Interest Rate Derivative Pricing with Stochastic Volatility

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
dinsdag 25 september 2012 om 15:00 uur

door

Bin CHEN
Master of Science in Engineering and Policy Analysis

geboren te Hangzhou, China

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr.ir. C.W. Oosterlee

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr.ir. C.W. Oosterlee,	TU Delft, promotor
Prof.dr. C. Vázquez Cendón,	University of Coruña, Spain
Prof.dr.ir. B. Koren,	TU Eindhoven
Prof.dr. M.H. Vellekoop,	University of Amsterdam
Drs. S. van Weeren,	Rabobank International
Prof.dr.ir. A.W. Heemink,	TU Delft
Prof.dr.ir. C. Vuik,	TU Delft
Prof. dr. F.H.J. Redig,	TU Delft, reservelid

Interest Rate Derivative Pricing with Stochastic Volatility.
Dissertation at Delft University of Technology.

The work described in this thesis was financially supported by Rabobank International and CWI Amsterdam.

ISBN 978-94-6203-052-7

Printed in The Netherlands by: Wöhrmann Print Service

Acknowledgements

This thesis concludes my four years of research in Centrum Wiskunde Informatica (CWI) and the Derivative Research & Validation (DR&V) team of Rabobank International. In the past four years, many people have helped me go through the painstaking process towards the final completion of this thesis. At this moment, I would like to express my thanks to all those who contributed in many ways to the success of this research.

First and foremost, I would like to thank Prof. dr. ir. Cornelis Oosterlee for his academic support, project management skills, and kind patience. Without his continuing advice, mathematically as well as personally, this thesis would not have been possible. I was especially amazed by his readiness to work at any time. Occasionally I would send emails at 11 o'clock in the evening or 6:30 in the morning, but I always received response within a very short period of time, even on weekends. His hard-working attitude was a great motivation for me.

My special words of thanks should also go to Drs. Sacha van Weeren for his sharp opinion, deep knowledge of financial products and no-nonsense attitude, which were a great inspiration driving my research. Most of the time, his no-nonsense attitude was combined with a high level of directness. A discussion with him could be tough if I presented an idea which I had not yet fully thought through or if my report did not provide strong enough evidence to support the argument. Thus I would have to constantly push further for a deeper understanding, better arguments and more thoughtful solutions, which was reflected in the content of my research papers.

Sincere gratitude goes to Dr. Lech Grzelak who acted as the co-author of one of my four research papers, but our cooperation went far beyond that paper. He is an amazingly fast learner with great technical skills. It was great pleasure to work with him closely and to have stimulating discussions with him on a lot of things inside and outside our research.

I am also grateful to Dr. Natalia Borovykh and Dr. Tim Dijkstra, who are scientists turned finance professionals. With them I could discuss many details of my work and they gave me many useful comments, which helped me to improve the scientific quality of my research.

I should not forget my friendly colleagues in CWI, Linda Plantagie, Benjamin Sanderse, Willem Haverkort, Shashi Jain, Marjon Ruijter and Bram van Es,

with whom I attend weekly scientific meetings and we enjoyed regular sport activities together. They made my stay in CWI so much more enjoyable.

My special words of appreciation also go to my other colleagues in DR&V Rabobank International, Erik van Raaij, Erik Hennink, Thomas Zelders, Marcel Wijnen and Maurice Lutterot, among others. It has been great to work with you!

Finally and most importantly, I would like to thank my family for their support, especially my wife Fei. It is her love, encouragement and devotion that have enabled me to go this far. For all her sacrifices and tolerance of my occasionally vulgar moods, I cannot express my gratitude enough through words. Thank you, Fei, for everything.

Bin Chen
Utrecht, August 2012

Summary

Interest Rate Derivative Pricing with Stochastic Volatility

Bin Chen

One purpose of exotic derivative pricing models is to enable financial institutions to quantify and manage their financial risk, arising from large books of portfolios. These portfolios consist of many non-standard exotic financial products. Risk is managed by means of the evaluation of sensitivity parameters, i.e. the so-called Greeks, the deltas, vegas, gammas and also volgas, vannas, and others. In practice, practitioners do not expect an exotic derivative pricing model to be a high precision *predictive* model. What is important is a high precision *replication of the hedging instruments*, as well as efficient computation with the model.

Plain vanilla interest rate options like swaptions and caps are liquidly traded instruments, serving as fundamental building blocks of hedging portfolios for exotic products. In the early twenty-first century, the so-called implied volatility skew and smile in the market became pronounced in the interest rate plain vanilla market. The *stochastic alpha beta rho (SABR)* model [46] then became widely accepted as the market standard to model this implied volatility skew/smile.

The model's popularity is due to the existence of an *accurate analytic approximation* for the implied volatilities, presented by Hagan et al., in [46]. This approximation formula is often used by practitioners to inter- and extrapolate the implied volatility surface. The application of the SABR model is so prevalent that one can even observe SABR-type implied volatility curves in the market nowadays (which means that the SABR model can perfectly resemble one set of market implied volatilities with different strike prices).

This PhD thesis considers the SABR model as its basis for further extension, and focuses on the various problems arising from the application of the SABR model in both plain vanilla and exotic option pricing, from a modelling as well as numerical point of view.

In Chapter 2, we present an analytic approximation to the convexity correction of Constant Maturity Swap (CMS) products under a two-factor SABR model by means of *small time asymptotic expansion* technique.

In Chapter 3, we apply the *small time asymptotic expansion* differently, to a problem of approximating the first and second moments of the integrated variance of the log-normal volatility process in the context of defining a low-bias discretization scheme for the SABR model. With the approximated moment information, we can approximate the density of the integrated variance by means of a log-normal distribution with the first two moments matched to that information. The conditional SABR process turns out to be a squared Bessel process, given the terminal volatility level and the integrated variance. Based on the idea of mixing conditional distributions and a direct inversion of the noncentral chi-square distributions, we propose the low-bias SABR Monte Carlo scheme. The low-bias scheme can handle the asset price process in the vicinity of the zero boundary well. The scheme is stable and exhibits a superior convergence behaviour compared to the truncated Euler scheme.

In Chapter 4, we extend the discretization scheme proposed in Chapter 3 towards a SABR model with stochastic interest rate in the form of a Hull-White short rate model, the SABR-HW model. The hybrid model is meant for pricing long-dated equity-interest-rate linked exotic options with exposure to both the interest rate and the equity price risk. To facilitate the calibration of the SABR-HW model, we propose a projection formula, mapping the SABR-HW model parameters onto the parameters of the *nearest* SABR model. The numerical inversion of the projection formula can be used to calibrate the model.

In Chapter 5, we focus on a version of the stochastic volatility LIBOR Market Model with time-dependent skew and volatility parameters. As a result of choosing time-dependent parameters, the model has the flexibility to match to the market quotes of an entire swaption cube (in terms of various combinations of expiry, tenor and strike), as observed in the current interest rate market. Thus, this model is in principle well-suited for managing the risk of a complete exotic option trading book in a financial institution, consisting of both exotic options and its plain vanilla hedge instruments.

The calibration of the model to the swaption quotes relies on a model-mapping procedure, which relates the model parameters (most often time-dependent) in a high-dimensional LMM model to swaption prices. The model-mapping procedure maps the high-dimensional swap rate dynamics implied by the model onto a one-dimensional displaced diffusion process with time-dependent coefficients. Those time-dependent parameters are subsequently averaged to obtain the effective constant parameters of the projected model. Two known projection methods that are available in the literature, the freezing projection and the more involved Markov projection, have been compared within the calibration process. The basic freezing projection achieves a good accuracy at significantly less computational cost in our tests, and it is thus applied within the calibration purpose.

A second advantage of the freezing projection formula is that it enables us to formulate the time-dependent skew calibration problem as a convex optimization problem. Our contribution in this chapter is the convex optimization formulation of the skew calibration problem. Based on the convex formulation, we are able to translate the calibration of a large number of free variables into a well-known quadratic programming problem formulation, for which efficient

algorithms are available. The convexity of the formulated optimization problem guarantees the obtained solution to be a global optimum. The stability of the procedure can be beneficial for application in the day-to-day derivative trading practice, i.e. the daily re-calibration and hedging.

Samenvatting

Interest Rate Derivative Pricing with Stochastic Volatility

Bin Chen

Eén van de doelstellingen van prijsmodellen voor financiële derivaten is het in staat stellen van financiële instellingen om hun financiële risico's, die voortvloeien uit grote portefeuilles, te kwantificeren en te beheren. Deze portefeuilles bestaan uit vele niet-standaard exotische financiële producten. Het risico wordt beheerd door middel van de evaluatie van gevoeligheidsparameters, i.e. de zogenaamde Grieken, de delta's, vega's, gamma's en ook volga's, vanna's, en anderen. In de praktijk blijkt dat beoefenaars niet verwachten dat een prijsmodel voor een exotisch derivaat een zeer precies *voorspellingsmodel* is. Wat belangrijk is, is een hoge precisie *replicatie van de indekkingsinstrumenten*, alsmede een efficiënte berekening met het model.

"Plain vanilla" rente-opties zoals swaptions en caps zijn liquide verhandelde instrumenten, die dienst doen als fundamentele bouwstenen van de indekkingsportefeuilles voor de exotische producten. In het begin van de eenentwintigste eeuw werden de zogenaamde *geïmpliceerde volatiliteits skew en smile* duidelijk geobserveerd in marktdata in de rentemarkt. Het *Stochastische Alpha Beta Rho (SABR)* model [46] werd vervolgens algemeen aanvaard als de marktstandaard om deze geïmpliceerde skew en smile te modelleren.

De populariteit van het model is te danken aan het bestaan van een *nauwkeurige analytische benadering* voor de geïmpliceerde volatiliteit, gepresenteerd door Hagan et al., in [46]. Deze benaderingsformule wordt vaak gebruikt door beoefenaars voor het inter- en extrapoleren van geïmpliceerde volatiliteitsoppervlakken. De toepassing van het SABR-model is zo overheersend dat men zelfs tegenwoordig geïmpliceerde volatiliteitscurves van het SABR-type in de markt kan waarnemen (wat betekent dat het SABR-model een set van markt geïmpliceerde volatiliteiten met verschillende uitoefenprijzen perfect kan repliceren). Dit proefschrift beschouwt het SABR-model als een basis voor verdere uitbreiding, en richt zich op de verschillende problemen die voortvloeien uit de toepassing

van het SABR-model in zowel "plain vanilla" als exotische optieprijzen, vanuit een modellerings- en numeriek oogpunt. In hoofdstuk 2 presenteren we een analytische benadering voor de convexiteitscorrectie van Constant Maturity Swap (CMS) producten onder een twee-factor SABR-model door middel van een *kleine tijdschaal asymptotische expansie*-techniek. In hoofdstuk 3 passen we de asymptotische expansie anders toe, op een probleem bij het benaderen van het eerste en tweede moment van de geïntegreerde variantie van het log-normale volatiliteitsproces om een lage-bias discretisatie voor het SABR-model te definiëren. Met de benaderde momenten-informatie, kunnen we de dichtheid van de geïntegreerde variantie door middel van een log-normale verdeling benaderen. Het SABR-proces blijkt een kwadratisch Bessel-proces te zijn, gegeven het eindtijd-volatiliteitsniveau en de geïntegreerde variantie. Gebaseerd op het idee om voorwaardelijke verdelingen en een directe inversie van de niet-centrale chi-kwadraat verdeling te vermengen, stellen wij het lage-bias SABR Monte Carlo schema voor. Het schema is stabiel en vertoont een beter convergentiegedrag in vergelijking met een aangepaste Euler-discretisatie.

In hoofdstuk 4 breiden we het discretisatie-schema dat voorgesteld werd in hoofdstuk 3 uit naar een SABR-model met stochastische korte-tijdrente in de vorm van een Hull-White model, het SABR-HW model. Het hybride model is bedoeld voor de prijsbepaling van langlopende equity-rente gekoppelde exotische opties met blootstelling aan zowel de rente als ook het equity-prijsrisico. Om de kalibratie van het SABR-HW-model te vereenvoudigen, stellen we een projectieformule voor, die de SABR-HW-modelparameters afbeeldt op de parameters van het *dichtstbijzijnde* SABR-model. Een numerieke inversie van deze projectieformule kan gebruikt worden om het model te kalibreren.

In hoofdstuk 5 richten we ons op een versie van het stochastische volatiliteit LIBOR marktmodel met tijdsafhankelijke skew en volatiliteitsparameters. Als gevolg van het kiezen van tijdsafhankelijke parameters heeft het model de flexibiliteit om marktnoteringen van een volledige swaption kubus (in termen van verschillende combinaties van uitoefendatum, looptijd en uitoefenprijs), zoals waargenomen in de hedendaagse rentemarkt te modelleren. Dus, dit model is in principe zeer geschikt voor beheersing van het risico van een complete exotische optie handelsportefeuille in een financiële instelling, bestaande uit zowel exotische opties als ook de "plain vanilla" indekkingsinstrumenten.

De kalibratie van het model op de swaption-noteringen is gebaseerd op een model-mapping procedure, die de modelparameters (meestal tijdsafhankelijk) koppelt in een hoog-dimensionaal LMM-model aan swaption-prijzen. De model-mapping procedure beeldt hoog-dimensionale swaprente-dynamica geïmpliceerd door het model af op een één-dimensionaal getransleerd diffusieproces met tijdsafhankelijke coëfficiënten. Vervolgens wordt het gemiddelde genomen van die tijdsafhankelijke parameters om *effectieve constante parameters* van het geprojecteerde model te verkrijgen. Twee bekende projectie-methoden die in de literatuur beschikbaar zijn, de freezing-projectie en de rekenintensievere Markov-projectie, werden vergeleken in het kalibreringsproces. De standaard freezing-projectie haalt een goede nauwkeurigheid tegen aanzienlijk lagere computationele kosten in onze tests, en wordt dus toegepast binnen het kalibratieproces.

Een tweede voordeel van de freezing-projectie-formule is dat het ons in staat stelt om het tijdsafhankelijke skew kalibratie-probleem te formuleren als een convex optimalisatieprobleem. Onze bijdrage in dit hoofdstuk is de convexe optimalisatie-formulering van het skew kalibratie-probleem. Op basis van de

convexe formulering, kunnen we de kalibratie van een groot aantal vrije variabelen vertalen in een bekende kwadratische programmeringsprobleemformulering, waarvoor efficiënte algoritmen beschikbaar zijn. De convexiteit van het geformuleerde optimaliseringsprobleem garandeert dat de verkregen oplossing een globaal optimum is. De stabiliteit van de procedure kan nuttig zijn voor toepassing in de dagelijkse handelspraktijk met derivaten, dat wil zeggen de dagelijkse re-kalibratie en indekking.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Prelude

> *Hurling brickbats at bankers is a popular pastime. The 'Occupy Wall Street' movement and its various offshoots complain that a malign* 1%*, many of them bankers, are ripping off the virtuous* 99%*.*
>
> –*Economist*, issue Jan. 7th, 2012

The world has never seen so much public anger towards speculation activities, banking bonuses, derivative trading, in short the banking industry. This popular anger is understandable, as the financial crisis of 2007-08 has produced the deepest recession since the 1930s. Many of the players at the heart of it seem to have got off scot-free. The biggest banks are bigger than ever, and bonuses are being paid once again.

Both the public and the regulators are worried about the size of the derivative trading business and its tight connection to the real economy. Fear remains for an 'interlocking fragility', since financial institutions are interrelated by their derivatives payments – when one fails, they all fail [104].

In the year 2008, the world witnessed the dramatic unfolding of a banking crisis, seemingly starting with the collapse of Lehman Brothers. We also witnessed the breakdown of ABN AMRO (a renowned major Dutch bank). Since that year, financial jobs have been cut, several prestigious firms have been wiped out and many exotic financial products have disappeared from the market.

Importantly, however, new rules and regulations aiming at curbing the oversized banking industry have been put on the table. These regulations have been created with sincere intentions, aiming to move the banking industry away from the "high-risk-high-return" investment banking business to become safer utility companies and serve their core social functions, i.e. taking deposits, giving out loans and transferring risk.

At the same time, there are several industry-wide trends changing the way transactions are done in the derivative business. One of these trends is the

strict implementation of a credit support annex (CSA), which requires over-the-counter [1] (OTC) derivatives to be cleared at a central clearing house and both parties of a financial contract would have to post collateral on the margins of the present value (PV) of the derivative contract. Another trend is the inclusion of the so-called credit value adjustment (CVA) in the valuation of the derivatives. The idea is to adjust the value of derivatives by the potential loss due to possible *counter party default*.

The joint effect of these two changes would be a substantial reduction in the profit margins (or, in financial terminology, a lower return on capital) of the derivative trading business. Moreover, there are very many new regulations from both national and super-national regulators aiming to curb the risk-taking activities seen in the derivatives trading business before the 2008 crisis. The combined effect of the above-mentioned trends will give a sharp downsizing of the derivative trading business. For some smaller players, it may be sensible to avoid the capital intensive business completely.

Given the fast changing landscape of the derivative trading business in the aftermath of the crisis, new understanding and new analytic tools are needed by practitioners, which sparks the need for new research.

1.1.1 Back to Basics

The trading volume of OTC interest rate derivatives reduced and then resurged quickly after the crisis. By June 2011, the global notional of the OTC interest rate derivatives was around $553 trillion, surpassing the pre-crisis peak of $458 trillion at June 2008 [2]. From the 2011 June figures we see, however, that the majority of OTC derivative notional volumes were *interest rate swaps* (around 80%), followed by interest rate options, like *swaptions, caps/floors*, etc. (10.19%) and forward rate agreements (FRAs) (10.08%). Among these OTC trades, Europe and the United States are by far the biggest markets with $219 and $170 trillion notional OTC interest rate derivatives, out of $553 trillion notional traded in the whole world [3].

Although notional values are not very meaningful in the derivative markets for assessing the total exposure of a market, they are indicative for the trading volumes in specific derivative instruments, and the industrial interest in each derivative. The statistics suggest that investors, corporations and banks are reaffirming interest rate derivatives as their major risk transferring tools. We would like to claim that complicated interest rate derivatives, that do not make much economical sense but mainly provide a high leverage for speculators like hedge funds, see a trading activity at historically low levels, whereas the trading in basic products which serve fundamental economical and financial purposes (e.g. risk transfer and hedging) increases.

