Improving risk neutral valuation techniques with applications in insurance

Singor, Stefan

DOI
10.4233/uuid:c70c38c8-af5f-4454-a612-38b92c773ea2

Publication date
2017

Document Version
Final published version

Citation (APA)
Singor, S. (2017). *Improving risk neutral valuation techniques with applications in insurance*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:c70c38c8-af5f-4454-a612-38b92c773ea2

IMPROVING RISK NEUTRAL VALUATION TECHNIQUES WITH APPLICATIONS IN INSURANCE

IMPROVING RISK NEUTRAL VALUATION TECHNIQUES WITH APPLICATIONS IN INSURANCE

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. ir. K. C. A. M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 20 november 2017 om 10:00 uur

door

Stefan Nico SINGOR

wiskundig ingenieur
geboren te Oud-Beijerland, Nederland.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. C. W. Oosterlee

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. C. W. Oosterlee	Technische Universiteit Delft, promotor
Prof. dr. ir. C. Vuik	Technische Universiteit Delft
Prof. dr. ir. H. X. Lin	Technische Universiteit Delft
Prof. dr. G. Deelstra	Université libre de Bruxelles, België
Prof. dr. A. de Schepper	Universiteit van Antwerpen, België
Prof. dr. M. Francke	Universiteit van Amsterdam
Dr. H. Steehouwer	Erasmus Universiteit Rotterdam / Ortec Finance
Prof. dr. ir. A. W. Heemink	Technische Universiteit Delft, reservelid

Improving risk neutral valuation techniques with applications in insurance
Dissertation at Delft University of Technology

ISBN 978-90-9030645-2

An electronic version of this dissertation is available at
http://repository.tudelft.nl/

Cover design by Ton Meulmeester, Ortec Finance

Contents

Summary

In times of market turmoil volatility increases and stock values and interest rates decrease, so that the risks in the balance sheets of insurance companies increase. An important part of these risks is due to the guarantees that are embedded in insurance policies. Life insurers sell products like unit-linked, profit sharing and variable annuity products. These contracts contain guarantees to the policyholders. Such contracts embedded in the insurers' liabilities are called embedded options.

The value and cash flow of these contracts are respectively relevant for the balance sheet and the profit and loss account. Typically in periods of volatile markets, the value of these embedded options increase, so that the insurance company must hold a larger liability value on the balance sheet in order to be able to pay out future cash flows. The valuation of these embedded options in insurance liabilities is therefore important to insurers for risk management applications. In this thesis we consider various topics regarding the valuation of these embedded options.

We consider the risk neutral valuation methodology to compute option values. The underlying economic variables, such as stocks, interest rates and inflation, are in this case modeled by a set of stochastic differential equations. The correlation structure among the economic variables makes the risk neutral model involved. Such models are called hybrid models. The derivation of semi-analytic formulas for such models for valuation and calibration is hard if not impossible. Alternatively, one has to resort to the numerically expensive Monte Carlo simulations for valuation. The derivation of semi-analytic formulas for valuation using hybrid risk neutral models and the application of Monte Carlo simulations for valuation are the main topics of this thesis.

In Chapters 2 and 3 we consider the derivation of semi-analytic valuation formulas. Such formulas lead to fast computation speed, which is desired from practice. The point of departure in this thesis is a Heston type inflation model in combination with a Hull-White model for nominal and real interest rates, in which all the correlations can be non-zero. Inflation is an important risk factor for pension funds and insurance companies. Due to the presence of the Heston dynamics our derived inflation model is able to capture the implied volatility skew/smile, which is present in the inflation option market data. We derive an efficient semi-closed form pricing formula to approximate the value of index- and year-on-year inflation options. The derived pricing formulas allow for an efficient calibration of the inflation model. We illustrate our approach using a real-life pension fund example, where the Heston Hull-White inflation model is used to determine the value of conditional future indexations.

In Chapter 3, we propose a valuation model for real estate derivatives. We aim to value derivatives that are coupled to real estate indices with a degree of autocorrelation. Since the underlying asset cannot be traded in a frictionless market, it is impossible to use classic pricing formulas for derivatives, because these formulas fully rely on no-arbitrage assumptions. We therefore first model the underlying efficient market price of real estate and then construct the observed index value. Using this real estate model, we

derive closed-form pricing solutions for forwards, swaps and European put and call options. We demonstrate the application of the model by valuing a put option on a house price index. Autocorrelation in the index returns appears to have a large impact on the option values.

To forecast the balance sheet and solvency capital requirements, the technique of real world scenarios is applied. A real world scenario is a possible evolution of the future state of economic variables, such as interest rates, inflation and equity returns, that is consistent with a clear set of assumptions. At each time step of a real world scenario one has to value the embedded option. The number of valuations increases enormously in this case, so that fast valuation methods are required.

Semi-analytic formulas are ideally used for valuation, because they are fast. These formulas do however not exist for more advanced risk neutral models, so that one has to apply numerical techniques for the purpose of valuation. Valuation by means of Monte Carlo simulations is the preferred method, but comes at a cost of large computation time. The application of Monte Carlo simulation may lead to nested simulations. The real world scenarios are called the outer scenarios and the risk neutral scenarios the inner scenarios. In Chapters 4, 5 and 6 we focus on accelerating and modeling of these nested Monte Carlo simulations.

In Chapter 4, we introduce the Option Interpolation Model for approximating embedded option values. The proposed method is based on interpolation with radial basis functions, which can be adopted to interpolate scattered data. To reduce computation time we present an inversion method to determine the interpolation function weights. The robustness, accuracy and efficiency of the Option Interpolation Model are analyzed by means of a number of numerical experiments. We show that the proposed approximation method results in highly accurate estimates.

In a nested simulation it is important that the risk neutral model at hand is consistent with the generated implied volatilities at each time point. In Chapter 5, we propose a modeling framework for risk neutral stochastic processes nested in a real world stochastic process. We make use of the class of State Space Hidden Markov models for modeling the joint behavior of the parameters of a risk neutral model and the dynamics of option market instruments. This modeling concept enables us to perform non-linear estimation, forecasting and robust calibration. The proposed method is applied to the Heston model for which we find highly satisfactory results. We use the estimated Heston model to compute the required capital of an insurance company under Solvency II.

In Chapter 6, we present a high performance computing framework to improve the performance of nested simulations. We aim to take full advantage of the parallelism of Graphical Processing Units (GPUs). The parallel structure of modern GPUs makes them more efficient than general-purpose Central Processing Units (CPUs) for algorithms where the processing of large blocks of data can be done in parallel. We manage to reduce the computation time of a nested simulation application from several hours to tens of minutes.

We note that all the work presented in this thesis is based on published or submitted papers written during the PhD research.

Samenvatting

In tijden van crisis neemt de volatiliteit in de financiële markten toe en dalen de aandelen en rentes, zodat de risico's toenemen op de balans van een verzekeraar. Een belangrijk deel van deze risico's is te wijten aan de garanties die ingebed zijn in verzekeringspolissen. Levensverzekeraars verkopen unit-linked, winstdeling en variable annuity producten. Deze contracten bevatten garanties aan de polishouder en worden ook wel ingebedde opties genoemd.

De waarde en de kasstromen van deze opties zijn respectievelijk belangrijk voor de balans en de winst- en verliesrekening. Typisch in tijden van hoge volatiliteit stijgen de waarden van deze ingebedde opties. Dit betekent dat de verzekeraar een hogere verplichting moet nemen op de balans om toekomstige kasstromen te kunnen uitbetalen. De waardering van deze ingebedde opties in verzekeraarsverplichtingen is daarom belangrijk voor verzekeraars voor risico managementtoepassingen. In dit proefschrift behandelen we verschillende onderwerpen met betrekking tot de waardering van ingebedde opties.

We gebruiken de methodologie van risico neutralewaardering voor het berekenen van optiewaarden. De onderliggende economische variabelen zoals aandelenrendementen, rentes en inflaties worden in dit geval gemodelleerd door een set van stochastische differentiaalvergelijkingen. De correlatiestructuur tussen de economische variabelen maakt het risico neutralemodel moeilijk oplosbaar. Dergelijke modellen worden hybride modellen genoemd. Het afleiden van semi-analytische formules voor zulke modellen voor waardering en kalibratie is moeilijk, zo niet onmogelijk. Als alternatief wordt de numeriek duurdere Monte Carlosimulatie gebruikt voor waardering. Het berekenen van optieprijzen op basis van semi-analytische formules en op basis van Monte Carlosimulaties staat centraal in dit proefschrift.

In de Hoofdstukken 2 en 3 bespreken we de afleiding van semi-analytische waarderingsformules. Het gebruik van deze formules resulteert in snelle rekentijden, wat zeer wenselijk is in de praktijk. Het startpunt in dit proefschrift is een Heston-type model voor inflatie gecombineerd met een Hull-White model voor de modellering van nominale en reële rentes, waarbij de correlaties ongelijk aan nul kunnen zijn. Inflatie is een belangrijke risicovariabele voor pensioenfondsen en verzekeraars. Vanwege de Heston dynamica in het inflatiemodel is het mogelijk om belangrijke eigenschappen in de optiemarkt data (smiles/skews) te modelleren. We bepalen een efficiënte half-gesloten waarderingsformule om de waarde te benaderen van index- en jaar-op-jaar inflatieopties. De waarderingsformules kunnen gebruikt worden voor efficiënte kalibratie van het inflatiemodel. We gebruiken een pensioenenvoorbeeld om de relevantie van het inflatiemodel te laten zien. We gebruiken het Heston Hull-White inflatiemodel voor de berekening van conditionele toekomstige indexaties.

In Hoofdstuk 3 presenteren we een waarderingsmodel voor vastgoedopties. Het doel is om opties te waarderen met als onderliggende een vastgoedindex met een hoge mate van autocorrelatie. Omdat de onderliggende index niet verhandeld kan worden, is het

niet mogelijk om klassieke waarderingsformules te gebruiken (omdat deze formules gebaseerd zijn op de aanname dat arbitrage niet voorkomt). Daarom modelleren we eerst de onderliggende efficiënte marktprijs van vastgoed en vervolgens de geobserveerde indexwaarde. Als toepassing gebruiken we het vastgoedmodel voor de waardering van een put-optie met de huisprijsindex als onderliggende. We laten zien dat autocorrelatie in de index een groot effect kan hebben op de optiewaarde.

Om de balans en kapitaalvereisten te simuleren, wordt de techniek op reële wereldscenario's toegepast. Een reële wereldscenario is een mogelijke toekomstige ontwikkeling van economische variabelen, zoals rentes, inflaties en aandelenrendementen, die consistent is met een set aannames in de reële wereld. Op elk toekomstig tijdstip van een scenario moet een waardering van de ingebedde opties gedaan worden. Het aantal waarderingen neemt daarom enorm toe zodat snelle waarderingsmethoden bijzonder gewenst zijn.

Semi-analytische formules zijn ideaal voor zulke waarderingen, omdat ze snel zijn. Deze formules bestaan alleen niet voor geavanceerde risico neutralemodellen waardoor men gebruik zal moeten maken van Monte Carlosimulaties. Waarderingen op basis van Monte Carlosimulaties zijn voldoende nauwkeurig, maar de bijbehorende rekentijden zijn hoog. De toepassing van Monte Carlosimulaties leidt tot zogeheten geneste Monte Carlosimulaties. De reële wereldscenario's worden de buitenscenario's genoemd en de risico neutralscenario's de binnenscenario's. In de Hoofdstukken 4, 5 en 6 ligt de focus op het versnellen en verbeterd modelleren van deze geneste Monte Carlosimulaties.

In Hoofdstuk 4 introduceren we het Optie Interpolatie Model voor de benadering van ingebedde optiewaarden. De methode is gebaseerd op interpolatie op basis van radiale basisfuncties, die ongestructureerde data kunnen interpoleren. Om rekentijden te verbeteren, stellen we een inversie rekenmethode voor om de interpolatie gewichten te bepalen. De robuustheid, nauwkeurigheid en efficiëntie van het Optie Interpolatie Model worden geanalyseerd door middel van een aantal numerieke experimenten. We laten zien dat de voorgestelde methode resulteert in nauwkeurige schattingen.

In een geneste simulatie is het belangrijk dat een risico neutraalmodel consistent is met de gegenereerde marktdata op elk tijdstip. In Hoofdstuk 5 stellen we een modelleringsraamwerk voor om op een consistente manier risico neutrale processen te modelleren in een reële wereld scenarioset. We maken gebruik van State Space Hidden Markovmodellen voor de modellering van de gecombineerde dynamica van de modelparameters en de optie marktdata. De methode stelt ons in staat nietlineaire en robuuste kalibratie door te voeren en model parameters en optieprijzen nauwkeurig te voorspellen. We passen de methode toe op het Hestonmodel, waarvoor we goede resultaten behalen. We gebruiken het geschatte Hestonmodel voor de berekening van het vereiste kapitaal van een verzekeraar onder Solvency II regulering.

In Hoofdstuk 6 presenteren we een wetenschappelijk rekenraamwerk voor voor het versnellen van de geneste simulaties. Het doel is om volledig gebruik te maken van de parallelle structuur op grafische kaarten (GPUs). De parallelle structuur van moderne GPUs maakt ze efficiënter dan Central Processing Units (CPUs) voor algoritmes waarvoor grote blokken data parallel tegelijk verwerkt kunnen worden. We verlagen de rekentijd van een geneste simulatietoepassing van enkele uren naar tientallen minuten.

We merken op dat het werk dat wordt gepresenteerd in dit proeschrift is gebaseerd

op gepubliseerde of reeds ingediende artikelen die geschreven zijn tijdens het PhD onderzoek.

CHAPTER 1

Introduction

1.1. INSURANCE

Insurance is about the protection against future losses and the agreements of the insurance are represented by a policy. In case of a loss, the policyholder receives protection from an insurance company. Individuals buy insurance policies based on their assessment of a possible loss in the future, and insurers offer them cover based on their assessment of the cost of covering any claims. So, the insurance industry works on the principle of risk. There are nowadays many types of insurance policies available. The most common types of personal insurance policies are car, health, homeowners and life insurance policies. See [41] for more information.

The insurance market is large with a total of €1,200 billion premiums turn over in 2015. Europe, United States and Asia are approximately equally large, each contributing about 30%. Compared to the non-life and health insurance markets, the life insurance market is largest with €730 billion premiums (60% market share).

Forms of insurance were already practiced by Chinese and Babylonian traders in the third and second millennia before Christ [135]. Chinese merchants distributed their wares across many vessels to limit the loss due to any single vessel's capsizing. If a merchant received a loan to fund his shipment, he would pay the lender an additional sum in exchange for the lender's guarantee to cancel the loan should the shipment be stolen, or lost at sea.

The first known insurance contract dates from Genoa in 1347, and property insurance as we know it today can for example be traced back to the Great Fire of London, which in 1666 destroyed more than 13,000 houses. The devastating effects of the fire converted the development of insurance from a matter of convenience into one of urgency.

Insurance became more sophisticated during the Enlightenment period (17th and 18th century) as the industrial changes called for insurance solutions. It provided the basis for accepting actuarial sciences as a rational means to conduct better business. Towards the end of the 18th century the first modern and global insurance company, The Phoenix, was founded in London.

As a big step ahead in time, the insurance market changed rapidly in the 1990s when the internet emerged. This, in response to consumer behavior, improved technology and the increase of available data, which is also known as big data. The internet has changed the way customers buy insurance and maintain contact with the insurance providers. Big data refers to the massively increasing volume, velocity and granularity of data sets that are being accessed. Big data applications gained interest in the last 5 years.

The ability to compile and analyze those granular data sets is now transforming the

way insurers see large pools of consumers and how they price risks. This has implications for the cost and availability of insurance for all consumers. Insurance provides financial protection to the community by pooling resources to manage risks. That approach is now being significantly refined by insurers' increased capacity to examine risks in greater detail through the analysis of large volumes of granular data. The outcome is that increased individual risk pricing will make premiums tailored in the way that they will be more reflective of that risk.

Given the rapid changes in the insurance market, the study of risk management becomes more essential.

1.2. EMBEDDED OPTIONS

Life insurance companies provide contracts that contain guarantees. They for example sell unit-linked (UL), profit sharing and variable annuity products. Such contracts embedded in the insurers' liabilities are called embedded options. Embedded options are contracts (rights) in life insurance policies that may provide a profit to policyholders, but never a loss. The value of these contracts depends on one or more underlying variables such as interest rates, inflation and equity. This makes the future cash flows uncertain.

A UL life insurance contract is a savings policy that pays at maturity date. The initial lump sum and premiums paid during the contract are invested according to a certain asset allocation with an embedded investment guarantee. This means that when the invested capital at maturity date is lower than the guaranteed capital, the insurer has to pay the difference to the policyholder. Otherwise the policyholder receives the invested capital. For the insurer it's important to accurately compute the current liability value in order to be able to pay out the (uncertain) cash-flow after 30 years. Answering this fundamental question is the main focus of this thesis.

In the life insurance market, 20% of the €730 billion premiums is part of UL type that often consist of guarantees. The modeling of these option structures in insurance liabilities is therefore essential to insurers for risk management applications.

The value and cash flow of these contracts are respectively relevant for the balance sheet and the profit and loss account. Due to the hybrid nature of these embedded options, different movements in the variables can influence the embedded option values. Especially in financial crisis the valuation of these embedded contracts is important. In times of crisis, the volatility increases and stock values go down and it's likely that the invested capital is lower than the guaranteed capital leading to high liability values for UL products on the balance sheet of a life insurer. Assigning a realistic value to these contracts is therefore important.

1.3. EMBEDDED OPTIONS IN PENSIONS

Pension is about saving for an income after retirement. Defined Benefit (DB) and Defined Contribution (DC) are the main pension schemes. In a DB plan an income in retirement is determined based on the employee's earnings history, tenure of service and age. In this case the pension funds bear the investment risk and can benefit from surpluses. In a DC plan, each member owns an investment account to build up an income after retirement. The contributions are fixed, but the future benefits fluctuate on the

basis of investment earnings. So, the pension fund members bear the investment risk (instead of the pension fund itself).

DB pension funds and life insurers are both providers of old-age income provisions. The main difference lies in the mitigation of risk. The risks are borne by either beneficiaries and the sponsor (in a DB plan) or the beneficiaries (in a DC plan). In the case of a life insurer it is borne by external shareholders, i.e. there is no risk for the policyholders (except in the case of bankruptcy).

Pension funds are also different from insurance companies because they have the ability to use risk-mitigating instruments such as steering mechanisms (e.g. higher contributions, additional sponsor support) and adjustment mechanisms (e.g. conditional indexation, cutting benefits) for improving the financial position of the fund.

On the so-called traditional balance sheet of a DB pension fund, the asset and liability values of the fund are stated to reflect the financial position. Steering and adjustment instruments are however not taken into account in the traditional balance sheet.

To deliver a more complete picture of the financial position of a pension fund and to be able to compare pension funds, EIOPA proposed the holistic balance sheet (HBS) in 2011 to assess pension funds [30]. The HBS is in essence similar to a traditional balance sheet as it shows all assets and liabilities in a single overview. However, in the HBS the value of additional steering mechanisms is also taken into account. The HBS and the valuation of steering mechanisms are new concepts and are still in the early stages of development.

For most steering mechanisms, the conditional pay-off depends on some underlying decision variables. The HBS approach does include the steering and adjustment instruments of a pension fund and values them as embedded options on the HBS. Hence, assigning accurate option values to these steering and adjustment instruments is important for risk management applications.

A transition from DB to DC has been underway for a couple of decades. The main reasons for this transition are (amongst others) the high employer costs, volatility and unpredictability of contributions and possibilities for tailoring (choose your own risk profile). Under a DC scheme there would be no need for a HBS framework because DC is individually based instead of collectively under DB. Given the recent increase in attention for DC schemes, it is uncertain how the development of HBS will evolve.

1.4. STRATEGIC RISK MANAGEMENT OF EUROPEAN INSURANCE COMPANIES

Risk management consists of identifying and analyzing loss exposures and to minimize the financial impact of the risks the insurance company are exposed to. This includes the analysis of the effects of financial risks to the organization. For example, changes in interest rates are an important financial risk. For the protection against these risks, hedging tools are used to manage the exposure to interest rate volatility. A hedge is an investment to reduce the risk of price movements in an asset. Normally, a hedge consists of taking an offsetting position in a related security, such as a futures contract. See [73] for more information.

Risk management computations in the insurance industry improved rapidly after the

World War II. First, because of the increased use of derivatives as risk management instruments in the 1970s and 1980s. Derivatives are contracts that protect the holder from certain risks and are popular hedging tools. Second, insurers began to consider portfolio and Asset Liability Management (ALM) [140] in the 1980s and 1990s.

The objective of ALM is the management of the assets, liabilities and solvency required capitals on the future balance sheet and to specify an investment strategy that given the specified risk limits maximizes the ambition of an insurance company. This can for example be to achieve a maximum dividend to shareholders or profit sharing to policyholders. The objective of strategic risk management is to manage these risks and returns in a changing environment.

International risk regulation began in the 1980s, and financial firms developed internal risk management models and capital calculation formulas to hedge against unanticipated risks and reduce regulatory capital. Governance of risk management became essential and the chief risk officer positions were created. See [39] for more information.

Solvency II is a European wide solvency regulation that came into effect on January 1, 2016. It's designed to ensure consistency of supervision and reserving across European insurers. Solvency II replaces the Solvency I framework that was introduced in 1973 and was in need of replacement due to all the changes over the years. Solvency II consists of three pillars: quantitative requirements (pillar 1), governance & supervision (pillar 2) and disclosure & transparency (pillar 3). In pillar 1 the solvency capital requirement is defined in such a way that insurers can meet their obligations to policyholders and beneficiaries over the following 12 months with a 99.5% probability. This should limit the probability of falling into financial ruin less than once in 200 cases.

Under Solvency II an insurer is required to value the liabilities at market value (instead of book value under Solvency I). Furthermore, at the latest January 1, 2021, all insurance companies across the world must report their income statements in terms of market value instead of book value, based on a new insurance industry accounting principle called the International Financial Reporting Standards 17 (IFRS 17). Typically for life insurers this results in valuation challenges as they sell products that contain guarantees. Due to the uncertain future cash flows it is a challenge to compute the market values of those products.

Most insurance companies are nowadays able to generate the necessary Solvency II and other reports that show the current balance sheet and financial position. Generating consistent forward-looking projections of the balance sheet and solvency capital requirements, which is necessary in ALM analysis to support strategic decision making in general, is still a challenge.

To find the desired investment strategy, ALM studies utilize the technique of so-called real world scenario analysis. Scenarios are future trajectories of financial variables that managers must take into account in their policy determination and evaluation. Examples of financial variables are inflation, interest rates, currencies, the returns of the various investment categories, and the development of instruments deduced from these, such as financial derivatives. ALM studies calculate, with the use of a corporate model of the insurance company, for every year and each scenario, what the consequences are of the strategy to the ambition of the insurance company.

Regulations, governance rules and risk management methods failed to prevent the

worldwide financial crisis that began in 2007. Improvements of regulation and risk management computations hence remains essential despite the progress seen in the last 50 years.

The worldwide financial crisis of 2007-2009 still has an enormous impact on the worldwide economy and has led to financial problems for insurance companies. In order to stimulate the economy most central banks intervened in the market by using the so-called Quantitative Easing (QE) monetary policy. QE is implemented by buying amounts of fixed income products from commercial banks and other financial institutions. The goal is to increase private-sector spending and return inflation to its target. Consequently, QE raises the prices of those fixed income products and lowers their yield, so that interest rates are artificially kept at a low level.

This low interest rate environment and the desire to keep profitability at acceptable levels force insurance companies to a search for yield. First of all, there is more investment focus on more risky asset classes, which in general provide a higher return. Insurers will have to take risk. Secondly, there is a liability focus on the profit margins of insurance products, which have deteriorated. The search for yield is therefore also concerned with the product mix an insurer offers.

The Solvency II framework puts high capital charges on most of the higher yield assets, because of the higher risks. Given the low yield environment and the high capital charges on most of the higher yield assets, the search for yield becomes more complex.

Asset allocation decisions directly impact the Solvency II required capital, but they also impact the available capital. That is, risky assets that provide for higher expected returns, would increase the available capital but would also increase the required capital. It is hence a challenge to manage the available and required capital and the interactions between these two in order to find the optimal asset mix.

The challenge becomes even bigger when the insurer sells products that contain guarantees. For such products it is difficult to incorporate the products risk profile in a consistent way. This will be a research topic in this thesis.

1.5. VALUATION OF EMBEDDED OPTIONS

Research in this thesis is about the valuation of embedded options in ALM for insurance companies. Computing accurate option values is important to insurance companies in order to be able to pay out future cash flows and perform accurate risk management. The computation of the values of embedded options (and their sensitivities) is a mandatory part of Solvency II for European insurance companies and becomes mandatory under IFRS 17 for insurance companies. An insurance company deals with the computation of present and future option values.

Present option values and the corresponding risks are computed at a certain (historical) point in time for reporting and internal steering. Besides the present value, additional valuations are performed to quantify risks with respect to the important risk drivers, such as interest rate and implied volatility.

Insurers also wish to compute accurate option values at a certain time point in the future, that are relevant for ex-ante (i.e. forward looking) risk management applications such as ALM. In such applications, one starts with the generation of real world projections of all relevant economic variables. At each time step of a scenario, one is interested

in computing all balance sheet items, so that the relevant risks can be assessed. Hence, at each time step of a scenario one needs to evaluate option values, because they are part of the balance sheet. The modeling of these option structures in insurance liabilities is essential to insurers for ex-ante risk management applications, as they become dominant on the balance sheet and the profit and loss account in times of stress (crisis).

Present prices of assets are obtained by the computation of the expected values of discounted future cash flows. The prices hence depend on the asset's risk. Investors for example demand more profit when they are exposed to more uncertainty. For consistent valuation, the calculated expected values need to be adjusted for an investor's risk preference. Unfortunately, discount rates vary between investors and the risk preference of an individual is difficult to quantify.

According to the fundamental theorem of asset pricing, there is an alternative way to price assets. Instead of first taking the (real world) expectation and then adjusting for an investor's risk preference, one can adjust the probabilities of future outcomes such that they incorporate all investors' risk premia, and then take the expectation under this new probability distribution, the risk neutral measure. Once the risk neutral probabilities are found, every asset can be priced by simply taking the present value of its expected payoff.

The risk neutral measure is different from the real world measure, because in the real world investors demand risk premia, whereas in a risk neutral world all individuals are indifferent to risk and expect to earn on all securities a return equal to the (instantaneous) risk-free rate and thus do not incorporate any such premia. The method of risk neutral pricing is considered a useful computational technique for valuing financial derivatives.

Since the risk neutral pricing methodology is widely accepted, and alternatives hardly exist, this risk neutral pricing framework has become the pricing standard for embedded options in the insurance industry. The valuation of embedded options consists of three steps: the choice of the risk neutral model and the calibration and valuation steps.

Modeling derivative products in finance often starts with the specification of a system of stochastic differential equations (SDEs). Such a SDE system consists of economic state variables like stock prices, inflation, nominal and real interest rates and volatility. By imposing a correlation structure (between the Brownian motions) on this system of SDEs one can use them for pricing exotic derivatives.

Calibration is the estimation of the unknown model parameters in a mathematical model. Once the risk neutral model is chosen, the model parameters are determined in such a way that the model replicates market prices as accurately as possible. Calibration cannot be done in a closed form and numerical optimization routines are used to solve the calibration problem. That is, the sum of squared errors (or another error measure) is used to measure the error between market and model prices. The calibrated model is then used to value the embedded option. Analytic formulas are highly desired for computing the option values because they are fast to evaluate.

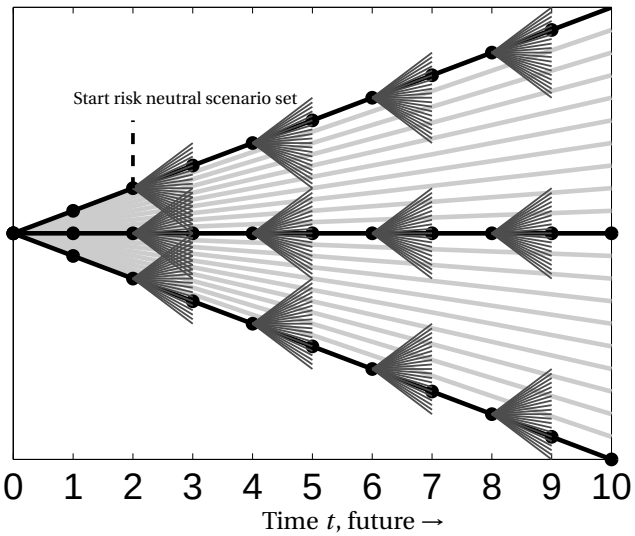
The most well-known example is the Black-Scholes formula for put/call equity index options. The Black-Scholes model is one of the most important concepts in modern financial theory, which was developed in 1973 and is still widely used. The authors received the 1997 Nobel Memorial Prize in Economic Sciences for their work.

Due to the complexity of embedded contracts, analytic formulas often don't exist.

Therefore, numerical Monte Carlo simulations [64] need to be employed to approximate the option values. Monte Carlo methods (or Monte Carlo experiments) form a broad class of computational algorithms that rely on random sampling to obtain numerical results. They are often used for problems in physics and mathematics and are useful when it is difficult or impossible to use other approaches.

Computing option values at present time $t = 0$ is rather straightforward. We can obtain option market data, calibrate the model parameters and compute the embedded option values. However, the calibration of the risk neutral model at a future time step and the computation of future option values is more involved. In each real world scenario the market option prices need to be simulated, which should be used for calibration. This means that in each scenario a numerical optimization routine should be performed in order to obtain the model parameters. Furthermore, because option values are ideally computed by using risk neutral Monte Carlo simulations, in each real world scenario one should generate a set of risk neutral Monte Carlo simulations. The combination of these two simulations is called a nested simulation. The nested simulation framework is illustrated in Figure 1.1. The real world scenarios are called the outer scenarios and the risk neutral scenario the inner scenarios.

Figure 1.1: The dashed lines refer to risk neutral scenarios and the straight lines refer to real world scenarios. At each time point, a risk neutral valuation should be performed. So, for 10.000 real world scenarios and 5 time steps per scenario, $(50.000 + 1)$ risk neutral valuations (including $t = 0$) should be performed.



1.6. NUMERICAL AND MODELING CHALLENGES

Given the importance of computing present and future option values in risk management applications, there are a number of numerical and modeling challenges an insurance company faces. In this thesis we focus on two challenges which we discuss below.

Extending the Black-Scholes model In order to compute accurate option values, risk neutral models are used that can model the relevant stylized facts in the option market data. The Black-Scholes model for example assumes that the log of the returns follows a normal distribution, but when markets are in stress, these assumptions may be violated. The Black-Scholes model is furthermore not able to reproduce the markets implied volatility smile/skew. The volatility value that produces a theoretical value which is exactly equal to the market price is called the implied volatility. The implied volatility smile/skew gives rise to an implied distribution of the underlying asset which has fatter tails than a log-normal distribution.

Model extensions are proposed based on the analysis of financial data. It's observed that for example interest rates and volatility are typically not deterministic but instead follow a stochastic process. Therefore, stochastic models have also been proposed for interest rates and volatility. A popular extension of the Black-Scholes model is the use of stochastic volatility or local volatility models for equity.

Although these risk neutral models can be easily defined, real use of these models is only guaranteed when they provide a satisfactory fit with the option market data.

The nested simulation problem Most nested simulation applications found in financial applications can run for several days on modern computers, an obvious bottleneck to their viability. Furthermore, these long-running simulations discourage any research on new models and new methodologies.

There are two procedures that make nested simulations expensive. First, Monte Carlo simulations [64] are used to compute the risk neutral option values. Applying analytic formulas is advantageous regarding the computation time, but restrictive when expanding the model set-up to hybrid models [65], that are generally required to obtain a high quality-of-fit with the option market data. In other words, analytic formulas can only be derived for specific cases, such as the Black-Scholes and Heston models for call/put options, but are often not available for hybrid models such as the Heston Hull-White model.

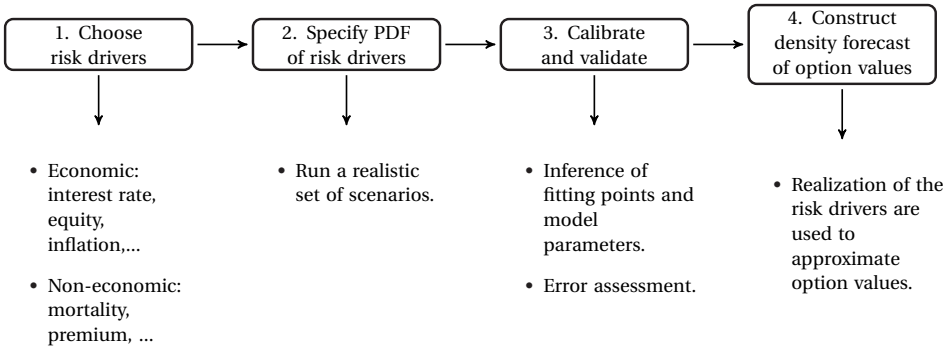
Second, the calibration of the model parameters at each time and in each scenario can not be done in a closed form. Numerical optimization techniques that can be used for calibration are expensive and even more time consuming in the case of using Monte Carlo simulations for valuations.

Many valuation techniques are available in the academic literature to accelerate nested simulations. These techniques have their pros and cons. The trade-off for choosing a valuation technique is between accuracy, monotonicity of solutions, computation time and ease of implementation. The trade-off generally depends on the application at hand. Ease of implementation is important for both applications. Popular valuation techniques are: the nested simulation, analytic formulas, curve fitting, regression methods and interpolation methods. In [6] an overview is given of existing valuation methods.

Future embedded option values are approximated in four steps, see Figure 1.2. First of all, the option's pay-off function to determine the relevant risk drivers is analyzed. Risk drivers can either be economic or non-economic variables. Examples of economic risk drivers are interest rates, inflation, equity and implied volatilities. Examples of non-economic risk drivers are costs, premiums, benefits and mortality. Secondly, the em-

pirical probability distribution function (PDF) of the risk drivers is identified. This PDF is used to generate relevant interpolation points for the subsequent option valuation. Thirdly, the interpolant is calibrated and validated for error assessment. Fourthly, the method is used for density forecasting in an ex-ante risk management application. This means that real world realizations of the underlying risk drivers are not used in the actual (Monte Carlo) valuation function, but the method is used to compute risk neutral option values.

Figure 1.2: Flowchart of approximating embedded options.



A way of accelerating nested simulations is by using High Performance Computing (HPC). This area of research is developing rapidly due to improved hardware. HPC becomes more and more essential in the insurance industry. HPC is a general term for techniques to make applications run faster than on regular desktops or workstations. Examples of HPC are: a grid of regular desktops, a cluster of compute nodes or offloading parts of the calculations to other available hardware or even a combination of all. The scope of HPC is still growing and gets more accessible for small companies and even individuals. For example, Amazon offers an on-demand service for using different HPC instances like GPU clusters, I/O clusters or storage optimized clusters. The scope of applications using such techniques is still growing.

Applying HPC solutions is not easy. Since preparing applications for HPC can be hard, the concepts of HPC are mainly applied by early adaptors in the industry. Companies have to consider the additional investment in developing and maintaining HPC applications against the performance gain.

1.7. OUTLINE OF THIS THESIS

In this thesis, we improve risk neutral valuation techniques for financial option products. The methods derived in the thesis are relevant for risk management for pension funds and insurance companies. This thesis is organized as follows.

In Chapters 2 and 3 we focus on the challenge of extending the Black-Scholes model and in particular on the modeling of- and calibration to options on inflation and real estate indices, which are important risk variables for insurance companies and pension funds.

In Chapter 2, we consider a Heston type inflation model in combination with a Hull-White model for nominal and real interest rates, in which all the correlations can be non-zero. Due to the presence of the Heston dynamics our derived inflation model is able to capture the implied volatility skew/smile, which is present in the inflation option market data. We derive an efficient approximate semi-closed pricing formula to approximate the value of two types of inflation dependent options: index and year-on-year inflation options. The derived pricing formulas allow for an efficient calibration of the inflation model. We also illustrate our approach using a real-life pension fund example, where the Heston Hull-White model is used to determine the value of conditional future indexations. This chapter contains essentially the contents of the article [121].

In Chapter 3, we propose a risk neutral valuation model for real estate derivatives. We demonstrate the application of the model by valuing a put option on a house price index. Autocorrelation in the index returns appears to have a large impact on the option value. We also study the effect of an over- or undervalued real estate market. The observed effects are significant. This chapter contains essentially the contents of the article [129].

In Chapters 4, 5 and 6 we focus on the challenge of nested Monte Carlo simulation.

In Chapter 4, we introduce the so-called Option Interpolation Model for accurate approximations of embedded option values in insurance liabilities. Such a method is important for ex-ante risk management applications. The method is based on interpolation with radial basis functions, which may be used to interpolate scattered data. To reduce computation time we present an inversion method to determine the appearing interpolation function weights. The robustness, accuracy and efficiency of the OIM are analyzed by means of a number of numerical experiments. We show that the proposed approximation method results in highly accurate estimates.

In Chapter 5, we propose a modeling framework for risk neutral stochastic processes nested in a real world stochastic process. The framework is important for insurers that deal with the valuation of embedded options and in particular at future points in time. We make use of the class of State Space Hidden Markov models for modeling the joint behavior of the parameters of a risk neutral model and the dynamics of option market instruments. This modeling concept enables us to perform non-linear estimation, forecasting and robust calibration. The proposed method is applied to the Heston model for which we find highly satisfactory results. We use the estimated Heston model to compute the required capital of an insurance company under Solvency II.

In Chapter 6, we present a HPC framework to improve the performance of nested simulations. We aim to take full advantage of the parallelism of GPUs. We manage to reduce the execution time of a nested simulation application from several hours to tens of minutes. This chapter contains essentially the contents of the article [33].

In Chapter 7 conclusions are presented, as well as an outlook for future research.

CHAPTER 2

Pricing inflation products with stochastic volatility and stochastic interest rates

We consider a Heston type inflation model in combination with a Hull-White model for nominal and real interest rates, in which all the correlations can be non-zero. Due to the presence of the Heston dynamics our derived inflation model is able to capture the implied volatility skew/smile, which is present in the inflation option market data. We derive an efficient approximate semi-closed pricing formula for two types of inflation dependent options: index and year-on-year inflation options. The derived pricing formulas allow for an efficient calibration of the inflation model. We also illustrate our approach using a real-life pension fund example, where the Heston Hull-White model is used to determine the value of conditional future indexations.

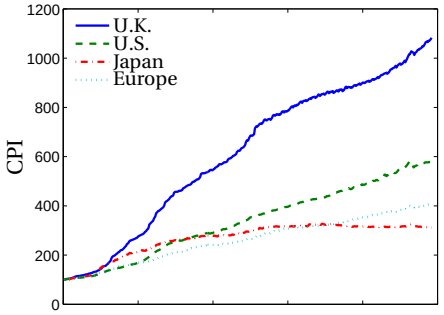
2.1. INTRODUCTION

Inflation-dependent derivatives are increasingly important in financial engineering. As a consequence, inflation markets are becoming more active, liquid and transparent. Inflation is defined as a rise in the general level of prices of goods and services in an economy over a certain period of time (usually one year). The price level is measured by a so-called Consumer Price Index (CPI), which reflects the actual price level of a basket of typical consumer goods. The inflation rate is then defined as the percentage change of the CPI. Inflation derivatives have been traded for over a decade starting in the U.K. in the early 1990s. Since 2000, the market for inflation derivatives has seen a rapid growth in volumes and in types of products across various markets and linked to various domestic and regional inflation indices, such as, French CPI, Euro-zone HICP, U.S. CPI, etc. Broker volumes increased substantially from late-2002, driven by a rise in the need to hedge, for example, retail products.

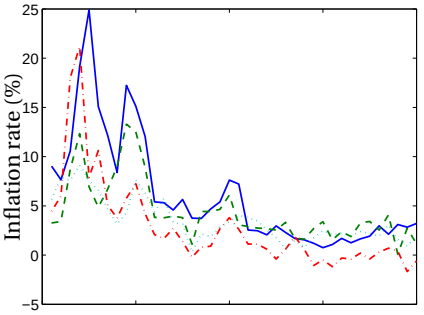
This chapter is based on the article 'Pricing Inflation Products with Stochastic Volatility and Stochastic Interest Rates', published in *Insurance: Mathematics and Economics*, 52(2):286–299, 2013 [121].

Figure 2.1: Historical overview of CPIs and inflation rates.

(a) CPIs (monthly data (31/01/1970-31/12/2010))



(b) Inflation rates (yearly data (31/12/1971-31/12/2010))

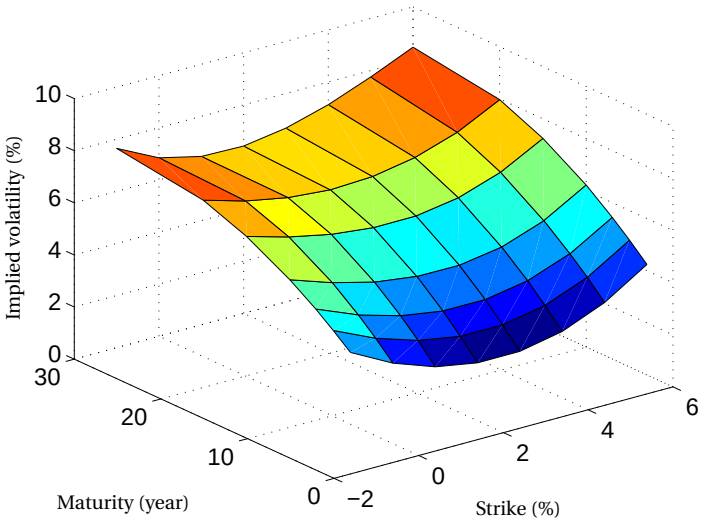


2

Many pension funds, life insurance companies and banks trade these inflation dependent derivatives. For life insurance companies it is important, due to (among others) regulation and new accounting standards, to value their liabilities, which contain inflation dependent embedded options, as market consistent as possible. Pension funds are, for example, interested in the conditional future indexation of pension rights, which can be viewed as an exotic derivative depending on the CPI.

The well-known Fisher equation [52] defines a relation between the nominal and real interest rates on the market and the break-even inflation rate. The break-even inflation rate is the yield spread between nominal and inflation-linked bonds and is a fundamental indicator of inflation expectations. The use of stochastic nominal and real interest rates is crucial for an accurate inflation pricing model. Furthermore, as it turns out, according to [84], there is a significant skew/smile present in the inflation option market data in the sense that the implied Black-Scholes (BS) volatilities are not constant for different strike levels and maturities (like in the stock or currency option markets). In Figure 2.2 the market implied volatility smile is clearly visible.

Figure 2.2: Market implied volatilities of (Euro) inflation indexed options as of September 30, 2010.



Because of this smile/skew effect in the inflation option market data, the Heston model [70] is often used in practice, as this model is capable of capturing this effect. The Heston model is for example well established for pricing stock and currency derivatives, however, not yet for pricing inflation derivatives. The variance process of the CPI is then modeled by a so-called Cox-Ingersoll-Ross (CIR) process (see [32]). Recently, much attention has also been devoted in the literature to stochastic volatility driven by a Schöbel-Zhu process (see for example [133]) in combination with stochastic interest rates to model the CPI. In [133] also a special case of the Heston model in combination with stochastic interest rates was investigated, where some correlations were assumed to be zero. However, the case of a full correlation structure is of particular interest in this chapter. It turns out that these correlation parameters can be influential when pricing exotic derivatives.

We model the CPI by the Heston model, coupled with stochastic nominal and real interest rate processes that are driven by the one-factor Hull-White model. A Hull-White model is a special case of a (multi-factor) Gaussian model (see [18, Chap. 3 and 4]). Our focus is on the fast valuation of inflation index cap/floor options and year-on-year (YoY) inflation cap/floor options[1], because for these products the speed of valuation is crucial for calibration. We derive an efficient pricing engine for these options, so that calibration of our inflation model can be done relatively fast. The key to obtaining the pricing formulas is the derivation of the discounted log-CPI characteristic function (ChF) under the T-forward measure. Since the ChF to be derived contains expressions which have to be evaluated numerically, efficient numerical techniques are developed as well.

This chapter is organized as follows. In Section 2.2 we discuss the coupled inflation-

[1] YoY cap/floor options are defined as a series of forward starting call/put options.

interest rate model and derive the model under the T-forward measure. In Section 2.3 we discuss the valuation of two inflation-dependent options: inflation index caps/floors and YoY inflation caps/floors. In Section 2.4 we present numerical results, which include calibration results. We also devote attention to the comparison between the Heston and the Schöbel-Zhu model. In Section 2.5 we illustrate our approach using a real-life pension fund example, where the Heston Hull-White model is used to determine the value of conditional future indexations. We conclude in Section 2.6.

2.2. SPECIFICATION OF THE INFLATION MODEL

We consider the Heston model in which interest rates are modeled by the one-factor Hull-White interest rate model (see [18, p. 71-80]) to model the CPI. We call this inflation model the Heston Hull-White inflation (HHWi) model.

2.2.1. THE HULL-WHITE INTEREST RATE MODEL

Term structure models, such as the Hull-White (HW) model, describe the evolution of the interest rate curve through time. Modeling the stochastic behavior of the interest rate term structure is particularly important when pricing interest rate-dependent derivatives. The HW model is an example of a no-arbitrage model, because it is designed to exactly fit today's term structure by producing an interest rate behavior which is consistent with this term structure at all times.

The HW model allows for the occurrence of negative rates. Because of the underlying Gaussian distributions it is possible to derive explicit formulas for a number of financial instruments, like interest rate derivatives and bond prices. The different model parameters also provide flexibility and give insight into the dynamic behavior of the term structure.

The nominal and real interest rates, r_n and r_r, under the risk neutral nominal and real economy measures $\mathbb{Q}_n$ and $\mathbb{Q}_r$, respectively, are modeled by one-factor HW models:

$$dr_l(t) = (\theta_l(t) - \kappa_l r_l(t))dt + \sigma_l dW^{r_l}(t), \quad r_l(0) \geq 0, \tag{2.1}$$

where κ_l is a mean-reversion parameter and σ_l a volatility parameter with $l \in \{n, r\}$. The time-dependent function $\theta_l(t)$ is determined by the nominal/real initial term structure as observed in the market via:

$$\theta_l(t) = \frac{\partial f_l(0,t)}{\partial t} + \kappa_l f_l(0,t) + \frac{\sigma_l^2}{2\kappa_l}\left(1 - \exp\left(-2\kappa_l t\right)\right), \quad l \in \{n, r\}. \tag{2.2}$$

The time-dependent function $f_l(t,T)$ $(0 \leq t \leq T)$ denotes the instantaneous forward curve at time t for maturity T. See [18, p. 73] for details.

Nowadays, the quadratic Gaussian and Libor Market Models (among others) are becoming increasingly important to model interest rates (see for example [4, 66]), because they can model an interest rate smile. However, the application of these models is not part of the present work.

2.2.2. THE HESTON HULL-WHITE INFLATION MODEL

We model the evolution of the CPI, denoted by I, and the coupled stochastic variance factor v by the Heston model under the nominal economy spot measure[2], $\mathbb{Q}_n$ (where the nominal and real interest rates follow a Hull-White model, see Eq. (2.1)). The dynamics are given by:

$$\begin{cases} dI(t) = & (r_n(t) - r_r(t))I(t)dt + & \sqrt{v(t)}I(t)dW^I(t), & I(0) \geq 0, \\ dv(t) = & \kappa_v(\bar{v} - v(t))dt + & \sigma_v\sqrt{v(t)}dW^v(t), & v(0) \geq 0, \end{cases} \tag{2.3}$$

where κ_v is a mean-reversion parameter, σ_v a volatility parameter and $\bar{v}$ denotes the long-term variance level. The inflation rate is defined as the percentage change of the CPI, i.e. $\frac{I(t)}{I(\widetilde{t})} - 1$ for $0 \leq \widetilde{t} < t$.

Remark.

- *There exists an analogy between our inflation model and the modeling of currencies, which is also remarked by [76]. It turns out that the inflation model can be used to model currencies by replacing the real interest rate by the foreign interest rate. The CPI then denotes the exchange rate. See for example [66] which employs a very similar model as our inflation model to model the exchange rate.*
- *The instantaneous inflation, $\left(r_n(t) - r_r(t)\right)dt$, in Eq. (2.3) is equal to the instantaneous break-even inflation, which is an important feature in our model.*
- *Seasonality in inflation rates can become important when modeling quarterly or monthly inflation rates. One way to model seasonality is to assume that we have already modeled the seasonally adjusted CPI, $I(t)$, using our inflation model. We can then add a seasonal component, say $\xi(t)$, to obtain the CPI value with seasonality, $\widetilde{I}(t)$. Different approaches can be used to estimate the $\xi(t)$ function, but this is outside the scope of the present section.*

We now need to determine the process for the real interest rate in the nominal economy. Therefore, we apply a change of measure (i.e. change of numéraire) from the risk neutral real economy measure, $\mathbb{Q}_r$, to the nominal economy measure, $\mathbb{Q}_n$. The authors in [18, p. 46] show that this change of measure is equivalent to a change of measure of the numéraire $M_r(t)$ to $M_n(t)/I(t)$, where $M_n(t)$ and $M_r(t)$ are money-savings accounts in the nominal and real economy, respectively, which evolve according to:

$$dM_l(t) = M_l(t)r_l(t)dt, \quad \text{with} \quad l \in \{n, r\}. \tag{2.4}$$

By applying the two-dimensional version of Itô's lemma we derive the following SDE of the numéraire $M_n(t)/I(t)$ under $\mathbb{Q}_n$:

$$d\left(\frac{M_n(t)}{I(t)}\right) = \left(\frac{M_n(t}{I(t)}\right)r_r(t)dt - \left(\frac{M_n(t)}{I(t)}\right)\sqrt{v(t)}dW^I(t), \qquad I(0) \geq 0.$$

[2] In the nominal economy this measure is generated by the nominal money-savings account, $M_n(t)$, which evolves according to Eq. (2.4).

Using [18, Prop. 2.3.1], we then obtain the following real interest rate dynamics under $\mathbb{Q}_n$:

$$dr_r(t) = (\theta_r(t) - \rho_{I,r}\sigma_r\sqrt{v(t)} - \kappa_r r_r(t))dt + \sigma_r dW^{r_r}(t),\ r_r(0) \geq 0.$$

The correlation structure between the Brownian motions $d\mathbf{W}_t = \left(dW_t^I, dW_t^v, dW_t^{r_n}, dW_t^{r_r}\right)^{\mathrm{T}}$ is defined by the following symmetric instantaneous correlation matrix:

$$d\mathbf{W}_t\left(d\mathbf{W}_t\right)^{\mathrm{T}} = \begin{pmatrix} 1 & \rho_{I,v} & \rho_{I,n} & \rho_{I,r} \\ . & 1 & \rho_{v,n} & \rho_{v,r} \\ . & . & 1 & \rho_{n,r} \\ . & . & . & 1 \end{pmatrix} dt. \tag{2.5}$$

2.2.3. INFLATION DYNAMICS UNDER THE T-FORWARD MEASURE

To value inflation-dependent derivatives it is convenient to use the inflation model under the T-forward nominal economy measure (instead of the spot measure), which we denote by $\mathbb{Q}_n^T$ (see for example [132]). This measure is generated by the nominal zero-coupon bond, $P_n(t,T)$. In other words, under the T-forward measure the forward CPI, I_T, is a martingale, i.e.

$$P_n(t,T)\mathbb{E}^T\left[I_T(T) \mid \mathscr{F}_t\right] = P_n(t,T)I_T(t) = P_r(t,T)I(t), \tag{2.6}$$

where $P_n(t,T)$ and $P_r(t,T)$ are nominal and real zero-coupon bonds, respectively. The inflation model under this T-forward measure is given in Proposition 2.2.1.

Proposition 2.2.1. *The inflation model under the T-forward nominal economy measure $\left(\mathbb{Q}_n^T\right)$, with a full matrix of correlations, is given by:*

$$\begin{cases} dI_T(t) = I_T(t)\left(\sqrt{v(t)}dW_T^I(t) + \sigma_n B_n(t,T)dW_T^n(t) - \sigma_r B_r(t,T)dW_T^r(t)\right), \\ dv(t) = \left(\kappa_v(\bar{v} - v(t)) - \sigma_v\sigma_n\rho_{v,n}B_n(t,T)\sqrt{v(t)}\right)dt + \sigma_v\sqrt{v(t)}dW_T^v(t), \end{cases}$$

where I_T denotes the forward CPI under the T-forward measure. The interest rate processes are given by:

$$\begin{cases} dr_n(t) = & \left(\theta_n(t) - \sigma_n^2 B_n(t,T) - \kappa_n r_n(t)\right)dt + & \sigma_n dW_T^{r_n}(t), \\ dr_r(t) = & \left(\theta_r(t) - \rho_{I,r}\sigma_r\sqrt{v(t)} - \sigma_n\sigma_r\rho_{n,r}B_n(t,T) - \kappa_r r_r(t)\right)dt + & \sigma_r dW_T^{r_r}(t), \end{cases}$$

where the time-dependent function $\theta_l(t)$ is given by Eq. (2.2) and

$$B_l(t,T) = \frac{1}{\kappa_l}\left(1 - \exp\left(-\kappa_l(T-t)\right)\right),$$

for $l \in \{n, r\}$.

