigma_r}\right)\Psi(t-s) + \int_s^t \Psi(t-u)dW_u^r, \tag{4.48}$$

where

$$\Psi(x) = \frac{1}{\kappa}(1 - e^{-\kappa x}). \tag{4.49}$$

Substituting the solution from (4.48) into (4.47) I obtain for Y_t

$$\begin{aligned} Y_t &= Y_s + \frac{\theta}{\sigma_S}(t-s) + \frac{1}{\sigma_S}\left(\sigma_r X_s - \theta\right)\Psi(t-s) - \frac{1}{2}\sigma_S(t-s) \\ &\quad + \int_s^t \frac{\sigma_r}{\sigma_S}\Psi(t-u)dW_u^r + \int_s^t dW_u^S \\ &= Y_s + \frac{\theta}{\sigma_S}(t-s) + \frac{1}{\sigma_S}\left(\sigma_r X_s - \theta\right)\Psi(t-s) - \frac{1}{2}\sigma_S(t-s) \\ &\quad + \int_s^t \frac{\sigma_r}{\sigma_S}\Psi(t-u) + \rho dW_u^r + \sqrt{1-\rho^2}\int_s^t dW_u^Z. \end{aligned} \tag{4.50}$$

Here, W_t^Z is a new Brownian motion independent of W_r, so that $W_t^S = \rho W_t^r + \sqrt{1-\rho^2}W_t^Z$.

Equation (4.50) shows that X_t and Y_t are jointly normally distributed. Applying the one and two-dimensional Itô-isometry to compute the variance and covariance terms, I obtain the following distribution for X_t and Y_t:

$$X_t, Y_t \mid X_s, Y_s \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \eta_X^2 & \tau^2 \\ \tau^2 & \eta_Y^2 \end{bmatrix}\right), \tag{4.51}$$

where

$$\begin{aligned}
\mu_X &= X_s e^{-\kappa(t-s)} + \frac{\theta}{\sigma_r}(1 - e^{-\kappa(t-s)}) \\
\mu_Y &= Y_s + \left(\frac{\theta}{\sigma_S} - \frac{1}{2}\sigma_S\right)(t-s) + \left(\frac{\sigma_r}{\sigma_S}X_s - \frac{\theta}{\sigma_S}\right)\Psi(t-s) \\
\eta_X^2 &= \frac{1}{2\kappa}(1 - e^{-2\kappa(t-s)}) \\
\eta_Y^2 &= t - s + \frac{2\rho\sigma_r}{\kappa\sigma_S}(t - s - \Psi(t-s)) \\
&\qquad + \frac{\sigma_r^2}{\kappa^2\sigma_S^2}(t - s - 2\Psi(t-s) + \frac{1}{2\kappa}(1 - e^{-2\kappa(t-s)})) \\
\tau^2 &= \left(\rho + \frac{\sigma_r}{\kappa\sigma_S}\right)\Psi(t-s) - \frac{\sigma_r}{2\kappa^2\sigma_S}\left(1 - e^{-2\kappa(t-s)}\right).
\end{aligned}$$

With X_t and Y_t being jointly normally distributed the densities $g(x)$ and $g(y|x)$ can be written as

$$\begin{aligned}
g(X_t|X_s) &= \varphi(X_t; \mu_X, \eta_X^2), \\
g(Y_t|X_t, X_s, Y_s) &= \varphi(Y_t; \mu_Y + \frac{\tau^2}{\eta_X^2}(X_t - \mu_X), (1-\tau^2)\eta_Y^2).
\end{aligned} \tag{4.52}$$

Next, I need to construct a two-dimensional grid for the iterative application of the one-dimensional quadrature rules. To this end, I first discretize the interval for X into suitable intervals with intermediate points $x_0, \ldots, x_{n_X}$, and then discretize Y for each x_j. The grid points for Y do not have to be the same across all values of X, but this choice simplifies the implementation.

Applying this approach with quadrature rules giving integration weights w_j^X for the outer integral, and $w_{j,l}^Y$ for the inner integral, I obtain the following formula for the approximation.

$$\begin{aligned}
&\mathbb{E}[v_i(S_{t_i}, r_{t_i}, v_{i+1}) \mid \mathcal{F}_{t_i}] \\
&\quad = \iint v_i(S_{t_i}, r_{t_i}, v_{i+1}(S,r)) f(S, r|S_{t_i}, r_{t_i}) d(S,r) \\
&\quad = \iint \tilde{v}_i(X_{t_i}, Y_{t_i}, \tilde{v}_{i+1}) g(X, Y|X_{t_i}, Y_{t_i}) d(X,Y) \\
&\quad \approx \sum_{j=0}^{n_X} w_j^X \left(\sum_{l=0}^{n_Y} w_{j,l}^Y \tilde{v}_i(X_{t_i}, Y_{t_i}, \tilde{v}_{i+1}) g(Y_l \mid X_j, X_{t_i}, Y_{t_i}) \right) g(X_j, \mid X_{t_i}).
\end{aligned} \tag{4.53}$$

Here, $\tilde{v}_i(x, y, \tilde{v}_{i+1}) = v(\phi_r^{-1}(x), \phi_S^{-1}(y), \tilde{v}_{i+1})$ is simply a short-hand notation for the payoff function in terms of the transformed underlyings. Equation (4.53) shows that to compute the option's value at time t_i, I need the option's value at time t_{i+1} at the grid-points for X_t and Y_t. To compute the option's value

at time t_i, I start at the final time T where the option's payoff only depends on the values of X_T and Y_T, and then repeatedly solve equation (4.53) until I arrive at the desired point in time.

4.4. Computational examples

This section provides two computational examples to demonstrate the performance of, and the differences between, the bivariate trinomial tree and the QUAD algorithm. To analyze the performance of the algorithms, it is desirable to compare them for options for which closed-form pricing formulas exist. In this section, I examine the performance of the two methods for two different options. The first option studied is a classical European call option as introduced by Black and Scholes (1973), but for the model with stochastic interest rates driven by an Ornstein-Uhlenbeck process. This option and its closed form solution are presented in Section 4.4.1. The second option is the so-called traffic light option, a correlation product whose payoff depends on both the stock price and the interest rate. As the payoff depends on the value of both processes, so does the value of the option. The dependence on two processes makes the payoff of such options more intricate and difficult to price analytically. The traffic light option is a comparatively simple correlation product, for which a closed-form pricing formula was derived in Løchte (2007). This option and its closed-form solution are presented in Section 4.4.2. As the value of this option depends strongly on the interest rate, the stock process, and their correlation, it makes for an interesting target for comparing the two pricing methods. Section 4.4.3 presents examples and the setup used for the two methods. Both options and valuation methods are applied to a number of different parameters and settings, and the accuracy and performance figures are discussed.

4.4.1. European call

The first option considered is a classical European call option. This option is characterized by the payoff function

$$V(S,T) = [S_T - \bar{S}]^+, \tag{4.54}$$

where $\bar{S}$ is the option's strike. While the closed form solution for the European call's price introduced in Black and Scholes (1973) assumes a fixed interest rate, the model was soon generalized. Rabinovitch (1989) derives a closed-form solution when the interest rate is given by an Ornstein-Uhlenbeck process, as in

equation (4.1). The analytical solution for the time-t price of a European Call is given by

$$V_t = S_t\mathcal{N}(d_1) - \bar{S}P_t(T)\mathcal{N}(d_2), \tag{4.55}$$

where

$$\begin{aligned}
d_1 &= \frac{\ln\left(\frac{S_t}{\bar{S}\cdot P_t(T)}\right) + \frac{\tau}{2}}{\sqrt{\tau}}, \\
d_2 &= d_1 - \sqrt{\tau}, \\
\tau &= \sigma_S^2(T-t) + \frac{\sigma_r^2}{\theta^2}((T-t) - 2\Psi(T-t) + \frac{1}{2\theta}(1 - e^{-2\theta(T-t)})) \\
&\qquad - 2\frac{\rho\sigma_S\sigma_r}{\theta}(T - t - \Psi(T-t)).
\end{aligned}$$

Furthermore, $P_t(T)$ denotes the time-t price of a zero coupon-bond paying 1 unit of currency at time t, which is given by

$$\begin{aligned}
P_t(T) &= \exp\left(A(T-t) - r_t\Psi(T-t)\right), \\
A(T-t) &= \left(\theta - \frac{\sigma_r^2}{2\kappa^2}\right)(\Psi(T-t) - T + t) - \frac{\sigma_r^2}{4\kappa^2}\Psi(T-t)^2
\end{aligned}$$

and $\Psi(T-t)$ is as given in equation (4.49).

4.4.2. Traffic light option

Traffic light options were introduced by London based investment banks as a response to a new regulatory regime introduced by the Danish Financial Supervisory Authorities (DFSA). At the beginning of the last decade, Danish life insurers were increasingly substituting stocks for their more conservative interest-rate based assets to meet the promised payments of products with high interest rate guarantees. As a response, the DFSA introduced a new system with a *red light* and *yellow light* scenario, which put the insurers' balance sheets under considerable stress. Insurance companies who were not able to withstand these stress tests were placed under tight supervisory control and monitored closely. London based investment banks, such as Goldman Sachs, soon came up with strategies to allow insurance companies to escape such harsh regulatory monitoring: as both the red and yellow light scenario entail significant drops in both stock prices and interest rates, it is relatively inexpensive to insure oneself against such adverse effects with correlation products, namely with options whose payoffs dependent on both the asset value and the interest rate. Because the option only gains value in the precise setting that is tested by the supervisor, they are less expensive than two independent options which each have a payoff when only one process drops in value.

As one might imagine, the value of such correlation products is very sensitive to the correlation between the two processes, and they are generally difficult to price. However, Løchte (2007) derived an analytical pricing formula for a simple traffic light option under a Black-Scholes model with stochastic interest rates driven by the Vasicek model. The dynamics of the two processes are thus given by

$$\begin{aligned} dr_t &= \kappa(\theta - r_t)dt + \sigma_r dW_t^r \\ dS_t &= r_t S_t dt + \sigma_S S_t dW_t^S. \end{aligned} \tag{4.56}$$

Incidentally, this is the same model as that in equation (4.41), for which I have extended the QUAD-algorithm. Løchte (2007) derives a pricing formula for a European-style traffic light option which has the following time T payoff

$$V(r, S, T) = [\bar{r} - r_T]^+ \cdot [\bar{S} - S_T]^+, \tag{4.57}$$

where $\bar{r}$ and $\bar{S}$ are the respective strike levels for the interest rate and the asset process. The payoff of this option is thus the product of a standard European equity put option and an interest rate floorlet. As previously mentioned, a consequence of this payoff specification is that a payoff occurs only if both the interest rate and the asset process are below their respective strike levels at the maturity date.

Even with this fairly simple payoff, the valuation of this option is quite difficult. Løchte (2007) changes the numéraire and the risk-neutral measure twice to obtain a closed-form solution. The price of the option with payoff (4.57) is given by

$$\begin{aligned} C(r, S, t) = P(r, t; T) \cdot \Bigg[& (\bar{r} - m_r)\bar{S} \cdot M\left(\frac{\bar{r} - m_r}{v_r}, \frac{\ln \bar{S} - m_S}{v_S}; \rho_{rS}\right) \\ & + (m_r + \sigma_{rS} - \bar{r}) e^{m_S + \frac{1}{2}v_S^2} \cdot \\ & \quad M\left(\frac{\bar{r} - m_r}{v_r} - \rho_{rS} v_S, \frac{\ln \bar{S} - m_S}{v_S} - v_S; \rho_{rS}\right) \\ & + v_r \left(e^{m_S + \frac{1}{2}v_S^2} I(v_S) - \bar{S} I(0)\right) \Bigg], \end{aligned}$$

where

$$
\begin{aligned}
m_r &= r_t + \kappa(\bar{\theta} - r_t)\Psi(T-t) - \frac{\sigma_r^2}{2}\Psi^2(T-t) \\
v_r^2 &= \frac{\sigma_r^2}{2\kappa}\left(1 - e^{-2\kappa(T-t)}\right) \\
m_S &= \ln S_t + \left(\bar{\theta} - \frac{\sigma_r^2}{\kappa^2} - \frac{\rho\sigma_S\sigma_r}{\kappa} - \frac{1}{2}\sigma_S^2\right)(T-t) \\
&\qquad - \left(\bar{\theta} - \frac{\sigma_r^2}{\kappa^2} - \frac{\rho\sigma_S\sigma_r}{\kappa} - r_t\right)\Psi(T-t) + \frac{\sigma_r^2}{2\kappa}\Psi^2(T-t) \\
v_s^2 &= \left(\frac{\sigma_r^2}{\kappa^2} + \sigma_S^2 + \frac{2\rho\sigma_r\sigma_S}{\kappa}\right)(T-t) \\
&\qquad - \left(\frac{\sigma_r^2}{\kappa^2} + \frac{2\rho\sigma_r\sigma_S}{\kappa}\right)\Psi(T-t) - \frac{\sigma_r^2}{2\kappa}\Psi^2(T-t) \\
\sigma_{rS} &= \rho\sigma_S\sigma_r\Psi(T-t) + \frac{\sigma_r^2}{2}\Psi(T-t) \\
\rho_{rS} &= \frac{\sigma_{rS}}{v_r v_S} \\
I(\alpha) &= \int_{-\infty}^{\frac{\bar{r}-m_r}{v_r} - \alpha\rho_{rS}} u \cdot \mathcal{N}\left(\frac{\frac{\ln \bar{S} - m_S}{v_S} - \alpha - \rho_{rS}u}{\sqrt{1-\rho_{rS}^2}}\right)\varphi(u)du
\end{aligned}
$$

Here, φ denotes the standard normal density, $\mathcal{N}$ the standard normal cumulative distribution function, and Ψ is defined as in equation (4.49).

4.4.3. Pricing European options with trees and the QUAD algorithm

In the example presented in this section, I price a European call and a traffic light option with parameters similar to those found in Løchte (2007). In Table 1 of his paper, Løchte prices the traffic light option for maturities between 0.25 and 20 years and for correlation values ρ between -0.99 and 0.99. As I discuss two options and a number of other parameters, I focus on correlations of -0.25, 0, and 0.25, and options with a maturity of one year. The remaining parameters are chosen as in Løchte (2007), and take the following values: $\kappa = 0.25$, $\tilde{\theta} = 0.04$, $\sigma_r = 0.02$, $\lambda_r = -0.1$, $\sigma_S = 0.20$, $\bar{r} = r_0 = 0.03$, $\bar{S} = S_0 = 100$. Løchte presents $\tilde{\theta}$ in terms of the real-world measure, the corresponding risk-neutral reversion level based on a market price of risk of $\lambda_r = -0.1$ is $\theta = 0.048$.

Pricing the options on a bivariate trinomial tree

When pricing the options on the bivariate trinomial tree with parameters as shown above, a number of special cases apply. First of all, the bivariate model

is the Black-Scholes model with stochastic interest rate dynamics as in the Vasicek model. Furthermore, the interest rate follows an Ornstein-Uhlenbeck process which is already homoskedastic. This means that the transformation given in equation (4.3) only divides the process r_t by σ_r, scaling the volatility of the transformed X_t process to one. A further consequence of this choice of interest rate process is that it is legal for the interest rate to take negative values. Therefore, I do not need to ensure that X_t reaches zero exactly and can simply choose the step size as in (4.7) with $b = \sqrt{3}$.

The asset process follows a geometric Brownian motion which is a special case of equation (4.2) with $\alpha = 1$. The transformation from equation (4.3) to transform S_t to a homoskedastic process Y_t is therefore given by $Y_t = \phi(S_t, t) = \frac{1}{\sigma_S} \ln(S_t)$. The transformed process Y_t is the same process as the one presented in equation (4.46). As the transformed process Y_t goes to negative infinity when the actual process tends towards zero, it is not necessary to calibrate the step size so that the process can hit zero exactly. Therefore, I can also choose the step size $b = \sqrt{3}$ for the stock process.

Pricing traffic light options with the extended QUAD algorithm

To compute the value of the option, I need to obtain the transition densities for this model and decide upon a grid and a quadrature rule. As was previously shown, the transformed X_t and Y_t processes have a joint bivariate normal distribution, so that the conditional densities can be written as in equation (4.52). In the case of the traffic light option, equation (4.57) shows that the payoff is zero if either $S \geq \bar{S}$ or $r \geq \bar{r}$. This makes $\bar{S}$ and $\bar{r}$ two natural starting points for the grid for this option. For the European call option, the grid can only be cut off below the strike of stock process, as the payoff is unaffected by values of the interest rate. Cutting off the grid can significantly limit the space that the grid has to cover, and thus speed up computation times considerably. Where the grid cannot be cut off because the payoff of an option is zero, I extend it up to values where the transition density becomes very small. As both transition densities are normal, their tails decrease exponentially. I decide to cut off the grid for values more than six standard deviations away from the mean, as the cumulative distribution function for these values less than 10^{-9}, and, therefore, values beyond these make a negligible contribution to the option's value.

The payoff functions are smooth in both X_T and Y_T, and since the rest of the integrand also consists of smooth functions, the entire integrand is in C^∞. To take advantage of this smoothness, I select the Simpson rule as quadrature rule. Since there are no other points that the grid needs to hit, I choose an equidistant grid, taking advantage of the faster convergence of the quadrature procedure in this case.

Results

Tables 4.1 and 4.2 show the results for the European call and the traffic light option, respectively. The three major columns in each table show results for out-of-the-money (OTM), at-the-money (ATM), and in-the-money (ITM) options. The three major rows show results obtained with the closed-form solution, the bivariate trinomial tree, and the QUAD method. For each data point, I report the absolute price as computed by the respective method, the %-error with respect to the corresponding closed-form solution, and the time taken. I have implemented both pricing methods as well as the analytical formulas in Haskell. The computation times reported were obtained on an Intel Core i5-2520M with 2.5 GHz and 4GB of RAM, and the Glasgow Haskell Compiler[6] 7.4.2 with optimization options (-O2).

For the European call presented in Table 4.1, the strike for the OTM option is at $\bar{S} = 110$, at $\bar{S} = 100$ for the ATM option, and at $\bar{S} = 90$ for the ITM option.

The bivariate trinomial tree tends to underprice the option, as almost all computed prices are lower than those obtained with the closed-form solution. The only exception is the OTM option with $\rho = 0$ and $n = 96$ time steps. Pricing errors decrease as the number of steps increases from $n = 48$ to $n = 192$. Prices in the case where correlation is absent are fairly accurate, as the absolute value of the relative errors does not exceed 0.11%. For $|\rho| = 0.25$ errors remain higher, but are below 1% of the option's price.

The QUAD method gives generally more accurate approximations than the tree, except for the OTM option with a step size of $dx = 0.4$. The accuracy of the method improves rapidly as the step size decreases. Nevertheless, the QUAD method also has more problems approximating prices when $\rho \neq 0$. While prices are fairly accurately approximated for the smallest step size $dx = 0.1$, errors are larger when $\rho \neq 0$: for $\rho = 0$ the absolute value of the relative error does not exceed 0.04%, while for $\rho \neq 0$ it goes up to 0.16%. Nevertheless, I conclude that the QUAD method is superior to the tree, as it gives more accurate approximations and the errors also decrease quickly as the step size is decreased.

Focusing on the time taken to compute the results, it is apparent that the QUAD method beats the bivariate trinomial tree by several orders of magnitude. The bivariate trinomial tree takes between 180 and 210 seconds to compute results with the most accurate settings, while the QUAD method takes less than 0.05 seconds and still outperforms the tree in terms of accuracy. The practical implications of this difference are considerable, as it is all but impossible to use

[6] `http://www.haskell.org/ghc/`

the bivariate trinomial tree for tasks such as calibration, where options need to be priced multiple times with different parameters. The QUAD method would still be quite practical for such a task. The large performance difference can be attributed to the different complexities of the algorithms: the tree needs steps in the time domain to increase the number of steps in the space domain which gives it a complexity of $O(n^3)$, where n is the desired number of steps in one space dimension. The result of the $O(n^3)$ complexity is directly visible in the computation times: for each doubling of the number of time steps, the computation time increases by a factor of $2^3 = 8$. The QUAD method only has a complexity of $O(n^2)$, since it does not need additional time steps to generate a finer grid in the space domain. This can also be seen in the computation times which increase roughly by a factor of 4 as the step size is split in half, and the number of steps per space dimension doubles.

For the results of the traffic light option presented in Table 4.2, a similar picture emerges. Both methods have difficulty in pricing this option, which is reflected in higher relative errors. Especially for the OTM option, the bivariate trinomial trees gives large errors for all numbers of time steps. The QUAD method has also some difficulties in this case, as can be seen by the large errors for step sizes $dx = 0.4$ and $dx = 0.2$. With a step size of $dx = 0.1$ the QUAD algorithm achieves again reasonable accuracy, albeit with larger errors than for the European option. The increased difficulty in pricing this option can be attributed to the fact that the payoff of the traffic light option is more intricate than that of the European call. First of all, the payoff is only positive in one quadrant of the space domain. Furthermore, large payoffs only occur when both the interest rate and the stock price dip simultaneously far below the strike level. Since a large contribution to the option's value comes from a relatively small area in the space domain, both methods have more difficulty producing accurate price estimates. Their performance depends on how well the option's payoff is modeled in this part of the space domain.

The computation times for the traffic light option are comparable to those of the European call as the computations only differ in the payoff function.

4.5. Conclusion

In this chapter, I extended two existing option pricing methods to a bivariate model with an asset process and a stochastic interest rate. The derivation and implementation of both methods was discussed in detail. The bivariate trinomial tree was extended to cover the CEV-CKLS model, which is a generalized form of the Black-Scholes-Vasicek model allowing for heteroskedastic processes. To this end, I adapted the approach introduced by Nelson and Ramaswamy

			European call								
			OTM, $\bar{S} = 110$			ATM, $\bar{S} = 100$			ITM, $\bar{S} = 90$		
			Value	% error	time	Value	% error	time	Value	% error	time
Closed form		$\rho = -0.25$	5.253	0.00%	1.8×10^{-4}s	9.390	0.00%	1.4×10^{-5}s	15.438	0.00%	6.0×10^{-6}s
		$\rho = 0.0$	5.343	0.00%	6.0×10^{-6}s	9.480	0.00%	5.0×10^{-6}s	15.506	0.00%	6.0×10^{-6}s
		$\rho = 0.25$	5.433	0.00%	1.0×10^{-5}s	9.568	0.00%	5.0×10^{-6}s	15.574	0.00%	9.0×10^{-6}s
Tree	$n = 48$	$\rho = -0.25$	5.207	-0.89%	2.8×10^{0} s	9.314	-0.81%	2.7×10^{0} s	15.383	-0.36%	2.7×10^{0} s
		$\rho = 0.0$	5.323	-0.38%	2.7×10^{0} s	9.438	-0.44%	2.7×10^{0} s	15.491	-0.10%	2.7×10^{0} s
		$\rho = 0.25$	5.371	-1.13%	2.8×10^{0} s	9.477	-0.95%	2.7×10^{0} s	15.506	-0.43%	2.7×10^{0} s
	$n = 96$	$\rho = -0.25$	5.232	-0.41%	2.2×10^{1} s	9.334	-0.59%	2.2×10^{1} s	15.384	-0.35%	2.2×10^{1} s
		$\rho = 0.0$	5.349	0.10%	2.2×10^{1} s	9.459	-0.22%	2.2×10^{1} s	15.492	-0.09%	2.2×10^{1} s
		$\rho = 0.25$	5.397	-0.65%	2.2×10^{1} s	9.499	-0.72%	2.2×10^{1} s	15.508	-0.42%	2.2×10^{1} s
	$n = 192$	$\rho = -0.25$	5.225	-0.54%	1.8×10^{2} s	9.345	-0.48%	1.8×10^{2} s	15.398	-0.26%	1.8×10^{2} s
		$\rho = 0.0$	5.342	-0.02%	1.9×10^{2} s	9.469	-0.11%	1.9×10^{2} s	15.506	0.00%	1.8×10^{2} s
		$\rho = 0.25$	5.391	-0.76%	1.9×10^{2} s	9.509	-0.61%	1.9×10^{2} s	15.522	-0.33%	2.1×10^{2} s
QUAD	$dx = 0.4$	$\rho = -0.25$	5.174	-1.52%	8.5×10^{-3}s	9.399	0.09%	3.2×10^{-3}s	15.425	-0.09%	3.2×10^{-3}s
		$\rho = 0.0$	5.262	-1.52%	3.2×10^{-3}s	9.484	0.05%	3.2×10^{-3}s	15.488	-0.12%	3.2×10^{-3}s
		$\rho = 0.25$	5.350	-1.52%	3.3×10^{-3}s	9.568	0.01%	3.2×10^{-3}s	15.551	-0.15%	3.2×10^{-3}s
	$dx = 0.2$	$\rho = -0.25$	5.261	0.14%	1.2×10^{-2}s	9.343	-0.50%	1.2×10^{-2}s	15.457	0.12%	1.2×10^{-2}s
		$\rho = 0.0$	5.348	0.08%	1.2×10^{-2}s	9.429	-0.53%	1.2×10^{-2}s	15.520	0.09%	1.2×10^{-2}s
		$\rho = 0.25$	5.434	0.03%	1.2×10^{-2}s	9.514	-0.56%	1.2×10^{-2}s	15.582	0.05%	1.2×10^{-2}s
	$dx = 0.1$	$\rho = -0.25$	5.261	0.14%	4.4×10^{-2}s	9.382	-0.09%	4.4×10^{-2}s	15.448	0.06%	4.4×10^{-2}s
		$\rho = 0.0$	5.341	-0.04%	4.4×10^{-2}s	9.481	0.01%	4.5×10^{-2}s	15.506	0.00%	4.4×10^{-2}s
		$\rho = 0.25$	5.435	0.04%	4.5×10^{-2}s	9.552	-0.16%	4.5×10^{-2}s	15.573	0.00%	4.5×10^{-2}s

Table 4.1.: This table shows approximate prices, relative errors, and time taken when pricing a European call option in the Black-Scholes-Vasicek model. Relative errors for the tree and QUAD methods are computed with respect to the exact solution provided by the analytical pricing formula. Values and computation times are based on a Haskell implementation. The computations were performed on an Intel Core i5-2520M with 2.5 GHz and 4GB of RAM, using the Glasgow Haskell Compiler 7.4.2 with optimization options (-O2).