Let us introduce some basic financial derivatives and explain the economical motivations and benefits of trading these products.

An *interest swap*, for example, translates one type of interest rate stream

[1] Over-the-counter, or *off-exchange*, trading refers to trades of financial instruments, such as stocks, bonds, commodities or derivatives, directly between two parties. It is contrasted with exchange trading, which occurs via facilities constructed for the purpose of trading (i.e. exchanges), such as futures exchanges or stock exchanges.

[2] See BIS Quarterly Review, June 2011, and the published notional amounts on the report.

[3] See BIS Quarterly Review, June 2008, and the published notional amounts on the report.

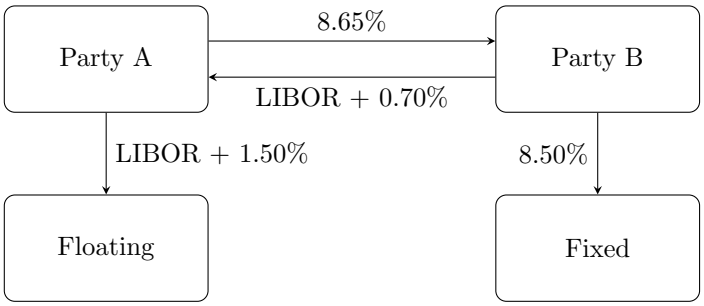

Figure 1.1: Borrower A uses a swap to convert a floating interest rate to a fixed rate with party B.

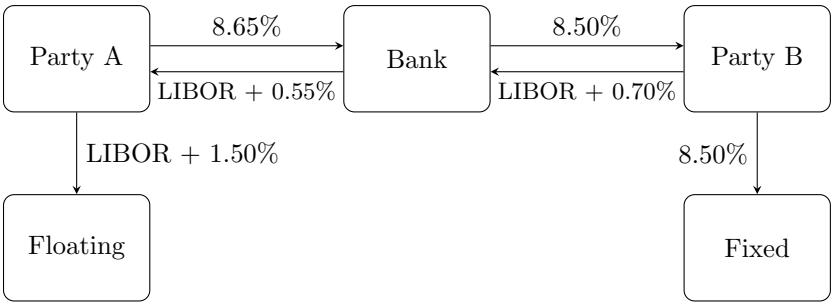

Figure 1.2: The cash flows of a swap contract when a bank serves as the financial intermediary connecting two parties of a swap contract.

into another, like a floating to a fixed rate, or vice versa. Each swap has two counter parties, where one party pays a fixed and receives a floating interest rate, while the other party receives a fixed and pays a floating rate (see Figure 1.1). Usually parties do not swap the payments directly, but each party starts a separate swap contract with a bank as intermediary. In order to bring these two parties together, a bank takes a so-called *spread in the swap payments* as its profit (see Figure 1.2).

The importance of interest rate swaps and related interest rate swaptions is based on the key role that swaps play when transferring and mitigating the interest rate risk to which a firm is exposed. Interest rate risk is the exposure of a firm to changing interest rates. It affects the profitability of a firm in various ways, like for example with changing costs due to changes in the interest rate. Companies with debt charged at variable rates (for example, based on the LIBOR rate, which measures the short maturity, say 3 months, borrowing costs of AA rated big international banks) will be exposed to increases in interest rates, whereas companies whose borrowing costs are completely, or partly, fixed will be exposed to a decrease of interest rates. The reverse is obviously true for companies with cash term deposits [101].

Of course, interest rates also have an impact on a firm's performance in a changing business environment (like a changing monetary policy by the central

bank) and have an impact on pension plans sponsored by a firm [4].

Interest rate swaps are arguably the most widely used and popular interest risk transfer instruments. Industrial sources estimate that more than 50% of new corporate debt issues are immediately *swapped from fixed into floating*, or from floating into fixed rates, by means of interest rate swaps [74].

Debt issuers that swap their debts typically want the right to cancel a swap at future points in time. Similarly, debt issuers that do not immediately swap their debts often want the right to enter into a pre-specified swap at some later date. A swaption provides the option to its holder to enter or stop a swap contract at a future time point. Thus, the swaption seems a natural instrument for debt issuers to provide them flexibility throughout the financing cycle [74].

However, not all financial contracts are benign in nature. For example, a contract called a *range accrual note* was once popular among hedge fund clients in the boom years. The contract paid a higher *coupon* than standard bond, when a so-called *benchmark rate* was within a certain range. The reason why this product was popular is that it offered investors the potential of receiving higher yields, especially when the volatility of interest rates was low for a longer time. In reality, it is costly and difficult to hedge such a product, due to its payment mechanism (based on multiple embedded digital options) and multiple hidden underlying risk exposures (highly sensitive to gamma exposure, due to the embedded digital options).

In contrast to the stable trading volumes of the basic derivative products, the exotic derivative business has witnessed a sharp decline after the crisis, and this market is not expected to rise any time soon. The most notable example is the rise and fall of the so-called collateralised debt obligation (CDO) contracts. In simple terms, a CDO can be thought of as a pool of bonds or other assets, which pay cash flows to investors in a prescribed sequence. If the cash collected from the pool of assets in the CDO is not sufficient to pay all of its investors, those in the *junior capital layers (tranches)* will suffer losses first, and after them the median layer investors. The CDO trading volumes grew significantly between 2000 and 2006, and then shrank dramatically during the subprime mortgage crisis, which began in 2007. Many of the assets held in these CDOs were subprime mortgage-backed bonds (more details are given in Lewis [73]).

The back-to-basics trend and dwindling demand for exotic options does not necessary make the finance profession a dull one. In fact, supposedly basic vanilla products, e.g. interest rate swaps, cross currency swaps and swaptions, have become more complicated than before (see Chapter 6.1 for a more detailed analysis). Nowadays, these products would have to be analyzed carefully with sophisticated models that were previously reserved for the exotic derivatives.

1.1.2 Financial Innovation

The cause of the 2008 financial crisis may not have been solely the excessive use of derivatives to gain higher, risky leverage, but also to some extent the financial infrastructure which was far behind the financial innovation. By financial infrastructure, we mean the exchanges (physical or electronic), legal rules, risk management systems and knowledge workforce, etc.

[4] Pension plans that carry liability and investment risk for a sponsor contain interest rate risk in a similar way as bonds.

As Scholes [99] argued, from an economic point of view 'infrastructure to support financial innovation must follow the innovation', as it would be too expensive to build all information links, legal rules, risk management control mechanisms, before the introduction of new products. Since successful innovations are difficult to predict, the infrastructure necessary to support innovation should lag behind the innovations themselves, which naturally increases the likelihood that control mechanisms will be insufficient at times to prevent a crisis happening. Failure, however, does not lead immediately to the conclusion that society will be better off with less innovation or vetting innovations before their initiation [99].

Instead of limiting the market mechanism to risk transfer and hedging, an improved financial infrastructure, in terms of better risk management systems, may give rise to lower costs and greater benefit to society. Scholes [99] pointed out that the most important lesson to be learned from the crisis is that regulators and senior management should understand much more about financial engineering. Board members of financial institutions should fully understand risk reports and the financial results, and demand clear explanations of the risks involved. This knowledge would not only be of benefit to regulators when making policy choices, but also to senior bank management in strategic decision-making for their organizations.

To provide a detailed report of the current state of development in the field of financial engineering is beyond the scope of this introduction, but we wish to provide a view of the context so that we do not lose the global picture when we start to deal with the finer details of financial engineering and stochastic analysis.

In the rest of the introduction, we will give an overview of derivative trading and its risk management, as practised by a typical fixed income derivative trading desk at a financial institution, discuss the role of financial models and motivate our choice of the target models.

1.2 Derivative Trading: Theory and Practice

> *The derivative trading business is ultimately the management of basis risk. Some do it better than others. No one did it well enough ahead of the crisis.*
>
> – Mr Chavez, Goldman Sachs

1.2.1 Basic Information Flow

To understand how a trading desk uses a financial model in practice, it is useful to introduce some notation. Here we follow closely the line of reasoning in Andersen & Piterbarg [10] and let $\Xi^{\text{mkt}}(t)$ be an N^{mkt}-dimensional vector containing observable market data at time t. For a fixed income desk, the components of $\Xi^{\text{mkt}}(t)$ may typically be interest rate swaps, caps, and swaption prices (or swaption implied volatilities) at multiple strike prices, tenor dates, and expiry dates. Secondly, let $\Xi^{\text{par}}(t)$ denote the set of N^{par} non-observable parameters. These parameters are usually obtained by calibration to historical information or to implied volatilities based on option prices. Sometimes these

parameters are fixed by traders based on their market judgement or beliefs. Examples of non-observable parameters are short rate mean reversion parameters in a stochastic differential equation (SDE), stochastic volatility mean reversion parameters, correlation parameters, and so on.

A first step in derivative pricing involves the calibration step to determine the parameter vectors Ξ^{mkt} and Ξ^{par}. This model calibration itself is based on at least two steps: (1) the construction of a discount bond curve, (2) the calibration of the model for the yield curve interest rate dynamics. Together with the discount bond curve, parameters Ξ^{mkt} and Ξ^{par} are inserted into a main model calibration technique, which results in Ξ^{mod}. We write

$$\Xi^{\text{mod}}(t) := C\left(\Xi^{\text{mkt}}(t); \Xi^{\text{par}}(t)\right),$$

where $C(\cdot)$ represents the calibration function.

Given the time t yield curve and a set of model parameters, we may proceed to use the model to price a given portfolio of derivative contracts. For this we should include contract data for a specified set of securities, and also additional parameters, $\Xi^{\text{num}}(t)$, that define the numerical mathematics methods used to determine derivative prices from the model. Examples of parameters in $\Xi^{\text{num}}(t)$ include the number of Monte Carlo paths, the size of a discretization time step of a finite difference grid or of an SDE discretization, and so on. With $V(t) = V_1(t) + \cdots + V_n(t)$ denoting the value of a portfolio of n derivatives, we write

$$V(t) = H\big(\Xi^{\text{mod}}(t); \Xi^{\text{num}}(t)\big), \tag{1.2.1}$$

for some 'transfer function', $H(\cdot)$, connecting the model parameters with financial derivative prices, based on arbitrage-free pricing principles via a chosen mathematical model. The flow of information is illustrated in Figure 1.3.

The theory of continuous-time arbitrage-free pricing [20] suggests that the function $H(\cdot)$ will assign a value which is equal to the costs of dynamic hedging through the portfolio's life time. However, in practice continuous hedging cannot be applied because of transaction costs, liquidity and other factors. The hedging is thus not complete. As a result, the hedging portfolio is not completely 'risk-free' and the cost of hedging does not always equal the portfolio's price.

In addition, real-life hedging contains an additional layer of complexity. At a time $t' > t$, the vector of model parameters, $\Xi^{\text{mod}}(t)$, will be discarded and the model is *re-calibrated*. The re-calibrated model parameters, $\Xi^{\text{mod}}(t')$, may not be consistent with those obtained and used at time t. Thus the previous hedging portfolio, determined by the parameters $\Xi^{\text{mod}}(t)$, will not hold true any more. Henceforth, a realistic hedging portfolio has to compensate for the effects of changes in vector $\Xi^{\text{mod}}(t)$ as much as possible.

For this, one typically relies on the Taylor expansion to expand the time-t model value in Eq. (1.2.1):

$$\Delta V(t) = H\big(\Xi^{\text{mod}}(t) + \Delta\big) \approx H\big(\Xi^{\text{mod}}(t)\big) + \frac{\partial H\big(\Xi^{\text{mod}}(t)\big)}{\partial \Delta} \cdot \Delta + \frac{1}{2}\Delta^T \cdot \text{Hes}(t) \cdot \Delta, \tag{1.2.2}$$

where Δ is a vector of small perturbations to the parameter vector, i.e. $\Delta = \{\delta_1, \ldots, \delta_{N^{\text{mkt}}}\}^T$, and Hes is an $(N^{\text{mkt}} \times N^{\text{mkt}})$ *Hessian* matrix, containing the

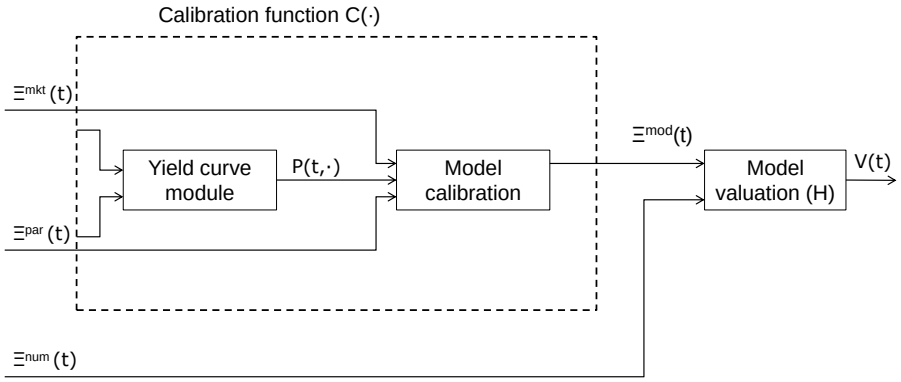

Figure 1.3: The basic information flow in derivatives pricing and model calibration for a typical fixed income derivatives trading desk, due to [10].

second-order partial derivative information of the function H. Note that we have omitted the effect of the parameters that determine the numerical scheme, $\Xi^{\text{num}}(t)$, in the above equation.

Loosely speaking, the first-order sensitivity vector contains delta values [5] and vega values [6], while the *Hessian* matrix, Hes, contains gamma values [7] and other second-order sensitivities. The trading desk will use these sensitivities to evaluate by how much the portfolio should be re-balanced to keep it to a large extent market-neutral and robust to market shocks [10]. Risk management will use the sensitivities to ensure that the exposures to individual market data components are well within given *risk limits*.

1.2.2 The Role of Models and Model Choice

In the previous section, we have briefly described the general principles of derivative pricing and calibration. Here, we discuss in some more detail the issue of pricing and hedging of exotic options, like so-called callable LIBOR exotics, path-dependent options, and so on.

Exotic interest rate derivatives are different from many of the basic, so-called *vanilla*, options because their values are not observable in the market

[5] Delta measures the rate of change of the option value with respect to changes in the underlying asset price.

[6] Vega measures the sensitivity to the volatility. It is the derivative of the option value with respect to the volatility of the underlying asset.

[7] Gamma measures the rate of change in the delta, with respect to changes in the underlying price. Gamma is the second derivative of the value function with respect to the underlying price.

(i.e. there is no regulated exchange with market quotes for these products). In the absence of market prices, we require mathematical models to determine their prices and these models have to be calibrated indirectly, i.e. to other market information which is relevant for the class of products under consideration. Due to their inherent complexity, exotic options have a non-trivial dependency on the dynamics of market rates and sophisticated term structure models are required for valuation and risk management. Roughly speaking, the purpose of the mathematical model for exotic options can be characterised as *performing a sophisticated extrapolation of information from a series of spanning vanilla market instruments* to compute a meaningful exotic option price.

The main purpose of exotic derivative pricing models is to enable financial institutions to manage their financial risks arising from large books of portfolios consisting of many non-standard exotic financial products and deals. Thus exotic derivatives pricing models are most often applied in the day-to-day task of evaluating the hedge ratios of the book of exotic derivatives, by means of the evaluation of the Greeks. Business reality requires the computation time to extract the risk metrics for a trading book to be below twenty-four hours [83], as otherwise traders and risk managers would not adopt the underlying mathematical model, no matter how attractive the model is from a theoretical point-of-view. In fact, practitioners do not expect an exotic derivatives pricing model to be a high precision *predictive* model. From a practical perspective, a successful exotic derivative pricing model is about high precision *replication of the hedging instruments*, as well as efficient computation with the model.

When the dynamics specified by the mathematical model often resemble the true underlying price process, the model price would be an accurate indication of the derivative price and the delta hedging strategy suggested by the model would very often be effective, leaving only some 'basis risk' uncovered. Basis risk is the risk that a hedge will not precisely match movements in the price of the underlying asset. In reality, most practical models are 'wrong', and basis risk thus always needs to be accounted for. In these circumstances, a trader can employ several trading strategies to minimize the risk because of model inaccuracies. One is to establish a robust static hedge portfolio. The drawback of a static approach is that it is usually too expensive. A more common approach is to construct a dynamic hedge portfolio around (1.2.1), in the Taylor expansion sense (1.2.2), aiming to reduce the movements in quantities which are assumed to be non-random.

A typical example is the so-called vega hedging strategy in a Black-Scholes model. Despite the fact that the Black-Scholes model is based on the assumption that the volatility in the model is constant, which is clearly not the case in reality, a trader can set up a hedge against moves in the volatility parameter. Empirical evidence shows that these strategies improve the robustness and effectiveness of hedging in real markets [103].

If a trader has sufficient information advantages and the pricing power to charge a spread between the long and short side positions, it does not matter whether or not the Black-Scholes model is inaccurate, because the hedging strategy is based on multiple options with different strike prices and maturity dates, which may also cancel out some hedging errors.

We have argued before that a practical exotic model is about high precision *replication of the hedging instruments* as well as efficient computation with the model. Based on these insights, it is not difficult to understand why the

so-called LIBOR market model (LMM), introduced in the mid 1990s by Brace, Gaterek and Musiela [24], rapidly gained wide popularity among practitioners. This was due to the fact that the LMM replicates prices produced by the Black model [19], which was the at-that-time-universally-accepted market standard for plain vanilla products. In the early twenty-first century, the so-called implied volatility skew and smile in the market became pronounced in the interest rate plain vanilla market. The *stochastic alpha beta rho (SABR)* model [46] then became accepted as the market standard to capture the implied volatility skew/smile observed in the plain vanilla market.

The model's popularity is mainly due to the existence of an *accurate analytic approximation* for the implied volatilities, presented by Hagan et al. [46]. This approximation formula is often used by practitioners to inter- and extrapolate the implied volatility surface.