The dynamics of the forward CPI are simplified by changing to logarithmic transformed coordinates, where we define $x_T(t) := \log I_T(t)$[3]*:*

$$\begin{aligned} dx_T(t) &= -\frac{1}{2}\Big(\nu(t)+\sigma_n^2 B_n^2(t,T)+\sigma_r^2 B_r^2(t,T)+2\rho_{\nu,n}\sigma_n B_n(t,T)\sqrt{\nu(t)} \\ &\quad -2\rho_{\nu,r}\sigma_r B_r(t,T)\sqrt{\nu(t)}-2\rho_{n,r}\sigma_n\sigma_r B_n(t,T)B_r(t,T)\Big)dt \\ &\quad +\sqrt{\nu(t)}dW_T^I(t)+\sigma_n B_n(t,T)dW_T^n(t)-\sigma_r B_r(t,T)dW_T^r(t). \end{aligned}$$

Proof. The general outline of the proof is as follows. From Eq. (2.6) it follows that

$$I_T(t) = I(t)\frac{P_r(t,T)}{P_n(t,T)}, \tag{2.7}$$

where the dynamics of $I(t)$ are given in Section 2.2.1. The dynamics of $I_T(t)$ are obtained by applying Itô's lemma to Eq. (2.7) in combination with the dynamics of $I(t)$ and the dynamics of the real and nominal zero-coupon bonds, $P_r(t,T)$ and $P_n(t,T)$, under the nominal economy measure $(\mathbb{Q}_n)$. Expressing the full model in terms of independent Brownian motions simplifies the derivation of the Radon-Nikodým derivative (see [18, p. 45 and 911]). By computing the Itô derivative of this Radon-Nikodým derivative the Girsanov kernel for the transition from $\mathbb{Q}_n$ to $\mathbb{Q}_n^T$ is derived and finishes the proof. For the full proof we refer to [66]. □

From Proposition 2.2.1 we note that under the T-forward nominal economy measure $\left(\mathbb{Q}_n^T\right)$ the forward CPI does not depend directly on the real and nominal interest rate processes, $r_r(t)$ and $r_n(t)$, but only depends on the Brownian motions $dW^{r_n}(t)$ and $dW^{r_r}(t)$. Actually the forward CPI depends on all the Brownian motions since the correlations can be non-zero. The key is the independence of the state variables.

2.3. PRICING FORMULAS

We discuss the pricing of two inflation dependent options. The pricing of inflation index options is discussed in Section 2.3.1 and the pricing of YoY inflation options is discussed in Section 2.3.2. In Section 2.3.3 we show numerical results of the derived pricing formulas of forward starting options.

2.3.1. INFLATION INDEXED OPTIONS

We briefly discuss the pricing of inflation indexed cap and floor options. The inflation model, which we use for option pricing, is given in Section 2.2.1 under the measure $\mathbb{Q}_n$ and by Proposition 2.2.1 under the measure $\mathbb{Q}_n^T$.

The model price of an inflation indexed cap/floor option maturing at time T with strike level[4] $K := (1+\tilde{k})^{\mathrm{T}}$ (the expression $(1+\tilde{k})^{\mathrm{T}}$ means $1+\tilde{k}$ to the power T) written on the inflation index (the CPI) (with $\omega = 1$ for a cap option and $\omega = -1$ for a floor option) is

[3] Note that this transformation is well defined since $I(0) > 0$ and, thus, $I_T(0) > 0$.
[4] The strike level $\tilde{k}$ is (market data) input.

given by

$$\eta_{\overline{\Pi}}(t,T,\tilde{k},\omega) := M_n(t)\mathbb{E}^{\mathbb{Q}_n}\left[\frac{\max\left(\omega\left(I(T)-K\right),0\right)}{M_n(T)} \mid \mathscr{F}_t\right], \tag{2.8}$$

where $M_n(t)$ indicates the nominal money-savings account, which evolves according to Eq. (2.4). We denote the market price by $\overline{\Pi}$. Since the stochastic expressions $M_n(T)$ and $\max\left(\omega\left(I(T)-K\right),0\right)$ are not independent, the computation of the expectation under the $\mathbb{Q}_n$ measure is rather involved.

It turns out that the complexity of the problem is greatly reduced under the T-forward measure. We then get the following pay-off structure:

$$\eta_{\overline{\Pi}}(t,T,\tilde{k},\omega) = P_n\left(t,T\right)\mathbb{E}^{\mathbb{Q}_n^T}\left[\max\left(\omega\left(I_T(T)-K\right),0\right) \mid \mathscr{F}_t\right]. \tag{2.9}$$

From the two pay-off structures in Eqs. (2.8) and (2.9) we note that the pay-off structure under the T-forward measure has a simpler form since the price of the pure discount bond at time $t=0$ is directly observable in the market.

$\eta_{\overline{\Pi}}(t,T,\tilde{k},\omega)$ in Eq. (2.9) can also be formulated in integral form:

$$\eta_{\overline{\Pi}}(t,T,\tilde{k},\omega) = P_n\left(t,T\right)\int_{\mathbb{R}}\max\left(\omega K\left(\exp\left(y\right)-1\right),0\right)\widetilde{f}(y|x)dy,$$

where $\widetilde{f}(y|x)$ denotes the probability density function of $y := \log\left(\frac{I_T(T)}{K}\right)$ given $x := \log\left(\frac{I_T(t)}{K}\right)$.

Fourier-based methods[5] can be used to compute these integrals in the case the density function is not known in advance. These methods rely on the existence of the ChF. The derivation of the ChF for this particular option is discussed in [66]. We denote the corresponding approximation of the full-scale HHWi model by HHWi-i. For this model we can employ Fourier-based methods for efficient pricing of inflation index options.

2.3.2. YEAR-ON-YEAR INFLATION OPTIONS

We discuss the pricing of YoY inflation cap/floor options by describing the general pricing methodology. In general, a cap/floor option, $\widetilde{\Pi}$, is defined by a series of so-called caplet/floorlet options, $\widehat{\Pi}$, i.e:

$$\widetilde{\Pi}(\omega,t,\tau,T,\bar{k}) = \sum_{k=1}^{n}\widehat{\Pi}(\omega,t,T_{k-1},T_k,\bar{k}),$$

where $\omega=1$ for a cap/caplet option and $\omega=-1$ for a floor/floorlet option. Furthermore, $\tau := T_k - T_{k-1}$ defines the tenor parameter with $T_0=0$ and $T_n=T$. The integer n denotes the number of caplets/floorlets in the cap/floor option. This integer is dependent on the tenor parameter, which is in practice often a fixed interval. The strike level is given by $\bar{k}$. So, the pricing of a YoY inflation cap/floor option reduces to the pricing of a series of YoY inflation caplet/floorlet options.

[5] See for example [28, 47].

The model price of a YoY inflation caplet/floorlet option starting at time T_{k-1} ($0 \le t \le T_{k-1}$) and maturing at time T_k ($T_{k-1} \le T_k$), written on the inflation index, is given by

$$\eta_{\widehat{\Pi}}(\omega, t, T_{k-1}, T_k, \bar{k}) = M_n(t)\mathbb{E}^{\mathbb{Q}_n}\left[\frac{\max\left(\omega\left(\frac{I(T_k)}{I(T_{k-1})} - (\bar{k}+1)\right), 0\right)}{M_n(T_k)} \mid \mathscr{F}_t\right],$$

where $M_n(t)$ indicates the nominal money-savings account, which evolves according to Eq. (2.4).

By changing the measure from $\mathbb{Q}_n$ to the T_k-forward measure, $\mathbb{Q}_n^{T_k}$, with $k = 1, \ldots, n$, and by using $K^* := 1 + \bar{k}$, we arrive at the following pricing problem:

$$\eta_{\widehat{\Pi}}(\omega, t, T_{k-1}, T_k, \bar{k}) = P_n(t, T_k)\mathbb{E}^{T_k}\left[\max\left(\omega\left(\frac{I(T_k)}{I(T_{k-1})} - K^*\right), 0\right)\Big|\mathscr{F}_t\right].$$

Since the T_k-forward CPI, $I_{T_k}(t) = \frac{P_r(t,T_k)}{P_n(t,T_k)} I(t)$, under the T_k-forward measure is a martingale with numéraire $P_n(t, T_k)$ and $I_{T_k}(T_k) = I(T_k)$, we can simply write:

$$\eta_{\widehat{\Pi}}(\omega, t, T_{k-1}, T_k, \bar{k}) = P_n(t, T_k)\mathbb{E}^{T_k}\left[\max\left(\omega\left(\frac{P_r(T_{k-1}, T_k)}{P_n(T_{k-1}, T_k)}\frac{I_{T_k}(T_k)}{I_{T_k}(T_{k-1})} - K^*\right), 0\right)\Big|\mathscr{F}_t\right].$$

The dynamics for $I_{T_k}(t)$ under the T_k-forward measure are given by Proposition 2.2.1.

For numerical experiments we make use of the put-call parity to price options of call type, so in this case caplet options. In other words, when for example a floorlet option, $\eta_{\widehat{\Pi}}(-1, t, T_1, T_2, \bar{k})$, with strike $\bar{k}$ and times $0 \le t \le T_1 < T_2$, is computed, the price of the corresponding caplet option $\eta_{\widehat{\Pi}}(1, t, T_1, T_2, \bar{k})$ is computed by:

$$\eta_{\widehat{\Pi}}(1, t, T_1, T_2, \bar{k}) = \eta_{\widehat{\Pi}}(-1, t, T_1, T_2, \bar{k}) + P_n(t, T_1)P_r(T_1, T_2) - P_n(t, T_2)\left(1 + \bar{k}\right),$$

where P_n and P_r are nominal and real zero-coupon bonds, respectively.

As already mentioned, to apply Fourier-based pricing methods we have to derive the (forward) ChF belonging to this option, which is the topic of the next section.

DERIVATION OF THE (FORWARD) CHARACTERISTIC FUNCTION

By setting[6]

$$X(T_{k-1}, T_k) = \frac{P_r(T_{k-1}, T_k)}{P_n(T_{k-1}, T_k)}\frac{I_T(T_k)}{I_T(T_{k-1})}, \quad \text{for} \quad \mathrm{k} = 1, \ldots, \mathrm{n},$$

we perform the log-transformation:

$$\begin{aligned} x(T_{k-1}, T_k) &:= \log X(T_{k-1}, T_k) = \log\left(\frac{P_r(T_{k-1}, T_k)}{P_n(T_{k-1}, T_k)}\frac{I_{T_k}(T_k)}{I_{T_k}(T_{k-1})}\right) \\ &= \log I_{T_k}(T_k) - \log I_{T_k}(T_{k-1}) + \log P_r(T_{k-1}, T_k) - \log P_n(T_{k-1}, T_k). \end{aligned}$$

[6] We note that the same approach, for deriving the forward ChF, is used as in [133]. For convenience, the notation is analogue to the notation in [133].

We derive the forward ChF for the process $x(T_{k-1}, T_k)$:

$$\phi_{YoY}(u,t,x(T_{k-1,},T_k)) := \mathbb{E}^{T_k}\left[\exp\left(iux(T_{k-1},T_k)\right)|\mathscr{F}_t\right]. \tag{2.10}$$

By substitution we have:

$$\begin{aligned}\phi_{YoY}(u,t,x(T_{k-1},T_k)) = \quad & \mathbb{E}^{T_k}\Bigg[\exp\Bigg(iu\Big(\log I_{T_k}(T_k) - \log I_{T_k}(T_{k-1}) \\ & + \log P_r(T_{k-1},T_k) - \log P_n(T_{k-1},T_k)\Big)\Bigg)|\mathscr{F}_t\Bigg].\end{aligned}$$

Now, by iterated expectations we find:

$$\begin{aligned}\phi_{YoY}(u,t,x(T_{k-1},T_k)) = \quad & \mathbb{E}^{T_k}\Bigg[\mathbb{E}^{T_k}\Bigg[\exp\Bigg(iu\Big(\log I_{T_k}(T_k) - \log I_{T_k}(T_{k-1}) \\ & + \log P_r(T_{k-1},T_k) - \log P_n(T_{k-1},T_k)\Big)\Bigg)|\mathscr{F}_{k-1}\Bigg]|\mathscr{F}_t\Bigg].\end{aligned}$$

Since $I_T(T_{k-1})$, $P_n(T_{k-1},T_k)$ and $P_r(T_{k-1},T_k)$ are $I_{T_{k-1}}$ measurable[7], we can write:

$$\begin{aligned}\phi_{YoY}(u,t,x(T_{k-1},T_k)) = \quad & \mathbb{E}^{T_k}\Bigg[\exp\Bigg(-iu\Big(\log I_{T_k}(T_{k-1}) - \log P_r(T_{k-1},T_k) \\ & + \log P_n(T_{k-1},T_k)\Big)\Bigg)\cdot\mathbb{E}^{T_k}\left[\exp\left(iu\log I_{T_k}(T_k)\right)|\mathscr{F}_{k-1}\right]|\mathscr{F}_t\Bigg].\end{aligned}$$

The last expectation equals the characteristic function for $\log I_{T_k}(T_k)$, i.e.

$$\phi_i(u,\log I_T(T_k),T_{k-1},T_k) := \mathbb{E}^{T_k}\left[\exp\left(iu\log I_{T_k}(T_k)\right)|\mathscr{F}_{k-1}\right].$$

In [66] an affine approximation is found for this ChF, i.e.:

$$\phi_{i,1} := \exp\Big(A(u,T_k-T_{k-1}) + iu\log I_{T_k}(T_{k-1}) + C(u,T_k-T_{k-1})v(T_{k-1})\Big), \tag{2.11}$$

with functions $A(u,\tau)$ and $C(u,\tau)$ given by Eqs. (2.12) and (2.13). By subscripts (like the $\phi_{i,1}$ in Eq. (2.11)) we indicate subsequent approximations.

In [66] it is noted that the Kolmogorov backward partial differential equation, for which ϕ in Eq. (2.10) is the solution, contains non-affine $\sqrt{v}$-terms, so that finding the solution is nontrivial. Approximation of these $\sqrt{v}$-terms by a linearization technique leads to an approximating closed-form solution of the ChF.

The functions $A(u,\tau)$ and $C(u,\tau)$ in Eq. (2.11) are given by:

$$\begin{aligned}A(u,\tau) \quad := \quad & \int_0^\tau \left(\kappa_v\bar{v} - \rho_{v,n}\sigma_v\sigma_n\varphi(s)B_n(s)(1-iu) - \rho_{v,r}\sigma_v\sigma_r\varphi(s)B_r(s)\right)C(s)ds \\ & + (u^2+iu)\int_0^\tau \Psi(s,\varphi(s))ds, \end{aligned} \tag{2.12}$$

$$C(u,\tau) \quad := \quad \frac{1-\exp\left(-d\tau\right)}{\sigma_v^2(1-g\exp\left(-d\tau\right))}\left(\kappa_v - \rho_{I,v}\sigma_v iu - d\right), \tag{2.13}$$

[7] See [18].

where $\varphi(t) := \mathbb{E}\left[\sqrt{v(t)}\right]$[8], $d := \sqrt{(\kappa_v - \rho_{I,v}\sigma_v iu)^2 - \sigma_v^2 iu(iu-1)}$ and $g := \frac{\kappa_v - \rho_{I,v}\sigma_v iu - d}{\kappa_v - \rho_{I,v}\sigma_v iu + d}$. Furthermore,

$$\begin{aligned}\Psi(t,\varphi(t)) \quad := \quad & (\rho_{I,r}\sigma_r B_r(t,T) - \rho_{I,n}\sigma_n B_n(t,T))\varphi(t) + \rho_{n,r}\sigma_n\sigma_r B_n(t,T)B_r(t,T) \\ & -\frac{1}{2}\left(\sigma_n^2 B_n^2(t,T) + \sigma_r^2 B_r^2(t,T)\right).\end{aligned}$$

The ChF, ϕ, is then approximated by:

$$\phi_{YoY,1} = \mathbb{E}^{T_k}\left[\exp\left(-iu\log I_{T_k}(T_{k-1}) + iu\log P_r(T_{k-1},T_k) - iu\log P_n(T_{k-1},T_k)\right)\phi_{i,1}\middle|\mathscr{F}_t\right].$$

Due to Eq. (2.11) we have:

$$\begin{aligned}\phi_{YoY,1} = \quad & \mathbb{E}^{T_k}\left[\exp\left(A(u,T_k - T_{k-1}) + C(u,T_k - T_{k-1})v(T_{k-1})\right) \times \right. \\ & \left. \exp\left(-iu\log P_n(T_{k-1},T_k)\right)\exp\left(iu\log P_r(T_{k-1},T_k)\right)\middle|\mathscr{F}_t\right]. \qquad (2.14)\end{aligned}$$

As the underlying nominal interest-rate model is the Hull-White model, the zero-coupon bond (ZCB) $P_n(T_{k-1},T_k)$ is given by, see [18, p. 75-78]:

$$P_n(T_{k-1},T_k) = \exp\left(A_n(T_{k-1},T_k) - B_n(T_{k-1},T_k)r_n(T_{k-1})\right), \qquad (2.15)$$

with analytically known functions $A_n(T_{k-1},T_k)$ and $B_n(T_{k-1},T_k)$. However, since we work under the nominal economy measure $\mathbb{Q}_n$, the dynamics of the real interest rate are not affine and, as a consequence, the dynamics of P_r are not affine. Hence, the derivation of the dynamics of P_r is nontrivial.

By approximating the variance process under $\mathbb{Q}_n$ (see Section 2.2.1) by its expectation, the process of the real interest rate, conditional on $\mathscr{F}_s$, is affine and normally distributed. Following the approach as outlined in [18, Chap. 3.3] we derive:

$$\begin{aligned}A_r(T_{k-1},T_k) = \quad & \log\frac{P_r(0,T_k)}{P_r(0,T_{k-1})}(B_r(T_{k-1},T_k)f_r(0,T_{k-1}) + \Lambda(T_{k-1},T_k) \\ & -\frac{\sigma_r^2}{4\kappa_r}(1-\exp\left(-2\kappa_r T_{k-1}\right))B_r(T_{k-1},T_k)^2), \\ B_r(T_{k-1},T_k) = \quad & \frac{1}{\kappa_r}\left(1-\exp\left(-\kappa_r(T_k - T_{k-1})\right)\right),\end{aligned}$$

where

$$\begin{aligned}\Lambda(T_{k-1},T_k) = \quad & \mathbb{E}\left[\sqrt{v(T_k)}\right]\frac{\rho_{I,r}\sigma_r}{\kappa_r}(T_k - T_{k-1} - B_r(T_{k-1},T_k) - B_n(T_{k-1},T_k) \\ & +\frac{1}{\kappa_n+\kappa_r}(1-\exp\left(-(\kappa_n+\kappa_r)(T_k - T_{k-1})\right))).\end{aligned}$$

[8] In [65] approximations are proposed for $\mathbb{E}\left[\sqrt{v(t)}\right]$, which are also used for the numerical experiments.

By substituting the nominal and real ZCB expressions into the expression in Eq. (2.14) the approximating ChF in Eq. (2.14) is now given by:

$$\begin{aligned}\phi_{YoY,1} = \quad & \exp\Big(iu\big(A_r(T_{k-1},T_k)-A_n(T_{k-1},T_k)\big)\Big)\exp\big(A(u,T_k-T_{k-1})\big)\times \\ & \mathbb{E}^{T_k}\Big[\exp\big(C(u,T_k-T_{k-1})v(T_{k-1})\big)\times \\ & \exp\Big(iu\big(B_n(T_{k-1},T_k)r_n(T_{k-1})-B_r(T_{k-1},T_k)r_r(T_{k-1})\big)\Big)\big|\mathscr{F}_t\Big]. \quad (2.16)\end{aligned}$$

The Laplace transform in Eq. (2.16) is of a very complicated form. In order to find a closed-form solution for Eq. (2.16), additional assumptions of independence between processes are required.

A basic approximation to Eq. (2.16) is given by:

$$\begin{aligned}\phi_{YoY,2} = \quad & \exp\Big(iu\big(A_r(T_{k-1},T_k)-A_n(T_{k-1},T_k)\big)+A(u,T_k-T_{k-1})\Big)\times \\ & \mathbb{E}^{T_k}\Big[\exp\big(C(u,T_k-T_{k-1})v(T_{k-1})\big)\big|\mathscr{F}_t\Big]\times \\ & \mathbb{E}^{T_k}\Big[\exp\Big(iu\big(B_n(T_{k-1},T_k)r_n(T_{k-1})-B_r(T_{k-1},T_k)r_r(T_{k-1})\big)\Big)\big|\mathscr{F}_t\Big] \quad (2.17)\end{aligned}$$

The approximation above consists of two expectations under the T_k-forward measure. Since the nominal and real interest rates, $r_n(T_{k-1})$ and $r_r(T_{k-1})$, are normally distributed, the sum of these two normally distributed random variables is also normally distributed and the ChF of this sum can be found analytically. Furthermore, since $v(T_{k-1})$ is non-central chi-square distributed the corresponding ChF can also be found analytically. Result 2.3.1 and Lemma 2.3.1 provide these solutions.

Result 2.3.1. *For given times $0 \le s \le t \le T$, nominal and real interest rate processes r_n and r_r, as defined in Proposition 2.2.1, and $Y(t,T) := B_n(t,T)r_n(t) - B_r(t,T)r_r(t)$, the following holds:*

$$\mathbb{E}^T\Big[\exp\big(iuY(t,T)\big)\big|\mathscr{F}_s\Big] \approx \exp\Big(iu\mathbb{E}^T\Big[Y(t,T)\big|\mathscr{F}_s\Big]-\frac{1}{2}u^2\mathbb{V}\mathrm{ar}^T\Big(Y(t,T)\big|\mathscr{F}_s\Big)\Big),$$

where r_n evolves under $\mathbb{Q}_n^T$ according to Proposition 2.2.1. To ensure that the real interest rate process is normally distributed under $\mathbb{Q}_n^T$, we assume that it evolves according to

$$dr_r(t) = \Big(\theta_r(t)-\rho_{I,r}\sigma_r\mathbb{E}\Big[\sqrt{v(t)}\Big]-\sigma_n\sigma_r\rho_{n,r}B_n(t,T)-\kappa_r r_r(t)\Big)dt+\sigma_r dW_T^{r_r}(t).$$

The random variable $Y(t,T)$ is then normally distributed with expectation and variance given by:

$$\begin{aligned}\mathbb{E}^T\Big[Y(t,T)\big|\mathscr{F}_s\Big] &= B_n(t,T)\mathbb{E}^T\Big[r_n(t)\big|\mathscr{F}_s\Big]-B_r(t,T)\mathbb{E}^T\Big[r_r(t)\big|\mathscr{F}_s\Big], \\ \mathbb{V}\mathrm{ar}^T\Big[Y(t,T)\big|\mathscr{F}_s\Big] &= B_n^2(t,T)\mathbb{V}\mathrm{ar}^T\Big[r_n(t)\big|\mathscr{F}_s\Big]+B_r^2(t,T)\mathbb{V}\mathrm{ar}^T\Big[r_r(t)\big|\mathscr{F}_s\Big] \\ &\quad -2B_n(t,T)B_r(t,T)Cov^T\Big[r_n(t),r_r(t)\big|\mathscr{F}_s\Big],\end{aligned}$$

with

$$Cov^T\left[r_n(t), r_r(t)\,\middle|\,\mathscr{F}_s\right] = \rho_{n,r}\sqrt{\mathbb{V}\mathrm{ar}^T\left[r_n(t)\,\middle|\,\mathscr{F}_t\right]\mathbb{V}\mathrm{ar}^T\left[r_r(t)\,\middle|\,\mathscr{F}_s\right]}.$$

Proof. By approximating the variance process under $\mathbb{Q}_n$ (see Section 2.2.1) by its expectation the process of the real interest rate, conditional on $\mathscr{F}_s$, is normally distributed.

Next, since the random variable $Y(t,T)$ is defined as a (weighted) sum of normally distributed random variables[9], $Y(t,T)$ is also normally distributed. The characteristic function for any normally distributed random variable X with expectation μ and variance σ^2 is given by

$$\phi_X(u) = \mathbb{E}[\exp(iuX] = \exp\left(iu\mu - \frac{1}{2}\sigma^2u^2\right).$$

The proof is finished by the appropriate substitutions. □

Lemma 2.3.1. *For $0 \le s \le t \le T$ the Laplace transform of* $\mathbb{E}^T\left[\exp\left(C(u,T-t)v(t)\right)\middle|\mathscr{F}_s\right]$ *is given by:*

$$\begin{aligned}\mathbb{E}^T\left[\exp\left(C(u,T-t)v(t)\right)\middle|\mathscr{F}_s\right] &= \psi(u,s,t,T)^{\frac{2\kappa_v\bar{v}}{\gamma^2}}\exp\Big(\psi(u,s,t,T)\exp\left(-\kappa_v(t-s)\right)\times \\ &\quad C(u,T-t)v(t)\Big),\end{aligned} \tag{2.18}$$

where

$$\psi(u,s,t,T) := \frac{1}{1-\frac{2\gamma^2}{4\kappa_v}\left(1-\exp\left(-\kappa_v(t-s)\right)\right)C(u,T-t)} \ge 0.$$

The function $C(u,T-t)$ is given in Eq. (2.13).

Proof. Since the variance process $v(t)$, conditional on $\mathscr{F}_s$, is distributed as a constant $c := \frac{\sigma_v^2(1-\exp(-\kappa_v(t-s)))}{4\kappa_v}$ times a non-central chi-square distribution with $d := \frac{4\kappa_v\bar{v}}{\sigma_v^2}$ degrees of freedom and non-centrality parameter $\lambda := \frac{4\kappa_v\exp(-\kappa_v(t-s))}{\sigma_v^2(1-\exp(-\kappa_v(t-s)))}$, the proof is straightforward, see [32]. □

We denote the approximation in Eq. (2.17) of the full-scale HHWi model by HHWi-YoY. For this model we can employ Fourier-based methods for efficient pricing of YoY inflation options.

2.3.3. NUMERICAL EXPERIMENT: VALUATION OF YEAR-ON-YEAR INFLATION OPTIONS

To analyze the performance of the approximations introduced for the YoY inflation options we compute the initial ($t = 0$) implied Black-Scholes volatilities for different strike levels using the full-scale HHWi model and the HHWi-YoY model. This is done by inverting the characteristic function using Fourier-based methods. We consider two test cases:

[9] $r_n(t)$ and $r_r(t)$, conditional on $\mathscr{F}_t$, are normally distributed [18, Chap. 3.3.1].

- Case I: the forward starting option starts at $T_1 = 4$ and matures at $T_2 = 5$.
- Case II: the forward starting option starts at $T_1 = 29$ and matures at $T_2 = 30$.

For the generation of risk neutral (RN) scenarios we make use of an advanced simulation scheme including exact simulation (also called unbiased simulation) for the interest rate and variance processes (see, for example, [3, 19]). To reduce the variance of the MC estimator we use 100.000 scenarios in combination with two variance reduction techniques (i) antithetic sampling and (ii) Empirical Martingale Simulation (EMS) (see [43, 64]).

As the base parameter setting we use the parameters as specified below:

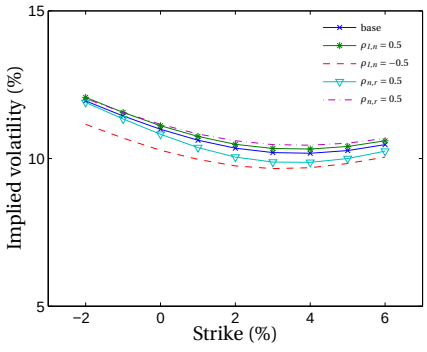
$$\kappa_v = 0.3, \quad v(0) = 0.04, \quad \bar{v} = 0.04, \quad \sigma_v = 0.6, \quad \rho_{I,v} = -0.7,$$

with interest rate volatilities $\sigma_n = 0.0089$, $\sigma_r = 0.0084$ and correlations $\rho_{I,n} = \rho_{I,r} = \rho_{v,n} = \rho_{v,r} = 0$ and $\rho_{n,r} = 0$ (unless stated otherwise). To test the pricing accuracy we use an extreme test case, i.e. the Feller condition, $2\kappa_v \bar{v} > \sigma_v^2$, is not satisfied, so that inflation volatilities can attain zero. These parameters are not calibrated to market data; this topic will be discussed in Section 2.4.

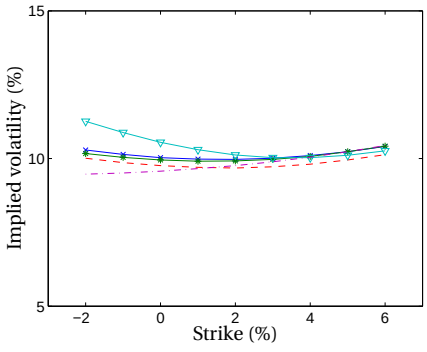
First we investigate the sensitivity of the pricing of YoY inflation options to the correlation parameters by performing a MC simulation. We therefore vary the correlations $\rho_{I,n}$ and $\rho_{n,r}$. The results for cases I and II are presented in Figures 2.3a and 2.3b, respectively.

Figure 2.3: Sensitivity to correlations using a Monte Carlo simulation to the full-scale HHWi model.

(a) Case I



(b) Case II

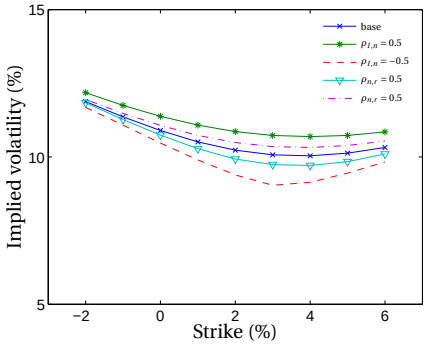


Observing Figures 2.3a and 2.3b we conclude that for this parameter setting the correlation parameters $\rho_{I,n}$ and $\rho_{n,r}$ are influential regarding the change in implied volatility.

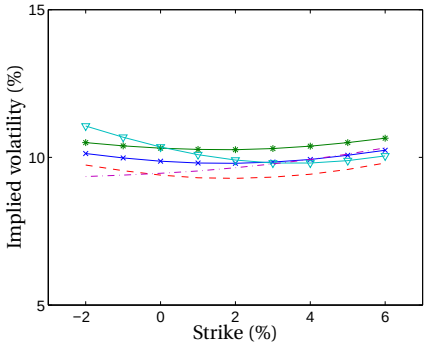
Next, we perform the same experiment using the HHWi-YoY model. The results for cases I and II are presented in Figures 2.4a and 2.4b, respectively.

Figure 2.4: Sensitivity to correlations using the HHWi-YoY model.

(a) Case I



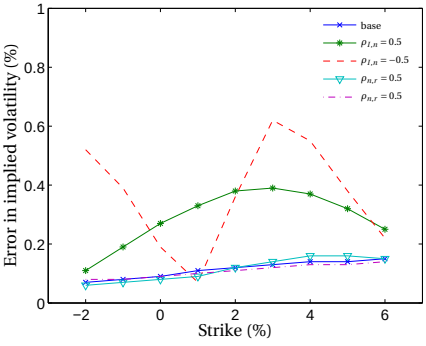
(b) Case II



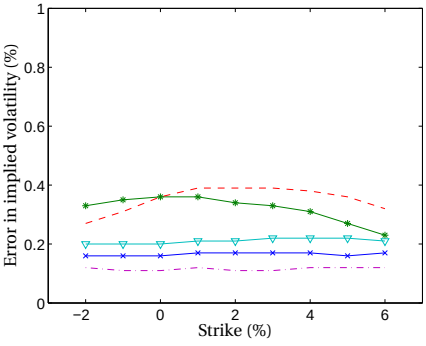
Figures 2.5a and 2.5b show the difference in implied volatility between the full-scale HHWi and HHWi-YoY model.

Figure 2.5: Difference between the HHWi and HHWi-YoY model.

(a) Case I



(b) Case II



From Figures 2.5a and 2.5b we can conclude that the maximum error for cases I and II is equal to 0.6% point and 0.4% point in terms of implied volatilities, respectively. In both cases we considered $\tau = 1$, which is common when YoY forward starting options are considered.

2.4. CALIBRATION RESULTS

A calibration procedure consists of the computation of $\min_{\Omega}\{\|C - \eta_C\|\}$, where C denotes the market price, η_C the model price, Ω the set of parameters (including constraints) and $\|\cdot\|$ some norm. In our case market data are available for pairs $\left(\overline{T}, \overline{K}\right)$,

with $\overline{T}$ denoting the option maturity and $\overline{K}$ the strike level. For the norm we take the Euclidean norm, so that calibration in our case consists of computing:

$$\min_{\Omega}\left\{\|C-\eta_C\|_p\right\} = \min_{\Omega}\left\{\left(\sum_{j=1}^{m}\sum_{k=1}^{n}\left|C(\overline{T}_j,\overline{K}_k)-\eta_C(\overline{T}_j,\overline{K}_k)\right|^p\right)^{\frac{1}{p}}\right\}, \tag{2.19}$$

where we use $p=2$. We note that also the p-norm of the difference of market and model implied volatilities could be minimized. However, since then in every iteration step of the optimization procedure an extra numerical inversion has to be performed, which may lead to numerical difficulties, this is not the method of choice. Market prices of plain vanilla options are often used for calibration, because this data is available.

In the calibration procedure it is possible to incorporate both types of inflation options. This is easily done when we specify the market option price C (and, thus, also the corresponding model value η_C) as an inflation indexed cap/floor or a YoY inflation caplet/floorlet with corresponding strike level and maturity. It is also possible to assign different weights to different calibration points.

The minimization problem in Eq. (2.19) is solved iteratively using a numerical minimization algorithm. We first sample random starting points and then we refine this solution using the well-known Levenberg-Marquardt least-squares algorithm, which is a local minimization method. This procedure is repeated and the best solution is kept.

We show calibration results for the full-scale HHWi model (see Section (2.2.1)). We first calibrate the one-factor Hull-White interest rate model to interest rate options, like swaptions and/or interest rate cap/floor options, see [18, Chap. 2 and 3], to determine the interest rate model parameters κ_n, κ_r, σ_n and σ_r (see Section 2.2.1). Conditional on the parameters of the interest rate model, we calibrate the inflation model to inflation indexed cap/floor options and/or YoY inflation caplet/floorlet options with Fourier-based methods.

For the correlation parameters we perform the following calibration:

1. The correlation parameters between 'observable' variables, i.e. $\rho_{I,n}=0.36$, $\rho_{I,r}=-0.29$ and $\rho_{n,r}=0.78$ are determined using historical information[10] in the sample period $1985-2009$.

2. The correlation parameter $\rho_{I,v}$ is determined in the calibration process. Appropriate bounds for this parameter are used in the calibration process so that the correlation matrix remains positive definite.

3. The correlation parameters, $\rho_{r,v}$ and $\rho_{n,v}$ are derived from a conditional sampling method.

Because of the procedure mentioned above, we start the calibration with the following

[10] This is industrial practice.

correlation matrix, which is defined in Eq. (2.5):

$$\begin{pmatrix} 1 & \rho_{I,v} & \rho_{I,n} & \rho_{I,r} \\ . & 1 & \rho_{v,n} & \rho_{v,r} \\ . & . & 1 & \rho_{n,r} \\ . & . & . & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{I,v} & 0.36 & -0.29 \\ . & 1 & \rho_{v,n} & \rho_{v,r} \\ . & . & 1 & 0.78 \\ . & . & . & 1 \end{pmatrix}, \tag{2.20}$$

where the correlation parameters $\rho_{I,v}$, $\rho_{v,n}$ and $\rho_{v,r}$ are to be determined.

The inflation option market data, as of September 30, 2010, which is used in this section for calibration consists of two inflation option products, namely inflation index caps/floors and YoY inflation caps/floors. For both options market data is available for a whole range of strikes and maturities and prices are quoted in terms of base points (bp.). To compare calibration results, option prices are expressed here in terms of implied Black-Scholes volatilities.

Since YoY inflation caps/floors are essentially a series of YoY caplets/floorlets we perform a so-called stripping method, which is explained in [18, p. 682], to obtain the market data for YoY inflation caplets/floorlets. Obviously, performing a calibration to YoY caplets/floorlets instead of to YoY caps/floors reduces the computation time significantly.

2.4.1. CALIBRATING THE INTEREST RATE MODEL

For the calibration of the Euro nominal interest rate model we use the zero-coupon interest rate curve of September 30, 2010. The zero-coupon real interest rate curve as of September 30, 2010 is constructed using available information about zero-coupon break-even inflation as derived from index-linked swaps[11] (as of September 30, 2010).

We then obtain an estimate of the initial real zero-coupon curve by applying the Fisher equation

$$r_r(t) = \frac{1 + r_n(t)}{1 + bei(t)} - 1,$$

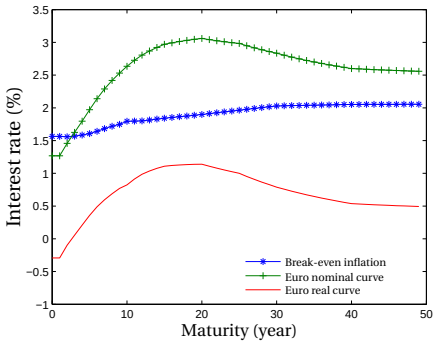
where bei denotes the break-even inflation, r_r the real interest rate and r_n the nominal interest rate. The resulting interest rate curves are shown in Figure 2.6a.

We calibrate the one-factor Hull-White model using market prices as of September 30, 2010 of forward-at-the-money options on Euro swap contracts (Euro swaptions). We calibrate the two parameters of the model, the mean-reversion and the volatility parameter, using a large set of swaptions, with option maturities ranging from 1 to 15 years and swap maturities ranging from 1 to 10 years. Swaptions with long maturities, > 15 years, and swap lengths, > 10 years, have deliberately been omitted from the calibration set. Liquidity for such contracts is often limited, which may result in not very representative market quotes. The optimal mean-reversion parameter is 0.0300; the optimal volatility parameter is 0.0089. A comparison between the model and market prices is shown in Figure 2.6b, where prices are expressed in terms of implied Black volatilities.

[11] The maturities of these swaps range from 1 to 50 years. We set the short break-even inflation equal to the 1-year break-even inflation. Missing maturities are approximated by linear interpolation.

Figure 2.6: Calibration results of interest rates.

(a) Overview of interest rate curves



(b) Quality of fit of the calibrated Hull-White model

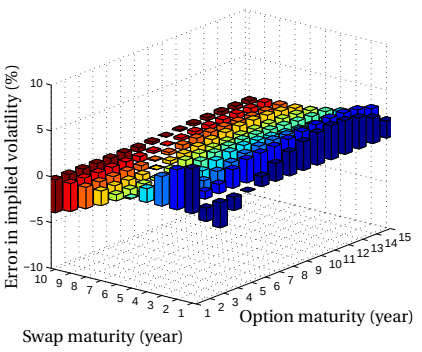


Figure 2.6b shows that the difference between model and market prices is very small. The average absolute error is 1.5% points. The fit is less accurate for short maturing options. This is due to the used objective function $\|C - \eta_C\|_2$ in our optimization procedure. Since the values of long maturing options are higher than the values of short maturing options the long maturing options automatically have a 'higher weight' in the optimization procedure. This can be overcome by introducing weights in the calibration procedure, however this refinement is outside the scope of the present work.

Option markets for real interest rates are still very limited. Therefore we set the mean-reversion parameter of the real interest rate model equal to the mean-reversion parameter of the nominal interest rate model.

Remark. *The choice of equal mean reversion parameters is justified when we estimate a Vasicek model (see [18, Chap. 3.2.1]) using a maximum likelihood estimation to historical nominal and real interest rates in the sample period* 1985 – 2009. *It turns out that the resulting mean reversion parameters are of the same order.*

The volatility parameter of the real interest rate model is determined by a scaling factor based on the volatility of historical nominal and real interest rates. The correlation parameter $\rho_{n,r}$ is also based on historical data (see Eq. (2.20)). The resulting parameters of the interest rate model are found to be:

$$\kappa_n = 0.0300,\ \kappa_r = 0.0300,\ \sigma_n = 0.0089,\ \sigma_r = 0.0084 \quad \text{and} \quad \rho_{\mathrm{n,r}} = 0.78.$$

2.4.2. CALIBRATION TO INFLATION MARKET DATA

The calibration of the inflation model can be performed using inflation market data. The specific inflation options were already explained in Section 2.3. To derive a reliable set of parameters, we use relevant liquid market data so that market conditions are captured well. We perform a calibration to YoY inflation caplets/floorlets. In the calibration routine the approximate model HHWi-YoY is applied.

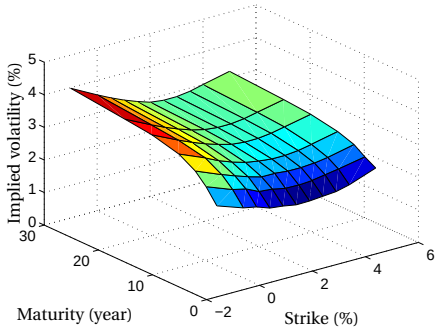
Note that a combined calibration to two different sets of inflation market data, namely to inflation index caps/floors and YoY inflation caplets/floorlets, can also be performed.

In such a calibration routine the approximate models HHWi-i and HHWi-YoY would be applied. It depends, however, on the 'problem at hand', which calibration is preferable. For example, when one is interested in pricing an out-of-the-money (inflation dependent) option, one should calibrate the inflation model to out-of-the-money options.

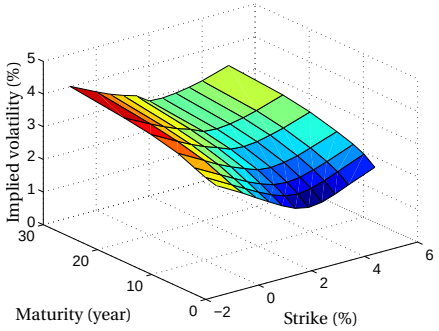
In Figures 2.7a and 2.7b the calibration results of the calibration to YoY inflation options are presented.

Figure 2.7: Quality of fit of the calibrated inflation model to YoY inflation options.

(a) Market implied volatilities (%)



(b) HHWi implied volatilities (%)



The calibration errors are very small; the average absolute error is 0.12% point and the maximum absolute error is 0.4% point, which indicates that the inflation model can be well calibrated to YoY inflation options.

The calibration results in the following model parameters:

$$\kappa_v = 0.095, \quad v(0) = 3.040 \cdot 10^{-4}, \quad \bar{v} = 2.401 \cdot 10^{-3}, \quad \sigma_v = 0.051,$$
$$\rho_{I,v} = -0.890, \quad \rho_{v,r} = 0.261 \text{ and } \rho_{v,\mathrm{n}} = -0.323.$$

Observing this parameter setting, we note that the Feller condition, $2\kappa_v \bar{v} > \sigma_v^2$, is not satisfied, hence,

$$\mathbb{P}\big(v(t) = 0 \mid t > 0\big) > 0.$$

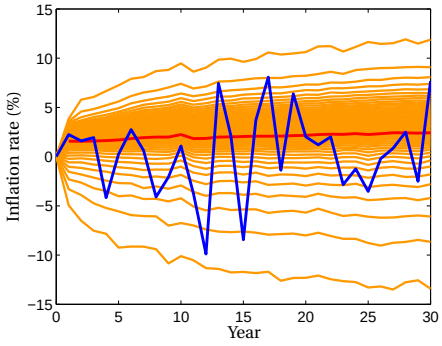
This implies that the variance process has a fat tailed distribution.

To get an impression of the results, 50 percentiles[12] for the inflation rate[13] and the volatility process are visualized in Figures 2.8a and 2.8b; the red line represents the average value over all scenarios and the blue line represents a randomly selected scenario.

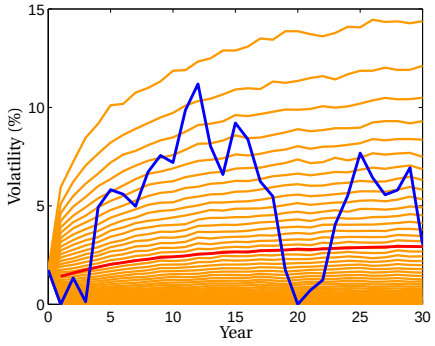
[12] We have used 10.000 scenarios, so that 200 scenarios are in between each lines in the scenario graphs.
[13] As already mentioned, the inflation rate is defined as the percentage change of the CPI.

Figure 2.8: Graphical impression of the generated risk neutral scenarios of the inflation rate and the volatility process for a horizon of 30 years.

(a) Inflation rate process



(b) Volatility process



The average inflation rate is approximately equal to the difference between the (average) nominal and real rates. The volatility of changes in the price inflation is high (1.6% in year 1 and 3.8% in year 30). As a result, the probability of negative inflation (deflation) is high (up to 20%). The fat tailed distribution of the volatility process is clearly visible in the percentile graphs.

2.4.3. MODEL COMPARISON: HESTON VS. SCHÖBEL-ZHU

As already mentioned, much attention has been devoted in the literature to stochastic volatility models driven by a Schöbel-Zhu (SZ) process (see for example [133]) in combination with stochastic interest rates to model the CPI. Therefore, we compare the SZ Hull-White inflation model (the SZHWi model) with our proposed model, the Heston Hull-White inflation (HHWi) model.

A summary of the differences between the Heston and the SZ model is listed below.

- Whereas in the Heston model the variance of the inflation is simulated, in the SZ model the volatility is simulated. The dynamics of the SZHWi model are given by:

$$\begin{cases} dI(t) = & (r_n(t) - r_r(t))I(t)dt + & \nu_{SZ}(t)I(t)dW^I(t), & I(0) \geq 0, \\ d\nu_{SZ}(t) = & \kappa_{\nu,SZ}(\bar{\nu}_{SZ} - \nu_{SZ}(t))dt + & \sigma_{\nu,SZ}dW^{\nu}_{SZ}(t), & \nu_{SZ}(0) \geq 0, \end{cases}$$

  where $\kappa_{\nu,SZ}$ is a mean-reversion parameter, $\sigma_{\nu,SZ}$ a volatility parameter, $\bar{\nu}_{SZ}$ denotes the long-term volatility level and $\nu_{SZ}(0)$ denotes the initial volatility level. The interest rate dynamics are given in Section 2.2.1.

- Since the volatility is conditional normally distributed, there is a positive probability of negative volatilities. Therefore, there is a positive probability that the sign of the instantaneous correlations with the Wiener process of the inflation process collapses in the simulation. This possibly leads to mispricing of (embedded) options.

- It turns out that there exists a direct relation between the Heston model and the SZ model[14]. More specifically, the SZ model is a specific Heston model, when the following parameter settings hold:

$$\kappa_v = 2\kappa_{v,SZ}, \qquad \sigma_v = 2\sigma_{v,SZ} \qquad \text{and} \qquad \bar{v} = \frac{\sigma^2_{v,\text{SZ}}}{2\kappa_{v,\text{SZ}}} = \frac{\sigma^2_v}{4\kappa_v}. \tag{2.21}$$

.

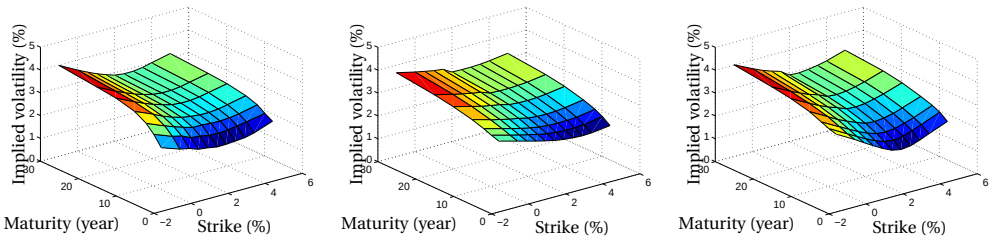
By using the Heston model to simulate the SZ model using the parameter specification in Eq. (2.21), we note that the Feller condition is always not satisfied, i.e.

$$\frac{2\kappa_v \bar{v}}{\sigma_v^2} = \frac{1}{2} < 1.$$

Using this relation between the Heston and the SZ model we calibrate the SZ model using the Heston model. In the Figures 2.9b and 2.9c calibration results are shown of the HHWi model and the SZHWi model.

Figure 2.9: Quality of fit of the HHWi and SZWHi model to YoY inflation options.

(a) Market implied volatilities (%) (b) SZHWi implied volatilities (%) (c) HHWi implied volatilities (%)



It turns out that by using the SZWHi model the average absolute error is 0.16% point and the maximum absolute error is 0.40% point, which indicates that (on average) the SZHWi model can also be well calibrated to YoY inflation options. However, after analysing the Figures 2.9b and 2.9c we observe that the HHWi model is better able to model the skew/smile effect in the market implied volatilties.

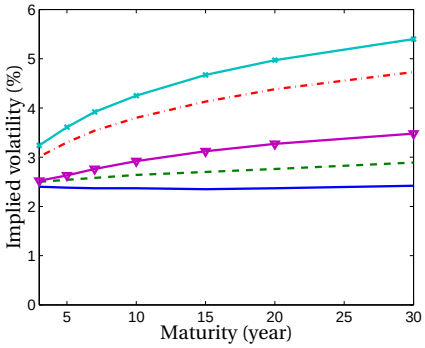
To gain more insight in the flexibility of modeling implied volatility skews/smiles by the HHWi and the SZHWi models, we again use the parameter relation given in Eq. (2.21). It turns out that the calibrated long term volatility of the SWHWi model is equal to 3.8%. Whereas in the case of the SZHWi model this parameter is fully determined by the mean reversion and volatility of variance parameter, this long term volatility can attain different values when using the HHWi model. By using different long term volatility parameters for the HHWi model, different types of implied volatility skews/smiles can be modeled. In this range of implied volatility skews/smiles only one implied volatility skew/smile is modeled by the SZHWi model, exactly when the parameter setting in Eq. (2.21) holds.

[14] See for more information [132].

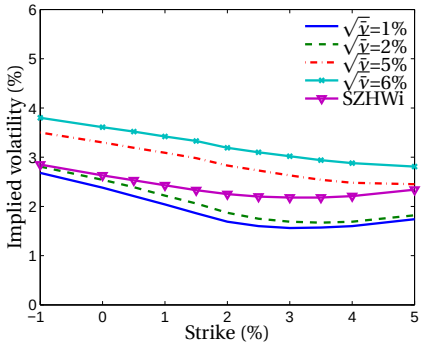
We perform a Monte Carlo experiment by using the calibrated SZHWi model and vary the long term volatility parameter in the HHWi model, i.e. this long term volatility is set to 1%, 2%, 5% and 6%. Using these parameter settings we value in Figure 2.10a a range of ATM YoY inflation options with different maturities and in Figure 2.10b we value a range of YoY inflation options with different strike levels all maturing after 5 years. All option prices are quoted in terms of implied volatilties.

Figure 2.10: Implied volatilities produced by the SHHWi model and different HHWi models.

(a) ATM IVs for different maturities



(b) 5-year IVs for different strike levels



The Figures 2.10a and 2.10b clearly show that the HHWi model is more flexible in modeling implied volatility skews/smiles.

2.5. VALUATION OF THE INDEXATION PROVISION OF A PENSION FUND

Risk neutral (RN) scenarios are mainly used for valuation purposes. Such special purpose scenarios can, for example, be used for a market-consistent valuation of premiums, benefits, and indexations of a pension fund (PF), to support strategic decision-making and provisioning. This valuation of premiums, benefits, and indexations is becoming increasingly important for risk management to assess the consequences of policy changes to the different stakeholders of a PF[15]. Furthermore, the valuation of indexations is important for hedging strategies. See for similar experiments, for example, [130].

We perform several MC simulations to obtain a value for the conditional indexations provision[16] of a PF. For this numerical experiment we use a stylized PF. The liabilities of this PF can be viewed as a general liability setting in the Netherlands. The initial funded ratio (FR) is equal to 110%. The PF makes use of a conditional indexation policy. Indexation is linear when the FR is between 105% and 115%; when the FR is below 105% pension rights are not indexed. We assume that the PF invests in three main investment

[15] Stakeholders of a PF are for example: pensioners, the sponsor and employees.
[16] The indexation provision of a PF can be viewed as an 'embedded option' on the balance sheet of the PF.

categories, 20% MSCI Europe stocks, 10% Euro direct real estate (RE) and 70% Euro government bonds.