			Traffic light option								
			OTM, $\bar{r} = 0.02, \bar{S} = 90$			ATM, $\bar{r} = 0.03, \bar{S} = 100$			ITM, $\bar{r} = 0.04, \bar{S} = 110$		
			Value	% error	time	Value	% error	time	Value	% error	time
Closed form		$\rho = -0.25$	0.373	0.00%	5.9×10^{-4}s	2.633	0.00%	5.8×10^{-4}s	11.149	0.00%	1.4×10^{-3}s
		$\rho = 0.0$	0.809	0.00%	5.6×10^{-4}s	4.201	0.00%	5.6×10^{-4}s	14.691	0.00%	5.9×10^{-4}s
		$\rho = 0.25$	1.431	0.00%	5.7×10^{-4}s	6.039	0.00%	6.1×10^{-4}s	18.429	0.00%	6.8×10^{-4}s
Tree	$n = 48$	$\rho = -0.25$	0.372	-0.26%	3.0×10^{0} s	2.621	-0.44%	3.0×10^{0} s	11.169	0.18%	2.9×10^{0} s
		$\rho = 0.0$	0.796	-1.57%	2.9×10^{0} s	4.150	-1.21%	3.0×10^{0} s	14.621	-0.48%	2.9×10^{0} s
		$\rho = 0.25$	1.390	-2.88%	2.9×10^{0} s	5.921	-1.95%	3.1×10^{0} s	18.245	-1.00%	3.1×10^{0} s
	$n = 96$	$\rho = -0.25$	0.379	1.59%	2.4×10^{1} s	2.646	0.50%	2.4×10^{1} s	11.252	0.93%	2.4×10^{1} s
		$\rho = 0.0$	0.806	-0.43%	2.4×10^{1} s	4.176	-0.61%	2.4×10^{1} s	14.696	0.03%	2.4×10^{1} s
		$\rho = 0.25$	1.400	-2.14%	2.4×10^{1} s	5.945	-1.56%	2.4×10^{1} s	18.309	-0.66%	2.4×10^{1} s
	$n = 192$	$\rho = -0.25$	0.380	1.86%	1.9×10^{2} s	2.658	0.98%	1.9×10^{2} s	11.234	0.77%	2.0×10^{2} s
		$\rho = 0.0$	0.807	-0.28%	1.8×10^{2} s	4.189	-0.30%	1.9×10^{2} s	14.678	-0.09%	1.9×10^{2} s
		$\rho = 0.25$	1.401	-2.06%	1.8×10^{2} s	5.957	-1.36%	1.8×10^{2} s	18.292	-0.74%	1.8×10^{2} s
QUAD	$dx = 0.4$	$\rho = -0.25$	0.356	-4.51%	3.1×10^{-3}s	2.642	0.37%	3.0×10^{-3}s	10.998	-1.35%	3.1×10^{-3}s
		$\rho = 0.0$	0.791	-2.26%	3.1×10^{-3}s	4.216	0.35%	3.0×10^{-3}s	14.551	-0.95%	3.0×10^{-3}s
		$\rho = 0.25$	1.415	-1.09%	3.1×10^{-3}s	6.058	0.31%	3.0×10^{-3}s	18.307	-0.66%	3.0×10^{-3}s
	$dx = 0.2$	$\rho = -0.25$	0.378	1.38%	1.1×10^{-2}s	2.571	-2.32%	1.1×10^{-2}s	11.188	0.35%	1.1×10^{-2}s
		$\rho = 0.0$	0.818	1.04%	1.1×10^{-2}s	4.140	-1.46%	1.1×10^{-2}s	14.733	0.29%	1.1×10^{-2}s
		$\rho = 0.25$	1.443	0.84%	1.1×10^{-2}s	5.983	-0.92%	1.1×10^{-2}s	18.476	0.25%	1.1×10^{-2}s
	$dx = 0.1$	$\rho = -0.25$	0.375	0.43%	4.1×10^{-2}s	2.621	-0.43%	4.3×10^{-2}s	11.171	0.19%	6.5×10^{-2}s
		$\rho = 0.0$	0.811	0.21%	4.2×10^{-2}s	4.200	-0.02%	4.4×10^{-2}s	14.712	0.14%	4.6×10^{-2}s
		$\rho = 0.25$	1.436	0.35%	4.2×10^{-2}s	6.035	-0.07%	4.4×10^{-2}s	18.465	0.19%	4.6×10^{-2}s

Table 4.2.: This table shows approximate prices, relative errors, and time taken when pricing a European traffic light option in the Black-Scholes-Vasicek model. Relative errors for the tree and QUAD methods are computed with respect to the exact solution provided by the analytical pricing formula. Values and computation times are based on a Haskell implementation. The computations were performed on an Intel Core i5-2520M with 2.5 GHz and 4GB of RAM, using the Glasgow Haskell Compiler 7.4.2 with optimization options (-O2).

(1990) and Nawalkha and Beliaeva (2007) to the bivariate case. Care has been taken to specify transition probabilities which correctly account for the correlation between processes. The QUAD method introduced by Andricopoulos et al. (2003) and extended in Andricopoulos et al. (2007) has also been extended to cover the bivariate Black-Scholes-Vasicek model with stochastic interest rates.

Both pricing approaches were used to price two options for a number of settings and parameter constellations. In the two cases which were demonstrated, the QUAD method outperforms the bivariate trinomial tree both in terms of accuracy and computation time. Based on these findings, I conclude that this method is superior in this particular setting, namely for pricing options in a model where a closed-form transition density is available and no or only a few time steps are needed to price the option. Under these circumstances, the QUAD method is vastly superior to the bivariate trinomial tree. The large reduction in computation time can make the application of the QUAD method feasible in situations where the application of other methods, such as the bivariate trinomial tree, would be impossible or impractical.

Nevertheless, the bivariate trinomial tree still fills an important role, namely in pricing options in models such as the CEV-CKLS model, where no closed-form transition density is available and where the QUAD method cannot be applied. Another setting where the bivariate trinomial tree would prove to be useful is the pricing of options which require the introduction of many time steps, for example American options. While the QUAD method would still maintain some of its advantages, such as the flexibility in choosing the right grid, it would loose most of them in terms of computation time, thus rendering the bivariate trinomial tree a viable alternative.

5. Risk management of implicit options under uncertainty

5.1. Introduction

In this chapter, I examine the risk of participating life-insurance contracts with cliquet-style guarantees from the perspective of the insurance company. The main contribution to the literature is the modeling of estimation uncertainty with respect to both the asset and the interest rate processes and the analysis of its impact on contracts with cliquet-style guarantees. The motivation for analyzing the impact of uncertainty on implicit options in life insurance contracts is twofold. First, given the long duration of life insurance contracts, I expect that uncertainty about risk parameters has a strong impact on risk measures such as the value-at-risk. Second, the insurance companies' investment portfolio is difficult to replicate and implicit options, which depend on this portfolio, are difficult to hedge. My approach is similar to Hardy (2002), in that I apply Bayesian methods to quantify such effects and study them under two risk management paradigms. In contrast to Hardy (2002), I focus on a contract with cliquet-style guarantees and analyze two models that incorporate stochastic interest rates. With the help of a Markov chain Monte Carlo (MCMC) algorithm, I derive posterior distributions for the relevant model parameters. Using these distributions, I determine value-at-risk measures that take parameter and model uncertainty into account. I compare these measures to standard measures, which assume the maximum-likelihood estimates of the parameters to be correct.

The analysis is done for two risk management strategies. In what I call *static reserving*, the insurance company invests the premium into the risk-less asset and holds this portfolio until the expiration of the contract. In the dynamic hedging approach, the insurance company implements a monthly updated delta hedge. Under both risk-management approaches, I find that risk measures increase significantly when accounting for parameter and model uncertainty. Under static reserving, the risk measures increase between 403% and 626% when allowing for parameter uncertainty. This increase can go up to 880% when model uncertainty is also accounted for. When hedging dynamically, the risk-measures in the base case where parameter uncertainty is ignored are much lower than

when reserving statically. These lower risk measures increase by between 382% and 598% when parameter uncertainty is introduced, roughly comparable to the increase under static reserving. However, once model uncertainty is introduced, the risk measures increase dramatically, for example by 8175% for the insurance contract. These findings differ significantly from Hardy, who finds that only the static reserving approach is affected by parameter uncertainty and that the impact on the dynamic hedging approach is negligible.

Participating life insurance contracts have received considerable attention by researchers in recent years. These contracts usually feature an interest rate guarantee and a bonus distribution rule, which determines how return surpluses are divided between policyholders and the insurance company. There are two major types of guarantees. Point-to-point guarantees, which are common in Anglo-Saxon countries, offer a minimum interest rate on insurance premia that is paid until the maturity of the contract. In contracts with cliquet-style guarantees, which are prevalent in other markets (e.g. Germany), a portion of the surplus is credited to the policyholder each year, and the guarantee applies not only to premia but also to distributed surpluses. Contracts with cliquet-style guarantees have been analyzed in Grosen and Løchte Jørgensen (2000), Bacinello (2001), Bacinello (2003), Tanskanen and Lukkarinen (2003), Bauer et al. (2006), Kling, Richter, and Ruß (2007), and Zaglauer and Bauer (2008). On the other hand, contracts with point-to-point guarantees have been studied by Briys and De Varenne (1997) and Grosen and Løchte Jørgensen (2002), among others.

Most papers dealing with contracts with cliquet-style guarantees focus on deriving correct prices for these contracts. Using the risk-neutral valuation paradigm, they consider different bonus schemes as well as different valuation models and numerical valuation approaches. Two exceptions to this are the works by Kling, Richter, and Ruß (2007) and Gatzert and Kling (2007), which both focus on the risk imposed by such contracts and their implicit options and not on valuation. Kling, Richter, and Ruß (2007) analyze how different regulatory settings and the insurance company's solvency influence the likelihood that the insurance company will be able to meet its obligations. Gatzert and Kling (2007) examine the shortfall risk of insurance contracts that are priced using risk-neutral valuation but which are not hedged. The authors argue that insurance companies will not or cannot implement the perfect hedging strategies assumed when applying risk-neutral valuation. They find that interest rate guarantees impose a considerable shortfall risk to the insurance companies when contracts are not hedged and that contract parameter constellations that lead to equal risk-neutral contract values can lead to very different risk profiles. This finding is particularly pronounced for cliquet-style guarantees, which can lead to a high shortfall risk.

The Bayesian methodology gained popularity in insurance literature as it provides a framework to study the impact of model and parameter uncertainty. Its application was first suggested by Cairns (2000), who criticizes in his study that model and parameter uncertainty are typically ignored in risk management calculations in insurance. The study by Hardy (2002) applies this insight and uses the Bayesian approach to analyze the impact of parameter uncertainty on the risk management of guaranteed minimum benefit contracts. Similarly to this study, Hardy (2002) derives posterior parameter distributions for her model and studies the impact of parameter uncertainty on two risk management approaches. The study by Luoma, Puustelli, and Koskinen (2008) uses Bayesian methods to analyze the role of the underlying asset and interest rate model in pricing life insurance policies. While also concerned with parameter and model uncertainty, this study focuses on the impact on pricing and not on risk management.

The remainder of this chapter is structured as follows: Section 5.2 introduces the term-fix insurance contract and describes the embedded options and guarantees. Section 5.3 details the two asset models I consider, a Black-Scholes model with a Vasicek short-rate process, and a constant elasticity of variance model where the short-rate follows the model introduced by Chan et al. (1992). This section also introduces the numerical pricing algorithm I develop to derive prices of the insurance contract. This algorithm is based on the extended QUAD algorithm presented in chapter 4. Section 5.4 presents the estimation procedures and covers the MCMC algorithm used to obtain the uncertainty distributions. The risk management approaches are covered in Section 5.5. This section also presents the impact of parameter uncertainty on the two risk management approaches and analyzes the impact of different drivers of uncertainty. Finally, Section 5.6 concludes.

5.2. A simple term-fix insurance contract

The contract I consider is a simplified term-fix insurance with a cliquet-style interest rate guarantee. Tanskanen and Lukkarinen (2003) show how a class of such contracts can be analyzed and priced using risk-neutral valuation. Bauer et al. (2006) use this framework to analyze two contracts. The first represents the legal requirements in Germany, while the second stands for the actual practice followed by German life insurance companies. These contracts are further analyzed in Zaglauer and Bauer (2008), who value both contracts in an asset model with stochastic interest rates.[1] I focus on the contract described as the MUST case in Zaglauer and Bauer (2008), which is oriented on the features

[1]The model used by Zaglauer and Bauer (2008) corresponds to the *base model* in this study.

Assets	Liabilities
A_t	L_t
	R_t
A_t	A_t

Table 5.1.: Simplified balance sheet

required by German legislation. As my goal is to analyze the financial risk of the contract, I follow Tanskanen and Lukkarinen (2003), Bauer et al. (2006), and Zaglauer and Bauer (2008) in ignoring biometric risk in my calculations. Therefore, the contract I analyze offers a single payment L_T at maturity.

I model this contract with a simplified balance sheet presented in Table 5.1. On the asset side, A_t represents the current value of the asset portfolio held to support future liabilities from the contract. On the liability side, the current obligation to the policy holder is denoted by L_t while $R_t = A_t - L_t$ denote possible reserves resulting from additional assets held in excess of the policyholder's account value.

The fair value of the contract depends strongly on how gains in the asset portfolio A_t are distributed between the policyholder's account, free reserves, and the insurance company's shareholders. This question is delicate in practice and depends on legal and strategic considerations as well as on market conventions. In the following, I describe the MUST case of Bauer et al. (2006), which corresponds to the current legal situation in Germany. This contract, and thus my approach, is based on the framework described by Tanskanen and Lukkarinen (2003), which allows for a wide range of distribution schemes.

According to German legislation, insurance companies are obliged to pay at least a minimum rate of interest g on the policyholder's account value. This minimum interest rate has been fixed at up to 4% from 1994 to 2000, but was lowered to 2.25% on January 1st, 2007 and further to 1.75% on January 1st 2012. However, the lowered rates apply only to new contracts. Existing contracts are subject to the guaranteed interest rates in force when the contracts are made. The average guaranteed rate in German life insurance portfolios is currently at about 3.4%.[2] In addition to the guaranteed interest rate, it is required that at least a minimum participation rate $\delta = 90\%$ of the earnings on book values is passed on to policyholders. The insurance company's earnings are subject to accounting rules and thus the earnings on book values are in general not equal to the earnings on market values. In particular, insurance companies have usually some discretion in how earnings on market values are accounted for. I assume that only $y = 50\%$ of market value earnings need to be displayed

[2]See for example Ockenga, Schmitt, and Großpietsch (2010)

in book values. Let ν denote the policy anniversaries and assume that without loss of generality, these occur at integer values of $\nu = 1, \ldots, T$. I can then describe the update rule of the policyholder's account by

$$L_{\nu+1} := (1+g)L_\nu + \left[\delta y \left(A^-_{\nu+1} - A^+_\nu\right) - gL_\nu\right]^+. \tag{5.1}$$

The first part $(1+g)L_\nu$ of this update rule accounts for the guaranteed interest rate g, by which the policyholder's account value increases regardless of the return obtained on the asset portfolio. The second part describes the additional return contributed by the asset portfolio, provided such returns are sufficiently high. Here, A^-_ν (A^+_ν) denotes the value of the asset portfolio before (after) dividend payments are made. Deferring for the moment how such dividends are calculated, $\delta y(A^-_{\nu+1} - A^+_\nu)$ describes the part of the gain on the asset portfolio that must be displayed in book values (y) and credited to the policyholder (δ). If this part of the earnings is higher than the guaranteed interest payment it is additionally credited to the policyholder's account. I briefly note that the update rule given in equation (5.1) is a special case of the following contract described by Tanskanen and Lukkarinen (2003):

$$L_\nu = \begin{cases} L_{\nu-1} & t \in [\nu-1, \nu) \\ \mathrm{Bon}_\nu(A^-_\nu, A^+_{\nu-1}, L_{\nu-1}) & t = \nu \end{cases}. \tag{5.2}$$

I will now turn to the evolution of the asset process A_t. As noted, I assume that dividend payments are made at policy anniversaries $\nu = 1, \ldots, T$ and I will denote these payments by d_ν. Following Bauer et al. (2006), I assume that if the return on the asset portfolio is so poor that the earnings displayed in book values ($y(A^-_{\nu+1} - A^+_\nu)$) are less than the guaranteed interest rate payment, no dividend is paid out. If the return on the asset portfolio is such that the gains displayed in book values cover the guaranteed interest rate, but do not require additional contributions to the policyholder's account, then the remaining gains on book values are paid out to shareholders. Should the gains in book values be so large that more than the guaranteed interest rate, i.e. $\delta y(A^-_{t+1} - A^+_t)$, is attributed to the policyholder's account, then the remainder $(1-\delta)y(A^-_{t+1} - A^+_t)$ is paid out as a dividend. Formalizing these rules I obtain the following formula for dividend payments:

$$\begin{aligned} d_{\nu+1} =& (1-\delta)y\left(A^-_{\nu+1} - A^+_\nu\right) \mathbb{1}_{\left\{\delta y\left(A^-_{\nu+1} - A^+_\nu\right) > gL_\nu\right\}} \\ &+ \left(y\left(A^-_{\nu+1} - A^+_\nu\right) - gL_\nu\right) \cdot \mathbb{1}_{\left\{\delta y\left(A^-_{\nu+1} - A^+_\nu\right) \le gL_\nu \le y\left(A^-_{\nu+1} - A^+_\nu\right)\right\}}. \end{aligned} \tag{5.3}$$

If at an anniversary ν the return on the investment portfolio A^-_ν is insufficient to meet the new obligations L_ν, even if the entire reserve R_ν is dissolved I assume that the asset portfolio is propped up by a capital shot c_ν to cover its

obligations, i.e.

$$c_\nu = [A_\nu^- - L_\nu]^+.$$

These capital shots constitute effectively the interest rate guarantee as they are responsible for ensuring that at least the minimum interest rate can be attributed to the portfolio at each point in time.

Taking these two streams of capital into account, the asset portfolio is therefore updated according to

$$A_{\nu+1}^+ = A_{\nu+1}^- - d_\nu + c_\nu. \tag{5.4}$$

Observing that c_ν and d_ν are in effect functions of A_ν^+, $A_{\nu+1}^-$ and L_ν, I can assert that the evolution of A_t^+ can also be described by a known function Div_ν, i.e.

$$A_t^+ = \begin{cases} A_t^- & t \in (\nu-1, \nu) \\ \mathrm{Div}_\nu(A_\nu^-, A_{\nu-1}^+, L_{\nu-1}) & t = \nu \end{cases}. \tag{5.5}$$

Tanskanen and Lukkarinen (2003) show that contracts adhering to the update rules given in (5.2) and (5.5) can be priced under the risk-neutral valuation paradigm. Given the different payment streams I have outlined above, Bauer et al. (2006) deconstruct the contract and show that it can be thought of as a combination of the initial premium and three implicit options. First, the risk-neutral value of the future dividend payments

$$D_0 = \mathbb{E}_\mathbb{Q}\left[\sum_{\nu=1}^{T} B_\nu^{-1} d_\nu\right],$$

where B_t denotes the time t value of one unit of currency invested at time 0. Second, the risk-neutral value of the interest rate guarantees,

$$C_0 = \mathbb{E}_\mathbb{Q}\left[\sum_{\nu=1}^{T} B_\nu^{-1} c_\nu\right],$$

and third, the risk-neutral expectation of the change in the reserve account

$$\triangle R_0 = \mathbb{E}_\mathbb{Q}\left[B_T^{-1} R_T\right] - R_T,$$

I call a contract equipped with these features a *European contract*.

The contract may contain in addition a surrender option, meaning that the policy holder may have the right to cancel the contract in exchange for the current account value at each policy anniversary $\nu \in \{1, \ldots, T\}$. This surrender option is a Bermudan option, therefore, I call a contract equipped with such an

option a *Bermudan contract*. The gain from surrendering in year t_0 is

$$w_{t_0} = \left[D_{t_0} + B_{t_0}\mathbb{E}_{\mathbb{Q}}\left[B_T^{-1}R_T \mid \mathcal{F}_{t_0}\right] - R_{t_0} - C_{t_0}\right]^+,$$

and the value of the surrender option is therefore

$$W_0 = \sup_{\tau\in\Upsilon[0,T]} \mathbb{E}_{\mathbb{Q}}\left[B_\tau^{-1}w_\tau\right],$$

where $\Upsilon\{0,T\}$ denotes all stopping times with values in $\{0,\ldots,T\}$. For a fair contract with initial premium P the equilibrium condition

$$P := \mathbb{E}_{\mathbb{Q}}\left[B_T^{-1}P\frac{L_T}{L_0}\right] \ (+W_0) = P + C_0 - D_0 - \triangle R_0 \ (+W_0),$$

should hold. In the following, I will consider pricing and hedging the compound contract and the options under various parameter constellations and risk management approaches.

5.3. Pricing the contract under stochastic interest rates

5.3.1. The asset model

This section describes the two asset models I use to analyze the impact of parameter and model uncertainty. In the *base model*, the short-rate process follows an Ornstein-Uhlenbeck process, while the asset process is a geometric Brownian motion. In the *extended model*, both processes are generalized by introducing one additional parameter per process. In the *extended model*, the interest rate process is given by the model introduced by Chan et al. (1992) and the asset process follows a constant elasticity of variance process as discussed in Cox and Ross (1976).

My motivation for using these extensions to the interest rate and asset model are twofold: first, both models have been used extensively in literature as generalizations of their simpler underlying counterparts. For example, the CKLS model has been discussed as a generalization of the most popular short-rate models in Nowman (1997), Bliss and Smith (1998), and Treepongkaruna and Gray (2003). The CEV model has also been covered in popular studies such as Beckers (1980), Emanuel and MacBeth (1982), and Lauterbach and Schultz (1990). In particular, the CEV model has been used before in literature to discuss model uncertainty in the insurance context in Luoma, Puustelli, and Koskinen (2008). Second, both processes are natural extensions of those used in the base model case and follow from extending them continuously. This

is convenient from both a model and analysis perspective, as it allows me to understand the effect of introducing this additional source of uncertainty.