However, most of the SABR applications are limited to approximating implied volatilities, irrespective of the underlying dynamics of the SABR model. In fact, for certain sets of parameters, the Hagan asymptotic formula is inaccurate regarding the implied volatilities, meaning that the dynamics suggested by a calibrated SABR model are not consistent with the terminal distribution implied by the market prices of the plain vanilla options used for calibration. This feature makes the SABR model less attractive for exotic option pricing, since it does not only require information about the terminal distribution, implied by European option prices, but also about the underlying (pathwise) dynamics.

This thesis focuses on various problems arising from the application of the SABR model in plain vanilla as well as in exotic option pricing, from a modelling as well as a numerical point of view. In Chapter 3 we discuss a low-bias discretization scheme for the SABR model. The purpose of the scheme is its use as a back-testing tool for the low-strike (close to zero) vanilla option prices in the SABR model. In other parts of the thesis, we also apply the SABR model to price exotic derivatives, trying to bridge the gap between plain vanilla and exotic options. For example, in Chapter 2 we introduce a two factor extension of the SABR model to price a convexity correction and in Chapter 4 we extend the SABR model and the low-bias discretization scheme to a framework with stochastic interest rates, and investigate how the SABR model can be cast as a model with time-dependent parameters to price long-dated equity-interest-rate hybrid exotics. In the final chapter with research results, we discuss a skew-smile-consistent extension of the Libor Market Model, by which we can fit the market quotes of the entire swaption cube. The local volatility function involved is an approximation of the local volatility function of the SABR model, under some technical conditions.

1.3 Organization of the Thesis

The thesis is organized as follows: In Chapter 2, we apply a two-factor SABR model and propose an analytic approximation of the convexity correction based on the technique of *small disturbance asymptotic expansion* from Yoshida [110] and Kunitomo & Takahashi [67]. Numerical experiments show the accuracy of the proposed scheme.

In Chapter 3, we develop a low-bias discretization scheme for the SABR model, which deals efficiently with (undesired) possible negative asset price val-

ues, and the discretization bias of commonly used Euler discretization schemes. The proposed algorithm is based on the analytic properties of the governing distribution. Experiments with realistic model parameters show that this scheme is robust for interest rate valuation.

Chapter 4 extends the discretization scheme proposed in the previous chapter to a SABR model with stochastic interest rates in Hull-White short rate model formulation. We propose a projection formula for the resulting SABR-HW model, mapping the SABR-HW model parameters onto the parameters of the *nearest* SABR model. The inverse of the projection formula enables a rapid calibration of the model. The purpose of the model is to price long-dated equity-interest-rate linked hybrid exotic options with exposure to both the interest rate and the equity price risks.

In Chapter 5, we consider the stochastic volatility extension of the LIBOR Market Model with a displaced diffusion local volatility function (which can be considered as an approximation of the SABR model under some technical conditions) aiming at matching the market quotes of an entire swaption cube. This model should be appropriate for exotic option pricing, as it sufficiently spans many relevant benchmark market prices so that the exotic option prices obtained from the model are consistent with the market information. Our contribution is the formulation of the calibration of the time-dependent skew parameters into a convex optimization problem, which can be solved very efficiently and the optimal solution is unique.

CHAPTER 2

Analytical Approximation to Constant Maturity Swap Convexity Correction in a Multi-Factor SABR Model

This chapter is adapted from the article
"Analytical Approximation to Constant Maturity Swap Convexity Correction in a Multi-Factor SABR Model"
accepted by *International Journal of Theoretical and Applied Finance*, in Volume 13, Issue 7, pp. 1019-1046 (2010) [31].

In this chapter we consider the convexity correction in a multi-factor SABR type stochastic volatility model, in which the volatility and the short-term forward rate are modelled as independent factors. In general, the convexity correction is not analytically tractable in a multi-factor model, but based on the assumption of linear swap rates an analytic solution is available. Linear swap rate models are popular among practitioners for their efficiency and their ability to capture the swaption volatility smile. For an efficient approximation of the solution, we adopt the small disturbance asymptotics technique and construct a stochastic Taylor series of the underlying process. Several numerical experiments compare the accuracy of the approximation with a Monte Carlo benchmark solution.

2.1 Introduction

The growing popularity of transactions of *constant maturity swap* (CMS) type in the fixed income market has increased the demand for accurate and efficient pricing methods. This research topic attracts efforts from academia and

practitioners alike.

The main lines of research for pricing methods seem to go basically in two directions. In the first, one deals with the problem by setting up a term-structure model under the T-forward measure, where the pricing originally occurs. For example, Brigo and Mercurio [25] model the bond prices associated with the CMS swap and quanto CMS swap by a G2++ model (2-factor Gaussian short rate model). The papers by Lu and Neftci [109] and Henrard [47] express the CMS swap as a collection of forward LIBOR rates under the forward measure and compute numerically the CMS price in a full-factor LIBOR market model. These approaches result in black-box computational schemes in which the risk sensitivities, e.g. the Vega, cannot be derived directly.

In the second line of research the pricing problem is formulated under the so-called *swap measure* and the given implied swaption volatilities are considered as the 'market distribution of the swap rates'. Since CMS products are mainly hedged by forward swaps and swaptions, the advantage of the measure change approach is consistency between the CMS products and their hedging instruments. Because of the measure change, from the forward to the swap measure, the Radon-Nykodym derivatives need to be approximated. Hunt and Kennedy [52] and Pelsser [85] approximate this measure change ratio in terms of a linear function of the swap rate (assuming that the yield curve is mainly driven by the swap rate) and obtain an analytic solution to the CMS price. Hagan [44] and Mercurio [79] succeed in statically replicating the CMS swap/options by European swaptions. Because of the popularity of the static replication approach, an increasing volume of swaption transactions for hedging purposes has been observed in the market, resulting in a more pronounced smile. A problem is the assumption of a one-factor yield curve, as only *parallel shifts* in the yield curve can then be taken into account. However, a CMS structure depends significantly on the slope of the yield curve, but it is not very sensitive to parallel shifts [16].

In this chapter, we adopt the Stochastic Alpha Beta Rho (SABR) model [46] to describe the dynamics of the underlying swap rate. The SABR model has the capability of generating rich skew/smile patterns and it is often used in the market [93]. We introduce an additional yield curve factor, next to the swap rate, in the measure change ratio, in order to take the dynamics at the short-end of the yield curve into account. Here, the CMS convexity correction is decomposed in two parts: A part driven by the variance of the swap rate, which is affected by the skew/smile in the implied swaption volatilities, and a second part related to the covariance between the swap and LIBOR rate, which is a result of the terminal decorrelation [1] between these two rates. One can view our pricing approach as a perturbation of the conventional CMS convexity correction away from the one-factor assumption.

We obtain an analytic approximation formula for the covariance, based on the well-known stochastic Taylor expansion [66]. Deriving the stochastic Taylor

[1] What influences the price of an exotic product, as Rebonato [92] states, is not the instantaneous correlation or volatility functions, but, the *terminal* (as opposed to instantaneous) decorrelation, $\bar{\rho}_{xy}(T)$, defined by

$$\bar{\rho}_{xy}(T) = \frac{\int_0^T \sigma_x(s)\sigma_y(s)\rho_{xy}(s)ds}{\sqrt{\int_0^T \sigma_x(s)^2 ds \int_0^T \sigma_y(s)^2 ds}}.$$

expansion by a repeated use of Itô's lemma is somewhat cumbersome when higher-orders terms are considered. We can simplify the derivation by adopting a *small disturbance asymptotics* technique (e.g. Yoshida [110] and Kunitomo [67]) to construct the Taylor series of the multi-factor SABR process.

There are a number of advantages to our approach. First of all, it models forward swap rates directly, and therefore achieves a very satisfactory agreement between the CMS contracts and their hedging instruments. Secondly, the SABR model can easily be calibrated to implied volatilities of the liquid swaptions. Thirdly, it reflects the CMS' price sensitivity to the yield curve forward correlation structure. And, finally, it provides an easy-to-implement approximation formula for the CMS convexity correction under the multi-factor model. Hence it can be used for a quick evaluation of the model risk resulting from the terminal decorrelation of the forward rates.

A less direct implication of our work is the following. The pricing of derivatives written on CMS contracts, such as CMS swaps and spread options, on the basis of underlying CMSs is impossible when the markets for the latter become illiquid, as in the recent financial distress. In such circumstances and as long as markets for plain interest rate swaps are still liquid, a conceptually sound and practically viable alternative is to price CMSs on the basis of the underlying swap prices and then price CMS derivative based on such 'synthetic CMS prices'.

The chapter is organized as follows. In Section 2, the pricing problem is formulated in an arbitrage-free way so that CMS-based derivatives are consistently priced across measures. Section 3 presents the stochastic Taylor expansion formula to the covariance of a two-factor stochastic volatility model. Examples and corresponding results for the approximate model are summarized in Section 4, where the approximate solutions are compared against short time step Monte Carlo prices. In the last section, conclusions are made.

2.2 Problem Formulation

A feature which distinguishes CMS-type contracts from plain vanilla contracts is that they pay a swap rate of one maturity, say 10 years, at each resetting time, as opposed to a regular swap, which pays the same coupon rate throughout a whole period. Hence to compute the CMS rate an adjustment has to be made to the forward swap rate implied by the swap rate curve. This adjustment is convex in the swap rate as its 'official' name, *convexity correction*, suggests. The convexity (in the swap rate) is the result of positive correlation in the yield curve [2]. The one-time-payment of the swap rate is always greater than, or equal to, the regular forward swap rate.

The above description is merely heuristic; the mathematical set-up for the CMS contract will be described in more detail in Section 2.4.2.

Because of the existence of multiple admissible pricing measures [3], it is im-

[2] Imagine the swap rate increases in value, then the discounting effect in the annuity, on which the forward swap is paid, will increase. As a result, the forward swap payoff may increase more slowly than the one-time-payment of the swap rate; On the other hand, when the swap rate decreases in value the discounting effect in the annuity may get smaller, and, consequently, the forward swap payoff may decrease slower than the one-time-payment swap rate.

[3] in this, as well as in many other interest rate derivative pricing problems.

portant to make sure that a product is consistently priced across measures without any *arbitrage* possibilities. The implication of this constraint is investigated in the following sections.

2.2.1 Measure change and arbitrage-free constraints

Girsanov's transformation gives rise to a *convexity correction*, and pricing problems in general, due to the change of measure:

$$\mathbb{E}^{Q*}[\phi\big(X(T)\big)|\mathcal{F}_t] = \frac{N(t)}{N^*(t)}\mathbb{E}^{Q}\Big[\frac{N^*(T)}{N(T)}\phi\big(X(T)\big)|\mathcal{F}_t\Big], \tag{2.2.1}$$

where $\phi\big(X(T)\big)$ is a payoff function; T is maturity time. We denote the value of $X(t)$'s natural numéraire at time t as $N(t)$ whereas $N^*(t)$ is the (unnatural) numeraire under which the payment terms are specified. Regarding the notation, we denote the filtration up to time t by a subscript t to the expectation sign, i.e. $\mathbb{E}^{Q}[\cdot|\mathcal{F}_t] = \mathbb{E}^{Q}_t[\cdot]$, whenever necessary. So, $\mathbb{E}^{Q}_0[\cdot]$ indicates an expectation w.r.t the filtration up to current time point, $t = 0$.

In order to satisfy the no-arbitrage conditions, we make the following assumption:

Assumption 2.2.1. *All rates are priced in an arbitrage-free way under their own natural pricing measure. So, the rate $X(t)$ is a martingale process under the natural measure Q. Related to the RHS of Eq. (2.2.1), this assumption excludes the possibility of arbitrage in the rate $X(t)$.*

By making use of the relation $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]+\mathbb{C}\text{ov}[X,Y]$, one finds that the convexity correction originates from the covariance between two stochastic processes:

$$\begin{aligned}
\mathbb{E}^{Q*}_t[\phi\big(X(T)\big)] &= \frac{N(t)}{N^*(t)}\mathbb{E}^{Q}_t\Big[\frac{N^*(T)}{N(T)}\phi\big(X(T)\big)\Big] \\
&= \frac{N(t)}{N^*(t)}\mathbb{E}^{Q}_t\Big[\frac{N^*(T)}{N(T)}\Big]\mathbb{E}^{Q}_t\Big[\phi\big(X(T)\big)\Big] + \frac{N(t)}{N^*(t)}\mathbb{C}\text{ov}^{Q}_t\Big[\frac{N^*(T)}{N(T)}, \phi\big(X(T)\big)\Big] \\
&= \mathbb{E}^{Q}_t\Big[\phi\big(X(T)\big)\Big] + \underbrace{\frac{N(t)}{N^*(t)}\mathbb{C}\text{ov}^{Q}_t\Big[\frac{N^*(T)}{N(T)}, \phi\big(X(T)\big)\Big]}_{\text{Convexity correction } Cc_t}.
\end{aligned} \tag{2.2.2}$$

The last equality in (2.2.2) is the result of the martingale property of the term $N^*(T)/N(T)$ which is due to the fact that it is a ratio of two tradable assets and the martingale property of $X(T)$ under its natural measure Q.

Let us focus on the numéraire ratio $N^*(T)/N(T)$. Because the yield curve is highly correlated, changes in rate $X(T)$ give rise to proportional movements of the natural numéraire $N(T)$. The numéraire $N^*(T)$ is, however, driven by another rate, which we denote by $Y(T)$. So, the numéraire ratio is a function of two rates, i.e.

$$\frac{N^*(T)}{N(T)} = f\big(X(T), Y(T)\big). \tag{2.2.3}$$

We further assume the following:

Assumption 2.2.2. *The function $f\big(X(s), Y(s)\big)$ in Eq. (2.2.3) is smooth and twice differentiable w.r.t $X(s)$ and $Y(s)$ with $s \in (t, T]$.*

By Itô's lemma [54], we have

$$\frac{N^*(T)}{N(T)} - \frac{N^*(t)}{N(t)} = \int_t^T \left\{ \frac{\partial f}{\partial X(s)} dX(s) + \frac{\partial f}{\partial Y(s)} dY(s) + O(ds) \right\}.$$

The CMS swap has a payoff which is linear in the swap rates, i.e. $\Phi\big(X(T)\big) = X(T)$ (we only consider this case in the present chapter). Then, we have the payoff in stochastic integral form

$$\Phi\big(X(T)\big) = X(T) = X(t) + \int_t^T dX(s), \tag{2.2.4}$$

and hence using (2.2.3) and (2.2.4) the convexity correction (2.2.2) simplifies:

$$\begin{aligned}
Cc(t) &= \frac{N(t)}{N^*(t)} \mathbb{C}\mathrm{ov}_t^Q \Big[\frac{N^*(T)}{N(T)}, \phi\big(X(T)\big) \Big] = \frac{N(t)}{N^*(t)} \mathbb{E}^Q \Big[\Big(\frac{N^*(T)}{N(T)} - \frac{N^*(t)}{N(t)} \Big) \big(X(T) - X(t) \big) \Big] \\
&= \frac{N(t)}{N^*(t)} \mathbb{E}_t^Q \Big[\Big\{ \int_t^T \frac{\partial f}{\partial X(s)} dX(s) \cdot \int_t^T dX(s) + \int_t^T \frac{\partial f}{\partial Y(s)} dY(s) \cdot \int_t^T dX(s) \\
&+ \int_t^T O(ds) ds \cdot \int_t^T dX(s) \Big\} \Big] \\
&\approx \frac{N(t)}{N^*(t)} \Big\{ \frac{\partial f}{\partial X(t)} \mathbb{E}_t^Q \big[\int_t^T dX(s) \cdot \int_t^T dX(s) \big] + \frac{\partial f}{\partial Y(t)} \mathbb{E}_t^Q \big[\int_t^T dY(s) \cdot \int_t^T dX(s) \big] \\
&+ \int_t^T O(ds) ds \underbrace{\mathbb{E}_t^Q \big[\int_t^T dX(s) \big]}_{=0} \Big\} \\
&= \frac{N(t)}{N^*(t)} \Big(\frac{\partial f}{\partial X(t)} \mathbb{V}\mathrm{ar}_t^Q [X(T)] + \frac{\partial f}{\partial Y(t)} \mathbb{C}\mathrm{ov}_t^Q [X(T), Y(T)] \Big).
\end{aligned} \tag{2.2.5}$$

Usually the partial derivatives of the numéraire ratio are smooth and slowly varying [50, 55]. A widely accepted approach is therefore to *freeze* them to their initial values, as demonstrated in Hull and White [50] and Jäckel and Rebonato [55]. This is also the approximation made in Eq. (2.2.5).

The form of Eq. (2.2.5) suggests that the convexity correction is driven by the terms $\mathbb{V}\mathrm{ar}_t^Q[X(T)]$ and $\mathbb{C}\mathrm{ov}_t^Q[X(T), Y(T)]$. The covariance-based formulation naturally combines with multi-factor arbitrage-free interest rate modelling. This is a useful property of (2.2.5) because traders tend to have a better formalized view about the correlation between two arbitrary interest rates than about the joint density of these two rates [92].

2.2.2 Model set-up and technical issues

The drawback of the covariance-based formulation (2.2.5) is of a technical nature, since the variance and covariance quantities are not easily computed, especially not when skew/smile features are taken into account. The main result of this chapter is, therefore, an expansion formula for the covariance quantity

in a multi-factor stochastic volatility model based on the Stochastic Alpha Beta Rho (SABR) model [46] under a reference measure, i.e. $\forall s \in (t, T]$

$$\begin{array}{rcll} dX(s) & = & & \sigma(s)\big(X(s)\big)^{\beta_x} dW^{(1)}(s) \\ dY(s) & = & \mu\big(s, Y(s)\big)ds + & \sigma(s)\big(Y(s)\big)^{\beta_y} dW^{(2)}(s) \\ d\sigma(s) & = & & \eta\sigma(s)dZ(s), \end{array} \tag{2.2.6}$$

with η the volatility-of-the-volatility. The three Brownian motions are correlated by the following correlation matrix

$$\begin{bmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{bmatrix}$$

with $dW^{(1)}(s)dZ(s) = \rho_{xz}ds, dW^{(2)}(s)dZ(s) = \rho_{yz}ds$ and $dW^{(1)}(s)dW^{(2)}(s) = \rho_{xy}ds$. Note that the model is defined under rate $X(s)$'s natural measure Q. Superscript Q is omitted for ease of presentation. In the rest of this chapter, the processes without specific superscripts are defined under the measure Q. Term $\mu\big(s, Y(s)\big)$ is the arbitrage-free drift of the rate $Y(s)$ whose natural pricing measure is Q^*.