Note that the inflation rate is the main driver of the initial indexation provision of a PF. We assume that indexation follows the price inflation for the inactive members of the PF and the wage inflation for the active members. In order to obtain the initial indexation provision we generate a consistent set of RN scenarios, so that all future indexation cash flows can be discounted with the nominal risk-free interest rate. The option price is then computed by:

$$\frac{1}{N}\sum_{k=1}^{N}\left(\sum_{t^*=t}^{T}\frac{M_{k,n}(t)}{M_{k,n}(t^*)}\widetilde{C}_k(t^*)\right),$$

where $t \le t^* \le T$, N denotes the number of scenarios, $\widetilde{C}_k(t^*)$ denotes the indexation cash flow in year t^* and scenario k, and $M_{k,n}$ denotes the nominal money-savings account (see Eq. (2.4)) in scenario k. We note that for this experiment we assume yearly time steps, i.e. $t^*, t, T \in \mathbb{N}$. In order to obtain an accurate option value, the number of scenarios N should be chosen as high as possible.

Since liquid inflation option market data only recently became available we take as the benchmark the fact that the price inflation model is calibrated to historical data. The historical volatility of the inflation rate is equal to 0.81%, which results in the following Heston parameters: $\kappa_\nu = 1$, $\nu(0) = \bar{\nu} = 0.46$ and $\sigma_\nu = 0$, as benchmark parameter setting. The full matrix of correlations is then also calibrated to historical data so that numerical inconsistencies are avoided.

Wage inflation, which is used for the (conditional) indexation of pension rights of active members, is modeled as price inflation plus 1% point. Furthermore, direct RE is modeled using a special purpose model, which is based on the Heston Hull-White model, where we explicitly model auto-correlation in the returns (see for more information [129]). The investment category MSCI Europe stocks is also modeled by a Heston Hull-White model and is, for simplicity, calibrated to the historical volatility. Furthermore, an appropriate underlying bond portfolio is used for the investment category government bonds.

We perform the following two numerical experiments:

Experiment I Since the indexation provision of the PF is based on the price and wage inflation, we apply several calibrations of our price inflation model[17]. Besides the inflation market data as of Q3 2010 we use a shifted set of market inflation option prices w.r.t. the market data of Q3 2010, using factors ±10% and 30%. The calibrated models are then used for a market consistent valuation of the indexation provision. The results are shown in Figure 2.11a.

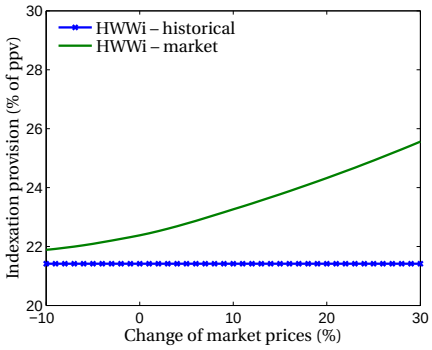
Experiment II To show the effect of different correlation parameters on the indexation provision, we perform a valuation of the indexation provision using different correlation values for $\rho_{I,n}$ $\rho_{I,r}$ and $\rho_{n,r}$. As a starting point we use the calibrated inflation model (see Section 2.4.2). The results are shown in Figure 2.11b.

[17] In Section 2.4.2 we have shown that our inflation model can be well calibrated to inflation option market data, so that market conditions are replicated well and, therefore, a realistic (market consistent) value of the indexation provision can be obtained.

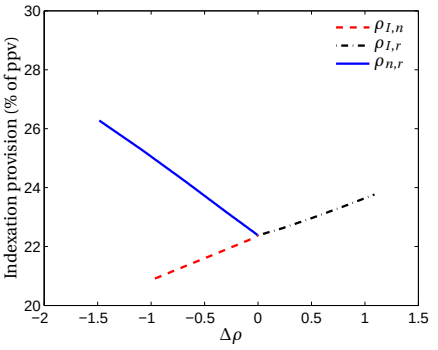
For the numerical experiment we use $N = 10.000$ scenarios[18], so that sufficiently accurate results are obtained (other specifics of the MC simulation can be found in Section 2.3.3). Since the horizon of the liabilities is long we use as a simulation horizon $T = 80$ years so that all indexation cash flows are included in the MC simulation.

Figure 2.11: Overview of numerical results. Option values are expressed in terms of the pension fund provision.

(a) Experiment I



(b) Experiment II



Observing Figure 2.11a we can conclude that calibrating the inflation model to inflation option market data results in different indexation provisions compared to the benchmark inflation models. The benchmark inflation models are insensitive to a change of the inflation option market prices, which justifies the usefulness of calibrating the inflation model to inflation option market data.

Observing Figure 2.11b we can conclude that changing the correlation parameters can have a significant effect on the indexation provision; especially when $\rho_{n,r}$ changes. When for example the correlation, $\rho_{n,r}$, changes from 0.78 to -0.7 then the indexation provision changes from 22.4% to 26.3%. Therefore, we can conclude that the indexation provision is influenced by the correlations, which confirms that all correlations should indeed be incorporated in a valuation model.

2.6. CONCLUSIONS

We derived an approximate closed-form solution of inflation indexed cap/ floor options and year on year inflation caplet/floorlet options, where the CPI follows a Heston model in which the nominal and real interest rates are modeled by one-factor Hull-White models. Using Fourier-based methods calibration can be done highly efficiently.

Using the developed models we performed a calibration of the inflation model to year-on-year inflation options. Our inflation model is able to model the market implied volatility skew accurately, so that market conditions are replicated well. Although the Schöbel-Zhu Hull-White inflation model can be well calibrated to inflation option market data, the proposed inflation model is better able to model the smile/skew effect in

[18] See Appendix 2.A for validation experiments.

the market implied volatilities. Furthermore, the proposed inflation model is more flexible in producing different sorts of implied volatility patterns.

Using the calibrated inflation model we performed a market consistent valuation of the conditional indexation provision of a stylized pension fund. It turns out that the results change significantly when performing a calibration to market inflation option data instead to historical data, so it is recommendable to use market data instead of historical data for valuation purposes. By changing the correlation parameters, indexation provisions may change significantly, which justifies the use of a full correlation matrix.

The focus in this chapter is on the modeling of inflation and the calibration to inflation options, since inflation is an important risk variable for insurance companies and pension funds. However, there are many other risk variables that are of importance. In this context we focus on the risk neutral modeling of direct real estate indices in the next chapter.

APPENDIX 2.A

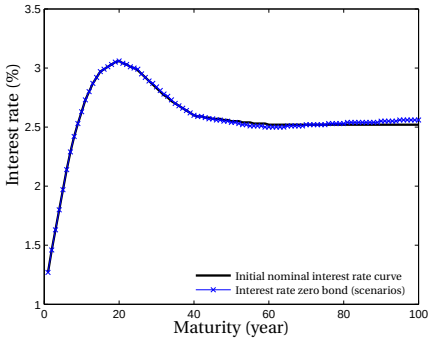
To illustrate the fact that the MC simulation (using 10.000 scenarios and a horizon of 100 years) of the HHWi model performed in Section 2.5 fulfills the martingale condition we perform two martingale tests. We first perform a (simple) MC experiment in which we price a series of zero-coupon bonds with different maturities. The payoff of this experiment is obviously equal to the principal of each bond for all scenarios. This payoff is then discounted back along the path of the short nominal interest rate for each scenario. The average discounted value (over all scenarios) then yields the MC price of each bond. This price can be converted into an equivalent interest rate for each maturity. If the generated scenario set is indeed arbitrage free, these interest rates should coincide with the initial nominal interest rate curve. The results are shown in Figure 2.12a.

As a second test, we price a series of index-linked zero-coupon bonds with different maturities. The principal of each bond is now indexed at the end of each year with the price inflation. The final payoff is then again discounted back along the path of the short nominal interest rate for each scenario. The average discounted value (over all scenarios) then yields the MC price of each index-linked bond. This price can subsequently be converted into an real interest rate for each maturity. If the generated scenario set is indeed arbitrage free, these interest rates should coincide with the initial real interest rate curve.

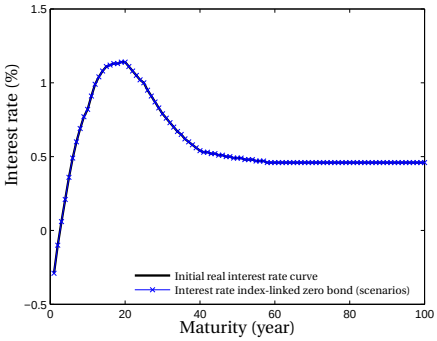
This comparison is made in Figure 2.12b. Note the perfect agreement between the real interest rates as implied by the scenario set and the initial real interest rate curve. This is due to the application of the empirical martingale simulation technique, which detects and corrects deviations from the desired martingale property.

Figure 2.12: Comparison between the nominal/real interest rates as determined by the scenario set and the initial nominal/real interest rate curve.

(a) Nominal interest rate



(b) Real interest rate



2

Observing Figure 2.12a a good agreement between the nominal interest rates as implied by the scenario set and the initial nominal curve is visible. This indicates that the interest rate scenarios are arbitrage free with respect to the initial nominal interest rate curve. The remaining differences will further diminish when a larger scenario set is used.

CHAPTER 3

Risk neutral valuation of real estate derivatives

We propose a risk neutral valuation model for real estate derivatives. We first model the underlying efficient market price of real estate and then construct the observed index value with an adaptation of the price update rule by [15]. We derive closed-form pricing solutions for forwards, swaps and European put and call options. We demonstrate the application of the model by valuing a put option on a house price index. Autocorrelation in the index returns appears to have a large impact on the option value. We also study the effect of an over- or undervalued real estate market. The observed effects are significant and as expected.

3.1. INTRODUCTION

Recently, the interest in real estate derivatives has surged. This interest has for instance been fueled by the introduction of real estate futures on the Chicago Mercantile Exchange (CME) in 2006. These futures give investors the opportunity to directly manage house price risk. Currently, trading is possible using 20 regional indices and two composite indices [11, 118]. The authors in [45, 46, 62] also provide an overview of other real estate derivatives markets, such as swap trading on the U.K. Investment Property Database index (IPD) or the U.S. NCREIF Property Index (NPI).

Currently, the most mature property derivatives market is the U.K. IPD derivatives market. At the end of 2008 some GBP 19.3 billion of swaps referenced IPD indices. In the beginning of 2009, trading in IPD derivatives has decreased significantly, however, mostly because fewer deals between banks were executed with Lehman Brothers exiting the market and several other banks cutting back on new business activities. The U.S. CME futures market does not yet have much liquidity, with only occasional trades. Property derivatives markets in France and Germany are also still very small.

We develop a risk neutral valuation model for real estate derivatives. Our main goal is to value derivatives that are coupled to private real estate indices with a significant degree of autocorrelation. It is well known from the real estate literature (see for example [63] for an overview) that autocorrelation can occur in appraisal-based indices because appraisers slowly update past prices with new market information. Transaction-based indices can exhibit a positive autocorrelation because private real estate markets are less informationally efficient than public securities markets. As a result, the price discovery and information aggregation functions of the private real estate market are less effective. This can cause noisy prices and inertia in asset values (and returns) (see [118, p. 4] and

This chapter is based on the article 'Risk Neutral Valuation of Real Estate Derivatives', published in *The Journal of Derivatives*, 23(1):89–110, 2015 [129].

the references given in that paper).

A significantly positive autocorrelation implies a (partial) predictability of future returns and opportunities for arbitrageurs. It is not possible, however, to trade the assets which constitute a private real estate index in a liquid market and at low costs. In practice, the index is thus not a tradable asset and arbitrage possibilities are very limited. This can also cause significant problems for suppliers of real estate derivatives since they cannot easily trade the underlying assets and (delta) hedge their positions. Nevertheless, derivatives markets for forward and swap contracts have been emerging in recent years.

The authors in [62] note, however, that a 2006 survey of (potential) market participants identified a lack of confidence in how real estate derivatives should be priced. They also note that this concern is understandable, since the underlying asset cannot be traded in a frictionless market. This makes it impossible to use classic pricing formulas for derivatives (such as the relationship between spot and forward prices), since these formulas only apply under the no-arbitrage assumption. The authors in [62] argue, however, that the valuation of real estate derivatives is still possible using equilibrium pricing rules, provided that the dynamic behavior of the underlying real estate index is properly taken into account. We take the next step by proposing a quantitative risk neutral valuation model, which can be used for actual pricing purposes.

A small body of related research exists in the equity option literature. The authors in [93] study the effect of predictability of asset returns in a continuous-time model. They propose an adjustment of the Black-Scholes pricing formula for stock options [14] to account for the effect of predictability. The paper [77] develops a discrete-time model to derive an analytic pricing formula for options on a stock index which exhibits positive correlation due to infrequent trading of the underlying stocks. There, it is assumed that the unobservable true liquidation value of the index follows a random walk process. The observed (autocorrelated) index is then modeled as the weighted average of current and past returns. More recently, the authors in [89] derived a closed-form formula for a European option on an asset with returns following a first-order moving average process. The paper [46] uses mean-reverting continuous-time models, that exhibit predictability for the drift term, for deriving closed form solutions of the main property derivatives traded in the financial markets.

The real estate literature also contains a few pioneering papers on risk neutral valuation. Early examples of risk neutral valuation techniques are given by [22, 25, 81]. The last paper is especially important, because it describes how a risk neutral valuation model can be used to value derivatives that are related to commercial real estate indices. Throughout their paper, these authors assume that the real estate index follows a random walk process with drift. By construction, such a process leads to uncorrelated index returns. The authors in [25] note, however, that their model can also be used in the case of autocorrelated indices provided that a proper transformation can be found to switch (back and forth) between the autocorrelated index (which is observed) and the uncorrelated variable (which is explicitly modeled).

A different approach is followed by [119] in their paper on home equity insurance. They first fit the observed real estate returns by a simple autoregressive (AR) model with one lag. Using this model, the conditional returns and volatilities can be determined analytically. The assumption is then made that options on the house price index can be

valued using an adaptation of the familiar Black-Scholes equation [14]. This adaptation consists of replacing the expected risk-free return with the expected real world return and the implied volatility by the estimated value from the AR model. One aspect of this approach is adopted by us, namely modeling the real estate returns with an AR model. We view the second, heuristic, step (in which the real world return is directly used as an input for a risk neutral valuation formula) as problematic, however.

Our approach circumvents this problem by following the approach of [77]. We thus explicitly model the (underlying) efficient market value of the real estate index and then construct the observed index. We argue that an adaptation of the price update rule proposed by [15] serves this purpose well. The first modification of this update rule is straightforward and consists of adding multiple lag terms. This leads to an AR model which can be estimated using standard econometric techniques. A second modification is more fundamental from a valuation perspective and consists of using the accrued value of past observations. Empirical results confirm that the volatility and autocorrelation of a (transaction-based) Dutch house price index can be replicated quite well (on an annual basis) with our model. As a second example, we consider monthly data for the U.S. 10 city S&P/Case-Shiller house price index. Our analysis shows that modeling seasonality and stochastic volatility is important for such monthly data.

The remainder of this chapter is organized as follows. In Section 3.2 we introduce our theoretical framework and analyze the properties of the real world and risk neutral process for real estate indices with autocorrelation. We also explain how the real estate model can be coupled to a stochastic interest rate model. Section 3.3 contains closed-form pricing formulas for forwards, swaps and European options. In Section 3.4 we estimate the real estate model using historical information for house price indices in the Netherlands and the U.S. Section 3.5 discusses the valuation of a European put option on a house price index using Monte Carlo simulation. We also assess the quality of the derived closed-form option pricing formula. Section 3.6 concludes.

3.2. MODELING FRAMEWORK

We present a risk neutral model that consists of a discrete-time model for the observed real estate index in combination with continuous-time models for the efficient market process of real estate and for interest rates. The current point in time is denoted as $t = 0$. Time is measured with respect to the period between two price updates of the real estate index. Unless stated otherwise (and without loss of generality), we assume that the time step between two price updates is equal to one year. Hence, $t = 1$ corresponds to one year ahead, $t = 2$ to two years ahead, etc. In the continuous-time models, the non-integer points in time are also sampled. To avoid confusion, we therefore denote the continuous-time variable with τ in the remainder of this section.

3.2.1. REAL WORLD PROCESS

PRICE UPDATE MODEL

We model the real world process of a real estate index by an adaptation of the price update rule proposed by [15][1]. They suggest that the new price is a weighted sum of the

[1] An earlier application of this update rule can be found in [21].

current market price and the last period's price. More precisely, they propose the following price update rule[2]:

$$a(t) = w_0\, y(t) + (1 - w_0)\, a(t-1), \tag{3.1}$$

where $a(t)$ is the current price, $a(t-1)$ is the previous price, $y(t)$ is the true market price and w_0 is a constant, $0 \leq w_0 \leq 1$. The parameter w_0 is commonly referred to as the confidence parameter. If w_0 is close to 1, the market price $y(t)$ is weighted heavily; if w_0 is small, the emphasis is more on the previous price $a(t-1)$. The simple price update rule in Eq. (3.1) is frequently used to model appraisal smoothing in real estate indices. See [63] for an extensive overview of research in this area. We show that this price update rule can also be used to describe the dynamic behavior of autocorrelated transaction-based indices.

The model in Eq. (3.1) is equivalent to an exponentially weighted moving average (EWMA) model, see [73, p. 479–480]. By substituting the expression for $a(t-1)$ in $a(t)$, the expression for $a(t-2)$ in $a(t-1)$, etc., we find that

$$a(t) = w_0 \sum_{i=1}^{m} (1 - w_0)^{i-1}\, y(t-i+1) + (1 - w_0)^m\, a(t-m), \tag{3.2}$$

where $1 \leq m \leq t$. This equation shows that the current value $a(t)$ partly consists of a basket of previous y-terms, where the weight of these terms decreases at an exponential speed (as controlled by the w_0 parameter). By setting m equal to t we also see that the weight of the index value at time 0, $a(0)$ is equal to $(1 - w_0)^t$ at time t.

It is important to note that the price update rule in Eq. (3.1) does not account properly for the time value of money because the previous value $a(t-1)$ is not accrued. From a valuation perspective this leads to a systematic underperformance of the real estate index. To correct for this effect, we adapt the price update rule in Eq. (3.1) and accrue the past index value with the expected (annual) return π:

$$a(t) = w_0\, y(t) + (1 - w_0)\,(1 + \pi)\, a(t-1). \tag{3.3}$$

Using accrued prices is common practice when appraisers set new prices for real estate objects. In this case, the reference price level is often formed by previous transactions for similar objects with a correction for the price increase (or decrease) of the real estate market up to the current point in time. The expected return is not modelled in detail at this point to keep the analysis as simple as possible. In practice the expected return may depend (positively) on the amount of risk associated with the real estate investment. In the risk neutral model the expected return is coupled directly to the level of the interest rate, as we explain in detail in Section 3.2.2.

Substitution of $a(t-1)$ in $a(t)$, $a(t-2)$ in $a(t-1)$, etc., again yields the EWMA form of Eq. (3.3):

$$a(t) = w_0 \sum_{i=1}^{m} (1 - w_0)^{i-1}\, y^*(t-i+1) + (1 - w_0)^m\, a^*(t-m), \tag{3.4}$$

[2]Equation (3.1) assumes that the price update rule is constant over time. Generalizations with time-varying parameters can be found in [20].

where $1 \leq m \leq t$ and

$$\begin{aligned} a^*(t-m) &\equiv a(t-m)(1+\pi)^m, \\ y^*(t-i+1) &\equiv y(t-i+1)(1+\pi)^{i-1}. \end{aligned} \tag{3.5}$$

Equation (3.4) is thus equivalent to Eq. (3.2) if we accrue past values.

We can easily determine the evolution of annual returns based on Eq. (3.3):

$$r^a(t) = w_0 \frac{y(t-1)}{a(t-1)} r^y(t) + (1-w_0) \frac{a(t-2)}{a(t-1)} (1+\pi) r^a(t-1), \tag{3.6}$$

where $r^a(t) \equiv a(t)/a(t-1)-1$ is the index return and $r^y(t) \equiv y(t)/y(t-1)-1$ is the unobserved return, both using annual compounding. A much simpler expression is derived when the index series are expressed in logarithms, see [63]. In this case, continuously compounded returns can be expressed as log differences:

$$r_c^a(t) = w_0^* r_c^y(t) + (1-w_0^*) r_c^a(t-1), \tag{3.7}$$

and thus

$$r_c^y(t) = \frac{1}{w_0^*} r_c^a(t) - \frac{1-w_0^*}{w_0^*} r_c^a(t-1), \tag{3.8}$$

where $r_c^a(t)$ is the index return and $r_c^y(t)$ is the unobserved market return, both using continuous compounding. The parameter w_0^* has a similar interpretation as the confidence parameter w_0. This parameter determines which fraction of the index return is explained by the unobserved market return (the remaining fraction is explained by the past index return). Note that the effect of accrual disappears when we take log differences (that is, when we assume that past values accrue with the same return π).

THE EFFICIENT MARKET PROCESS

We now assume that the underlying market returns follow a random walk process with drift:

$$r_c^y(t) = \pi + \epsilon(t), \tag{3.9}$$

where $\epsilon(t)$ is a normally-distributed, serially-uncorrelated noise term with zero mean and variance σ_ϵ^2. Note that the drift parameter π is assumed to be constant here to keep the analysis as simple as possible. However, in successive periods of appreciation and depreciation of the price levels this assumption is not always valid. A more appropriate specification would then be to allow π to change over time. An example of such a model is a local linear trend model, which is for example used in [58].

The confidence parameter w_0^* can be calculated from the first-order autoregressive (AR) process that we obtain by substituting Eq. (3.9) in Eq. (3.7):

$$r_c^a(t) = w_0^* \pi + (1-w_0^*) r_c^a(t-1) + w_0^* \epsilon(t). \tag{3.10}$$

The parameter w_0^* is thus equal to 1 minus the first-order autocorrelation of the index returns. In practice, we can also neglect the difference between w_0 in Eq. (3.6) and w_0^* in Eq. (3.7) because (on average) $y(t-1) \approx a(t-1)$ and $\frac{a(t-2)}{a(t-1)} \approx \frac{1}{1+\pi}$. Under these simplifying assumptions, the functional form of Eqs. (3.6) and (3.7) becomes the same. Annually

and continuously compounded returns also have almost the same first-order autocorrelation.[3] It thus follows that $w_0 \approx w_0^*$.

A word of caution is appropriate at this point. The assumption that the underlying market returns follow a random walk with drift is probably too strong for the private real estate market since these markets are less informationally efficient than public securities markets, see also [63]. We should therefore be careful not to directly equate the underlying random walk process with the true market process. A better interpretation would be to state that the observed index returns can be modeled using an underlying efficient market process in combination with the price update rule in Eq. (3.1). In the remainder of this section, we therefore refer to $y(t)$ as the efficient market price at time t. We will show in Section 3.2.3 that assuming an underlying efficient market process makes it possible to analyze the properties of the constructed real estate index with autocorrelation. This facilitates the derivation of several pricing formulas in Section 3.3.

PRICE UPDATE RULES WITH MULTIPLE LAGS

The price update rule in Eq. (3.1) can easily be extended with multiple lag terms. For the general case of p lags (with $p \geq 1$) we have:

$$a(t) = w_0 y(t) + \sum_{i=1}^{p} \omega_i a(t-i), \text{ where } w_0 + \sum_{i=1}^{p} \omega_i = 1. \tag{3.11}$$

The generalization of Eq. (3.10) then becomes:

$$r_c^a(t) = w_0^* \pi + \sum_{i=1}^{p} \omega_i^* r_c^a(t-i) + w_0^* \epsilon(t), \text{ where } w_0^* + \sum_{i=1}^{p} \omega_i^* = 1. \tag{3.12}$$

Equation (3.12) is an AR model of order p. To estimate the expected return π, the weights ω_i^* and the variance σ_ϵ^2 of an AR(p) model different approaches can be followed, see [98] or [123, p. 43–129] for detailed overviews. Note that the restriction that the sum of the weights should be equal to one does not complicate the estimation of the model since both w_0^* and π are free parameters.

A simple approach is to estimate the model parameters with an ordinary least squares (OLS) regression method. This method basically minimizes the one-step-ahead prediction errors. An alternative approach is to choose the weights in such a way that the autocovariance function of the AR process is exactly equal to the autocovariance function of the observed real estate index. This correspondence can be achieved by using the Yule-Walker equations, see [123, p. 46].

To decide which model is most appropriate several order selection criteria have been proposed in the literature, see for example [123, p. 82–84]. These selection criteria typically choose the model order in such a way that the prediction error is minimized while putting a penalty on the number of parameters estimated. The estimation of the real estate model is discussed in detail in Section 3.4.[4]

[3] We can, at any given point in time, convert a continuously compounded return $r_c^a(t)$ into an annually compounded return $r^a(t)$ with the relation $r^a(t) = \exp(r_c^a(t)) - 1$. Let us now consider a given time series for $r_c^a(t)$. We can convert these returns into annually compounded returns (using this relation).The temporal correlation between the two time series is almost the same since changes in $r^a(t)$ and $r_c^a(t)$ are in first-order approximation the same.

[4] A natural extension of the price update model is to model the constant w_0 by a time-dependent function, so

SEASONALITY

Seasonality in real estate returns can become important when modeling quarterly or monthly returns. Let us assume that we have already modeled the seasonally adjusted index $a(t)$ using Eq. (3.1) or (3.11). We can then add a seasonal component $g(t)$ to obtain the index value with seasonality, $\tilde{a}(t)$:

$$\log(\tilde{a}(t)) = \log(a(t)) + g(t), \tag{3.13}$$

or, equivalently:

$$\tilde{a}(t) = a(t)\exp(g(t)). \tag{3.14}$$

Different approaches can be used to estimate the $g(t)$ function. A natural assumption is to assume that seasonality does not have a net effect on an annual basis. For simplicity, one can also assume that the seasonal pattern is constant over time. Given these assumptions one could then use so-called dummy variables in the OLS regression. These dummy variables are equal to one for the respective periods. For example, a January dummy is equal to one for all January (log) returns and zero for all other months; a February dummy is equal to one for all February returns and zero otherwise, etc. The $g(t)$ function is then easily constructed using the estimated weights of the dummy variables. Another (even simpler) method consists of detrending the log index and then fitting a (shifted) sine function with a period of one year to the data. The first approach (i.e., a regression on monthly dummies) is used in Section 3.4.4.

3.2.2. RISK NEUTRAL PROCESS

PROCESS FOR INTEREST RATES

We model the evolution of the short interest rate by the familiar one-factor Hull-White (HW) model, see [73, p. 688–689]. Within the large family of interest rate models, the HW model is a typical example of a no-arbitrage model. Such a model produces interest-rate scenarios that are consistent with the current term structure. This no-arbitrage feature is extremely important for option pricing applications, since a small error in the underlying bond prices can cause large errors in the price of interest-rate options, see [73, p. 686].

Technically speaking, the one-factor HW model assumes that the risk neutral process for the nominal short rate r is as follows:

$$dr(\tau) = \big(\theta(\tau) - \kappa_r r(\tau)\big)\,d\tau + \sigma_r\, dW^r(\tau). \tag{3.15}$$

We denote time in this equation by the symbol τ to indicate that we now use a continuous-time model. This model assumes that the short interest rate fluctuates around the mean reversion level $\theta(\tau)/\kappa_r$. The parameter κ_r controls the amount of mean-reversion. The θ-function is deterministic and chosen in such a way that the model satisfies the no-arbitrage constraint. The one-factor HW model is in fact an extension of the model in [134] in the sense that the mean reversion level is time-dependent instead of constant. Parameter σ_r controls the volatility of the Wiener process W^r.

that the constant w_0 can change with market circumstances. This extension is, however, not investigated in this section.

PRICE UPDATE MODEL

We now derive the risk neutral process for real estate indices with autocorrelation. The risk neutral process for the evolution of the index value can be derived analogously to Eq. (3.3):

$$a(t) = w_0 y(t) + (1-w_0)\pi(t)a(t-1),$$
$$\text{where } \pi(t) \equiv \exp\left(\int_{t-1}^{t} r(\tau)d\tau\right)\exp(-q). \tag{3.16}$$

The expected return π is thus a time-dependent function in a risk neutral world and depends on the level of the (short) interest rate and the direct return. More precisely, the term $\exp\left(\int_{t-1}^{t} r(\tau)d\tau\right)$ is the risk-free return on a bank account between time $t-1$ and t. The term $\exp\left(-q\right)$ is a correction for the direct return q associated with real estate investments. By setting q equal to zero a total return index is modeled.

Substitution of $a(t-1)$ in $a(t)$, $a(t-2)$ in $a(t-1)$, etc., again yields the following EWMA form of Eq. (3.16):

$$a(t) = w_0 \sum_{i=1}^{m} (1-w_0)^{i-1} y^*(t-i+1) + (1-w_0)^m a^*(t-m), \tag{3.17}$$

where $1 \le m \le t$ and

$$\begin{aligned} a^*(t-m) &\equiv a(t-m)\exp\left(\int_{t-m}^{t} r(\tau)d\tau\right)\exp\left(-qm\right), \\ y^*(t-i+1) &\equiv y(t-i+1)\exp\left(\int_{t-i+1}^{t} r(\tau)d\tau\right)\exp\left(-q(i-1)\right). \end{aligned} \tag{3.18}$$

Equation (3.16) can be extended for the general price update model with p lag terms:

$$a(t) = w_0 y(t) + \sum_{i=1}^{p} \omega_i a^*(t-i) \text{ where } w_0 + \sum_{i=1}^{p} \omega_i = 1. \tag{3.19}$$

The EWMA form of Eq. (3.19) can also be derived. Let us assume that $T > t \ge 0$. We can now substitute $a^*(T-1)$ in $a(T)$, $a^*(T-2)$ in $a^*(T-1)$, etc., until an expression is obtained with only the terms $y^*(t+1), \cdots, y^*(T)$ and $a^*(t), \cdots, a^*(t-p+1)$:

$$a(T) = \sum_{i=1}^{T-t} c_i y^*(t+i) + \sum_{i=1}^{p} d_i a^*(t-i+1), \tag{3.20}$$

where $c_{T-t} = w_0$. Explicit expressions for the c_i and d_i coefficients of this equation can be determined using a software package which is able to perform symbolic algebra calculations.

THE EFFICIENT MARKET PROCESS

We also need to specify the risk neutral process for the underlying efficient market price. Analogously to Eq. (3.9), we use a random walk process with drift (geometric Brownian motion):

$$dy(\tau) = (r(\tau)-q)y(\tau)d\tau + \sigma_y y(\tau)dW^y(\tau), \tag{3.21}$$

where the volatility σ_y is constant and W^y follows a Wiener process. By means of Ito's lemma it can be shown that $\log y(\tau)$ is governed by the following process, see [73, p. 270–271]:

$$d\log y(\tau) = \left(r(\tau) - q - \sigma_y^2/2\right)d\tau + \sigma_y dW^y(\tau). \tag{3.22}$$

For numerical reasons Eq. (3.22) is commonly used in practice instead of Eq. (3.21). Note that these equations are equivalent to the Black-Scholes price process [14] for a dividend paying stock in case of stochastic interest rates.

MODEL EXTENSIONS: REAL INTEREST RATES, INFLATION, STOCHASTIC VOLATILITY

It is also possible to model real interest rates and inflation in a consistent way. [18, p. 646–647] for example develop a consistent risk neutral model for nominal and real interest rates as well as the CPI index (see also Chapter 2). To keep the analysis as simple as possible, we do not consider such an extended model. Including inflation may be very important for practical applications, however, since real estate cash flows (like rental income or maintenance costs) are often inflation-linked.

Another extension consists of modeling stochastic volatility. This is especially important when considering high-frequency data, like monthly or quarterly returns. An example is given in Section 3.4.4 for monthly U.S. house price data. A quite general stochastic volatility model is the constant elasticity of variance (CEV) model:

$$\begin{cases} dy(\tau) = & (r(\tau) - q)y(\tau)d\tau + \sqrt{v(\tau)}y(\tau)dW^y(\tau), \\ dv(\tau) = & \kappa_v(\bar{v} - v(\tau))dt + \sigma_v v(\tau)^{\beta_v} dW^v(\tau). \end{cases}$$

This model is a natural extension of the geometric Brownian motion in Eq. (3.21) and has for example been studied by [78]. The κ_v parameter controls the speed of mean reversion of the variance $v(\tau)$. The $\bar{v}$ parameter denotes the long variance level and σ_v controls the volatility of the variance process. The initial variance should, of course, also be specified as a boundary condition. The elasticity parameter β_v must satisfy $0.5 \leq \beta_v \leq 1.0$ to retain the uniqueness of option prices. Both limiting cases are in fact well-known stochastic volatility models. For $\beta_v = 1/2$ we have the model of [70] and for $\beta_v = 1$ we have the continuous-time GARCH model as in [102]. Maximum likelihood estimation of the CEV model parameters, based on option prices, is discussed in [2].

3.2.3. MARTINGALE PROPERTIES

THE EFFICIENT MARKET PROCESS

If there are no arbitrage opportunities, the expected price of a traded security has to increase in the same way as a bank account in a risk neutral world, see [73, p. 630]. To verify this no-arbitrage restriction, we consider the realization of the efficient market price $y(t)$ and the nominal bank account $M(t) \equiv M(0)\exp\left(\int_0^t r(\tau)d\tau\right)$ up to time t and determine the expected value of the ratio $y(T)/M(T)$ for $T > t \geq 0$. Let us first consider the situation where all direct returns are reinvested in the index (i.e., we have a total return index). This situation can also be modeled by setting the direct return q equal to zero. We then have that (see Appendix 3.A for the proof):

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{y(T)}{M(T)}\middle|\mathscr{F}_t\right] = \frac{y(t)}{M(t)}, \tag{3.23}$$

where $\mathbb{E}^{\mathbb{Q}}\left[y(T)/M(T)\,\middle|\,\mathscr{F}_t\right]$ means that the expected value of $y(T))/M(T)$ in a risk neutral world and conditional on the filtration up to time t is considered. The expected value of $y(T)/M(T)$ is thus constant for $T > t \geq 0$. That is, this ratio is a zero-drift (martingale) process. A total return index thus satisfies the martingale requirement for traded securities if its dynamics is governed by Eq. (3.22).

Note that a price index with $q > 0$ is not a tradable asset, comparable to the situation for an index of dividend-paying stocks. Consequently, the martingale property is not satisfied by a price index if $q > 0$. In this case:

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{y(T)}{M(T)}\,\middle|\,\mathscr{F}_t\right] = \frac{y(t)}{M(t)}\exp\left(-q(T-t)\right), \tag{3.24}$$

see again Appendix 3.A. A price index is thus not a tradable asset if direct returns are paid out.

THE REAL ESTATE INDEX PROCESS

We now consider the realization of the real estate index $a(t)$ up to time t and determine the expected value of the ratio $a(T)/M(T)$ for $T > t \geq 0$. In Appendix 3.B we also prove that

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{a(T)}{M(T)}\,\middle|\,\mathscr{F}_t\right] = \frac{\exp\left(-q(T-t)\right)}{M(t)}\left[y(t)(1-\alpha_{w_0,T}(t)) + a(t)\alpha_{w_0,T}(t)\right], \tag{3.25}$$

where $\alpha_{w_0,T}(t) \equiv (1-w_0)^{T-t}$. $a(T)/M(T)$ is thus a martingale if $a(t) = y(t)$ and $q = 0$. Since $a(t) = y(t)$ holds in general only when $w_0 = 1$, Eq. (3.16) does not represent (the risk neutral process of) a tradable asset when $w_0 < 1$. Arbitrage opportunities would thus exist in case of a complete market when trading an autocorrelated real estate index. The reverse argument also holds: the index value may well be different from the efficient market price, but active trading in the index is not possible in this case: otherwise arbitrageurs would quickly force the index value toward the efficient market price.

Another important observation is that the future development of a total return real estate index with autocorrelation (i.e., $q = 0$ and $w_0 < 1$) is unbiased if the index is in equilibrium at time t (i.e., when $a(t) = y(t)$). With unbiased we here mean that $\mathbb{E}^{\mathbb{Q}}[a(T)/M(T)] = a(t)/M(t)$ for $T > t$. Stated otherwise, if the real estate index starts from an equilibrium situation, the expected return is in line with the return on a risk-free bank account. As a consequence, the pricing formulas for linear instruments (forwards and swaps) all collapse to the classic no-arbitrage formulas if a total return real estate index is in equilibrium at the valuation date. This will be proved more formally in the next section.

A generalization of Eq. (3.25) also exists for the price update model with more than one lag term, as specified in Eq. (3.19). Using the same procedure as in Appendix 3.B, the counterpart of Eq. (3.25) follows:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\frac{a(T)}{M(T)}\,\middle|\,\mathscr{F}_t\right] &= \frac{\exp\left(-q(T-t)\right)}{M(t)}\left[y(t)\sum_{i=1}^{T-t}c_i + \sum_{i=1}^{p}d_i\,\hat{a}(t-i+1)\right],\\ \text{where } \hat{a}(t-i+1) &\equiv a(t-i+1)\exp\left(\int_{t-i+1}^{t} r(\tau)\,d\tau\right)\exp\left(-q(i-1)\right).\end{aligned}$$

Incorporating seasonality is also straightforward (see Eq. (3.14)):

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{\tilde{a}(T)}{M(T)}\,|\mathscr{F}_t\right] = \exp(g(T))\mathbb{E}^{\mathbb{Q}}\left[\frac{a(T)}{M(T)}\,|\mathscr{F}_t\right], \tag{3.26}$$

where $\tilde{a}(t)$ is the index value with seasonality. We developed a risk neutral model for autocorrelated real estate indices, which can be coupled to existing risk neutral models for interest rates, inflation, stochastic volatility, etc. By studying the martingale properties of the real estate index we find that the no-arbitrage restriction is only satisfied under very specific conditions (i.e., for total return indices without autocorrelation). In general, arbitrage possibilities thus exist. These cannot be exploited easily, however, since the underlying index cannot be traded actively. We will use the derived results in the next section to derive pricing formulas for various real estate derivatives.

3.3. PRICING FORMULAS

We derive pricing formulas for derivatives that are linked to autocorrelated real estate indices. We first present results for the simple price update model with one lag term and then for a model with multiple lags.

3.3.1. FORWARDS

We can easily determine the price of a forward contract on a real estate index. Let us assume that the forward contract expires at time $T > t$ and that the agreed-upon delivery price is $F_T(t)$. For the owner of the forward contract, the payoff at time T is then equal to the difference between $F_T(t)$ and the index $a(T)$. If we denote the price of this contract at time t as $\eta_f(t)$ (f denotes the market price), we have that:

$$\eta_f(t) = M(t)\mathbb{E}^{\mathbb{Q}}\left[\frac{F_T(t)-a(T)}{M(T)}\,|\mathscr{F}_t\right]. \tag{3.27}$$

Using Eq. (3.25) we then arrive at the following result:

$$\eta_f(t) = P(t,T)F_T(t) - \exp\big(-q(T-t)\big)\left[y(t)(1-\alpha_{w_0,T}(t)) + a(t)\alpha_{w_0,T}(t)\right],$$

where $P(t,T)$ denotes the price at time t of a zero-coupon bond which matures at time $T > t$. The scaling factor $\alpha_{w_0,T}(t)$ is equal to $(1-w_0)^{T-t}$. The price of this forward contract is equal to zero if

$$F_T(t) = \frac{\exp\big(-q(T-t)\big)}{P(t,T)}\left[y(t)(1-\alpha_{w_0,T}(t)) + a(t)\alpha_{w_0,T}(t)\right] \;\text{ if } \eta_f(t) = 0. \tag{3.28}$$

When the index value $a(t)$ is also equal to the efficient market price $y(t)$ we obtain the classic relationship between the (spot) value of the index and the forward price:

$$F_T(t) = \frac{\exp\big(-q(T-t)\big)}{P(t,T)}a(t) \;\text{ if } y(t) = a(t) \;\text{ and } \eta_f(t) = 0.$$

The above analysis is easily extended to more general price update models with seasonality and multiple lag terms. By substituting Eq. (3.26) in Eq. (3.27) we arrive at:

$$\tilde{F}_T(t) = \frac{\exp\big(g(T)\big)\exp\big(-q(T-t)\big)}{P(t,T)}\left[y(t)\sum_{i=1}^{T-t}c_i + \sum_{i=1}^{p}d_i\hat{a}(t-i+1)\right],$$

with $\tilde{F}_T$ the forward price including the seasonal component.

The authors of [62] note that the forward market can signal that the real estate market is over- or undervalued. Equation (3.28) makes this price discovery function of the forward market explicit: using an estimate for the confidence parameter w_0, together with the actual forward price $F(T)$ and the index value $a(t)$, the underlying efficient market price $y(t)$ can be derived. The accuracy of the extracted efficient market price is of course strongly depending on the degree of liquidity and density in the forward market. A reliable price reporting system is also crucial. The authors of [62] mention that the U.K. IPD swap market appears to be performing the price discovery function well since IPD swap prices have fallen dramatically in 2006, even as the IPD index itself has continued to climb. The IPD swap market has thus correctly signaled overvaluation in the U.K. property market.

If there is a liquid forward market it also becomes possible to (delta) hedge movements of the underlying efficient market price. For example, if we calculate the sensitivity $\partial F_T(t)/\partial y(t)$ using Eq. (3.28) we find that:

$$\partial F_T(t)/\partial y(t) = \frac{\exp\big(-q(T-t)\big)}{P(t,T)}[1-\alpha_{w_0,T}(t)+w_0\alpha_{w_0,T}(t)],$$

where we have used Eq. (3.16) to determine $\partial a(t)/\partial y(t)$. Changes of the efficient market price $y(t)$ are thus directly reflected in changes of the forward price $F_T(t)$. This is an important result because we implicitly assumed (see Section 3.2.2) that continuous trading in the underlying efficient market index is possible. This, obviously, cannot be achieved by trading in the primary real estate market (due to a limited liquidity and high trading costs). Using forward contracts it however becomes possible to replicate the efficient market process in accurate approximation, provided this secondary market is sufficiently liquid. This also provides a method to replicate the cash flows of more complicated real estate derivatives (like options) using delta hedging. A liquid forward market would thus serve as the foundation of the risk neutral valuation method developed in this section.

We should also note that in case of stochastic interest rates forward and futures prices are not equal. This is caused by the daily settlement procedure for futures contracts. Assume for instance that the real estate index is strongly positively correlated with interest rates. When the real estate index increases, the gain of a long futures contract is invested with a high probability at an above average interest rate. The opposite holds when the real estate index drops and the resulting loss probably needs to be financed at a below average interest rate. It thus follows that in case of a positive correlation between the real estate index and interest rates a long futures contract will be more attractive than a long forward contract. Other factors may also cause significant differences between forward and futures contracts (like taxes and transaction costs).

3.3.2. SWAPS

We assume that the swap contract starts at time $T_0 \geq t$ and ends at time $T_n > T_0$. To fix the notation: the owner of a receiver swap receives the price return of the real estate index in each period and pays the floating rate. The floating payments are based on the index values at the beginning of each period. The floating rate can, for example, be

the LIBOR spot rate. We also assume (without loss of generality) that the cash flows are swapped annually. Results for a total return index can be obtained by setting q equal to zero in the equations below. We denote the market price of a swap by Π and the market price of a swaplet by Π_k, where $1 \leq k \leq n$.

We first determine the model value of the swaplet which is active during the time interval $[T_{k-1}, T_k]$, where $1 \leq k \leq n$. If we denote the model price of this swaplet by $\eta_{\Pi_k}(t)$, we can use the following result by [12][5]:

$$\eta_{\Pi_k}(t) = LM(t)\mathbb{E}^{\mathbb{Q}}\left[\frac{a(T_k)}{M(T_k)} - \frac{a(T_{k-1})}{M(T_{k-1})}\Big|\mathscr{F}_t\right],$$

where L is a scaling parameter which can be used to set the notional amount of the swap to the right amount[6]. The total value of the swap, $\eta_\Pi(t)$, is thus equal to:

$$\eta_\Pi(t) = \sum_{k=1}^{n}\eta_{\Pi_k}(t) = LM(t)\mathbb{E}^{\mathbb{Q}}\left[\frac{a(T_n)}{M(T_n)} - \frac{a(T_0)}{M(T_0)}\Big|\mathscr{F}_t\right]. \tag{3.29}$$

Substituting Eq. (3.25) and rearranging terms, we find that

$$\begin{aligned}\eta_\Pi(t) &= L\exp\big(-q(T_N-t)\big)\big[y(t)(1-\alpha_{w_0,T_N}(t)) + a(t)\alpha_{w_0,T_N}(t)\big] \\ &- L\exp\big(-q(T_0-t)\big)\big[y(t)(1-\alpha_{w_0,T_0}(t)) + a(t)\alpha_{w_0,T_0}(t)\big].\end{aligned} \tag{3.30}$$

When $a(t) = y(t)$ and $q = 0$ the value of the swap contract is equal to zero. The same holds if $w_0 = 1$ and $q = 0$. This is in line with the result obtained by [12], which proves that the value of a total return real estate swap is exactly equal to zero if the real estate index follows a random walk process with drift[7].

Results for the general model with multiple lags and seasonality can easily be derived by substituting Eq. (3.26) in Eq. (3.29). If the swap market is sufficiently liquid it also becomes possible to (delta) hedge movements of the underlying efficient market price. For example, if we calculate the sensitivity $\partial\eta_{\Pi_T}(t)/\partial y(t)$ using Eqs. (3.30) and (3.16), we find that:

$$\begin{aligned}\partial\eta_\Pi(t)/\partial y(t) = \quad & L\exp\big(-q(T_N-t)\big)\big[1-\alpha_{w_0,T_N}(t) + w_0\alpha_{w_0,T_N}(t)\big] \\ & -L\exp\big(-q(T_0-t)\big)\big[1-\alpha_{w_0,T_0}(t) + w_0\alpha_{w_0,T_0}(t)\big].\end{aligned}$$

A liquid swap market can thus also be used to delta hedge more complicated derivatives. If there is no access to either a liquid forward or swap market the risk neutral valuation approach cannot be applied. One should then resort to methods developed for pricing in incomplete markets. A good example of this approach is given in [126]. They study the effect of market frictions (like transaction costs, transaction time, and short sale constraints) to explain why property returns are swapped against a rate that can deviate significantly from LIBOR.

[5] We do not assume (as is the case in [12]) that direct returns are reinvested. This situation is, however, easily obtained by setting q equal to zero in the derived equations.

[6] The notional is thus not a fixed amount but is adjusted periodically by the appreciation and depreciation of the index, see the description in [25, p. 22].

[7] The authors in [106] have extended this result by considering the effect of counterparty default risk. They find that the swap price is no longer equal to zero in this case because a compensation for the additional risk is required.

3.3.3. EUROPEAN OPTIONS

We value the option at time t. We consider a European option which expires at time $T > t \geq 0$ and cannot be exercised before that date. If $w_0 = 1$, an exact pricing formula exists. This formula is a modification of the familiar equation in [13]. The crucial modification is an adjustment of the implied volatility parameter to account for the effect of stochastic interest rates. This adjusted volatility, denoted by $\sigma_y^*(t,T)$, can be calculated as follows, see [18, p. 888]:

$$\begin{aligned}\sigma_y^{*2}(t,T) &= \frac{\sigma_r^2}{\kappa_r^2}\left(T-t+\frac{2}{\kappa_r}exp(-\kappa_r(T-t))-\frac{1}{2\kappa_r}exp(-2\kappa_r(T-t))-\frac{3}{2\kappa_r}\right)\\ &\quad +{\sigma_y}^2(T-t)+2\rho\frac{\sigma_r\sigma_y}{\kappa_r}\left(T-t-\frac{1}{\kappa_r}(1-exp(-\kappa_r(T-t))\right),\end{aligned}$$

with ρ the correlation between the Wiener processes for the short interest rate (see Eq. (3.15)) and the efficient market process (see Eq. (3.22)).

If $w_0 < 1$, a simple, approximating pricing formula can be derived. Equation (3.17) shows that the index value at time T is equal to a weighted sum of $T-t$ log-normal distributions. We thus have an Asian basket option. To value this option, we first calculate the first moment μ_1 and the second moment μ_2 of the exact probability distribution at time T, see [73, p. 578–579]:

$$\begin{aligned}\mu_1 &= \mu_{1,0}+\sum_{i=1}^{T-t}\mu_{1,i},\\ \mu_{1,0} &= \exp\left((r_{N,T}(t)-q)(T-t)\right)a(t)(1-w_0)^{(T-t)},\\ \mu_{1,i} &= \exp\left((r_{N,T}(t)-q)(T-t)\right)y(t)w_0(1-w_0)^{i-1},\end{aligned} \tag{3.31}$$

and

$$\mu_2=\sum_{i=1}^{T-t}\mu_{1,i}^2\exp\left(\sigma_y^{*2}(t,t+i)\right)+\sum_{i<j}\mu_{1,i}\mu_{1,j}\exp\left(\sigma_y^{*2}(t,t+i)\right). \tag{3.32}$$

Using these moments, we can fit the exact distribution with an approximating log-normal distribution. This approach was first proposed by [87]. The forward price $F_T(t)$ and the implied volatility σ can then be approximated using the following equation, see also [73, p. 565]:

$$F_T(t)=\mu_1,\quad \sigma=\sqrt{\frac{1}{T-t}\log\left(\frac{\mu_2}{\mu_1}\right)}.$$

Closed-form pricing formulas for European put and call options are then given by the familiar price for an option [13] on a forward contract:

$$\begin{aligned}\eta_p(t) &= \exp\left(-r_{N,T}(t)(T-t)\right)\left[KN(-d_2)-F_T(t)N(-d_1)\right],\\ \eta_c(t) &= \exp\left(-r_{N,T}(t)(T-t)\right)\left[F_T(t)N(d_1)-KN(d_2)\right],\end{aligned}$$

where

$$d_1=\frac{\log(F_T(t)/K)+\sigma^2(T-t)/2}{\sigma\sqrt{T-t}},\ d_2=d_1-\sigma\sqrt{T-t}. \tag{3.33}$$

$\eta_p(t)$ here denotes the price of a put option, $\eta_c(t)$ the price of a call option and K the strike price. The market prices of a put and call option are denotes by p and c, respectively.

Equations (3.32)-(3.33) are also valid for models with multiple lag terms and seasonality. Equation (3.31) should be generalized, however, in this case. The proper form is:

$$\begin{aligned}
\mu_1 &= \mu_{1,0} + \sum_{i=1}^{T-t} \mu_{1,i}, \\
\mu_{1,0} &= \exp(g(T)) \exp\big((r_{N,T}(t) - q)(T-t)\big) \sum_{i=1}^{p} d_i \hat{a}(t-i+1), \\
\mu_{1,i} &= \exp(g(T)) \exp\big((r_{N,T}(t) - q)(T-t)\big) y(t) c_i.
\end{aligned}$$

The accuracy of this option pricing formula is tested in Section 3.5.2. It is important to note that alternative pricing techniques for Asian options are available in the literature. The paper [95] provides an overview of the current state-of-the-art in this field. We recommend these techniques, however, when option pricing with a very high accuracy is required. We also note that an accurate valuation of American options is possible using the least squares Monte Carlo method by [94].

The accuracy of the proposed valuation model is of course strongly depending on the ability of the valuation model to properly capture the dynamic behavior of the real estate index. Model selection and estimation issues are therefore discussed in detail in the next section.

3.4. ESTIMATION OF THE MODEL

Our real estate model could be calibrated (at least partly) using market data if a liquid real estate option market would exist. The standard approach for equity option models is for example to fit the model parameters as good as possible to prices of equity options with different maturities and strike levels (see also the discussion in Chapter 2). The market-implied volatilities (and sometimes also the market-implied correlations) are thus the key inputs for the model estimation. This approach is currently not feasible for real estate, however, due to a lack of trading in real estate options. We thus have to resort to an estimation of the model using historical index data.

We therefore base our analysis on historical data for the Dutch and U.S. residential markets. We first provide an overview of the data in Section 3.4.1. Calibration results for the Dutch and U.S. interest rate models are then presented in Section 3.4.2. We subsequently present calibration results for the Dutch and U.S. real estate models in Section 3.4.3 and 3.4.4.

We mainly focus on annual historical data for the Dutch real estate market and monthly historical data for the U.S. real estate model. Note that it would of course be possible to do everything at the monthly interval. This would, however, limit the accessibility of this section since the analysis at the monthly interval requires much more advanced model estimation techniques and does not directly lead to analytical prices for real estate derivatives (due to seasonality and stochastic volatility which become important at the monthly interval).

3.4.1. DESCRIPTION OF THE DATA AND MODEL ASSUMPTIONS

Calibration results highly depend on the underlying data[8]. Therefore, it is important to investigate the data before calibrating the model. As a first example, we consider a transaction-based index of Dutch house prices. We use monthly returns for the period 12/1973 – 3/2011[9]. The index for the period 12/1973 – 12/1994 is based on data from the Dutch Association of Realtors and Property Consultants (NVM), see [131]. The index for the period 1/1995 – 3/2011 is from Statistics Netherlands (CBS). For more information about the latter index, see [36].