The real-world dynamics of both processes can be described by the following stochastic differential equations:

$$\begin{aligned} dr_t &= (\beta - \kappa r_t)dt + \sigma_r r_t^{\xi} dW_t^{\mathbb{P}} \\ dA_t &= \mu A_t dt + \sigma_A A_t^{\alpha} dZ_t^{\mathbb{P}}, \end{aligned} \tag{5.6}$$

with $dW_t^{\mathbb{P}} \cdot dZ_t^{\mathbb{P}} = \rho dt$. To obtain the base model, I set $\xi = 0$ and $\alpha = 1$, while in the extended model I allow both parameters to vary.

To change to the risk-neutral measure, I assume that the market price of risk is of the form $\lambda_t = \lambda r_t^{1-\xi} dt$. This is a standard assumption for the Vasicek ($\xi = 0$) and Cox-Ingersoll-Ross ($\xi = 1/2$) models, as discussed in Brigo and Mercurio (2007). The advantage of this form for the market price of risk is that the dynamics have the same structure under the objective and the risk-neutral measure. Under $\mathbb{Q}$ I obtain the following dynamics

$$\begin{aligned} dr_t &= (\beta - \tilde{\kappa} r_t)dt + \sigma_r r_t^{\xi} dW_t^{\mathbb{Q}} \\ dA_t &= r_t A_t dt + \sigma_A A_t^{\alpha} dZ_t^{\mathbb{Q}}, \end{aligned} \tag{5.7}$$

where $\tilde{\kappa} = \kappa - \lambda\sigma_r$ and the instantaneous correlation is $dW_t^{\mathbb{Q}} \cdot dZ_t^{\mathbb{Q}} = \rho dt$.

5.3.2. Numerical pricing

I now study the numerical pricing approach, which I use to price the insurance contract and implicit options under the model defined in equation (5.7). To this end, I first apply the insights gained by Tanskanen and Lukkarinen (2003) to derive a PDE for the term fix contracts. Then, I show how I numerically solve this PDE with a modified QUAD algorithm which was introduced in Andricopoulos et al. (2003) and extended in Andricopoulos et al. (2007).

Derivation of a value function

Let V_t denote the value of the European contract at time t. The value of this contract will in general depend on $t, r_t, A_t^-, A_{\lfloor t \rfloor}^+$ and $L_{\lfloor t \rfloor}$. If I have a function V, such that $V_t = V(t, r_t, A_t^-, A_{\lfloor t \rfloor}^+, L_{\lfloor t \rfloor})$ I call V a value function. Note that at policy anniversaries (for integer values of t), I have $A_t^- = A_{\lfloor t \rfloor}^+$ and thus $V_t = V(t, r_t, A_{\lfloor t \rfloor}^+, A_{\lfloor t \rfloor}^+, L_{\lfloor t \rfloor})$.

Define $F_\nu(r, a, l) := V(\nu, r, a, a, l)$. Using a no-arbitrage argument, Tanskanen and Lukkarinen 2003 show that

$$\begin{aligned}\lim_{t\to\nu^-} V(t, r_t, A_t, A_{\nu-1}, L_{\nu-1}) &= V(\nu, r_\nu, A_\nu, A_\nu, L_\nu) \\ &= F_\nu(r_\nu, A_\nu, Bon_\nu(A_\nu, A_{\nu-1}, L_{\nu-1})).\end{aligned}$$

If the value of F is known for a given $t_0 \in (\nu - 1, \nu)$, then for fixed values $A_{\nu-1}$ and $L_{\nu-1}$ the problem of finding a fair value function on the interval $[\nu - 1, t_0)$ reduces to solving the following PDE:

$$\begin{aligned}&\frac{\partial f}{\partial t} + rA\frac{\partial f}{\partial A} + \frac{1}{2}\sigma_A^2 A^{2\alpha}\frac{\partial^2 f}{\partial A^2} \\ &\qquad + (\beta - \kappa r)\frac{\partial f}{\partial r} + \frac{1}{2}\sigma_r^2 r^{2\xi}\frac{\partial^2 f}{\partial r^2} + \rho\sigma_A\sigma_r A^\alpha r^\xi \frac{\partial^2 f}{\partial A \partial r} - rf = 0. \quad (5.8)\end{aligned}$$

I know that there is a unique solution to this PDE and that this solution satisfies my requirements for a value function. Taking the limit $t_0 \to \nu$ I obtain a value function for the interval $[\nu - 1, \nu)$ if the value at $t = \nu$ is known.

I have now obtained a backward-iterative algorithm for finding a value function. Since the value of the contract at maturity is known, I can solve equation (5.8) for any given triple (r, a, l). This gives me a function $f(t, r, a, l)$ which must satisfy $V(T - 1, r, a, a, l) = f(T - 1, r, a, a, l) = F_{T-1}(r, a, l)$. Given this value, I can now iterate backward until I reach $T = 0$.

Numerical solutions with a modified QUAD algorithm

While I now have an algorithm for finding prices to the insurance contract and the implicit options, the problem of how to numerically obtain solutions for the contract values remains. In principal, I can obtain risk-neutral prices for the contract and the embedded options via either Monte Carlo methods or by approximating solutions to the PDE (5.8).

Monte Carlo methods are appealing, because of their relative simplicity compared to other methods. However, I will require frequent evaluation of the value function and its implicit options for many different parameter values, points in time, and values of r, A, and L. Especially troubling is that to obtain the computations under dynamic hedging, I need to compute derivatives in 120 points for each path, each requiring a new simulation with a large number of paths to obtain sufficient accuracy. Since I want to repeat this specific exercise a large number of times, it becomes clear that Monte Carlo approaches are computationally too burdensome for this problem. The alternative is thus to approximate solutions to the PDE (5.8). The easiest and most established approach is to apply finite differences as used for example by Zaglauer and Bauer

(2008). I have found that this technique also leads to complications, as I need to discretize the solution space in time, r, A, and L.

Instead, I have opted to adapt the extended QUAD algorithm presented in chapter 4. The basic idea of the algorithm is to solve a pricing PDE via the computation of the risk-neutral expectations and to split the time and space domain of the expectation in such a way that all integrals can be approximated with little error. This algorithm has two main advantages in my setting. The first advantage is that I do not need to discretize the time domain very finely, as I only need one step per policy anniversary. The second is that the flexible nature of the algorithm allows me to avoid one other source of interpolation error.

To construct the grid needed for the approximation, there is one more challenge to overcome. While between policy anniversaries, the dynamics of the asset portfolio are the same as that of the stock process, this is not the case at policy anniversaries, where dividend payments or capital shots alter the asset portfolio. Furthermore, the evolution of A_t^+ and L_t is highly dependent upon each other. As I would like to minimize interpolation errors and thus directly map to nodes from the previous time step, it is advantageous to construct the grid not in A_t^+ but in the reserve ratio $x_t = \frac{A_t^+}{L_t}$. The update rule for the reserve ratio x_t can be expressed independently of L_t, which facilitates the construction of the grid and reduces the need for interpolation. If I denote by $\triangle_t = \frac{A_t^-}{A_{t-1}^+} - 1$ the return on the asset portfolio before any adjustments, then the update rule for the reserve ratio is given by

$$x_t = \begin{cases} 1 & (1+\triangle_t)x_{t-1} < 1+g \\ \frac{(1+\triangle_t)x_{t-1}}{1+g} & 1+g-x_{t-1} \leq \triangle_t x_{t-1} < \frac{g}{y} \\ \frac{(1+\triangle_t(1-y))x_{t-1}+g}{1+g} & \frac{g}{y} \leq \triangle_t x_{t-1} < \frac{g}{\delta y} \\ \frac{(1+\triangle_t(1-y+\delta y))x_{t-1}}{1+\delta y \triangle_t x_{t-1}} & \frac{g}{\delta y} \leq \triangle_t x_{t-1} \end{cases}. \tag{5.9}$$

At the same time, I can recover the original variables A_t, L_t, and r_t, and can therefore apply A_t's density between policy anniversaries to obtain a solution on this grid.

To compute a solution, first, split the time domain at the policy anniversaries $t = 0, 1, \ldots, T$. At each of these points in time, I construct a grid in r and x and each of these grid points has a vector of values for L. At time T, the payoff and thus the values of the contract and the options are known. Going back to time $T-1$ I first pick a grid point $(r_{T-1}, x_{T-1}, L_{T-1})$ and then select all attainable[3] grid points at time T, $\{(r_T^i, x_T^j)\}_{(i,j)}$. Now the value at $(r_{T-1}, x_{T-1}, L_{T-1})$ can

[3]This is to say grid points that can be reached with sufficiently high probability.

be approximated by

$$
\begin{aligned}
&F_{T-1}(r_{T-1}, x_{T-1}, L_{T-1}) = \mathbb{E}_{\mathbb{Q}}\left[e^{-\int_{T-1}^{T} r_t dt} F\left(r_T, \frac{A_T^+}{L_T}, L_T\right) \mid \mathcal{F}_{T-1}\right] \\
&= \int_{\mathcal{D}} e^{-\int_{T-1}^{T} r_t dt} F_T\left(r_T, \frac{\mathrm{Div}_T(A_T^-, A_{T-1}^+, L_{T-1})}{\mathrm{Bon}_T(A_T^-, A_{T-1}^+, L_{T-1})}, \mathrm{Bon}_T(A_T^-, A_{T-1}^+, L_{T-1})\right) \\
&\qquad \cdot f\left(r_T, A_T^- \mid r_{T-1}, A_{T-1}^+\right) d(r_T, A_T^-) \\
&\approx \sum_i \sum_j F\left(r_T^i, x_T^j, L_T(x_T^j, L_{T-1})\right) \\
&\qquad \cdot f(x_T^j \mid x_{T-1}, r_{T-1}, r_T^j) f(r_T^j \mid r_{T-1}) \omega_j \omega_i, \qquad (5.10)
\end{aligned}
$$

where the $f(.|.)$ denote the respective conditional risk-neutral densities and ω_i and ω_j denote the integration weight, depending on the method used to discretize the integrals.[4]

This algorithm is an elegant solution in my case. The first advantage is that I do not need to calculate time points between policy anniversaries. Secondly, using a Simpson rule, the numeric integrals converge quite quickly, therefore reducing the need for extremely fine grids. Finally, once a grid has been constructed and the value of the contracts or options have been calculated, I can calculate a price for any time point t and values r_t, A_t, and L_t. If the point is not close to a grid point, I simply evaluate the numeric integral once again. This does not introduce additional interpolation errors and is efficient.

5.4. Estimating parameters and uncertainty

In this section, I describe the estimation techniques employed to obtain parameter estimates or estimates of parameter distributions. In my analysis of the impact of parameter uncertainty, I will consider these three cases:

1. No accounting for parameter uncertainty. In this case, I assume $\alpha = 1$ and $\xi = 0$ while the other parameters of model (5.6) are estimated using maximum likelihood. In the ensuing analysis, these estimates are treated as if they were the correct parameters.

2. Parameter uncertainty is incorporated by estimating posterior distributions for the parameters in model (5.6), while again $\alpha = 1$ and $\xi = 0$ are held fixed. This distribution is estimated with a MCMC algorithm as described in Section 5.4.2.

[4] I use the adapted compound Simpson rule presented in chapter 4. This rule is applicable to non-homogeneous intervals.

3. Both parameter and model uncertainty are considered. In this case I also use the MCMC algorithm, but this time α and ξ are allowed to vary. The posterior distribution therefore encompasses all eight parameters.

5.4.1. Maximum likelihood estimation

In the first case the parameters of the base model are estimated via maximum likelihood. Given equidistant observations $(r_0, \ldots, r_n)$ of the interest rate process and the asset price process $(A_0, \ldots, A_n)$, the differences of the tuples $((r_0, \ln A_0), \ldots, (r_n, \ln A_n))$ follow a bivariate normal distribution, leading to the following log-likelihood function:

$$\ln L(\mu, \sigma_A, \kappa, \beta, \sigma \mid A_0, \ldots, A_n, r_0, \ldots, r_n) = \\ -\frac{n}{2} \ln\left((1-\rho^2)\eta_A^2\eta_r^2\right) - \frac{1}{2(1-\rho^2)} \sum_{j=1}^{n} \left(\frac{x_j^2}{\eta_A^2} - 2\rho \frac{x_j y_j}{\eta_A \eta_r} + \frac{y_j^2}{\eta_r^2} \right) \tag{5.11}$$

with $x_j = \ln(A_j) - \ln(A_{j-1}) - (\mu - \frac{\sigma_A^2}{2})t$, $\eta_A^2 = \sigma_A^2 \Delta t$, and $y_j = r_j - r_{j-1}e^{-\kappa\Delta t} + \frac{\beta}{\kappa}(1 - e^{-2\kappa\Delta t})$, $\eta_r^2 = \frac{\sigma_r^2}{2\kappa}(1 - e^{-2\kappa\Delta t})$.

To obtain maximum likelihood estimates for the extended model, I derive a log-likelihood function by applying an Euler-discretization to the model. In both cases the maximum likelihood estimators are derived by minimizing the negative log-likelihood functions numerically.

5.4.2. Bayesian estimation

When considering parameter or model uncertainty, I use observations and a prior distribution to obtain a posterior parameter distribution. Given a prior distribution P for the parameters ϑ and data $\mathbf{x}$, I can obtain a posterior distribution $P(\vartheta \mid \mathbf{x})$ via Bayes' theorem:

$$P(\vartheta \mid x) = \frac{P(\vartheta, x)}{P(x)} = \frac{P(\vartheta) \cdot P(x \mid \vartheta)}{\int_\Theta P(\vartheta) P(x \mid \vartheta) d\vartheta} \propto P(\vartheta) \cdot L(\vartheta \mid x). \tag{5.12}$$

I use uninformative priors for all parameters. Specifically, I use improper priors of the form $\mathbb{1}_{\{.\in\mathbb{R}\}}$ for the parameters μ and β, $\mathbb{1}_{\{.>0\}}$ for σ_S, α, κ, σ_r, and ξ and $\frac{1}{2} \cdot \mathbb{1}_{\{.\in[-1,1]\}}$ for ρ. While the density function of the posterior distribution can be written down explicitly, it does not belong to a known distribution. To sample from this distribution, I employ a Markov chain Monte Carlo algorithm, namely a random walk Metropolis-Hastings algorithm.

	base model	extended model
μ	0.0290 (0.0212)	0.0238 (0.0333)
σ_S	0.0365 (0.0045)	8.4426 (27.0715)
α	n/a	0.3282 (0.9705)
κ	0.1390 (0.3318)	0.2319 (0.4368)
θ	0.0141 (0.2707)	0.0137 (3.2865)
σ_r	0.0033 (0.0004)	0.0054 (0.2022)
ξ	n/a	0.4942 (0.2806)
ρ	0.0836 (0.1615)	-0.1088 (0.2428)

Table 5.2.: Maximum likelihood estimates for the base and extended model (standard errors in parenthesis).

The random walk MCMC algorithm works by generating a Markov chain whose ergodic distribution is the desired distribution. The algorithm I employ is similar to the one presented in Section 3.2.4, except that now the parameters of both models need to be sampled. The difficulty in constructing a well mixing chain is once again that the parameters in both volatility terms, σ_S and α, as well as σ_r and ξ, exhibit a strong non-linear dependence. This makes the construction of well-mixing Markov chains quite difficult.

To overcome this challenge, I use the same approach as in chapter 3. The key insight is that while the volatility parameters vary substantially when seen together, the variation in the instantaneous volatility, $\sigma_S S_0^\alpha$ and $\sigma_r r_0^\xi$, is relatively small. Here, S_0 and r_0 are some conveniently fixed values, for example the averages of the observed time series. Therefore, I once again adapt the approach of first sampling candidates for α and ξ, and only then sampling candidates for the volatility parameters σ_S and σ_r. This way, the proposal distribution can account for the non-linear dependence between the parameters.

To sample the volatility parameters, I first sample the level effect parameters (α and ξ) and then perturb the volatility parameters (σ_S and σ_r) conditionally on the candidates for α and ξ. Specifically, given a previous value in the chain $\vartheta = (\mu, \sigma_S, \alpha, \kappa, \theta, \sigma_r, \xi)$, I first draw candidates $\tilde{\alpha} \sim \mathcal{N}(\alpha, \delta_\alpha)$ and $\tilde{\xi} \sim \mathcal{N}(\xi, \delta_\xi)$, which are centered at the previous values. Given these two candidates, I then

draw candidates $\tilde{\sigma}_S$ and $\tilde{\sigma}_r$, by perturbing the instantaneous volatility that results from the candidates $\tilde{\alpha}$ and $\tilde{\xi}$:

$$\tilde{\sigma_S} \sim \mathcal{N}(\sigma_S S_0^{-\alpha}, \delta_{\sigma_S}),$$
$$\tilde{\sigma_r} \sim \mathcal{N}(\sigma_r r_0^{-\xi}, \delta_{\sigma_r}).$$

The densities of $\tilde{\sigma_S}$ and $\tilde{\sigma_r}$ are thus normal when conditioned on $\tilde{\alpha}$ and $\tilde{\xi}$, respectively. This gives the following joint proposal densities for (σ_S, α) and (σ_r, ξ):

$$q_{\sigma_S,\alpha}(\tilde{\sigma_S}, \tilde{\alpha} \mid \sigma_S, \alpha) = \phi(\tilde{\alpha}\,;\, \alpha, \delta_\alpha) \cdot \phi(\tilde{\sigma}_S\,;\, \sigma_S S_0^{-\tilde{\alpha}}, \delta_{\sigma_S}),$$
$$q_{\sigma_r,\xi}(\tilde{\sigma_r}, \tilde{\xi} \mid \sigma_r, \xi) = \phi(\tilde{\xi}\,;\, \xi, \delta_\xi) \cdot \phi(\tilde{\sigma_r}\,;\, \sigma_r r_0^{-\tilde{\xi}}, \delta_{\sigma_r}).$$

The proposal distribution for the drift parameters μ, κ, and θ is not as delicate, therefore I just sample them from a multivariate normal distribution centered around their previous value in the chain with an appropriately chosen covariance matrix Σ:

$$(\tilde{\mu}, \tilde{\kappa}, \tilde{\theta}) \sim \mathcal{N}((\mu, \kappa, \theta), \Sigma).$$

The full proposal density for $\tilde{\vartheta} = (\tilde{\mu}, \tilde{\sigma_S}, \tilde{\alpha}, \tilde{\kappa}, \tilde{\theta}, \tilde{\sigma_r}, \tilde{\xi})$ thus given by

$$\begin{aligned} q(\tilde{\vartheta} \mid \vartheta) =& \phi((\tilde{\mu}, \tilde{\kappa}, \tilde{\theta})\,;\, (\mu, \kappa, \theta), \Sigma) \\ & \cdot \phi(\tilde{\alpha}\,;\, \alpha, \delta_\alpha) \cdot \phi(\tilde{\sigma}_S\,;\, \sigma_S S_0^{-\tilde{\alpha}}, \delta_{\sigma_S}) \\ & \cdot \phi(\tilde{\xi}\,;\, \xi, \delta_\xi) \cdot \phi(\tilde{\sigma_r}\,;\, \sigma_r r_0^{-\tilde{\xi}}, \delta_{\sigma_r}). \end{aligned} \tag{5.13}$$

Given this specially constructed proposal density, the random walk MCMC algorithm produces a well-mixing chain. The algorithm is presented formally in Algorithm 4.

5.4.3. Risk-neutral dynamics

To calculate prices and hedge ratios for my simulations, I require both the $\mathbb{P}$- and $\mathbb{Q}$-dynamics of the model. As in Hardy (2002), I assume that the parameters under the $\mathbb{Q}$-measure are fixed during the life of the contract. Typically, $\mathbb{Q}$-dynamics are obtained by fitting a model to observed market prices or are based on the expertise of a trader. To price the insurance contract and the implicit options, it is not immediately clear what prices should be used to calibrate the $\mathbb{Q}$-dynamics. Such contracts are not traded publicly and therefore a proxy for calibrating the measure would be needed. Furthermore, since I am

Algorithm 4 Random walk Metropolis-Hastings for CEV-CKLS model with special proposal distribution.

1. Choose a start value $\vartheta^{(0)} = (\mu^{(0)}, \sigma_S^{(0)}, \alpha^{(0)}, \kappa^{(0)}, \theta^{(0)}, \sigma_r^{(0)}, \xi^{(0)})$
2. For $i = 0, \ldots$
   a) Draw a new candidate $\tilde{\vartheta}$ with the following procedure:
      - Draw $(\tilde{\mu}, \tilde{\kappa}, \tilde{\theta}) \sim \mathcal{N}((\mu^{(i)}, \kappa^{(i)}, \theta^{(i)}), \Sigma)$.
      - Draw $\tilde{\alpha} \sim \mathcal{N}(\alpha, \delta_\alpha)$.
      - Draw $\tilde{\xi} \sim \mathcal{N}(\xi, \delta_\xi)$.
      - Draw $\tilde{\sigma_S} \sim \mathcal{N}(\sigma_S S_0^{-\alpha}, \delta_{\sigma_S})$.
      - Draw $\tilde{\sigma_r} \sim \mathcal{N}(\sigma_r r_0^{-\xi}, \delta_{\sigma_r})$.

   b) If $\min\left\{\frac{p(\tilde{\vartheta}) q_{\vartheta}(\vartheta^{(i)}|\tilde{\vartheta})}{p(\vartheta^{(i)}) q_{\vartheta}(\tilde{\vartheta}|\vartheta^{(i)})}, 1\right\} > u$, with $u \sim U[0,1]$, set $\vartheta^{(i+1)} = \tilde{\vartheta}$, otherwise set $\vartheta^{(i+1)} = \vartheta^{(i)}$.

   c) If $i =$ MAXSTEPS stop, otherwise go to step 2(a).

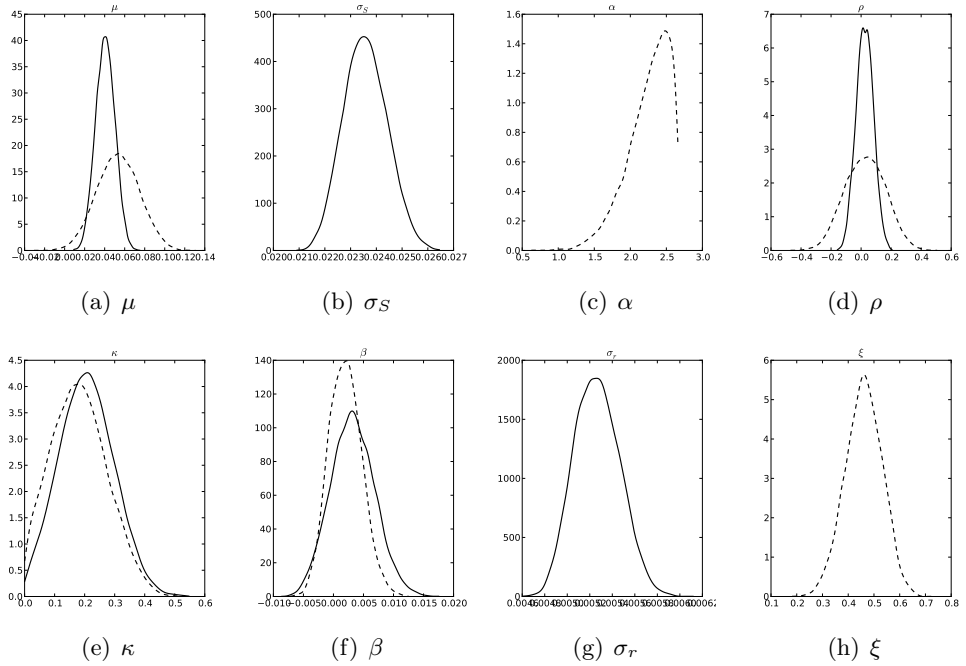
interested in the relative performance of the hedging strategy under the different uncertainty scenarios, the choice of the particular $\mathbb{Q}$-dynamics is of lesser importance. While different parameter choices would lead to different performances of the hedging strategy, their relative performance under uncertainty should be preserved. Therefore, I have opted to let the $\mathbb{Q}$-dynamics coincide with the $\mathbb{P}$-dynamics obtained from the maximum likelihood estimation. As the choice of the $\mathbb{Q}$-dynamics would otherwise be somewhat arbitrary and its impact on the relative performance of the risk-management approach should be small, this is the most transparent choice.