Practitioners often choose fewer volatility factors than the number of state variables, like a single volatility factor in, e.g., Piterbarg [86], Andersen and Andreasen [5]. In this chapter, we also use a single volatility factor in (2.2.6). It serves as a first multi-factor academic model for the techniques proposed.

The convexity correction in a two-factor model has been derived for bivariate log-normal models in [105]. In a bivariate log-normal distribution, the covariance can easily be computed by integrating over the *terminal* bi-variate distribution of the rate with respect to the *modified* [4] payoff function of the two rates involved, $X(t)$ and $Y(t)$. However, when stochastic volatility is considered, the integration over the terminal bi-variate distribution does not result in the correct values, because the process $[X(t), Y(t)]$ is not Markovian and the volatility realized along the path has a non-trivial impact on the convexity correction.

Even if we obtain the joint densities of the triplets $[X(t), Y(t), \sigma(t)]$ correctly, there is no guarantee that we can directly integrate the terminal joint density over the payoff function when including the *arbitrage-free constraints*. In the next section, we therefore use a different method to approximate the covariance, based on the stochastic Taylor expansion.

2.3 Stochastic Taylor Expansion to the Two-Factor SABR Model

Usually, multi-factor SABR prices are computed by a short time step Monte Carlo procedure, which is time consuming. In this section, we derive an approximation to the covariance quantity. The method used is based on the well-known Itô-Taylor expansion, described in Kloeden and Platen [66], and references therein. Instead of deriving the formula by directly applying Itô's

[4] It is modified, because the relative numéraire is also included in the expected value of the payoff.

lemma, we rely on the *small disturbance asymptotics* technique, described in Kunitomo and Takahashi [67], to construct the Taylor series of the processes $X(t)$ and $Y(t)$. This technique has been applied to interest rate derivative pricing problems by Kawai [64, 65] and Hagan [45]. Its theoretical validity was discussed in detail in [68].

2.3.1 Stochastic Taylor expansion to asset dynamics

We express the solution in terms of *successive* terms with different orders of growth in time t. We first reformulate the system (2.2.6) by specifying a time rescaling $t = \epsilon^2\tau$, so that the processes $\epsilon W^{(\cdot)}(\tau)$, $\epsilon Z(\tau)$ have the same variances as $W^{(\cdot)}(t)$ and $Z(t)$, respectively.

$$\begin{array}{rcl} dX^{(\epsilon)}(\tau) & = & \epsilon\sigma^{(\epsilon)}(\tau)\big(X^{(\epsilon)}(\tau)\big)^{\beta_x} dW^{(1)}(\tau) \\ dY^{(\epsilon)}(\tau) & = \epsilon^2\mu\big(\tau, Y(\tau)^{(\epsilon)}\big)d\tau + & \epsilon\sigma^{(\epsilon)}(\tau)\big(Y^{(\epsilon)}(\tau)\big)^{\beta_y} dW^{(2)}(\tau) \\ d\sigma^{(\epsilon)}(\tau) & = & \epsilon\eta\sigma^{(\epsilon)}(\tau)dZ(\tau) \end{array} \tag{2.3.1}$$

The covariance of the time rescaling processes, $X^{(\epsilon)}(\tau)$ and $Y^{(\epsilon)}(\tau)$, does not change, i.e.

$$dX^{(\epsilon)}(\tau)dY^{(\epsilon)}(\tau) \propto \epsilon^2\rho_{xy}d\tau = \rho_{xy}dt \propto dX(t)dY(t).$$

Since we do not know the distribution of system (2.3.1) explicitly, we consider the stochastic expansion around a deterministic process $[X^{(0)}(\tau), Y^{(0)}(\tau), \sigma^{(0)}(\tau)]$ when the time rescaling parameter, ϵ, goes to zero, as required by the adopted small disturbance asymptotic technique. We substitute the time rescaling process in the definition of the covariance and truncate the solution up to the desired order of accuracy to obtain an approximation formula for the covariance.

Proposition 1. *The stochastic Taylor expansion of the volatility process, $\sigma^{(\epsilon)}(\tau)$, up to fourth-order, reads*

$$\sigma^{(\epsilon)}(\tau) = \sigma^{(0)}(\tau) + \epsilon\sigma^{(1)}(\tau) + \epsilon^2\sigma^{(2)}(\tau) + \epsilon^3\sigma^{(3)}(\tau) + O(\epsilon^4),$$

where

$$\begin{aligned} &\sigma^{(0)}(\tau) = \sigma(0), \\ &\sigma^{(1)}(\tau) := \frac{\partial\sigma^{(\epsilon)}(\tau)}{\partial\epsilon}\Big|_{\epsilon=0} = \int_0^\tau \eta\sigma(0)dZ(s), \\ &\sigma^{(2)}(\tau) := \frac{1}{2}\frac{\partial^2\sigma^{(\epsilon)}(\tau)}{\partial\epsilon^2}\Big|_{\epsilon=0} = \int_0^\tau \eta\int_0^{s_1}\eta\sigma^{(0)}(s_2)dZ(s_2)dZ(s_1), \\ &\sigma^{(3)}(\tau) := \frac{1}{6}\frac{\partial^3\sigma^{(\epsilon)}(\tau)}{\partial\epsilon^3}\Big|_{\epsilon=0} = \int_0^\tau \eta\int_0^{s_1}\eta\int_0^{s_2}\eta\sigma^{(0)}(s_3)dZ(s_3)dZ(s_2)dZ(s_1). \end{aligned}$$

Proof of Proposition 1. This is a well-known result. The Taylor expansion of $\sigma^{(\epsilon)}(\tau)$ around $\epsilon = 0$ gives

$$\sigma^{(\epsilon)}(\tau) = \sigma^{(0)}(\tau) + \epsilon\frac{\partial\sigma^{(\epsilon)}(\tau)}{\partial\epsilon}\Big|_{\epsilon=0} + \frac{1}{2}\epsilon^2\frac{\partial^2\sigma^{(\epsilon)}(\tau)}{\partial\epsilon^2}\Big|_{\epsilon=0} + \frac{1}{6}\epsilon^3\frac{\partial^3\sigma^{(\epsilon)}(\tau)}{\partial\epsilon^3}\Big|_{\epsilon=0} + O(\epsilon^4).$$

From the volatility process in integral form,

$$\sigma^{(\epsilon)}(\tau) = \sigma(0) + \epsilon \int_0^\tau \eta \sigma^{(\epsilon)}(s) dZ(s),$$

one finds that $\sigma^{(0)}(\tau) = \sigma(0)$.

Then

$$\begin{aligned}
\frac{\partial \sigma^{(\epsilon)}(\tau)}{\partial \epsilon}\Big|_{\epsilon=0} &= \Big[\int_0^\tau \eta \sigma^{(\epsilon)}(s) dZ(s) + \epsilon \int_0^\tau \eta \frac{\partial \sigma^{(\epsilon)}(s)}{\partial \epsilon} dZ(s)\Big]\Big|_{\epsilon=0} = \int_0^\tau \eta \sigma(0) dZ(s); \\
\frac{\partial^2 \sigma^{(\epsilon)}(\tau)}{\partial \epsilon^2}\Big|_{\epsilon=0} &= \Big[2\int_0^\tau \eta \frac{\partial \sigma^{(\epsilon)}(s)}{\partial \epsilon} dZ(s) + \epsilon \int_0^\tau \eta \frac{\partial^2 \sigma^{(\epsilon)}(s)}{\partial^2 \epsilon}\Big]\Big|_{\epsilon=0} \\
&= 2\int_0^\tau \eta \frac{\partial \sigma^{(0)}(s)}{\partial \epsilon} dZ(s) = 2\int_0^\tau \eta \int_0^{s_1} \eta \sigma^{(0)}(s_2) dZ(s_2) dZ(s_1).
\end{aligned}$$

Similarly, we find

$$\frac{\partial^3 \sigma^{(\epsilon)}(\tau)}{\partial \epsilon^3}\Big|_{\epsilon=0} = 6\int_0^\tau \eta \int_0^{s_1} \eta \int_0^{s_2} \eta \sigma^{(0)}(s_3) dZ(s_3) dZ(s_2) dZ(s_1).$$

□

Proposition 2. *The stochastic Taylor expansion of $X^{(\epsilon)}(\tau)$, up to fourth-order, can be expressed as follows:*

$$X^{(\epsilon)}(\tau) = X^{(0)}(\tau) + \epsilon X^{(1)}(\tau) + \epsilon^2 X^{(2)}(\tau) + \epsilon^3 X^{(3)}(\tau) + O(\epsilon^4), \qquad (2.3.2)$$

where

$$\begin{aligned}
X^{(0)}(\tau) =& X(0), \\
X^{(1)}(\tau) :=& \frac{\partial X^{(\epsilon)}(\tau)}{\partial \epsilon}\Big|_{\epsilon=0} \\
=& \int_0^\tau \sigma(0) \big(X(0)\big)^{\beta_x} dW^{(1)}(s), \\
X^{(2)}(\tau) :=& \frac{1}{2} \frac{\partial^2 X^{(\epsilon)}(\tau)}{\partial \epsilon^2}\Big|_{\epsilon=0} \\
=& \int_0^\tau \sigma(0) \int_0^{s_1} \sigma(0) \beta_x \big(X(0)\big)^{\beta_x - 1} dW^{(1)}(s_2) dW^{(1)}(s_1) \\
&+ \int_0^\tau \big(X(0)\big)^{\beta_x} \int_0^{s_1} \eta \sigma(0) dZ(s_2) dW^{(1)}(s_1), \\
X^{(3)}(\tau) :=& \frac{1}{6} \frac{\partial^3 X^{(\epsilon)}(\tau)}{\partial \epsilon^3}\Big|_{\epsilon=0} \\
=& \int_0^\tau \sigma(0) \int_0^{s_1} \sigma(0) \int_0^{s_2} \sigma(0) \beta_x (\beta_x - 1) \big(X(0)\big)^{\beta_x - 2} dW^{(1)}(s_3) dW^{(1)}(s_2) dW^{(1)}(s_1) \\
&+ \int_0^\tau \sigma(0) \int_0^{s_1} \beta_x \big(X(0)\big)^{\beta_x - 1} \int_0^{s_2} \eta \sigma(0) dZ(s_3) dW^{(1)}(s_2) dW^{(1)}(s_1) \\
&+ \int_0^\tau \big(X(0)\big)^{\beta_x} \int_0^{s_1} \eta \int_0^{s_2} \eta \sigma(0) dZ(s_3) dZ(s_2) dW^{(1)}(s_1) \\
&+ \int_0^\tau \int_0^{s_1} \eta \big(\sigma(0)\big)^2 \beta_x \big(X(0)\big)^{\beta_x - 1} Z(s_1) W^{(1)}(s_1) dW^{(1)}(s_1).
\end{aligned}$$

Proof of Proposition 2. We make a Taylor expansion of the process $X^{(\epsilon)}(\tau)$ around $\epsilon = 0$:

$$X^{(\epsilon)}(\tau) = X^{(0)}(\tau) + \epsilon \frac{\partial X^{(\epsilon)}(\tau)}{\partial \epsilon}\Big|_{\epsilon=0} + \frac{1}{2}\epsilon^2 \frac{\partial^2 X^{(\epsilon)}(\tau)}{\partial \epsilon^2}\Big|_{\epsilon=0} + \frac{1}{6}\epsilon^3 \frac{\partial^3 X^{(\epsilon)}(\tau)}{\partial \epsilon^3}\Big|_{\epsilon=0} + O(\epsilon^4).$$

It is again easy to see that $X^{(0)}(\tau) = X(0)$. Following the arguments in Kunitomo [67], we have

$$\begin{aligned}
&\frac{\partial X^{(\epsilon)}(\tau)}{\partial \epsilon}\Big|_{\epsilon=0} \\
=&\Big[\int_0^\tau \sigma^{(\epsilon)}(s)\big(X^{(\epsilon)}(s)\big)^{\beta_x} dW^{(1)}(s) + \epsilon \int_0^\tau \sigma^{(\epsilon)}(s) \frac{\partial \big(X^{(\epsilon)}(s)\big)^{\beta_x}}{\partial \epsilon} dW^{(1)}(s) \\
&+ \epsilon \int_0^\tau \big(X^{(\epsilon)}(s)\big)^{\beta_x} \frac{\partial \sigma^{(\epsilon)}(s)}{\partial \epsilon} dW^{(1)}(s)\Big]\Big|_{\epsilon=0} \\
=&\int_0^\tau \sigma^{(0)}(s)\big(X^{(0)}(s)\big)^{\beta_x} dW^{(1)}(s) = \int_0^\tau \sigma(0)\big(X(0)\big)^{\beta_x} dW^{(1)}(s); \\
&\frac{\partial^2 X^{(\epsilon)}(\tau)}{\partial \epsilon^2}\Big|_{\epsilon=0} \\
=&\Big[2\int_0^\tau \sigma^{(\epsilon)}(s) \frac{\partial \big(X^{(\epsilon)}(s)\big)^{\beta_x}}{\partial \epsilon} dW^{(1)}(s) + 2\int_0^\tau \big(X(s)^{(\epsilon)}\big)^{\beta_x} \frac{\partial \sigma^{(\epsilon)}(s)}{\partial \epsilon} dW^{(1)}(s) \\
&+ \epsilon \int_0^\tau \sigma^{(\epsilon)}(s) \frac{\partial^2 \big(X^{(\epsilon)}(s)\big)^{\beta_x}}{\partial \epsilon^2} dW^{(1)}(s) \\
&+ \epsilon \int_0^\tau \big(X^{(\epsilon)}(s)\big)^{\beta_x} \frac{\partial^2 \sigma^{(\epsilon)}(s)}{\partial \epsilon^2} dW^{(1)}(s) + 2\epsilon \int_0^\tau \frac{\partial \big(X(s)^{(\epsilon)}\big)^{\beta_x}}{\partial \epsilon} \frac{\partial \sigma^{(\epsilon)}(s)}{\partial \epsilon} dW^{(1)}(s)\Big]\Big|_{\epsilon=0} \\
=&2\int_0^\tau \sigma(0) \int_0^{s_1} \sigma(0)\beta_x \big(X(0)\big)^{\beta_x - 1} dW^{(1)}(s_2) dW^{(1)}(s_1) \\
&+ 2\int_0^\tau \big(X(0)\big)^{\beta_x} \int_0^{s_1} \eta\sigma(0) dZ(s_2) dW^{(1)}(s_1).
\end{aligned}$$

Recursive application of this scheme gives

$$
\begin{aligned}
&\frac{\partial^3 X^{(\epsilon)}(\tau)}{\partial \epsilon^3}\Big|_{\epsilon=0} \\
=&\Big[3\int_0^\tau \sigma^{(\epsilon)}(s)\frac{\partial^2\big(X^{(\epsilon)}(s)\big)^{\beta_x}}{\partial\epsilon^2}dW^{(1)}(s) + 3\int_0^\tau \big(X^{(\epsilon)}(s)\big)^{\beta_x}\frac{\partial^2\sigma^{(\epsilon)}(s)}{\partial\epsilon^2}dW^{(1)}(s) \\
&+ 6\int_0^\tau \frac{\partial\big(X^{(\epsilon)}(s)\big)^{\beta_x}}{\partial\epsilon}\frac{\partial\sigma^{(\epsilon)}(s)}{\partial\epsilon}dW^{(1)}(s)\Big]\Big|_{\epsilon=0} \\
=&3\int_0^\tau \sigma^{(0)}(s)\frac{\partial^2\big(X^{(0)}(s)\big)^{\beta_x}}{\partial\epsilon^2}dW^{(1)}(s) + 3\int_0^\tau \big(X^{(0)}(s)\big)^{\beta_x}\frac{\partial^2\sigma^{(0)}(s)}{\partial\epsilon^2}dW^{(1)}(s) \\
&+ 6\int_0^\tau \frac{\partial\big(X^{(0)}(s)\big)^{\beta_x}}{\partial\epsilon}\frac{\partial\sigma^{(0)}(s)}{\partial\epsilon}dW^{(1)}(s) \\
=&6\int_0^\tau \sigma(0)\int_0^{s_1}\sigma(0)\int_0^{s_2}\sigma(0)\beta_x(\beta_x-1)\big(X(0)\big)^{\beta_x-2}dW^{(1)}(s_3)dW^{(1)}(s_2)dW^{(1)}(s_1) \\
&+ 6\int_0^\tau \sigma(0)\int_0^{s_1}\beta_x\big(X(0)\big)^{\beta_x-1}\int_0^{s_2}\eta\sigma(0)dZ(s_3)dW^{(1)}(s_2)dW^{(1)}(s_1) \\
&+ 6\int_0^\tau \big(X(0)\big)^{\beta_x}\int_0^{s_1}\eta\int_0^{s_2}\eta\sigma(0)dZ(s_3)dZ(s_2)dW^{(1)}(s_1) \\
&+ 6\int_0^\tau\int_0^{s_1}\eta\big(\sigma(0)\big)^2\beta_x\big(X(0)\big)^{\beta_x-1}Z(s_1)W^{(1)}(s_1)dW^{(1)}(s_1).
\end{aligned}
$$

□

Proposition 3. *The stochastic Taylor expansion of* $Y^{(\epsilon)}(\tau)$*, up to fourth-order, gives:*

$$Y^{(\epsilon)}(\tau) = Y^{(0)}(\tau) + \epsilon Y^{(1)}(\tau) + \epsilon^2 Y^{(2)}(\tau) + \epsilon^3 Y^{(3)}(\tau) + O(\epsilon^4),$$

where

$$
\begin{aligned}
Y^{(0)}(\tau) =& Y(0),\\
Y^{(1)}(\tau) :=& \frac{\partial Y^{(\epsilon)}(\tau)}{\partial \epsilon}\Big|_{\epsilon=0} = \int_0^\tau \sigma(0)\big(Y(0)\big)^{\beta_y} dW^{(2)}(s),\\
Y^{(2)}(\tau) :=& \frac{1}{2}\frac{\partial^2 Y^{(\epsilon)}(\tau)}{\partial \epsilon^2}\Big|_{\epsilon=0}\\
=& \int_0^\tau \mu\big(0, Y(0)\big)d\tau + \int_0^\tau \sigma(0)\int_0^{s_1} \sigma(0)\beta_y\big(Y(0)\big)^{\beta_y-1} dW^{(2)}(s_2)dW^{(2)}(s_1)\\
&+ \int_0^\tau \big(Y(0)\big)^{\beta_y}\int_0^{s_1}\eta\sigma(0)dZ(s_2)dW^{(2)}(s_1),\\
Y^{(3)}(\tau) :=& \frac{1}{6}\frac{\partial^3 Y^{(\epsilon)}(\tau)}{\partial \epsilon^3}\Big|_{\epsilon=0}\\
=& \int_0^\tau \sigma(0)\int_0^{s_1}\sigma(0)\int_0^{s_2}\sigma(0)\beta_y(\beta_y-1)\big(Y(0)\big)^{\beta_y-2} dW^{(2)}(s_3)dW^{(2)}(s_2)dW^{(2)}(s_1)\\
&+ \int_0^\tau \sigma(0)\int_0^{s_1}\beta_y\big(Y(0)\big)^{\beta_y-1}\int_0^{s_2}\eta\sigma(0)dZ(s_3)dW^{(2)}(s_2)dW^{(2)}(s_1)\\
&+ \int_0^\tau \big(Y(0)\big)^{\beta_y}\int_0^{s_1}\eta\int_0^{s_2}\eta\sigma(0)dZ(s_3)dZ(s_2)dW^{(2)}(s_1)\\
&+ \int_0^\tau\int_0^{s_1}\eta\big(\sigma(0)\big)^2\beta_y\big(Y(0)\big)^{\beta_y-1}Z(s_1)W^{(2)}(s_1)dW^{(2)}(s_1).
\end{aligned}
$$

The proof is similar to that of Proposition 2.