As a second example, we consider the S&P/Case-Shiller Home Price Indices[10]. These indices measure the residential housing market in metropolitan regions across the United States. All indices are constructed using the repeat sales pricing technique. This methodology collects data on single-family home re-sales, and captures re-sold sale prices to form sale pairs. The S&P/Case-Shiller Home Price Indices are calculated monthly and published with a two month lag. The index point for each reporting month is based on sales pairs found for that month and the preceding two months. This index family consists of 20 regional indices and two composite indices as aggregates of the regions. We here consider the 10 city composite index, which tracks the house price in the original 10 S&P/Case-Shiller indices, because more historical data is available (compared to the 20 city composite index). We use the monthly index levels for the period 2/1987 – 3/2011. Figure 3.1 contains an overview of the statistics of the Dutch house prices and the S&P/Case-Shiller Home Price Indices.

[8] We note that the choice of historical data (for example, the region, type etc.) depends on the application at hand.

[9] The annual returns are calculated using the monthly returns in a certain year.

[10] More information about the Case-Shiller indices, including historical data and information about the index construction, can be found on http://www.homeprice.standardandpoors.com.

Table 3.1: Summary statistics

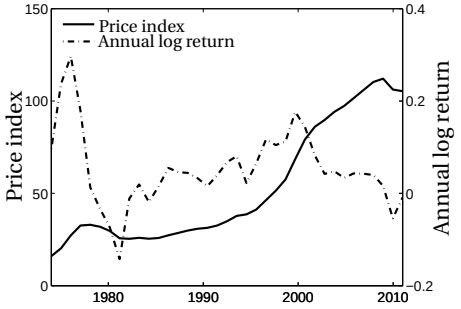
		Yearly		Monthly	
	Period	1973 – 2010	1987 – 2010	12/1973 – 3/2011	2/1987 – 3/2010
NL.	Average	0.0536	0.0563	0.0535	0.0556
	St. dev.	0.0809	0.0483	0.1178	0.0720
	Skewness	0.6677	0.4112	1.0660	0.2738
	Excess Kurt.	1.5437	0.5992	4.3780	0.2321
	Acf(1)	0.7316	0.6809	0.3814	0.3276
	Acf(2)	0.3161	0.3398	0.2864	0.4679
	Acf(3)	0.0455	0.2579	0.2977	0.4375
	Acf(6)			0.3265	0.4883
	Acf(12)			0.2550	0.4346
	Acf(24)			0.1276	0.2983
U.S.	Average		0.0379		0.0365
	St. dev.		0.0874		0.1088
	Skewness		-0.7899		-0.7671
	Excess Kurt.		0.7526		1.0274
	Acf(1)		0.6915		0.9372
	Acf(2)		0.3329		0.8200
	Acf(3)		-0.0519		0.6777
	Acf(6)				0.4127
	Acf(12)				0.6305
	Acf(24)				0.3752

In Figure 3.1a we show the development of the transaction-based index of Dutch house prices and the corresponding annual log returns. We observe that house prices increased until 1977, when they experienced a sharp fall. From the mid-1980s until the early 2000s, house prices exhibited a sharp increase. The most recent period, the late 2000s, which coincides with the global financial crisis, witnesses a decrease of house prices.

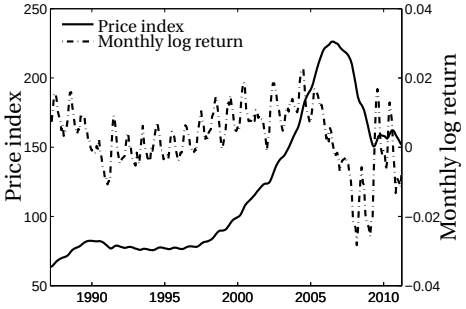
Figure 3.1b shows the development of the S&P/Case-Shiller Home Price Indices and the corresponding monthly log returns. Notice the sharp increase (more than 250%) of the house price in the period 1997 through 2006, followed by the sharp decline in prices in the last years. Looking more closely, we can also see that the log returns exhibit an oscillatory pattern with a period of approximately one year. This is due to a seasonal effect. In addition, the volatility of the log returns appears to be non-stationary over time. For example, the volatility in the quiet upward trending market (1995-2005) was much lower than during the recent market crash.

Figure 3.1: Overview of Dutch and U.S. house price index development.

(a) The CBS / NVM index of Dutch house prices between December 1973 and March 2011



(b) Development of the 10 city composite S&P/Case-Shiller index between February 1987 and March 2011



The total expected return on owner-occupied housing is the expected house price appreciation plus a convenience yield, see [34]. A convenience yield represents the (non-monetary) benefits from the housing services. When we assume that the convenience yield is a constant fraction of the house value we can model this aspect by setting the direct return q equal to the convenience yield. The authors in [34] refer to the convenience yield as an imputed rent and give an estimate of 0.67% per year for the U.S. housing market. The same percentage is used for the Dutch housing market. Note that this estimation of the convenience yield is lower than a typical rental rate. This is due to related expenses for house owners, such as depreciation, maintenance and repairs, property taxes, insurance and mortgage interest payments.

It is also important to specify the ratio of the initial index price level $a(0)$ and the efficient market price $y(0)$. If $a(0) > y(0)$, the house market is overvalued; if $a(0) < y(0)$, the house market is undervalued. The question whether the Dutch housing market is overvalued or not has been investigated by [59] using different models. Unfortunately, all models estimate the overvaluation of the Dutch market differently, ranging between approximately 0% and 12% overvaluation. Because the precise amount of overvaluation, thus, cannot be determined very accurately, we first take a neutral stance in Section 3.5.1 by assuming that the house market is in equilibrium. The effect of over- or undervaluation is then studied in a sensitivity analysis in Section 3.5.2.

3.4.2. CALIBRATION OF THE INTEREST RATE MODEL

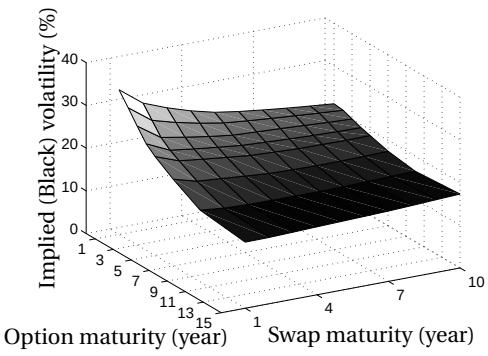
The valuation date is March 31, 2011. For the Euro area, we use the Euro (zero-coupon) swap curve as published by Bloomberg as the reference nominal interest rate curve. We estimate the stochastic interest rate model, a continuous-time one-factor Hull-White model, using market prices of forward-at-the-money options on Euro swap contracts (data also from Bloomberg). The two parameters of the one-factor Hull-White model (the mean-reversion parameter κ_r and the volatility parameter σ_r) are estimated using a large set of swaptions, with option and swap maturities ranging from 1 to 15 years.

Swaption prices are typically quoted in terms of implied (Black) volatilities. Figure 3.2a gives a graphical overview of these volatilities.

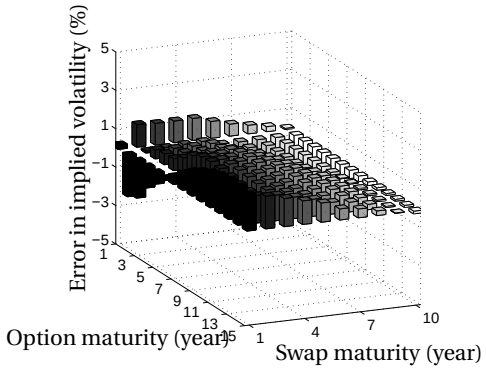
We use the Levenberg-Marquardt least-squares algorithm to find the optimal model parameters. The Hull-White parameters with the best fit are a mean reversion κ_r of $\approx$ 0.0341 and a volatility σ_r of $\approx$ 0.0097. A comparison between the model and market prices is shown in Figure 3.2b. In this figure we show the difference between the model and market implied volatility for the entire set of swaptions. The average absolute error

Figure 3.2: Calibration results of the Hull-White model.

(a) Market-implied volatilities



(b) Quality of fit of the estimated Hull-White model



is equal to 0.68-% point; the maximum absolute error is 2.16-% point (for a 2-year option on a 1 year swap). We also used a more elaborate two-factor Hull-White model. This does not improve the results significantly, however, so we continue with the one-factor model. The correlation between the Wiener processes for the short interest rate (see Eq. (3.15)) and the efficient market process (see Eq. (3.22)) is estimated using historical data. More precisely, we use the correlation between historical changes in the short interest rate and the derived efficient market returns. This correlation is equal to 0.16.

The U.S. interest rate model is calibrated in the same way as the Euro interest rate model, i.e. the same data range for swaptions is used and also the same optimization procedure. The Hull-White parameters with the best fit are a mean reversion equal to $\approx$ 0.0625 and a volatility parameter equal to $\approx$ 0.0146. The correlation parameter is also estimated using historical data and is equal to 0.45.

3.4.3. EXAMPLE 1: DUTCH HOUSE PRICE INDEX

Using annual historical data, we compare the quality of price update models with up to 3 lags. The parameters of Eq. (3.12) are estimated with ordinary least squares (OLS) regression using annual log returns for the period 1977 – 2010. By applying the augmented Dickey-Fuller test, the null hypothesis of a unit root is rejected at the 5% level for these returns, i.e. the process is stationary. The main estimation results are summarized in Table 3.2.

Table 3.2: Characteristics of the estimated price update models.

	1 lag	2 lags	3 lags
w_0^*	0.394	0.458	0.530
w_1^*	0.606	0.878	0.902
w_2^*	N.A.	-0.336	-0.387
w_3^*	N.A.	N.A.	-0.045
π	0.025	0.035	0.028
$w_0^*\pi$	0.010	0.016	0.015
σ_ϵ	0.102	0.093	0.082
$w_0^*\sigma_\epsilon$	0.040	0.043	0.043
Durbin-Watson	1.380	1.870	1.870

Because we assume in our valuation model that the efficient market returns follow a random-walk process with drift, it is important to check if the residuals indeed have a serial correlation close to zero. In this case, the Durbin-Watson test statistic should be close to 2. Table 3.2 shows that this is indeed the case for the models with 2 or 3 lags.

There exist several order selection criteria to select the best AR model. These criteria typically choose the model order in such a way that the prediction error is minimized while putting a penalty on the number of parameters estimated. The prediction error is here measured using the maximum likelihood residual variance. This variance is not corrected for the number of parameters estimated. The number of parameters is equal to the order of the AR model plus an additional parameter for the strength of the random walk process. We use the final prediction error criterion (FPE), the Akaike information criterion (AIC), the Schwarz criterion (SC) and the Hannan and Quinn criterion (HQ)[11]. The order for which the value of the criterion is minimized is seen as the model which is closest to the true model and is therefore the optimal order. Each of the criteria assumes that the models are estimated including a constant term. Table 3.3 shows that the model with 2 lags is unanimously selected.

When we select a model with two lags, one additional historical observation (1976) can be used (compared to the model with three lags). We therefore reran the calibration for the model with two lags, including this additional observation. This results in the following model parameters: $w_0^* = 0.404$, $w_1^* = 1.028$, $w_2^* = -0.431$, $\pi = 0.041$, $w_0^*\pi = 0.017$, $\sigma_\epsilon = 0.111$, $w_0^*\sigma_\epsilon = 0.045$. This model is used in the remainder of this section.

Table 3.3: Selecting the optimal order of the estimated price update models.

	1 lag	2 lags	3 lags	selected order
1. FPE	0.038	0.034	0.036	2
2. AIC	-3.276	-3.386	-3.330	2
3. SC	-3.186	-3.252	-3.151	2
4. HQ	-3.245	-3.340	-3.269	2

[11] The constant c that is used in the HQ criterion is set equal to 1.1.

3.4.4. EXAMPLE 2: U.S. HOUSE PRICE INDEX

We calibrate a monthly price update model of the U.S. S&P/Case-Shiller Home Price Index in this section. The price update model that we consider is thus different than the annual price update model that we used in the previous section. For higher frequency data aspects like stochastic volatility and seasonality become more important and these effects are therefore explicitly modeled here.

We model seasonality and stochastic volatility as follows. First, we estimate an autoregressive (AR) model with seasonal dummies using OLS (see the description in Section 3.2.1). We allow for at most 14 lags, i.e. a lookback period of at most 14 months[12]. We estimate the optimal model using the automatic model selection option in PCGive[13]. We then remove all AR coefficients that are not significant and extend this model with a GARCH(1,1) stochastic volatility model, see [16]. This model can be written in the following form:

$$(\sigma(t))^2 = \alpha_0 + \alpha_1 (r_c^a(t-1))^2 + \beta_1 (\sigma(t-1))^2,$$

where $\sigma(t)$ is the volatility of the monthly log return $r_c^a(t)$. The AR coefficients, the coefficients of the seasonal dummies and the GARCH(1,1) coefficients (α_0, α_1, and β_1) are then determined by maximum likelihood estimation in PCGive. In order to generate risk neutral scenarios using the GARCH(1,1) model the corresponding parameters α_0, α_1, and β_1 are projected to the continuous counterpart. For more information we refer to [73, p. 482].

The characteristics of the estimated model are summarized in Table 3.4. The parameters of Eq. (3.12) are estimated with ordinary least squares (OLS) regression using monthly log returns for the period 3/1988 – 3/2011.

Table 3.4: Characteristics of the estimated Case-Shiller model, including seasonality and stochastic volatility. The AR model consists of 14 lags, of which only the 5 significant coefficients are included. Three seasonal dummies (for the months February, March and June) are also included. The α_0, α_1 and β_1 parameters are the coefficients of the GARCH(1,1) stochastic volatility model.

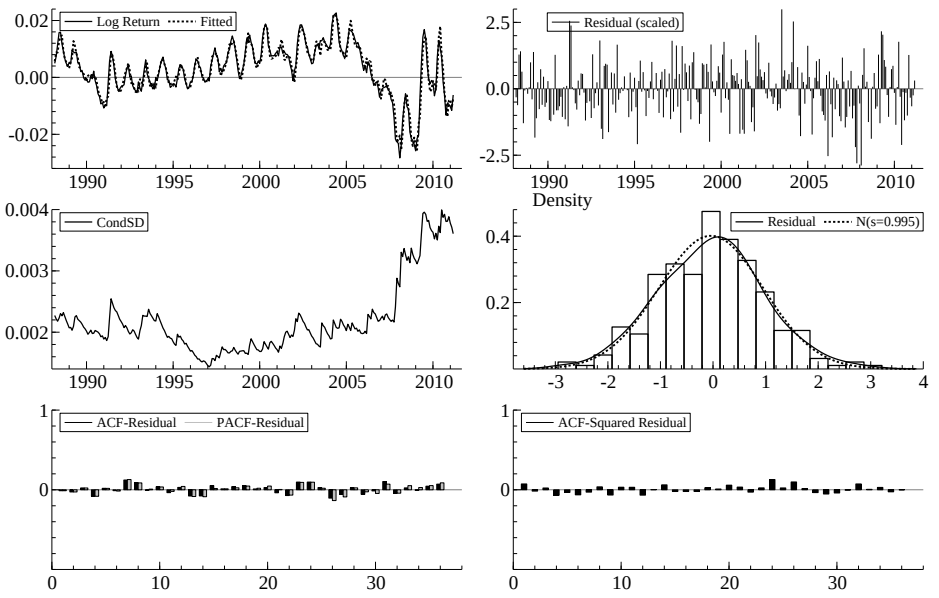
Model with 14 lags	Coefficients	Standard Error	t-Stat	p-value
$r_c^a(t-1)$	1.145950	0.03714	33.1	0.000
$r_c^a(t-3)$	-0.367669	0.07172	-5.76	0.000
$r_c^a(t-4)$	0.127003	0.05567	2.55	0.011
$r_c^a(t-11)$	0.182534	0.03585	5.15	0.000
$r_c^a(t-13)$	-0.130242	0.03579	-3.53	0.000
$r_c^a(March)$	0.00142174	0.00046	2.34	0.020
$r_c^a(April)$	0.00167801	0.00046	3.10	0.002
$r_c^a(July)$	-0.00143169	0.00048	-2.80	0.005
α_0	3.599e-8	9.621e-8	0.450	0.653
α_1	0.061430	0.02898	2.07	0.040
β_1	0.937042	0.04248	24.1	0.000

The residuals of the regression have a serial correlation close to zero: the Portmanteau test statistic is equal to 36.1 (with p-value 0.2418). The residuals are also (in accurate

[12] Results do not improve significantly when we use more than 14 lags.
[13] See [40].

approximation) normally distributed: the normality test statistic is equal to 0.63 (with a P-value of 0.72). The results are shown in more detail in Figure 3.3.

Figure 3.3: Estimation results for the 10 city composite S&P/Case-Shiller index.



The top left figure shows how well the model fits the historical data period. The top right figure displays the (scaled) residuals. The evolution of conditional volatility is displayed in the middle left figure. Notice the increasing volatility in the most recent period. The histogram in the middle right figure shows the histogram of the residuals. Notice that these are (approximately) normally distributed. The bottom figures display the autocorrelation functions (ACF) of the residuals. These autocorrelations should be close to zero if the model fits the data well. This is indeed the case. The applied GARCH(1,1) model thus successfully describes the stochastic volatility component that is present in the data.

3.5. MODEL APPLICATION: DERIVATIVE PRICING

We consider a house owner who buys a (hypothetical) at-the-money put option with a maturity of 10 years on his/her house. This option can only be exercised at the maturity date, i.e. it is a European option. The underlying index is the Dutch house price index that we introduced in Section 3.4.3. We use the price update model with two lags. The direct return q (i.e. the convenience yield) is set equal to 0.67%.

It is important to note that a property derivative market does currently not exist in the Netherlands (a liquid derivatives market does also not exist in the U.S at this moment). The example given in this section is thus only provided to illustrate the developed

valuation framework. We should also note that our valuation model assumes that continuous trading in the underlying efficient market price is possible. We explained before (see Section 3.3.1 and 3.3.2) that it is possible to replicate continuous trading in the underlying index using forward or swap contracts. Once a liquid forward or swap market has emerged in the Netherlands, the applied valuation framework can thus be applied to value (arbitrary complex) property derivatives. In practice, it is of course also important to distinguish between global and more local real estate risk. Derivatives trading typically focuses on the main (metropolitan) areas, that can make it more difficult to hedge house price risk in local areas using derivatives.

As an interesting side line: in the Netherlands a real-life option on the house price exists. Under certain (specific) conditions the so-called 'Waarborgfonds Eigen Woningen' (WEW) pays out to the bank (the lender) when the owner of the mortgage is not able to fulfill his payments. This guarantee is backed up by the Dutch government if the buffer of the WEW turns out to be insufficient in the future.

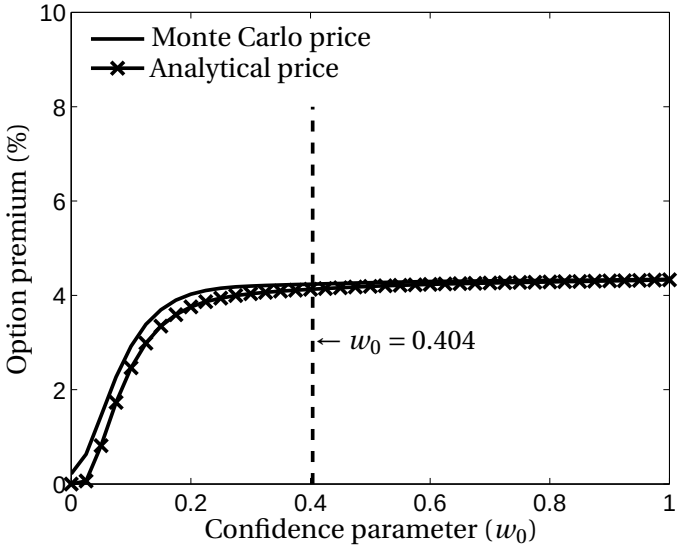
3.5.1. OPTION PRICING RESULTS

We value the 10-year at-the-money put option using Monte Carlo simulation. This is done by generating 1,000,000 risk neutral scenarios and discounting the option payoffs back to time zero along the path of the short interest rate[14]. To determine the effect of autocorrelation in the index returns on the option value we also generated results with alternative model parameters. For these alternative models, the confidence parameter w_0 is varied between zero and one. The other weights (w_1 and w_2) are proportionally scaled up or down in order to keep the sum of all weights equal to one. Note that we assume in this section that the current real estate index level $a(0)$ is equal to the current efficient market level $y(0)$, see Eq. (3.19). The impact of overvaluation (when $a(0) > y(0)$) or undervaluation (when $y(0) < a(0)$) will be studied in the next section.

The results are shown in Figure 3.4. Recall that returns are highly correlated if $w_0 = 0$. If $w_0 = 1$, returns are almost completely uncorrelated. Also keep in mind that the estimated value of w_0 is equal to 0.404, as indicated in the graph. The option premiums are expressed as percentages of the notional amount. Figure 3.4 clearly shows that the option premiums decrease when the autocorrelation of the returns increases (i.e. w_0 decreases). This is due to the smaller (cumulative) volatility of the autocorrelated real estate returns.

[14] We use antithetic sampling to reduce the standard error of the Monte Carlo estimate.

Figure 3.4: Price of a 10-year at-the-money put option on a house price index as a function of the confidence parameter w_0.

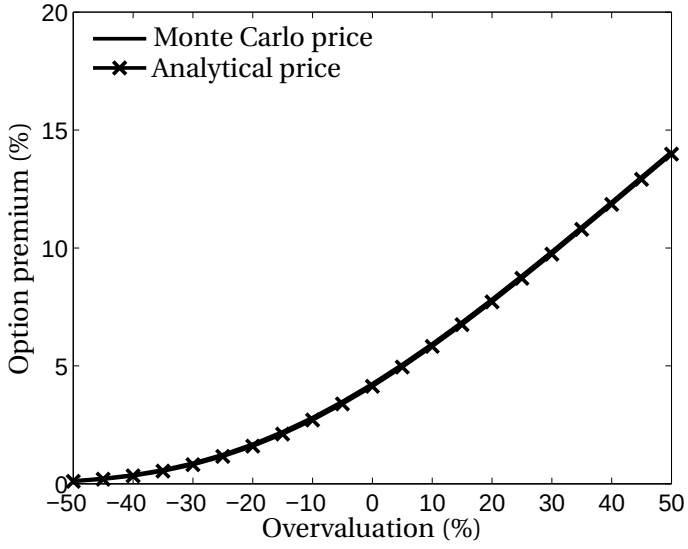


The quality of the approximate analytical pricing formula that we derived in Section 3.3.3 is also investigated in Figure 3.4. This analytical pricing formula is very accurate for high values of the confidence parameter w_0. In this case, the terminal probability distribution of the index value is determined to a large extent by only a few log-normal distributions. The terminal distribution function can be fitted well with a single log-normal distribution in this case, so the approximation error is small. When w_0 decreases, the terminal probability distribution of the index value becomes the sum of a series of different log-normal functions. As a result, the fit with one log-normal function deteriorates. However, the quality of the approximation remains quite satisfactory. For alternative approximation methods, the interested reader is referred to the overview in [95].

3.5.2. EFFECT OF OVER- OR UNDERVALUATION

We can model overvaluation and undervaluation of the real estate market by setting the current efficient market level $y(0)$ lower respectively higher than the current index level $a(0)$, see Eq. (3.19). The results are shown in Figure 3.5.

Figure 3.5: Effect of over- or undervaluation on the option price.



In this figure, the confidence parameter is set equal to the default value (0.404). Overvaluation is measured as $(a(0) - y(0))/y(0)$. The option premiums increase, to be expected, when the initial index level is higher than the efficient market price (and vice versa). Figure 3.5 also demonstrates that the agreement between the Monte Carlo price and the analytical price is very good in the case of under- or overvaluation. Information about the degree of under- or overvaluation of the real estate market may be obtained by using information in the forward or swap market (see [62] and Section 3.3) or by using information in the public real estate market, see [63].

3.6. CONCLUSIONS

We proposed a risk neutral valuation model for real estate derivatives that are linked to autocorrelated indices. Following [77], we first model the (unobserved) underlying market fluctuations using a simple random walk process with drift. We then reconstruct the observed index using an adaptation of the price update rule by [15].

The first modification of the update rule by [15] is straightforward and consists of adding multiple lag terms. This leads to an autoregressive (AR) model which can be estimated using standard econometric techniques. A second modification is more fundamental from a valuation perspective and consists of using the accrued value of past observations. We show, using real (annual and monthly) data, that this model is able to reproduce the dynamic behavior of a transaction-based house price index with autocorrelation. For high-frequency data (like monthly house prices) aspects like seasonality and stochastic volatility can become important. It is possible to model such phenomena as well within the developed framework. We analyzed the resulting model analytically and we derived closed-form pricing solutions for forwards, swaps and European

put and call options. The developed model can be applied once a liquid forward or swap market has been established. In this case it becomes possible to (approximately) replicate the underlying efficient market process. The risk neutral assumption of continuous trading in the underlying asset is then satisfied and arbitrarily complex derivatives can be hedged and priced. Given actual market prices for forwards or swaps, the derived pricing formulas can also be used to estimate the difference between the current index level and the efficient market price. This facilitates the price discovery process: information about market over- or undervaluation can be extracted from the derivatives markets for forwards or swaps. This, in turn, can also make the primary real estate market more efficient because the price update process is facilitated.

We valued a European put option on a house price index. We first generate benchmark (Monte Carlo) results and then test our approximate closed-form pricing formula. This example highlights the strong effect of autocorrelation in the underlying index on the option price. As is well known from the real estate literature, a high degree of autocorrelation reduces the (annual) volatility of the real estate index returns, compared with the (annual) volatility of the underlying true market price. This causes lower option prices, since the time value of the option decreases in this case. Using the proposed model, the effect of over- or undervaluation of the real estate market is also studied.

The proposed real estate model can be applied for different purposes. First, it can be used to price existing derivatives in real estate markets, see the examples in [11, 25, 62]. Second, it can also be used for the valuation of so-called hybrid forms of sales, see [83]. In this case a housing corporation sells a house with a discount to the tenant. In addition, there is a profit and loss sharing mechanism when the house is sold in the future. By determining the present value of the future profits and losses, the corporation can determine if the initial discount (given to the home buyer) is reasonable. This information can also be used when the corporation reports on a pure market-value basis and includes the present value of future profits and losses on the balance sheet.

The focus of the first two chapters is on the modeling and calibration of economic series. For calibration we constructed (semi-)analytical formulas to achieve fast computation times. These chapters belong to the first challenge of extending the Black-Scholes model, as described in the Section 1.6. In the Chapters 4, 5 and 6 we focus on the second challenge, which is about the nested Monte Carlo simulation issue.

APPENDIX 3.A

Using Eq. (3.22) we find that

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{y(T)}{M(T)}\,\middle|\,\mathscr{F}_t\right]=\mathbb{E}^{\mathbb{Q}}\left[\frac{y(t)\exp\left(\int_t^T r(\tau)d\tau-\sigma_y^2(T-t)/2-q(T-t)+\sigma_y W^y(T-t)\right)}{M(t)\exp\left(\int_t^T r(\tau)d\tau\right)}\,\middle|\,\mathscr{F}_t\right]. \tag{3.34}$$

The right-hand side of this expression can be simplified to

$$\frac{y(t)}{M(t)}\exp\left(-q(T-t)\right)\exp\left(-\sigma_y^2(T-t)/2\right)\mathbb{E}^{\mathbb{Q}}\left[\exp\left(\sigma_y W^y(T-t)\right)\,\middle|\,\mathscr{F}_t\right]. \tag{3.35}$$

Since we also have that

$$\mathbb{E}^{\mathbb{Q}}\left[\exp\left(\sigma_y W^y(T-t)\right)\big|\mathscr{F}_t\right] = \exp\left(\sigma_y^2(T-t)/2\right), \tag{3.36}$$

we arrive at Eq. (3.24).

APPENDIX 3.B

The proper starting point for the analysis is Eq. (3.17), since this equation enables us to write $a(T)$ as a basket of previous (accrued) efficient market prices and the (accrued) index value at time t. This becomes clear when we set t equal to T and m equal to $T-t$ in Eq. (3.17):

$$a(T) = w_0 \sum_{i=1}^{T-t} (1-w_0)^{i-1} y^*(T-i+1) + (1-w_0)^{T-t} a^*(t). \tag{3.37}$$

Let us first determine whether the accrued prices $y^*(T-i+1)$, at time T and conditional on the filtration up to time t, are martingales for $1 \le i \le T-t$ if $q=0$. This is indeed the case, since

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\frac{y^*(T-i+1)}{M(T)}\big|\mathscr{F}_t\right] &= \mathbb{E}^{\mathbb{Q}}\left[\frac{y(T-i+1)\exp\left(\int_{T-i+1}^{T} r(\tau)d\tau\right)}{M(T-i+1)\exp(\int_{T-i+1}^{T} r(\tau)d\tau)}\exp(-q(i-1))\big|\mathscr{F}_t\right] \\ &= \mathbb{E}^{\mathbb{Q}}\left[\frac{y(T-i+1)}{M(T-i+1)}\exp(-q(i-1))\big|\mathscr{F}_t\right]. \end{aligned} \tag{3.38}$$

Using Eq. (3.24) we also find that

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{y^*(T-i+1)}{M(T)}\big|\mathscr{F}_t\right] = \frac{y(t)}{M(t)}\exp\left(-q(T-t)\right). \tag{3.39}$$

The following result then easily follows:

$$\mathbb{E}^{\mathbb{Q}}\left[\frac{a(T)}{M(T)}\big|\mathscr{F}_t\right] = \frac{\exp\left(-q(T-t)\right)}{M(t)}\left(y(t)w_0\sum_{i=1}^{T-t}(1-w_0)^{i-1} + a(t)(1-w_0)^{T-t}\right). \tag{3.40}$$

Since $w_0\sum_{i=1}^{T-t}(1-w_0)^{i-1} = 1-(1-w_0)^{T-t}$, Eq. (3.25) is obtained.

CHAPTER 4

Approximation of insurance liability contracts using radial basis functions

We present the Option Interpolation Model (OIM) for accurate approximation of embedded option values in insurance liabilities. Accurate approximation is required for ex-ante risk management applications. The OIM is based on interpolation with radial basis functions, which can interpolate scattered data, and does not suffer from the curse of dimensionality. To reduce computation time we present an inversion method to determine the interpolation function weights. The robustness, accuracy and efficiency of the OIM are investigated in several numerical experiments. We show that the OIM results in highly accurate approximations.

4.1. INTRODUCTION

Classical interpolation methods suffer from the curse of dimensionality in higher dimensions and can often only be applied in low-dimensional settings (i.e. up to two or three dimensions). RBFs have the major advantage of being able to interpolate scattered data and therefore they do not suffer from the curse of dimensionality, which is an important requirement in practical risk management. Implementation is intuitive and the method also performs well in higher dimensions. Interpolation using RBFs is widely used in the field of time series prediction, control of nonlinear systems exhibiting a sufficiently simple chaotic behavior and 3D reconstruction in computer graphics. Overviews of interpolation using RBFs are presented in [23, 55] and textbooks [48, 74, 75, 138].

RBFs often contain a shape parameter, which influences the shape of the basis functions. RBF interpolants are known to be sensitive towards the trade-off between accuracy and numerical stability [49, 56, 57]. Numerical rounding errors can spoil the results of theoretically accurate interpolations, especially for small values of the shape parameter. Research has been dedicated to this trade-off, resulting in techniques to perform stable computations for a wide range of values for the shape parameter [49, 56, 57].

Popular calibration techniques for the shape parameter are the Leave-One-Out-Cross-Validation (LOOCV) [109] method and Maximum Likelihood Estimation (MLE) [113]. In LOOCV, the shape parameter is calibrated by minimizing a cost function which mimics the behavior of the root mean square error between the interpolant and the (unknown) function from which the data is sampled. MLE selects the shape parameter as the value for which the observed data is deemed most likely to occur. The LOOCV method is considered in this chapter.

This chapter is based on the article 'Approximation of insurance liability contracts using radial basis functions', submitted for publication, 2017.

We introduce the Option Interpolation Model (OIM) to approximate risk neutral option values. OIM uses radial basis functions (RBFs) [74, 100] for interpolation. The number of input variables, i.e. the risk drivers, for option valuation can be large, so that standard interpolation methods cannot be applied. Only the most important risk drivers are typically used in OIM, but there are typically more than five problem dimensions involved.

Smoothing techniques [27] can account for a possible loss in accuracy if not all risk drivers are used to determine the option value. Smoothing is also applied when the option data contains measurement noise. We will also consider smoothing in the OIM to improve stability. The modeling framework therefore also involves a smoothing parameter. The calibration of the smoothing parameter can be well incorporated in the LOOCV method which is used for calibration.

The specification of the interpolation points (as scattered data) is somewhat involved. In most publications, the interpolation points are chosen as a low-discrepancy (or quasi-random) point set [103], such as Chebyshev nodes and Halton sequences. These point sets approximate a uniform discretization of the high-dimensional domain. Next to these data-independent sets, there are data-dependent approaches to specify the interpolation points. The so-called Greedy algorithm that produces near-optimal point sets is proposed in [35]. Such sets are constructed by recursively adding points to minimize an error bound.

In [72] the RBF interpolant is used to numerically solve the Black-Scholes partial differential equation (PDE). The resulting approximation scheme is compared to the finite difference method and the true solution, the Black-Scholes formula. The method [72] leads by construction to an interpolant for the Black-Scholes formula and for its derivatives. The authors in [72] show that the method results in accurate mesh-free approximations for both European and American equity options. Instead of solving the PDE, Monte Carlo integration is nowadays the standard method for solving such pricing problems.

We evaluate the modeling framework by several numerical experiments. As a realistic test function we use the Black-Scholes formula for modeling option values. The advantage of such a test is that we can compare the approximation values with analytic reference values. In particular, we study accuracy, monotonicity and numerical stability of the OIM. We find that interpolation via RBFs results in highly satisfactory results.

The remainder of this chapter is organized as follows. In Section 4.2, we introduce the OIM and describe the related properties. In Section 4.3, we propose an inversion method to determine the appearing weights of the interpolant. This method is particularly useful for an iterative calibration method and results in rapid computations. In Section 4.4, we explain the ex-ante risk management application, the related approximation problem and we test the proposed methodology. We conclude in Section 4.5.

4.2. MODELING FRAMEWORK

We consider a d_X-dimensional random vector $X \in \mathscr{X} \subset \mathbb{R}^{d_X}$, where $d_X \in \mathbb{N}_+$ denotes the number of risk drivers for the embedded option value and $\mathscr{X}$ the domain of X. The random variable X represents the set of risk drivers and is governed by a PDF $f_X : \mathscr{X} \to \mathbb{R}$. The mapping $\eta : \mathscr{X} \to \mathscr{Y}$ that computes the option value Y given the realizations of X, so

that $Y := \eta(X)$, with $Y \in \mathscr{Y} \subset \mathbb{R}$ a one-dimensional random variable. The corresponding PDF of Y, f_Y, can generally not be obtained in closed form, due to the involved mapping η representing the option value. Option values Y are often determined by Monte Carlo simulations [64], which is market standard. In the proposed Option Interpolation Model (OIM) we approximate the expensive mapping η by a computationally cheaper mapping $\widetilde{\eta}$. The method is outlined in the sections below.

Remark. *In practice, the number of dimensions d_X is often too large to construct the approximation function $\widetilde{\eta}$. Therefore, a smaller number of dimensions than d_X is used to construct the approximation function. That is, only the relevant risk drivers are used. However, here we consider an equal number of dimensions d_X for both mappings η and $\widetilde{\eta}$.*

4.2.1. INTERPOLATION OF SCATTERED DATA USING RADIAL BASIS FUNCTIONS

Consider N instances of the random variable X denoted by $x = \{x_1, \dots x_N\} \subset \mathscr{X}$. Each instance is a combination of the d_X risk drivers. Now $y_i = \eta(x_i)$ for $i = 1, \dots, N$ denote the true option values, such that $(x, y) = \{(x_i, y_i)\}_{i=1}^N$ are available data points. We assume that there exists a univariate function $\phi : [0, \infty) \to \mathbb{R}$ with the property $\phi(X) = \phi(\|X\|)$. The function ϕ is called the radial basis function (RBF). Radial basis functions are used to build up function approximations of the form:

$$\widetilde{\eta}(x) = \sum_{k=1}^{N} w_k \phi\left(\|x - x_k\|\right), \text{ for } x \in \mathscr{X}. \tag{4.1}$$

The approximation function $\widetilde{\eta}(x)$ is represented by a sum of N radial basis functions, each associated with a different center x_k and weighted by to be estimated weights w_k.

Eq. (4.1) is known as the interpolation equation. We refer to this model as the benchmark Option Interpolation Model (OIM), which is used as a benchmark model in Section 4.4. In Section 4.2.2 we discuss three techniques for improvement. We note that $\widetilde{\eta}$ depends on a number of arguments such as the shape parameter, the interpolation points and the RBF. For notational convenience and readability, we omit the dependency on these arguments in $\widetilde{\eta}$.

Table 4.1 lists several popular choices of the RBF ϕ; these are also the most commonly used choices in the academic literature. There are other choices of RBFs, such as the Matérn family; for a more elaborate overview of these choices we refer to [23].

Table 4.1: Common choices of radial basis functions.

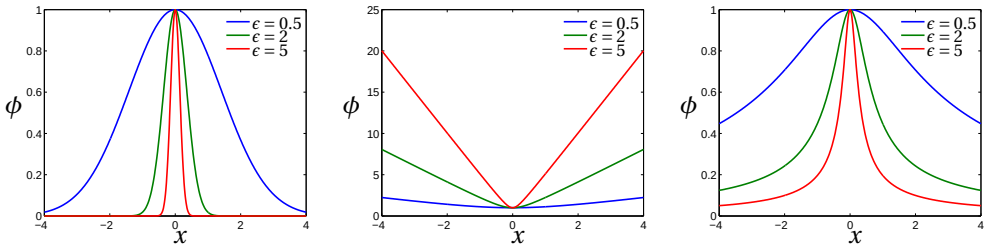
RBF	Definition
Gaussian	$\phi(r,\epsilon) = \exp -(\epsilon r)^2$
Multiquadric (MQ)	$\phi(r,\epsilon) = \sqrt{1+(\epsilon r)^2}$
Inverse multiquadric (IMQ)	$\phi(r,\epsilon) = {}^1/_{\sqrt{1+(\epsilon r)^2}}$

The optimal RBF is case-specific and depends upon three aspects: numerically stability, accuracy and monotonicity. With 'optimal' we mean that a small number of interpolation points can be used for an acceptable degree of numerical stability and accuracy. The first two requirements are common to many applications, and are often referred to as the trade-off between stability and accuracy. Accuracy comes at the cost of numerical instability. Monotonicity is another important requirement here for hedging (risk management) applications. In Section 4.4, we further explore these properties.

The RBFs listed in Table 4.1 contain the so-called shape parameter, $\epsilon \in \mathbb{R}_+$. In the majority of publications a single shape parameter is chosen. We illustrate the influence of the shape parameter in Figure 4.1 for the Gaussian, MQ and IMQ RBFs. The IMQ RBF has a similar shape as the Gaussian RBF. We consider $\phi(x,\epsilon)$ for different values of $x \in \mathbb{R}$ around one data center, the zero-point. Generally, the RBF tends to get wider for smaller values of the shape parameter ϵ.

Figure 4.1: Shape of the Gaussian RBF and MQ RBF for different values of the shape parameter ϵ.

(a) Gaussian (b) MQ (c) IMQ



The area in which the RBF is different from zero is also known as the region of influence. As Eq. (4.1) suggests, if $\phi\left(\left\|x - x_k\right\|, \epsilon\right) \neq 0$ the interpolant $\widetilde{\eta}(x)$ at $x \in \mathscr{X}$ is affected by data center x_k; otherwise it is not. For the RBFs listed in Table 4.1, the region of influence is the whole $\mathbb{R}^{d_X}$. However, as Figure 4.1 shows, the Gaussian and IMQ RBFs tend to zero, such that in a numerical sense, these RBFs are compactly supported. A 'wide' RBF gives rise to a higher probability that $x \in \mathscr{X}$ is affected by one or more data centers. This is achieved by choosing a small shape parameter[1]. Consequently, small shape parameters are henceforth the standardized approach. This comes however at the cost of numerical stability, see Section 4.2.3.

[1] This depends generally on the definition of the RBF.

As Figure 4.1 shows, the shape parameter ϵ regulates the shape of the RBF ϕ. For the Gaussian and IMQ RBF, we have $0 \leq \phi(r,\epsilon) \leq 1$ and

$$\lim_{\epsilon r \to \infty} \phi(r,\epsilon) = 0, \quad \lim_{\epsilon r \downarrow 0} \phi(r,\epsilon) = 1. \tag{4.2}$$

For the MQ RBF, we have $\phi(r,\epsilon) \geq 1$ and

$$\lim_{\epsilon r \to \infty} \phi(r,\epsilon) = \infty, \quad \lim_{\epsilon r \downarrow 0} \phi(r,\epsilon) = 1. \tag{4.3}$$

For the Gaussian and IMQ RBF, the interpolant may tend to zero due to lack of data influencing the interpolant, as $\phi(\|x - x_k\|, \epsilon)$ tends to zero. To avoid this behavior, the shape parameter is relatively small. For the MQ RBF, the interpolant does not tend to zero due to lack of data influencing the interpolant. More precisely, for the MQ RBF, at any $x \in \mathcal{X}$ we have $\phi(\|x - x_k\|, \epsilon) \geq 1$ for each data center x_k and any shape parameter $\epsilon > 0$. Thus, each data center influences the interpolant $\widetilde{\eta}(x)$ (see Figure 4.1b). As $\lim_{\epsilon r \to \infty} \phi(r,\epsilon) = \infty$, it is again preferred that the value of the shape parameter be relatively small, this time to avoid overfitting.

The interpolation weights $w = \left(w_1, \ldots, w_N\right)^T$ in Eq. (4.1) are determined by requiring that $\widetilde{\eta}$ be an interpolant of η, i.e.

$$\widetilde{\eta}(x_k) = y_k, \quad \text{for } k = 1, \ldots, N. \tag{4.4}$$

Using the abbreviation $\phi_{ij} := \phi(\|x_i - x_j\|, \epsilon)$, the interpolation statement (4.4) can be written as $\Phi w = y$, where the interpolation matrix Φ is defined as

$$\Phi = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1N} \\ \vdots & \ddots & \vdots \\ \phi_{N1} & \cdots & \phi_{NN} \end{pmatrix}. \tag{4.5}$$

Note that the diagonal elements of Φ equal 1, independent of the shape parameter, for all RBFs listed in Table 4.1. Furthermore, for these RBFs the interpolation matrix Φ is guaranteed to be non-singular if the data centers $x_1, \ldots, x_N$ are unique. However, ill-posedness can be a serious problem. For the Gaussian and IMQ RBF this was proved in [114], for the MQ RBF in [100]. Non-singularity of Φ implies existence and uniqueness of the interpolation weights w, so that the interpolant $\widetilde{\eta}$ is well-defined if the data centers $x_1, \ldots, x_N$ are unique.

For the Gaussian and IMQ RBFs, the interpolation matrix Φ is symmetric and positive definite (SPD). This does not depend on the (unique) locations of the data centers. For the MQ RBF this is not the case; the interpolation matrix is not positive definite, but symmetric. If Φ is SPD, efficient numerical techniques for solving the system $\Phi w = y$ can be used, such as the Cholesky decomposition. If Φ is not SPD, an *LU* decomposition can be used to solve the system $\Phi w = y$. A Cholesky decomposition is approximately twice as efficient as an *LU* decomposition for solving a system of linear equations [108]. The Gaussian RBF will be used from this point on.

Extrapolation behavior The RBF interpolant can also be used for extrapolation purposes. Generally, approximating within the so-called convex hull is more accurate than approximating outside the convex hull. For $\mathbf{x} = \{x_1, \ldots, x_N\} \in \mathbb{R}^{d_X}$, the convex hull of the data centers is defined as

$$\mathscr{C} = \left\{ x \in \mathbb{R}^{d_X} : x = \sum_{k=1}^{N} \lambda_k x_k \text{ for } 0 \leq \lambda_1, \ldots, \lambda_N \leq 1 \text{ such that } \sum_{k=1}^{N} \lambda_k = 1 \right\}. \tag{4.6}$$

In the case of extrapolation, $\widetilde{\eta}$ is evaluated at some $x \in \mathscr{X} \setminus \mathscr{C}$. Hence, when x is further from the convex hull $\mathscr{C}$, the distances $\|x - x_1\|, \ldots, \|x - x_N\|$ will be larger. The limiting behavior of $\phi(r, \epsilon)$ as $r \to \infty$ determines the extrapolation behavior. For the Gaussian and IMQ RBF, Eq. (4.2) gives $\lim_{\epsilon r \to \infty} \phi(r, \epsilon) = 0$. As a consequence, $\lim_{r \to \infty} \Phi = I$ independent of the shape parameter, with I the $N \times N$ identity matrix, and $\lim_{r \to \infty} w = y$. This shows in the limit $r \to \infty$, the interpolant $\widetilde{\eta}$ will tend to zero. For the Gaussian and IMQ RBF, the extrapolation thus tends to zero. For the MQ RBF we have $\lim_{r \to \infty} \phi(r, \epsilon) = \infty$, such that the extrapolation tends to infinity.

As extrapolation is typically less accurate than interpolation, the region $\mathscr{X} \setminus \mathscr{C}$ is preferably as small as possible. To reduce the necessity of extrapolation, we specify boundaries on the d_X risk drivers, such that $X \in \mathscr{D}$, with $\mathscr{D} \subset \mathscr{X}$ a bounded domain. As a consequence, the region of extrapolation $\mathscr{D} \setminus \mathscr{C}$ is bounded as well. Values outside this bounded domain are simply truncated.

Similar to polynomial interpolation, RBF interpolation suffers from Runge's Phenomenon [110], an observation of divergence near the boundary of the interpolation domain. In [17, 54], several techniques to reduce Runge's Phenomenon in RBF interpolation are discussed. An effective technique, which also extends to higher dimensions, is the addition of a low-order polynomial, see Section 4.2.2.

4

4.2.2. REDUCING OSCILLATORY EFFECTS TO IMPROVE STABILITY

A stable and accurate approximation function is important for risk management applications. For hedging applications, the option derivative behavior should also be well incorporated within the OIM. Oscillations in the OIM are highly undesirable, which can be analyzed by the derivatives of the interpolant with respect to the relevant risk drivers.

As the RBF interpolant (see Eq. (4.1)) has an attractive form, a closed-form expression of the derivatives is easily obtained. See Appendix 4.A for a short overview. The RBF interpolant as in Eq. (4.1) may exhibit an oscillatory behavior. There are three techniques to overcome this phenomenon; these will be described below. In Section 4.4, we test those techniques via numerical experiments.

Data-dependent shape parameter Alternatively to the specification in Eq. (4.1), a shape parameter can be specified for each individual data point. Although this leads to more flexibility, the number of parameters to calibrate increases as well. Instead, we will apply a data-dependent shape parameter [53], which has the following form:

$$\epsilon_k = \epsilon \frac{\max_j d_j^{\min}}{d_k^{\min}}, \tag{4.7}$$

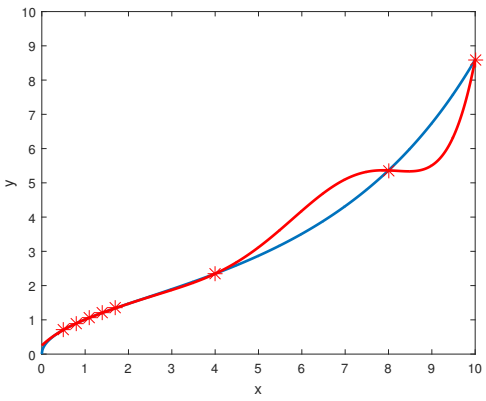
where $\epsilon > 0$ serves as a scaling parameter, $d_k^{\min}$ is the Euclidean distance between data center x_k and its nearest neighboring data center; the numerator in Eq. (4.7) is the maximum separation distance between any two neighboring data centers. The fraction in Eq. (4.7) thus never falls below 1. This choice of shape parameter ϵ_k relates to the distribution of the data centers. When the nodes are close to each other, the corresponding ϵ_k are large, leading to tight RBF kernels. When the nodes are further away, the corresponding ϵ_k are small, leading to wide RBF kernels.

The SPD property of Φ is not preserved in case of the data-dependent shape parameter, so that a Cholesky decomposition can no longer be used to determine the interpolation weights. Furthermore, the Φ matrix can no longer be guaranteed to be non-singular [53]. However, the authors in [53] find that independent of the distribution of the interpolation points and the ϵ parameter in Eq. (4.7), the eigenvalues follow a regular pattern. This would not be the case if the matrix Φ were singular. A formal proof is missing in this context.

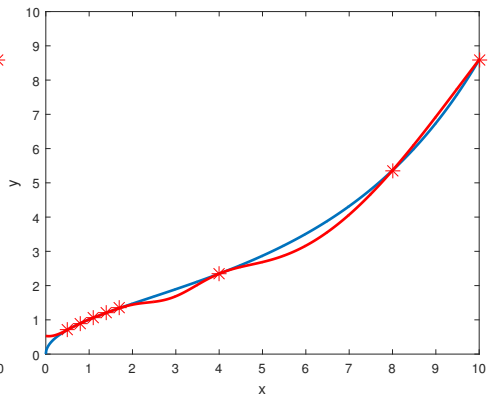
Figure 4.2 shows two interpolants, for the data-independent and the data-dependent shape parameters.

Figure 4.2: The effect of data-dependent shape parameters. Two Gaussian interpolants (in red) based on $\epsilon = 0.1$ and 8 data centers (red asterisks) and the actual function (in blue) are shown, for the data-independent (left) and the data-dependent (right) shape parameter.

(a) Data-independent



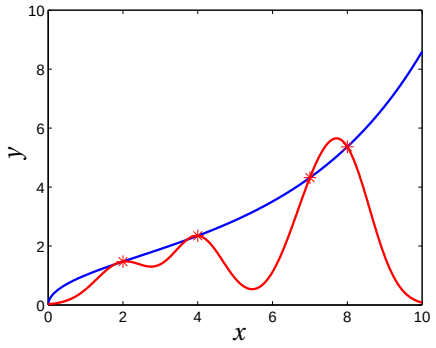
(b) Data-dependent



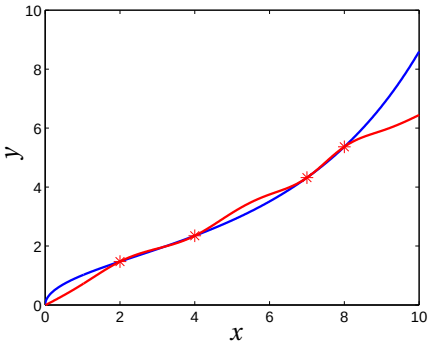
Affine term The addition of a polynomial term to the interpolation equation in Eq. (4.1) [54] has two main advantages. Undesired oscillatory effects are reduced, i.e. it reduces the effects related to Runge's phenomenon. This fits well with the option pricing context because the option's derivative behavior is important for hedging applications. We choose to add a first-order polynomial, as typically low-order polynomials reduce the effects of Runge's phenomenon better than higher-order polynomials. In Figures 4.3a and 4.3b we illustrate the impact of an affine term in combination with the interpolant, for a specific test case.

Figure 4.3: The effect of adding an affine term. Two Gaussian interpolants (in red) based on $\epsilon = 1$ and 4 data centers (red asterisks) and the actual function (in blue) are shown, without (left) and with (right) the addition of an affine term.

(a) Excl. affine term



(b) Incl. affine term



Let $\mathrm{P} : \mathscr{X} \to \mathbb{R}$ be a first-order polynomial. With the addition of this affine term, the interpolation equation becomes

$$\tilde{\eta}(x) = \sum_{k=1}^{N} w_k \phi\left(\left\|x - x_k\right\|, \epsilon\right) + \mathrm{P}(x), \quad \text{for } x \in \mathscr{X}. \tag{4.8}$$

4

The affine mapping P can be written as

$$\mathrm{P}(x) = v \cdot \begin{pmatrix} 1 \\ x \end{pmatrix} = v_0 + \sum_{k=1}^{d_X} v_k x^k,$$

with $x = \left(x^1, \ldots, x^{d_X}\right)^T$ and where $v = \left(v_0, \ldots, v_{d_X}\right)^T$ are the affine weights. The interpolation equation then reads

$$\tilde{\eta}(x) = \sum_{k=1}^{N} w_k \phi\left(\left\|x - x_k\right\|, \epsilon\right) + v_0 + \sum_{k=1}^{d_X} v_k x^k. \tag{4.9}$$

Adding an affine term often improves both the accuracy and the monotonicity of the interpolant, because it reduces oscillations in the interpolant. These oscillations originate from the first summation in Eq. (4.9). The addition of an affine term also reduces the oscillations near boundaries of the convex hull of the data centers as a result of Runge's phenomenon [54].