5.4.4. Data set and estimation results

As a proxy for the short-rate, I use German money call rates[5] with monthly frequency from December 31, 1996 to December 31, 2006. Insurance companies typically hold only a relatively low equity proportion in their investment portfolios. I use quotes of the German stock market index DAX over the same period to obtain a combined portfolio consisting of 15% stocks and 85% of a money market account as asset portfolio. This is the same data set used by Zaglauer and Bauer (2008). The maximum-likelihood estimates for both models for this time period are presented in Table 5.2.

To estimate the posterior distributions using the previously described random

[5] I use the OECD Germany Interest Rates Intermediate Call Money Total Bank Rate which I obtained from Bloomberg.



(a) μ (b) σ_S (c) α (d) ρ

(e) κ (f) β (g) σ_r (h) ξ

Figure 5.1.: Posterior parameter distributions for the base model (solid lines) and extended model (dashed lines). The distributions of σ_S and σ_r under the extended model are omitted as they exhibit large peaks close to 0 and long tails.

walk Metropolis-Hastings algorithm, I discard 50 000 draws as burn-in parameters and generate 10 000 000 subsequent draws which I thin to 10 000 for further processing. The initial acceptance ratios of the chains (before thinning) lay between 70% and 85%. The resulting posterior distributions are shown in Figure 5.1. The plot shows that the posterior distribution for the parameters in the drift terms (μ, κ, β) and the correlation parameter ρ are almost identical for both models, underlining that the differences lie only in the volatility terms. For the extended model, I omit plotting the posterior distributions of the volatility parameters σ_S and σ_r. Both distributions have very large peaks close to zero and exhibit long tails. This makes it difficult to discern them in the charts and obscures the distributions of these parameters in the base model.

5.5. Risk management techniques

I will now study the impact of parameter and model uncertainty on two risk management techniques. I refer to the first approach, which does not actively hedge risks, as *static reserving*. The second approach implements a delta hedge which is rebalanced on a monthly basis. I call this approach *dynamic hedging*.

I analyze how the two risk management approaches fare for a European contract

	Risk-neutral price
European contract	9811.86
Interest rate guarantee	43.94
Value of dividends	81.53
Change in reserve	108.30
Surrender option	138.43

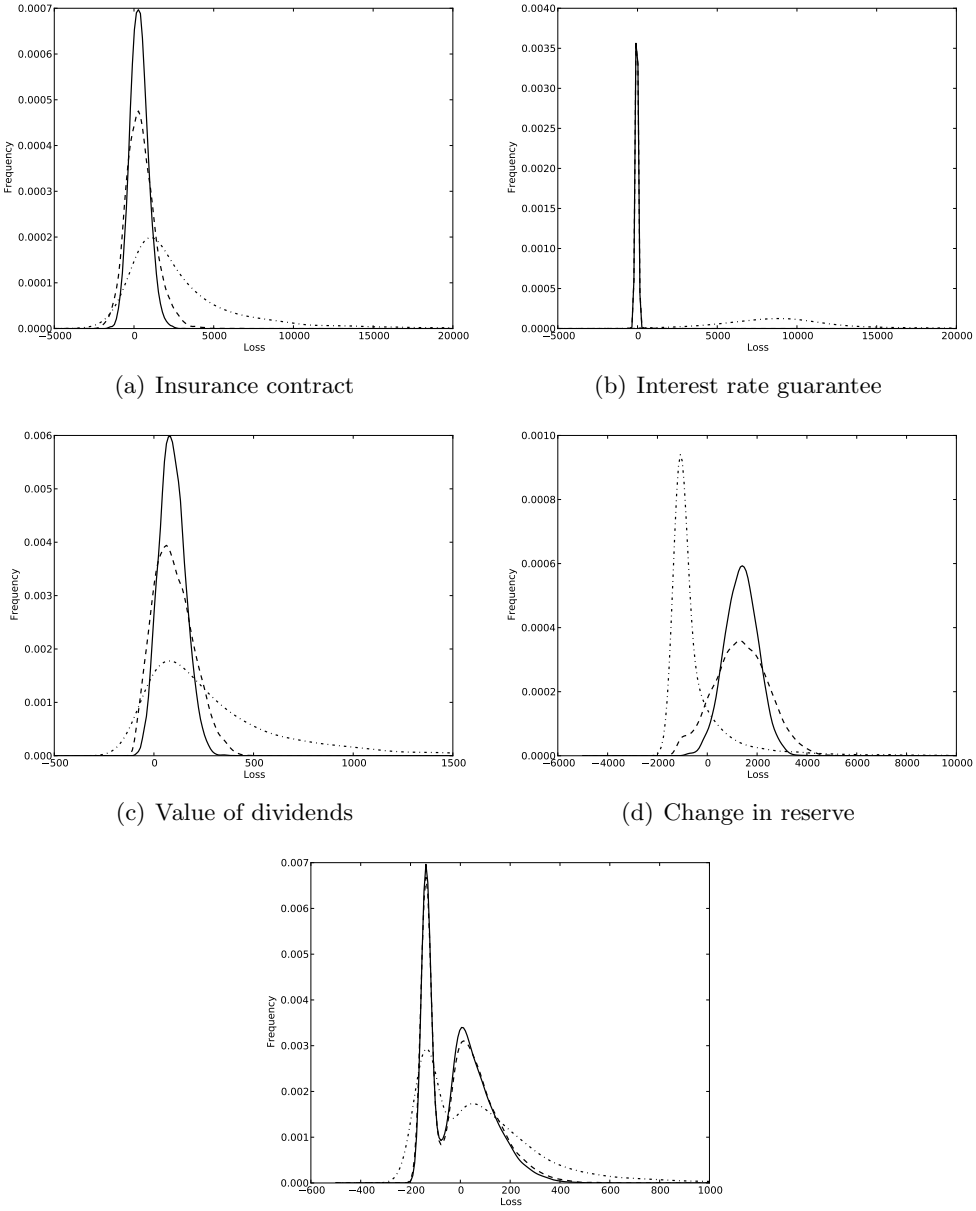
Table 5.3.: Risk-neutral values of the European contract and its embedded options with parameters based on maximum likelihood estimates.

and its embedded options. I assume an initial premium P_0 of 10 000, a guaranteed interest rate of $g = 2.25\%$, a participation rate of $\delta = 0.9$ and assume that the proportion of market value gains that the insurer must display in book values is $y = 0.5$. The parameters δ and y are as in Bauer et al. (2006) and Zaglauer and Bauer (2008), while the guaranteed interest rate is based on the rate effective in Germany from January 1st, 2007 on. The risk-neutral values of the European contract and its embedded options are reported in Table 5.3. Note that the contract was not calibrated so that its price matches the initial premium payment at inception. Instead, the initial values displayed in Table 5.3 are the fair values based on the risk-neutral parameters estimated from the observed data. I also analyze the results for contracts with alternative guaranteed interest rates in Tables 5.6 and 5.9.

5.5.1. Static reserving

Under this approach, an insurance company does not actively hedge the risks it is exposed to but reserves a certain amount of risk capital in order to cover incurred losses with a high probability. I implement this approach as follows: at time $t = 0$ the company sells the contract, receiving a premium of P_0 and investing this premium into a money market account as risk capital. These positions are held up to time T when a liability of L_T becomes due while the money market account has incurred interest and is now worth $P_0 B_T$, with $B_T = e^{\int_0^T r_t dt}$. The final loss from this strategy is $X_T = L_T - P_0 B_T$ and the time $t = 0$ value is thus $X_0 = B_T^{-1} L_T - P_0$. A standard risk measure is the value-at-risk x that will not be exceeded with probability α: $\text{VaR}_\alpha = \inf\{x : P(X_0 \leq x) \geq \alpha\}$. As the loss distributions I observe in the different cases vary substantially in their means, I will concentrate on quantiles of the distribution of unexpected losses. This can be thought of as the value-at-risk of a loss distribution after expected gains or losses have been subtracted.

To study the impact of parameter and model uncertainty, I simulate paths from



(a) Insurance contract

(b) Interest rate guarantee

(c) Value of dividends

(d) Change in reserve

(e) Surrender option

Figure 5.2.: Predicted loss distributions under static reserving. Solid line: no uncertainty, dashed line: with parameter uncertainty, dash-dotted line: with model and parameter uncertainty.

		mean	stddev	90%q UL	99%q UL
Insurance contract	noPU	323.01	571.10	739.77	1466.99
	PU	427.05	920.01	1190.25	2560.45
	MU	3039.73	4370.95	4434.70	16 912.84
Interest rate guarantee	noPU	-43.89	2.72	-0.02	0.01
	PU	-37.03	63.42	-6.87	253.18
	MU	9426.16	5230.70	4693.05	18 571.14
Value of dividends	noPU	92.15	64.91	85.66	160.96
	PU	98.71	98.92	136.93	256.31
	MU	343.03	492.84	519.27	1902.00
Change in reserve	noPU	1369.48	667.67	851.44	1557.48
	PU	1356.68	1082.98	1396.93	2532.51
	MU	-411.21	1781.92	1243.69	7579.59
Surrender option	noPU	-5.83	121.39	165.02	323.69
	PU	0.48	129.29	175.11	349.39
	MU	91.90	316.89	295.12	1219.70

Table 5.4.: Descriptive statistics of the loss resulting from static reserving. The mean and standard deviation are based on the unadjusted loss distribution while the 90% and 99% quantiles are of the unexpected losses, i.e. calculated on a loss distribution centered at 0.

time 0 to time T. In the case where parameter uncertainty is neglected, the basic model with the maximum likelihood estimates of the parameters is used for the simulation. In the following tables I label this case *noPU*. In the case where parameter uncertainty is considered, the paths are generated with the base model with parameters drawn from a posterior distribution for each path (PU). Finally, for the case where both parameter and model uncertainty are considered, the paths are generated from the extended model where parameters are once again drawn from posterior parameter distributions (MU).

Results for this strategy are presented in Table 5.4 and Figure 5.2. The solid, dashed, and dash-dotted lines show the loss distributions obtained by neglecting parameter uncertainty, considering parameter uncertainty, and considering both model and parameter uncertainty, respectively. All distributions are shown so that they reflect a short position in the contract. One should note however, that the insurance company holds a short position in the insurance contract, the interest rate guarantee, and the surrender option, and a long position in the value of dividends, and the change in reserve. Similarly to the graphs, the tables also show the 90%- and 99%-quantiles of the unexpected losses to simplify comparing the width of the loss distributions. It is however also noteworthy that the means of the loss distributions under parameter and model uncertainty are significantly different from zero. This is due to the asymmetric impact that parameter and model uncertainty have upon loss distributions. Another

varying parameters	Insurance contract		Interest rate guarantee		Value of dividends		Change in reserve		Surrender option	
	base	extended	base	extended	base	extended	base	extended	base	extended
none	1466.99	254.95	0.01	711.97	160.96	0.02	1557.48	17.04	323.69	138.35
	100.00%	17.38%	100.00%	11 429 682.57%	100.00%	0.01%	100.00%	1.09%	100.00%	42.74%
all	2560.45	16 912.84	253.18	18 571.14	256.31	1902.00	2532.51	7579.59	349.39	1219.70
	174.54%	1152.89%	4 064 414.27%	298 135 215.18%	159.23%	1181.63%	162.60%	486.66%	107.94%	376.81%
μ	1836.02	254.95	239.44	711.97	256.35	0.02	2430.69	17.04	322.33	138.35
	125.16%	17.38%	3 843 960.54%	11 429 682.57%	159.26%	0.01%	156.07%	1.09%	99.58%	42.74%
σ_S, α	1465.67	16 403.84	-0.03	18 896.92	162.83	1917.76	1577.41	7589.79	323.53	1381.95
	99.91%	1118.20%	-449.96%	303 365 183.07%	101.16%	1191.43%	101.28%	487.31%	99.95%	426.93%
μ, σ_S, α	1858.87	16 403.84	253.93	18 896.92	256.96	1917.76	2450.76	7589.79	321.91	1381.95
	126.71%	1118.20%	4 076 545.39%	303 365 183.07%	159.64%	1191.43%	157.35%	487.31%	99.45%	426.93%
θ	2970.82	1730.39	0.01	740.35	163.27	0.05	1793.15	15.54	403.25	242.43
	202.51%	117.96%	142.71%	11 885 370.24%	101.43%	0.03%	115.13%	1.00%	124.58%	74.89%
κ	1918.74	1169.11	0.01	746.00	163.25	0.03	1627.04	17.15	399.69	245.75
	130.79%	79.69%	156.49%	11 976 112.00%	101.42%	0.02%	104.47%	1.10%	123.48%	75.92%
κ, θ	2424.45	1346.10	0.01	727.64	163.39	0.03	1680.69	15.86	346.77	233.86
	165.27%	91.76%	109.33%	11 681 301.76%	101.51%	0.02%	107.91%	1.02%	107.13%	72.25%
σ_r, ξ	1473.78	1108.50	0.01	729.76	160.90	0.07	1555.38	16.24	330.28	386.74
	100.46%	75.56%	116.29%	11 715 312.76%	99.96%	0.05%	99.86%	1.04%	102.04%	119.48%
$\kappa, \theta, \sigma_r, \xi$	2406.87	1648.73	0.01	738.98	163.26	0.07	1680.79	15.85	347.34	423.17
	164.07%	112.39%	128.49%	11 863 394.31%	101.42%	0.05%	107.92%	1.02%	107.31%	130.73%
ρ	1451.13	254.95	0.01	711.97	160.36	0.02	1552.90	17.04	320.64	138.35
	98.92%	17.38%	101.17%	11 429 682.57%	99.63%	0.01%	99.71%	1.09%	99.06%	42.74%

Table 5.5.: This table shows the 99% quantile of unexpected losses incurred when reserving statically. The percentages compare the loss quantiles to the base case where parameter and model uncertainty are ignored.

Timeseries	Insurance contract base	Insurance contract extended	Interest rate guarantee base	Interest rate guarantee extended	Value of dividends base	Value of dividends extended	Change in reserve base	Change in reserve extended	Surrender option base	Surrender option extended
20y, w, g=2.25%,	2560.45	16 912.84	253.18	18 571.14	256.31	1902.00	2532.51	7579.59	349.39	1219.70
10% stocks	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
20y, w, g=1.75%,	2623.93	16 605.01	-1.89	18 195.64	238.38	1872.08	2380.44	7360.21	368.61	1166.89
10% stocks	102.48%	98.18%	-0.75%	97.98%	93.01%	98.43%	94.00%	97.11%	105.50%	95.67%
20y, w, g=3.4%,	2607.76	17 576.68	1262.49	19 566.72	280.94	1983.35	2970.34	7965.25	130.08	699.80
10% stocks	101.85%	103.93%	498.66%	105.36%	109.61%	104.28%	117.29%	105.09%	37.23%	57.37%
10y, w, g=2.25%,	3476.51	27 687.80	0.00	31 440.20	214.63	3131.52	2139.45	10 518.75	579.34	1884.70
10% stocks	135.78%	163.71%	0.00%	169.30%	83.74%	164.64%	84.48%	138.78%	165.82%	154.52%
30y, w, g=2.25%,	3017.81	4052.27	2772.04	4026.16	342.16	459.34	3768.23	2301.63	409.00	572.77
10% stocks	117.86%	23.96%	1094.90%	21.68%	133.50%	24.15%	148.79%	30.37%	117.06%	46.96%
20y, d, g=2.25%,	1914.84	14 129.64	-2.64	16 748.29	228.50	1619.95	2178.78	6336.76	278.80	1228.20
10% stocks	74.79%	83.54%	-1.04%	90.18%	89.15%	85.17%	86.03%	83.60%	79.80%	100.70%
20y, m, g=2.25%,	3892.58	13 952.21	342.46	16 173.77	277.54	1626.81	2859.97	5278.13	442.22	858.75
10% stocks	152.03%	82.49%	135.26%	87.09%	108.28%	85.53%	112.93%	69.64%	126.57%	70.41%
20y, w, g=2.25%,	2006.89	15 862.23	0.01	18 611.44	164.93	1803.32	1259.35	7083.15	384.93	1437.30
5% stocks	78.38%	93.79%	0.00%	100.22%	64.35%	94.81%	49.73%	93.45%	110.17%	117.84%
20y, w, g=2.25%,	4516.62	1819.19	4826.42	3294.08	555.56	227.81	6021.65	1587.25	555.45	552.06
15% stocks	176.40%	10.76%	1906.35%	17.74%	216.76%	11.98%	237.77%	20.94%	158.98%	45.26%

Table 5.6.: This table shows the 99% quantile of unexpected losses incurred when reserving statically. It compares the effect of estimating the parameters on time series of different lengths (10, 20, and 30 years) and different frequencies (daily, weekly, and monthly), with different proportions of stock and varying the guaranteed interest rate g. The percentages compare the loss quantiles to the base case in the upper left cell of each major column.

interesting feature is the bimodality of the loss distribution of the surrender option, which can be seen in Figure 5.2 (e). Often, the surrender option will never turn into the money, i.e. it will expire without providing any payout over its lifetime. These scenarios drive the left peak of the loss distribution. When the option does attain a positive value, its payoff is usually substantially larger than zero. The value of the option in this case is the difference between the guaranteed value of the contract and the current value of the assets. As both are substantially larger than the options value, values substantially larger than the options value are quickly attained. These scenarios drive the right peak of the loss distribution.

The plots in Figure 5.2 and the data in Table 5.4 show that considering parameter and model uncertainty leads to wider predicted loss distributions. For all parts of the contract, the 99%-quantile of the unexpected loss distributions is larger when parameter uncertainty or model uncertainty is considered. This is generally also true for the 90%-quantile with the only exception being the interest rate guarantee where the 90%-quantile of unexpected losses is actually smaller when parameter uncertainty is considered. This observation can easily be explained as the loss distribution is strongly peaked and the 90%-quantile is still part of this peak. The 99%-quantile of unexpected losses increases by 75% (1053%), 59% (1082%), 63% (387%) and 8% (277%) of the value of the base measure for the insurance contract, the value of dividends, the risk-neutral change in reserve, and the surrender option when parameter (model) uncertainty is considered. Providing percentage increases for the interest rate guarantee makes little sense as the base measure in this case is close to 0, but the risk-measures also increase substantially in this case. The high increases in these cases are due to the fact that uncertainty in the parameters is in fact quite large and paths derived under parameter or model uncertainty span a much wider range of possible outcomes than when parameter uncertainty is ignored.

Because I let the $\mathbb{P}$- and $\mathbb{Q}$-dynamics coincide, one could expect that the mean in the no uncertainty case should be fairly close to 0. Deviations from this point to the fact that the two measures are in fact not identical, even though the parameters are equal. The drift of the stock under the risk-neutral measure $\mathbb{Q}$ is given by the risk-free rate, whereas under $\mathbb{P}$ it is given by μ. Table 5.4 shows that this effect is indeed quite strong - the mean of all parts of the contract is pronouncedly different from 0. Comparing the different cases of parameter and model uncertainty, I observe that the means in the case without uncertainty and in the case with parameter uncertainty are fairly similar, while the means in the case with model uncertainty differ significantly from the other two. The more flexible structure of the extended model has a strong impact on the value of the contract.

Table 5.5 shows in more detail which uncertainties are driving the increase in the risk measures. Specifically, the table lists the 99%-quantiles of the unexpected losses as obtained by simulations with the base and extended model when some parameters are allowed to vary while all others are held fixed at their maximum likelihood estimates. The percentages indicate the size of the 99%-quantile of unexpected losses relative to the top left cell, i.e. to the base model where all parameters are held fixed (no uncertainty). In case of the base model, the increase in risk measures is mostly driven by the uncertainty in the parameters κ and β. However, for the extended model, uncertainty in the parameters σ_S and α contributes the most to increases in the risk measures. This is true for all parts of the contract. In some cases only allowing the stock parameters to vary freely leads to even larger increases in the risk measures than when all parameters are allowed to vary, indicating that some uncertainties have an offsetting effect.

Table 5.6 describes how the risk measures change when the parameters are observed over different observation horizons, with different frequencies, or when the stock proportion or the guaranteed interest rate of the contract are varied. The level of the guaranteed interest rate has the strongest effect on the interest rate guarantee under the base model. Lowering the guaranteed rate to 1.75% eliminates all risk from the interest rate guarantee under the base model, while increasing it to 3.4% substantially increases the risk measure. This effect does not carry over to the extended model, because the loss distributions are very wide in this case. Increasing the guaranteed rate to 3.4% also lowers the risk of the surrender option under both the base and the extended model. This is due to the fact that when the guaranteed interest rate is high, surrendering the contract is not very valuable, since the contract is expected to gain more value in the future. Lengthening the observation horizon or increasing the observation frequency leads to less uncertainty while shortening the observation horizon or using less frequent observations usually leads to more uncertainty. This is broadly reflected in the results, however there are some exceptions, which are due to the fact that the parameters in each time series are different and therefore the results are not always fully comparable. Finally, one would expect that lower stock proportion leads to less volatile portfolios while high stock proportion leads to a more volatile portfolio. This observation holds true for the base model where lower (higher) stock proportion indeed corresponds to lower (higher) risk measures. However, I observe the opposite behavior for the extended model. The reason for this is that the volatility parameters σ_S and α are harder to estimate for low stock portfolios, leading to more uncertainty and thus higher risk measures.

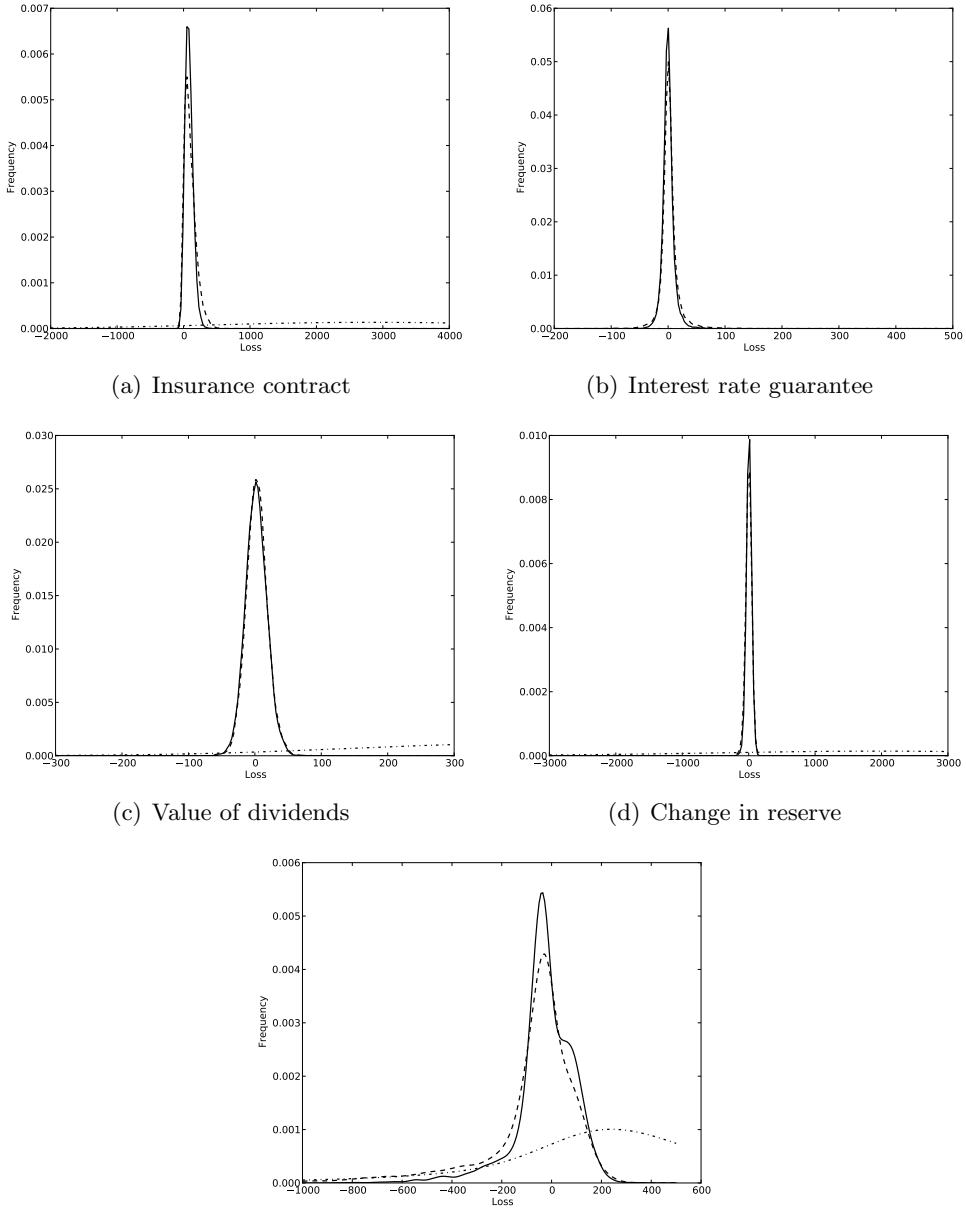
		mean	stddev	90%q UL	99%q UL
Insurance contract	noPU	83.19	60.57	80.33	172.40
	PU	99.54	87.92	119.39	275.99
	MU	4617.40	11 548.91	4823.19	18 758.66
Interest rate guarantee	noPU	0.13	10.02	10.71	31.35
	PU	2.15	15.65	13.71	56.77
	MU	7996.32	6339.80	6522.64	24 754.82
Value of dividends	noPU	2.07	16.20	20.58	40.04
	PU	3.03	15.97	20.23	40.66
	MU	613.72	1356.67	520.33	2228.33
Change in reserve	noPU	4.84	39.94	49.45	91.45
	PU	0.54	44.79	55.43	99.86
	MU	1879.97	11 918.77	3110.12	10 675.74
Surrender option	noPU	-26.88	121.53	134.14	216.83
	PU	-76.27	192.94	175.48	277.36
	MU	297.16	1106.96	581.84	2346.24

Table 5.7.: Descriptive statistics of the loss resulting from dynamic hedging. The mean and standard deviation are based on the unadjusted loss distribution while the 90% and 99% quantiles are of the unexpected losses, i.e. calculated on a loss distribution centered at 0.