2.3.2 Expansion solution to the covariance

We recall the definition of the covariance and substitute the expansions $X^{(\epsilon)}(\tau)$ and $Y^{(\epsilon)}(\tau)$ of $X(\tau)$ and $Y(\tau)$, respectively. This way we obtain a stochastic Taylor expansion formula for the covariance, which is given by the following lemma.

Lemma 2.3.1. *For a multi-factor parametric stochastic volatility model of the form (2.3.1), the stochastic Taylor expansion of the covariance,* $\mathbb{C}\mathrm{ov}_0[X^{(\epsilon)}(t), Y^{(\epsilon)}(t)]$, *is given by*

$$
\mathbb{C}\mathrm{ov}_0[X^{(\epsilon)}(t), Y^{(\epsilon)}(t)] = \nu^2 t + (\Lambda + \Gamma + \Sigma)\frac{t^2}{2} + O(t^6), \tag{2.3.3}
$$

where

$$
\begin{aligned}
\nu^2 =& \big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy},\\
\Lambda =& \big(\sigma(0)\big)^4\beta_x\big(X(0)\big)^{\beta_x-1}\beta_y\big(Y(0)\big)^{\beta_y-1}(\rho_{xy})^2,\\
\Gamma =& \eta\big(\sigma(0)\big)^3\beta_x\big(X(0)\big)^{\beta_x-1}\big(Y(0)\big)^{\beta_y}\rho_{xy}\rho_{xz} + \eta\big(\sigma(0)\big)^3\big(X(0)\big)^{\beta_x}\beta_y\big(Y(0)\big)^{\beta_y-1}\rho_{xy}\rho_{yz}\\
&+ \eta\big(\sigma(0)\big)^3\big(X(0)\big)^{\beta_x}\beta_y\big(Y(0)\big)^{\beta_y-1}\rho_{yz}\rho_{xy} + \eta\big(\sigma(0)\big)^3\beta_x\big(X(0)\big)^{\beta_x-1}\big(Y(0)\big)^{\beta_y}\rho_{xz}\rho_{xy},\\
\Sigma =& \eta^2\big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}.
\end{aligned}
$$

Proof of Lemma 2.3.1. To facilitate the proof, we firstly recall the product formula for two Itô integrals:

$$\mathbb{E}\Big[\int_0^t f dW^{(1)}(s)\cdot\int_0^t g dW^{(2)}(s)\Big]=\int_0^t (f\cdot g)\rho ds, \tag{2.3.4}$$

with $dW^{(1)}(s)dW^{(2)}(s)=\rho dt$.

Since the process $X(\tau)$ is a martingale, we have $\mathbb{E}_0[X(\tau)]=X(0)$. For process $Y(\tau)$ we have

$$\mathbb{E}_0[Y(\tau)]=Y(0)+\int_0^\tau \mu\big(s_1,Y(s_1)\big)ds_1.$$

We substitute these two expectations and the expansions for $X(\tau)$ and $Y(\tau)$ in the definition of the covariance:

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}_0^{(\epsilon)}[X^{(\epsilon)}(\tau),Y^{(\epsilon)}(\tau)]\\
=&\mathbb{E}\big[\big(X^{(\epsilon)}(\tau)-\mathbb{E}_0[X^{(\epsilon)}(\tau)]\big)\big(Y^{(\epsilon)}(\tau)-\mathbb{E}_0[Y^{(\epsilon)}(\tau)]\big)\big]\\
\approx&\mathbb{E}\Big[\big(X^{(0)}(\tau)+\epsilon X^{(1)}(\tau)+\epsilon^2X^{(2)}(\tau)+\epsilon^3X^{(3)}(\tau)-X(0)\big)\cdot\\
&\Big(Y^{(0)}(\tau)+\epsilon Y^{(1)}(\tau)+\epsilon^2Y^{(2)}(\tau)+\epsilon Y^{(3)}(\tau)-Y(0)-\epsilon^2\int_0^t\mu\big(s_1,Y^{(\epsilon)}(s_1)\big)ds_1\Big)\Big]\\
=&\mathbb{E}\big[\big(\epsilon^2X^{(1)}(\tau)Y^{(1)}(\tau)+\epsilon^3X^{(1)}(\tau)Y^{(2*)}(\tau)+\epsilon^4X^{(1)}(\tau)Y^{(3)}(\tau)+\epsilon^3X^{(2)}(\tau)Y^{(1)}(\tau)+\\
&\epsilon^4X^{(2)}(\tau)Y^{(2*)}(\tau)+\epsilon^4X^{(3)}(\tau)Y^{(1)}(\tau)+O(\epsilon^5)\big)\big],
\end{aligned}\tag{2.3.5}$$

where we have eliminated the drift term from $Y^{(2)}(\tau)$ and defined

$$\begin{aligned}
Y^{(2)}(\tau):=&\int_0^\tau\sigma(0)\int_0^{s_1}\sigma(0)\big(Y(0)\big)^{\beta_y}dW^{(2)}(s_2)dW^{(2)}(s_1)\\
&+\int_0^\tau\big(Y(0)\big)^{\beta_y}\int_0^{s_1}\eta\sigma(0)dZ(s_2)dW^{(2)}(s_1).
\end{aligned}$$

By Itô product formula (2.3.4), we find that

$$\begin{aligned}
&X^{(1)}(\tau)Y^{(1)}(\tau)\\
=&\int_0^\tau\big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}ds=\big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}\tau,
\end{aligned}\tag{2.3.6}$$

$$\begin{aligned}
&X^{(1)}(\tau)Y^{(2*)}(\tau)\\
=&\int_0^\tau\big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\int_0^{s_1}\sigma(0)\big(Y(0)\big)^{\beta_y}dW^{(2)}(s_2)\rho_{xy}ds_1\\
&+\int_0^\tau\sigma(0)\big(X(0)\big)^{2\beta_x}\big(Y(0)\big)^{\beta_y}\int_0^{s_1}\eta\sigma(0)dZ(s_2)\rho_{xy}ds_1\\
=&\big(\sigma(0)\big)^3\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}\int_0^\tau\int_0^{s_1}dW^{(2)}(s_2)ds_1\\
&+\eta\big(\sigma(0)\big)^2\big(X(0)\big)^{2\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}\int_0^\tau\int_0^{s_1}dZ(s_2)ds_1.
\end{aligned}\tag{2.3.7}$$

It is not difficult to see that the two terms in (2.3.7) are Itô integrals with mean zero. After taking the expectation, the term $X^{(1)}(\tau)Y^{(2*)}(\tau)$ disappears, i.e., $\mathbb{E}_0[X^{(1)}(\tau)Y^{(2*)}(\tau)] = 0$. The same holds for all terms of odd order, e.g. for $\epsilon^3, \epsilon^5, \ldots$. Hence we only have to consider the even terms in the expansion of the covariance, for example:

$$
\begin{aligned}
& X^{(1)}(\tau)Y^{(3)}(\tau) \\
= & \int_0^\tau \big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\int_0^{s_1}\sigma(0)\int_0^{s_2}\sigma(0)\big(Y(0)\big)^{\beta_y}dW^{(2)}(s_3)dW^{(2)}(s_2)\rho_{xy}ds_1 \\
+ & \int_0^\tau \big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\int_0^{s_1}\big(Y(0)\big)^{\beta_y}\int_0^{s_2}\eta\sigma(0)dZ(s_3)dW^{(2)}(s_2)\rho_{xy}ds_1 \\
+ & \int_0^\tau \sigma(0)\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\int_0^{s_1}\eta\int_0^{s_2}\eta\sigma(0)dZ(s_3)dZ(s_2)\rho_{xy}ds_1 \\
+ & \int_0^\tau \sigma(0)\big(X(0)\big)^{\beta_x}\int_0^{s_1}\eta\big(\sigma(0)\big)^2\big(Y(0)\big)^{\beta_y}\rho_{yz}ds_2\rho_{xy}ds_1 \\
= & \big(\sigma(0)\big)^4\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}\int_0^\tau\int_0^{s_1}\int_0^{s_2}dW^{(2)}(s_3)dW^{(2)}(s_2)ds_1 \\
+ & \big(\sigma(0)\big)^3\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\eta\rho_{xy}\int_0^\tau\int_0^{s_1}\int_0^{s_2}dZ(s_3)dW^{(2)}(s_2)ds_1 \\
+ & \big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\eta^2\rho_{xy}\int_0^\tau\int_0^{s_1}\int_0^{s_2}dZ(s_3)dZ(s_2)ds_1 \\
+ & \eta\big(\sigma(0)\big)^3\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}\rho_{yz}\frac{\tau^2}{2}. \qquad (2.3.8)
\end{aligned}
$$

The computation of the expectation of $X^{(1)}(\tau)Y^{(3)}(\tau)$ requires that we deal with three double Itô integrals, i.e. $\mathbb{E}\Big[\int_0^{s_1}\int_0^{s_2}dW^{(2)}(s_3)dW^{(2)}(s_2)\Big]$, $\mathbb{E}\Big[\int_0^{s_1}\int_0^{s_2}dZ_{s_3}dW^{(2)}(s_2)\Big]$ and $\mathbb{E}\Big[\int_0^{s_1}\int_0^{s_2}dZ(s_3)dZ(s_2)\Big]$. For the first integral in (2.3.8), we can compute its expectation by applying Itô's lemma:

$$
\begin{aligned}
& W^{(2)}(s_2)dW^{(2)}(s_2) = \frac{1}{2}d\big(W^{(2)}(s_2)\big)^2 + \frac{1}{2}ds_2, \text{ or,} \\
& \int_0^{s_1} W^{(2)}(s_2)dW^{(2)}(s_2) = \frac{1}{2}\int_0^{s_1} d\big(W^{(2)}(s_2)\big)^2 + \int_0^{s_1}\frac{1}{2}ds_2 = \frac{1}{2}\big(W^{(2)}(s_1)\big)^2 + \frac{1}{2}s_1 \\
& \qquad \Rightarrow \mathbb{E}\Big[\int_0^{s_1}\underbrace{\int_0^{s_2}dW^{(2)}(s_3)}_{=W^{(2)}(s_2)-W^{(2)}(0)}\,dW^{(2)}(s_2)\Big] = \mathbb{E}\big[\frac{1}{2}\big(W^{(2)}(s_1)\big)^2 + \frac{1}{2}s_1\big] = 0.
\end{aligned}
$$

Similarly, we find that $\mathbb{E}\Big[\int_0^{s_1}\int_0^{s_2}dZ(s_3)dZ(s_2)\Big] = 0$ and $\mathbb{E}\big[\int_0^{s_1}\int_0^{s_2}dZ(s_3)dW^{(2)}(s_2)\big] = 0$.

The computations of $\mathbb{E}\big[X^{(2)}(\tau)Y^{(2)}(\tau)\big]$ and $\mathbb{E}\big[X^{(3)}(\tau)Y^{(1)}(\tau)\big]$ are performed

in the same manner. Finally, we substitute the result in Eq. (2.3.5) and obtain

$$\begin{aligned}\mathbb{C}\mathrm{ov}_0[X(\tau),Y(\tau)] =&\epsilon^2\big(\sigma(0)\big)^2\tau+\epsilon^4\big(\big(\sigma(0)\big)^4\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}(\rho_{xy})^2\frac{\tau^2}{2}\\&+2\epsilon^4\eta\big(\sigma(0)\big)^3\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}(\rho_{xz}+\rho_{yz})\frac{\tau^2}{2}\\&+\epsilon^4\eta^2\big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}\frac{\tau^2}{2}+O(\epsilon^6)\\=&\big(\sigma(0)\big)^2t+\big(\big(\sigma(0)\big)^4\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}(\rho_{xy})^2\frac{t^2}{2}\\&+2\eta\big(\sigma(0)\big)^3\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}(\rho_{xz}+\rho_{yz})\frac{t^2}{2}\\&+\eta^2\big(\sigma(0)\big)^2\big(X(0)\big)^{\beta_x}\big(Y(0)\big)^{\beta_y}\rho_{xy}\frac{t^2}{2}+O(t^6),\end{aligned}$$

where the last equality is a consequence of the time rescaling we defined earlier, i.e. $\epsilon^2\tau = t$. As stated earlier, the rescaled system $[X^{(\epsilon)}(t), X^{(\epsilon)}(t), \sigma^{(\epsilon)}(t)]$ preserves the variance and covariance. Therefore, we obtain the desired approximation for the covariance in the original time scale. □

The terms in Eq. (2.3.3) can be interpreted as follows:

- The first term, ν^2, is the leading term of the covariance which grows linearly with time. It is also the solution we would obtain by assuming a constant volatility;
- Correction term Λ is due to the first-order sensitivity of the covariance w.r.t to the forward rate dynamics;
- Term Σ quantifies the impact of stochastic volatility. It is positive, hence it adds a positive contribution to the covariance;
- Finally, term Γ is related to the interaction between the forward rate and the volatility dynamics [45] and hence it can be of either sign, depending on the correlation parameters, ρ_{xz}, ρ_{yz}, in the model.

2.4 Examples

In this section, we present some examples of the expansion formula for the covariance.

2.4.1 Two-factor log-normal model

We begin by analyzing the accuracy of the stochastic Taylor expansion formula for the two-factor log-normal model. This is a special case of the two-factor SABR model, with $\beta_x = \beta_y = 1$ and volatility-of-volatility parameter, η, equal to zero, i.e.

$$\begin{aligned}dX(t) &= \sigma_x X(t) dW^{(1)}(t),\\ dY(t) &= \sigma_y Y(t) dW^{(2)}(t).\end{aligned} \tag{2.4.1}$$

Since an analytic solution for the covariance is available for this model, the stochastic Taylor expansion solution (2.4.5) is compared to the exact solution. It is shown that the stochastic Taylor expansion of the covariance between two log-normally distributed variables agrees well with the Taylor expansion of the exact solution of the same quantity.

We reformulate system (2.4.1) by making the time rescaling $t = \epsilon^2\tau$, so that the processes $\epsilon W^{(1)}(\tau)$, $\epsilon W^{(2)}(\tau)$ have the same variances as $W^{(1)}(t)$ and $W^{(2)}(t)$, respectively,

$$\begin{aligned} dX^{(\epsilon)}(\tau) &= \epsilon\sigma_x X(\tau) dW^{(1)}(\tau), \\ dY^{(\epsilon)}(\tau) &= \epsilon\sigma_y Y(\tau) dW^{(2)}(\tau), \end{aligned} \tag{2.4.2}$$

with $dW^{(1)}(\tau)dW^{(2)}(\tau) = \rho dt$.