With the addition of an affine term, the interpolation statement (4.4) reads $\Phi w + P v = y$, where

$$P = \begin{pmatrix} 1 & -x_1^T- \\ \vdots & \vdots \\ 1 & -x_N^T- \end{pmatrix} \tag{4.10}$$

is an $N \times (d_X + 1)$ matrix. The additional equations necessary to solve the interpolation statement for w and v come from the requirement that affine data should be reproduced exactly. In [88] the equation $w^T P = 0$ is shown to imply exactly this. With the affine addition, the interpolation statement becomes

$$\widetilde{\Phi}\begin{pmatrix} w \\ v \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}, \text{ with } \widetilde{\Phi} = \begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix}. \tag{4.11}$$

Applying the affine term leads to a symmetric $\widetilde{\Phi}$ matrix, but positive definiteness and non-singularity are not preserved.

Smoothing technique Another extension to the RBF interpolant is the smoothing technique. By replacing the interpolation statement $\Phi w = y$ by

$$(\Phi + \alpha I)\, w = y, \tag{4.12}$$

with $\alpha \in \mathbb{R}$ the smoothing parameter (see [88]), we are often able to improve the systems condition number and 'smooth' the interpolation points. That is, by using such a regularization technique we correct the option data $y_1, \ldots, y_N$ by smoothed option data $\tilde{y}_1, \ldots, \tilde{y}_N$ satisfying

$$y_k = \tilde{y}_k + \alpha_k, \quad \big(k = 1, \ldots, N\big). \tag{4.13}$$

So, in fact we force the interpolant to pass through the $\tilde{y}_k$ instead of the y_k values. To avoid overfitting we choose the values $\alpha = \big(\alpha_1, \ldots, \alpha_N\big)^T$ as $\alpha = -\alpha w$, with smoothing parameter $\alpha \in \mathbb{R}$. In Section 4.2.3 we analyze the impact of the smoothing parameter on the numerical stability.

It depends on the smoothing parameter whether or not the SPD property is preserved for $(\Phi + \alpha I)$. It turns out that if $\alpha > -\lambda_{\min}$ where $\lambda_{\min}$ is the minimum eigenvalue for Φ, then the matrix $(\Phi + \alpha I)$ is SPD. Here we have used the fact that all eigenvalues are positive for an SPD matrix.

Remark. *Smoothing is applied when the option data is noisy. In the option valuations context, data is noisy when the dimension of the risk drivers of $\widetilde{\eta}$ is smaller than the true dimension of the risk drivers or when the option values are obtained by a small number of risk neutral scenarios, hence contain statistical noise. The latter is typically encountered within the LSMC method context [7].*

4.2.3. NUMERICAL STABILITY

The shape parameter regulates the width of the RBFs and thus controls the range of influence of the data centers. The optimal value of the shape parameter with respect to a certain error measure appears to be case-specific. That is, it depends on the function to be approximated. The widths of these RBFs depend on the number of data centers and their locations. The RBFs should ideally be peaked (large shape parameter) where the data centers lie close to each other, and wide (small shape parameters) where the data centers are further away from each other. This is achieved by shape parameters of the form displayed in Eq. (4.7). However, choosing smaller shape parameters comes at the

cost of numerical stability. This is known as the trade-off between accuracy and numerical stability [56, 136]. In theory, the shape parameter is preferably as small as possible; in practice numerical rounding errors will spoil the results for too small values of the shape parameter.

Numerical stability for $\Phi w = y$ is quantified by the condition number of the interpolation matrix Φ. When the condition number is large, a small error in y may cause a large error in w. For an $N \times N$ matrix Φ, the condition number is defined as

$$\kappa(\Phi) = \frac{\sigma_{\max}(\Phi)}{\sigma_{\min}(\Phi)}, \tag{4.14}$$

with $\sigma_{\min}$ and $\sigma_{\max}$ denoting the smallest and largest singular values of Φ, respectively. When Φ is SPD, Eq. (4.14) is simplified to

$$\kappa(\Phi) = \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)}, \tag{4.15}$$

with $\lambda_{\min} > 0$ and $\lambda_{\max} > 0$ denoting the smallest and largest eigenvalues. When Φ is symmetric but not positive definite, Eq. (4.14) is simplified to

$$\kappa(\Phi) = \left| \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)} \right|. \tag{4.16}$$

The shape and smoothing parameters and the addition of an affine term all influence the numerical stability and the accuracy of the interpolant. Combinations of the shape and smoothing parameters that lead to large condition numbers should be avoided, because numerical rounding errors will corrupt the results. We therefore specify a threshold $\kappa_{\max}$ for the condition number. During the calibration, combinations of shape and smoothing parameters that lead to a condition number larger than $\kappa_{\max}$ will not be considered. In other words, the parameters will be calibrated in the parameter set

$$\left\{(\epsilon, \alpha) \in \mathbb{R}^2 : \kappa(\Phi) \leq \kappa_{\max}\right\} \text{ or } \left\{(\epsilon, \alpha) \in \mathbb{R}^2 : \kappa(\widetilde{\Phi}) \leq \kappa_{\max}\right\},$$

depending on whether the affine term is added.

Impact of the shape parameter We assume the interpolant $\widetilde{\eta}$ is defined by Eq. (4.1), i.e. without the addition of an affine term and without smoothing[2]. As stated in Section 4.2.1, the interpolation matrix Φ is non-singular for the RBFs listed in Table 4.1 if the data centers are unique. However, Φ is a dense matrix, and it can become ill-conditioned. As Φ depends on the shape parameter, this parameter influences the numerical stability of the system $\Phi w = y$. For the Gaussian and IMQ RBFs, we have

$$\lim_{\epsilon \to \infty} \Phi = I, \tag{4.17}$$

with I the identity matrix, so that $\lim_{\epsilon \to \infty} \kappa(\Phi) = \kappa(I) = 1$. On the other hand, we have

$$\lim_{\epsilon \downarrow 0} \Phi = J, \tag{4.18}$$

with J an all-ones matrix. Hence, $\lim_{\epsilon \downarrow 0} \kappa(\Phi) = \kappa(J) = \infty$.

[2] The behavior of the data-dependent shape parameter is similar as for the data-independent shape parameter.

Impact of the smoothing parameter The smoothing parameter also influences the condition number of the interpolation matrix and therefore the numerical stability of the interpolation matrix Φ as well. If λ is an eigenvalue of Φ, then for any $\alpha \in \mathbb{R}$, $\lambda + \alpha$ is an eigenvalue of $\overline{\Phi} := \Phi + \alpha I$. If Φ is SPD, the condition number of $\overline{\Phi} := \Phi + \alpha I$ reads

$$\kappa(\overline{\Phi}) = \left| \frac{\lambda_{\max}(\overline{\Phi})}{\lambda_{\min}(\overline{\Phi})} \right| = \left| \frac{\lambda_{\max}(\Phi) + \alpha}{\lambda_{\min}(\Phi) + \alpha} \right|, \quad \big(\alpha \neq -\lambda_{\min}(\Phi)\big). \tag{4.19}$$

To understand the influence of the smoothing parameter α on the condition number of $\overline{\Phi}$, the relation between $\kappa(\Phi)$ and $\kappa(\overline{\Phi})$ is derived. To do so, we consider the following regions for α:

$$\begin{aligned} I_1 &= \big(-\lambda_{\min}(\Phi), 0\big), \\ I_2 &= \big(-\lambda_{\max}(\Phi), -\lambda_{\min}(\Phi)\big), \\ I_3 &= \mathbb{R} \backslash \big(-\lambda_{\max}(\Phi), 0\big). \end{aligned}$$

Note that for $\alpha \in I_1 \cup I_3$ the condition number reads $\kappa(\overline{\Phi}) = \frac{\lambda_{\max}(\Phi)+\alpha}{\lambda_{\min}(\Phi)+\alpha}$.
For $\alpha \in I_1$ we have

$$\begin{aligned} \kappa(\Phi) &= \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)} \cdot \frac{\alpha}{\lambda_{\min}(\Phi)+\alpha} + \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)+\alpha} \\ &< \frac{\alpha}{\lambda_{\min}(\Phi)+\alpha} + \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)+\alpha} = \frac{\lambda_{\max}(\Phi)+\alpha}{\lambda_{\min}(\Phi)+\alpha} = \kappa(\overline{\Phi}), \end{aligned} \tag{4.20}$$

because $\frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)} > 1$ and $\frac{\alpha}{\lambda_{\min}(\Phi)+\alpha} < 0$. For $\alpha \in I_1$ smoothing thus decreases the numerical stability.

For $\alpha \in I_2$ no relation between $\kappa(\overline{\Phi})$ and $\kappa(\Phi)$ can be derived.
For $\alpha \in I_3$ we have

$$\begin{aligned} \kappa(\Phi) &= \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)} \cdot \frac{\alpha}{\lambda_{\min}(\Phi)+\alpha} + \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)+\alpha} \\ &> \frac{\alpha}{\lambda_{\min}(\Phi)+\alpha} + \frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)+\alpha} = \frac{\lambda_{\max}(\Phi)+\alpha}{\lambda_{\min}(\Phi)+\alpha} = \kappa(\overline{\Phi}), \end{aligned} \tag{4.21}$$

because $\frac{\lambda_{\max}(\Phi)}{\lambda_{\min}(\Phi)} > 1$ and $\frac{\alpha}{\lambda_{\min}(\Phi)+\alpha} > 0$. For $\alpha \in I_3$ smoothing thus improves the numerical stability.

Impact of affine term The addition of an affine term also affects the numerical stability. To derive a relation between $\kappa(\Phi)$ and $\kappa(\widetilde{\Phi})$, with $\widetilde{\Phi} = \begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix}$, a relation between the eigenvalues of Φ and $\widetilde{\Phi}$ is required. Since Φ is SPD and $\widetilde{\Phi}$ is symmetric, according to [9] (Theorem 3.5) the eigenvalues of $\widetilde{\Phi}$ belong to the union of I_1 and I_2, with

$$I_1 = \left[\frac{1}{2}\left(\lambda_{\min}(\Phi) - \sqrt{\lambda^2_{\min}(\Phi) + 4\sigma^2_{max}}\right), \frac{1}{2}\left(\lambda_{\max}(\Phi) - \sqrt{\lambda^2_{\max}(\Phi) + 4\sigma^2_{min}}\right) \right] \tag{4.22}$$

$$I_2 = \left[\lambda_{\min}(\Phi), \frac{1}{2}\left(\lambda_{\max}(\Phi) + \sqrt{\lambda^2_{\max}(\Phi) + 4\sigma^2_{max}}\right) \right], \tag{4.23}$$

where $\lambda_{\min}(\Phi)$ and $\lambda_{\max}(\Phi)$ denote the smallest and largest eigenvalues of Φ and σ_{min} and σ_{max} denote the smallest and largest singular values of P, respectively. Theorem 3.5 applies when $\lambda_{min}, \sigma_{min} > 0$.

We use the bounds in Eqs. (4.22) and (4.23) to derive a relation between $\kappa(\Phi)$ and $\kappa(\widetilde{\Phi})$. We note that there is no assurance that these bounds are attained.

Since $\lambda_{\min}(\Phi)$, $\lambda_{\max}(\Phi)$, σ_{min}, σ_{max} are all positive, the smallest and largest eigenvalues of $\widetilde{\Phi}$ are given by

$$\lambda_{\min}(\widetilde{\Phi}) = \frac{1}{2}\left(\lambda_{\min}(\Phi) - \sqrt{\lambda^2_{\min}(\Phi) + 4\sigma^2_{max}}\right), \tag{4.24}$$

$$\lambda_{\max}(\widetilde{\Phi}) = \frac{1}{2}\left(\lambda_{\max}(\Phi) + \sqrt{\lambda^2_{\max}(\Phi) + 4\sigma^2_{max}}\right). \tag{4.25}$$

The condition number of $\widetilde{\Phi}$ is then computed as

$$\begin{aligned}\kappa(\widetilde{\Phi}) &= \left|\frac{\lambda_{\max}(\widetilde{\Phi})}{\lambda_{\min}(\widetilde{\Phi})}\right|, \\ &= \left|\frac{\frac{1}{2}\left(\lambda_{\max}(\Phi) + \sqrt{\lambda^2_{\max}(\Phi) + 4\sigma^2_{max}}\right)}{\frac{1}{2}\left(\lambda_{\min}(\Phi) - \sqrt{\lambda^2_{\min}(\Phi) + 4\sigma^2_{max}}\right)}\right| = \left|\frac{\lambda_{\max}(\Phi) + \sqrt{\lambda^2_{\max}(\Phi) + 4\sigma^2_{max}}}{\lambda_{\min}(\Phi) - \sqrt{\lambda^2_{\min}(\Phi) + 4\sigma^2_{max}}}\right|.\end{aligned}$$

Since

$$\lambda_{\max}(\Phi) < \lambda_{\max}(\Phi) + \sqrt{\lambda^2_{\max}(\Phi) + 4\sigma^2_{max}}, \tag{4.26}$$

$$\lambda_{\min}(\Phi) > \lambda_{\min}(\Phi) - \sqrt{\lambda^2_{\min}(\Phi) + 4\sigma^2_{max}}, \tag{4.27}$$

we have $\kappa(\widetilde{\Phi}) > \kappa(\Phi)$ when adding the affine term. So, conditionally on the shape parameter the numerical stability worsens in this case. This may not be the case with respect to accuracy, which is tested in Section 4.4.

4.3. COMPUTING INTERPOLATION WEIGHTS

In practice, the computational budget of an insurer does not facilitate the generation of many input option data for approximating the option function. So, the goal is to calibrate the interpolant as accurately as possible, while keeping the number of interpolation points as small as possible. An iterative (adaptive) calibration method is desired to determine the interpolation points and the shape and smoothing parameters. In this way, a smaller number of points is used for approximation, which meets the demands of an insurer better. We constructed a method to compute the interpolation weights where we add new information to the interpolant. This method is applicable to a fixed shape parameter. That is, we assume $k = 1, \ldots, N_{iter}$ iterations (with $N_{iter} \in \mathbb{N}_+$) and in each iteration we add m data centers to the interpolant. The m data centers are adaptively added by optimizing a relevant error measure (such as the LOOCV).

To form the RBF interpolant, a linear system of equations needs to be solved. Preconditioning techniques are proposed to speed up this process [67, 69, 92]. The analysis

in Section 4.2 focuses on the RBF method itself and its properties, instead of solving linear systems of equations. The decomposition method that we propose in this section can be combined with preconditioning. However, the decomposition method aims at efficiently solving an iteratively growing system of equations rather than solving a single system of equations. To properly analyze the effect and the computational gain of the proposed decomposition method itself, we choose not to apply additional preconditioning here.

In Section 4.3.1 we discuss the method and in Section 4.3.2 its numerical complexity. The method is constructed on the basis of *LU* decomposition.

4.3.1. THE DECOMPOSITION METHOD

The interpolation weights w are determined by solving a linear system of equations. We discuss a solution method for the system $\Phi w = y$, and also extend the results to the system corresponding to the affine addition (see Eq. (4.11)). Smoothing only alters the diagonal elements of the matrix Φ, and hence does not affect the methods discussed. Since the matrix Φ is non-singular, an *LU* decomposition can be defined, $\Phi = LU$, such that the interpolation system $\Phi w = y$ can be solved as[3]

$$w = U^{-1}u \quad \text{where} \quad \mathrm{u} = \mathrm{L}^{-1}\mathrm{y}. \tag{4.28}$$

Based on N data centers, Φ_1 is of the form $\Phi_1 = \begin{pmatrix} \phi_{11} & \cdots & \phi_{1N} \\ \vdots & \ddots & \vdots \\ \phi_{N1} & \cdots & \phi_{NN} \end{pmatrix}$.

Assuming $m \geq 1$ new data centers are added, i.e. $x_{N+1}, \ldots, x_{N+m}$, the interpolation system then reads

$$\left(\begin{array}{ccc|ccc} & & & \phi_{1,N+1} & \cdots & \phi_{1,N+m} \\ & \Phi_1 & & \vdots & \ddots & \vdots \\ & & & \phi_{N,N+1} & \cdots & \phi_{N,N+m} \\ \hline \phi_{N+1,1} & \cdots & \phi_{N+1,N} & \phi_{N+1,N+1} & \cdots & \phi_{N+1,N+m} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \phi_{N+m,1} & \cdots & \phi_{N+m,N} & \phi_{N+m,N+1} & \cdots & \phi_{N+m,N+m} \end{array}\right)\left(\begin{array}{c} w_1 \\ \vdots \\ w_N \\ w_{N+1} \\ \vdots \\ w_{N+m} \end{array}\right) = \left(\begin{array}{c} y_1 \\ \vdots \\ y_N \\ y_{N+1} \\ \vdots \\ y_{N+m} \end{array}\right). \tag{4.29}$$

The solution to the interpolation equation can be based on the solution of the previous interpolation equation, which saves significant computation time. Note that when extra option data points are added, the matrix Φ_1 does not change for fixed shape and smoothing parameters. In other words, the interpolation system is augmented with m rows and m columns. We write Eq. (4.29) as:

$$\begin{pmatrix} \Phi_1 & \Phi_2 \\ \Phi_3 & \Phi_4 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} y \\ y_2 \end{pmatrix}. \tag{4.30}$$

[3] Eq. (4.28) contains inverses, though in practice these inverses are never computed explicitly, as solving the whole system is more accurate and faster than explicitly computing the inverse, as is well-known.

Using the *LU* decomposition we get

$$\begin{pmatrix} \Phi_1 & \Phi_2 \\ \Phi_3 & \Phi_4 \end{pmatrix} = \begin{pmatrix} L_1 & 0 \\ L_3 & L_4 \end{pmatrix} \begin{pmatrix} U_1 & U_2 \\ 0 & U_4 \end{pmatrix}, \tag{4.31}$$

so that solving Eq. (4.30) is equivalent to solving

$$\begin{pmatrix} L_1 & 0 \\ L_3 & L_4 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} y \\ y_2 \end{pmatrix} \text{ and } \begin{pmatrix} U_1 & U_2 \\ 0 & U_4 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}. \tag{4.32}$$

Note that $L_1 U_1 = \Phi_1$, where $L_1 \equiv L$ and $U_1 \equiv U$, with L and U the original lower and upper triangular matrices in the *LU* decomposition of Φ_1. Therefore, $u_1 = L_1^{-1} y = L^{-1} y = u$. To solve Eq. (4.30), part of the *LU* decomposition is thus known from the previous iteration. We do not need to compute the matrices $L_1 = L$ and $U_1 = U$ and the auxiliary vector $u_1 = u$, thus saving computation time compared to a full *LU* decomposition. The solution to Eq. (4.30) is then obtained as

$$\begin{aligned} L_3 u_1 + L_4 u_2 = y_2 &\implies u_2 = L_4^{-1} \left(y_2 - L_3 u \right), \\ U_4 w_2 = u_2 &\implies w_2 = U_4^{-1} u_2, \\ U_1 w_1 + U_2 w_2 = u_1 &\implies w_1 = U_1^{-1} \left(u - U_2 w_2 \right). \end{aligned} \tag{4.33}$$

Addition of affine term With the addition of an affine term, the interpolation system becomes

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} w \\ v \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}, \tag{4.34}$$

with corresponding *LU* decomposition

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} = \begin{pmatrix} L_1 & 0 \\ L_3 & L_4 \end{pmatrix} \begin{pmatrix} U_1 & U_2 \\ 0 & U_4 \end{pmatrix}. \tag{4.35}$$

The solution to Eq. (4.34) is obtained by

$$\begin{pmatrix} L_1 & 0 \\ L_3 & L_4 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} U_1 & U_2 \\ 0 & U_4 \end{pmatrix} \begin{pmatrix} w \\ v \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{4.36}$$

as

$$\begin{aligned} L_1 u_1 = y &\implies u_1 = L_1^{-1} y, \\ L_3 u_1 + L_4 u_2 = 0 &\implies u_2 = -L_4^{-1} \left(L_3 u_1 \right), \\ U_4 v = u_2 &\implies v = U_4^{-1} u_2, \\ U_1 w + U_2 v = u_1 &\implies w = U_1^{-1} \left(u_1 - U_2 v \right). \end{aligned} \tag{4.37}$$

When m data centers are added, matrix Φ is again augmented with m rows and m columns, similar to Eq. (4.30). Matrix P is augmented with m rows to $\begin{pmatrix} P \\ P_2 \end{pmatrix}$, where P_2 is an $m \times (d +$

1) matrix. The new interpolation system thus reads

$$\begin{pmatrix} \Phi_1 & \Phi_2 & P \\ \Phi_3 & \Phi_4 & P_2 \\ P^T & P_2^T & 0 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \nu \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ 0 \end{pmatrix}, \tag{4.38}$$

with *LU* decomposition

$$\begin{pmatrix} \Phi_1 & \Phi_2 & P \\ \Phi_3 & \Phi_4 & P_2 \\ P^T & P_2^T & 0 \end{pmatrix} = \begin{pmatrix} J_1 & 0 & 0 \\ J_4 & J_5 & 0 \\ J_7 & J_8 & J_9 \end{pmatrix} \begin{pmatrix} V_1 & V_2 & V_3 \\ 0 & V_5 & V_6 \\ 0 & 0 & V_9 \end{pmatrix}. \tag{4.39}$$

Using this *LU* decomposition, solving Eq. (4.38) is equivalent to solving

$$\begin{pmatrix} J_1 & 0 & 0 \\ J_4 & J_5 & 0 \\ J_7 & J_8 & J_9 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} V_1 & V_2 & V_3 \\ 0 & V_5 & V_6 \\ 0 & 0 & V_9 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}. \tag{4.40}$$

The construction of the *LU* decomposition can again be based on results of the previous iteration:

$$\begin{aligned} L_1 U_1 = \Phi_1 = J_1 V_1 &\implies J_1 = L_1 \text{ and } V_1 = U_1, \\ L_1 U_2 = P = J_1 V_3 = L_1 V_3 &\implies V_3 = U_2, \\ L_3 U_1 = P^T = J_7 V_1 = J_7 U_1 &\implies J_7 = L_3. \end{aligned} \tag{4.41}$$

Furthermore, $t_1 = J_1^{-1} y = L_1^{-1} y = u_1$. To solve Eq. (4.38), we do not need to recompute the matrices J_1, V_1, V_3 and J_7 and the auxiliary vector u_1, thus again saving computation time compared to a full *LU* decomposition.

4.3.2. COMPUTATIONAL COMPLEXITY

To gain insight into the computational complexity we study the number of floating-point operations (flops) necessary to compute the interpolation weights. The gain in computational cost of the decomposition method proposed in the previous section is determined. To do so, the computational cost of performing a full *LU* decomposition of the $(N+m)\times(N+m)$ matrix Φ is compared to the computational cost of performing an *LU* decomposition of the $(N+m)\times(N+m)$ matrix Φ given the *LU* decomposition of the $N\times N$ matrix Φ. This analysis is performed for N, $m \geq 1$. Note that for $m = 0$, the computational cost of the proposed decomposition method is zero, as no new data centers are added, so that no new *LU* decomposition needs to be calculated. We focus on a certain iteration j. The cumulative gain can be analyzed by considering all iterations $j = 1, \ldots, N_{iter}$. Two steps are required for solving Eq. (4.30): forming the *LU* decomposition and performing a forward/backward substitution (for more information we refer to [24]).

The computational complexity of the *LU* decomposition of an $N \times N$ matrix Φ is equal to $\frac{2}{3}N^3 - \frac{2}{3}N$. This cost is determined as follows. During the *LU* decomposition, $N-1$ iterations of Gaussian elimination are performed. In the ith iteration, the lower right block of size $(N-i)\times(N-i+1)$ is altered. Each iteration consists of a subtraction

and a multiplication. In the ith iteration, the computational cost is thus equal to $2(N-i)(N-i+1)$, so that the total computational costs of the LU decomposition are equal to,

$$2\sum_{i=1}^{N-1}(N-i)(N-i+1)=\frac{2}{3}N^3-\frac{2}{3}N. \tag{4.42}$$

After the LU decomposition is formed, the computational cost of solving the system $\Phi w = y$ through a forward/backward substitution (see Eq. (4.28)) is equal to $2N^2-2N$, because in both the forward and the backward solve, N steps are performed, with $i-1$ subtractions and $i-1$ multiplications in the ith iteration. The forward and backward solves thus have a computational cost equal to

$$\sum_{i=1}^{N}2(i-1)=N^2-N, \tag{4.43}$$

so that the total cost equals $2N^2-2N$. Solving Eq. (4.30), where the matrix is of size $(N+m)\times(N+m)$ instead of $N\times N$, using a full LU decomposition would thus involve a computational cost of

$$\frac{2}{3}(N+m)^3-\frac{2}{3}(N+m)+2(N+m)^2-2(N+m)=\frac{2}{3}(N+m)^3+2(N+m)^2-2\frac{2}{3}(N+m). \tag{4.44}$$

However, as seen in the previous section, a full LU decomposition is not necessary, as matrices L_1 and U_1 in Eq. (4.31) are already known. Therefore, in the first N iterations of the Gaussian elimination process, only $(N+m-i)(N+m-i+1)-(N-i)(N-i+1)$ elements need to be altered. In iterations $i=N+1,\ldots,N+m$ again $(N+m-i)(N+m-i+1)$ elements are altered. The computational cost of efficiently computing the LU decomposition thus equals

$$2\sum_{i=1}^{N+m-1}(N+m-i)(N+m-i+1)-2\sum_{i=1}^{N}(N-i)(N-i+1)=\frac{2}{3}m^3+2N^2m+2Nm^2-\frac{2}{3}m. \tag{4.45}$$

After this LU decomposition, the solution is obtained through the forward/backward substitution of Eq. (4.33). Matrices L_4 and U_4 are $m\times m$, so that the computational cost of determining u_2 and w_2 is $2m^2-2m$. Similarly, as matrix U_1 is $N\times N$, the computational cost of determining w_1 is N^2-N. The total computational cost of the forward/backward substitution of Eq. (4.33) is thus equal to N^2+2m^2-N-2m, and the computational cost of efficiently solving Eq. (4.30) equals

$$\frac{2}{3}m^3+2N^2m+2Nm^2+N^2+2m^2-N-2\frac{2}{3}m. \tag{4.46}$$

The computational complexity of solving Eq. (4.30) is thus reduced from Eq. (4.44) to Eq. (4.46).

Let $F:\mathbb{N}_+\times\mathbb{N}_+\mapsto\mathbb{R}$ be the reduction factor of computation time, given by:

$$F(N,m)=\frac{\frac{2}{3}(N+m)^3+2(N+m)^2-2\frac{2}{3}(N+m)}{\frac{2}{3}m^3+2N^2m+2Nm^2+N^2+2m^2-N-2\frac{2}{3}m}. \tag{4.47}$$

When the current data set is of size N and m new data points are added, the proposed LU method to compute the new interpolation weights is $F(N,m)$ times more efficient than computing the new interpolation weights using a full LU decomposition.

Let $\bar{n}(F)$ denote the numerator and $\bar{d}(F)$ the denominator of the fraction in Eq. (4.47), such that $F(N,m) = {}^{\bar{n}(F)}/_{\bar{d}(F)}$. Then the difference between $\bar{n}(F)$ and $\bar{d}(F)$ is given by

$$\bar{n}(F) - \bar{d}(F) = \frac{2}{3}N^3 + N^2 + 4Nm - \frac{5}{2}N + \frac{8}{3}m, \tag{4.48}$$

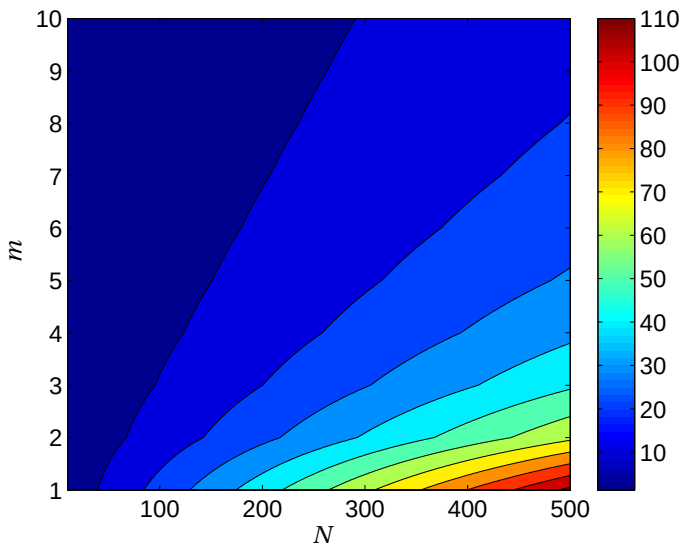
which implies $\bar{n}(F) > \bar{d}(F)$ and hence $F(N,m) > 1 \;\forall\; N,m \in \mathbb{N}_+$. So, there is always a gain in computation time when using the proposed method.

Note that

$$\lim_{N\uparrow\infty} F(N,m) = \infty \quad \text{and} \quad \lim_{m\uparrow\infty} F(N,m) = 1. \tag{4.49}$$

The gain in computation time is thus largest for large N and small m. This is as expected, and according to the practical use of this technique in the insurance industry, as the reduction factor describes the gain in computation time when the current data set is of size N and m new data points are added. Figure 4.4 illustrates the reduction factor $F(N,m)$ for different values of N and m.

Figure 4.4: Theoretical reduction factor of computational complexity for computing the interpolation weights.



4.4. NUMERICAL EXPERIMENTS

We show the impact of the data-dependent shape parameter, the addition of an affine term and smoothing to the benchmark OIM in Eq. (4.1) by means of numerical experiments. We focus on the Gaussian RBF, see Section 4.2.1. As a test case we use the

Black-Scholes option formula, see Eq. (4.50). The Black-Scholes formula can be seen as a generic option pricing formula for calls and puts. Such options are comparable to unit-linked guarantee options, that are sold by insurance companies. Black-Scholes option values [14] are computed by

$$V := V(\sigma, r, q, A, K, \tau, \omega) = \omega \left(A \exp\left(-q\tau\right) \mathcal{N}(\omega\, d_1) - K \exp\left(-r\tau\right) \mathcal{N}(\omega\, d_2) \right), \tag{4.50}$$

with K the strike level, $\tau > 0$ the time to maturity, r the deterministic interest rate, σ the volatility, q the fixed dividend yield, A the asset price index, and

$$d_1 = \left(\log\left(\frac{A}{K}\right) + (r - q + \frac{1}{2}\sigma^2)\tau \right) / \left(\sigma\sqrt{\tau}\right), \quad d_2 = d_1 - \sigma\sqrt{\tau}. \tag{4.51}$$

Furthermore, $\mathcal{N}(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and $\omega = 1$ for a call and $\omega = -1$ for a put option.

We assume $q = 0$ and a 10-year maturing put option, with $K = 120\%$. Hence, we restrict ourselves to the economic variables (i.e. the risk drivers) interest rate and volatility, i.e. $X = \left(r, \sigma\right)^T$. The function to be approximated is

$$V := V(\sigma, r, 0, 1, 1.2, 10, -1). \tag{4.52}$$

After applying the chain-rule the derivatives of the Black-Scholes formula are known in closed-form as well. In Eqs. (4.53) and (4.54) the formulas of the first-order derivatives with respect to interest rate and volatility are given, because they will be used in the numerical experiments to follow.

$$\partial V / \partial r = \omega K \tau \exp\left(-r\tau\right) \mathcal{N}(\omega\, d_2), \tag{4.53}$$

$$\partial V / \partial \sigma = K\sqrt{\tau} \exp\left(-r\tau\right) \mathcal{N}(d_2), \tag{4.54}$$

In Sections 4.4.1 and 4.4.2 we assume the risk drivers to follow a uniform distribution with $r \in [-0.01, 0.1]$ and $\sigma \in [0, 0.6]$, so that $\mathcal{D} = [-0.01, 0.1] \times [0, 0.6]$. Based on historical data such a distribution would be unrealistic; therefore we consider a realistic distribution in an insurance case in Section 4.4.3.

We aim to approximate this Black-Scholes formula using the OIM. We define the following 4 models for comparison:

OIM1 the benchmark OIM as in Eq. (4.1);

OIM2 the OIM1 extended by the polynomial term;

OIM3 the OIM1 extended by the data-dependent shape parameter;

OIM4 the OIM1 extended by the smoothing parameter;

4.4.1. TRADE-OFF BETWEEN ACCURACY AND STABILITY

The number of interpolation points that can be used in the OIM depends on the computational budget of the insurer. These interpolation points are computed on the basis of Monte Carlo valuations, which are generally expensive. An important requirement for

constructing the OIM is therefore to keep the number of valuations as small as possible. So, rapid convergence is crucial for the method to work in practice. From this perspective, an iterative (adaptive) sampling method may be preferred, because each iteration one can benefit from the information obtained in previous iterations. There are several ways of performing such iterative sampling method.

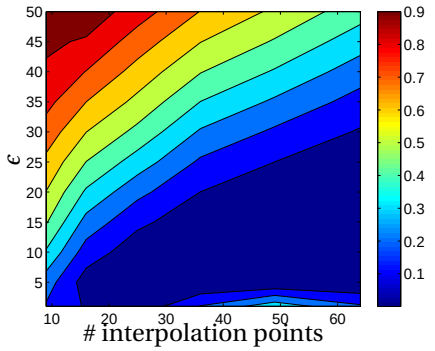
Error bounds [112] and convergence rates are available for interpolation based on RBFs. However, the precise convergence rate depends very much on the problem at hand and the structure of the interpolation points. As it turns out, quantifying convergence factors in the general setting is involved due to the presence of the shape (and possibly the smooth) parameter. Instead of analyzing convergence, we analyze the trade-off between accuracy and stability. This trade-off is highly non-linear and sensitive to the number of interpolation points.

In this numerical experiment we gain insight in the trade-off between numerical stability and accuracy for OIM1 and OIM2. We vary the number of interpolation points with $N = 9, 16, 25, 36, 49, 64$ and the shape parameter between 1 and 50 with step size 5. The interpolation points are evenly distributed in the domain $\mathscr{D}$. For each combination of N and ϵ, we compute the LOOCV error measure (see Appendix 4.A) and the condition number. The advantage of the LOOCV error measure is that it is based on out-of-sample testing.

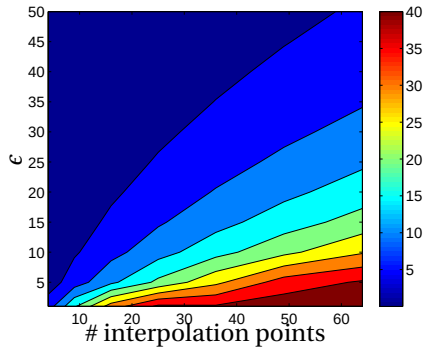
In Figures 4.5a, 4.5b and 4.5c, 4.5d we illustrate the trade-off between accuracy and stability for OIM1 and OIM2, respectively. We compute the log of the condition number for better visibility.

Figure 4.5: Trade-off between accuracy and stability.

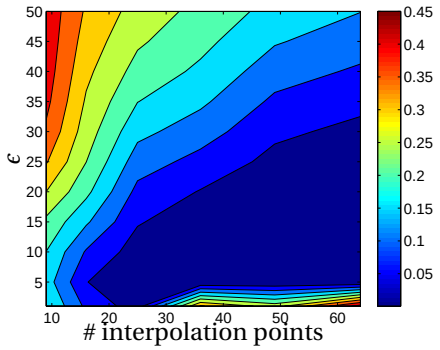
(a) LOOCV error (OIM1)



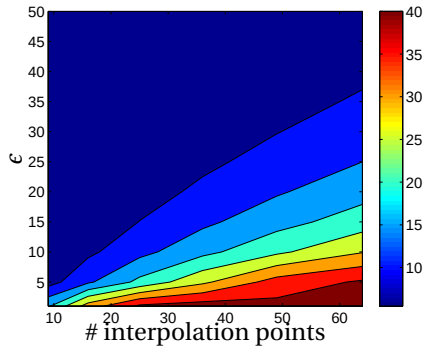
(b) Log condition number (OIM1)



(c) LOOCV error (OIM2)



(d) Log condition number (OIM2)



We observe that the accuracy gets worse for larger shape parameter values, and even more so for small numbers of interpolation points. Accuracy also gets worse when the number of interpolation points increases for small values of the shape parameter. The condition number increases when the number of interpolation points increase, hence it becomes nontrivial to calibrate the shape parameter when a large sample size is used. For a (relatively) large number of interpolation points large shape parameters result in stable systems. However, this comes at the cost of less accurate results.

We observe in this experiment that the OIM2 has a much better trade-off between accuracy and stability than OIM1. This means that OIM2 can be calibrated with higher accuracy for a given numerical stability measured by the condition number.

4.4.2. COMPARISON OF INTERPOLATION MODELS

The calibration of the interpolation points and the model parameters is involved due to the trade-off between accuracy and stability, but also due to the dimensionality. Each interpolation data point may be interpreted as a parameter to be estimated. A large number of interpolation points gives rise to a higher-dimensional optimization problem. In order to avoid high-dimensional calibrations we generate 100 uniformly distributed interpolation points sets with $N = 9$. Since we randomly generate the interpolation points, they are not evenly distributed (as in Section 4.4.1). We use a small number of interpolation points because this is to be preferred in practice.

The advantage of such an experiment is that the set of interpolation points is the same for each model, which enables fair comparison of the interpolation capabilities of the different models. We note that calibration results can be improved for each model by carefully choosing the interpolation points. The performance of the calibration depends on the target function used for calibration. For each set of interpolation points and model, we again calibrate the shape parameter and, if applicable, the smoothing parameter using the LOOCV method. We avoid combinations of the shape and the smoothing parameters for which $\kappa_{\max} = 10^6$, so that stability is sufficiently guaranteed.

To compare the models OIM1-OIM4 we make use of the RMSE measure. The RMSE is defined by

$$RMSE\left(\widetilde{\eta},\eta\right) := \sqrt{\frac{1}{N_{RMSE}} \sum_{j=1}^{N_{RMSE}} (\widetilde{\eta}(x_j) - \eta(x_j))^2}, \tag{4.55}$$

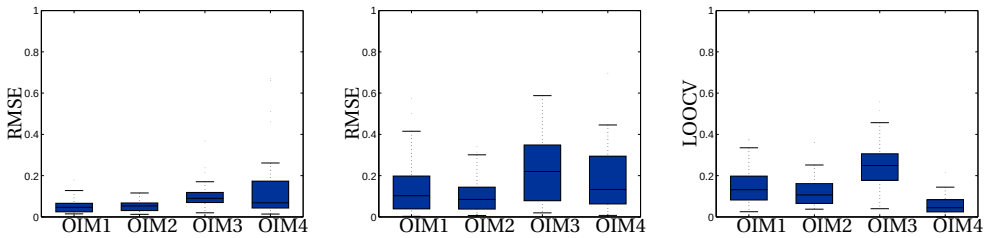
where $x_j \in \mathscr{D}$ with $j = 1,\ldots,N_{RMSE}$ and $N_{RMSE} = 10^5$ are test points chosen in such a way that they discretize the region $\mathscr{D}$. Next to this RMSE quantity, computed in $\mathscr{D}$, we also compute the RMSE in the tail region $\mathscr{D}_{\text{tail}} \subset \mathscr{D}$, which we define such that η is highest in $\mathscr{D}_{\text{tail}}$. The Black-Scholes formula gives the highest option values for low interest rates and high volatility and therefore we choose

$$\mathscr{D}_{\text{tail}} = \left\{(r,\sigma) \in \mathscr{D} : r \in [-0.01,-0.0045] \text{ and } \sigma \in [0.57,0.6]\right\},$$

which corresponds to a 95%-tail event. A high-quality fit in these tails is required in practice. The RMSE in $\mathscr{D}_{\text{tail}}$ is then calculated by generating a large and fixed sample that discretizes the region $\mathscr{D}_{\text{tail}}$. Figure 4.6 shows the box-plots of the RMSE in $\mathscr{D}$ and $\mathscr{D}_{\text{tail}}$ for each OIM.

Figure 4.6: Insight in the accuracy.

(a) RMSE in $\mathscr{D}$ (b) RMSE in $\mathscr{D}_{\text{tail}}$ (c) LOOCV

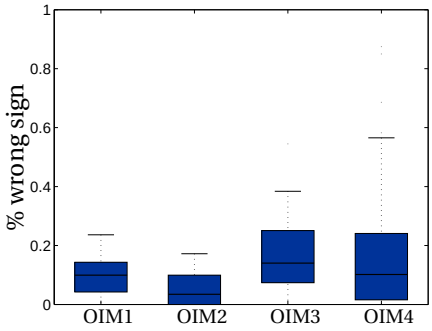


We observe that there is no model that clearly outperforms the other models. The OIM4 results in the best calibration results, i.e. small LOOCV errors, and second best is the OIM2. The quality of fit in $\mathscr{D}_{\text{tail}}$ depends on the number of test points outside the convex-hull and the ability of the model to deal with extrapolation. In this experiment we test the latter as the interpolation points are the same for each model. The OIM2 model results in the best accuracy in $\mathscr{D}_{\text{tail}}$. Hence, the addition of the affine term results in highly satisfactory results here.

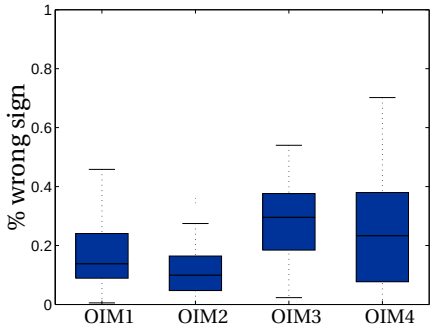
To gain insight in the monotonicity, we compare derivatives of the interpolant to the true derivatives. We focus on the sign of the derivative, as this gives important information on the behavior of the interpolant (i.e. indicating increase or decrease). For each model and for each test point $x_j \in \mathscr{D}$, the exact first derivatives with respect to r and σ are compared to the corresponding derivatives of the interpolant. The exact derivatives are given by Eqs. (4.53) and (4.54). The derivatives of the interpolant can be computed in closed-form as well (see Appendix 4.B). We measure the monotonicity by the percentage of derivatives that has the wrong sign. Hence, the smaller this percentage the better the monotonicity. Figure 4.7 shows the monotonicity for each model, depicted as the percentage of approximated first derivative that has the wrong sign, for both risk drivers r and σ.

Figure 4.7: Insight in the monotonicity.

(a) Derivative r



(b) Derivative σ



With respect to monotonicity there is more variety in the results. The model OIM2 outperforms the other models with respect to monotonicity. This model results in a small percentage and has a small average percentage of wrong signs and low variance. So, the addition of the affine term also improves the monotonicity of the interpolant here.

4.4.3. INSURANCE CASE

We test the OIM for a fictional insurance company XYZ to approximate future embedded option values in an ex-ante risk management application.

The balance sheet We assume the insurer sells unit-linked products of which we consider a simplified version. The starting balance sheet in Euros of insurance company XYZ is summarized in Table 4.2. We assume a perfectly matching asset portfolio for the guaranteed liability cash flows, an asset portfolio (denoted by A) and an embedded option (denoted by V).

Table 4.2: Balance sheet of XYZ at $t = 0$.

Assets		Liabilities	
		Matching portfolio	€700
Assets	€200	Embedded option	€100
		Surplus	€100
Total	€900	Total	€900

We consider the unit-linked product to be of European type, where capital is invested in a single equity series. The unit-linked product then becomes an equity index put option, and the value is relevant to the insurer's balance sheet. We consider again a

120% in-the-money put option with maturity equal to 10-years, which we value using the Black-Scholes formula. The function to be approximated is found in Eq. (4.52).

Projection of the balance sheet To project the balance sheet for $t \geq 0$, the asset process $\{A(t)\}_{t\in[0,T]}$ and the embedded option process $\{V(t)\}_{t\in[0,T]}$ are relevant, because the surplus (available capital) $\{S(t)\}_{t\in[0,T]}$ is computed by the difference of the value of the asset and the value of the liabilities: $S(t) = A(t) - V(t)$. Other balance sheet items are neglected in this simplified example. Using the set-up of the balance sheet in Table 4.2 we are able to directly compare the approximation to $V(t)$ and hence also $S(t)$.

We assume that $\big(A(t), v(t), r(t)\big)$ are modeled by a Heston-Vasicek model [18], where $\sigma(t) := \sqrt{v(t)}$.

$$\begin{aligned} dA(t) &= (r(t) + \mu_A)\, A(t)dt + \sqrt{v(t)}\, A(t)\, dW^A(t), \\ dv(t) &= \kappa_v\, (\bar{v} - v(t))dt + \sigma_v \sqrt{v(t)}\, dW^v(t), \\ dr(t) &= \kappa_r\, (\bar{r} - r(t))dt + \sigma_r\, dW^r(t), \end{aligned}$$

with

$$\begin{aligned} A(0) = 200,\ \mu_A = 0.02,\ v(0) = 0.02,\ \bar{v} = 0.04,\ \kappa_v = 0.02, \\ \sigma_v = 0.01,\ r(0) = 0.02,\ \bar{r} = 0.04,\ \kappa_r = 0.02,\ \sigma_r = 0.01, \end{aligned}$$

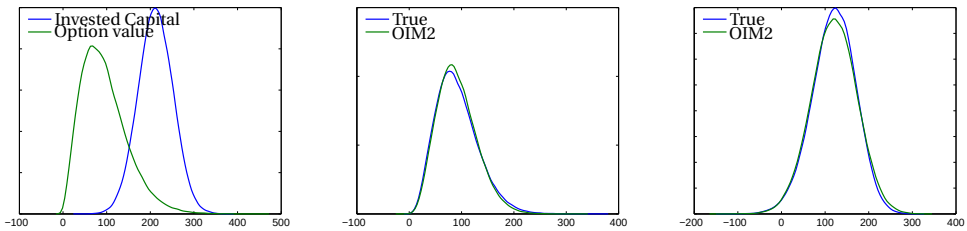
and where $W^A(t), W^r(t), W^v(t)$ are correlated Wiener process with correlation matrix Σ

$$\Sigma = \begin{pmatrix} 1 & & \\ \rho_{A,\sigma} & 1 & \\ \rho_{A,r} & \rho_{\sigma,r} & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ -0.75 & 1 & \\ 0.20 & -0.10 & 1 \end{pmatrix}. \tag{4.56}$$

Approximation of the embedded option Here, we use the true option value in Eq. (4.50) and the approximation OIM2 to compute the embedded option process $V(t)$. We favor the OIM2 from Section 4.4.2 based on the monotonicity and quality of fit. We select the optimal set of interpolation points and corresponding shape parameter from the 100 randomly generated sets interpolation points based on the LOOCV error measure. These interpolation points are visualized in Figure 4.9. We generate a large sample of one million scenarios of $\big(A(t), r(t), \sigma(t)\big)$ with a $t = 1$-year horizon. In Figures 4.8b and 4.8c we compare the embedded option and surplus processes ($V(1)$ and $S(1)$).

Figure 4.8: Comparison of V_1 and S_1.

(a) True V_1 and A_1 (b) Approximation of V_1 (c) Approximation of S_1



We observe a highly satisfactory resemblance for the embedded option and surplus processes. These approximation results are based on a randomly generated set of interpolation points with $N = 9$. The results can be further improved by extending the number of interpolation points or by carefully selecting the interpolation points.

Comparison of computation times For ease of implementation, we make use of the exact Black-Scholes formula for numerical approximation. However, in practice such a closed form formula of an embedded option does not exist due to the complexity of the option structure. Monte Carlo simulations are therefore used as numerical approximation. To compare computation times, we therefore make use of a nested Monte Carlo simulation approach.

For numerical implementation we use the Euler scheme of the Black-Scholes stochastic differential equation (SDE) in combination with a small discretization step $\Delta t = 1/100$. We note that the implementation of the nested Monte Carlo simulation is not optimized with respect to high performance computing (HPC) [33] although the simulation code is vectorized. The comparison of computation times is performed in the computation program Matlab.

We find that for a single valuation the OIM2 is 42 times faster than the Monte Carlo simulation using 500 risk neutral scenarios and 82 times faster using 1,000 risk neutral scenarios. Although the computation time of a single valuation of the Monte Carlo simulation is still highly satisfactory (0.0162s and 0.0296s respectively), in a nested framework this computation time increases with the number of valuations. An increase in the number of valuations from 1 to 1,000 or from 1 to 10,000 results in an increase of the computation by factors 1,000 and 10,000 respectively. The resulting computation times then become too high.

Given the interpolation points we can pre-compute the interpolation weights for OIM2. Next, increasing the number of interpolations in OIM2 from 1 to 1,000 results in an increase of the the computation time of a factor 3 (instead of 1,000). An increase in the number interpolation points from 1 to 10,000 results in an increase of the computation time by a factor 14 (instead of 10,000). These factors are thus much smaller than those of the nested simulation framework.

Furthermore, we find that increasing the number of interpolation points from 9 to 64 results in an increase of the computation time by a factor 5. Given the exponential

convergence rates of RBF interpolants [139], this leads to satisfactory trade-offs between accuracy and computation time.

Computing the required capital Next, we analyze the impact of computing the required capital at $t = 0$, which is comparable to the Solvency Capital Requirement (SCR) under Solvency II. Insurers use internal models to compute the required capital for internal steering and/or reporting to the regulator. The required capital is the amount of capital the insurer must hold against unforeseen losses during a one-year period. The required capital is the VaR of a loss-function on a certain horizon (often the 1-year horizon). A common loss function (see [7]) is given by

$$L_{t+\Delta} = S_t\,(1 + r_{t,\Delta}) - S_{t+\Delta} \quad \text{with} \quad \Delta = 1, \tag{4.57}$$

with $r_{t,\Delta}$ the Δ-year risk-free rate in year t. This loss function will be used in our experiments. For a certain confidence level $\alpha \in [0,1]$, the required capital R_α at $t = 0$ is computed by:

$$\mathbb{P}\left(L_1 \geq R_\alpha\right) \leq 1 - \alpha, \quad \left(L_1 = S_0(1 + r_{0,1}) - S_1\right) \tag{4.58}$$

where we set $r_{0,1} = 0.025$ as the one-year risk-free rate at $t = 0$. Hence, the probability that the loss in one year exceeds the required capital is less than or equal to $1 - \alpha$. In practice, this confidence level is often set to 99.5%.

In Table 4.3 we give an overview of numerical values of the required capital at $t = 0$ for different confidence levels. Insurance companies with a higher credit rating are obliged to compute their required capital with a lower confidence level α. The values in parenthesis are the solvency ratios, i.e. the surplus divided by required capital. We also show the results of an improved version of OIM2, where we add 4, 9 and 16 interpolation points. We sample these additional points evenly distributed in the 90% tail-event of σ_1 and 10% tail-event of r_1. In this way we provide insight in the accuracy gain when we sample additional points in the tail of the distribution. We compute the SSE of the approximation models based on confidence levels $\alpha = 99.5\%, 97.5\%, 95\%, 90\%$. In Figure 4.9, the marked region refers to the relevant tail-event, the 16 additional sampled interpolation points are marked by x and the existing interpolation points are marked by $*$.

Figure 4.9: Additional 16 interpolation points.

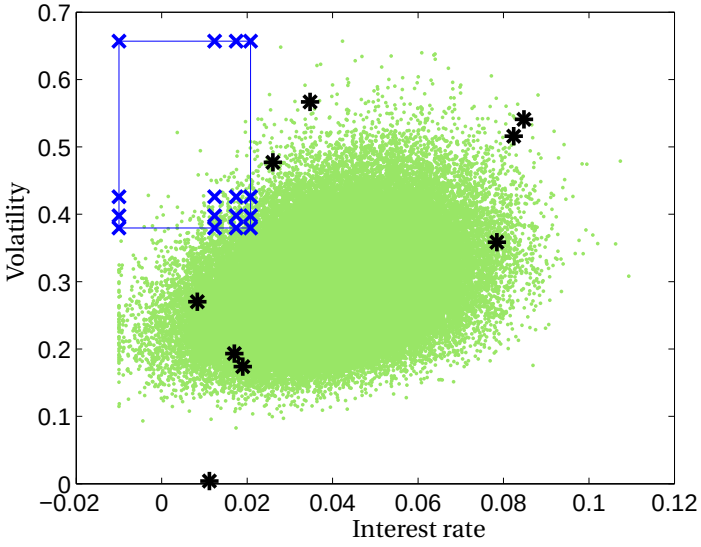


Table 4.3: Required capital at $t = 0$

	SSE	99.5%	97.5%	95.0%	90.0%
True		122.0 (0.82)	85.1 (1.18)	66.9 (1.50)	46.7 (2.14)
OIM2	12.5	118.9 (0.84)	85.2 (1.17)	67.8 (1.47)	48.3 (2.07)
OIM2 + 4	2.7	121.7 (0.82)	84.3 (1.19)	66.0 (1.52)	45.6 (2.19)
OIM2 + 9	0.4	122.3 (0.82)	85.3 (1.17)	67.2 (1.49)	47.1 (2.12)
OIM2 + 16	0.1	121.9 (0.82)	85.0 (1.18)	66.9 (1.49)	47.1 (2.13)

In this fictional example, the insurance company XYZ has solvency issues for $\alpha = 99.5\%$, because in that case the required capital is larger than the available capital. The OIM2 gives highly satisfactory results up to a VaR of 97.5%. We again note that these results are already obtained by a randomly generated set of interpolation points with $N = 9$. Accuracy improves considerably when we increase the number of additional interpolation points in the relevant tail of the distribution. The SSE decreases from 12.5 for OIM2 to 0.1 when we add 16 additional interpolation points.