5.5.2. Dynamic hedging

In the dynamic hedging risk management strategy, the insurance company tries to delta-hedge the contract or the implicit options against changes in the asset and interest rate process. In the classical option-pricing theory, it is assumed that the financial agent constantly readjusts his hedges and can trade in a frictionless market with no transaction costs. These assumptions are unrealistic and I relax them by assuming that the insurance company only adjusts its hedges on a monthly basis. It would be easy to include transaction costs in this model as is done for example by Hardy (2002), but I refrain from doing so to focus on the effects that are due to parameter and model uncertainty.

Under these assumptions, the dynamic hedging risk management approach can be described as follows. At time t_0, the contract is sold for a premium P_{t_0} and a hedge portfolio $P_{t_0} = H_{t_0}^+ = m_{t_0} M_{t_0} + u_{t_0} U_{t_0}$ is acquired. H_t^- (H_t^+) denotes the values of the hedge portfolio at time t before (after) it is adjusted to the new hedges. M_{t_0} and U_{t_0} denote the t_0-price of one unit of the money market account or the underlying, respectively, and m_{t_0} and u_{t_0} denote the respective number of units acquired. This portfolio is held until t_1 when its value will have changed to $H_{t_1}^- = m_{t_0} M_{t_1} + u_{t_0} U_{t_1}$. The weights m_{t_0} and u_{t_0} are adjusted to m_{t_1} and u_{t_1}, so that the new hedge portfolio is now worth $H_{t_1}^+ = m_{t_1} M_{t_1} + u_{t_1} U_{t_1}$. In



(a) Insurance contract

(b) Interest rate guarantee

(c) Value of dividends

(d) Change in reserve

(e) Surrender option

Figure 5.3.: Predicted loss distributions under dynamic hedging. Solid line: no uncertainty, dashed line: with parameter uncertainty, dash-dotted line: with model and parameter uncertainty.

varying parameters	Insurance contract		Interest rate guarantee		Value of dividends		Change in reserve		Surrender option	
	base	extended	base	extended	base	extended	base	extended	base	extended
none	172.40	12.23	31.35	51.55	40.04	7.57	91.45	43.50	216.83	99.23
	100.00%	7.10%	100.00%	164.45%	100.00%	18.90%	100.00%	47.56%	100.00%	45.76%
all	275.99	18 758.66	56.77	24 754.82	40.66	2228.33	99.86	10 675.74	277.36	2346.24
	160.09%	10 880.77%	181.10%	78 970.65%	101.55%	5565.00%	109.19%	11 673.39%	127.91%	1082.06%
μ	259.78	12.23	41.13	51.55	39.03	7.57	90.84	43.50	226.25	99.23
	150.68%	7.10%	131.20%	164.45%	97.47%	18.90%	99.33%	47.56%	104.34%	45.76%
σ_S, α	172.01	18 971.84	35.57	24 368.42	40.96	2102.74	98.82	10 211.93	217.23	2323.41
	99.77%	11 004.41%	113.47%	77 738.00%	102.29%	5251.37%	108.05%	11 166.23%	100.18%	1071.53%
μ, σ_S, α	270.13	18 971.84	50.00	24 368.42	39.38	2102.74	97.24	10 211.93	227.00	2323.41
	156.68%	11 004.41%	159.51%	77 738.00%	98.34%	5251.37%	106.33%	11 166.23%	104.69%	1071.53%
θ	178.11	15.90	38.88	81.76	40.70	8.98	92.37	66.97	336.08	231.53
	103.31%	9.23%	124.04%	260.83%	101.64%	22.43%	101.01%	73.23%	154.99%	106.78%
κ	174.53	14.48	34.77	65.33	40.72	8.19	92.22	63.39	330.14	189.39
	101.23%	8.40%	110.91%	208.40%	101.69%	20.46%	100.84%	69.31%	152.26%	87.35%
κ, θ	175.99	14.27	34.07	71.32	40.82	8.54	92.32	57.43	259.85	168.12
	102.08%	8.28%	108.69%	227.52%	101.94%	21.34%	100.95%	62.80%	119.84%	77.54%
σ_r, ξ	171.75	17.09	32.76	62.05	39.59	7.86	90.73	52.50	224.32	247.72
	99.62%	9.91%	104.50%	197.94%	98.88%	19.64%	99.21%	57.41%	103.45%	114.25%
$\kappa, \theta, \sigma_r, \xi$	174.76	19.44	34.76	77.45	40.12	8.87	92.14	68.16	267.71	310.09
	101.37%	11.27%	110.89%	247.08%	100.21%	22.15%	100.75%	74.53%	123.47%	143.01%
ρ	170.37	12.23	34.95	51.55	40.61	7.57	91.92	43.50	219.22	99.23
	98.82%	7.10%	111.49%	164.45%	101.43%	18.90%	100.51%	47.56%	101.10%	45.76%

Table 5.8.: This table shows the 99% quantile of unexpected losses incurred when hedging dynamically. The percentages compare the loss quantiles to the base case where parameter and model uncertainty are ignored.

Timeseries	Insurance contract base	Insurance contract extended	Interest rate guarantee base	Interest rate guarantee extended	Value of dividends base	Value of dividends extended	Change in reserve base	Change in reserve extended	Surrender option base	Surrender option extended
20y, w, g=2.25%,	275.99	18 758.66	56.77	24 754.82	40.66	2228.33	99.86	10 675.74	277.36	2346.24
10% stocks	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
20y, w, g=1.75%,	289.50	18 857.74	31.61	22 936.65	33.32	2141.62	103.51	9719.85	403.56	2519.78
10% stocks	104.89%	100.53%	55.68%	92.66%	81.94%	96.11%	103.65%	91.05%	145.50%	107.40%
20y, w, g=3.4%,	190.17	20 961.47	93.81	28 740.76	49.19	2327.07	89.02	12 724.91	41.41	1439.12
10% stocks	68.90%	111.74%	165.25%	116.10%	120.98%	104.43%	89.15%	119.19%	14.93%	61.34%
10y, w, g=2.25%,	253.47	30 172.07	32.34	37 365.75	35.25	3562.61	99.32	15 107.09	653.63	4047.04
10% stocks	91.84%	160.84%	56.97%	150.94%	86.70%	159.88%	99.46%	141.51%	235.66%	172.49%
30y, w, g=2.25%,	346.52	2506.14	148.47	2908.33	37.66	288.59	121.67	1232.08	312.53	480.29
10% stocks	125.55%	13.36%	261.53%	11.75%	92.60%	12.95%	121.84%	11.54%	112.68%	20.47%
20y, d, g=2.25%,	244.97	17 616.84	33.74	23 618.93	40.13	2014.08	105.08	9374.56	208.95	2402.28
10% stocks	88.76%	93.91%	59.43%	95.41%	98.70%	90.39%	105.23%	87.81%	75.34%	102.39%
20y, m, g=2.25%,	311.70	15 934.58	111.82	19 750.46	44.45	1760.79	118.88	7407.74	384.35	1505.63
10% stocks	112.94%	84.95%	196.98%	79.78%	109.33%	79.02%	119.05%	69.39%	138.58%	64.17%
20y, w, g=2.25%,	124.66	22 290.25	12.13	26 083.14	46.68	33 314.67	66.50	987.56	307.87	2818.38
5% stocks	45.17%	118.83%	21.36%	105.37%	114.81%	1495.05%	66.59%	9.25%	111.00%	120.12%
20y, w, g=2.25%,	671.01	935.43	254.32	1480.25	43.67	143.38	186.09	553.52	402.56	446.97
15% stocks	243.13%	4.99%	447.98%	5.98%	107.40%	6.43%	186.36%	5.18%	145.14%	19.05%

Table 5.9.: This table shows the 99% quantile of unexpected losses incurred when hediging dynamically. It compares the effect of estimating the parameters on time series of different lengths (10, 20, and 30 years) and different frequencies (daily, weekly, and monthly), with different proportions of stock and varying the guaranteed interest rate g. The percentages compare the loss quantiles to the base case in the upper left cell of each major column.

general, the cost of adjusting the portfolio (i.e. changing the weights from m_{t_0} and u_{t_0} to m_{t_1} and u_{t_1}) will be non-zero for two reasons: The adjustments are made infrequently and are computed with a model which may be misspecified. For both reasons, the hedges will not be perfect and the adjustments to the weights at time t_1 will, in general, not be self-financing. Therefore, $H_{t_1}^+ \neq H_{t_1}^-$ and $H_{t_1}^- - H_{t_1}^+$ additional capital needs to be injected (or may be withdrawn if $H_{t_1}^- - H_{t_1}^+$ is negative). This procedure is repeated until $t = T$ when the liability L_T becomes due and the hedge portfolio is liquidated, paying out H_T^-. Therefore, the total loss at time T from this strategy is

$$X_T = L_T - P_{t_0} B_T + \sum_{i=1}^{n} \frac{B_T}{B_{t_i}} \left(H_{t_i}^+ - H_{t_i}^- \right),$$

and the corresponding time $t = 0$ value is $X_0 = B_T^{-1} X_T$.

Under this risk management approach, I calculate loss distributions under the assumption that parameter and model uncertainty - if they are considered - are used to simulate paths but not to derive hedge ratios. The rationale for this is that it allows me to analyze the additional risk that an insurance company is exposed to by parameter and model uncertainty if it follows a standard risk management approach. Furthermore, even if the insurance company considers parameter and model uncertainty, it needs to parameterize its models in a specific way. Therefore, in the case where parameter uncertainty is considered (PU), the paths are simulated with parameters obtained from the posterior distribution of the base model and in the case where model uncertainty is considered (MU), the parameters are drawn from posterior distributions of the extended model. In both cases the hedge ratios are based on the base model with parameters obtained from maximum likelihood estimation. In the case where parameter uncertainty is ignored, both paths and hedge ratios are obtained using the base model with maximum likelihood estimates ($noPU$). This approach is similar to the one taken by Hardy (2002).

Results for this strategy are presented in Table 5.7 and Figure 5.3. As one would expect, the loss distributions are generally not as wide as under static reserving. Furthermore, in the cases where uncertainty is ignored or only parameter uncertainty is considered, the loss distributions are much more symmetric, reflecting that they are due to hedging errors which can be attributed to the infrequent readjustments. Table 5.7 shows that under parameter uncertainty, the risk measures increase only moderately, between 2% for the risk-neutral value of dividends and 81% for the interest rate reserve. While 81% is a strong increase in relative terms, the absolute amount lost is small, so these increases are moderate. In relative terms, the increases are comparable to those seen under the static reserving approach.

However, when model uncertainty is considered, the risk measures increase dramatically. With respect to the base case without parameter uncertainty, the 99% quantile of unexpected losses increases by factors between ten for the surrender option and about 790 for the interest rate guarantee. These dramatic losses occur because the volatility of the stock and interest rate process behave very differently under the extended model and the base model. Because of this, hedges derived under the base model may simply not work when the path generating process follows the extended model. This effect is particularly strong when stock values or interest rates are high. Under the extended model, which features heteroskedasticity, the volatility increases substantially and the hedges computed on the base model are inadequate. A few wrong hedges are enough to drastically increase the loss in such a scenario. The data in Table 5.8 supports this observation. Comparing the effects that uncertainty in the different parameters has, I see that allowing uncertainty in the volatility parameters σ_S and α have by far the largest impact on the increase in the quantiles of unexpected losses. This is true of both the base and the extended model, but the effect is much stronger in the extended model, as it allows a much wider range of volatility parameters.

Note that under dynamic hedging, I would expect the mean of the loss distribution to be relatively close to zero when I neglect parameter or model uncertainty. However, it may not be quite zero due to systematic hedging errors caused by the infrequent readjustments. Table 5.7 shows that these errors have a considerable impact on the mean, most pronouncedly on the insurance contract and the surrender option. As with static reserving, the effect of parameter and model uncertainty on the mean of the loss distribution is difficult to predict. Deviations in this case will be driven by systematic hedging errors, due to the fact that the underlying process behaves differently from what is assumed in the hedge calculations. Table 5.7 shows that under parameter uncertainty, there are a limited amount of systematic hedging errors. However, under model uncertainty, the mean of the loss distributions deviates substantially from zero in all cases. This underlines the impact that model uncertainty can have on the performance of the hedging strategy.

Tables 5.8 and 5.9 further illustrate the impact that uncertainty in the different parameters and the observation frequency and parameters of the contract have on the risk measures. As previously mentioned, Table 5.8 shows that the stock process volatility parameters σ_S and α have by far the strongest impact on the risk measures. In fact, under the extended model, they are solely responsible for the dramatic increases that I see in the risk measures. Under the base model, the drift parameter μ also has some influence. Table 5.9 describes how the risk measures change when the parameters are observed over different observation horizons, with different frequencies, or when the stock proportion or the guar-

anteed interest rate of the contract are varied. The guaranteed interest rate does not have a strong impact on the risk measures. The most pronounced effect is that the surrender option is less risky when the guaranteed interest rate is higher, which can again be explained by the observation that it is more attractive to keep holding the contract under these circumstances. The rest of the table can basically be summarized by the observation that longer and more frequent observations lead to less uncertainty in the parameters and therefore to lower risk measures. For the time series with lower and higher stock proportions than in the base portfolio, I observe again an inverse effect between the base and the extended model. This can be explained by the observation that portfolios with lower stock proportions lead in fact to more estimation uncertainty in the σ_S and α parameter pair. This means that the path generating model has a very different volatility term from the hedging model, leading again to higher risk estimates.

5.6. Conclusion

In this chapter, I analyzed the impact of parameter and model uncertainty on the risk management of implicit options in life insurance contracts. To this end, I have used MCMC algorithms to derive posterior distributions capturing parameter and model uncertainty and have shown how different risk management approaches are impacted by uncertainty. Furthermore, I have introduced an efficient pricing algorithm to value the contract and its implicit options.

The results indicate that both risk management approaches are strongly impacted by parameter and model uncertainty. Under both approaches, the consideration of parameter uncertainty leads to a substantial increase in the quantiles of unexpected losses by between 8% and 75% under static reserving and between 2% and 81% under dynamic hedging. These increases are relative to the base measures which were computed assuming that the initial risk measures are correct. When additionally introducing model uncertainty, risk measures under static reserving increase further, up to a factor of ten of the original measures in case of the insurance contract or the interest rate guarantee. While this increase is substantial, it pales in comparison to the increases observed when model uncertainty is considered when hedging dynamically. The risk measures in these cases increase dramatically, up to a factor of 790, in the case of the interest rate guarantee. This leads me to conclude that hedging infrequently with the wrong model simply does not work. Comparing the impact of estimation uncertainty on the two risk management approaches, I conclude that both approaches are impacted by estimation uncertainty and that the impact of parameter uncertainty on the two approaches is comparable. Introducing

model uncertainty has a much more dramatic impact on dynamic hedging than on static reserving.

The results presented here are quite dramatic, and they may overstate the effect that parameter and model uncertainty have on risk calculations in practice. First of all, my modelling approach has a few constraining factors which would likely be addressed in practice. For example, few risk managers would hold fast to their model when it has produced completely incorrect hedges several times in a row. Nevertheless, these results illustrate how large the impact of parameter and model uncertainty can be. Specifically, choosing an incorrectly specified model can lead to vastly underestimated risk measures and can suggest risk management approaches that simply don't work for the true underlying model. As was shown in chapter 3, the Vasicek model gives rarely a good fit to interest rate time series. Nevertheless, it is often used because it is well known and leads to convenient formulas. The results form this chapter suggest that using the wrong model can have quite dramatic consequences. While the methods presented here may lead to some scenarios that overstate uncertainty or are deemed unrealistic in practice, the Bayesian approach I employed has the benefit of making it easy to include additional information, for example the opinion of experienced risk managers. As my results show that too much trust in estimation results and the risk measures based upon them can be fatal, I conclude that under both risk management approaches, cautious evaluations should take parameter and model uncertainty into account.

6. Conclusion

6.1. Summary and conclusions

In this thesis, I studied the impact of estimation uncertainty upon risk measures. I applied Bayesian methods to quantify model and parameter uncertainty and to study their impact in two case studies. Additionally, I introduced two extensions to pricing models as they were mandated by the requirements of these studies.

Chapter 2 presented the Bayesian methodology I used throughout this thesis. This chapter motivated the use of the Bayesian approach and presented the most important results as they were used in the later chapters. I found the Bayesian approach to be a suitable tool, as it is able to easily and consistently integrate estimation uncertainty and allows for the application of a similar risk measurement machinery as standard approaches. The two biggest drawbacks from its use are that it is not always simple to arrive at results from a computational perspective, and that it has yet to gain the same widespread acceptance as methods based on the frequentist view of statistics have.

In Chapter 3, I applied the Bayesian approach presented in the previous chapter to the estimation of short-rate models nested within the Chan-Karolyi-Longstaff-Sanders model. This chapter was influenced by the widely varying estimates for these models found in literature. However, the hypothesis that estimation uncertainty in the volatility parameters could be responsible for these varying estimates had to be rejected. While Bayesian methods which include model and parameter uncertainty suggested additional uncertainty in the estimates when compared to standard frequentist estimates, they could not explain the full difference between estimates. The estimation uncertainty can also not sufficiently improve the predictive performance of the CKLS model to justify its use in predicting quantiles of the short-rate.

Chapter 4 introduced two extensions to existing pricing methods. The development of these extensions was required by the study in Chapter 5, which demanded a fast method for pricing options within the Black-Scholes-Vasicek model. One of the methods, namely the extension of the QUAD algorithm, showed very promising performance characteristics when applied to pricing options in the Black-Scholes-Vasicek model. The other method allows the pricing of options in the more general CEV-CKLS model.

Chapter 5 presented the second case study of the effects of parameter and model uncertainty. It compared the effect of these uncertainties on two risk management methods, namely managing the options embedded in a participating life-insurance contract by only holding capital or by holding a hedging portfolio which is updated once per month. The results show that the effect of parameter and model uncertainty can be quite dramatic in both cases. While some results seen in this Chapter probably overstate the real risk caused by model and parameter uncertainty, they should still caution risk managers of potential risks entailed in parameter and model uncertainty.

With these results, I now return to the two research questions raised in the introduction of this thesis:

- *How large is estimation uncertainty in short-term interest rate models?*

The amount of estimation uncertainty in the CKLS model was studied in Chapter 3 and Chapter 5. The results from Chapter 3 indicate that there is some estimation uncertainty in the estimates, which is indicated by the lower test statistics obtained with Bayesian methods. However, these estimation uncertainties are not large enough to explain the disparate estimation results found in literature. Furthermore, estimation uncertainty decreases substantially when moving to daily observations and long estimation horizons. Chapter 5 also indicates moderate estimation uncertainty in the parameter estimates. Overall, these results indicate that there is a moderate amount of estimation uncertainty, which decreases with the length of time series and observation frequencies.

- *Does the predictive power of short-term interest rate models improve when estimation uncertainty is considered?*

While estimation uncertainty does increase the uncertainty inherent in predictions and thus leads to wider distributions and more careful risk estimates, I found that it cannot sufficiently increase the predictive power of the CKLS model for it to accurately predict the one step ahead distribution of the interest rates studied. I therefore conclude that while including estimation uncertainty does lead to more careful predictions, this additional uncertainty is not enough for the CKLS model to predict the future distribution of short-rates sufficiently carefully.

- *How can the impact of estimation uncertainty in these models be measured and how large is it?*

The results from the previous chapters have shown that the Bayesian approach provides a consistent framework that allows incorporating estimation uncertainties into the target profit and loss distribution which is subsequently used in all risk management applications. It is thus possible to use and compare the

distributions incorporating estimation uncertainties to those who do not. The Bayesian approach provides thus a methodology for integrating and studying the effects of these additional risks. Another advantage of this approach is that it will work with much of the existing machinery present when measuring and managing financial risk.

The disadvantage of this approach is that it often relies on computational results and that these computational results are not always easy to produce. As was shown in Chapter 2, the simple question of convergence of a Markov chain can be quite delicate to answer in practice, and Chapter 4 showed that a straightforward application of these methods can even mandate the development of more efficient pricing methods.

While the results from Chapter 3 and 5 indicate that the amount of estimation uncertainty in the parameters is moderate, the impact of model and parameter uncertainty upon risk measures can be quite drastic. The study concerning the uncertainty in estimates in the CKLS-model indicated that estimation uncertainty cannot explain all of the uncertainties involved, while the study on the risk measures for implicit options showed that even so, actual risk-management results can be quite dramatic.

Clearly, two interacting factors are at play when determining the impact of model and parameter uncertainty upon risk measures: first, the extent of uncertainty in the model and parameter estimates, and second, the impact of this uncertainty on the measures of interest. Given the possibly dramatic results found in Chapter 5, I conclude that further study of these effects is warranted.

6.2. Outlook

While this thesis explored several aspects concerning the impact of parameter and model uncertainty in risk management, it is clear that this effort is only a small contribution in the quest to answer these important questions. Having gained some familiarity with the field, I conclude that there are many open questions that need to be addressed.

First of all, the consideration of model and parameter uncertainty in risk management so far lacks an overarching and consistent theory. While the inclusion of parameter and model uncertainty in the Bayesian methodology is a good start, it really only covers the first half of the equation. The second half, specifically how these uncertainties affect the results of risk measurement, is addressed in this thesis in a practical manner, but it could clearly benefit from a more theoretical assessment. To this end, I believe that it is necessary to quantify estimation objectives via a loss function. This approach has been applied when

the Bayesian approach was introduced in portfolio selection theory. A similar approach to obtain a consistent theory to manage parameter and model uncertainty in risk management would be desirable.

A second open question is the separation of the two drivers of the impact of estimation uncertainty on risk measures. An intuitive approach is to separate the total impact of estimation uncertainty on risk measures into estimation uncertainty and the sensitivity of risk measures to this uncertainty. However, it is so far not clear how this separation can be done so that one could take the two results and combine them in a meaningful way without too much computational effort.

Due to these concerns, the approach presented in this thesis has relied on computational techniques, which can be quite demanding in programmer ability and computational resources. It may then at times be necessary for existing computational techniques to be improved upon, as I have done in Chapter 4.

A. Additional results for Chapter 3

A.1. Additional tables

Table A.1.: This table shows the p-values and pseudo p-values for the estimates of the CKLS model over the ten year window given in Table 3.4. P-values for the maximum likelihood estimates (MLE) and pseudo p-values for the maximum a-posteriori estimates (MAP) were obtained with the method described in Section 3.3.2. P-values below 0.1 are marked with a single star and p-values below 0.05 and 0.01 are marked with two and three stars, respectively.