The asymptotic expansion of $X^{(\epsilon)}(\tau)$, up to fourth-order, reads:

$$\begin{aligned} & X^{(\epsilon)}(\tau) \\ = \; & X(0) + \epsilon\sigma_x X(0)\int_0^\tau dW^{(1)}(s) + \frac{1}{2}\epsilon^2 2\sigma_x^2 X(0)\int_0^\tau\int_0^{s_1} dW^{(1)}(s_2)dW^{(1)}(s_1), \\ & + \frac{1}{6}\epsilon^3 6\sigma_x^3 X(0)\int_0^\tau\int_0^{s_1}\int_0^{s_2} dW^{(1)}(s_3)dW^{(1)}(s_2)dW^{(1)}(s_1) + O(\epsilon^4). \end{aligned} \tag{2.4.3}$$

Similarly, we have

$$\begin{aligned} Y^{(\epsilon)}(\tau) =& Y(0) + \epsilon\sigma_y Y(0)\int_0^\tau dW^{(1)}(s) + \epsilon^2\sigma_y^2 Y(0)\int_0^\tau\int_0^{s_1} dW^{(1)}(s_2)dW^{(1)}(s_1) \\ & + \epsilon^3\sigma_y^3 Y(0)\int_0^\tau\int_0^{s_1}\int_0^{s_2} dW^{(1)}(s_3)dW^{(1)}(s_2)dW^{(1)}(s_1) + O(\epsilon^4). \end{aligned} \tag{2.4.4}$$

Substituting Eqs. (2.4.3) and (2.4.4) in the definition of the covariance, we find

$$\begin{aligned} & \mathbb{C}\mathrm{ov}_0\left[X^{(\epsilon)}(\tau), Y^{(\epsilon)}(\tau)\right] \\ =& \mathbb{E}\left[\left(X^{(\epsilon)}(\tau) - \mathbb{E}[X^{(\epsilon)}(\tau)]\right)\left(Y^{(\epsilon)}(\tau) - \mathbb{E}[Y^{(\epsilon)}(\tau)]\right)\right] \\ =& \mathbb{E}\left[\left(\epsilon^2 X^{(1)}(\tau)Y^{(1)}(\tau) + \frac{1}{4}\epsilon^4 X^{(2)}(\tau)Y^{(2)}(\tau) + \frac{1}{6}\epsilon^4 X^{(1)}(\tau)Y^{(3)}(\tau) + \frac{1}{6}\epsilon^4 X^{(3)}(\tau)Y^{(1)}(\tau)\right)\right] \\ & + O(\epsilon^6), \end{aligned}$$

where

$$\begin{aligned} X^{(1)}(\tau)Y^{(1)}(\tau) &= \sigma_x\sigma_y X(0)Y(0)\rho\tau, \\ X^{(2)}(\tau)Y^{(2)}(\tau) &= 4\sigma_x^2\sigma_y^2 X(0)Y(0)\rho^2\frac{\tau^2}{2}, \\ X^{(1)}(\tau)Y^{(3)}(\tau) &= 6\sigma_x\sigma_y^3 X(0)Y(0)\rho\tau\int_0^{s_1}\int_0^{s_2} dW^{(2)}(s_3)dW^{(2)}(s_2), \\ X^{(3)}(\tau)Y^{(1)}(\tau) &= 6\sigma_y\sigma_x^3 X(0)Y(0)\rho\tau\int_0^{s_1}\int_0^{s_2} dW^{(1)}(s_3)dW^{(1)}(s_2). \end{aligned}$$

Since $\int_0^{s_1}\int_0^{s_2} dW^{(1)}(s_3)dW^{(1)}(s_2)$ and $\int_0^{s_1}\int_0^{s_2} dW^{(2)}(s_3)dW^{(2)}(s_2)$ have zero expectation, we find

$$\begin{aligned}\mathbb{C}\mathrm{ov}_0\big[X^{(\epsilon)}(\tau), Y^{(\epsilon)}(\tau)\big] &= \frac{1}{2}\epsilon^2 X^{(1)}(\tau)Y^{(1)}(\tau) + \frac{1}{4}\epsilon^4 X^{(2)}(\tau)Y^{(2)}(\tau)\\ &= X(0)Y(0)\Big(\epsilon^2\sigma_x\sigma_y\rho\tau + \epsilon^4\sigma_x^2\sigma_y^2\rho^2\frac{\tau^2}{2}\Big)\\ &= X(0)Y(0)\Big(\sigma_x\sigma_y\rho t + \sigma_x^2\sigma_y^2\rho^2\frac{t^2}{2}\Big). \qquad (2.4.5)\end{aligned}$$

The first term in (2.4.5) is the Gaussian approximation which grows linearly with time t. The second term acts as the convexity correction and accounts for the non-Gaussian part of the distribution.

Due to the tractability of log-normally distributed random variables, their covariance can be computed directly by:

$$\begin{aligned}&\mathbb{C}\mathrm{ov}_0\big[X(t), Y(t)\big]\\ =&\mathbb{E}\big[(X(t)-\mathbb{E}[X(t)])(Y(t)-\mathbb{E}[Y(t)])\big] = \mathbb{E}\big[X(t)Y(t)\big] - \mathbb{E}[X(t)]\mathbb{E}[Y(t)]\\ =&X(0)Y(0)e^{-\frac{1}{2}\sigma_x^2 t-\frac{1}{2}\sigma_y^2 t}\mathbb{E}\big[e^{\sigma_x Z^{(1)}(t)+\sigma_y(\rho Z^{(1)}(t)+\sqrt{1-\rho^2}Z^{(2)}(t)}\big] - X(0)Y(0)\\ =&X(0)Y(0)e^{-\frac{1}{2}\sigma_x^2 t-\frac{1}{2}\sigma_y^2 t}\mathbb{E}\big[e^{(\sigma_x+\sigma_y\rho)Z^{(1)}(t)}e^{\sigma_y\sqrt{1-\rho^2}Z^{(2)}(t)}\big] - X(0)Y(0)\\ =&X(0)Y(0)e^{-\frac{1}{2}\sigma_x^2 t-\frac{1}{2}\sigma_y^2 t}e^{\frac{1}{2}(\sigma_x+\sigma_y\rho)^2 t+\frac{1}{2}\left(\sigma_y\sqrt{1-\rho^2}\right)^2 t} - X(0)Y(0)\\ =&X(0)Y(0)e^{\rho\sigma_x\sigma_y t} - X(0)Y(0), \qquad (2.4.6)\end{aligned}$$

since $\mathbb{E}\big[e^{(\sigma_x+\sigma_y\rho)Z^{(1)}(t)-\frac{1}{2}(\sigma_x+\sigma_y\rho)^2 t}e^{\sigma_y\sqrt{1-\rho^2}Z^{(2)}(t)-\frac{1}{2}\left(\sigma_y\sqrt{1-\rho^2}\right)^2 t}\big] = 1$, and $Z^{(1)}(t)$ and $Z^{(2)}(t)$ are independent Brownian motions.

The expansion solution (2.4.5) for the variables $X(t)$ and $Y(t)$ in (2.4.1) agrees with the first two terms of the Taylor expansion of the solution in Eq. (2.4.6). We denote the term $\rho\sigma_x\sigma_y t$ in (2.4.5) by "Expn. O1" (the first-order term), the term $\frac{1}{2}\big(\rho\sigma_x\sigma_y t\big)^2$ in (2.4.5) by "Expn. O2" (the second-order term), etc.

Numerical experiment

In this section, we evaluate numerically the accuracy of the expansion solution (2.4.5) for two different sets of parameters, given in Table 2.1. The correlation between the two Brownian motions is set to $\rho_{xy} = 0.6$, for both experiments.

Table 2.1: Two parameter sets for the evaluation of the expansion in the log-normal case.

	$X(0)$	σ_x	$Y(0)$	σ_y
High vol.	1	40%	1	45%
Low vol.	1	20%	1	25%

The results of the Taylor approximation are presented in Table 2.2. The expansion is accurate, especially when the volatilities are small. In order to

Table 2.2: Comparison of the accuracy of the expansion for the covariance against the exact covariance, for the log-normal case.

Expn.	O2	O3	O4	Exact	O2	O3	O4	Exact
	Low volatility				High volatility			
2 yr	0.0618	0.0618	0.0618	0.0618	0.2393	0.2410	0.2411	0.2411
5 yr	0.1613	0.1618	0.1618	0.1618	0.6858	0.7120	0.7156	0.7160
10 yr	0.3450	0.3495	0.3498	0.3499	1.6632	1.8732	1.9298	1.9447
15 yr	0.5513	0.5664	0.5681	0.5683	2.9322	3.6408	3.9278	4.0531
20 yr	0.7800	0.8160	0.8214	0.8221	4.4928	6.1724	7.0794	7.6711
25 yr	1.0313	1.1016	1.1147	1.1170	6.3450	9.6255	11.8398	13.8797
30 yr	1.3050	1.4265	1.4538	1.4596	8.4888	14.1575	18.7492	24.5337

Unit: basis points

obtain a satisfactory accuracy in the case of high volatility, one has to expand up to terms of higher-order. In the interest rate derivative pricing problems, however, the value of an underlying is typically only a few percentage points. Figure 2.1 displays the accuracy results graphically. The results of the left-hand side in Table 2.2 suggest that the expansion up to order fourth-order is sufficient even for long time to maturity.

2.4.2 Constant maturity swap

We will now analyze the approximation for the CMS contract, which we first describe in some more detail.

Recall that the computation of the CMS convexity correction is reduced to the approximation:

$$Cc(t) \approx \frac{N(t)}{N^*(t)}\Big(\frac{\partial f}{\partial X(t)}\mathbb{V}\mathrm{ar}_t^Q[X(T)] + \frac{\partial f}{\partial Y(t)}\mathbb{C}\mathrm{ov}_t^Q[X(T),Y(T)]\Big), \tag{2.4.7}$$

where the numéraire ratio is a function of the rates $X(s)$ and $Y(s)$, i.e., $\frac{N^*(s)}{N(s)} = f\big(X(s),Y(s)\big)$, $s \in (t,T]$.

The CMS pricing formula reads:

$$\begin{aligned} \mathrm{CMS}(t) &= P(t,T_{\mathrm{pay}})\mathbb{E}^{Q^{T_{\mathrm{pay}}}}[\mathrm{SR}(T_0)] \\ &= A(t)\mathbb{E}^{Q^A}\left[\frac{\mathrm{SR}(T_0)P(T_0,T_{\mathrm{pay}})}{A(T_0)}\Big|\mathcal{F}_t\right], \end{aligned} \tag{2.4.8}$$

where

- t denotes the current time point,
- T_0 is the starting (or expiry) time of the CMS contract,
- T_{pay} is the delayed payment time of the CMS contract, i.e. $T_{\mathrm{pay}} = T_0 + \tau$ where τ is the time fraction of the payment delay,
- $A(T_0)$ is the T_0-value of the annuity of the reference swap $\mathrm{SR}(T_0)$, i.e. $A(T_0) = \sum_{i=1}^m \delta_i P(T_0,T_i)$ with accrual factors δ_i,

- T_i $(i = 1, 2, 3, \ldots, m)$ represents a series of m resetting dates for the underlying reference swap,

- $\mathrm{SR}(T_0)$ stands for the T_0-value of a swap starting from T_0 with maturity T_m, i.e. $\mathrm{SR}(T_0) = \mathrm{SR}(T_0, T_0, T_m)$,

Two measures are involved in the CMS pricing problem:

- The *T-forward measure*, which is denoted by $Q^{T_{\text{pay}}}$, is associated to zero coupon bonds with some maturity T;

- The *annuity measure*, denoted by Q^A, is the natural martingale measure for (forward starting) swaps and swaptions. The annuity pays 1 Euro at each coupon day of the swap, accrued according to the swap day count conventions.

Note that the swap rate, $\mathrm{SR}(t)$, corresponds to the rate $X(t)$ in Eq. (2.4.7). There is no drift term under the annuity measure. Since the LIBOR rate on the payment date, $L(t, T_{\text{pay}})$, corresponds to the rate $Y(t)$ in our problem formulation, it is, in general, not a martingale process under the swap measure.

The variance/covariance quantity in (2.4.7) can be approximated by the Taylor expansion formula once the parameters are calibrated. The numéraire ratio is problem-specific and the partial derivatives in (2.4.7) have to be determined according to the payment features. The numéraire ratio is here given by: $N(t)/N^*(t) = P(t, T_{\text{pay}})/A(t)$.

$P(t, T_{\text{pay}})$ is driven by a LIBOR rate, so that:

$$P(t, T_{\text{pay}}) = P(t, T_0) \frac{1}{1 + \tau L(t, T_{\text{pay}})}.$$

The swap annuity is defined by $A(t) := \sum_{i=1}^{m} \delta_i P(t, T_i)$. This expression is approximated by the following relation

$$P(t, T_i) \approx P(t, T_0) \prod_{j=1}^{i} \frac{1}{1 + \delta_j \mathrm{SR}(t)}, \quad i = 1, \ldots, m.$$

Then, the annuity reads

$$A(t) = \sum_{i=1}^{m} \delta_i P(t, T_i) \approx P(t, T_0) \sum_{i=1}^{m} \left(\delta_i \prod_{j=1}^{i} \frac{1}{1 + \delta_j \mathrm{SR}(t)} \right).$$

So, the numéraire ratio considered in the CMS pricing problem reads

$$\begin{aligned} \frac{N_t^*}{N_t} =& f\Big(\mathrm{SR}(t), L(t, T_{\text{pay}})\Big) \\ =& \frac{P(t, T_{\text{pay}})}{A(t)} = \frac{1}{\big(1 + \tau L(t, T_{\text{pay}})\big) \sum_{i=1}^{m} \left(\delta_i \prod_{j=1}^{i} \frac{1}{1 + \delta_j \mathrm{SR}(t)} \right)}. \end{aligned} \tag{2.4.9}$$

The partial derivatives to the numéraire ratio (2.4.9) w.r.t the swap rate and the LIBOR rate then read:

$$\frac{\partial f}{\partial \mathrm{SR}(t)} = \sum_{i=1}^{m} \delta_i \sum_{j=1}^{i} \frac{\delta_j}{1+\delta_j \mathrm{SR}(t)} \cdot \frac{1}{\prod_{j=1}^{i}\left(1+\delta_j \mathrm{SR}(t)\right)} \cdot \left(\frac{1}{\sum_{i=1}^{m} \delta_i \prod_{j=1}^{i} \frac{1}{1+\delta_j \mathrm{SR}(t)}}\right)^2, \tag{2.4.10}$$

$$\frac{\partial f}{\partial L(t,T_{\mathrm{pay}})} = -\frac{\tau}{\left(1+\tau L(t,T_{\mathrm{pay}})\right)^2} \frac{1}{\sum_{i=1}^{m}\left(\delta_i \prod_{j=1}^{i} \frac{1}{1+\delta_j \mathrm{SR}(t)}\right)}. \tag{2.4.11}$$

Numerical experiment for CMS with SABR model

Here, we compare numerically the accuracy of the approximation for the CMS convexity correction with a reference solution, generated by the Monte Carlo method, and also with other approximations available in the literature. We are also interested in the price impact of the factor decorrelation. The SABR model is popular among practitioners and often used as an "implied volatility interpolation tool" for swaptions and caplets.

The CMS contract priced here pays a 10 years, annually reset, par swap rate with a 6 month payment delay.

For the multi-factor SABR model described in Eq. (2.2.6), the parameters chosen for the processes $X(t)$ and $Y(t)$ are given in Table 2.3.

Table 2.3: Parameters for the CMS experiments.

X_0	vol-of-vol	Corr. (ρ_{xz})	skew (β)
3.4%	0.2	-0.4	1 and 0.6
Y_0	vol-of-vol	Corr. (ρ_{xz})	skew (β)
3.0%	0.2	-0.5	1 and 0.6

In the Monte Carlo method for the benchmark prices for the convexity correction, we choose the Euler time discretization scheme and use a grid of 10 steps per year.

The following methods are compared in this section:

1. *MC* represents the short time step Monte Carlo method for the two-factor SABR model;

2. *Expn.* is the expansion solution derived in this chapter. The features captured by this method are the skew/smile surface and the factor decorrelation;

3. *Gausn. App.* denotes the Gaussian approximation method, obtained by assuming that the underlying diffusion processes are Gaussian; This approximation can model a terminal decorrelation but not a smile or skew.

4. *1fSK* is a one-factor model with skew/smile features. More specifically, we consider Mercurio's method [79] in this experiment. The skew/smile is captured, but the terminal decorrelation is not modeled.

In order to investigate the price impact of factor decorrelation, we set up two experiments with different correlations between the swap rate and the LIBOR rate, ρ_{xy}: One with a *positive* correlation, $\rho_{xy} = 0.3$, and another in which a *negative* correlation is chosen, $\rho_{xy} = -0.3$. The numerical results obtained are summarized in Figure 2.2 and Table 2.4 for $\beta = 1$, and in Table 2.5 for $\beta = 0.6$.

The expansion solutions derived in this chapter, Expn, agree well with the Monte Carlo benchmark prices for these two experiments, see Tables 2.4 and 2.5. The one-factor-with-skew model, 1fSK, returns, by construction, the same price in the two experiments, which is an obvious drawback of that model. For short expiry times, e.g. smaller than 10 years, all four methods provide more-or-less the same level of accuracy for $\beta = 1$. This is different for $\beta = 0.6$, for which only approximation "Expn" agrees well with the benchmark prices. For longer expiry times, the differences between the Monte Carlo prices and the one-factor, as well as the Gaussian, approximation increase. Our approximation, Expn, resembles the benchmark prices rather well, even up to expiration times of 30 years.

When $\beta \approx 1$, the skew/smile feature has a more significant impact than the terminal decorrelation. Figure 2.2 shows that the one-factor model with skew is preferred over the two-factor model, Gausn. App., for all expiry times farther than 5 years.

The reason is that a significant part of the change in $f\big(X(t), Y(t)\big) = P(t, T_{\text{pay}})/A(t)$ is related to changes in the level of the annuity. The changes in the numéraire ratio are therefore highly *correlated* to the movements in the swap rate. In the present experiment, the partial derivative $\partial f/\partial X$ is significantly larger than $\partial f/\partial Y$ ($\partial f/\partial X = 0.595602318 >> \partial f/\partial Y = -0.058036757$). So, even if the covariance between the two rates is low by specifying a strongly negative correlation between the swap and the LIBOR rate, the terminal decorrelation of the convexity correction, $\frac{\partial f}{\partial Y}\mathbb{C}\text{ov}[X, Y]$, is still very small compared to the overall convexity correction.

However, when $\beta \leq 0.5$ the terminal decorrelation has a more significant impact and the accuracy of the expansion reduces quickly with increasing time. Small values of β imply a stronger variability in the underlying process, and the approximation obtained is not sufficiently accurate then.