4.5. CONCLUSIONS

We introduced the Option Interpolation Model (OIM) for modeling option values in ex-ante risk management applications. The method is easy to implement and results in fast computation times compared to the nested Monte Carlo simulation method. The OIM is based on interpolation using radial basis functions. This interpolation method does not suffer from the curse of dimensionality, which is a requirement in practice. The usage of nested Monte Carlo simulations is avoided using OIM.

We explained the different settings for the OIM and showed the impact of the parameters on accuracy and numerical stability, as well as the trade-off between accuracy and stability. In particular, we analyzed the shape parameter, smoothing and the addition of an affine term. We constructed a matrix inversion method to obtain interpolation weights. The method is especially useful in the case of an adaptive calibration procedure. Such a procedure is often preferred in practice. The method is most efficient when the number of additional data points is small.

All method components are compared with the Black-Scholes option formula. We find highly satisfactory improvements compared to benchmark settings. The addition of the affine term improves monotonicity, which is a desired property in practice. However, we note that the calibration of the interpolation points and the shape and smoothing parameters is a challenging topic. The accuracy results depend on the criteria to be optimized, and the calibration of the interpolation points is a high-dimensional problem, which can result in large computation times. Calibration is more difficult when the condition number is high.

We tested the OIM to an insurance case, where we computed the required capital at $t = 0$. The method already gives highly satisfactory results for only $N = 9$ interpolation points. The approximation results are improved by carefully sampling extra interpolation points. This confirms that the OIM is a good choice as an approximation method for ex-ante risk management applications for insurance companies.

The OIM is based on pre-computed option values, that are computed in pre-selected economic states. The OIM is used to approximate the option value in a real world scenario. In the next section we consider also the nested simulation framework and in particular the calibration of a risk neutral model in a certain real world scenario.

APPENDIX 4.A

In general it is quite easy to construct a model with good quality in-sample fits, but to construct a model with good quality out-of-sample fits is more involved. We make use of the Leave-One-Out Cross-Validation (LOOCV) method [109], which is a powerful method for out-of-sample error forecasting. In the LOOCV method we consider N instances of the random variable X, denoted by $x = \{x_1, \ldots, x_N\} \subset \mathscr{X}$, and the corresponding interpolation $\widetilde{\eta}$, based on some RBF ϕ. For $k = 1, \ldots, N$, let $x^{(k)} = x \backslash \{x_k\}$ be the set of data centers obtained by removing x_k from the original data set. Let $\widetilde{\eta}^{(k)}$ denote the interpolation based on data set $x^{(k)}$ and consider the following error vector

$$E = \begin{pmatrix} E_1 \\ \vdots \\ E_N \end{pmatrix}, \text{ with } E_k = \left| \eta(x_k) - \widetilde{\eta}^{(k)}(x_k) \right|. \tag{4.59}$$

E_k is thus the absolute difference between the function value $\eta(x_k) = y_k$ and the interpolated value at x_k, where the interpolation is based on all data centers except x_k. The optimal shape parameter is then chosen as $\arg\min_{\epsilon>0} \text{error}(\epsilon)$, where $\text{error}(\epsilon) = \|E\|$, for some norm $\|\cdot\|$. The evaluation of $\text{error}(\epsilon)$ is computationally expensive; N systems of size $(N-1) \times (N-1)$ need to be solved for a single evaluation of $\text{error}(\epsilon)$. In [109] an attractive form of E_k is proposed which reduces this computational burden.

Note that E_k depends on the shape parameter ϵ, as E_k assesses the interpolation $\widetilde{\eta}^{(k)}$. In [109], the LOOCV was therefore introduced as a method to calibrate the shape parameter, by minimizing error(ϵ). This can be extended to include the smoothing parameter. With smoothing, $\widetilde{\eta}$ depends on both the shape parameter ϵ and the smoothing parameter α, such that the optimal parameters can be chosen as the pair $(\epsilon, \alpha) \in \mathbb{R}_{>0} \times \mathbb{R}$ minimizing error$(\epsilon, \alpha) = \|E\|$.

The LOOCV algorithm for selecting a value for the shape parameter as presented in [109] is a specific case of Leave-p-Out Cross-Validation (LPOCV), where not one, but p data centers are removed from the original data set. In LPOCV, the error is thus measured as in Eq. (4.59), where $x^{(k)}$ is obtained by removing p data centers from the original data set. This enables the use of other error measures (such as the R^2 statistic) and the use of more holdout sets. Other interesting adaptations of the LOOCV framework are a maximum likelihood estimation and a comparison between actual and interpolated values using a local cost function (see [113] and [111]).

APPENDIX 4.B

We again consider N instances of the random variable X, denoted as $x = \{x_1, \ldots, x_N\} \subset \mathscr{X}$, and the corresponding interpolation $\widetilde{\eta}$, based on some RBF ϕ. Closed-form expressions of the derivatives of the interpolation equation in Eq. (4.1) can be obtained. The corresponding expressions are derived for the first-order derivatives of the RBFs listed in Table 4.1, see also [99]. We consider the first derivative of the interpolation $\widetilde{\eta}$ with respect to the j-th dimension, where $j = 1, \ldots, d_X$ (denoted by superscripts):

$$\frac{\partial}{\partial X_j}\widetilde{\eta}(x) = \frac{\partial}{\partial X_j}\sum_{k=1}^{N} w_k \phi\left(\|X - x_k\|, \epsilon\right) = \sum_{k=1}^{N} w_k \frac{\partial}{\partial X_j}\phi\left(\|X - x_k\|, \epsilon\right), \tag{4.60}$$

where

$$\frac{\partial}{\partial X_j}\phi\left(\|X - x_k\|, \epsilon\right) = \left.\frac{\partial}{\partial r}\phi(r, \epsilon)\right|_{r=\|X-x_k\|} \cdot \frac{\partial}{\partial X_j}\|X - x_k\| = \left.\frac{\partial}{\partial r}\phi(r, \epsilon)\right|_{r=\|X-x_k\|} \cdot \frac{(X - x_k)^j}{\|X - x_k\|}. \tag{4.61}$$

Here $(X - x_k)^j$ denotes the j-th entry of the point $X - x_k$. This leads to the following first order derivatives:

- Gaussian: $$\frac{\partial}{\partial X_j}\widetilde{\eta}(X) = -2\sum_{k=1}^{N} \epsilon^2 w_k \cdot (X - x_k)^j \cdot \exp - \left(\epsilon \|X - x_k\|\right)^2, \tag{4.62}$$

- MQ: $$\frac{\partial}{\partial X_j}\widetilde{\eta}(X) = \sum_{k=1}^{N} \epsilon^2 w_k \frac{(X - x_k)^j}{\sqrt{1 + \left(\epsilon \|X - x_k\|\right)^2}}, \tag{4.63}$$

- IMQ: $$\frac{\partial}{\partial X_j}\widetilde{\eta}(X) = -\sum_{k=1}^{N} \epsilon^2 w_k \frac{(X - x_k)^j}{\left(1 + \left(\epsilon \|X - x_k\|\right)^2\right)^{3/2}}. \tag{4.64}$$

With the addition of an affine term, the first-order derivatives of Eqs. (4.62), (4.63), (4.64) are augmented by an extra term v_k.

CHAPTER 5

On the modeling of nested risk neutral stochastic processes with applications in insurance

We propose a modeling framework for risk neutral stochastic processes nested in a real world stochastic process. The framework is important for insurers that deal with the valuation of embedded options and in particular at future points in time. We make use of the class of State Space Hidden Markov models for modeling the joint behavior of the parameters of a risk neutral model and the dynamics of option market instruments. This modeling concept enables us to perform non-linear estimation, forecasting and robust calibration. The proposed method is applied to the Heston model for which we find highly satisfactory results. We use the estimated Heston model to compute the required capital of an insurance company under Solvency II and we find large differences compared to a basic calibration method.

5.1. INTRODUCTION

We present a modeling framework to compute values of options in insurance that start at a future point in time. Such a modeling framework is important for ex-ante risk management for insurance companies. Since valuation via Monte Carlo simulations [64] is the market standard, this gives rise to a so-called nested Monte Carlo simulation problem, as discussed in Chapter 1. Instead of solving the nested simulation problem [6, 7, 33], here we propose a different modeling framework for the risk neutral models to compute future option values. In order to compute future option values at given monitoring dates, we need a risk neutral model that is consistent with the generated real world dynamics of interest rates and implied volatility at each monitoring date.

In [8] a mathematical framework is provided for the derivation of the required risk capital under the European regulatory framework Solvency II. The Option Interpolation Model presented in Chapter 4 can also be used in this context. Different alternatives for the numerical implementation based on nested simulations are reviewed. The model parameters should vary in each scenario in order to obtain consistency with the real world scenarios. The calibration to the simulated real world dynamics of option market prices would be too time consuming to perform for each real world scenario set and hence avoiding such calibrations is desired in practice. We specify and calibrate a simulation model for the model parameters, so that calibration time is reduced.

In [50] (and related articles [124, 125]) the computation of exposure distributions at a

This chapter is based on the article 'On the Modeling of Nested Risk Neutral Stochastic Processes with Applications in Insurance', published in *Applied Mathematical Finance,* 1-35, 2017 [122].
The notation in this chapter is slightly different than in the other chapters.

future point in time is discussed. These distributions are required for the computation of a credit valuation adjustment and potential future exposure, that are relevant for banks. Nested simulations are avoided by deriving an efficient approximation technique based on regression. First, risk neutral scenarios are generated, that are used to calculate option values via regression. Second, real world scenarios are generated and linked to the earlier regressed option values for the risk neutral scenarios. Our methodology is different, we directly forecast the model parameters under the real world measure and hence we embed the risk neutral measure dynamics in the real world measure dynamics.

The problem of modeling the dynamics of equity index option prices (often in terms of implied volatility) has been investigated intensively in the academic literature. The focus is however typically on the empirical facts in implied volatility surfaces and not on the link with an underlying risk neutral model that can be used for out-of-sample valuation. The latter is relevant to be able to determine the future value of embedded options and is therefore considered in this chapter. In [71] a survey of methodologies is carried out for constructing implied volatility surfaces at a certain time t. In [61] the Stochastic Volatility Inspired (SVI) parametrization method for equity option implied volatilities (IVs) is discussed. In [31] a study to the dynamics of the IV surfaces is performed and it is shown that they are driven by a small number of random factors. A factor model is proposed, which is compatible with the empirical observations. In [29] a partial differential equation (PDE) approach is presented to model the dynamics of IVs.

We make use of the class of the well-known State Space Hidden Markov (SSHM) models [42], that provide us a general modeling framework. SSHM models gained interest in the last 20 years due to the increased computational power and the improved estimation techniques. SSHM models are applied in many applications like navigation, time series analysis or robotic motion planning and control. By using the class of SSHM models, we can connect the dynamics of the model parameters of a risk neutral model to the dynamics of the option market prices. By modeling the model parameters as stochastic processes under the real world measure we enforce consistency with the option market price dynamics and the risk neutral model can be used for (out-of-sample) valuation at a monitoring date.

During the calibration of the SSHM models we reduce the dimension of the calibration problem to improve the numerical stability of the model parameters with respect to the option market data, i.e. we wish to avoid overfitting issues. Numerical stability is an important requirement in practice, because stable balance sheet valuations are desired. We reduce the dimension by distinguishing static and dynamic model parameters, where static model parameters do not change over time. This is also beneficial for present value calibrations (and valuations) because only the dynamic part needs to be calibrated.

We apply the proposed framework to the risk-neutral Heston model [70], which is popular in practice due to its flexibility in modeling implied volatility structures, and use S&P-500 index option data for calibration. Although the Heston model can be well calibrated to market option data, we show that some industry standard numerical techniques suffer from numerical instability, which results in undesired noisy calibrations. Using the proposed modeling framework we reduce the 5 risk factors, i.e. the mean reversion, initial variance, long term variance, volatility of variance and correlation, of the

Heston model to one single hidden risk factor. The calibration fit to the market data is highly satisfactory with respect to accuracy and stability.

The remainder of this chapter is organized as follows. In Section 5.2 we outline the problem and formulate the mathematical framework. In Section 5.3 we propose the modeling framework for predicting the risk neutral model parameters under the real world probability measure. In Section 5.4 we apply the proposed methodology to the Heston equity model and we show the impact of the proposed methodology by valuing an unit-linked insurers product. We conclude in Section 5.5.

5.2. METHODOLOGY

5.2.1. MATHEMATICAL FRAMEWORK

Risk neutral valuation Let $(\Omega, \mathscr{F}, \mathbb{Q})$ be a probability space, where Ω represents the space of all possible states in the financial market, $\mathbb{Q}$ is the so-called risk neutral probability measure and the filtration $\{\mathscr{F}_u\}_{u\in[0,T_u]}$ represents all information about the financial market up to time u where $T_u \in \mathbb{R}_+$ is the maturity of the longest-term option in the option portfolio. We assume that investors can trade continuously in a frictionless financial market. We introduce a d_Z-dimensional Markov process $\{Z_\Theta(u)\}_{u\in[0,T_u]}$, to model the uncertainty in the financial market required for option valuation, where $Z_{u,\Theta}$ evolves under $\mathbb{Q}$ according to the stochastic differential equation (SDE) in Eq. (5.1) and is parametrized by a d_Θ-dimensional vector $\Theta \in \mathbb{R}^{d_\Theta}$,

$$dZ_\Theta(u) = \mu_{Z,\Theta}(Z_\Theta)\,du + \sigma_{Z,\Theta}(Z_\Theta)\,dW^Z(u), \quad \Big(Z_\Theta(0) \in \mathbb{R}^{d_Z}, 0 \le u \le T_u\Big), \tag{5.1}$$

with model parameters $\Theta \in \mathbb{R}^{d_\Theta}$. Here $\mu_{Z,\Theta} : \mathbb{R}^{d_Z} \mapsto \mathbb{R}^{d_Z}$ denotes the drift process, $\sigma_{Z,\Theta} : \mathbb{R}^{d_Z} \mapsto \mathbb{R}^{d_Z \times d_Z}$ the volatility process and W^Z are (correlated) standard Wiener processes.

We consider the valuation of tradable options in the market and embedded options in insurance liabilities. We let V denote a set of d_V tradable options and analogously we let $\tilde{V}$ represent a set of $d_{\tilde{V}}$ embedded options that are part of the balance sheet of an insurance company. We are interested in the valuation of V and $\tilde{V}$ given the risk neutral model $Z_\Theta(u)$ with model parameters Θ. For notational convenience, we define the valuation mapping of V as $\eta_V : \mathbb{R}^{d_\Theta} \mapsto \mathbb{R}^{d_V}$. That is, the function η_V gives the values of the options in V with respect to the risk neutral model $Z_\Theta(u)$, given a set of model parameters Θ. Analogously, we define the mapping $\eta_{\tilde{V}}$ as the valuation function of the embedded options $\tilde{V}$. Prices for V are observable in the market, but this is not the case for the embedded options $\tilde{V}$.

Let us denote the pay-off functions of V and $\tilde{V}$ respectively by $p_V : \mathbb{R}^{d_Z} \times \mathbb{R} \mapsto \mathbb{R}$ and $p_{\tilde{V}} : \mathbb{R}^{d_Z} \times \mathbb{R} \mapsto \mathbb{R}$, so that the future cash flows of the options are given by, respectively,

$$C_V(u) = p_V(u, Z_\Theta), \quad C_{\tilde{V}}(u) = p_{\tilde{V}}(u, Z_\Theta).$$

The most well-known example is an equity put option with pay-off function $C_V(u) = \max\big(K - S(T_u - u), 0\big)$, where K denotes the strike level, T_u the option maturity and $S(u)$ the equity index. The cash flows $C_{\tilde{V}}(u)$ are in practice generated by the actuarial system of an insurer. Under $\mathbb{Q}$, option payments can be computed as expected discounted cash flows with respect to a chosen numéraire process, e.g. the bank account $\{M(u)\}_{u\in[0,T_u]}$,

with $M(u) = \exp\left(\int_0^u r(v)\,dv\right)$, where $\{r(u)\}_{u\in[0,T_u]}$ is the risk neutral risk-free interest rate process. We furthermore define the discount factor by $M^{-1}(u) = \frac{1}{M(u)}$. In the risk neutral world all individuals are indifferent to risk and expect to earn on all assets a return equal to the instantaneous risk free short rate. The risk neutral option value is then computed as the expected value under the risk neutral measure:

$$\eta_V(\Theta) = \mathbb{E}^{\mathbb{Q}}\left[\int_0^{T_u} M^{-1}(u)\,C_V(u)\,du\right], \quad \eta_{\tilde{V}}(\Theta) = \mathbb{E}^{\mathbb{Q}}\left[\int_0^{T_u} M^{-1}(u)\,C_{\tilde{V}}(u)\,du\right].$$

Given a set of observed option prices V, we would like to find the parameters Θ such that $\eta_V(\Theta)$ is the best approximation of V. We define the calibration function $\chi : \mathbb{R}^{d_V} \mapsto \mathbb{R}^{d_\Theta}$ by

$$\chi(V) = \arg\min_{\Theta\in\mathscr{D}_\Theta} \|V - \eta_V(\Theta)\|, \tag{5.2}$$

where $\mathscr{D}_\Theta \subset \mathbb{R}^{d_\Theta}$ denotes the relevant domain of Θ. So, $\Theta = \chi(V)$ is the best fit to V with respect to norm $\|\cdot\|$. We will use the Euclidean norm, which in particular means that all option values are weighted equally, but note that it can be useful to apply weights for extreme in or out of the money options. This calibration function is computed using numerical optimization techniques, because closed form solutions usually do not exist.

Once we have fitted the model parameters to the data, i.e. when we have computed $\chi(V)$, the values of the embedded options readily follow from

$$\eta_{\tilde{V}}(\chi(V)) = \mathbb{E}^{\mathbb{Q}}\left[\int_0^{T_u} M^{-1}(u)\,C_{\tilde{V}}(u)\,du\right]. \tag{5.3}$$

In Section 5.3.4 we discuss the numerical stability of this function, and in particular how it relates to η_V.

5

Nested simulation We let $\mathbb{P}$ denote the real world (physical) probability measure and $\{Y(t)\}_{t\in[0,T_t]}$ with $T_t \in \mathbb{R}_+$ be a d_Y-dimensional Markov process to model the uncertainty of the real world financial market. We let T_t denote the horizon of the real world simulation and t denotes time in the real world simulation. The market variables, i.e. the interest rate curve R and option instruments V, that are required for calibrating Θ follow hence stochastic processes: $\{R(t)\}_{t\in[0,T_t]}$ and $\{V(t)\}_{t\in[0,T_t]}$. Since the model parameters Θ of $Z_\Theta(u)$ are connected to these stochastic market variables as shown in Eq. (5.2), they cannot be fixed given the uncertainty of the real world market. Therefore, we model them by a stochastic process $\{\Theta(t)\}_{t\in[0,T_t]}$.

This means that at each time $t \in [0, T_t]$ a risk neutral model is parametrized, as

$$\Theta(t) \mapsto \{Z_{\Theta(t)}(u)\}_{u\in[0,T_u]}.$$

By modeling the stochastic process $\Theta(t)$ we generalize the nested simulation framework in [8] because we also take the implied volatility risk factor into account. In [8] the parameters in Θ are kept constant during the real world scenario simulation. The proposed modeling framework leads to a more realistic modeling of market IVs and model parameters of a risk neutral model. We note that the embedded option values also follow a

stochastic process $\left\{\tilde{V}(t)\right\}_{t\in[0,T_t]}$ given the link with Θ in Eq. (5.3), and the IV of an equity index option is obtained by solving for the volatility parameter of the Black-Scholes formula.

During numerical computation, we simulate a stochastic process given a discretization in the time domain $t \in [0, T_t]$; the random sample is called a scenario set. This can be applied to generate $N_{\mathbb{P}}$ independent paths of the real world variables $Y^j(t)$ with $j = 1,\ldots,N_{\mathbb{P}}$. However, in order to calculate $\eta_V(\Theta_t)$ using $N_{\mathbb{Q}}$ Monte Carlo simulations given a scenario set for $\Theta(t)$, we have in turn to simulate $\left\{Z_{\Theta^{(j)}(t)}(u)\right\}_{u\in[0,T_u]}$ for each $t \in [0, T_t]$ and scenario $j = 1,\ldots,N_{\mathbb{P}}$ of $\Theta(t)$. This is a nested simulation, see Chapter 1. The scenarios for $\left\{Z_{\Theta^{(j)}(t)}(u)\right\}_{u\in[0,T_u]}$ are the inner scenarios, those for $Y(t)$ the outer scenarios. In the nested modeling structure in Figure 1.1, the dark gray lines refer to the stochastic process $\{Z_{\Theta(t)}(u)\}_{u\in[0,T_u]}$ given a scenario set of $\Theta(t)$ and the black (and light gray) lines refer to the real world stochastic process $\{Y(t)\}_{t\in[0,T_t]}$.

In the proposed modeling framework, which we discuss in Section 5.3, we avoid the numerical calibration procedure to obtain Θ_t given the market variables Y_t, see Eq. (5.2). We make use of the class of State Space Hidden Markov (SSHM) models to directly model the stochastic process Θ_t. The observable process V_t is subsequently modeled via the valuation mapping $\eta_V(\Theta_t)$. We also discuss numerical stability of the model parameters with respect to changes in the market option data for calibration.

5.3. MODELING FRAMEWORK

We explain the proposed modeling framework. In Section 5.3.1 we first elaborate on the State Space Hidden Markov (SSHM) models. In Section 5.3.2 we discuss the estimation of SSHM models. In Section 5.3.3 we discuss our proposed modeling framework. An important requirement is to improve numerical stability of the model parameters Θ with respect to the option market data, which we discuss in Section 5.3.4.

5.3.1. STATE SPACE HIDDEN MARKOV MODELS

State Space Hidden Markov (SSHM) models provide a general and flexible framework for modeling time-series in a broad range of applications. A thorough introduction into the field can be found in the books [26, 42]. These models provide a flexible framework in financial modeling for stochastic processes. The most popular SSHM model is the linear Gaussian model, which can be estimated in closed form using the well-known Kalman filter [44]. However, in most cases the Kalman filtering method can not be applied due to the non-linear behavior of the observable process.

For notation[1] we let $X_n := X_{t_n}$ and $X_{1:n} := \{X_k\}_{k=t_1}^{t_n}$ (and similar for V). In an SSHM modeling structure we consider a hidden state process $X_n \in \mathcal{X}$ that evolves according to

$$X_1 \sim f_{1,X,\psi}(X_1), \quad X_n \mid \left(X_{n-1} = x_{n-1}\right) \sim f_{X,\psi}(x_n \mid x_{n-1}), \tag{5.4}$$

where X_n is the state at time t_n, $f_{1,X,\psi}$ a probability density function (PDF) and $f_{X,\psi}$ describes the transition from x_{n-1} to x_n, where x_n is a realization of X_n, with unknown parameters $\psi \in \Psi \subset \mathbb{R}^{d_\psi}$. The hidden states are conditionally independent.

[1] We use similarly notation as in [42].

We are interested in the hidden states $X_{1:N}$, but in our case we can only observe the $V_n \in \mathcal{V}$ process for $n = 1, \ldots, N$. Conditional on $X_{1:n}$, the observations V_n are assumed to be independent and their marginal densities are given by

$$V_n \mid \big(X_n = x_n\big) \sim f_{V,\psi}(v_n \mid x_n), \tag{5.5}$$

where $f_{V,\psi}$ describes the transition from x_n to v_n, with v_n a realization of V_n. The non-linear relation between x_n and v_n makes the SSHM model involved. In general, the PDFs $f_{X,\psi}$ and $f_{V,\psi}$ are chosen to be well-known distribution functions. We assume that all information in the observed data is explained by the underlying hidden process $X_{1:N}$, so that specification of the correlation between X_n and V_n is redundant. Such an assumption is standard in classical state space models. An overview of alternative representations is provided in [101].

5.3.2. ESTIMATION

State inference Given the observed process $V_{1:N} = v_{1:N}$ and fixed parameters ψ, we are interested in inferring the states $X_{1:N} = x_{1:N}$. In a Bayesian framework, inference of $X_{1:n}$ given a realization $v_{1:n}$ of $V_{1:n}$ relies on the posterior distribution $f_\psi(x_{1:n} \mid v_{1:n})$:

$$f_\psi(x_{1:n} \mid v_{1:n}) = \frac{f_\psi(v_{1:n}, x_{1:n})}{f_\psi(v_{1:n})} = \frac{f_\psi(v_{1:n} \mid x_{1:n})\, f_\psi(x_{1:n})}{f_\psi(v_{1:n})}, \tag{5.6}$$

with $f_\psi(v_{1:n} \mid x_{1:n})$ the likelihood, $f_\psi(x_{1:n})$ the prior, $f_\psi(v_{1:n})$ the evidence and $f_\psi(v_{1:n}, x_{1:n})$ is often referred to as the complete data likelihood. The prior, likelihood and evidence (also known as the marginal likelihood) are defined by

$$f_\psi(x_{1:n}) = f_{1,X,\psi}(x_1) \prod_{k=2}^{n} f_{X,\psi}\big(x_k \mid x_{k-1}\big), \tag{5.7}$$

$$f_\psi(v_{1:n} \mid x_{1:n}) = \prod_{k=1}^{n} f_{V,\psi}\big(v_k \mid x_k\big), \tag{5.8}$$

$$f_\psi(v_{1:n}) = \int f_\psi(v_{1:n}, x_{1:n})\, dx_{1:n}. \tag{5.9}$$

For the linear Gaussian model the posterior distribution is a Gaussian distribution whose mean and covariance can be computed in closed form [44]. For most non-linear non-Gaussian models, it is not possible to compute these distributions in closed-form and we need to employ numerical methods.

We are interested in the filtering and marginal likelihood computations, i.e. the sequential approximation of the distributions $\left\{f_\psi(x_{1:n} \mid v_{1:n})\right\}_{n\geq 1}$ and marginal likelihoods $\left\{f_\psi(v_{1:n})\right\}_{n\geq 1}$, because directly solving Eq. (5.6) is problematic due to the high dimensionality. In this sequential computation, one often relies on the so-called prediction and update equations [42]; the recursion satisfies the marginal distribution $f_\psi(x_n, v_{1:n})$.

The prediction and update equations are respectively given by

$$\text{Prediction}: \quad f_\psi(x_n \mid v_{1:n-1}) = \int f_{X,\psi}(x_n \mid x_{n-1})\, f_\psi(x_{n-1} \mid v_{1:n-1})\, dx_{n-1}, \tag{5.10}$$

$$\text{Update}: \quad f_\psi(x_n \mid v_{1:n}) = \frac{f_{V,\psi}(v_n \mid x_n)\, f_\psi(x_n \mid v_{1:n-1})}{f_\psi(v_n \mid v_{1:n-1})}, \tag{5.11}$$

where

$$f_\psi(v_n \mid v_{1:n-1}) = \int f_{V,\psi}(v_n \mid x_n)\, f_{X,\psi}(x_n \mid x_{n-1})\, f_\psi(x_{n-1} \mid v_{1:n-1})\, dx_{n-1:n}. \tag{5.12}$$

If we can compute $f_\psi(x_{1:n} \mid v_{1:n})$ and thus $f_\psi(x_n \mid v_{1:n})$ sequentially, then the evidence $f_\psi(v_{1:n})$ can also be evaluated recursively using

$$f_\psi(v_{1:n}) = f_\psi(v_1) \prod_{j=2}^{n} f_\psi(v_n \mid\ v_{1:n-1}),$$

where the $f_\psi(v_n \mid v_{1:n-1})$ are computed by Eq. (5.12).

Once the filtering stage is completed, one may smooth the filtered state process. Filtering is the estimation of the distribution of the current state x_n based upon the observations received up until current time $v_{1:n}$. Smoothing implies estimating the distribution of the state x_n given all observations up to some later time $v_{1:N}$, i.e. $f_\psi(x_n \mid v_{1:N})$. In general, smoothing is computationally more challenging than filtering. The trajectory estimates obtained by such methods, as a result of the additional information available, tend to be smoother than those obtained by filtering.

Non-linear Hidden Markov models Unfortunately, in most applications the SSHM model is non-linear and non-Gaussian. Therefore, various numerical approximations were developed over the years. The well-known Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) were introduced shortly after the Kalman Filter. In general, the UKF can acquire more accurate estimation results than the EKF, but can lead to serious errors for non-Gaussian distributions [68, 120]. Since we want our approach to be as general as possible we make use of so-called Sequential Monte Carlo (SMC) methods.

SMC methods represent a set of flexible and powerful simulation-based methods which provide samples distributed approximately according to posterior distributions and facilitate the approximate calculation of $f_\psi(v_{1:n})$. The main idea behind SMC methods is to obtain a large collection of weighted random samples, named particles, whose empirical distribution converges to the distribution we wish to sample from, the posterior distribution $f_\psi(x_{1:n} \mid v_{1:n})$. For this reason SMC methods are also referred to as Particle Filters in the filtering context. They have become a popular class of methods for inference in non-linear non-Gaussian state space models. An overview of the theory and applications of different SMC methods can be found in [42].

Combined state and parameter inference Because the transitional PDFs in Eqs. (5.4) and (5.5) are parametrized by ψ we are also interested in inferring these parameters given the observed process $V_{1:N} = v_{1:N}$. The Expectation-Maximization (EM) algorithm

[38] is an efficient method for combined state and parameter inference. The EM is an iterative method for maximizing the likelihood $f_\psi(v_{1:n})$ in Eq. (5.9). This method is useful when it is not possible to evaluate and optimize this likelihood directly. The method can be used for combined state and parameter inference in SSHM models. The algorithm is based on the insight that the auxiliary function

$$Q(\psi,\psi') = \int \log f_\psi(v_{1:n}, x_{1:n})\, f_{\psi'}(x_{1:n} \mid v_{1:n})\, dx_{1:n}, \tag{5.13}$$

may be used as a surrogate for $f_\psi(v_{1:n})$, because increasing $Q(\psi,\psi')$ forces an increase of $f_\psi(v_{1:n})$ [26]. The EM algorithm is initialized by $\psi_0 \in \Psi$ and $x_{1:N}^{(0)}$ and iterates between an expectation (E) step, which facilitates state inference, and a maximization (M) step for parameter inference:

$$\begin{aligned} &\text{E-step}: \text{Conditional on } \psi_{k-1} \text{ compute } Q(\psi,\psi_{k-1}),\\ &\text{M-step}: \text{Conditional on } x_{1:N}^{(k)} \text{ compute } \psi_k = \arg\max_{\psi\in\Psi} Q(\psi,\psi_{k-1}). \end{aligned}$$

This EM-algorithm generates a sequence $\left\{\psi_k, x_{1:N}^{(k)}\right\}_{k=0}^{N_{EM}}$, which converges to a stationary point of the likelihood for $N_{EM}\to\infty$, with the number of iterations $N_{EM}\in\mathbb{N}_+$. We solve the M-step of the EM algorithm using optimization techniques from the Matlab library.

Note that since the computation of the E-step includes a complicated multi-dimensional integral, we can approximate it by using Monte Carlo integration. Assuming we can sample from $f_{\psi_{k-1}}(x_{1:N}\mid v_{1:N})$ using a particle filtering method, we replace the E-step by the simulation of $N_{par}\in\mathbb{N}_+$ realizations $\{X_{1:N}^j\}_{j=1}^{N_{par}}$ from $f_{\psi_{k-1}}(x_{1:N}\mid v_{1:N})$ and the computation of

$$\tilde{Q}_k(\psi) = \frac{1}{N_{par}} \sum_{j=1}^{N_{par}} \log f_\psi(X_{1:N}^j, v_{1:N}).$$

This leads to the Monte Carlo EM algorithm (MCEM). Unfortunately, a drawback of this approach is that it requires the number of particles N_{par} to grow with each new iteration of the algorithm $k \le N_{EM}$ [26]. Besides this, we need to sample a whole new set of realizations of the hidden states $\{X_{1:N}^j\}_{j=1}^{N_{par}}$ at each iteration, that are not re-used in later iterations. The Stochastic Approximation EM (SAEM) algorithm [37] makes more efficient use of the simulated variables by replacing $\tilde{Q}_k(\theta)$ with a stochastic averaging procedure

$$\hat{Q}_k(\psi) = (1-\gamma_k)\hat{Q}_{k-1}(\psi) + \gamma_k \left(\frac{1}{N_{par}} \sum_{j=1}^{N_{par}} \log f_\psi(X_{1:N}^j, v_{1:N}) \right),$$

where $\{\gamma_k\}_{k\ge 0}$ is a decreasing sequence of weights that satisfy $\sum\gamma_k=\infty$ and $\sum\gamma_k^2<\infty$. As $k\to\infty$, the SAEM algorithm converges to a local maximum of the likelihood function [37]. The computational advantage of the SAEM algorithm is especially significant in problems where maximization is much cheaper than simulation.

The MCEM and SAEM methods cannot be applied directly, because for non-linear non-Gaussian SSHM models it is not possible to directly sample from the posterior distribution. Using a particle filter method for sampling from the posterior distribution leads to an SMC-analogue of the previous methods, the PSEM method described in [115]. If we take the SAEM approach, it is sufficient to generate a single sample each iteration.

In [85] it is shown that for convergence of the SAEM algorithm, it is not necessary to sample exactly from the posterior distribution. We can also sample from a family of so-called Markov kernels $\{MK_\psi(x_{1:N} \mid x'_{1:N})\}_{\psi\in\Psi}$ on $\mathscr{X}^N$ that leaves the family of posterior distributions invariant. Assume that we have such a family and let in iteration $k \leq N_{EM}$ of the SAEM method $x_{1:N}[k-1]$ be the previous draw from the Markov kernel. We then sample $X_{1:N}[k] \sim MK_{\psi_{k-1}}(x_{1:N} \mid x_{1:N}[k-1])$ and update $\hat{Q}$ according to

$$\hat{Q}_k(\psi) = (1-\gamma_k)\hat{Q}_{k-1}(\psi) + \gamma_k \log f_\psi(X_{1:N}[k], v_{1:N}).$$

The next approximation of ψ is then obtained by maximizing this quantity w.r.t. $\psi \in \Psi$ (the M-step) using conditional particle filters (CPF) as the required Markov kernel. We use the SAEM method using conditional particle filters (CPF), which is referred to as CPF-SAEM [90, 91]. In this method the Markov kernel is constructed by running a SMC sampler in which one particle trajectory $x'_{1:N}$ is specified a priori, a so-called CPF. We can think of this reference trajectory as guiding the simulated particles to a relevant region of the state space. The path $x'_{1:N}$ is ensured to survive all re-sampling steps.

5.3.3. MODELING CONCEPT

For notational convenience, we denote time $t \leq T_t$ and $u \leq T_u$ via subscript. The process Y_t models the uncertainty of the real world financial market, and consists amongst others of the processes R_t and V_t. To avoid a numerical calibration procedure to obtain Θ_t, see Eq. (5.2), we model the process Θ_t directly. Given the process Θ_t, we compute the observable process by $V_t = \eta_V(\Theta_t)$ and the embedded option values by $\tilde{V}_t = \eta_{\tilde{V}}(\Theta_t)$.

Modeling the joint behavior of possibly a large number of variables in Θ_t without any constraints is generally problematic, because estimating an underlying model may result in a high-dimensional problem. To reduce the dimension of the problem, we assume the d_Θ-dimensional process $\{\Theta_t\}_{t\in[0,T_t]}$ to be generated by a latent (hidden) d_X-dimensional process $\{X_t\}_{t\in[0,T_t]} \in \mathscr{X}$, with

$$\Theta_t = \Upsilon(X_t), \quad \Upsilon : \mathbb{R}^{d_X} \mapsto \mathbb{R}^{d_\Theta}. \tag{5.14}$$

The dimension of the calibration problem is controlled by $d_X \leq d_\Theta$. We aim to vary the model parameters that are most significant with respect to the calibration fit, hence by keeping d_X as small as possible. The parameters of Υ (including X_t) are still determined by maximizing the calibration fit to the option market data V_t. This is beneficial for interpretation and can improve numerical stability.

When $d_X \leq d_\Theta$, we achieve an effective dimension-reduction, since the model parameters Θ_t are then driven by a lower-dimensional process X_t. Several ways of defining the mapping Υ exist and three requirements for Υ should be taken into account. First, the mapping should lead to a satisfactory fit to the market data. Second, the mapping

should lead to a realistic representation of the model parameters. Third, the mapping should lead to numerically stable results. A basic assumption is a linear mapping between X_t and Θ_t in Eq. (5.15), i.e.

$$\Theta_t = \Upsilon(X_t) := a_\Upsilon X_t + b_\Upsilon, \quad \left(a_\Upsilon \in \mathbb{R}^{d_X \times d_\Theta}, b_\Upsilon \in \mathbb{R}^{d_\Theta}\right) \tag{5.15}$$

where a_Υ and b_Υ are unknown parameters. The mapping in Eq. (5.14) can be split into a dynamic part X_t and static part $\{a_\Upsilon, b_\Upsilon\}$. That is, the static part does not change over time and is calibrated once. Calibration to available market data then only takes place with respect to the dynamic part (while the static part is kept fixed). This reduces computation times of calibration, because (by construction) the parameter space is smaller. In Section 5.4 we estimate these dynamic and static parts for the Heston model.

The mapping Υ should be specified in such a way that Θ_t belongs to domain $\mathscr{D}_\Theta \subset \mathbb{R}^{d_\Theta}$. For example, the correlation matrix should be positive definite, the correlation parameters should be in $[-1,1]$ and volatility parameters should be non-negative. To ensure that Υ maps values of X_t to the correct domain for Θ_t, we truncate values that are outside the domain $\mathscr{D}_\Theta$, i.e.

$$\Theta_t \in \left[\mathscr{D}_{\Theta,min}, \mathscr{D}_{\Theta,max}\right],$$

where the minimum and maximum domains are respectively denoted by $\mathscr{D}_{\Theta,min} \in \mathbb{R}^{d_\Theta}$ and $\mathscr{D}_{\Theta,max} \in \mathbb{R}^{d_\Theta}$.

Using this approach we compute the process V_t by the following composite function:

$$V_t = \eta_V(\Upsilon(X_t)), \quad \left(\Theta_t = \Upsilon(X_t)\right). \tag{5.16}$$

Next, we let the hidden states X_t evolve through time via the transition PDF $f_{X,\psi}$ and given a realization $X_t = x_t$, the option price $V_t = v_t$ is computed via Eq. (5.16). We model the (calibration) error between the valuation function η_V and the observed process V_t as a stochastic process by means of the PDF $f_{V,\psi}$. This is required because the underlying risk neutral model is generally not able to perfectly fit the market option data. The Black-Scholes model can for example only model flat IV surfaces, whereas the IV typically varies per strike and maturity. Secondly, we introduce an error by reducing the number of degrees of freedom in the risk neutral model for calibration: $d_X \leq d_\Theta$. A basic assumption for $f_{V,\psi}$ would be a normal distribution. In this case the error $v_t - \eta_V(\Upsilon(x_t))$ is symmetric around zero.

The proposed SSHM model is given by

$$\text{State}: \quad X_1 \sim f_{1,X,\psi}(X_1), \quad X_n \,|\, \left(X_{n-1} = x_{n-1}\right) \sim f_{X,\psi}(x_n \,|\, x_{n-1}), \tag{5.17}$$

$$\text{Observation}: \quad V_n \,|\, \left(X_n = x_n\right) \sim f_{V,\psi}(v_n \,|\, \eta_V(\Upsilon(x_n))), \tag{5.18}$$

The SSHM model defined in Eqs. (5.17)-(5.18) is non-linear due to the non-linearity of the valuation function η_V. Hence, analytic solutions do not exist. We condition on the state x_n (as in Eq. (5.5)), but we incorporate the valuation function η_V in the PDF $f_{V,\psi}$, which we highlight by conditioning on $\eta_V(\Upsilon(x_n))$, $n \leq N$.

Given the observed process $V_{1:N} = v_{1:N}$, we wish to infer states $X_{1:N} = x_{1:N}$ and parameters ψ, where the static part of Υ (i.e. a_Υ and b_Υ) are part of ψ. We use the EM algorithm (see Section 5.3.2) for combined state and parameter inference. That is, the

dynamic part of Y, the hidden states X_t, is determined via the expectation step and the fixed model parameters ψ are determined via the maximization step. We aim to keep the dimension d_X as low as possible, but by optimizing the fit to the market data via maximizing the Q-function in Eq. (5.13).

5.3.4. NUMERICAL STABILITY

We wish to achieve a satisfactory fit to the option market data V by using as few parameters as possible, but we also wish to achieve stable valuations of the embedded options $\tilde{V}$, i.e. we wish to avoid overfitting issues. For notational convenience we drop the subscript t in this section. Numerical stability is important in practice because the calibrated model parameters are used by insurers to value the (out-of-sample) embedded options on their balance sheet. If changes in the market data imply undesired large changes in the model parameters this can result in noisy valuations of embedded options. Overfitting occurs when a model has too many parameters relative to the number of observations or has an incorrect model structure. Overfitting may lead to poor predictive performance, as it typically results in excessive reactions to minor fluctuations in the option market data used for calibration.

On the other hand, a modeling setup that does not react to changes in the market data at all may be numerically very stable, but will not result in realistic embedded option values. We therefore also require that our modeling setup fits market prices well enough. So, there is a trade-off to be made. Below we will focus on the numerical stability, but we emphasize that the quality-of-fit should also be checked when evaluating the modeling approach.

In Section 5.2.1 we have explained that the model parameters are determined via the calibration function χ in Eq. (5.2). These calibrated model parameters are then used to compute the embedded option values via Eq. (5.3). By applying the chain rule to the composite function $\eta_{\tilde{V}} \circ \chi$ in Eq. (5.3) we have

$$D_{\eta_{\tilde{V}} \circ \chi} = D_{\eta_{\tilde{V}}} D_{\chi},$$

where D denotes the Jacobian matrix. Hence, the sensitivity of $\eta_{\tilde{V}}$ to the option market data V can be decomposed to the sensitivity of $\eta_{\tilde{V}}$ to Θ and the sensitivity of Θ to V.

We use the operator norm $\|\cdot\|^{op}$ to quantify numerical stability, where we recall that the operator norm is defined as

$$\|A\|^{op} = \sup\left\{ \frac{\|Ax\|}{\|x\|} : x \in \mathbb{R}^n,\, x \neq 0 \right\},$$

for any linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, and any $n, m > 0$. We use the Euclidean norm to compute $\|\cdot\|$. We also recall that the operator norm equals the largest singular value of A. The operator norm quantifies the direction that is amplified the most by A.

A small change in the market option prices δV will result in a change in embedded option values $D_{\eta_{\tilde{V}} \circ \chi} \delta V$, up to first order. The operator norm therefore provides an upper bound to how strong embedded option values change as a result of a small change in observed market prices, and is therefore a measure of numerical stability. Using the chain rule and the Cauchy-Schwartz inequality, we obtain an upper bound of $\|D_{\eta_{\tilde{V}} \circ \chi}\|^{op}$:

$$\|D_{\eta_{\tilde{V}} \circ \chi}\|^{op} = \|D_{\eta_{\tilde{V}}} D_{\chi}\|^{op} \leq \|D_{\eta_{\tilde{V}}}\|^{op} \|D_{\chi}\|^{op}.$$

The sensitivity of $\eta_{\tilde{V}}$ to $\Theta = \chi(V)$ follows from the actuarial system of an insurer (see Section 5.2.1). We focus on the sensitivity of the calibration function χ in Eq. (5.2), as we assume the actuarial system of the insurance company as given. More concretely, to keep $\|D_{\eta_{\tilde{V}}\circ\chi}\|^{op}$ as small as possible, we require the operator norm $\|D_\chi\|^{op}$ to be as small as possible for stable calibrations of Θ.

Since we have no closed form expression for χ, and numerical calibration of its derivative is expensive, we relate D_χ to D_{η_V} for which we don't have to perform numerical optimizations. In order to do this, we assume that η_V is locally an embedding near Θ, so in particular D_{η_V} has maximal rank. This is a very modest assumption because a zero vector of D_{η_V} would be a salient over-parametrization: a change in parameter would not lead to any change in option values up to first order.

We proceed to decompose the tangent space $T_{\eta_V(\Theta)}$ into the image of $D_{\eta_V(\Theta)}$, $\mathrm{im}\,D_{\eta_V(\Theta)}$, and its ortho-complement, $(\mathrm{im}\,D_{\eta_V(\Theta)})^\perp$. Let us denote the orthogonal projection onto $\mathrm{im}\,D_{\eta_V(\Theta)}$ by π_{η_V}

$$\pi_{\eta_V}(\mathrm{im}\,D_{\eta_V(\Theta)})^\perp = 0, \quad \pi_{\eta_V}(x) = x,\ \forall x \in \mathrm{im}\,D_{\eta_V(\Theta)}.$$

Since $D_{\eta_V(\Theta)}$ has a trivial kernel, the following defines a pseudo-inverse of $D_{\eta_V(\Theta)}$:

$$D^+_{\eta_V(\Theta)} = (D^\top_{\eta_V(\Theta)} D_{\eta_V(\Theta)})^{-1} D^\top_{\eta_V(\Theta)}.$$

This is also a left-inverse of $D_{\eta_V(\Theta)}$, and moreover $D_{\eta_V(\Theta)} D^+_{\eta_V(\Theta)}$ is the orthogonal projection on the image of $D_{\eta_V(\Theta)}$, i.e.

$$D_{\eta_V(\Theta)}\, D^+_{\eta_V(\Theta)} = \pi_{\eta_V}.$$

We now claim that

$$D_{\chi(V)} = D^+_{\eta_V(\Theta)}, \quad \big(\Theta = \chi(V)\big). \tag{5.19}$$

To see this, consider a first order perturbation δV of the option market data V. This decomposes into

$$\delta V = \pi_{\eta_V}(\delta V) + (\delta V)^\perp, \quad (\delta V)^\perp := \delta V - \pi_{\eta_V}(\delta V) \in (\mathrm{im}\,D_{\eta_V})^\perp.$$

Let us write $\delta\Theta = D^+_{\eta_V(\Theta)}(\delta V)$, so that

$$D_{\eta_V(\Theta)}\delta\Theta = D_{\eta_V(\Theta)} D^+_{\eta_V(\Theta)}(\delta V) = \pi_{\eta_V}(\delta V).$$

Using this, we have up to first order that

$$\begin{aligned}
\chi(V+\delta V) &= \arg\min_{\Theta\in\mathscr{D}_\Theta} \|V + \delta V - \eta_V(\Theta)\|, \\
&= \arg\min_{\Theta\in\mathscr{D}_\Theta} \|V + \delta V - D_{\eta_V(\Theta)}\delta\Theta - \eta_V(\Theta - \delta\Theta)\| + \mathscr{O}^2(\delta V), \\
&= \arg\min_{\Theta\in\mathscr{D}_\Theta} \|V + (\delta V)^\perp - \eta_V(\Theta - \delta\Theta)\| + \mathscr{O}^2(\delta V).
\end{aligned} \tag{5.20}$$

5

Now, to determine for which Θ the minimum is attained, we observe that

$$\begin{aligned}\left.\frac{d}{d\Theta}\right|_{\Theta+\delta\Theta} \|V+(\delta V)^{\perp}-\eta_V(\Theta-\delta\Theta)\|^2 &= -2\Big(V+(\delta V)^{\perp}-\eta_V(\Theta)\Big)D_{\eta_V(\Theta)},\\ &= -2\big(V-\eta_V(\Theta)\big)D_{\eta_V(\Theta)},\\ &= \left.\frac{d}{d\Theta}\right|_{\Theta}\|V-\eta_V(\Theta)\|^2.\end{aligned}$$

Since the last term equals zero for $\Theta=\chi(V)$, it follows that the minimum in Eq. (5.20) is attained at $\Theta=\chi(V)+\delta\Theta$, and so

$$\chi(V+\delta V)=\chi(V)+\delta\Theta+\mathcal{O}^2(\delta V),$$

which proves Eq. (5.19).

Note that this implies that the singular values of $D_{\chi(V)}$ are equal to the reciprocals of the singular values of $D_{\eta_V(\Theta)}$ (which is a general fact about pseudo-inverses). Consequently,

$$\|D_\chi\|^{op}=\sigma_{D_\chi,max}=\frac{1}{\sigma_{D_{\eta_V},min}},$$

where $\sigma_{D_\chi,max}$ and $\sigma_{D_{\eta_V},min}$ are respectively the largest singular value of D_χ and smallest singular value of D_{η_V}. We conclude that for numerical stability of the calibration part, we would like $\sigma_{D_{\eta_V},min}$ to be as large as possible. In Section 5.4 we compute $\sigma_{D_{\eta_V},min}$ for the Heston model.

Dimension reduction We use the parametrization Υ in Eq. (5.14) to reduce the dimension of Θ. More precisely, we use the following adjusted calibration function

$$\chi^{\Upsilon}(V)=\Upsilon\left(\arg\min_{x\in\mathbb{R}^{d_X}}\|V-\eta_V(\Upsilon(x))\|,\right).$$

We assume that Υ is an embedding, and therefore $\eta_V\circ\Upsilon$ as well, so that we can apply exactly the same arguments as used above to prove Eq. (5.19), but now applied to $\eta_V\circ\Upsilon$ instead of η_V. We omit the points in which derivatives are taken for brevity. This results in

$$D_{\chi^{\Upsilon}}=D_\Upsilon\, D\left(\arg\min_{x\in\mathbb{R}^{d_X}}\|V-\eta_V(\Upsilon(x))\|,\right)=D_\Upsilon\,(D_{\eta_V\circ\Upsilon})^{+}.$$

We argue that using Υ to parametrize Θ generically improves numerical stability. To this end, observe that $D^{+}_{\eta_V}D_{\eta_V}=I$, and so

$$\begin{aligned}D_{\chi^{\Upsilon}} &= D^{+}_{\eta_V}D_{\eta_V}D_\Upsilon\,(D_{\eta_V\circ\Upsilon})^{+}=D^{+}_{\eta_V}D_{\eta_V\circ\Upsilon}\,(D_{\eta_V\circ\Upsilon})^{+},\\ &= D_\chi\,\pi_{\eta_V\circ\Upsilon},\end{aligned}\tag{5.21}$$

where we have denoted the orthogonal projection operator onto the image of $D_{\eta_V\circ\Upsilon}$ by $\pi_{\eta_V\circ\Upsilon}$. Eq. (5.21) can be used to compare the operator norms of D_χ and $D_{\chi\circ\Upsilon}$. On the

one hand we have that

$$\begin{aligned}\|D_\chi\|^{op} &= \|D_\chi \circ \pi_{\eta_V}\|^{op} = \sup\left\{\frac{\|D_\chi \pi_{\eta_V} x\|}{\|x\|} : x \in \mathbb{R}^{d_X},\, x \neq 0\right\},\\ &= \sup\left\{\frac{\|D_\chi x\|}{\|x\|} : x \in \mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}},\, \mathrm{x} \neq 0\right\},\end{aligned} \tag{5.22}$$

whereas we also have

$$\begin{aligned}\|D_{\chi^\Upsilon}\|^{op} &= \sup\left\{\frac{\|D_\chi \pi_{\eta_V \circ \Upsilon} x\|}{\|x\|} : x \in \mathbb{R}^{d_X},\, x \neq 0\right\},\\ &= \sup\left\{\frac{\|D_\chi x\|}{\|x\|} : x \in \mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}\circ\Upsilon},\, \mathrm{x} \neq 0\right\}.\end{aligned} \tag{5.23}$$

Since $\mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}\circ\Upsilon} \subseteq \mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}}$, we can conclude from Eqs. (5.22) and (5.23) that

$$\|D_{\chi^\Upsilon}\|^{op} \leq \|D_\chi\|^{op}. \tag{5.24}$$

Since the intention of using Υ is to reduce the dimension of the parameter space, we can safely assume that $\mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}\circ\Upsilon}$ is a strict subset of $\mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}}$, in practice in fact of high co-dimension. Under this assumption, Eq. (5.24) will generically be a strict inequality. The only case in which this does not hold, is when the right-singular vector of D_χ, corresponding to the largest singular value, lies in the image of $D_{\eta_V\circ\Upsilon}$. Obviously, this is generically not the case for a subspace of positive co-dimension.