	MLE				MAP, informative prior				MAP, flat prior			
Timeseries	α	β	σ	ξ	α	β	σ	ξ	α	β	σ	ξ
USD, 10y, d	0.7107	0.2373	0.0000***	0.0000***	0.7107	0.2373	0.0000***	0.0000***	0.7107	0.2373	0.0000***	0.0000***
USD, 10y, w	0.8300	0.6638	0.0000***	0.0000***	0.8300	0.6638	0.0000***	0.0000***	0.8300	0.6638	0.0000***	0.0000***
USD, 10y, m	0.7868	0.5751	0.0017***	0.0000***	0.7868	0.5751	0.0017***	0.0000***	0.7868	0.5751	0.0017***	0.0000***
GBP, 10y, d	0.4695	0.0003***	0.0000***	0.0000***	0.4695	0.0003***	0.0000***	0.0000***	0.4695	0.0003***	0.0000***	0.0000***
GBP, 10y, w	0.8173	0.0114**	0.0000***	0.0000***	0.8173	0.0114**	0.0000***	0.0000***	0.8173	0.0114**	0.0000***	0.0000***
GBP, 10y, m	0.8001	0.5405	0.0025***	0.0000***	0.8001	0.5405	0.0025***	0.0000***	0.8001	0.5405	0.0025***	0.0000***
JPY, 10y, d	0.2473	0.7425	0.0000***	0.0000***	0.2473	0.7425	0.0000***	0.0000***	0.2473	0.7425	0.0000***	0.0000***
JPY, 10y, w	0.4066	0.9435	0.0002***	0.0000***	0.4066	0.9435	0.0002***	0.0000***	0.4066	0.9435	0.0002***	0.0000***
JPY, 10y, m	0.0896*	0.4271	0.0024***	0.0000***	0.0896*	0.4271	0.0024***	0.0000***	0.0896*	0.4271	0.0024***	0.0000***
FRF, 10y, d	0.0768*	0.1150	0.0000***	0.0000***	0.0768*	0.1150	0.0000***	0.0000***	0.0768*	0.1150	0.0000***	0.0000***
FRF, 10y, w	0.2768	0.2187	0.0000***	0.0000***	0.2768	0.2187	0.0000***	0.0000***	0.2768	0.2187	0.0000***	0.0000***
FRF, 10y, m	0.4244	0.4269	0.0015***	0.0000***	0.4244	0.4269	0.0015***	0.0000***	0.4244	0.4269	0.0015***	0.0000***
DEM, 10y, d	0.3590	0.1553	0.0000***	0.0000***	0.3590	0.1553	0.0000***	0.0000***	0.3590	0.1553	0.0000***	0.0000***
DEM, 10y, w	0.1748	0.0951*	0.0000***	0.0000***	0.1748	0.0951*	0.0000***	0.0000***	0.1748	0.0951*	0.0000***	0.0000***
DEM, 10y, m	0.4445	0.2249	0.0038***	0.0000***	0.4445	0.2249	0.0038***	0.0000***	0.4445	0.2249	0.0038***	0.0000***
ITL, 10y, d	0.8991	0.8590	0.0000***	0.0000***	0.8991	0.8590	0.0000***	0.0000***	0.8991	0.8590	0.0000***	0.0000***
ITL, 10y, w	0.9074	0.5899	0.0000***	0.0000***	0.9074	0.5899	0.0000***	0.0000***	0.9074	0.5899	0.0000***	0.0000***
ITL, 10y, m	0.6034	0.9841	0.0000***	0.0000***	0.6034	0.9841	0.0000***	0.0000***	0.6034	0.9841	0.0000***	0.0000***
CHF, 10y, d	0.4743	0.0208**	0.0000***	0.0000***	0.4743	0.0208**	0.0000***	0.0000***	0.4743	0.0208**	0.0000***	0.0000***
CHF, 10y, w	0.4897	0.0453**	0.0000***	0.0000***	0.4897	0.0453**	0.0000***	0.0000***	0.4897	0.0453**	0.0000***	0.0000***
CHF, 10y, m	0.0486**	0.0661*	0.0013***	0.0000***	0.0486**	0.0661*	0.0013***	0.0000***	0.0486**	0.0661*	0.0013***	0.0000***
AUD, 10y, d	0.6603	0.6971	0.0000***	0.0001***	0.6603	0.6971	0.0000***	0.0001***	0.6603	0.6971	0.0000***	0.0001***
AUD, 10y, w	0.9219	0.9928	0.0014***	0.0023***	0.9219	0.9928	0.0014***	0.0023***	0.9219	0.9928	0.0014***	0.0023***
AUD, 10y, m	0.2958	0.3939	0.1443	0.0297**	0.2958	0.3939	0.1443	0.0297**	0.2958	0.3939	0.1443	0.0297**
EUR, 10y, d	0.6942	0.0222**	0.0000***	0.0000***	0.6942	0.0222**	0.0000***	0.0000***	0.6942	0.0222**	0.0000***	0.0000***
EUR, 10y, w	0.3459	0.0190**	0.0000***	0.0000***	0.3459	0.0190**	0.0000***	0.0000***	0.3459	0.0190**	0.0000***	0.0000***
EUR, 10y, m	0.8176	0.3987	0.0003***	0.0000***	0.8176	0.3987	0.0003***	0.0000***	0.8176	0.3987	0.0003***	0.0000***

Table A.2.: This table shows the p-values and pseudo p-values for the estimates of the CKLS model over the 25 year window given in Table 3.5. P-values for the maximum likelihood estimates (MLE) and pseudo p-values for the maximum a-posteriori estimates (MAP) were obtained with the method described in Section 3.3.2. P-values below 0.1 are marked with a single star and p-values below 0.05 and 0.01 are marked with two and three stars, respectively.

	MLE				MAP, informative prior				MAP, flat prior			
Timeseries	α	β	σ	ξ	α	β	σ	ξ	α	β	σ	ξ
USD, 25y, d	0.1368	0.8194	0.0000***	0.0000***	0.1368	0.8194	0.0000***	0.0000***	0.1368	0.8194	0.0000***	0.0000***
USD, 25y, w	0.3598	0.3666	0.0000***	0.0000***	0.3598	0.3666	0.0000***	0.0000***	0.3598	0.3666	0.0000***	0.0000***
USD, 25y, m	0.9395	0.3142	0.0000***	0.0000***	0.9395	0.3142	0.0000***	0.0000***	0.9395	0.3142	0.0000***	0.0000***
GBP, 25y, d	0.3768	0.0020***	0.0000***	0.0000***	0.3768	0.0020***	0.0000***	0.0000***	0.3768	0.0020***	0.0000***	0.0000***
GBP, 25y, w	0.0584*	0.5077	0.0000***	0.0000***	0.0584*	0.5077	0.0000***	0.0000***	0.0584*	0.5077	0.0000***	0.0000***
GBP, 25y, m	0.5670	0.4791	0.0000***	0.0000***	0.5670	0.4791	0.0000***	0.0000***	0.5670	0.4791	0.0000***	0.0000***
JPY, 25y, d	0.0003***	0.3785	0.0000***	0.0000***	0.0003***	0.3785	0.0000***	0.0000***	0.0003***	0.3785	0.0000***	0.0000***
JPY, 25y, w	0.0004***	0.5570	0.0000***	0.0000***	0.0004***	0.5570	0.0000***	0.0000***	0.0004***	0.5570	0.0000***	0.0000***
JPY, 25y, m	0.0024***	0.5131	0.0000***	0.0000***	0.0024***	0.5131	0.0000***	0.0000***	0.0024***	0.5131	0.0000***	0.0000***
CHF, 25y, d	0.0205**	0.2822	0.0000***	0.0000***	0.0205**	0.2822	0.0000***	0.0000***	0.0205**	0.2822	0.0000***	0.0000***
CHF, 25y, w	0.5955	0.9029	0.0000***	0.0000***	0.5955	0.9029	0.0000***	0.0000***	0.5955	0.9029	0.0000***	0.0000***
CHF, 25y, m	0.4056	0.4137	0.0000***	0.0000***	0.4056	0.4137	0.0000***	0.0000***	0.4056	0.4137	0.0000***	0.0000***
AUD, 25y, d	0.3967	0.2979	0.0000***	0.0000***	0.3967	0.2979	0.0000***	0.0000***	0.3967	0.2979	0.0000***	0.0000***
AUD, 25y, w	0.2714	0.1605	0.0000***	0.0000***	0.2714	0.1605	0.0000***	0.0000***	0.2714	0.1605	0.0000***	0.0000***
AUD, 25y, m	0.0707*	0.0301**	0.0000***	0.0000***	0.0707*	0.0301**	0.0000***	0.0000***	0.0707*	0.0301**	0.0000***	0.0000***
DEM, 25y, d	0.5315	0.1896	0.0000***	0.0000***	0.5315	0.1896	0.0000***	0.0000***	0.5315	0.1896	0.0000***	0.0000***
DEM, 25y, w	0.8317	0.7268	0.0000***	0.0000***	0.8317	0.7268	0.0000***	0.0000***	0.8317	0.7268	0.0000***	0.0000***
DEM, 25y, m	0.4071	0.1474	0.0000***	0.0000***	0.4071	0.1474	0.0000***	0.0000***	0.4071	0.1474	0.0000***	0.0000***

Table A.3.: This table shows descriptive statistics obtained for the MCMC chains used to compute the Bayes factors for time series observed over ten years. *Accept.* stands for the acceptance ratio, *mean* for the mean of all samples, *SE* for the standard error as described in Section 2.3.3, and *95%-HPD* for the smallest interval comprising 95% of all samples from the posterior distribution.

Timeseries	Accept.	α				β				σ				ξ			
		mean	SE	95%-HPD		mean	SE	95%-HPD		mean	SE	95%-HPD		mean	SE	95%-HPD	
USD, 10y, d	0.3640	-0.0003	0.0000	-0.0012	0.0006	-0.1334	0.0013	-0.2505	-0.0063	0.1155	0.0001	0.1005	0.1310	0.8403	0.0003	0.8095	0.8718
USD, 10y, w	0.3785	0.0003	0.0000	-0.0018	0.0024	-0.2382	0.0031	-0.4883	-0.0001	0.2447	0.0008	0.1617	0.3352	0.8530	0.0008	0.7717	0.9387
USD, 10y, m	0.3527	0.0018	0.0000	-0.0013	0.0049	-0.2669	0.0030	-0.5305	-0.0003	0.5332	0.0040	0.2219	0.9199	0.9098	0.0016	0.7542	1.0673
GBP, 10y, d	0.3738	0.0006	0.0000	-0.0001	0.0013	-0.2273	0.0011	-0.3208	-0.1411	0.3634	0.0005	0.3209	0.4094	1.3236	0.0004	1.2899	1.3570
GBP, 10y, w	0.3801	0.0002	0.0000	-0.0015	0.0018	-0.2195	0.0017	-0.3885	-0.0566	0.2991	0.0009	0.2080	0.3995	1.1187	0.0009	1.0288	1.2064
GBP, 10y, m	0.3453	0.0009	0.0000	-0.0027	0.0045	-0.1994	0.0023	-0.4185	-0.0001	0.2855	0.0022	0.0905	0.5192	0.9122	0.0019	0.6994	1.1333
JPY, 10y, d	0.3146	0.0002	0.0000	0.0000	0.0004	-0.1489	0.0023	-0.3555	0.0000	0.3632	0.0005	0.3149	0.4150	0.9778	0.0002	0.9569	0.9980
JPY, 10y, w	0.3281	0.0003	0.0000	0.0000	0.0005	-0.1905	0.0028	-0.4357	-0.0001	1.4699	0.0072	0.8710	2.1183	1.1439	0.0007	1.0801	1.2081
JPY, 10y, m	0.3611	0.0003	0.0000	0.0000	0.0005	-0.1863	0.0033	-0.4357	0.0000	5.4181	0.0535	1.3882	10.7711	1.3092	0.0014	1.1774	1.4505
FRF, 10y, d	0.3243	0.0081	0.0001	-0.0006	0.0177	-0.2040	0.0026	-0.4531	0.0000	2.8980	0.0041	2.5619	3.2728	1.8687	0.0005	1.8280	1.9114
FRF, 10y, w	0.3562	0.0108	0.0001	-0.0013	0.0235	-0.2879	0.0033	-0.6195	-0.0001	2.9640	0.0100	2.2246	3.8386	1.8513	0.0012	1.7532	1.9438
FRF, 10y, m	0.3523	0.0142	0.0002	-0.0007	0.0300	-0.3419	0.0046	-0.7131	-0.0004	1.5199	0.0112	0.5574	2.8068	1.5437	0.0026	1.2924	1.8167
DEM, 10y, d	0.3621	0.0046	0.0001	-0.0016	0.0111	-0.1348	0.0014	-0.2598	-0.0017	0.0647	0.0001	0.0554	0.0741	0.7688	0.0005	0.7204	0.8163
DEM, 10y, w	0.3578	0.0072	0.0001	-0.0021	0.0164	-0.2150	0.0023	-0.4131	-0.0011	0.1513	0.0007	0.0943	0.2125	0.9878	0.0015	0.8531	1.1140
DEM, 10y, m	0.3419	0.0077	0.0001	-0.0030	0.0187	-0.2143	0.0022	-0.4252	-0.0018	0.0798	0.0007	0.0265	0.1538	0.7038	0.0027	0.4246	0.9712
ITL, 10y, d	0.3571	0.0050	0.0001	-0.0027	0.0139	-0.1353	0.0015	-0.2920	0.0000	0.6532	0.0005	0.6028	0.7056	1.4185	0.0003	1.3885	1.4480
ITL, 10y, w	0.3456	0.0064	0.0001	-0.0045	0.0186	-0.1891	0.0022	-0.4020	0.0000	0.6360	0.0015	0.4942	0.7845	1.3687	0.0009	1.2831	1.4552
ITL, 10y, m	0.3450	0.0105	0.0001	-0.0038	0.0262	-0.2282	0.0028	-0.5055	0.0000	1.0270	0.0054	0.5949	1.5065	1.4180	0.0021	1.2515	1.5937
CHF, 10y, d	0.3584	0.0005	0.0000	-0.0002	0.0011	-0.2785	0.0022	-0.4870	-0.0442	0.1403	0.0002	0.1215	0.1593	0.8021	0.0003	0.7756	0.8276
CHF, 10y, w	0.3871	0.0008	0.0000	-0.0003	0.0019	-0.4111	0.0040	-0.7913	-0.0596	0.2561	0.0009	0.1785	0.3396	0.8414	0.0007	0.7770	0.8989
CHF, 10y, m	0.3778	0.0010	0.0000	-0.0002	0.0022	-0.3879	0.0043	-0.7440	-0.0010	0.8218	0.0063	0.3289	1.4411	1.0039	0.0013	0.8689	1.1383
AUD, 10y, d	0.3479	0.0069	0.0001	-0.0008	0.0159	-0.1309	0.0016	-0.2857	-0.0001	0.0067	0.0000	0.0051	0.0085	0.1206	0.0008	0.0341	0.2043
AUD, 10y, w	0.3507	0.0133	0.0002	-0.0012	0.0298	-0.2517	0.0031	-0.5390	-0.0001	0.0161	0.0001	0.0070	0.0274	0.2698	0.0020	0.0311	0.4830
AUD, 10y, m	0.3391	0.0194	0.0003	-0.0011	0.0407	-0.3659	0.0049	-0.7543	0.0000	0.0272	0.0005	0.0077	0.0685	0.3101	0.0044	0.0004	0.6926
EUR, 10y, d	0.3821	0.0006	0.0000	-0.0005	0.0017	-0.1126	0.0008	-0.1906	-0.0304	0.0535	0.0001	0.0476	0.0594	0.8014	0.0003	0.7725	0.8287
EUR, 10y, w	0.3792	0.0013	0.0000	-0.0014	0.0041	-0.1855	0.0016	-0.3367	-0.0066	0.0605	0.0002	0.0417	0.0796	0.6757	0.0008	0.5968	0.7554
EUR, 10y, m	0.3548	0.0032	0.0000	-0.0013	0.0078	-0.2614	0.0026	-0.5123	-0.0009	0.1252	0.0009	0.0539	0.2225	0.7035	0.0018	0.5325	0.8850

Table A.4.: This table shows descriptive statistics obtained for the MCMC chains used to compute the Bayes factors for time series observed over 25 years. *Accept.* stands for the acceptance ratio, *mean* for the mean of all samples, *SE* for the standard error as described in Section 2.3.3, and *95%-HPD* for the smallest interval comprising 95% of all samples from the posterior distribution.

	Accept.	α				β				σ				ξ			
Timeseries		mean	SE	95%-HPD		mean	SE	95%-HPD		mean	SE	95%-HPD		mean	SE	95%-HPD	
USD, 25y, d	0.3549	-0.0006	0.0000	-0.0014	0.0002	-0.0430	0.0006	-0.0913	0.0000	0.0989	0.0001	0.0934	0.1048	0.8112	0.0002	0.7952	0.8279
USD, 25y, w	0.3508	0.0000	0.0000	-0.0024	0.0023	-0.0743	0.0008	-0.1623	0.0000	0.0823	0.0001	0.0686	0.0956	0.6219	0.0005	0.5748	0.6690
USD, 25y, m	0.3724	0.0029	0.0000	-0.0025	0.0083	-0.1209	0.0012	-0.2479	-0.0018	0.0384	0.0002	0.0240	0.0556	0.3324	0.0012	0.2125	0.4506
GBP, 25y, d	0.3769	-0.0002	0.0000	-0.0008	0.0004	-0.0853	0.0006	-0.1416	-0.0284	0.3392	0.0002	0.3181	0.3623	1.2975	0.0002	1.2759	1.3186
GBP, 25y, w	0.3589	-0.0009	0.0000	-0.0026	0.0006	-0.0664	0.0007	-0.1333	0.0000	0.1595	0.0003	0.1313	0.1889	0.9477	0.0006	0.8881	1.0086
GBP, 25y, m	0.3582	0.0004	0.0001	-0.0047	0.0054	-0.0733	0.0009	-0.1573	0.0000	0.0483	0.0002	0.0318	0.0675	0.4527	0.0011	0.3286	0.5745
JPY, 25y, d	0.3589	0.0009	0.0000	0.0005	0.0014	-0.1570	0.0018	-0.3009	-0.0019	0.1222	0.0001	0.1130	0.1318	0.6681	0.0001	0.6534	0.6829
JPY, 25y, w	0.3709	0.0015	0.0000	0.0007	0.0022	-0.2350	0.0022	-0.4497	-0.0059	0.1569	0.0003	0.1267	0.1905	0.6706	0.0004	0.6338	0.7109
JPY, 25y, m	0.3742	0.0011	0.0000	0.0004	0.0018	-0.2040	0.0018	-0.3931	-0.0102	0.1219	0.0004	0.0836	0.1654	0.6525	0.0007	0.5912	0.7184
CHF, 25y, d	0.3391	0.0005	0.0000	-0.0002	0.0012	-0.0860	0.0011	-0.1932	-0.0002	0.2385	0.0001	0.2225	0.2544	0.8459	0.0001	0.8295	0.8618
CHF, 25y, w	0.3628	0.0010	0.0000	-0.0003	0.0022	-0.1313	0.0018	-0.2969	0.0000	0.2738	0.0005	0.2302	0.3236	0.8118	0.0004	0.7688	0.8507
CHF, 25y, m	0.3661	0.0015	0.0000	-0.0001	0.0033	-0.1391	0.0019	-0.2951	0.0000	0.1655	0.0007	0.1083	0.2281	0.6936	0.0009	0.6102	0.7858
AUD, 25y, d	0.3766	0.0067	0.0001	0.0009	0.0127	-0.1318	0.0012	-0.2352	-0.0237	0.1805	0.0001	0.1659	0.1947	1.1838	0.0003	1.1548	1.2137
AUD, 25y, w	0.3834	0.0102	0.0001	0.0010	0.0199	-0.1971	0.0016	-0.3598	-0.0401	0.1533	0.0003	0.1200	0.1848	1.0199	0.0007	0.9424	1.0969
AUD, 25y, m	0.3679	0.0135	0.0001	0.0025	0.0240	-0.2530	0.0020	-0.4301	-0.0711	0.1337	0.0007	0.0752	0.2017	0.9225	0.0018	0.7474	1.0990
DEM, 25y, d	0.3647	0.0002	0.0000	-0.0006	0.0011	-0.0375	0.0005	-0.0802	0.0000	0.1316	0.0001	0.1229	0.1406	0.9814	0.0002	0.9612	1.0004
DEM, 25y, w	0.3545	0.0008	0.0000	-0.0012	0.0029	-0.0621	0.0008	-0.1344	-0.0001	0.0892	0.0002	0.0740	0.1056	0.7608	0.0006	0.7095	0.8111
DEM, 25y, m	0.3640	0.0028	0.0001	-0.0020	0.0077	-0.0982	0.0012	-0.2074	-0.0004	0.0299	0.0001	0.0168	0.0440	0.3432	0.0012	0.2080	0.4787

Table A.5.: This table shows convergence diagnostics presented in Section 2.3.3 for the MCMC chains with flat priors for the ten year window.

	$\rho(20)$				KS				Gelman&Rubin					Effective size			
Timeseries	α	β	σ	ξ	α	β	σ	ξ	α	β	σ	ξ	chain	α	β	σ	ξ
USD, 10y, d	0.08	0.09	0.06	0.07	0.71	0.71	0.37	0.55	1.00	1.00	1.00	1.00	1.00	2807	2696	2997	2845
USD, 10y, w	0.09	0.06	0.04	0.05	0.91	0.91	0.47	0.40	1.00	1.00	1.00	1.00	1.00	2553	2900	2811	2791
USD, 10y, m	0.10	0.08	0.07	0.06	0.80	0.80	0.96	0.81	1.00	1.00	1.00	1.00	1.00	2575	2564	2696	2725
GBP, 10y, d	0.07	0.07	0.03	0.04	0.93	0.93	0.22	0.10	1.00	1.00	1.00	1.00	1.00	2727	2723	3179	3114
GBP, 10y, w	0.07	0.09	0.04	0.04	0.25	0.25	0.47	0.72	1.00	1.00	1.00	1.00	1.00	2932	2717	3033	2899
GBP, 10y, m	0.12	0.06	0.06	0.06	0.90	0.90	0.16	0.47	1.00	1.00	1.00	1.00	1.00	2263	2934	2883	2789
JPY, 10y, d	0.10	0.09	0.06	0.06	0.74	0.74	0.73	0.90	1.00	1.00	1.00	1.00	1.00	2400	2568	2688	2654
JPY, 10y, w	0.07	0.08	0.08	0.07	0.31	0.31	0.45	0.11	1.00	1.00	1.00	1.00	1.00	2693	2555	2475	2514
JPY, 10y, m	0.10	0.11	0.09	0.09	0.90	0.90	0.02	0.38	1.00	1.00	1.00	1.00	1.00	2575	2469	2467	2399
FRF, 10y, d	0.07	0.07	0.07	0.08	0.99	0.99	0.52	0.53	1.00	1.00	1.00	1.00	1.00	2838	2903	2552	2522
FRF, 10y, w	0.08	0.07	0.07	0.07	0.38	0.38	0.63	0.92	1.00	1.00	1.00	1.00	1.00	2642	2897	2639	2491
FRF, 10y, m	0.07	0.08	0.05	0.05	0.65	0.65	0.57	0.61	1.00	1.00	1.00	1.00	1.00	2638	2628	2780	2659
DEM, 10y, d	0.06	0.06	0.03	0.03	0.44	0.44	0.87	0.96	1.00	1.00	1.00	1.00	1.00	3052	3045	2906	2868
DEM, 10y, w	0.08	0.06	0.05	0.05	0.74	0.74	0.76	0.54	1.00	1.00	1.00	1.00	1.00	2858	2999	2723	2778
DEM, 10y, m	0.06	0.06	0.03	0.02	0.59	0.59	0.75	0.90	1.00	1.00	1.00	1.00	1.00	2872	2983	3058	2971
ITL, 10y, d	0.06	0.06	0.04	0.05	0.66	0.66	0.95	0.91	1.00	1.00	1.00	1.00	1.00	2932	3060	2857	2844
ITL, 10y, w	0.09	0.06	0.03	0.03	0.50	0.50	0.95	0.79	1.00	1.00	1.00	1.00	1.00	2654	2869	2963	2847
ITL, 10y, m	0.07	0.05	0.08	0.08	0.54	0.54	0.85	0.79	1.00	1.00	1.00	1.00	1.00	2705	2970	2568	2536
CHF, 10y, d	0.09	0.06	0.05	0.05	0.54	0.54	0.95	0.65	1.00	1.00	1.00	1.00	1.00	2582	2946	2932	2817
CHF, 10y, w	0.10	0.08	0.07	0.07	0.71	0.71	0.47	0.88	1.00	1.00	1.00	1.00	1.00	2569	2769	2819	2734
CHF, 10y, m	0.09	0.07	0.06	0.06	0.72	0.72	0.10	0.24	1.00	1.00	1.00	1.00	1.00	2713	2473	2652	2748
AUD, 10y, d	0.07	0.07	0.06	0.06	0.47	0.47	0.20	0.51	1.00	1.00	1.00	1.00	1.00	2748	2742	2680	2670
AUD, 10y, w	0.08	0.08	0.04	0.04	0.56	0.56	0.50	0.79	1.00	1.00	1.00	1.00	1.00	2738	2795	2872	2848
AUD, 10y, m	0.07	0.07	0.05	0.07	0.38	0.38	0.40	0.48	1.00	1.00	1.00	1.00	1.00	2658	2673	2611	2457
EUR, 10y, d	0.09	0.07	0.07	0.07	0.81	0.81	0.05	0.64	1.00	1.00	1.00	1.00	1.00	2775	2960	2749	2647
EUR, 10y, w	0.07	0.05	0.04	0.05	0.97	0.97	0.54	0.57	1.00	1.00	1.00	1.00	1.00	2672	2771	3234	3031
EUR, 10y, m	0.09	0.06	0.04	0.04	0.63	0.63	0.57	0.57	1.00	1.00	1.00	1.00	1.00	2526	2780	2766	2810

Table A.6.: This table shows convergence diagnostics presented in Section 2.3.3 for the MCMC chains with flat priors for the 25 year window.