Table 2.4: The CMS convexity corrections with $\beta_x = \beta_y = 1$ and two different swap-LIBOR correlations. Left-hand side: $\rho_{xy} > 0$; Right-hand side: $\rho_{xy} < 0$

	MC	*Expn.*	Gausn. App.	*1fSk*	*MC*	*Expn.*	Gausn. App.	*1fSk*
	Positive Correlation $\rho_{xy} = 0.3$				Negative Correlation $\rho_{xy} = -0.3$			
5 yr	16.75	17.30	16.49	15.98	17.33	18.07	17.36	15.98
10 yr	35.36	36.23	32.98	33.36	37.54	37.05	34.73	33.36
15 yr	59.18	56.78	49.47	52.15	59.97	58.41	52.09	52.15
20 yr	82.38	78.95	65.96	72.34	82.73	80.70	69.45	72.34
25 yr	102.88	102.75	82.45	93.94	122.24	104.38	86.81	93.94
30 yr	140.71	128.17	98.94	116.94	152.49	129.48	104.18	116.94

Unit: basis points

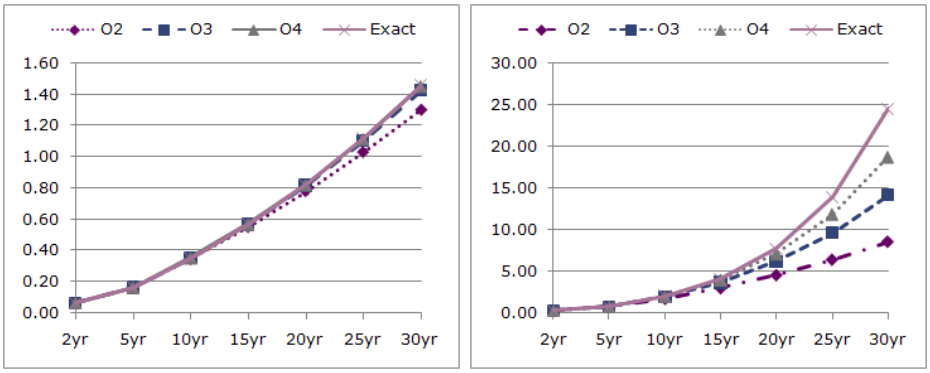

Figure 2.1: Accuracy of the expansion solution (2.4.5) for two-factor log-normal model. Left-hand side: low volatility, i.e. $\sigma_x = 20\%$ and $\sigma_y = 45\%$; Right-hand side: high volatility with $\sigma_x = 40\%$ and $\sigma_y = 45\%$.

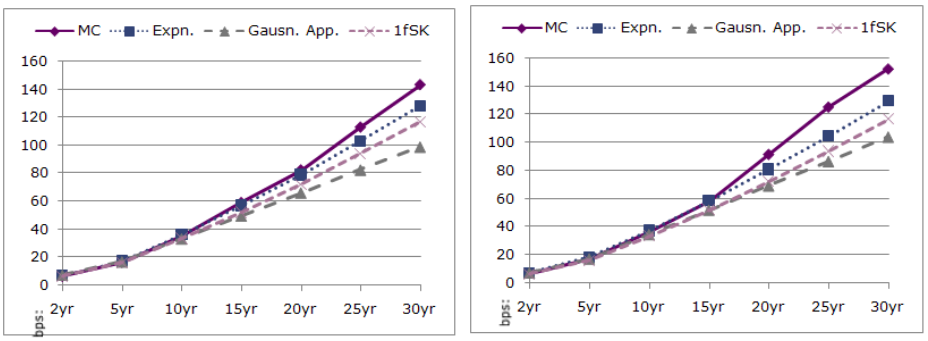

Figure 2.2: Comparison of several convexity correction methods with $\beta_x = \beta_y = 1$ and two different swap-LIBOR correlations. Left-hand side: the convexity correction in time for $\rho_{xy} > 0$. Right-hand side: convexity correction for $\rho_{xy} < 0$.

Table 2.5: The CMS convexity corrections with $\beta_x = \beta_y = 0.6$ and two different swap-LIBOR correlations. Left-hand side: $\rho_{xy} > 0$; Right-hand side: $\rho_{xy} < 0$

	MC	*Expn.*	Gausn. App.	*1fSk*	*MC*	*Expn.*	Gausn. App.	*1fSk*
	Positive Correlation $\rho_{xy} = 0.3$				Negative Correlation $\rho_{xy} = -0.3$			
5 yr	46.40	45.93	58.15	42.11	47.17	46.91	59.89	42.11
10 yr	119.00	121.10	169.98	110.34	122.20	121.54	173.47	110.34
15 yr	237.53	225.52	335.49	204.70	214.52	223.89	340.72	204.70
20 yr	371.47	359.18	554.68	325.18	345.88	353.96	561.66	325.18
25 yr	535.45	522.08	827.55	471.79	529.24	511.74	836.27	471.79
30 yr	740.33	714.23	1154.10	644.53	728.20	697.25	1164.57	644.53

Unit: basis points

Decomposition of the error

Two approximations that we made that may have an impact on the accuracy of our pricing formula are the following: We keep the partial derivatives $\frac{\partial f}{\partial X(s)}, \frac{\partial f}{\partial Y(s)}$ constant, at their initial values, and we approximate the variance/covariance term by the stochastic Taylor expansion. Here, we consider the impact of each of these simplifications separately, so that we can indicate directions for further improvement. We consider here the case $\beta = 1$.

First of all, we focus on the error which originates from the assumption of constant partial derivatives. We compare a CMS convexity correction *with frozen partial derivatives* sampled with the Monte Carlo method with the Monte Carlo results for the true convexity correction. More specifically, given the same set of Monte Carlo paths, we compute the convexity correction, with constant partial derivatives, by

$$\begin{aligned} &Cc^{(1)}(t) \\ =&\frac{1}{M}\frac{\partial f}{\partial X(t)}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\Big(\big(\sigma^{(k)}(s_{i-1})\big)^2\big(X^{(k)}(s_{i-1})\big)^{2\beta_x}\Delta\Big)\Big]+ \\ &\frac{1}{M}\frac{\partial f}{\partial Y(t)}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\Big(\sigma^{(k)}(s_{i-1})\big(X^{(k)}(s_{i-1})\big)^{\beta_x}\sigma^{(k)}(s_{i-1})\big(Y^{(k)}(s_{i-1})\big)^{\beta_y}\rho_1\rho_2\Delta\Big)\Big], \end{aligned}$$

where $i = 1, 2, \ldots, N$ represents the index for the time steps and k denotes for the number of trails.

We then compute the convexity correction by a step-wise approximation of the time varying partial derivatives, i.e.,

$$\begin{aligned} &Cc^{(2)}(t) \\ =&\frac{1}{M}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\frac{\partial f}{\partial X(s_{i-1})}\Big(\big(\sigma^{(k)}(s_{i-1})\big)^2\big(X^{(k)}(s_{i-1})\big)^{2\beta_x}\Delta\Big)\Big]+ \\ &\frac{1}{M}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\frac{\partial f}{\partial Y(s_{i-1})}\Big(\sigma^{(k)}(s_{i-1})\big(X^{(k)}(s_{i-1})\big)^{\beta_x}\sigma^{(k)}(s_{i-1})\big(Y^{(k)}(s_{i-1})\big)^{\beta_y}\rho_1\rho_2\Delta\Big)\Big]. \end{aligned}$$

Since we choose the same set of paths, the distributional statistics for the two formulas above are exactly the same. The difference in the prices, $Cc^{(1)}(t)$ and $Cc^{(2)}(t)$, therefore comes from the assumption of constant partial derivatives. The results for $\beta = 1$ are summarized in Table 2.6.

The approximation with constant partial derivatives consistently undervalues the convexity correction. The error grows almost linearly in time (see Figure 2.3). This error can be explained as follows: In the derivation of the convexity correction formula (2.2.5), constant values for the partial derivatives are set, so that

$$\mathbb{E}_t^Q\Big[\int_t^T \frac{\partial f}{\partial X(s)} dX(s) \cdot \int_t^T dX(s)\Big] \approx \frac{\partial f}{\partial X(t)} \mathbb{E}_t^Q\Big[\int_t^T dX(s) \cdot \int_t^T dX(s)\Big].$$

This can be interpreted as a first-order approximation to the stochastic integral $\int_t^T \frac{\partial f}{\partial X(s)} dX(s)$. The accuracy will improve if we include higher-order terms, i.e.,

$$\begin{aligned}
\int_t^T \frac{\partial f}{\partial X(s)} dX(s) &= \int_t^T \Big(\frac{\partial f}{\partial X(t)} + \int_t^s \big(\frac{\partial^2 f}{\partial X(s_1)^2}\big) dX(s_1) + O(s)\Big) dX(s) \\
&= \frac{\partial f}{\partial X(t)} \int_t^T dX(s) + \int_t^T \int_t^s \big(\frac{\partial^2 f}{\partial X(s_1)^2}\big) dX(s_1) dX(s) + o(T) \\
&= \frac{\partial f}{\partial X(t)} \int_t^T dX(s) + R,
\end{aligned}$$

where R is the remainder resulting from the approximation with constant partial derivatives. Thus, the integral related to the second-order partial derivatives, which is the leading term of the remainder R, forms the basis of the pricing error in Table 2.6. Under the assumption in Section 2.2.1 that the numéraire ratio is varying slowly and thus that the second-order partial derivatives in the swap rate are small, these second and higher-order partial derivative terms can be neglected.

Table 2.6: Approximation error due to constant partial derivatives, $\beta = 1$.

	Positive Correlation $\rho_{xy} = 0.3$			Negative Correlation $\rho_{xy} = -0.3$		
	Reference	Prox.	Error	Reference	Prox.	Error
2 yr	6.59	6.55	-0.03	6.5878	6.87	-0.04
5 yr	16.56	16.22	-0.34	17.33	16.90	-0.43
10 yr	35.36	32.67	-2.69	36.35	34.16	-2.19
15 yr	59.18	51.72	-7.46	58.74	52.48	-6.26
20 yr	82.38	69.80	-12.58	91.59	76.05	-15.54
25 yr	112.92	91.64	-21.28	125.42	99.68	-25.74
30 yr	143.16	104.89	-38.26	152.64	117.97	-34.67

Unit: basis points

We also consider the accuracy of the variance/covariance approximation by comparing the expansion solution (2.3.3) with the following Monte Carlo statis-

Table 2.7: Error in the variance and in the covariance approximation for positive and negative correlations, $\beta = 1$.

	Positive Correlation $\rho_{xy} = 0.3$					
	Variance approximation			Covariance approximation		
	MC	Expn.	Err.	MC	Expn.	Err.
2 yr	1.43	1.47	0.04	0.36	0.37	0.01
5 yr	3.55	3.78	0.23	0.80	0.88	0.08
10 yr	7.10	7.89	0.79	1.31	1.61	0.30
15 yr	11.00	12.33	1.33	1.66	2.20	0.53
20 yr	15.00	17.11	2.11	1.90	2.63	0.73
25 yr	20.00	22.21	2.21	2.05	2.91	0.86
30 yr	24.00	27.65	3.65	2.18	3.05	0.87
	Negative Correlation $\rho_{xy} = -0.3$					
	Variance approximation			Covariance approximation		
	MC	Expn.	Err.	MC	Expn.	Err.
2 yr	1.44	1.47	0.03	-0.34	-0.36	-0.01
5 yr	3.54	3.78	0.24	-0.73	-0.79	-0.07
10 yr	7.18	7.89	0.71	-1.12	-1.26	-0.13
15 yr	12.00	12.33	0.33	-1.34	-1.39	-0.05
20 yr	16.00	17.11	1.11	-1.48	-1.20	0.28
25 yr	21.00	22.21	1.21	-1.56	-0.67	0.89
30 yr	25.00	27.65	2.65	-1.63	0.18	1.81

Unit: basis points

tics:

$$\mathbb{V}\mathrm{ar} \approx \frac{1}{M}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\Big(\big(\sigma^{(k)}(s_{i-1})\big)^2 (X^{(k)}(s_{i-1}))^{2\beta_x}\Delta\Big)\Big],$$

$$\mathbb{C}\mathrm{ov} \approx \frac{1}{M}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\Big(\sigma^{(k)}(s_{i-1})\big(X^{(k)}(s_{i-1})\big)^{\beta_x}\sigma^{(k)}(s_{i-1})\big(Y^{(k)}(s_{i-1})\big)^{\beta_y}\rho_1\rho_2\Delta t\Big)\Big].$$

We find that the approximation of the variance by the expansion over-estimates the true variance (see Figure 2.4). In the current model, the variance component dominates the convexity correction as the partial derivatives of the numéraire ratio w.r.t the swap rate are approximately a factor of 10 larger in size than the partial derivatives of the numéraire ratio w.r.t the LIBOR rate. Hence the approximation errors in the variance component and the corresponding partial derivatives represent the dominant part of the error.

This over-estimation of the variance cancels out, to a large extent, the under-valuation due to the constant approximation of the partial derivatives, and, therefore, the overall error is smaller than the individual errors generated by each approximation. Table 2.7 presents the details of this test for $\beta = 1$.

Finally, we found that the approximation for the covariance is not stable for contracts with long expiry times and that the error increases with cubic order in time (see Figure 2.5). This suggests that an approximation of quadratic

order (in time) is not sufficient for the estimation of the covariance and that higher-order terms (e.g. $O(t^6)$ terms) need to be included for a more accurate approximation.

2.5 Conclusion

In this chapter we have focused on the convexity correction for CMS products under a two-factor SABR model. We derived an approximation for the convexity correction by applying the *small time asymptotics* technique to the Wiener processes involved. An efficient and easy-to-implement approximation formula for the CMS convexity correction is the result of this work.

By numerical experiments, comparing with the corresponding Monte Carlo prices, we find that the approximations result in satisfactory accuracy for β-values larger than one-half. In order to detail the impact of the various approximations in our pricing approach, we have set up numerical experiments to determine the individual errors of each approximation. Apparently, two significant parts of the approximation error cancel out, to a large extent. However, a fourth-order expansion formula for the covariance as presented here does not appear to be fully sufficient for contracts with very long expiration times, like thirty years.

The approximation has been derived for payoffs that are linear in the swap rate. For more general contracts, the constant partial derivatives approximation requires improvement. Furthermore, only a small amount of terminal decorrelation can be captured by adding an additional factor on the payment leg, since the annuity is still driven by one factor only. A two-factor model which has the two factors that are principal components of the empirical covariance matrix could in principle describe the terminal decorrelation of forward and swap rates in a more realistic way.

Appendix A: Description of the Monte Carlo Scheme

We explain the set up of the Monte Carlo simulation of the covariance. The implementation is done in a short time step procedure.

Given the SABR dynamics of X_t and Y_t described in Eq. (2.2.6), where the stochastic volatility process of both rates is driven by Brownian motion Z_t.

By the Euler discretization of the stochastic differential equation system, we have the following discretization for X_t in time:

$$\begin{aligned} X_{s_i} &= X_{s_{i-1}} + \sigma(s_{i-1})(X_{s_{i-1}})^{\beta_x}\left(\lambda_x Z_i + \sqrt{1-\lambda_x^2}U_i\right)\sqrt{\Delta}, \text{ and} \\ \sigma(s_i) &= \sigma(s_{i-1}) + \eta_x \sigma(s_{i-1}) Z_i \sqrt{\Delta}, \qquad \forall s_i = t + i\Delta \leq T, \end{aligned}$$

where Z_i and U_i are independent Gaussian pseudo-random numbers (with zero

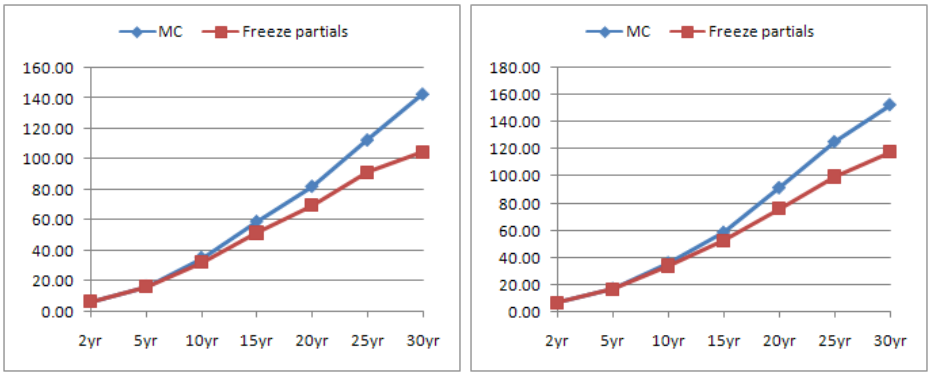

Figure 2.3: Error due to constant partial derivatives, with $\beta = 1$. Left-hand side: positive correlation between swap rate and the LIBOR rate. Right-hand side: negative correlation between these two rates.

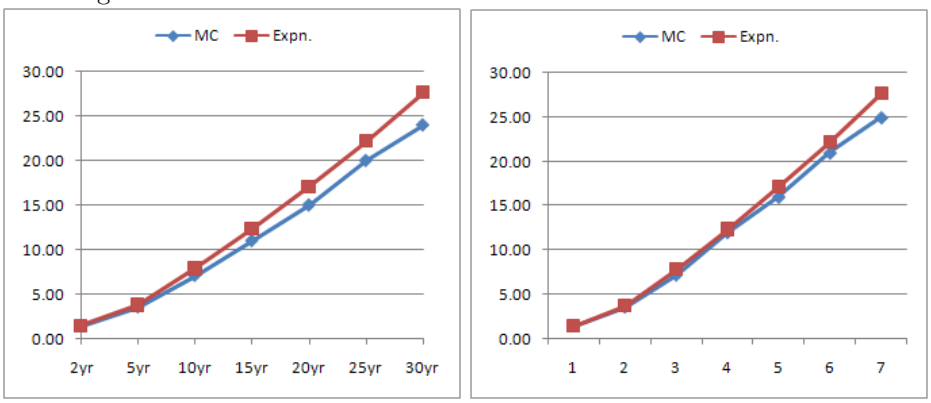

Figure 2.4: Error due to the expansion for the variance, $\beta = 1$. Left-hand side: positive correlation between swap rate and the LIBOR rate; Right-hand side: negative correlation between these two rates.