The extent to which $\|D_{\chi^\Upsilon}\|^{op}$ is smaller than $\|D_\chi\|^{op}$, and so the extent to which we have improved numerical stability using Υ as parametrization, of course depends very much on the details of the situation. In the special case that we use only a single dimension to parametrize Θ, i.e. $d_X = 1$, we can make the operator norm $\|D_{\chi^\Upsilon}\|^{op}$ more explicit. Let us now consider this univariate case.

5

We use Eq. (5.23) to compute the operator norm, so suppose $x \in \mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}\circ\Upsilon}$. This is of the form $x = \lambda\, D_{\eta_V} D_\Upsilon$ for some non-zero $\lambda \in \mathbb{R}$, so

$$D_{\chi^\Upsilon} x = D_\Upsilon\, (D_{\eta_V\circ\Upsilon})^+ x = D_\Upsilon\, (D_{\eta_V} A)^+\, (\lambda D_{\eta_V} D_\Upsilon) = \lambda D_\Upsilon,$$

because $(D_{\eta_V} D_\Upsilon)^+ D_{\eta_V} D_\Upsilon = 1$. It follows that

$$\frac{\|D_{\chi^\Upsilon} x\|}{\|x\|} = \frac{|\lambda|\|D_\Upsilon\|}{|\lambda|\|D_{\eta_V} D_\Upsilon\|} = \frac{\|D_\Upsilon\|}{\|D_{\eta_V} D_\Upsilon\|},$$

for any $x \in \mathrm{im}\,\mathrm{D}_{\eta_\mathrm{V}\circ\Upsilon}$. By Eq. (5.23) it then follows that

$$\|D_{\chi^\Upsilon}\|^{op} = \frac{\|D_\Upsilon\|}{\|D_{\eta_V} D_\Upsilon\|}. \tag{5.25}$$

Given the formulation of Υ in Eq. (5.15), we have

$$D_\Upsilon = a_\Upsilon \in \mathbb{R}^{d_X \times d_\Theta}.$$

In Section 5.4 we will use the expression in Eq. (5.25) to quantify numerical stability of the proposed method.

5.4. APPLICATION

We perform a calibration to historical equity option market data. We use the Heston model [70] for numerical experiments. We calibrate by using a basic calibration method ($Basic$), wherein we calibrate the full parameter space Θ, and the proposed method ($Proposed$) in respectively Sections 5.4.1 and 5.4.2. We compare the results and we also analyze the numerical stability. In Section 5.4.3 we compare the calibrated models by computing the present and future embedded option values of a fictive insurance company.

The Heston model Under the Heston model we model the stock index S_u, the variance process v_u and the money market account M_u under the risk neutral measure $\mathbb{Q}$ by

$$\begin{cases} dS_u = & (r-q)\,S_u\,du + & \sqrt{v_u}\,S_u\,dW_u^S, & S_0 = 1, \\ dv_u = & \kappa_v(\bar{v}-v_u)\,du + & \sigma_v\sqrt{v_u}\,dW_u^v, & v_0 \geq 0, \\ dM_u = & r\,M_u\,du, & & M_0 = 1, \end{cases} \tag{5.26}$$

with $Z_u = \{S_u, v_u, M_u\}$ and $u \in [0, T_u]$, r denotes the risk free interest rate, q the dividend yield, κ_v the mean reversion, $\bar{v}$ the long term variance level, σ_v the volatility of variance and ρ the correlation between the Wiener processes dW^S and dW^v.

We consider the dividend yield to be zero and the interest rate r is equal to the observed interest rate in the market, such that we have $\Theta = \{\kappa_v, v_0, \bar{v}, \sigma_v, \rho\}$. The volatility parameter σ_v affects the kurtosis (peak) of the probability distribution function of equity (log) returns. The lower the volatility of variance parameter, the higher the kurtosis (peak). The correlation parameter ρ affects the skewness of the probability distribution function of equity (log) returns. The lower the correlation parameter (between the stock index and the variance process), the higher the skewness, i.e. heavy tails to the left. Hence, these parameters affect the modeling of IVs in the strike dimension. In equity, ρ is often negative. The Heston model enables semi-analytic pricing of equity index put/call options via the so-called COS method [47]. We use this method for valuation.

The calibration data We use IV surfaces of S&P-500 equity index options. Multiple S&P-500 index options are quoted with different strike levels and maturities, but we use the most liquid option data. For calibration we use strikes $K = \{0.8, 1, 1.2\}$ and maturities $T = \{0.25, 1, 2\}$, with $d_K = 3$ and $d_T = 3$ and hence $d_V = 9$. For notation we use $V_t^{i,j}$ with $i = 1, \ldots, d_K$ and $j = 1, \ldots, d_T$ to refer to the option value with strike K_i and maturity T_j. We define a discrete set of (historical) times $\mathcal{T} = \{t_1, t_2, \ldots, t_N\}$ for which we can observe the process V_t for $t \in \mathcal{T}$, where N denotes the number of observations. We assume that the index $n = 1, \ldots, N$ refers to time $t_n \in \mathcal{T}$. We use historical data available from June 2005 until March 2014 with a monthly frequency, which leads to $N = 106$ observations.

In the literature much research has been performed to study the empirical behavior of IV surfaces, see for example [31] and the references therein. It turns out that for a given maturity the IVs increase when relative strike levels decrease[2]. Long maturing options

[2] Possible explanations for this phenomenon are the (see [80]) negative correlation between asset returns and volatility changes (leverage effect). Big jumps in the asset (spot) price tend to be downwards rather than upwards.

are less volatile than short maturing options. The volatility of the IV is level dependent. That is, the higher the (average) level of the IV, the more skewed the IV surface and hence more volatile. This may be due to the fact that investors sell their call options and buy put options for protection in case of high volatility, i.e. in times of crisis.

Given these empirical facts, IV surfaces tend to dynamically change over time. That is, the skew or smile in the strike dimension and the term structure of IVs changes dynamically over time. In times of crisis the (average) IV level is typically higher, the mean reversion from short term IVs to long term IVs is faster and the IV surface is more skewed. Opposite IV structures hold in less volatile periods. The VIX index is a popular measure of the implied volatility of short maturing S&P 500 index options. It represents a measure of the market's expectation of equity market volatility over the next 30 day period.

5.4.1. THE BASIC CALIBRATION APPROACH

In the basic calibration approach, we calibrate the full parameter space Θ_t at each historical time t by using the calibration function χ in Eq. (5.2). We let $\eta_{i,j}$ refer to the option with strike K_i and maturity T_j with $i = 1,\dots,d_K$ and $j = 1,\dots,d_T$. We set the calibration bounds equal to $\mathscr{D}_{\Theta,min} = \{0.01, 0.001, 0.001, 0.1, -1\}$ and $\mathscr{D}_{\Theta,max} = \{5, 0.5, 0.5, 1, -0.1\}$. We specify small lower limits for the volatility parameters to avoid numerical difficulties and the upper limits are specified based on expert opinion. For each time we (independently) calibrate 5 parameters to obtain the best fit with the market data, so in total we calibrate $5 \times 106 = 530$ parameters.

Quality-of-fit The Heston model fits well to the available data. The R^2 errors are > 0.99 for all times $t \in \mathscr{T}$, see Figure 5.4a, which is highly satisfactory. To gain insight in the quality-of-fit of separate parts of the option market data, we show in Table 5.1 the sum squared errors (SSE) per option and for all $t \in \mathscr{T}$.

Table 5.1: Quality-of-fit of Heston model.

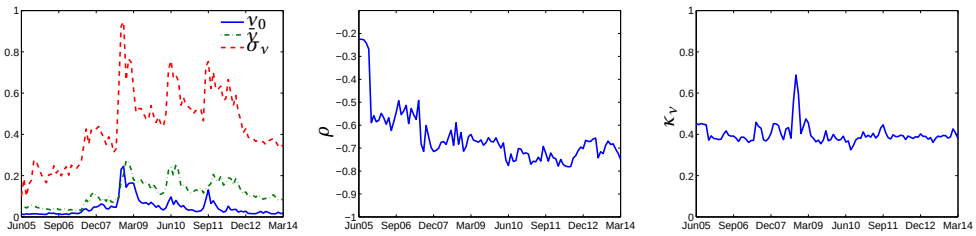
	$K = 0.8$			$K = 1.0$			$K = 1.2$		
	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$
Av. SSE ($\times 10^{-5}$)	0.25	0.15	0.15	0.35	0.26	0.04	0.08	0.09	0.05
Max. SSE ($\times 10^{-4}$)	0.49	0.07	0.11	0.16	0.13	0.02	0.20	0.09	0.09

The fit is worse for shorter maturing options. This is due to the fact that the skew in the maturity dimension from the Heston model decreases linearly [60], whereas the market data contains more curved behavior. The Heston model is hence not powerful enough to model such behavior.

Numerical stability of Θ The calibrated model parameters Θ are visualized in Figure 5.1.

Figure 5.1: Model parameters Θ_t.

(a) v_0, $\bar{v}$ and σ_v (b) ρ (c) κ_v



The parameters κ_v and ρ show a constant, but spiky, behavior around 0.4 and -0.7 respectively. The parameters v_0, $\bar{v}$, σ_v show common behavior, see Figure 5.1a, with correlations

$$\begin{pmatrix} 1 & & \\ \rho_{v_0,\bar{v}} & 1 & \\ \rho_{v_0,\sigma_v} & \rho_{\bar{v},\sigma_v} & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ 0.65 & 1 & \\ 0.76 & 0.90 & 1 \end{pmatrix}. \tag{5.27}$$

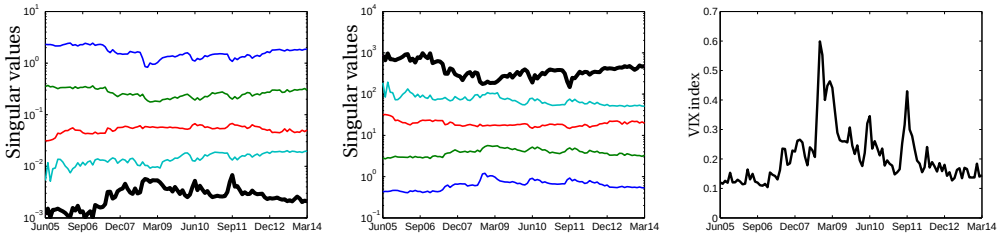
The determinant of this correlation matrix is close to zero: 0.08. The determinant of a correlation matrix approaches zero when some of the variables are strongly correlated. The long term variance level shows lagged behavior relative to the initial variance, the lag one cross-correlation is 0.75.

We recall from Section 5.3.4 that $\|D_{\eta_V}\|^{op}$ measures the numerical stability of V to Θ and $\|D_\chi\|^{op}$ measures the numerical stability of Θ to V. The operator norm is equal to the maximum singular value and, in particular, we have showed that the singular values of D_χ are equal to the reciprocals of the singular values of D_{η_V}. This means that, the higher the minimum singular value $\sigma_{\eta_V,min}$ of D_{η_V}, the lower the maximum singular value $\sigma_{\chi,max}$ of D_χ and hence the better the numerical stability.

In Figure 5.2a we show the singular values of D_{η_V}, where we emphasize the minimum singular value $\sigma_{\eta_V,min}$ in black. In Figure 5.2b we show the singular values of D_χ, where we emphasize the maximum singular value $\sigma_{\chi,max}$ in black. For reference, we show the VIX index in Figure 5.2c. Using the VIX index we gain insight in the dynamic behavior of numerical stability.

Figure 5.2: Numerical stability.

(a) Singular values of D_{η_V} (b) Singular values of D_χ (c) The VIX index



We find that there is less variation in the Heston model in volatile periods and vice versa. This means that the Heston model is less sensitive to market changes in volatile periods. There is a significant negative correlation between the market volatility, represented by the VIX index, and maximum singular value of D_χ. The Pearson correlation coefficient over this period is -0.71.

In times of a low VIX index, the IV surfaces in the market are typically relatively flat. The Heston model has too many degrees of freedom in this case and is therefore sensitive to changes in the market data, see [70]. In times of a high VIX index, the IV surfaces are typically not flat. In this case more degrees of freedom in the Heston model are used to obtain a good fit. Therefore, in times of high values of the VIX index, the Heston model is less sensitive to changes in the market data.

5.4.2. THE PROPOSED CALIBRATION APPROACH

We apply the proposed methodology in Section 5.3 to the Heston equity model in Eq. (5.26). The methodology can be applied to other models as well. By using the proposed methodology we aim at reducing the number of parameters in the risk neutral model to improve stability, but in such a way that the fit to the option market data is satisfactory.

Model specification We use the mapping Υ in Eq. (5.15) to reduce the dimension of Θ. We consider a 1-dimensional hidden state process, i.e. $d_X = 1$ and vectors $a_\Upsilon, b_\Upsilon \in \mathbb{R}^5$. To model the hidden state process X_t and the observable process V_t we need to specify the PDFs $f_{X,\psi}$ and $f_{V,\psi}$. For both PDFs, we consider a (simple) normal distribution and an (advanced) non-normal distribution. This results in 4 cases, which we describe below.

We consider the Ornstein-Uhlenbeck (OU) and Cox-Ingersoll-Ross (CIR) models [18] for modeling the hidden process $X_{1,t}$ and $X_{2,t}$, which are respectively presented by the following Stochastic Differential Equations (SDEs):

$$\begin{aligned} dX_{1,t} &= \kappa_{1,X}\left(\bar{X}_1 - X_{1,t}\right)dt + \sigma_{1,X}\,dW^{X_1}(t), \quad X_{1,0} \in \mathbb{R}, \\ dX_{2,t} &= \kappa_{2,X}\left(\bar{X}_2 - X_{2,t}\right)dt + \sigma_{2,X}\sqrt{X_{2,t}}\,dW^{X_2}(t), \quad X_{2,0} \geq 0, \end{aligned} \tag{5.28}$$

where $\kappa_{\cdot,X} \in \mathbb{R}_+$ is a mean reversion parameter, $\bar{X}_\cdot \in \mathbb{R}_+$ a long term state parameter and $\sigma_{\cdot,X} \in \mathbb{R}_+$ a volatility parameter (with $\cdot = 1,2$). The transitional PDFs for respectively the

OU and CIR model are known in analytic form:

$$f_{X,\psi}(x_{1,n}\,|\,x_{1,n-1}) = f_{norm}\Big(x_{1,n}; x_{1,n-1}\exp\left(-\kappa_{1,X}\Delta t\right) + \frac{\bar{X}_1}{\kappa_{1,X}}\left(1-\exp\left(-\kappa_{1,X}\Delta t\right)\right), \frac{\sigma_{1,X}^2}{2\kappa_{1,X}}(1-\exp\left(-2\kappa_{1,X}\Delta t\right))\Big), \tag{5.29}$$

$$f_{X,\psi}(x_{2,n}\,|\,x_{2,n-1}) = \frac{1}{C} f_{nccs}\left(x_{2,n}/C; d, \lambda_{X,t}\right), \tag{5.30}$$

where f_{norm} denotes the PDF of a normal distribution and f_{nccs} denotes the PDF of a Non-Central Chi-Squared (NCCS) distribution where

$$C = \frac{\sigma_{2,X}^2(1-\exp\left(-\kappa_{2,X}\Delta t\right))}{4\kappa_{2,X}}, \quad d = \frac{4\kappa_{2,X}\bar{X}_2}{\sigma_{2,X}^2} \quad \text{and} \quad \lambda_{X,t} = \frac{4\kappa_{2,X}x_{2,n-1}\exp\left(-\kappa_{2,X}\Delta t\right)}{\sigma_{2,X}^2(1-\exp\left(-\kappa_{2,X}\Delta t\right))}, \tag{5.31}$$

with $\Delta t = t_n - t_{n-1} = \frac{1}{12}$, because we use a monthly frequency of the historical data.

We consider a normal and a skewed normal distribution for modeling the observable process V_t conditional on the hidden process X_t. The two choices for modeling the observable process denoted by $V_{1,t}$ and $V_{2,t}$ are given by the following PDFs:

$$f_{V_1,\psi}(v_{1,n}\,|\,x_{.,n}) = f_{norm}\left(v_{1,t}; \eta_V(\Upsilon(x_{.,t})), \sigma_{1,V}\right), \tag{5.32}$$

$$f_{V_2,\psi}(v_{2,n}\,|\,x_{.,n}) = \frac{2}{\sigma_{2,V}} f_{norm}\left(\frac{v_{2,t}-\eta_V(\Upsilon(x_{.,t}))}{\sigma_{2,V}}; 0, 1\right) \cdot F_{norm}\left(\alpha_V\left(\frac{v_{2,t}-\eta_V(\Upsilon(x_{.,t}))}{\sigma_{2,V}}\right); 0, 1\right), \tag{5.33}$$

with $\cdot = 1,2$ and $\alpha_V \in (-1,1)$. The skewed normal PDF equals the normal PDF when $\alpha_V = 0$. In the normal distribution the error between the observed option price and the model option price are centered symmetrically around zero. Under the skewed normal distribution we are able to model a skewed distribution of these errors via the parameter α_V. This is for example relevant if the Heston model is only able to model certain combinations of K and T of the option market data. Negative option values are truncated to zero.

We consider four cases in which we use either the OU or the CIR model for the state process, and either the normal or the skewed normal distributions for the errors of the option prices with respect to the fit. Summarizing, we consider the following 4 cases:

- **Case I**: The state process follows the PDF in Eq. (5.29) and the observable process follows the PDF in Eq. (5.32). The model parameters to be estimated are $\psi = \left\{a_\Upsilon, b_\Upsilon, \kappa_{1,X}, \bar{X}_1, \sigma_{1,X}, \sigma_{1,V}\right\}$. This results in the estimation of 120 parameters (including the hidden states for $t \in \mathcal{T}$).
- **Case II**: The state process follows the PDF in Eq. (5.29) and the observable process follows the PDF in Eq. (5.33). The model parameters to be estimated are $\psi = \left\{a_\Upsilon, b_\Upsilon, \kappa_{1,X}, \bar{X}_1, \sigma_{1,X}, \sigma_{2,V}, \alpha_V\right\}$. This results in the estimation of 121 parameters (including the hidden states for $t \in \mathcal{T}$).

- **Case III**: The state process follows the PDF in Eq. (5.30) and the observable process follows the PDF in Eq. (5.32). The model parameters to be estimated are $\psi = \left\{a_Y, b_Y, \kappa_{2,X}, \bar{X}_2, \sigma_{2,X}, \sigma_{1,V}\right\}$. This results in the estimation of 120 parameters (including the hidden states for $t \in \mathcal{T}$).

- **Case IV**: The state process follows the PDF in Eq. (5.30) and the observable process follows the PDF in Eq. (5.33). The model parameters to be estimated are $\psi = \left\{a_Y, b_Y, \kappa_{2,X}, \bar{X}_2, \sigma_{2,X}, \sigma_{2,V}, \alpha_V\right\}$. This results in the estimation of 121 parameters (including the hidden states for $t \in \mathcal{T}$).

In each case, we hence use a much smaller number of parameters than in the basic calibration method where we used 530 parameters. We reduce the number of parameters by a factor ≈ 4.4.

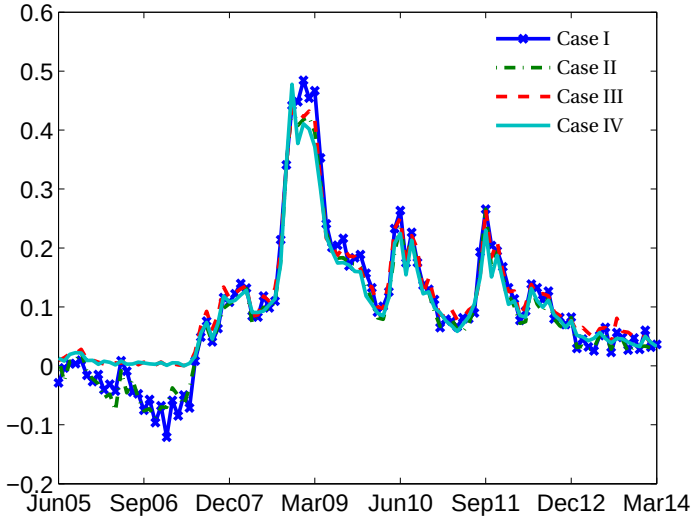
Next, in Section 5.3.2 we have discussed the estimation of SSHM models. We use the EM algorithm in combination with particle filtering. We have implemented the numerical methods in Matlab. In particle filtering, the estimation results depend on the EM iterations N_{EM}, the number of particles N_{par} and the initial parameter settings. Unfortunately, it is beforehand difficult to specify N_{EM} and N_{par}. Ideally, we specify those values as large as possible to guarantee convergence, but this would slow down the estimation process. There is hence a trade-off to be made. We take $N_{EM} = 50$, $N_{par} = 10$ and set the initial parameters equal to:

$$a_Y = (0.0, 0.0, 0.0, 0.0, 0.0), \quad b_Y = (0.35, 0.1, 0.1, 0.4, -0.7),$$
$$\sigma_{\cdot,X} = \sigma_{\cdot,V} = \kappa_{\cdot,X} = \bar{X}_{\cdot,X} = 0.1, \quad \alpha_V = 0,$$

with $\cdot = 1, 2$. We find that this parameter setting leads to satisfactory results for all experiments.

Estimated parameters The estimated state processes $X_t = x_t$ are shown in Figure 5.3 for Cases I-IV.

Figure 5.3: Estimated state process.



We observe similar estimations. The estimated processes for Cases III and IV are positive due to the underlying CIR process; the NCCS distribution has a positive domain. It turns out that the state processes are highly correlated with the VIX index; the Pearson correlations are

$$\text{Case I}: 0.884, \quad \text{Case II}: 0.922, \quad \text{Case III}: 0.929, \quad \text{Case VI}: 0.936. \tag{5.34}$$

In Case IV we find the highest correlation with the VIX index. The estimated parameters ψ are given in Table 5.2.

Table 5.2: Overview of model parameters ψ.

	κ_X	$\bar{X}$	σ_X	σ_V	α_V
Case I	0.8379	0.1048	0.1539	0.0050	-
Case II	0.8813	0.0835	0.1364	0.0039	-0.0991
Case III	0.3600	0.1162	0.3506	0.0043	-
Case IV	0.3387	0.1230	0.3461	0.0039	-0.3450

	a_1	a_2	a_3	a_4	a_5	b_1	b_2	b_3	b_4	b_5
Case I	0.1210	0.2725	0.3122	0.3625	-0.06300	0.1914	0.0132	0.1232	0.3475	-0.6647
Case II	0.1100	0.3529	0.3631	0.4043	-0.07100	0.2302	0.0114	0.1133	0.3779	-0.6751
Case III	0.1300	0.3182	0.5336	0.3988	-0.06400	0.2678	0.0111	0.0750	0.3970	-0.6780
Case IV	0.1400	0.3852	0.4296	0.6328	-0.09000	0.4584	0.0111	0.0562	0.4132	-0.6820

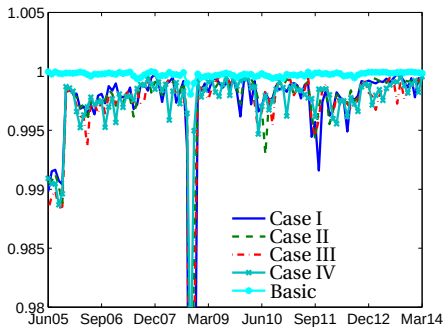
The estimated skewness parameter α_V is negative for Cases II and IV, so that the PDF $f_{V,\psi}$ is negatively skewed. This means that market prices are overestimated by the Heston model, i.e. the model prices are (on average) higher than the market prices. The PDF $f_{V,\psi}$ corrects for this via the negative α_V parameter.

The estimated long term states $\bar{X}$ are comparable, but the $\bar{X}$ is highest for Case IV. The mean reversion and correlation parameter are almost constant, i.e. the corresponding elements in a_Υ are approximately zero. The estimated ρ and κ_v are comparable although Cases III and IV estimate some higher mean reversion parameters than Cases I and II. The parameters v_0, $\bar{v}$ and σ_v change dynamically over time because the corresponding elements in the vector a_Υ are non-negative and hence these parameters are directly connected to the dynamic part x_t. We observe that the elements in a_Υ of v_0, $\bar{v}$ and σ_v are all positive which indicates positive correlations between v_0, $\bar{v}$ and σ_v. This coincide with the correlations computed in Eq. (5.27).

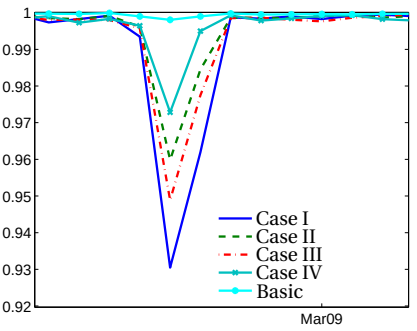
Quality-of-fit The R^2-error for Cases I-IV and the basic calibration method is visualized in Figures 5.4a and 5.4b.

Figure 5.4: Quality-of-fit.

(a) R^2



(b) R^2 (zoom-in around the spike at October 2008)



5

We observe in Figure 5.4a highly satisfactory results, i.e. $R^2 > 0.93$ and by removing the spike at October 2008 we have $R^2 > 0.98$ for all time points. The data quality of October 2008 is bad due to illiquidity in the market. These results are obtained for $d_X = 1$. The results typically improve when we increase the dimension d_X; for $d_X = 5$ the results would be similar to the basic calibration approach since all model parameters are used in that case.

The largest error is when the VIX index is highest, i.e. when there is turmoil in the market. In times of market stress the IV surface is typically more skewed due to (amongst others) the higher demand for put options for protection. Typical in those periods, the correlation parameter becomes more relevant. In the proposed methodology this parameter is estimated by approximately a constant and hence there is not enough flexibility to model this extreme skew.

The Cases I and II perform worst. This is due to the underlying normal distribution for the state process X_t, which is not capable of modeling the heavy tails in the data. We find in Figure 5.4b that the Cases III and IV perform better than Cases I and II respectively, because the underlying CIR process is able to model the heavy tails in the data.

Case IV performs best given the fit to the available data. The mean $R^2 = 0.9974$ is slightly smaller than the R^2 of the benchmark model $R^2 = 0.9997$. Therefore, we favor Case IV. We obtain a similar fit as the basic calibration method, but by using a much smaller parameter space. We recall that we reduce the number of parameters by a factor of 4.4 (from 530 to 121 in case of Case IV).

To gain insight in the performance of separate parts of the option surface we show in Table 5.3 the sum squared errors (SSE) per option type and for all $t \in \mathcal{T}$. The results are shown for Case IV.

Table 5.3: Quality-of-fit of the Heston model.

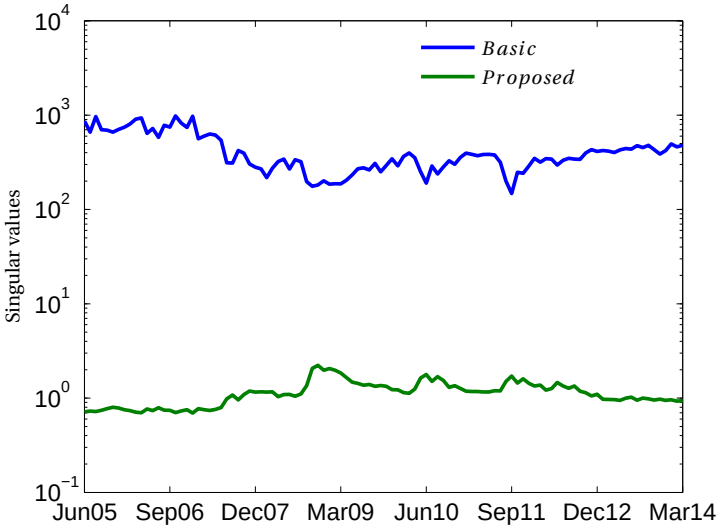
	$K = 0.8$			$K = 1.0$			$K = 1.2$		
	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$	$T = 0.25$	$T = 1$	$T = 2$
Av. SSE ($\times 10^{-4}$)	0.07	0.09	0.23	0.35	0.10	0.17	0.02	0.13	0.18
Max. SSE ($\times 10^{-3}$)	0.31	0.17	0.21	0.37	0.24	0.10	0.03	0.15	0.21

The overall fit to the market data is still satisfactory. Since we use a much smaller parameter space, we obviously lose flexibility with respect to the basic calibration method. We observe that the model is less capable of calibrating the longer maturing options.

From Table 5.2 we observe that the dynamic part, i.e. the non-zero element of a_Y, is related to the parameters v_0, $\bar{v}$ and σ_v. The parameters κ_v and ρ are approximately constant. This parametrization of Θ is able to model the height of the term structure of IVs (via v_0 and $\bar{v}$), but with approximately a fixed mean reversion κ_v. This might be restrictive in extreme market circumstances. We are able to model dynamic skews in the IV surface via σ_v, but perhaps not well enough in extreme events since the parameters v_0, $\bar{v}$ and σ_v are linked to the same (single) risk driver x_t. Since the correlation parameter is approximately fixed over time we model fixed skews in the IV surface. This can be restrictive in extreme cases, because skewness is important in high volatile periods; see Figures 5.4a and 5.4b.

Numerical stability By using the proposed methodology we also aim at improving the numerical stability, see Section 5.3.4. Since we assume $d_X = 1$ we have 1 singular value (instead of 5 in Section 5.4.1). We compute the operator norm of D_{χ^Y} by using Eq. (5.25) and is illustrated in Figure 5.5. For comparison, we also show the operator norm of D_χ from the basic calibration method in Section 5.4.1 denoted by $Basic$. We show the results of Case IV, but the results for Cases I-IV are similar.

Figure 5.5: Numerical stability of the χ and χ^{Υ} mappings.

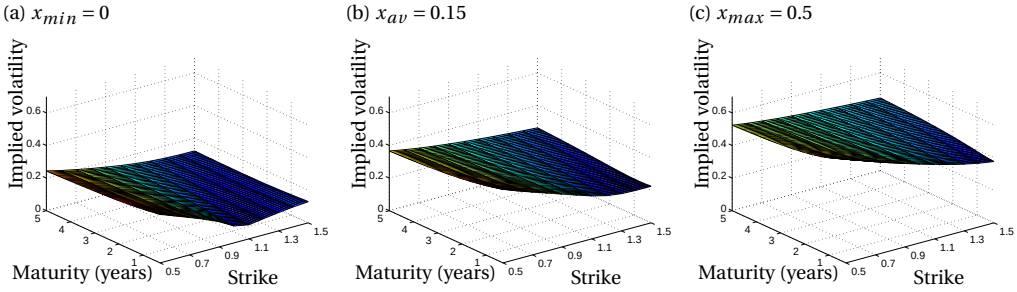


We observe a major improvement of numerical stability with factors up to 1000 with respect to the $Basic$ method. Contrary to the basic calibration method, there is more variation in the Heston model in volatile periods and vice versa. This is due to the lower number of risk drivers, i.e. $d_X = 1 \leq 5 = d_\Theta$. This means that the Heston model under the proposed method is more sensitive to market changes in volatile periods. There is a significant positive correlation between the market volatility (represented by the VIX index) and the (maximum) singular value of $D_{\chi^{\Upsilon}}$. The Pearson correlation coefficient between $\|D_{\chi^{\Upsilon}}\|^{op}$ and the VIX index over this period is 0.89. Despite the variation in volatile periods, the $Proposed$ method is for all $t \in \mathcal{T}$ more stable than the $Basic$ method.

Matching empirical facts We use the Υ mapping in Eq. (5.15) to reduce the dimension of the Θ space. As mentioned in Section 5.3.3 the mapping consists of a static part and a dynamic part X_t, which are estimated as is described in Section 5.4.2. We have showed that the proposed method leads to a highly satisfactory fit, obtained with numerical stability. We aim to get insight in the ability of the proposed method to model the relevant empirical facts in the data. The results described below are generated using the settings from Case IV.

In Figure 5.6 we show the IV surfaces belonging to $x_{min} = 0$, $x_{av} = 0.15$ and $x_{max} = 0.5$, i.e. the surface belonging to the minimum, average and maximum values of the hidden states $X_t = x_t$.

Figure 5.6: IV structures in the maturity dimension.



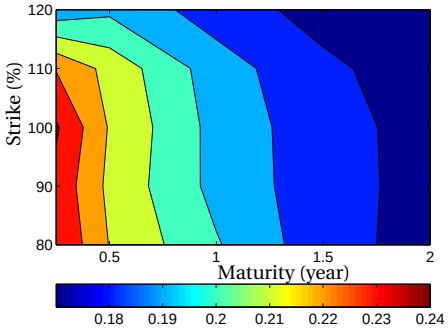
In the market data we observe that the level of the IV surface is high in times of turmoil, i.e. in times of a high VIX index. The estimated hidden states are highly correlated with the VIX index, see Eq. (5.34). We also observe that when the level of the IV surface is high, then there is more skew for the short term options. This is also observed in historical option market data.

To gain more insight in the match with the empirical facts we perform a principal component analysis (PCA) to the historical S&P-500 equity index IV as seen in the market and the corresponding IVs generated by the proposed model. We investigate whether the PCA factors are comparable. PCA factors are orthogonal (zero correlation) linear combinations of a time series that explain the largest part of the total variance. These factors are determined by assigning a weight (loading) on each of the input time series. The weights are determined such that the resulting factors describe the largest part of the joint movements (correlations) of the input time series.

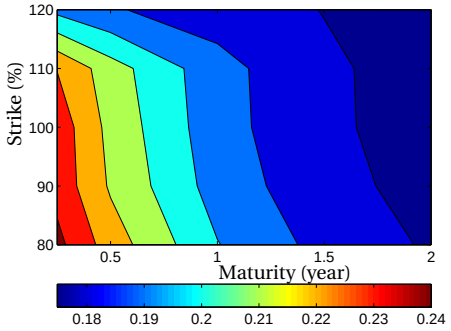
We perform a PCA to the historical S&P-500 equity index IV market data and to the IV surfaces calibrated by the $Proposed$ model. We use the option maturities 0.25, 0.50, 1.00, 1.50 and 2.00 years and strike levels 80%, 90%, 100%, 110% and 120%. It turns out that the first PCA component of the historical data accounts for 97.6% and the first PCA component of the $Proposed$ model 98.0%. The first components, that account for most of the variance, are very similar, so that the proposed model is able to model the dynamics in the historical IV surfaces rather well. The first PCA factors are shown in Figure 5.7.

Figure 5.7: First PCA factor.

(a) Historical data



(b) *Proposed* model

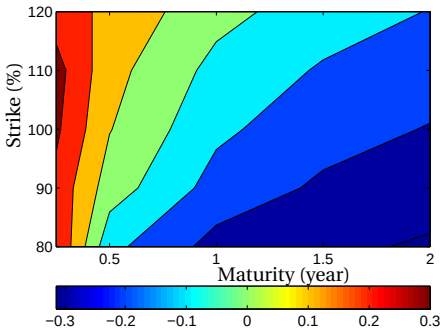


We observe indeed a satisfactory resemblance of the first PCA factor. The most significant skews are observed for out-of-the-money strike levels (for put options) and short maturing options. We find an almost flat/linear implied volatility structure for the in-the-money strike levels (for put options) and long maturing options.

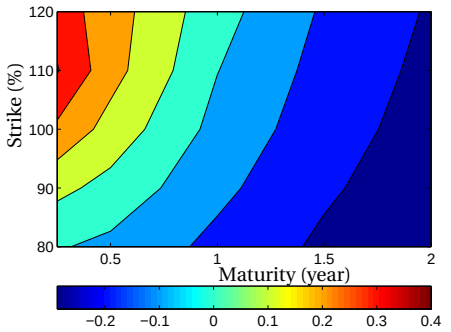
The second PCA factors account for 1.5% and 1.4% of the total variance of respectively the historical data and the *Proposed* model. The second PCA factors are shown in Figure 5.8.

Figure 5.8: Second PCA factor.

(a) Historical data



(b) *Proposed* model



Although the IV pattern is similar for different maturities and strikes, we observe some deviations in the second PCA factors. We note however that, the second PCA factor only contributes for a small part to the total variance, so that these deviations are less important.

5.4.3. IMPACT STUDY

In Sections 5.4.1 and 5.4.2 we have calibrated the $Basic$ and $Proposed$ methods to option market data. We found that the calibration fit is highly satisfactory for both methods. However, we have showed that for the $Proposed$ method the model parameters are less sensitive to changes in the option market data. Therefore, we favor the $Proposed$ method. We compare both methods by valuing an embedded option of an insurance company. That is, we compute the composite functions $\eta_{\tilde{V}} \circ \chi$ and $\eta_{\tilde{V}} \circ \chi^{\Upsilon}$, which we discussed in Section 5.3.4. We first focus on computing present embedded option values and after that we compute future embedded option values.

We assume a fictive insurer XYZ which sells unit-linked products of which we consider a simplified version. We assume the unit-linked product is of European type, where policyholders capital is invested in a single equity series. The unit-linked product then becomes an equity index put option. We consider a 120% in-the-money put option with maturity equal to 10-years. The embedded option value is relevant to the insurers balance sheet.

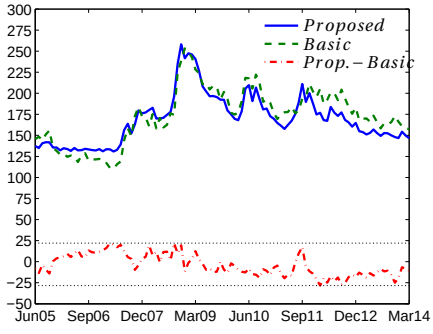
COMPUTING PRESENT EMBEDDED OPTION VALUES

For comparison, we use the $Basic$ and $Proposed$ methods to compute the embedded option value of insurance company XYZ, which we respectively denote by $\tilde{V}^B$ and $\tilde{V}^P$. For the proposed method we use Case IV.

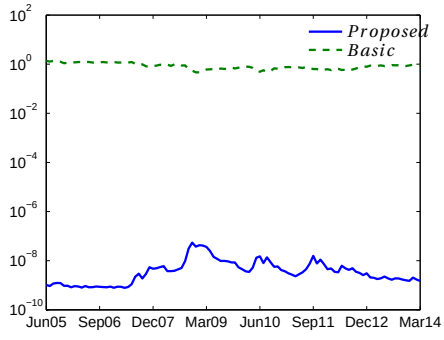
We first consider the computation of present embedded option values, and an option notional of 500. We use the operator norm to measure numerical stability, see Section 5.3.4. In Figure 5.9a we compare the embedded option values $\tilde{V}^B$ and $\tilde{V}^P$. We gain insight in the $Basic$ and $Proposed$ models in Figure 5.9b by computing the maximum singular values of $\|D_{\eta_{\tilde{V}} \circ \chi}\|^{op}$ and $\|D_{\eta_{\tilde{V}} \circ \chi^{\Upsilon}}\|^{op}$ (see Section 5.3.4).

Figure 5.9: Comparison of the $Basic$ and $Proposed$ models.

(a) Embedded option



(b) Numerical stability



From Figure 5.9a we observe that differences between the basic and proposed methods can be substantial. For the embedded option value we observe differences up to €25, which is around 17% relative difference.

It depends on how large the embedded option value is with respect to the total liability value to gain insight in the impact to the surplus, for example. The surplus is the difference between the assets and the liabilities on the insurer's balance sheet and is an important quantity of an insurer. The larger the embedded option value with respect to the total liability value the larger the impact of the differences between $Basic$ and $Proposed$ to the surplus. Considering the balance sheet in Table 5.4, the impact of a 17% difference in the embedded option value would result in relative difference of 20% in the surplus value.

We observe in Figure 5.9b that, as expected, the $Proposed$ method is more stable than the $Basic$ method.

COMPUTING FUTURE EMBEDDED OPTION VALUES

We compare the proposed methodology ($Proposed$) with a basic modeling approach ($Basic$), which is often applied in practice, to compute future embedded option values. The starting balance sheet in Euros of insurance company XYZ is summarized in Table 5.4. We assume a perfectly matching asset portfolio for the guaranteed liability cash flows, an asset portfolio (denoted by A) and an embedded option (denoted by $\tilde{V}$). The notional of the embedded option is determined in such a way that the embedded option value is equal to €100 for the $Basic$ and $Proposed$ models.

Table 5.4: Balance sheet of XYZ at $t = 0$.

Assets		Liabilities	
		Matching portfolio	€700
Equity	€200	Embedded option	€100
		Surplus	€100
Total	€900	Total	€900

To project the balance sheet for $t \geq 0$, the asset process A_t and the embedded option process $\tilde{V}_t$ are relevant. The surplus (available capital) S_t is computed by the difference of the value of the asset and the value of the liabilities:

$$S_t = A_t - \tilde{V}_t. \tag{5.35}$$

Other balance sheet items are neglected in this simplified example.

We show the impact of the $Basic$ and $Proposed$ methods to the embedded option value, the surplus of the balance sheet and the required capital. Special attention is devoted to the $t = 1$-year horizon because this is important for Solvency II computations. Under $Proposed$ we again consider the estimated model Case IV from Section 5.4.2 for the model parameters. Under $Proposed$ we forecast the model parameters Θ_t using the transitional PDF $f_{X,\Psi}$ in combination with Eq. (5.15).

Under the $Basic$ method, valuation is based on a fixed parameter setting (as in [8, 50]), so independent of $t \geq 0$: $\Theta_t = \Theta_0$. For comparison we use the same parameter settings as the initial parameters of Case IV:

$$\Theta_0 = \{\kappa_v, v_0, \bar{v}, \sigma_v, \rho\}_0 = a_Y x_0 + b_Y = \{0.4630, 0.0236, 0.0702, 0.4338, -0.6830\},$$

where a_Y and b_Y are found in Table 5.2. Using the $Proposed$ and $Basic$ valuation models we are able to compare the approximation of the embedded option $\tilde{V}_t^P$ and $\tilde{V}_t^B$. Via Eq. (5.35) we can also compare the surplus values which we denote by S_t^P and S_t^B.

In this stylized example we assume that $Y_t = \{A_t, r_t, X_t\}$ models the uncertainty in the real world financial market. Therefore, these processes are modeled under the $\mathbb{P}$ measure, see Section 5.2.1. We assume that the return equity portfolio A_t is modeled by a Black-Scholes OU model. So, we assume a fixed volatility, but stochastic interest rates generated by the OU model [18]. The state process X_t follows the estimated CIR process of Case IV from Section 5.4.2.

$$\begin{aligned} dA_t &= \mu_A A_t dt + \sigma_A A_t dW_t^A, \\ dr_t &= \kappa_r(\bar{r} - r_t)dt + \sigma_r dW_t^r, \\ dX_t &= \kappa_X(\bar{X} - X_t)dt + \sigma_X\sqrt{X_t}dW_t^X, \end{aligned}$$

with

$$\begin{aligned} A_0 = 200,\ \mu_A = 0.06,\ \sigma_A = 0.2\ r_0 = 0.02,\ \bar{r} = 0.04,\ \kappa_r = 0.02,\ \sigma_r = 0.01, \\ X_0 = 0.0326,\ \bar{X} = 0.1230,\ \kappa_X = 0.3387,\ \sigma_X = 0.3461. \end{aligned}$$

Since we are interested in computing the embedded option values $\tilde{V}_t$, we only consider the estimated PDF $f_{X,\psi}$ from Case IV to generate the model parameters Θ via Eq. (5.15). The PDF $f_{V,\psi}$ is not needed in this experiment.

We remark that the $Basic$ model only depends on the stochastic interest rates, because the model parameters are kept fixed during the simulation. Besides the stochastic interest rates, implied volatility risk is also taken into account for the $Proposed$ model because the model parameters follow a stochastic process. We assume the following correlation matrix Σ between the Wiener processes (W_t^A, W_t^r, W_t^X):

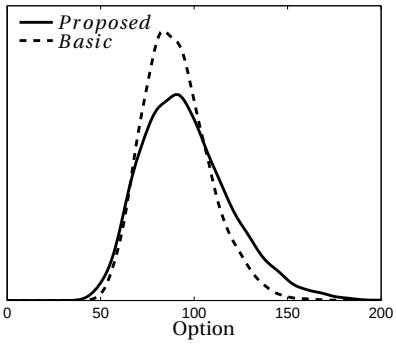
$$\Sigma = \begin{pmatrix} 1 & & \\ \rho_{A,r} & 1 & \\ \rho_{A,X} & \rho_{r,X} & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ 0.10 & 1 & \\ -0.60 & -0.05 & 1 \end{pmatrix}.$$

The joint modeling of X_t and r_t is important, because both risk drivers affect the embedded option value. We assume $\rho_{A,X}$ is negative, because the process X_t is highly correlated with implied volatility, which in turn are negatively correlated with the asset returns. In the numerical experiments, we vary the correlation $\rho_{r,X}$ between -0.8 and 0.7 to gain insight in the impact. We generate $N_{\mathbb{P}} = 10,000$ real world scenarios of $\left(A_t, r_t, X_t\right)$ with a $T_t = 1$-year horizon. For each time point we value the unit-linked option under the $Proposed$ and $Basic$ methods using the Heston pricing formula [47].

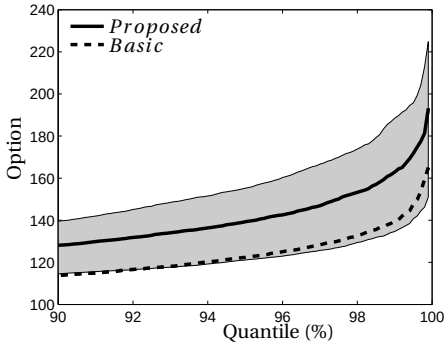
The embedded option process In Figures 5.10a and 5.10b we compare the embedded option process for the $Proposed$ and $Basic$ methods. In Figure 5.10a a density plot is shown. In Figure 5.10b insight is provided in the (right) tail risk. We compute tail risk by means of Value-at-Risk (VaR), i.e. we compute the value $\overline{V}$ such that $\mathbb{P}\left(\tilde{V}_1 \leq \overline{V}\right) = \alpha$, where we vary $\alpha \in [0.9, 1]$ (the right tail is relevant). The gray filled area in Figure 5.10b is computed by varying the correlation parameter $\rho_{r,X}$ between -0.8 and 0.7.

Figure 5.10: Comparison of the embedded option values.

(a) PDF



(b) Insight in right tail

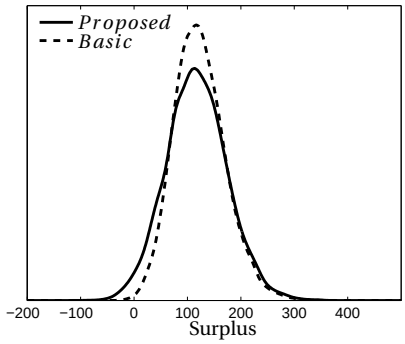


We observe that the PDF of $Proposed$ has a wider range of embedded option values, i.e. the embedded option process is more volatile under the $Proposed$ method. This is due to the extra stochastic process for modeling the equity volatility part of the Heston model. Due to the wider PDF, the right tail is more heavier in the $Proposed$ method, which results in differences up to €30. Varying the correlation parameter $\rho_{X,r}$ results in different valuations of the $Proposed$ method. The upper line of the filler area belongs to $\rho_{X,r} = 0.7$ and the lower line to $\rho_{X,r} = -0.8$.

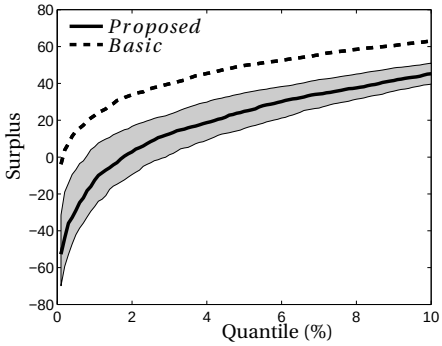
The surplus process In Figures 5.11a and 5.11b we compare the surplus process S_t for $Proposed$ and $Basic$. In Figure 5.11a a density plot is shown and in Figure 5.11b insight is provided in the tail risk. We again compute tail risk by means of Value-at-Risk (VaR), i.e. we compute the value $\overline{S}$ such that $\mathbb{P}\left(S_1 \leq \overline{S}\right) = \alpha$, where we vary $\alpha \in [0, 0.1]$ (the left tail is relevant). The gray filled area in Figure 5.11b is again computed by varying the correlation parameter $\rho_{r,X}$ between -0.8 and 0.7.

Figure 5.11: Comparison of the surplus.

(a) PDF



(b) Insight in left tail



We again observe that the PDF of the *Proposed* method has a wider range of values, although the impact is smaller compared to the embedded option process. Due to the wider PDF, the left tail is more heavier using the *Proposed* method, which results in differences up to €25. The choice of the correlation $\rho_{A,X}$ is crucial in this case. When A_t decreases, then X_t increases due to the negative correlation, so that the value of the embedded option $\tilde{V}_t$ increases. The latter results in a decrease of the surplus.

The required capital Lastly, we analyze the impact of computing the required capital (RC), which is comparable to the Solvency Capital Requirement (SCR) under Solvency II. Insurers use internal models to compute required capital for internal steering and/or reporting to the regulator. The required capital is the amount of capital the insurer must hold against unforeseen losses during a one-year period. The required capital is the VaR of a loss-function on a certain horizon (often the 1-year horizon). A common loss function (see [7]) is

$$L_t = S_{t-\Delta}(1 + R_{t-\Delta,\Delta}) - S_t \quad \text{for} \quad \big(\Delta = 1, \mathrm{t} \geq \Delta\big),$$

with $R_{t-\Delta,\Delta}$ the the Δ-year risk-free rate in year $t-\Delta$. This loss function will be used in our experiments. Based on a one-year horizon and for a certain confidence level $\alpha \in [0,1]$, the RC, RC_α, is computed by:

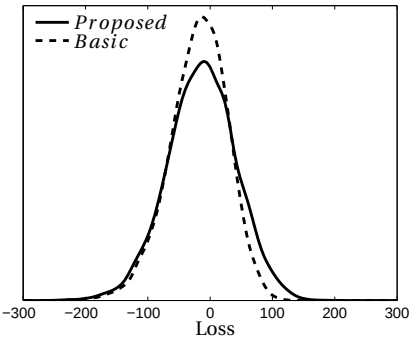
$$\mathbb{P}\big(L_1 \geq RC_\alpha\big) \leq 1-\alpha, \quad \big(L_1 = S_0(1+R_{0,1}) - S_1\big),$$

where we set $R_{0,1} = 0.025$ is the one-year risk-free rate at $t = 0$. Hence, the probability that the loss over one year exceeds the RC is less or equal to $1-\alpha$. In practice, this confidence level is often set to 99.5%.

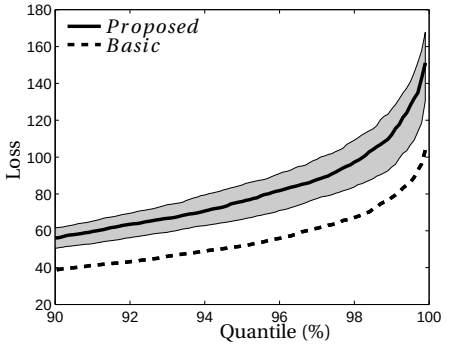
In Figures 5.12a and 5.12b we compare the loss function for *Proposed* and *Basic*. In Figure 5.12b the PDF is illustrated of the loss function. In Figure 5.12b insight is provided in the right tail of the distribution. The gray filled area in Figure 5.12b is again computed by varying the correlation parameter $\rho_{r,X}$ between −0.8 and 0.7.

Figure 5.12: Comparison of the loss function.

(a) PDF



(b) Insight in left tail



We again observe that the PDF of the *Proposed* method has a wider range of values compared to the *Basic* method. Due to the wider PDF, the right tail is more heavy in the *Proposed* method, which results in differences up to €40.

In Table 5.5 we give an overview of numerical values of the required capital for different confidence levels. Insurance companies with an higher credit rating are obligated to compute their required capital with a lower confidence level α. The values in parenthesis are the solvency ratio, i.e. the ratio between the surplus and required capital.

Table 5.5: Comparison of the required capital (in Euro).

	99.5%	97.5%	95.0%	90.0%
Proposed	121(0.81)	92(1.08)	74(1.34)	56(1.78)
Basic	86(1.16)	64(1.53)	51(1.92)	37(2.67)

We observe that the absolute difference in RC between *Proposed* and *Basic* becomes smaller when the confidence value α decreases from 99.5% to 90.0%. For α = 99.5% the difference is €35, which is 35% of the surplus. The differences in the solvency ratio is even larger: 81% for *Proposed* and 116% for *Basic*.

These differences show the relevance of taking the implied volatility risk factor into account, i.e. to apply the *Proposed* method. The impact of the differences between *Basic* and *Proposed* heavily depends on the initial balance sheet settings and in particular the value of the embedded option portfolio with respect to the total liability value, which is in this simplified example 14% (100/700).