	$\rho(20)$				KS				Gelman&Rubin					Effective size			
Timeseries	α	β	σ	ξ	α	β	σ	ξ	α	β	σ	ξ	chain	α	β	σ	ξ
USD, 25y, d	0.08	0.06	0.03	0.02	0.73	0.73	0.35	0.60	1.00	1.00	1.00	1.00	1.00	2606	2933	2820	2796
USD, 25y, w	0.11	0.06	0.05	0.05	0.54	0.54	0.50	0.18	1.00	1.00	1.00	1.00	1.00	2439	2842	2642	2634
USD, 25y, m	0.10	0.05	0.07	0.07	0.62	0.62	0.90	0.94	1.00	1.00	1.00	1.00	1.00	2422	2744	2927	2821
GBP, 25y, d	0.08	0.08	0.05	0.06	0.62	0.62	0.04	0.04	1.00	1.00	1.00	1.00	1.00	2790	2882	2742	2640
GBP, 25y, w	0.09	0.06	0.03	0.03	0.41	0.41	0.64	0.71	1.00	1.00	1.00	1.00	1.00	2652	2869	3153	3114
GBP, 25y, m	0.08	0.08	0.04	0.05	0.30	0.30	0.50	0.31	1.00	1.00	1.00	1.00	1.00	2514	2590	2971	2843
JPY, 25y, d	0.07	0.08	0.06	0.06	0.22	0.22	0.10	0.03	1.00	1.00	1.00	1.00	1.00	2820	2581	3090	3002
JPY, 25y, w	0.09	0.06	0.06	0.06	0.26	0.26	0.08	0.34	1.00	1.00	1.00	1.00	1.00	2430	3036	2970	2890
JPY, 25y, m	0.11	0.05	0.06	0.07	0.66	0.66	0.24	0.96	1.00	1.00	1.00	1.00	1.00	2458	3010	2754	2700
CHF, 25y, d	0.11	0.07	0.03	0.03	0.74	0.74	0.10	0.19	1.00	1.00	1.00	1.00	1.00	2394	2881	3238	3172
CHF, 25y, w	0.10	0.10	0.08	0.08	0.34	0.34	0.53	0.35	1.00	1.00	1.00	1.00	1.00	2439	2702	2593	2586
CHF, 25y, m	0.08	0.08	0.08	0.08	0.33	0.33	0.16	0.61	1.00	1.00	1.00	1.00	1.00	2626	2715	2612	2565
AUD, 25y, d	0.04	0.04	0.04	0.04	0.09	0.09	0.82	0.88	1.00	1.00	1.00	1.00	1.00	3150	3082	2831	2738
AUD, 25y, w	0.06	0.06	0.03	0.03	0.51	0.51	0.25	0.53	1.00	1.00	1.00	1.00	1.00	2865	2977	3113	3117
AUD, 25y, m	0.05	0.06	0.04	0.04	0.55	0.55	0.79	0.64	1.00	1.00	1.00	1.00	1.00	3036	3037	2683	2819
DEM, 25y, d	0.08	0.07	0.07	0.06	0.10	0.10	0.80	0.55	1.00	1.00	1.00	1.00	1.00	2833	3049	2524	2557
DEM, 25y, w	0.08	0.10	0.05	0.05	0.71	0.71	0.15	0.08	1.00	1.00	1.00	1.00	1.00	2688	2659	2627	2580
DEM, 25y, m	0.06	0.07	0.05	0.05	0.63	0.63	0.40	0.81	1.00	1.00	1.00	1.00	1.00	2689	2733	2734	2617

B. Additional results for Chapter 4

B.1. The drift term for the transformed interest rate process

The short-rate r_t has the following dynamics:

$$dr_t = \kappa(\theta - r_t)dt + \sigma_r r_t^{\xi} dW_t. \tag{B.1}$$

If $\xi = 1$, the transformed process takes the form

$$X_t = \frac{1}{\sigma_r} \ln(r_t). \tag{B.2}$$

An application of Itô's lemma yields:

$$dX_t = d\left(\frac{1}{\sigma_r} \ln(r_t)\right) = \left(\frac{\kappa\theta}{r_t\sigma_r} - \frac{\kappa}{\sigma_r} - \frac{1}{2}\sigma_r\right) dt + dW_t. \tag{B.3}$$

Equation (B.2) implies $r_t = \exp\{\sigma_r X_t\}$. Substituting this into (B.3) I obtain

$$dX_t = \left(\frac{\kappa\theta}{\sigma_r} \exp\{-\sigma_r X_t\} - \frac{\kappa}{\sigma_r} - \frac{1}{2}\sigma_r\right) dt + dW_t. \tag{B.4}$$

If $\xi \neq 1$, X_t is obtained by the transformation

$$X_t = \frac{r_t^{1-\xi}}{\sigma_r(1-\xi)}. \tag{B.5}$$

By Itô's lemma the stochastic differential equation for X_t is

$$dX_t = \left(\frac{\kappa\theta}{\sigma_r} r_t^{-\xi} - \frac{\kappa}{\sigma_r} r_t^{1-\xi} - \xi\sigma_r r_t^{\xi-1}\right) dt + dW_t. \tag{B.6}$$

With $r_t = (X_t\sigma_r(1-\xi))^{\frac{1}{1-\xi}}$ as implied by equation (B.5), X_t has the following dynamics:

$$dX_t = \left(\frac{\kappa\theta}{\sigma_r} (X_t\sigma_r(1-\xi))^{\frac{\xi}{\xi-1}} - \kappa X_t(1-\xi) - \frac{\xi}{X_t(1-\xi)}\right) dt + dW_t. \tag{B.7}$$

B.2. The drift term for the transformed stock process

The transformation of the stock process is very similar to that of the interest rate process. It's dynamics are given by

$$dS_t = r_t S_t dt + \sigma_S S_t^{\alpha} dW_t^S. \tag{B.8}$$

I again distinguish the transformation by the value of α. For $\alpha = 1$

$$Y_t = \frac{1}{\sigma_S} \ln S_t. \tag{B.9}$$

By applying Itô's lemma, I obtain

$$dY_t = \left(\frac{1}{\sigma_S} r_t - \frac{1}{2} \sigma_S \right) dt + dW_t^S. \tag{B.10}$$

Of course, Y_t should depend on X_t and not on r_t. Substituting for r_t yields

$$dY_t = \begin{cases} \left(\frac{1}{\sigma_S} \exp\{\sigma_r X_t\} - \frac{1}{2} \sigma_S \right) dt + dW_t^S & \xi = 1 \\ \left(\frac{1}{\sigma_S} \left(X_t \sigma_r (1-\xi) \right)^{\frac{1}{1-\xi}} - \frac{1}{2} \sigma_S \right) dt + dW_t^S & \xi \neq 1 \end{cases}. \tag{B.11}$$

For $\alpha \neq 1$, I can compute Y_t as

$$Y_t = \frac{S_t^{1-\alpha}}{\sigma_S (1-\alpha)}. \tag{B.12}$$

As usual an application of Itô's lemma yields the new dynamics for the Y_t process:

$$dY_t = \left(\frac{1}{\sigma_S} r_t S_t^{1-\alpha} - \frac{1}{2} \alpha \sigma_S S_t^{\alpha-1} \right) dt + dW_t^S. \tag{B.13}$$

Substituting the values for r_t and S_t, I obtain the new dynamics of Y_t in terms of X_t and Y_t:

$$dY_t = \begin{cases} \left((1-\alpha) \exp\{\sigma_r X_t\} Y_t - \frac{\alpha}{2(1-\alpha) Y_t} \right) dt + dW_t^S & \xi = 1 \\ \left((1-\alpha) \left(X_t \sigma_r (1-\xi) \right)^{\frac{1}{1-\xi}} Y_t - \frac{\alpha}{2(1-\alpha) Y_t} \right) dt + dW_t^S & \xi \neq 1 \end{cases}. \tag{B.14}$$

Bibliography

Aït-Sahalia, Yacine (1996). "Testing Continuous-Time Models of the Spot Interest Rate". In: *The Review of Financial Studies* 9.2, pp. 385–426.

— (1999). "Transition Densities for Interest Rate and Other Nonlinear Diffusions". In: *The Journal of Finance* 54.4, pp. 1361–1395.

— (2002). "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach". In: *Econometrica* 70.1, pp. 223–262.

— (2008). "Closed-form likelihood expansions for multivariate diffusions". In: *The Annals of Statistics* 36.2, pp. 906–937.

Andricopoulos, Ari D., Martin Widdicks, Peter W. Duck, and David P. Newton (2003). "Universal option valuation using quadrature methods". In: *Journal of Financial Economics* 67.3, pp. 447–471.

Andricopoulos, Ari D., Martin Widdicks, David P. Newton, and Peter W. Duck (2007). "Extending quadrature methods to value multi-asset and complex path dependent options". In: *Journal of Financial Economics* 83.2, pp. 471–499.

Bacinello, Anna R. (2001). "Fair Pricing of Life Insurance Participating Policies with a Minimum Interest Rate Guaranteed". In: *ASTIN Bulletin* 31.2, pp. 275–297.

— (2003). "Fair Valuation of a Guaranteed Life Insurance Participating Contract Embedding a Surrender Option". In: *Journal of Risk and Insurance* 70.3, pp. 461–487.

Basel Committee (1999). *A new capital adequacy framework*. Tech. rep. March.

— (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Tech. rep. June.

Bauer, Daniel, Rüdiger Kiesel, Alexander Kling, and Jochen Ruß (2006). "Risk-neutral valuation of participating life insurance contracts". In: *Insurance: Mathematics and Economics* 39.2, pp. 171–183.

Bayes, Thomas (1764). "An essay towards solving a problem in the doctrine of chances". In: *Philosophical Transactions of the Royal Society* 53, pp. 370–418.

Beckers, Stan (1980). "The constant elasticity of variance model and its implications for option pricing". In: *Journal of Finance* 35.3, pp. 661–673.

Berger, James O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer.

Berkowitz, Jeremy and James O'Brien (2002). "How Accurate Are Value at Risk Models at Commercial Banks?" In: *The Journal of Finance* 57.3, pp. 1093–1111.

Bernardo, José M. and Adrian F. M. Smith (2001). *Bayesian theory*. Chichester: John Wiley & Sons.

Black, Fischer and Robert Litterman (1991). "Asset Allocation: Combining investor views with market equilibrium". In: *Journal of Fixed Income* 1.2, pp. 7–18.

Black, Fischer and Myron Scholes (1973). "The pricing of options and corporate liabilities". In: *The Journal of Political Economy*, pp. 637–654.

Bliss, Robert R. and David C. Smith (1998). "The elasticity of interest rate volatility: Chan, Karolyi, Longstaff, and Sanders revisited". In: *Journal of Risk* 1.1, pp. 21–246.

Box, George E. P. and David R. Cox (1964). "An analysis of transformations". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 26.2, pp. 211–252.

Boyle, Phelim P. (1988). "A lattice framework for option pricing with two state variables". In: *Journal of Financial and Quantitative Analysis* 23.1, pp. 1–12.

Brailsford, Tim and Krishnan Maheswaran (1998). "The Dynamics of the Australian Short-Term Interest Rate". In: *Australian Journal of Management* 23.2, pp. 213–234.

Brennan, Michael J. and Eduardo S. Schwartz (1978). "Finite Difference Methods and Jump Processes Arising in the Pricing of Contingent Claims". In: *The Journal of Financial and Quantitative Analysis* 13.3, pp. 461–474.

— (1980). "Analyzing Convertible Bonds". In: *Journal of Financial and Quantitative Analysis* 15.4, pp. 907–929.

Brigo, Damiano and Fabio Mercurio (2007). *Interest Rate Models - Theory and Practice*. 2nd ed. Berlin: Springer.

Briys, Eric and François De Varenne (1997). "On the risk of insurance liabilities: debunking some common pitfalls". In: *Journal of Risk and Insurance* 64.4, pp. 673–694.

Brown, Stephen J. (1978). "The portfolio choice problem: Comparison of certainty equivalent and optimal Bayes portfolios". In: *Communications in Statistics: Simulation and Computation* 7.4, pp. 321–334.

Cairns, Andrew J. G. (2000). "A discussion of parameter and model uncertainty in insurance". In: *Insurance: Mathematics and Economics* 27.3, pp. 313–330.

Chan, Kalok C., G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders (1992). "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate". In: *The Journal of Finance* 47.3, pp. 1209–1227.

Chib, Siddhartha and Edward Greenberg (1996). "Markov chain Monte Carlo simulation methods in econometrics". In: *Econometric theory* 12.03, pp. 409–431.

Christoffersen, Peter F. (1998). "Evaluating interval forecasts". In: *International economic review* 39.4, pp. 841–862.

Cont, Rama (2006). "Model uncertainty and its impact on the pricing of derivative instruments". In: *Mathematical Finance* 16.3, pp. 519–547.
Coval, Joshua, Jakub Jurek, and Erik Stafford (2009). "The Economics of Structured Finance". In: *Journal of Economic Perspectives* 23.1, pp. 3–25.
Cox, John C. (1975). "Notes On Option Pricing I: Constant Elasticity of Variance Diffusions".
Cox, John C., Johnathan E. Ingersoll, and Stephen A. Ross (1985). "A theory of the term structure of interest rates". In: *Econometrica: Journal of the Econometric Society* 53.2, pp. 385–407.
Cox, John C. and Stephen A. Ross (1976). "The valuation of options for alternative stochastic processes". In: *Journal of Financial Economics* 3.1, pp. 145–166.
Cox, John C., Stephen A. Ross, and Mark Rubinstein (1979). "Option Pricing: A Simplified Approach". In: *Journal of Financial Economics* 7.3, pp. 229–263.
Dai, Tian-Shyr, Guan-Shieng Huang, and Yuh-Dauh Lyuu (2005). "An efficient convergent lattice algorithm for European Asian options". In: *Applied Mathematics and Computation* 169.2, pp. 1458–1471.
Derman, Emanuel (1996). "Model Risk". In: *Goldman Sachs Quantitative Strategies Research Notes* April.
Dickey, James M. (1971). "The weighted likelihood ratio, linear hypotheses on normal location parameters". In: *The Annals of Mathematical Statistics* 42.1, pp. 204–223.
Dothan, L. Uri (1978). "On the term structure of interest rates". In: *Journal of Financial Economics* 6.1, pp. 59–69.
Draper, David (1995). "Assessment and Propagation of Model Uncertainty". In: *Journal of the Royal Statistical Society. Series B* 57.1, pp. 45–97.
Elerian, Ola, Siddhartha Chib, and Neil Shephard (2001). "Likelihood inference for discretely observed nonlinear diffusions". In: *Econometrica* 69.4, pp. 959–993.
Emanuel, David C. and James D. MacBeth (1982). "Further Results on the Constant Elasticity of Variance Call Option Pricing Model". In: *The Journal of Financial and Quantitative Analysis* 17.4, pp. 533–554.
Eraker, Bjørn (2001). "MCMC Analysis of Diffusion Models With Application to Finance". In: *Journal of Business and Economic Statistics* 19.2, pp. 177–191.
Fernandez, Carmen, Eduardo Ley, and Mark F. J. Steel (2001). "Model uncertainty in cross-country growth regressions". In: *Journal of Applied Econometrics* 16.5, pp. 563–576.
French, Kenneth R. and Richard Roll (1986). "Stock return variances: The arrival of information and the reaction of traders". In: *Journal of Financial Economics* 17.1, pp. 5–26.

Fusai, Gianluca and Maria C. Recchioni (2007). "Analysis of quadrature methods for pricing discrete barrier options". In: *Journal of Economic Dynamics and Control* 31.3, pp. 826–860.

Gamerman, Dani and Hedibert F. Lopes (2006). *Markov Chain Monte Carlo*. 2nd ed. Boca Raton: Chapman and Hall.

Gatzert, Nadine and Alexander Kling (2007). "Analysis of participating life insurance contracts: A unification approach". In: *Journal of Risk and Insurance* 74.3, pp. 547–570.

Gelfand, Alan E. and Adrian F. M. Smith (1990). "Sampling based approaches to calculating marginal densities". In: *Journal of the American Statistical Association* 85.410, pp. 398–409.

Gelman, Andrew (1996). "Inference and monitoring convergence". In: *Markov chain Monte Carlo in Practice*. Ed. by Walter. R. Gilks, Sylvia Richardson, and David J. Spiegelhalter. New York: Champman Hall, pp. 131–143.

Gelman, Andrew and Donald B. Rubin (1992). "Inference from iterative simulation using multiple sequences (with discussion)". In: *Statistical Science* 7.4, pp. 457–511.

Geman, Stuart and Donald Geman (1984). "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images". In: *IEEE Transactions on Pattern Analysis Matching Intelligence* 6, pp. 721–741.

George, Edward I. and Merlise Clyde (2004). "Model Uncertainty". In: *Statistical Science* 19.1, pp. 81–94.

Gibson, Rajna, François-Serge Lhabitant, Nathalie Pistre, and Denis Talay (1998). "Interest rate model risk: an overview". In: *Journal of Risk* 1.3, pp. 37–62.

Glasserman, Paul (2003). *Monte Carlo Method in Financial Engineering*. New York, NY: Springer.

Gössl, Christoff (2005). "Predictions based on certain uncertainties - a Bayesian credit portfolio approach".

Gray, Philip (2005). "Bayesian Estimation of Short-Rate Models". In: *Australian Journal of Management* 30.1, pp. 1–22.

Green, T. Clifton and Stephen Figlewski (2002). "Market risk and model risk for a financial institution writing options". In: *The Journal of Finance* 54.4, pp. 1465–1499.

Grosen, Anders and Peter Løchte Jørgensen (2002). "Life Insurance Liabilities at Market Value: An Analysis of Insolvency Risk, Bonus Policy, and Regulatory Intervention Rules in a Barrier Option Framework". In: *Journal of Risk and Insurance* 69.1, pp. 63–91.

Grosen, Andres and Peter Løchte Jørgensen (2000). "Fair valuation of life insurance liabilities: the impact of interest rate guarantees, surrender options, and bonus policies". In: *Insurance: Mathematics and Economics* 26.1, pp. 37–57.

Han, Cong and Bradley P. Carlin (2001). "MCMC methods for computing Bayes factors: a comparative review." In: *American Statistical Association* 96.455, pp. 1122–1132.

Hansen, Nikolaus and Andreas Ostermeier (1996). "Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation". In: *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation*, pp. 312–317.

Hardy, Mary R. (2002). "Bayesian risk management for equity-linked insurance". In: *Scandinavian Actuarial Journal* 2002.3, pp. 185–211.

Hastings, W. Keith (1970). "Monte Carlo sampling methods using Markov chains and their application". In: *Biometrika* 57.1, pp. 97–109.

Heitfield, Erik (2009). "Parameter uncertainty and the credit risk of collateralized debt obligations".

Hilliard, Jimmy E., Adam L. Schwartz, and Alan L. Tucker (1996). "Bivariate binomial options pricing with generalized interest rate processes". In: *Journal of Financial Research* 19.4, pp. 585–602.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky (1999). "Bayesian Model Averaging: A Tutorial". In: *Statistical Science* 14.4, pp. 382–417.

Hsu, William Wei-Yuan and Yuh-Dauh Lyuu (2007). "A convergent quadratic-time lattice algorithm for pricing European-style Asian options". In: *Applied Mathematics and Computation* 189.2, pp. 1099–1123.

Hull, John and Wulin Suo (2002). "A methodology for assessing model risk and its application to the implied volatility function model". In: *Journal of Financial and Quantitative Analysis* 37.2, pp. 297–318.

Hull, John C. and Alan D. White (1994a). "Numerical Procedures for Implementing Term Structure Models I". In: *The Journal of Derivatives* 2.1, pp. 7–16.

— (1994b). "Numerical Procedures for Implementing Term Structure Models II: Two-Factor Models". In: *The Journal of Derivatives* 2.2, pp. 37–48.

Jeffreys, Harold (1961). *Theory of Probability*. 3rd ed. Oxford.

Jones, Christopher S. (1998). "Bayesian estimation of continuous-time finance models".

Kamrad, Bardia and Peter Ritchken (1991). "Multinomial approximating models for options with k state variables". In: *Management Science* 37.12, pp. 1640–1652.

Kass, Robert E. and Adrian E. Raftery (1995). "Bayes Factors". In: *Journal of the American Statistical Association* 90.430, pp. 773–795.

Kerkhof, Franciscus L. J. (2003). "Model Risk Analysis for Risk Management and Option Pricing". PhD thesis. Tilburg University.

Kerkhof, Jeroen, Bertrand Melenberg, and Hans Schumacher (2010). "Model risk and capital reserves". In: *Journal of Banking & Finance* 34.1, pp. 267–279.

Klassen, Timothy R. (2001). "Simple, fast, and flexible pricing of Asian options". In: *Journal of Computational Finance* 4.3, pp. 89–124.
Kling, Alexander, Andreas Richter, and Jochen Ruß (2007). "The Impact of Surplus Distribution on the Risk Exposure of With Profit Life Insurance Policies Including Interest Rate Guarantees". In: *Journal of Risk* 74.3, pp. 571–589.
Koop, Gary and Lise Tole (2004). "Measuring the health effects of air pollution: to what extent can we really say that people are dying from bad air?" In: *Journal of Environmental Economics and Management* 47.1, pp. 30–54.
Laplace, Pierre S. (1774). "Mémoir sur la probabilité des causes par le évènements". In: *Mémoires de mathématique et de physique presentés à l'Académie royale des sciences, par divers savans, lûs dans ses assemblées* 6, pp. 621–656.
Lauterbach, Beni and Paul Schultz (1990). "Pricing warrants: an empirical study of the Black-Scholes model and its alternatives". In: *Journal of Finance* 45.4, pp. 1181–1209.
Løchte, Peter (2007). "Traffic Light Options". In: *Journal of Banking & Finance* 31.12, pp. 3698–3719.
Löffler, Gunter (2003). "The effects of estimation error on measures of portfolio credit risk". In: *Journal of Banking & Finance* 27.8, pp. 1427–1453.
Luoma, Arto, Anne Puustelli, and Lasse Koskinen (2008). "Bayesian analysis of participating life insurance contracts with American-style options". In: *Proceedings of the AFIR Colloquium, Rome*.
Madigan, David and Adrian E. Raftery (1994). "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occams Window". In: *Journal of the American Statistical Association* 89.428, pp. 1535–1546.
Markowitz, Henry M. (1952). "Portfolio selection". In: *Journal of Finance* 7.1, pp. 77–91.
Marsh, Terry A. and Eric R. Rosenfeld (1983). "Stochastic processes for interest rates and equilibrium bond prices". In: *The Journal of Finance* 38.2, pp. 635–646.
McNeil, Alexander J., Rüdiger Frey, and Paul Embrechts (2005). *Quantitative Risk Management*. Princeton: Princeton University Press.
Merton, Robert C. (1973). "Theory of rational option pricing". In: *The Bell Journal of Economics and Management Science* 4.1, pp. 141–183.
Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller (1953). "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092.
Meucci, Attilio (2005). *Risk and Asset Allocation*. New York: Springer.
Meyn, Sean P. and Richard L. Tweedie (1993). *Markov Chains and Stochastic Stability*. Berlin: Springer.
Milev, Mariyan and Aldo Tagliani (2010). "Numerical valuation of discrete double barrier options". In: *Journal of Computational and Applied Mathematics* 233.10, pp. 2468–2480.

Morales, Knashawn H., Joseph G. Ibrahim, Chien-Jen Chen, and Louise M. Ryan (2006). "Bayesian Model Averaging With Applications to Benchmark Dose Estimation for Arsenic in Drinking Water". In: *Journal of the American Statistical Association* 101.473, pp. 9–17.

Nawalkha, Sanjay K. and Natalia A. Beliaeva (2007). "Efficient Trees for CIR and CEV Short Rate Models". In: *The Journal of Alternative Investments* 10.1, pp. 71–90.

Nelson, Daniel B. and Krishna Ramaswamy (1990). "Simple binomial processes as diffusion approximations in financial models". In: *Review of Financial Studies* 3.3, pp. 393–430.

Nocedal, Jorge (1980). "Updating Quasi-Newton Matrices with Limited Storage". In: *Mathematics of Computation* 35.151, pp. 773–782.

Nowman, K. Ben (1997). "Gaussian Estimation of Single-Factor Continuous Time Models of The Term Structure of Interest Rates". In: *The Journal of Finance* 52.4, pp. 1695–1706.

— (1998). "Continuous-time short term interest rate models". In: *Applied Financial Economics* 8.4, pp. 401–407.

Ockenga, Tim, Christoph Schmitt, and Axel Großpietsch (2010). *Deutsche Lebensversicherer: Keine Atempause*. Tech. rep. Fitch Ratings, pp. 1–14.

Pastor, Lubo (2000). "Portfolio selection and asset pricing models". In: *Journal of Finance* 55.1, pp. 179–223.

Pedersen, Asger Roer (1995). "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations". In: *Scandinavian Journal of Statistics* 22.1, pp. 55–71.

Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines (2006). "CODA: Convergence Diagnosis and Output Analysis for MCMC". In: *R News* 6.1, pp. 7–11.

Rabinovitch, Ramon (1989). "Pricing Stock and Bond Options when the Default-Free Rate is Stochastic". In: *Journal of Financial and Quantitative Analysis* 24.4, pp. 447–457.

Rachev, Svetlozar T., John S. J. Hsu, Biliana S. Bagasheva, and Frank J. Fabozzi (2008). *Bayesian Methods in Finance*. Hoboken: John Wiley & Sons.

Raftery, Adrian E. (1996). "Hypothesis testing and model selection via posterior simulation". In: *Markov chain Monte Carlo in practice*, pp. 163–188.

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting (1997). "Bayesian Model Averaging for Linear Regression Models". In: *Journal of the American Statistical Association* 92, pp. 179–191.

Robert, Christian P. (2007). *The Bayesian Choice*. Springer Texts in Statistics. New York, NY: Springer New York.

Robert, Christian P. and George Casella (2004). *Monte Carlo Statistical Methods*. 2nd ed. New York: Springer.

Robert, Christian P., Nicolas Chopin, and Judith Rousseau (2009). "Harold Jeffreys Theory of Probability Revisited". In: *Statistical Science* 24.2, pp. 141–172.
Rubinstein, Mark (2000). "On the Relation Between Binomial and Trinomial Option Pricing Models". In: *The Journal of Derivatives* 8.2, pp. 47–50.
Sakia, R. M. (1992). "The Box-Cox Transformation Technique: A Review". In: *The Statistician* 41.2, pp. 169–178.
Sanford, Andrew D. and Gael M. Martin (2006). "Bayesian comparison of several continuous time models of the Australian short rate". In: *Accounting and Finance* 46.2, pp. 309–326.
Schwarz, Hans-Rudolf and Norbert Köckler (2011). *Numerische Mathematik*. 8th ed. Wiesbaden: Vieweg+Teubner.
Stigler, Stephen M. (1986). "Laplace's 1774 Memoir on Inverse Probability". In: *Statistical Science* 1.3, pp. 359–378.
Tanskanen, Antti J. and Jani Lukkarinen (2003). "Fair valuation of path-dependent participating life insurance contracts". In: *Insurance: Mathematics and Economics* 33.3, pp. 595–609.
Tarashev, Nikola (2010). "Measuring portfolio credit risk correctly: Why parameter uncertainty matters". In: *Journal of Banking & Finance* 34.9, pp. 2065–2076.
Tarashev, Nikola and Haibin Zhu (2008). "Specification and calibration errors in measures of portfolio credit risk: The case of the ASRF model". In: *International Journal of Central Banking* 4.2, pp. 129–173.
Treepongkaruna, Sirimon and Stephen Gray (2003). "On the robustness of short-term interest rate models". In: *Accounting and Finance* 43.1, pp. 87–121.
Tse, Anson H. T., David B. Thomas, and Wayne Luk (2009a). "Accelerating quadrature methods for option valuation". In: *17th IEEE Symposium on Field Programmable Custom Computing Machines*. IEEE, pp. 29–36.
— (2009b). "Option pricing with multi-dimensional quadrature architectures". In: *International Conference on Field-Programmable Technology, 2009.* Pp. 427–430.
Vasicek, Oldrich (1977). "An equilibrium characterization of the term structure". In: *Journal of Financial Economics* 5.2, pp. 177–188.
Vega, Joseph de la (1966). "Confusión de confusiones". In: *Extraordinary Popular Delusions and the Madness of Crowds*. Ed. by Charles MacKay. Wiley, pp. 125–211.
Verdinelli, Isabella and Larry Wasserman (1995). "Computing Bayes factors using a generalization of the Savage-Dickey density ratio". In: *Journal of the American Statistical Association* 90.430, pp. 614–618.
Wilks, Samuel S. (1938). "The large-sample distribution of the likelihood ratio for testing composite hypotheses". In: *The Annals of Mathematical Statistics* 9.1, pp. 60–62.

Wintle, Brendan A., Michel A. McCarthy, Chris T. Volinsky, and Rodney P. Kavanagh (2003). "The Use of Bayesian Model Averaging to Better Represent Uncertainty in Ecological Models". In: *Conservation Biology* 17.6, pp. 1579–1590.

Yeung, Ka Yee, Roger E. Bumgarner, and Adrian E. Raftery (2005). "Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data". In: *Bioinformatics* 21.10, pp. 2394–2402.

Yu, Bin (1994). "Monitoring the convergence of Markov samplers based on estimated L^1 error". Berkley.

Yu, Bin and Per Mykland (1998). "Looking at Markov samplers through cusum path plots: A simple diagnostic idea". In: *Statstics and Computing* 8.3, pp. 275–286.

Zaglauer, Katharina and Daniel Bauer (2008). "Risk-neutral valuation of participating life insurance contracts in a stochastic interest rate environment". In: *Insurance: Mathematics and Economics* 43.1, pp. 29–40.

Zellner, Arnold and V. Karuppan Chetty (1965). "Prediction and decision problems in regression models from the Bayesian point of view". In: *Journal of the American Statistical Association* 60.310, pp. 608–615.

Zusammenfassung

In der vorliegenden Arbeit beschäftige ich mich mit der Frage wie statistische Schätzunsicherheiten, die sich bei der Bestimmung von parametrischen Modellen ergeben, bei der Berechnung von Risikokennzahlen berücksichtigt werden können und welche Effekte die Berücksichtigung dieser zusätzlichen Unsicherheiten hat. Diese Frage wird anhand von zwei Fallstudien untersucht, die z.T. eine erhebliche technische Vorarbeit erfordern. Neben einer kurzen Einleitung gliedert sich die Arbeit in vier Hauptkapitel und zwei Anhänge, deren wichtigste Ergebnisse im Folgenden kurz dargelegt werden.

Kapitel 2 führt die Bayesianischen statistischen Methoden ein, die in Kapitel 3 und 5 verwendet werden, und fasst die für diese Kapitel wichtigsten Resultate zusammen. Zunächst wird dieser Ansatz motiviert und die wichtigsten Gründe für seine Verwendung werden dargelegt. Mehrere Gründe sprechen für den Einsatz von Bayesianischen Methoden: Zunächst ist die Theorie von Bayesianischen Methoden weit entwickelt und der Ansatz ist konsistent und ermöglicht eine einfache Berücksichtigung von Parameter- und Modellunsicherheit. Des Weiteren passt dieser Ansatz zu der bereits bestehenden Maschinerie, die in der Industrie zur Bestimmung von Risikokennzahlen verwendet wird. Abschliessend sind durch die ansteigende Rechenleistung von Computern viele Methoden, die für den Bayesianischen Ansatz erforderlich sind, nun einfacher zu handhaben, was zu einer steigenden Popularität dieses Ansatzes geführt hat. Nach der Motivation wird im ersten Teil des Kapitels die Literatur zu Bayesianischen Methoden und ihrer Anwendung in der Ökonomie und im Riskiomanagement zusammengefasst.

Im nächsten Teil des zweiten Kapitels werden die Bayesianischen Methoden vorgestellt, wobei die Definition eines Bayesianischen Modells zusammen mit a-priori und a-posteriori Verteiliungen dargestellt werden. Außerdem wird mit dem *maximum a-posteriori estimator* (MAP) ein Bayesianischer Punktschätzer präsentiert, der auch in Kapitel 3 eingesetzt wird. In diesem Teil werden auch die *Bayes factors* diskutiert, bei denen es sich um ein objektives Bayesianisches Verfahren handelt. Dieser Abschnitt von Kapitel 2 schliesst mit der Diskussion von Marginalverteilungen und dem wichtigen Verfahren des *Bayesian model averaging*, also der Bayesianischen Durchschnittsbildung mehrerer Modelle.

Zusammenfassung

Abschnitt 2.3 beschreibt Berechnungsmethoden, die beim Einsatz des Bayesianischen Ansatzes ein zentrales Problem darstellen. Im Herzen dieser Methoden liegen so genannte *Markov chain Monte Carlo* (MCMC) Algorithmen, also Monte Carlo Simulationen bei denen die Realisierungen nicht voneinander unabhängig, sondern als Teil einer Markovkette generiert werden. Die Theorie der Markovketten ist daher eine Voraussetzung, um die Funktionsweise von MCMC Algorithmen zu verstehen. Aus diesem Grund werden die wichtigsten Ergebnisse aus der Markovkettentheorie kurz zusammengefasst. Danach werden die beiden wichtigsten MCMC Algorithmen vorgestellt, der *Gibbs Sampler* und der *Metropolis-Hastings* (MH) Algorithmus. Die Frage nach der Konvergenz einer MCMC Kette ist deutlich schwieriger zu beantworten, als dies etwa bei normalen Monte Carlo Simulationen der Fall ist. Daher befasst sich der letzte Teil dieses Kapitels mit verschiedenen Methoden zur Überprüfung der Konvergenz. Zunächst werden informale Methoden vorgestellt, die sich verschiedene grafische Eigentschaften von konvergierten MCMC Ketten zu Nutze machen. Danach wird auf formellere Methoden eingegangen, die an unterschiedlichen Testmethoden aus der klassischen Statistik angelehnt sind.

Kapitel 3 wendet Bayesianische Methoden auf die Schätzung verschiedener Short-Rate Modelle an, genauer gesagt auf Modelle die in dem Chan-Karolyi-Longstaff-Sanders Modell (Chan et al. (1992)) enthalten sind. Die Motivation dieses Kapitels besteht in den unterschiedlichen Schätzergebnissen für diese Modelle, die in der Literatur beispielsweise in den Studien von Chan et al. (1992), Brailsford and Maheswaran (1998), Treepongkaruna and Gray (2003), Gray (2005) und Sanford and Martin (2006) zu finden sind. Die Ergebnisse unterscheiden sich vor allem in den beiden Parametern, die die Ausprägung des Volatilitätsterms kontrollieren. Da beide Parameter eine starke, nicht-lineare Abhängigkeit aufweisen, liegt die Vermutung nahe, dass die unterschiedlichen Schätzergebnisse auf Parameterunsicherheit in diesen Parametern zurückzuführen sind.

Nach der Diskussion der bisherigen Ergebnisse in der Literatur konzentriert sich dieses Kapitel auf eine Reihe von methodologischen Entwicklungen. Zunächst wird eine Approximationsmethode, die von Aït-Sahalia (2002) entwickelt wurde, auf die Übergangsdichten des CKLS-Modells angewendet. Diese approximierte Übergangsdichte wird dann zur Schätzung von Maximum Likelihood Schätzern und bei der Simulation von a-posteriori Verteilungen verwendet. Bei der Simulation durch MCMC Algorithmen führe ich eine Neuparametrisierung des CKLS Modells ein, um die starke nicht-lineare Abhängigkeit zwischen den beiden Volatilitätsparametern zu reduzieren. Diese Neuparametrisierung verringert Schwierigkeiten, die sich normalerweise bei der Simulation der Parameter des CKLS Modells ergeben, erheblich und führt zu schnell konvergierenden Markovketten. Um die Aussagekraft der verschiedenen Modelle zu überprüfen,

berechne ich im Weiteren Bayes factors. Hierbei setze ich auf einer von Kass and Raftery (1995) beschriebenen Annäherungsmethode auf, und verwende zusätzlich eine Box-Cox Transformierung um die Annäherung zu verbessern. Dieses Verfahren wurde von Verdinelli and Wasserman (1995) beschrieben, aber hat meines Wissens noch keine Anwendung bei der Schätzung von Short-rate Modellen gefunden. In diesem Kapitel wird außerdem getestet, ob die Berücksichtigung von Schätzunsicherheiten die Vorhersagekraft des CKLS Models verbessern kann. Zu diesem Zweck wird ein Out-of-Sample Test verwendet, der zum Testen von Value-at-Risk Modellen entwickelt wurde.

Nach der Präsentation der Methodik wende ich diese auf die Schätzung der Modelle an. Die Parameter der Modelle werden auf Zeitreihen in neun verschiedenen Währungen angepasst, nämlich dem US-amerikanischen Dollar, dem Britischen Pfund, dem Euro, dem Schweizer Franken, dem Japanischen Yen, dem Australischen Dollar, der Deutschen Mark sowie dem Französischen Franc und der Italienischen Lira. Für diese Währungen werden zwei verschiedenen Zeitfenster von zehn und 25 Jahren sowie drei verschiedene Beobachtungshäufigkeiten verwendet (tägliche, wöchentliche und montaliche Beobachtungen).

Die Schätzergebnisse stimmen im Wesentlichen mit den Ergebnissen aus der Literatur überein. Es gibt nur schwache Indizien für die Existenz eines sogenannten *mean-reversion* Effekts. Dieser fällt jedoch etwas stärker als in den meisten bisherigen Studien aus, was auf die stetig fallenden Zinswerte während des Beobachtungszeitraums zurückzuführen ist. Die Schätzergebnisse für die Volatilitätsparameter schwanken zwischen den verschiedenen Zeitreihen, Beobachtungszeiträumen und sogar Beobachtungsdichten deutlich. Dies spricht nicht für die Konsistenz und Aussagekraft der untersuchten Modelle und stimmt weitestgehend mit den Aussagen der bisherigen Literatur überein. Insbesondere kann die Annahme, dass die in der Literatur auftretenden unterschiedlichen Schätzergebnisse auf Parameterunsicherheit zurückgeführt werden können, nicht belegt werden. Unter Anwendung der Bayesianischen Methoden kann zwar eine stärkere Unsicherheit festgestellt werden, als dies bei herkömmlichen statistischen Verfahren der Fall ist, allerdings reicht diese zusätzlichen Unsicherheit nicht zur Erklärung der unterschiedlichen Ergebnisse. Die Tests zur Vorhersagekraft des Modells ergeben, dass das Einbeziehen von Schätzunsicherheiten die Vorhersagekraft des Modells zwar leicht erhöht, diese insgesamt aber weiterhin unbefriedigend ist.

Kapitel 4 beschreibt die Erweiterung zweier Pricingverfahren, nämlich die Anwendung eines bivariaten Trinomialbaums auf das CEV-CKLS Modell sowie die Erweiterung des QUAD-Algorithmuses auf ein Black-Scholes-Vasicek Modell. Diese Erweiterungen sind aus der Anwendung in Kapitel 5 heraus motiviert, bei der schnelle Pricingverfahren zur wiederholten Bewertung von Optionen benötigt werden.

Zusammenfassung

Der erste Teil dieses Kapitels fasst zunächst die bestehende Literatur zu diesen beiden Bewertungsalgorithmen zusammen. Im zweiten Teil wird die Anwendung des Trinomialbaums auf das CEV-CKLS Modells beschrieben. Hierbei kommt es zu verschiedenen Schwierigkeiten, die überwunden werden müssen. Erstens können die in diesem Modell vorhandenen stochastischen Prozesse bei geeigneter Parameterwahl den Nullpunkt erreichen, diesen jedoch nicht unterschreiten. Daher benötigt man zur Konstruktion des Baumes ein geeignets Gitter, das das exakte Erreichen der Null zulässt, eine Unterschreitung dieses Wertes aber unmöglich macht. Zweitens sind beide Prozesse heteroskedastisch, d.h. ihre Varianz hängt von ihrem aktuellen Wert ab. Diese Eigenschaft erschwert die Konstruktion eines rekombinierbaren Baumes, da diese zunächst transformiert werden müssen, was bei der Gitterkonstruktion wiederum berücksichtigt werden muss. Schliesslich muss die Korrelation zwischen den beiden Prozessen berücksichtigt und bei der Wahl der Übergangswahrscheinlichkeiten richtig wiedergegeben werden.

Der dritte Teil des vierten Kapitels beschreibt die Erweiterung des QUAD-Algorithmuses auf ein Black-Scholes-Vasicek Modell mit stochastischen Zinsen. Zunächst wird der reguläre QUAD-Algorithmus, so wie er von Andricopoulos et al. (2003) präsentiert wurde, beschrieben. Der Algorithmus wird dann auf das Black-Scholes-Vasicek Modell erweitert, d.h. ein Modell in dem die Drift einer geometrischen Brown'schen Bewegung durch einen Vasicek-Zinsprozess dargestellt wird. Zur Erweiterung des Algorithmuses ist die korrekte Herleitung der bivariaten Übergangsdichte von entscheidender Bedeutung. Außerdem werden in diesem Teil des Kapitels einige Tricks erläutert, mit denen mit diesem Algorithmus Optionen bewertet werde können, die durch ihre Struktur (z.B. aufgrund von Dividendenzahlungen) eigentlich nicht auf dem Gritter darstellbar wären. Der Versicherungsvertrag aus Kapitel 5 ist ein Beispiel für einen solchen Kontrakt.

Der letzte Teil des vierten Kapitels wendet die beiden Bewertungsmethoden auf zwei verschiedene Optionen im Black-Scholes-Vasicek Modell an. Bei der ersten Option handelt es sich um eine europäische Call Option, bei der zweiten um eine sogenannte *Traffic light option*. Geschlossene Lösungen für beide Optionen wurden durch Rabinovitch (1989) und Løchte (2007) vorgestellt. In diesem Rahmen schneidet der zweidimensionale QUAD-Algorithmus deutlich besser ab als der bivariate Trinomialbaum, da er genauere Lösungen produziert und in deutich kürzerer Zeit läuft. Dies deutet darauf hin, dass der QUAD-Algorithmus unter diesen und ähnlichen Umständen zu bevorzugen ist. Allerdings hat der bivariate Trinomialbaum dennoch durchaus sinnvolle Anwendungen, z.B. lässt sich der QUAD-Algorithmus nicht in dem algemeineren CEV-CKLS Modell anwenden und die Geschwindigkeit und Genauigkeit könnte für andere Optionen, die häufigere Zeitschritte benötigen, durchaus vergleichbar mit der des QUAD-

Algorithmuses sein.

Kapitel 5 untersucht den Einfluss von Parameter- und Modellunsicherheit auf Risikomaße für implizite Optionen, die z.B. in partizipierenden Lebensversicherungsverträgen auftreten. Es ist daher die zweite Anwendung Bayesianischer Methoden zur Abschätzung des Effekts von Schätzunsicherheiten auf Risikomaße. Dieses Kapitel ist durch die Beobachtung von Zaglauer and Bauer (2008) motiviert, die beschreiben, dass die Bewertung des in der Studie untersuchten Versicherungskontrakts sensibel auf stochastischen Zinsen reagiert. Da die Parameter in Zinsprozessen in der Regel schwer zu schätzen sind, stellt sich die Frage, wie sehr die Bewertung und das Risiko dieses Kontraktes durch Schätzunsicherheit beeinflusst werden. Die Studie von Hardy (2002) untersucht eine ähnliche Problemstellung, wodurch sich zeigt, dass in der Versicherungsliteratur auch ein gewisses Interesse an dieser Fragestellung besteht.

Der erste Teil von Kapitel 5 motiviert das Problem und fasst kurz die relevante Literatur zu dem Thema zusammen. Im nächsten Teil wird der Versicherungsvertrag präsentiert und dargestellt, wie dieser in fünf implizite Optionen aufgeteilt werden kann. Der dritte Teil stellt das Bewertungsmodell dar, wobei es sich hier um ein CEV-CKLS Modell, wie es in Kapitel 4 untersucht wurde, handelt. Anschliessend beschreibe ich, wie Preise für diesen Kontrakt numerisch über den QUAD-Algorithmus bestimmt werden können. Hierbei handelt es sich um eine Erweiterung des in Kapitel 4 vorgestellten Algorithmuses, da zur Bestimmung der Preise eine zusätzliche Zustandsvariable berücksichtigt werden muss. Die Beschreibung dieser weiteren Zustandsvariablen im Modell erfordert einige Besonderheiten, die hier genauer dargelegt werden. Da dieser Algorithmus eine deutliche Performanceverbesserung gegenüber anderen Methoden darstellt, ist er eine interssante Alternative um ähnliche Kontrakte zu bewerten. Der folgende Abschnitt stellt dar, wie die Paramter für das Modell mit und ohne Berücksichtigung von Schätzunsicherheiten bestimmt werden. Im Weiteren werden die zwei Methoden zum Risikomanagement vorgestellt, anhand derer die Auswirkungen von Parameterunsicherheiten untersucht werden sollen. Bei der ersten Methode handelt es sich um eine Art *Buy-and-Hold*-Strategie, bei der der Wert des Kontraktes in ein Portfolio investiert wird aus dem eventuelle Leistungen bis zum Ende der Laufzeit des Kontraktes bestritten werden. Die zweite Strategie hedgt die impliziten Optionen des Kontraktes auf monatlicher Basis. Die Resultate aus diesem Kapitel zeigen, dass beide Risikomanagementansätze stark unter der Berücksichtigung von Parameter und Modellunsicherheit leiden. Die Verlustverteilungen unter beiden Ansätzen verbreiten sich deutlich wenn Parameter und Modellunsicherheit berücksichtigt werden.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr. Gunter Löffler for his support, his guidance, and his patience. Without him, this thesis would not have been possible. I would also like to thank my second supervisor Prof. Dr. Rüdiger Kiesel for his supportive and helpful advice.

Furthermore, I would like to thank my former colleagues from the Graduate Training Group 1100 and from the Institute of Finance, without whom this undertaking would have been a lot less fun and would have had much fewer memorable moments. I thank Montse Carty for proofreading and correcting this work.

I would like to thank my family for their continued love and support which helped me tremendously in difficult moments. Finally, I owe the biggest debt to my girlfriend Mònica, who has loved and supported me and has never lost patience through nights and weekends as the termination date stubbornly stayed a few months in the future.

This page contained a curriculum vitae which has been removed.

Declaration

I hereby declare that this thesis was performed and written on my own and that references and resources used within this work have been explicitly indicated.

I am aware that making a false declaration may have serious consequences.

Ulm, May 27, 2013 _____

(Signature)