5.5. CONCLUSIONS

We presented a method for modeling risk neutral models in a real world scenario model to perform nested Monte Carlo simulations. This is important for ex-ante risk management applications for insurance companies. In such applications, an insurer is required to compute their embedded option values at a future point in time. We make use of the

well-known State Space Hidden Markov models, which provide for a flexible modeling framework. By introducing a hidden state process we are able to reduce the dimension of the calibration problem. In this way we reduce the computation time and we improve the numerical stability of the model parameters with respect to option market data. The latter is desired in practice for transparency and stable valuations of embedded options.

We applied the proposed method to the well-known risk neutral Heston model. Although the Heston model consists of five model parameters, i.e. the mean reversion, initial variance, long term variance, volatility of variance correlation parameters, we show that in our proposed methodology a one-dimensional state process already results in highly satisfactory calibration results. That is, we obtain a maximum dimension reduction. We also show that numerical stability of the model parameters with respect to option market data is greatly improved. We measure numerical stability by means of the operator norm.

To show the relevance of our method, we compared the estimated Heston model from the proposed methodology to a basic parameter setting. We used both models to compute the present and future values of a (simplified) unit-linked product of a fictive insurer. The results differ substantially and especially the tails of the distributions differ, which are important in practice. Given the large differences in this simplified case study, we advice to use the proposed methodology for calibration, valuation and simulation.

Now we have a modeling framework for risk neutral models available, we seek a method to accelerate nested simulations. Such a method would also speed-up the calibration process as the method is also based on many Monte Carlo simulations. Therefore, we focus in the next chapter on high performance computing techniques to accelerate valuations based on Monte Carlo simulations.

CHAPTER 6

FiNS: A framework for accelerating nested simulations on heterogeneous platforms

To keep computation times of nested simulations within acceptable limits high performance computing is required. We present a framework designed to significantly improve the performance of nested simulations by using heterogeneous computing. Specifically, we use modern features from CUDA - streams, Hyper-Q, and Multi-Process Service, to take advantage of the parallelism of Graphical Processing Units. We manage to reduce the execution time of a nested simulation application from several hours to tens of minutes.

6.1. INTRODUCTION

Our aim is to improve the performance of nested simulations (see Chapter 1), making them feasible for both production and more empirically-driven research. Given that Graphical Processing Units (GPUs) are a proven technology in finance for performing Monte Carlo valuation simulations [1, 86], we aim to make use of these massively parallel architectures to accelerate financial nested simulations. The main challenge is improving efficiency, because the multiple layers of parallelism of nested simulations require a tight collaboration of the CPU and the GPU.

We introduce the Financial Nested Simulations (FiNS) framework; a CPU-GPU heterogeneous solution for improving the performance of nested simulations in financial applications. FiNS is driven by two important requirements: performance improvement and ease-of-use for financial specialists.

FiNS makes use of a set of advanced CUDA abstractions available in the latest NVIDIA architectures (Kepler and newer) to improve performance: CUDA streams [96], Hyper-Q [104], and Multi-Process Service (MPS) [105] are all used to efficiently offload simulations to the GPU. FiNS is built as a skeleton that can be easily adapted to different applications, which provides flexibility to financial specialists.

Financial nested simulations are increasingly important due to new regulations, so their performance becomes a production-level concern. Therefore, the authors in [8] describes several numerical methods for reducing the computational intensity of Solvency Capital Requirement (SCR) calculations, making it computationally feasible. Complementary to their study, our work demonstrates that the FiNS framework can render the same simulations feasible by using CPU-GPU heterogeneous computing.

The utility of GPUs in Monte Carlo valuation methods is becoming a proven technology in finance [86]. Such work is focusing strictly on improving the performance of risk

This chapter is based on the article 'FiNS: A Framework for Accelerating Nested Simulations on Heterogeneous Platforms', published in *European Conference on Parallel Processing*, 246–257, 2015 [33].

neutral simulation. In [1], a CPU-GPU performance comparison is made, achieving up to 10 times speed-up for both European and American contracts. Additionally, a lot of research has been dedicated to GPU-accelerated solutions for several risk neutral models [10, 51, 79, 127, 128]; the observed speed-ups range between 4 and 150. In our work, we focus on simulation models that require the cooperation of the CPU and the GPUs towards efficient nested simulations.

Using heterogeneous computing for large scale simulations is already established as a feasible solution to improve performance for many classes of applications. Systems such as Glinda, Qilin, or Insieme [82, 97, 116, 117] focus on static partitioning of one workload to multiple devices, under the assumption that the GPU is overloaded. Such systems are not suitable for our nested simulations, because in our scenarios the GPU is "underloaded". An alternative is to use a runtime-based system for heterogeneous computing, such as OmpSS or StarPU [5, 107]. However, none of these approaches supports sharing devices by multiple kernels, which is an essential performance booster for FiNS.

To demonstrate both the performance and usability of FiNS, we build a mock-up model of an existing Asset Liability management (ALM) tool, which emulates the behavior of a full nested simulation. Our results show significant performance improvement over the sequential code, with speed-ups ranging between 26 and 6 for light and heavy cases, respectively. This significant gain is due to our efficient use of both the CPU and the GPU. Although the reference sequential code is by no means optimized, FiNS brings a significant improvement in the way nested simulations can be used in production and research.

The main contribution of this work is threefold. First, we propose an original way to exploit streams for increasing the efficiency of heterogeneous CPU-GPU platforms in the case of applications with multiple layers of moderate parallelism. Second, we design FiNS, a generic framework for using heterogeneous platforms in financial nested simulations. Third, we demonstrate how FiNS can be used to accelerate ALM, a specific case of nested simulation used in risk management for institutional investors like insurance companies or pension funds.

The remainder of this chapter is organized as follows. In Section 6.2 we provide a high-level background to GPUs. In Section 6.3 we present the FiNS framework. We apply FiNS to a ALM application in Section 6.4. We conclude in Section 6.5.

6.2. GPU BACKGROUND

GPUs[1] are massively parallel processing units, originally designed for graphics. A GPU has multiple streaming processors (SMs), each grouping tens of simple cores. With hundreds to thousands such cores, the performance to be achieved can easily reach a couple of TFLOPs for applications with enough concurrency. Additionally, GPUs have a layered memory system, including private memory per core, shared memory and L1 cache per SM, L2 cache and a global, off-chip memory. Memory bandwidth is typically significantly higher than for CPUs, but it is still the limiting factor for performance in many applications.

[1] We focus on NVIDIA GPUs and we make heavy use of CUDA concepts. In theory all other GPUs have the required features, yet programming them remains a challenge.

GPUs are working as accelerators, i.e. they are not stand-alone processing units, but require a host to manage their involvement in computation. Such host is typically a CPU; in such a CPU-GPU platform, the GPU is called a device. Note that the host and the device run separate codes: the host code is the main application from which parts are being offloaded for computation by kernels running on the device. Also note that the memory spaces of the CPU and GPU are separated, which means that any application that offloads computation kernels to the GPU might also need to copy data from host to device and/or the other way.

For programming these GPU kernels, the most popular solution is CUDA, a proprietary programming model from NVIDIA. While portable models like OpenCL and higher level models like OpenACC exist and can be successfully used for many applications, they are not suitable for this work because the special features we are using are not yet available in these models. We further describe these features in the paragraphs below.

6.2.1. CUDA STREAMS AND HYPER-Q

A CUDA stream is an abstraction of a series of tasks run by the GPU. By tasks we mean (1) memory copies, (2) synchronization, and (3) kernels (i.e., computational tasks). The tasks in a single stream are ordered, but they are independent from tasks in different streams.

Using streams can improve the concurrency of an application. For example, within nested simulation we repeat a sequence of tasks for every outer scenario in every period. By embedding this sequence of tasks in a stream, and launching a new stream for each node in the outer simulation, we have an elegant solution to launch multiple inner simulations that will not interfere with each other.

The most important features of streams necessary for this work are: first, stream launches can be asynchronous, allowing the CPU to compute while the GPU is running. Furthermore, tasks for different hardware engines within streams can run concurrently, i.e. computational tasks (kernels), executed by the GPU SMs and memory copies (D2H and H2D), executed by the GPU copy engines. And finally, streams can run concurrently on the device. We further detail the way we exploit these features in the following paragraphs.

CPU-GPU concurrency Stream launches are asynchronous by default. To ensure an asynchronous launch, care must be taken that memory copies are initialized with the asynchronous API and that the host memory allocations are pinned [104]. If synchronization between device and host is required, this can be accomplished by either several CUDA synchronization methods or by implicit synchronization. Unintentionally synchronizing the streams on the device is the main difficulty of working with streams.

Compute and memory accesses overlap Within a stream, it is possible to overlap the memory transfers with computational work by invoking the tasks with the asynchronous API. If this is crucial for the application performance, one needs to chunk the work such that the overlap is optimal. Furthermore, memory copies of a stream can overlap the computations of another stream.

We note that, implicitly, this solution increases application parallelism by decreasing the granularity of the tasks and making use of the engine parallelism in the hardware platform.

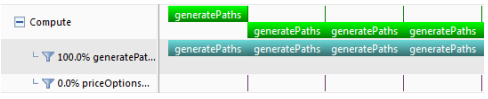
Concurrent Kernel execution The latest developments in NVIDIA cards increase the concurrency possibilities of the streams within a work queue. This is accomplished by NVIDIA's Hyper-Q feature [104].

Concurrent kernel execution is strictly bound by computational capacity and the device architecture. The latter is at the time of writing a dominant factor. The Kepler (and newer) cards support Hyper-Queuing, which has an important effect on performance.

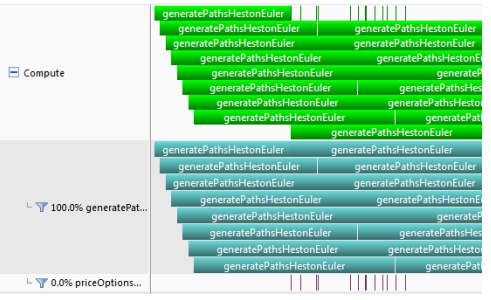
To illustrate this effect, we present in Figures 6.1a and 6.1b a comparison of the Fermi and Kepler architectures for streams concurrency. The streams contain two kernels: a large `generatePaths` kernel followed by the tiny `priceOptions` kernel. In Figure 6.1a we observe a two-way concurrency; because the final (tiny) task of a stream is running concurrently with the first (much larger) task of the subsequent stream. Hardware utilization in this case is low. On the contrary, Figure 6.1b displays the benefits of the hyper-Q allowing higher concurrency between streams, resulting in higher hardware utilization.

Figure 6.1: Concurrent Kernel execution.

(a) Stream concurrency on Fermi architecture



(b) Stream concurrency on Kepler architecture



6

6.2.2. MULTI PROCESSING SERVICE

Another way to increase the utilization of GPUs is to share the device for kernels from different (local) processes. In order to manage GPU sharing between processes, we used NVIDIA's Multi Processing Service (MPS) [105]. This software layer provides a context manager to handle work launched from different processes. MPS is exclusively available on Linux and is only provided by NVIDIA Tesla cards with compute capability 5 or higher. Although these restrictions limit applicability, it is a relatively cheap way to explore the concept of kernel offloading from multiple processes to a single GPU. This is an important feature when a single host process cannot generate sufficient work for the GPU, as it allows multiple cores or even multiple machines to collaborate in keeping a single device busy.

We note that previous work on offloading streams in a multi-threaded or multiprocessing environment [137] showed significant GPU utilization in benchmark cases. We implemented the same ideas as [137] for local Python processes. Section 6.4 provides more detail on how this feature affects our FiNS framework.

6.3. FRAMEWORK ARCHITECTURE

An ideal scenario is for the outer and inner scenarios to run in parallel. In this case, a perfect overlap provides optimal performance. Specifically, this means that within the duration of a real world simulation step (typically tens to hundreds of milliseconds), we must complete a full risk neutral simulation. This requirement demands a heterogeneous solution, which matches CPU + GPU architectures quite well, as seen in Figure 6.4. Our work therefore focuses on building a framework that significantly outperforms existing solutions for nested simulations, but is flexible enough to support multiple types of such applications, where the analysis and end-results of the inner and outer simulations can vary in complexity.

The key to efficient heterogeneous programming is in designing the right solution to utilize the available hardware efficiently. In our case, the main challenge is GPU utilization: one inner simulation offloaded to the GPU can not, for most applications, fully utilize a GPU on its own. In order to increase GPU utilization multiple inner simulations will have to run concurrently.

Without the concept of streams a custom implementation is needed, and it can become quite complex as one needs to build an aggregated kernel, as an artificial concatenation of kernels, which limits their flexibility in accepting different data sources or data types. When using CUDA streams these kernels can remain independent - thus flexible and fully reusable - and the task of concurrently executing them is offloaded to the device itself. From the perspective of flexibility this is an ideal solution for a generic framework, even if it comes with a small performance penalty.

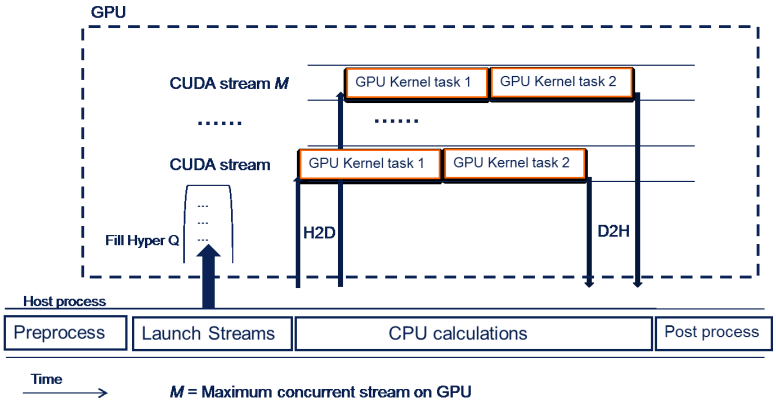
The Framework FiNS offers a skeleton-like infrastructure for the designers of nested simulation applications. Essentially, we provide a high-concurrency template that needs to be instantiated for a specific application. Using FiNS a developer needs to focus only on the implementation of the outer and inner simulation functionality. The framework will make sure that the mapping of these tasks on the real heterogeneous platform will be optimized for massive parallelism and efficient usage for both the CPU and the GPU.

To achieve this high level of flexibility we make use of CUDA streams, as seen in Figure 6.2, displaying the concurrency between host and a number of streams running on the device. The CPU prepares the workload for the GPU and launches the work in streams to the device. The device receives the streams and stacks the work in a Hyper-Q. Note here that the issue order of the streams is not necessarily the order of execution, which clearly requires the computations in streams to be completely independent. If this is not the case, FiNS cannot be used. As mentioned earlier, it is important to have all host memory allocations page-locked or else the CPU-GPU concurrency will break when memory transfers are initialized.

Furthermore, the Hyper-Q takes care of optimal hardware utilization by scheduling

the streams concurrently. All streams are launched asynchronously, so that the GPU computations can overlap with the CPU tasks. These tasks can consist of, for example, calculating real-world simulation steps, calculating statistics or memory flushes to the hard drive.

Figure 6.2: Using streams in FiNS. The CPU and all GPU streams can run concurrently. The level of concurrency achieved in practice is determined by the actual hardware.



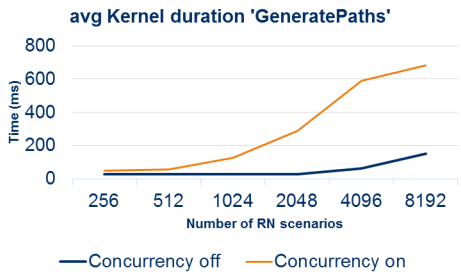
Streams in practice To show the performance behavior of kernel concurrency with streams, we set up a sample workload of 256 streams for the `generatePaths` kernel. We evaluated different sizes of offloaded work, determined by the number of risk neutral scenarios. To understand the performance gain for concurrency, we launched the sample both with and without kernel concurrency.

We observe that the average duration of the kernel increases when streams are running concurrently (Figure 6.3a). However, for a lower number of scenarios (till around 2,000), the concurrency contributes such that it is faster than the non-concurrent variant (Figure 6.3b). Moreover, for larger numbers of scenarios there is no performance penalty.

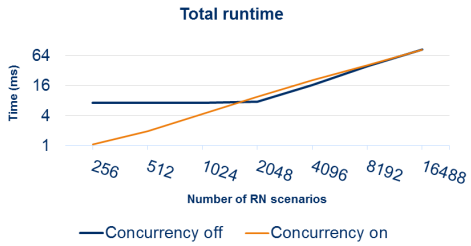
This behavior indicates that the CUDA Work Scheduler assigns more resources to single sequential streams than it does to concurrent streams. As a result, kernel concurrency in streams only benefits performance when the offloaded kernels are not large enough to fully utilize the GPU, whereas for larger cases performance remains equal.

Figure 6.3: Overview of performance.

(a) Average kernel duration



(b) Total runtime



6.4. NESTED SIMULATION FOR ALM TOOLING: A CASE STUDY

For performance reasons the available ALM software is not equipped with the nested simulation features. Instead, we use analytical methodologies, often less accurate. This also means that we have no real reference code to compare against. Therefore, we build a mock-up model of an ALM tool. In this mock-up model outer simulations are emulated by a sleep statement. Since we are interested in the impact on the user-time performance (wall-clock performance) for this application, we assumed three benchmark cases. They differ in the duration of a real world simulation step per scenario per period. We assumed normally distributed duration with means 75, 150 and 300 milliseconds for respectively a light, medium and a heavy case and a 5 millisecond standard deviation.

Figure 6.4: Overlapping real and nested simulations. Note that all RNs are different, yet independent, thus they can be executed concurrently.

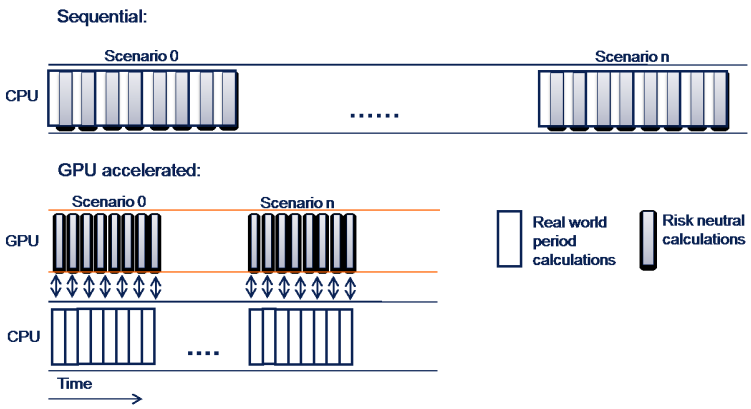


Figure 6.4 displays the concept of offloading risk neutral calculations to the GPU in comparison with a sequential version. With this concept the goal is to perform all risk neutral tasks within the runtime of a real world period. This is important since

ALM models are used in a decision-making process. Any additional runtime on current methodologies is unwanted by its users.

In contrast with Figure 6.4, modern ALM tools perform real world simulations in a multiprocess or multi-threading setting. This makes it more challenging finishing the offloaded risk neutral simulation within the duration of its parent real world simulation step, since the GPU will be receiving tasks from all processes simultaneously.

The framework presented in Section 6.3 serves as the greatest common divisor of several FiNS applications. For the ALM applications we extended the framework with multiprocessing support. This way, each host process is launching streams for its risk neutral calculation tasks to a daemon process hosted by the MPS (6.2.2). This daemon process manages all GPU requests from its slave processes and queues all received streams in a single hyper-Q. Using a single hyper-Q results in concurrent stream execution over the different processes.

Results Consider, again, the case of 2,000 real world scenarios with a horizon of 5 years (annual frequency), which are common dimensions for an insurer's ALM study. In a single process setting, simulations for the benchmark cases take resp. 12.5, 25 and 50 minutes. Note that we assumed perfect performance scaling for higher numbers of processes.

We measured that a single inner simulation of 1,000 scenarios with a horizon of 100 years and a 120 time steps per year takes 1.875 seconds on a state-of-the-art CPU. These simulation dimensions are representative for current CPU models. Given that the inner simulations are run $10,000 + 1$ (including $t = 0$) times for the complete run, the inner simulations represent a workload of over 5 hours in a single process setting.

For evaluating the use case on FiNS we used a NVIDIA K20 GPU. Table 6.1 displays the measured runtime of the mockup model with the heterogeneous framework. We observe that for the smaller number of cores the GPU is keeping up with the offloaded work. For a larger number of cores offloading work to the GPU, we observe that tasks stack in the hyper-Q and the CPU has to wait for the GPU results to be finished. Note that the single core results are below the theoretical reference for the Light and Medium case; this is caused by the assumption of normally distributed benchmark timings for the outer simulation.

The maximum speed-up of the heterogeneous framework versus the theoretical sequential runtime is bound by the fraction of the tasks to be offloaded. Amdahl's law tells us that the maximum achieved speed-up is defined by $\frac{1}{B}$, where B represents the fraction of time the models is strictly serial. This results in theoretical maximum speed-ups of 26.0, 13.5 and 7.25 for all processes in resp. the Light, Medium and Heavy benchmark cases. Table 6.2 displays the speed-up of the GPU accelerated model versus the theoretical CPU runtime. The results show that for most of the cases a near to maximum speed-up is reached.

Table 6.1: Runtime full nested simulation on GPU framework

Case # Cores	Light	Medium (in minutes)	Heavy
1	12.3	24.9	51.3
2	6.6	12.8	25.3
4	3.7	6.5	12.7
8	3.7	3.9	6.4
16	3.3	3.4	3.9

Table 6.2: User time speed-ups (maximum theoretical speed-up)

Case # Cores	Light *(26.0)*	Medium *(13.5)*	Heavy *(7.25)*
1	26.46	13.53	7.07
2	24.68	13.17	7.16
4	22.03	12.95	7.12
8	11.01	10.83	7.05
16	6.10	6.15	5.76

The results in Table 6.1 indicate that, if resources can be scaled (usage of multiple GPUs), the proposed architecture would in theory be able to close in to the theoretical maximum speed-up for every case defined in Table 6.2. Although scalability of the architecture is not implemented yet, we can conclude that the proposed architecture is most promising, since streams are easily distributable over multiple GPUs and MPS has multiple GPU support. To reveal the importance of NVIDIA MPS we also run the same tests with MPS disabled. We observed that speed-ups as displayed in Table 6.2 were up to 40% lower.

6.5. CONCLUSIONS

Due to the compute-intensive nature of nested simulations, CPU-only implementations lack the ability to provide sufficient performance, while GPU-only solutions are unable to efficiently utilize the hardware and lead to disappointing results. We proposed FiNS, which is a flexible heterogeneous framework based on modern technologies such as CUDA Streams, Hyper-Q and NVIDIAs MPS. FinNS is able to utilize both the CPU and the GPU to accelerate nested simulation applications.

We demonstrated the use of FiNS for an Asset Liability Management application, which required multiple CPU processes to offload calculations to the same GPU. To tackle this challenge, we customized the MPS functionality to handle local Python processes and achieved concurrency between streams owned by different local processes. The results demonstrate that FiNS achieves very good parallel efficiency.

Although the usage of HPC is an essential part of solving the nested simulation problem, the risk management software used by insurers is nowadays not flexible enough to use the latest hardware in the HPC area. Although this area is developing rapidly, approximation techniques that can be evaluated cheaply, such as the Option Interpolation Model discussed in Chapter 4, are often preferred.

CHAPTER 7

Conclusions and Outlook

7.1. CONCLUSIONS

In this thesis, we solved various problems with respect to the calibration and simulation of risk neutral models and the valuation of embedded options in risk management applications for insurance companies.

In Chapters 2 and 3, we focused on extending the Black-Scholes model. We discussed the modeling and calibration of inflation and real estate indices, that are relevant risk variables for insurance companies and pension funds.

In Chapter 2, we derived an approximate closed-form solution to compute the value of inflation indexed cap/ floor options and year on year inflation caplet/floorlet options, where the CPI follows a Heston model in which the nominal and real interest rates are modeled by one-factor Hull-White models. Using Fourier-based methods, calibration can be done highly efficiently. The inflation model is able to model the market implied volatility skew accurately, so that market conditions are replicated well. Using the calibrated inflation model we performed a market consistent valuation of the conditional indexation provision of a stylized pension fund. It turns out that the results change significantly when performing a calibration to market inflation option data instead to historical data. By changing the correlation parameters, indexation provisions change significantly, which justifies the use of a full correlation matrix.

In Chapter 3, we proposed a risk neutral valuation model for real estate derivatives. We valued a hypothetical European put option on a house price index. This example highlights the strong effect of autocorrelation in the underlying index on the option price. Using the proposed model, the effect of over- or undervaluation of the real estate market is also studied. The observed effects are significant.

In Chapters 2 and 3, we focused on the derivation of analytic formulas for fast calibration. Since Monte Carlo simulations are more flexible they became the preferred method for valuation. This results in the challenge of the nested simulation problem, which we consider in Chapters 4, 5 and 6.

In Chapter 4, we introduced the Option Interpolation Model (OIM) for modeling option values in ex-ante risk management applications. The usage of nested simulation is avoided using OIM. The method is easy to implement and has fast computation speed. The OIM is based on interpolation using radial basis functions. This interpolation method does not suffer from the curse of dimensionality. We explained the different settings in the OIM and showed the impact of the parameters on accuracy and numerical stability. This insight is required for a trade-off between accuracy and stability. In particular, we have analyzed the shape parameter, smoothing and the addition of an affine term. We constructed a matrix inversion method to obtain interpolation weights. The

method is especially useful in the case of an adaptive calibration procedure, which is often preferred in practice. All the relevant settings are tested to the Black-Scholes option formula. We find satisfactory results compared to benchmark settings. This confirms that the OIM is a useful technique as an approximation method for ex-ante risk management applications for insurance companies.

In Chapter 5, we presented a method for modeling risk neutral models in a real world scenario model to perform nested Monte Carlo simulations. We make use of the well-known State Space Hidden Markov models. By introducing a hidden state process we are able to reduce the dimension of the calibration problem. In this way we reduce the computation time and we improve the numerical stability of the model parameters with respect to option market data. The latter is desired in practice for transparency and stable valuations of embedded options. We apply the proposed method to the well-known risk neutral Heston model and show that in our proposed methodology a one-dimensional state process already results in highly satisfactory calibration results. We also show that numerical stability of the model parameters with respect to option market data is greatly improved. To show the relevance of our method, we compared the estimated Heston model from the proposed methodology to a basic parameter setting. We use both models to compute the present and future values of a unit-linked product of a fictive insurer. We find large differences in this simplified case study, and therefore we advice to use the proposed methodology for calibration, valuation and simulation.

In Chapter 6, we presented the Financial Nested Simulations (FiNS) framework, which is a high performance computing framework. FiNS is able to utilize both the CPU and the GPU to accelerate nested simulations. Due to the compute-intensive nature of nested simulations, CPU-only implementations lack the ability to provide sufficient performance, while GPU-only solutions are unable to efficiently utilize the hardware and lead to disappointing results. We demonstrated the use of FiNS for an Asset Liability Management application, which required multiple CPU processes to offload calculations to the same GPU. The results demonstrate that FiNS achieves very good parallel efficiency.

7.2. OUTLOOK

Due to more strict regulatory requirements of insurance companies, improved quality of computation hardware and increased availability of data, models for valuation and risk management are constantly improving. After completing the research in this thesis we came up with the following ideas for future research.

We choose basic risk neutral models, such as the Heston and Hull-White models, for solving calibration and valuation problems. Other models may also be considered, that are, for example, better able to model the implied volatility structures observed in the option market data.

When we considered the valuation of options, we always applied the risk neutral valuation methodology. However, the assumption of a complete market does not hold in this context. The Black-Scholes model assumes that there is always enough liquidity in the market for hedging, i.e. a complete market. In reality, this is unfortunately not the case. Furthermore, the market of insurance products is far from complete. This is on the one hand due to the relative small size of the market and on the other hand due to fact that insurance companies compete with each other to provide the best insurance prod-

ucts to their current and future policyholders. So, the assumption of a complete market may not be realistic in this context. Much is known about the valuation in incomplete markets although an industrial standard valuation methodology is still missing.

The calibration of the risk neutral models is a crucial step in the valuation process since the value of the embedded options heavily depends on those parameters. It's market practice to use option market data for the calibration of risk neutral models. However, such data is only available for trade-able assets. An interesting research area is how to calibrate variables that are not trade-able such as direct real estate, infrastructure and private equity. The complete market assumption should also be relaxed here. Furthermore, an interesting area is to not only focus on available option market data for calibration, but also on data of embedded options. In this way the risks of the insurance product could be better valued.

There are a number of research directions with respect to high performance computing. The implementation on any many-core architecture, like Intel MIC. The possibility of providing an intuitive front-end for a computation framework together with a computational infrastructure (e.g., in the cloud), enabling financial specialists to use it as a computational service.

The stability of the radial basis function (RBF) is important for accurate calibration to option data. We have explored the standard RBFs, but other RBFs may also be considered that lead to even better calibrations of the shape parameter, i.e. with high accuracy and numerical stability.

The big data application increasingly gains popularity in practice and machine learning techniques are developing fast. The combination of these two can be used to extract important information from the available data. The relevant signals can for example be used in combination with the Option Interpolation Model discussed in Chapter 4 or to improve the linear mapping that is used in Chapter 5.

References

[1] L.A. Abbas-Turki, S. Vialle, B. Lapeyre, and P. Mercier. Pricing derivatives on graphics processing units using Monte Carlo simulation. *Concurrency and Computation: Practice and Experience*, 26(9):1679–1697, 2014.

[2] Y. Aït-Sahalia and R. Kimmel. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83(2):413–452, 2007.

[3] L. Andersen. Simple and efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance*, 11:1–42, 2008.

[4] L. Andersen and J. Andreasen. Volatile volatilities. *Risk*, 5:163–168, 2002.

[5] C. Augonnet, S. Thibault, R. Namyst, and P. Wacrenier. StarPU: a unified platform for task scheduling on heterogeneous multicore architectures. *Concurrency and Computation: Practice and Experience*, 23(2):187–198, 2011.

[6] D. Bauer, D. Bergmann, and R. Kiesel. On the risk neutral valuation of life insurance contracts with numerical methods in view. *Astin Bulletin*, 40(1):65–95, 2010.

[7] D. Bauer, D. Bergmann, and A. Reuss. Solvency II and nested simulations - a least-squares Monte Carlo approach. In *Proceedings of the 2010 ICA congress*, 2010.

[8] D. Bauer, A. Reuss, and D. Singer. On the calculation of the solvency capital requirement based on nested simulations. *Astin Bulletin*, 42(2):453–499, 2012.

[9] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.

[10] A. Bernemann, R. Schreyer, and K. Spanderen. Accelerating exotic option pricing and model calibration using GPUs. *Available at SSRN 1753596*, 2011.

[11] M. Bertus, H. Hollans, and S. Swidler. Hedging house price risk with CME futures contracts: The case of Las Vegas residential real estate. *Journal of Real Estate Finance and Economics*, 37(3):265–279, 2008.

[12] T. Björk and E. Clapham. On the pricing of real estate index linked swaps. *Journal of Housing Economics*, 11:418–432, 2002.

[13] F. Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3:167–179, 1976.

[14] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[15] G. F. Blundell and C. W. R. Ward. Property portfolio allocation: a multi-factor model. *Journal of Property Research*, 4(2):145–156, 1987.

[16] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[17] J. P. Boyd. Six strategies for defeating the Runge phenomenon in Gaussian radial basis functions on a finite interval. *Computers & Mathematics with Applications*, 60(12):3108–3122, 2010.

[18] D. Brigo and F. Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Finance, Berlin, 2007.

[19] M. Broadie and O. Kaya. Exact simulation of stochastic volatility and other affine jump processes. *Operations Research*, 54(2):217–231, 2006.

[20] G. Brown and G. Matysiak. Valuation smoothing without temporal aggregation. *Journal of Property Research*, 15(2):89–103, 2002.

[21] R. G. Brown. *Statistical forecasting for inventory control*. McGraw-Hill, New York, 1959.

[22] G. W. Buetow and J. D. Albert. The pricing of embedded options in real estate lease contracts. *Journal of Real Estate Research*, 15:253–265, 1998.

[23] M. D. Buhmann. *Radial basis functions: theory and implementations*. Cambridge University Press, Cambridge, 2003.

[24] J. R. Bunch and J. E. Hopcroft. Triangular factorization and inversion by fast matrix multiplication. *Mathematics of Computation*, 28(125):231–236, 1974.

[25] R. J. Buttimer and J. B. Kau. A model for pricing securities dependent upon a real estate index. *Journal of Housing Economics*, 6:16–30, 1997.

[26] O. Cappé, E. Moulines, and T. Rydén. Inference in hidden Markov models. In *Proceedings of EUSFLAT Conference*, pages 14–16, 2009.

[27] J. C. Carr, R. K. Beatson, B. C. McCallum, W. R. Fright, T. J. McLennan, and T. J. Mitchell. Smooth surface reconstruction from noisy range data. In *Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 119–126. ACM, 2003.

[28] P. Carr and D. Madan. Option valuation using the Fast Fourier Transform. *Journal of Computational Finance*, 2(4):61–73, 1999.

[29] P. Carr and L. Wu. A new simple approach for for constructing implied volatility surfaces. *Preprint available at SSRN*, 2010.

[30] Z. Chen, A. Pelsser, and E. Ponds. Evaluating the UK and Dutch Defined Benefit pension policies using the holistic balance sheet framework. *Insurance: Mathematics and Economics*, 58:89–102, 2014.

[31] R. Cont and J. Da Fonseca. Dynamics of implied volatility surfaces. *Quantitative Finance*, 2(1):45–60, 2002.

[32] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 53:385–407, 1985.

[33] J. Cramwinckel, S.N. Singor, and A. L. Varbanescu. Fins: a framework for accelerating nested simulations on heterogeneous platforms. In *European Conference on Parallel Processing*, pages 246–257. Springer, 2015.

[34] F. De Jong, J. Driessen, and O. van Hemert. Hedging house price risk: Portfolio choice with housing futures, 2008.

[35] S De Marchi. On optimal center locations for radial basis function interpolation: computational aspects. *Universitae Politecnico di Torino*, 61(3):343–358, 2003.

[36] P. De Vries, G. Mariën, J. de Haan, and E. Van der Wal. A house price index based on the SPAR method. *Journal of Housing Economics*, 18(3):214–223, 2009.

[37] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94–128, 1999.

[38] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, pages 1–38, 1977.

[39] G. Dionne. Risk management: History, definition, and critique. *Risk Management and Insurance Review*, 16(2):147–166, 2013.

[40] J. A. Doornik and D. F. Hendry. *Empirical Econometric Modelling using PcGive: Volumes I-III*. Timberlake Consultants Press, London, 5th edition, 2007.

[41] M. S. Dorfman and D. A. Cather. *Introduction to Risk Management and Insurance*. Prentice-Hall, New Jersey, 2013.

[42] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *Handbook of Nonlinear Filtering*, 12:656–704, 2009.

[43] J.C. Duan and J.G. Simonato. Empirical martingale simulation for asset prices. *Management Science*, 44(9):1218–1233, 1998.

[44] J. Durbin and S. J. Koopman. *Time series analysis by state space methods*. Oxford University Press, Oxford, 2012.

[45] F. J. Fabozzi, R. J. Shiller, and R. S. Tunaru. Hedging real-estate risk. *Journal of Portfolio Management*, 35(5):92–103, 2009.

[46] F. J. Fabozzi, R. J. Shiller, and R. S. Tunaru. A pricing framework for real estate derivatives. *European Financial Management*, 18(5):762–789, 2012.

[47] F. Fang and C. W. Oosterlee. A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826–848, 2008.

[48] G. E. Fasshauer. *Meshfree approximation methods with MATLAB*, volume 6. World Scientific, Singapore, 2007.

[49] G. E. Fasshauer and M. J. McCourt. Stable evaluation of Gaussian radial basis function interpolants. *SIAM Journal on Scientific Computing*, 34(2):A737–A762, 2012.

[50] Q. Feng, S. Jain, P. Karlsson, D. Kandhai, and C. W. Oosterlee. Efficient computation of exposure profiles on real world and risk neutral scenarios for Bermudan swaptions. *Journal of Computational Finance, Forthcoming*.

[51] J.L. Fernández, A.M. Ferreiro, J.A. García-Rodríguez, A. Leitao, J.G. López-Salas, and C. Vázquez. Static and dynamic SABR stochastic volatility models: calibration and option pricing using GPUs. *Mathematics and Computers in Simulation*, 94:55–75, 2013.

[52] I. Fisher. The theory of interest. *New York*, 43, 1930.

[53] N. Flyer and E. Lehto. Rotational transport on a sphere: Local node refinement with radial basis functions. *Journal of Computational Physics*, 229(6):1954–1969, 2010.

[54] B. Fornberg, T.A. Driscoll, G. Wright, and R. Charles. Observations on the behavior of radial basis function approximations near boundaries. *Computers & Mathematics with Applications*, 43(3):473–490, 2002.

[55] B. Fornberg and N. Flyer. *A Primer on Radial Basis Functions with Applications to the Geosciences*, volume 87. SIAM, 2015.

[56] B. Fornberg, E. Larsson, and N. Flyer. Stable computations with Gaussian radial basis functions. *SIAM Journal on Scientific Computing*, 33(2):869–892, 2011.

[57] B. Fornberg and C. Piret. A stable algorithm for flat radial basis functions on a sphere. *SIAM Journal on Scientific Computing*, 30(1):60–80, 2007.

[58] M. K. Francke. Repeat sales index for thin markets. *The Journal of Real Estate Finance and Economics*, 41(1):24–52, 2010.

[59] M. K. Francke, S. Vujic, and G. A. Vos. Evaluation of house prices models using an ECM approach: The case of the Netherlands. Ortec Finance Methodological Working Paper, N0. 2009-05, 2009.

[60] J. Gatheral. *The volatility surface: a practitioner's guide*, volume 357. John Wiley & Sons, New Jersey, 2006.

[61] J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

[62] D. Geltner and J.D. Fisher. Pricing and index considerations in commercial real estate derivatives. *The Journal of Portfolio Management*, 33(5):99–118, 2007.

[63] D. Geltner, B. D. MacGregor, and G. M. Schwann. Appraisal smoothing and price discovery in real estate markets. *Urban Studies*, 40(5–6):1047–1064, 2003.

[64] P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53. Springer, New York, 2004.

[65] L. A. Grzelak and C. W. Oosterlee. On the Heston model with stochastic interest rates. *SIAM Journal on Financial Mathematics*, 2(1):255–286, 2011.

[66] L. A. Grzelak and C. W. Oosterlee. On cross-currency models with stochastic volatility and correlated interest rates. *Applied Mathematical Finance*, 19(1):1–35, 2012.

[67] N. A. Gumerov and R. Duraiswami. Fast radial basis function interpolation via preconditioned Krylov iteration. *SIAM Journal on Scientific Computing*, 29(5):1876–1899, 2007.

[68] S. S. Haykin. *Kalman filtering and neural networks*. Wiley Online Library, 2001.

[69] A. Heryudono, E. Larsson, A. Ramage, and L. von Sydow. Preconditioning for radial basis function partition of unity methods. *Journal of Scientific Computing*, 67(3):1089–1109, 2016.

[70] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.

[71] C. Homescu. Implied volatility surface: Construction methodologies and characteristics. *arXiv preprint arXiv:1107.1834*, 2011.

[72] Y.C. Hon and X.Z. Mao. A radial basis function method for solving options pricing models. *Journal of Financial Engineering*, 8:31–50, 1999.

[73] J. C. Hull. *Options, futures, and other derivatives*. Pearson Education India, 2006.

[74] A. Iske. Scattered data modelling using radial basis functions. In *Tutorials on Multiresolution in Geometric Modelling*, pages 205–242. Springer, 2002.

[75] A. Iske. *Multiresolution methods in scattered data modelling*, volume 37. Springer Science & Business Media, 2004.

[76] R. Jarrow and Y. Yildirim. Pricing treasury inflation protected securities and related derivatives using an HJM model. *Journal of Financial and Quantitative Analysis*, 38(2):337–358, 2003.

[77] E. Jokivuolle. Pricing European options on autocorrelated indices. *Journal of Derivatives*, 6(2):39–52, 1998.

[78] C. S. Jones. The dynamics of stochastic volatility: evidence from underlying and option markets. *Journal of Econometrics*, 116:181–224, 2003.

[79] M. S. Joshi. Graphical Asian options. *Wilmott Journal*, 2(2):97–107, 2010.

[80] M. Kamal and J. Gatheral. Implied volatility surface. *Encyclopedia of Quantitative Finance*, 2010.

[81] J. D. Kau, D. C. Keenan, W. J. Muller, and J. F. Epperson. Pricing commercial mortgages and their mortgage-backed securities. *Journal of Real Estate Finance and Economics*, 3(4):333–356, 1990.

[82] K. Kofler, I. Grasso, B. Cosenza, and T. Fahringer. An automatic input-sensitive approach for heterogeneous task partitioning. In *ICS 2013*, 2013.

[83] B. Kramer. Hybrid forms of sale: Valuation and risk for housing corporations. *Property Research Quarterly*, 4:49–56, 2008.

[84] S. Kruse. Pricing of inflation-indexed options under the assumption of a lognormal inflation index as well as under stochastic volatility, 2007.

[85] E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics*, 8:115–131, 2004.

[86] A. Lee, C. Yau, M. B. Giles, A. Doucet, and C. C. Holmes. On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods. *Journal of Computational and Graphical Statistics*, 19(4), 2010.

[87] E. Levy. Pricing European average rate currency options. *Journal of International Money and Finance*, 14:474–491, 1992.

[88] J.P. Lewis, F. Pighin, and K. Anjyo. Scattered data interpolation and approximation for computer graphics. In *ACM SIGGRAPH ASIA 2010 Courses*, page 2. ACM, 2010.

[89] S. L. Liao and C. C. Chen. The valuation of European options when asset returns are autocorrelated. *Journal of Futures Markets*, 26(1):85–102, 2006.

[90] F. Lindsten. An efficient stochastic approximation EM algorithm using conditional particle filters. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6274–6278. IEEE, 2013.

[91] F. Lindsten, M. I. Jordan, and T. B. Schön. Particle Gibbs with ancestor sampling. *Journal of Machine Learning Research*, 15(1):2145–2184, 2014.

[92] L. Ling and E. J. Kansa. A least-squares preconditioner for radial basis functions collocation methods. *Advances in Computational Mathematics*, 23(1):31–54, 2005.

[93] A. W. Lo and J. Wang. Implementing option pricing models when asset returns are predictable. *Journal of Finance*, 50(1):87–129, 1995.

[94] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001.

[95] R. Lord. Partially exact and bounded approximations for arithmetic Asian options. *Journal of Computational Finance*, 10(2):1–52, 2006.

[96] J. Luitjens. CUDA streams - best practices and pitfalls. Presentation GTC '14.

[97] C.K. Luk, S. Hong, and H. Kim. Qilin: exploiting parallelism on heterogeneous multiprocessors with adaptive mapping. In *MICRO 2009*, 2009.

[98] Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, 2006.

[99] N. Mai-Duy and T. Tran-Cong. Approximation of function and its derivatives using radial basis function networks. *Applied Mathematical Modelling*, 27(3):197–220, 2003.

[100] C. A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive approximation*, 2(1):11–22, 1986.

[101] K. P. Murphy. *Dynamic Bayesian networks: representation, inference and learning*. PhD thesis, University of California, Berkeley, 2002.

[102] D. B. Nelson. ARCH models as diffusion approximations. *Journal of Econometrics*, 45:7–38, 1990.

[103] H. Niederreiter. Quasi-Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society*, 84(6):957–1041, 1978.

[104] NVIDIA. Kepler GK110 architecture whitepaper, v1.0, 2012.

[105] NVIDIA. Sharing a GPU between MPI processes: Multi-process service (MPS) overview, 2013. Technical Brief TB-06737-003.

[106] K. Patel and R. Pereira. Pricing property index linked swaps with counterparty default risk. *Journal of Real Estate Finance and Economics*, 36(1):5–21, 1996.

[107] J. Planas, R. M. Badia, E. Ayguadé, and J. Labarta. Self-adaptive ompSs tasks in heterogeneous environments. In *IPDPS 2013*, 2013.

[108] S. A. Press, W. H.and Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*. Cambridge University Press, Cambridge, 1996.

[109] S. Rippa. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. *Advances in Computational Mathematics*, 11(2-3):193–210, 1999.

[110] C. Runge. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. *Zeitschrift für Mathematik und Physik*, 46(224-243):20, 1901.

[111] Y.V.S.S. Sanyasiraju and C. Satyanarayana. On optimization of the RBF shape parameter in a grid-free local scheme for convection dominated problems over non-uniform centers. *Applied Mathematical Modelling*, 37(12):7245–7272, 2013.

[112] R. Schaback. Error estimates and condition numbers for radial basis function interpolation. *Advances in Computational Mathematics*, 3(3):251–264, 1995.

[113] M. Scheuerer. An alternative procedure for selecting a good value for the parameter c in RBF-interpolation. *Advances in Computational Mathematics*, 34(1):105–126, 2011.

[114] I. J. Schoenberg. Metric spaces and completely monotone functions. *Annals of Mathematics*, pages 811–841, 1938.

[115] T. B. Schön, A. Wills, and B. Ninness. System identification of nonlinear state-space models. *Automatica*, 47(1):39–49, 2011.

[116] J. Shen, A. L. Varbanescu, and H. Sips. Look before you leap: Using the right hardware resources to accelerate applications. In *HPCC*, 2014.

[117] J. Shen, A. L. Varbanescu, P. Z., Y. Lu, and H. Sips. Improving performance by matching imbalanced workloads with heterogeneous platforms. In *ICS*, 2014.

[118] R. J. Shiller. Derivatives markets for home prices. Discussion Paper 1648, National Bureau of Economic Research, 2008.

[119] R. J. Shiller and A. N. Weiss. Home equity insurance. *Journal of Real Estate Finance and Economics*, 19(1):21–47, 1999.

[120] D. Simon. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

[121] S. N. Singor, L. A. Grzelak, D. D.B. Van Bragt, and C. W. Oosterlee. Pricing inflation products with stochastic volatility and stochastic interest rates. *Insurance: Mathematics and Economics*, 52(2):286–299, 2013.

[122] S.N. Singor, A. Boer, J.S.C. Alberts, and C.W. Oosterlee. On the modelling of nested risk-neutral stochastic processes with applications in insurance. *Applied Mathematical Finance*, pages 1–35, 2017.

[123] H. Steehouwer. *Macroeconomic Scenarios and Reality: A Frequency Domain Approach for Analyzing Historical Time Series and Generating Scenarios for the Future*. PhD thesis, Free University of Amsterdam, 2005.

[124] H. J. Stein. Joining risks and rewards. *Available at SSRN 2368905*, 2013.

[125] H. J. Stein. Fixing underexposed snapshots–proper computation of credit exposures under the real world and risk neutral measures. *Available at SSRN 2365540*, 2014.

[126] J. M. Syz and P. Vanini. Arbitrage free price bounds for property derivatives. *Journal of Real Estate Finance and Economics*, 43(3):281–298, 2011.

[127] Y. Tian, Z. Zhu, F. C. Klebaner, and K. Hamza. Option pricing with the SABR model on the GPU. In *High Performance Computational Finance (WHPCF), 2010 IEEE Workshop on*. IEEE, 2010.

[128] Y. Tian, Z. Zhu, F. C. Klebaner, and K. Hamza. Pricing barrier and american options under the sabr model on the graphics processing unit. *Concurrency and Computation: Practice and Experience*, 24(8), 2012.

[129] D. Van Bragt, M. K. Francke, S. N. Singor, and A. Pelsser. Risk neutral valuation of real estate derivatives. *The Journal of Derivatives*, 23(1):89–110, 2015.

[130] D. Van Bragt, H. Steehouwer, and B. Waalwijk. Market consistent alm for life insurers—steps toward solvency ii. *The Geneva Papers on Risk and Insurance-Issues and Practice*, 35(1):92–109, 2010.

[131] A. van Bussel and R. Mahieu. A repeat sales index for residential property in the Netherlands. *Available at SSRN: http://ssrn.com/abstract=607 or http://dx.doi.org/10.2139/ssrn.607*, 1996.

[132] A. van Haastrecht, R. Lord, A. Pelsser, and D. Schrager. Pricing long-dated insurance contracts with stochastic interest rates and stochastic volatility. *Insurance: Mathematics and Economics*, 45(3):436–448, 2009.

[133] A. van Haastrecht and A. Pelsser. Generic pricing of FX, inflation and stock options under stochastic interest rates and stochastic volatility. *Quantitative Finance*, 11(5):665–691, 2011.

[134] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.

[135] E. J. Vaughan and T. Vaughan. *Fundamentals of risk and insurance*. John Wiley & Sons, 2007.

[136] G. Wahba. *Spline models for observational data*. SIAM, 1990.

[137] F. Wende, T. Steinke, and F. Cordes. Multi-threaded Kernel Offloading to GPGPU Using Hyper-Q on Kepler Architecture. *ZIB-Rep. 14-19 June 2014*, 2014.

[138] H. Wendland. *Scattered data approximation*, volume 17. Cambridge University Press, Cambridge, 2004.

[139] Z. Wu and R. Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA journal of Numerical Analysis*, 13(1):13–27, 1993.

[140] S. A. Zenios and W. T. Ziemba. *Handbook of Asset and Liability Management: Applications and case studies*, volume 2. Elsevier, 2007.

Curriculum Vitæ

Stefan Nico SINGOR

02-12-1985 Born in Oud-Beijerland, the Netherlands.

EDUCATION

1998-2004 VWO
C.S.G. Willem van Oranje, Oud-Beijerland, the Netherlands

2004-2007 Bachelor Mathematics
Utrecht University, Utrecht, the Netherlands

2007-2009 Master Applied Mathematics (Cum laude)
Delft University of Technology, Delft, the Netherlands

2010-2017 PhD Applied Mathematics
Delft University of Technology, Delft, the Netherlands
Thesis: Improving risk neutral valuation techniques with applications in insurance
Promotor: Prof. dr. ir. C.W. Oosterlee

List of Publications

5. **S.N. Singor**, A. Boer, J.S.C. Alberts and C.W. Oosterlee, On the modeling of nested risk neutral stochastic processes, *Applied Mathematical Finance*, 1-35, 2017.

4. **S.N. Singor**, E. Schols and C.W. Oosterlee, Approximation of the value of insurance liability contracts using radial basis functions, submitted for publication, 2017.

3. J. Cramwinckel, **S.N. Singor**, A.L. Vabanescu, FiNS: a framework for accelerating nested simulations on heterogeneous platforms, *European Conference on Parallel Processing*, pages 246-257, Springer, 2015.

2. D.B. Van Bragt, M.K. Francke, **S.N. Singor** and A. Pelsser, Risk neutral valuation of real estate derivatives, *Journal of Derivatives*, 23(1):89-110, 2015.

1. **S.N. Singor**, L.A. Grzelak, D.B. Van Bragt and C.W. Oosterlee, Pricing inflation products with stochastic volatility and stochastic interest rates, *Insurance: Mathematics and Economics*, 52(2):286-299, 2013.

List of Attended Conferences with Presentation

4. SIAM Conference on Financial Mathematics & Engineering, Austin, Texas, United States, November 2016.

3. SIAM Conference on Financial Mathematics & Engineering, Chicago, Illinois, United States, November 2014.

2. SIAM Conference on Financial Mathematics & Engineering, Minnesota, Minneapolis, United States, July 2012.

1. SIAM Conference on Financial Mathematics & Engineering, San Francisco, California, United States, November 2010.

Acknowledgement

This thesis concludes my PhD research at the Delft University of Technology. I would like to thank a number of people that contributed with their motivation, expertise and work to the realization of this thesis.

First, I would like to thank my supervisor Kees Oosterlee for the guidance and support during the PhD project. Kees is a true motivator and he encouraged and supported me along the project. I'm also grateful for his valuable advise and insights during the PhD project. Thanks to Kees Oosterlee I could participate in a number of mini-symposiums at SIAM conferences to present my research findings.

Ortec Finance gave me the opportunity to work as a part-time PhD-researcher and as a consultant in financial risk management for which I'm very grateful. Working at Ortec Finance is a pleasure and it gave me the flexibility to discuss with colleagues about practical insights relevant for research. I applied and tested my research findings in practice and commercialized the constructed techniques. In particular, I would like to thank Guus Boender, who unfortunately past away in September 2014, Hens Steehouwer and Wendy Montulet for their motivation and support and David van Bragt and Alex Boer also for the in-depth discussions and their valuable insights.

I would like to thank the other authors Lech Grzelak, Marc Francke, Antoon Pelsser, Joris Cramwinckel, Anna Vabanescu, Eric Schols and Pien Alberts for their contributions.

Without the support of my family I could never complete this PhD project. My gratitude goes out to my parents who supported me during my school education. I would like to thank my wife Sabine for her support and patience during the periods off-work while I was working on my research. Combining research with young children is not an easy combination. Although my daughters unintentionally kept me up at nights, the extra time was useful to complete the research. Their happiness and enthusiasm was a good motivation for finishing the PhD research.

Stefan Singor
November 2017