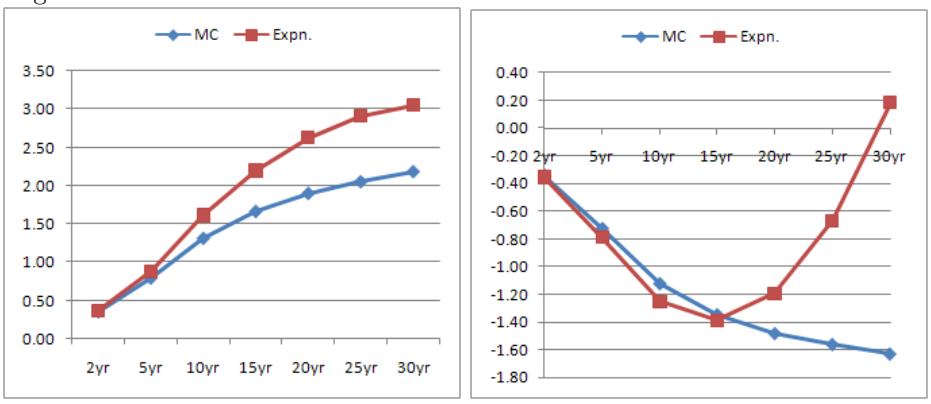

Figure 2.5: Error due to the approximation for the covariance, $\beta = 1$. Left-hand side: positive correlation between swap rate and the LIBOR rate; Right-hand side: negative correlation between these two rates.

mean and unit variance). Similarly, the process Y_t, in discrete time, reads

$$\begin{aligned} Y_{s_i} =& Y_{s_{i-1}} + \Big(\frac{A(s_{i-1})}{P(s_{i-1},T_{\text{pay}})} - \frac{A(s_{i-2})}{P(s_{i-2},T_{\text{pay}})}\Big)\Big(Y_{s_{i-1}} - Y_{s_{i-2}}\Big) \\ &+ \sigma(s_{i-1})(Y_{s_{i-1}})^{\beta_y}\big(\lambda_y Z_i + \sqrt{1-\lambda_y^2}V_i\big)\sqrt{\Delta}, \end{aligned}$$

and

$$\sigma(s_i) = \sigma(s_{i-1}) + \eta_y \sigma(s_{i-1}) Z_i \sqrt{\Delta}, \qquad \forall s_i = t + i\Delta \leq T,$$

where Z_i represents the same set of random numbers used before, and the V_i represents another set of random numbers. As a result, for the variance and covariance quantities in Eq. (2.4.7), in discrete time, we find:

$$\begin{aligned} &\frac{\partial f}{\partial X_t}\mathbb{V}\text{ar}_t^Q[X_T] + \frac{\partial f}{\partial Y_t}\mathbb{C}\text{ov}_t^Q[X_T, Y_T] \\ \approx &\frac{\partial f}{\partial X_t}\mathbb{E}\Big[\sum_{i=1}^{N}\Big(\sigma(s_{i-1})^2 (X_{s_{i-1}})^{2\beta_x}\Delta\Big)\Big] \\ &+ \frac{\partial f}{\partial Y_t}\mathbb{E}\Big[\sum_{i=1}^{N}\Big(\sigma(s_{i-1})(X_{s_{i-1}})^{\beta_x}\sigma(s_{i-1})(Y_{s_{i-1}})^{\beta_y}\rho_1\rho_2\Delta\Big)\Big] \\ \approx &\frac{1}{M}\frac{\partial f}{\partial X_t}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\Big(\big(\sigma^{(k)}(s_{i-1})\big)^2 (X_{s_{i-1}}^{(k)})^{2\beta_x}\Delta\Big)\Big] \\ &+ \frac{1}{M}\frac{\partial f}{\partial Y_t}\sum_{k=1}^{M}\Big[\sum_{i=1}^{N}\Big(\sigma^{(k)}(s_{i-1})(X_{s_{i-1}}^{(k)})^{\beta_x}\sigma^{(k)}(s_{i-1})(Y_{s_{i-1}}^{(k)})^{\beta_y}\rho_1\rho_2\Delta\Big)\Big], \end{aligned}$$

where the superscript denotes the k-th path of the Monte Carlo simulation. The expectation is approximated by the average of a large number of paths. Note that we have applied the Itô product formula to obtain the time discretization scheme above.

CHAPTER 3

A low-bias simulation scheme for the SABR stochastic volatility model

This chapter is adapted from the article
"A low-bias simulation scheme for the SABR stochastic volatility model"
accepted by *International Journal of Theoretical and Applied Finance*,
in Volume 15, Issue 02, pp. 1-37, (2012) [32].

The Stochastic Alpha Beta Rho (SABR) stochastic volatility model is widely used in the financial industry for the pricing of fixed income instruments. In this chapter we develop a low-bias simulation scheme for the SABR model, which deals efficiently with (undesired) possible negative values of the asset price process, the martingale property of the discrete scheme and the discretization bias of commonly used Euler discretization schemes. The proposed algorithm is based the analytic properties of the governing distribution. Experiments with realistic model parameters show that this scheme is robust for interest rate valuation.

3.1 Introduction

The Stochastic Alpha Beta Rho (SABR) model by Hagan [46] is a popular model in the financial industry because of the availability of an analytic asymptotic implied volatility formula. Practical applications of the SABR model include interpolation of volatility surfaces and the hedging of volatility risk. In the context of pricing interest rate derivatives, the combination of the SABR model and the market standard Libor Market Models (LMM) [94] is of particular interest. Other references on this topic include Morini & Mercurio [81], Hagan & Lesniewski [45] or Labordere [69].

The constant elasticity of variance (CEV) process, introduced by Cox [33], is an important ingredient of the SABR model. The CEV process has appeared in several other models in finance, including the CEV LMM by Andersen & Andreasen [4].

Despite the fact that the CEV model has been introduced more than 30 years ago and that various researchers have shown evidence of significant bias in the basic Euler scheme for the CEV model, only a few references devising efficient unbiased Monte Carlo schemes were found in the literature.

It was shown by Schroder [100] that the CEV process, representing the asset price dynamics of the SABR model, equals a space transformed *squared Bessel* process. As the volatility process in the SABR model is driven by a geometric Brownian motion, a close relation between the SABR model and the Heston model [48] exists: In the Heston model the asset price dynamics follow geometric Brownian motion, whereas the volatility is governed by a squared Bessel process. Due to this, it seems natural to generalize the unbiased simulation schemes for the Heston stochastic volatility model to the SABR case. Broadie & Kaya [26]'s so-called *exact simulation scheme* (the BK scheme) is based on the insight in Willard [108] that the conditional distribution, given the terminal volatility and the integrated variance in a time interval, is log-normal. In their scheme, an acceptance-rejection technique is employed to sample the variance process, and a Fourier inversion technique is applied to recover the variance process. Although the BK scheme is free of bias by construction, its practical application is hampered by its computational speed. Andersen [3] developed two efficient low-bias variants of the BK scheme, the truncated Gaussian (TG) and the quadratic exponential (QE) schemes, that are both based on the moment matching technique. Essentially, the noncentral chi-square distribution is approximated by a distribution whose moments are matched with those of the exact distribution. Since the QE scheme is based on transformations to uniform and normal random numbers, it can be implemented efficiently [43].

A direct application of the QE scheme to the SABR model does not work well, because the QE scheme is based on a squared Bessel process with a *reflecting* boundary at zero volatility, which gives rise to a sub-martingale process. It is therefore not suited to model SABR's asset price dynamics. Instead, a squared Bessel process with an *absorbing* boundary is the specification which is in agreement with the arbitrage-free constraints (and thus produces a true martingale process). Accurate handling of the absorbing boundary behavior is nontrivial, as the transition density of the absorbed process does not integrate to unity and the moments are not known in closed form.

Some simulation algorithms for the squared Bessel processes exist in the literature. Andersen & Andreasen [4] investigated basic Euler as well as log-Euler schemes for the CEV model in a Monte Carlo setting and mentioned that 'the simulated prices of caps, floors and swaptions exhibit a bias relative to the continuous-time prices ... even for an infinite number of Monte Carlo trials'. Kahl & Jäckel [63] proposed a higher-order Monte Carlo scheme based on an implicit Milstein discretization for the CEV process. This scheme does not perform well when zero is attainable, due to the discontinuity of the first derivative of the diffusion coefficient. Campolieti & Makarov [28] developed a scheme based on an acceptance-rejection sampling of the Bessel bridge process. The Bessel bridge scheme is however quite complex and its computational time is relatively high. Lord et al. [75] consider an Euler scheme in combination with

certain rules to deal with negative paths produced by the Euler scheme. The authors conclude that the computational efficiency of the Euler scheme with these fixes is superior to the more complicated schemes. For certain relevant parameter configurations the scheme may produce a significant bias for practical sizes of the time steps.

In this chapter we propose a low-bias path simulation scheme for the continuous-time CEV and SABR models, also based on Willard's [108] idea of mixing the conditional distributions [72] of a stochastic volatility model (given the terminal volatility and integrated variance). Our contribution is threefold. First of all, we derive the conditional distribution of the SABR model over a discrete time step and show that, conditioned on the terminal volatility and the integrated variance, it is a space transformed squared Bessel process with a shifted initial condition. Secondly, we propose an efficient easy-to-implement algorithm to simulate the squared Bessel process with an absorbing boundary at zero. Thirdly, we provide a simple approximation formula for the conditional moments of the integrated variance by means of the small disturbance expansion method (see Kunitomo & Takahashi [67] and previous chapter also), which facilitates effective sampling from the joint distribution of the terminal volatility and the integrated variance.

This chapter is organized as follows. In Section 3.2 we describe the basic SABR model and summarize some analytic properties that are relevant for simulation. In Section 3.3, we review some existing discretization schemes, for later comparison. In Section 3.3.3 we present the low-bias discretization scheme to simulate the asset and the variance processes. Section 3.4 discusses the performance of the whole algorithm. Section 3.5 concludes.

3.2 Some analytic features of the SABR model

Given a time interval Δ and an arbitrary set $\mathcal{T}$ of discrete times $s < s+\Delta \ldots < s+N\Delta$ and a stochastic process $X = \{X(t); t > 0\}$, a discretized simulation scheme generates a skeleton $X(s), X(s+\Delta), \ldots$ of a sample path of the stochastic process X. To device such a scheme, we start sampling from the marginal distribution of $X(s+\Delta)$. A repetition of such a one-period scheme may produce the full time-discrete paths for X. Since here we consider the discrete scheme generating paths for a stochastic volatility model, $X(t) = (S(t), \sigma(t))$, for all $t \in \mathcal{T}$, the asset price process $S(t)$ itself is not a Markov process. The fundamental question, as argued by Andersen [3], is how to generate a random sample of $S(s+\Delta)$ from the conditional distribution of $S(s+\Delta)$, given $(S(s), \sigma(s), \sigma(s+\Delta), \int_s^{s+\Delta} \sigma(u)^2 du)$.

In this section, we focus on the analysis of the conditional distribution function; the sampling technique is discussed in the next section. For notational convenience, we do not use the notation s and $s+\Delta$ for the initial and terminal time points, respectively, but use simply 0 for the initial time point, t, s represent the running time and Δ is the terminal time for each discrete time interval.

The Stochastic Alpha Beta Rho Stochastic Volatility (SABR SV) model [46] is given by the following system of stochastic differential equations (SDEs) with

constant parameters α and β:

$$\begin{aligned} dS(t) &= \sigma(t)S(t)^{\beta}dW_1(t), \\ d\sigma(t) &= \alpha\sigma(t)dW_2(t), \end{aligned} \tag{3.2.1}$$

where $dW_1(t)dW_2(t) = \rho dt$. Since the asset price $S(t)$ itself follows a CEV process, one can expect that the conditional SABR process, given $\sigma(\Delta)$ and $\int_0^{\Delta}\sigma(s)^2ds$, is a CEV process as well. The next step will be to 'mix' the conditional CEV process with the joint distribution of $\sigma(\Delta)$ and $\int_0^{\Delta}\sigma(s)^2ds$.

We will show, in the subsequent section, that, conditional on $\sigma(\Delta)$ and $\int_0^{\Delta}\sigma(s)^2ds$, the coordinate transformed asset price process defined by the invertible transformation $X(S) = S^{1-\beta}/(1-\beta)$ is a time-changed Bessel process of dimension $(1-2\beta-\rho^2(1-\beta))/((1-\beta)(1-\rho^2))$, starting at $S(0)^{1-\beta}/(1-\beta)$ given that $\beta \neq 1$. Based on this, the analytic distribution function for $S(\Delta)$ will be derived. After that we show that we can sample this conditional distribution efficiently by a direct inversion scheme.

3.2.1 The distribution of the CEV process

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space generated by $\{W(t)\}$, a one-dimensional Brownian motion. For all $0 \leq t \leq T$, the CEV process can be described by the following stochastic differential equation:

$$dS(t) = \sigma(t)S(t)^{\beta}dW(t), \tag{3.2.2}$$

with initial condition $S(0) = \xi_0$ which we assume to be $\mathcal{F}_0$-measurable.

Here we simply choose $\sigma(t)$ to be a constant, i.e. $\sigma(t) \equiv \sigma$. Following Schroder [100], we consider an invertible transformation, $X(t) = S(t)^{1-\beta}/(1-\beta)$, for $\beta \neq 1$. Application of Itô's lemma gives us the following SDE for $X(t)$, which we recognize as a time-changed Bessel process:

$$\begin{aligned} dX(t) &= (1-\beta)\frac{S(t)^{-\beta}}{1-\beta}\sigma S(t)^{\beta}dW_1(t) - \frac{1}{2}\beta(1-\beta)\frac{S(t)^{-1-\beta}}{1-\beta}\sigma^2 S(t)^{2\beta}dt \\ &= \sigma dW(t) - \frac{\beta\sigma^2}{(2-2\beta)X(t)}dt. \end{aligned} \tag{3.2.3}$$

Then we define a second transformation, $Y(t) = X(t)^2$, which results in a time-changed squared Bessel process of dimension $\delta := (1-2\beta)/(1-\beta)$, that thus satisfies the following SDE:

$$dY(t) = 2\sqrt{|Y(t)|}\sigma dW(t) + \delta\sigma^2 dt. \tag{3.2.4}$$

Let $\nu(t)$ be a time-change function, so that $\nu(t) = \sigma^2 t$. Then, $Y(t) = Z\big(\nu(t)\big)$, where $\{Z(t)\}$ is a δ-dimensional squared Bessel process, i.e., the strong solution of the SDE:

$$dZ(t) = 2\sqrt{|Z(t)|}dW(t) + \delta dt \tag{3.2.5}$$

with degree of freedom, δ. The squared Bessel process is a Markov process and its transition density is known explicitly.

The next step is to sample random numbers from the analytic transition density in Z-space and to apply an inverse variable transformation to obtain the random numbers in the original coordinates.

First, a few technical details need to be discussed. They are presented in the form of results and propositions below.

Result 3.2.1. *[4]: For a standard squared Bessel process, as defined by SDE (3.2.5), the following statements hold true:*

1. *All solutions to SDE (3.2.5) are* non-explosive.
2. *For $\delta < 2$, $Z = 0$ is an attainable boundary for Process (3.2.5).*
3. *For $\delta \geq 2$ SDE (3.2.5) has a unique solution and zero is not attainable.*
4. *For $0 < \delta < 2$ the SDE (3.2.5) does not have a unique solution, unless the boundary condition is specified for the solution at $Z = 0$.*
5. *For $\delta \leq 0$, there is a unique strong solution to the SDE (3.2.5), and boundary condition zero is absorbing.*

Proof. The results have been proved in Appendix A of [4] based on the theory presented in [21]. □

For the latter two cases, the transition densities are known:

Result 3.2.2 (Transition density for squared Bessel process)**.** *The transition density, $q^\delta(t, x, y)$, for the squared Bessel process reads:*

1. *For $\delta \leq 0$ and for $0 < \delta < 2$ in Eq. (3.2.5) but only when the boundary is absorbing at $y = 0$:*

$$q^\delta(t,x,y) = \frac{1}{2t}\Big(\frac{y}{x}\Big)^{\frac{\delta-2}{4}} \exp\Big(-\frac{x+y}{2t}\Big) I_{|\frac{\delta-2}{2}|}\Big(\frac{\sqrt{xy}}{t}\Big), \quad y \geq 0, t > 0. \tag{3.2.6}$$

2. *For $0 < \delta < 2$ when $y = 0$ is a reflecting boundary:*

$$q^\delta(t,x,y) = \frac{1}{2t}\Big(\frac{y}{x}\Big)^{\frac{\delta-2}{4}} \exp\Big(-\frac{x+y}{2t}\Big) I_{\frac{\delta-2}{2}}\Big(\frac{\sqrt{xy}}{t}\Big), \quad y \geq 0, t > 0. \tag{3.2.7}$$

Here we denote by $I_a(x)$ the Bessel function, defined by

$$I_a(x) := \sum_{j=0}^{\infty} \frac{(x/2)^{2j+a}}{j!\Gamma(a+j+1)},$$

and by $\Gamma(x)$ the Gamma function, $\Gamma(x) := \int_0^\infty u^{x-1}e^{-u}du$.

Proof. See Borodin [21] p. 136 for squared Bessel process transition densities. □

By solving a series of inequalities (see the results in Table 3.1), we find essentially three different parameter ranges which determine the behavior of the CEV process at the boundary and the form of the transition densities:

Table 3.1: The mapping of three parameter ranges.

CEV exponent	Squared Bessel δ
$-\infty < \beta < \frac{1}{2}$	$0 < \delta < 2$
$\frac{1}{2} \leq \beta < 1$	$-\infty < \delta \leq 0$
$\beta > 1$	$2 < \delta < \infty$

1. For $\beta > 1$, SDE (3.2.2) has a unique solution and boundary condition zero is not attainable. The density function integrates to unity over $S \in (0, \infty)$ for all $t \geq 0$ and the process $S(t)$ is a strict *local martingale*.
2. For $\beta < \frac{1}{2}$, SDE (3.2.2) does not have a unique solution, unless a separate boundary condition is specified for the boundary behavior at $S = 0$.
   - The density integrates to unity, if the boundary is *reflecting* and process $S(t)$ is a strict *sub-martingale*.
   - The density will not integrate to unity, if the boundary at $S = 0$ is *absorbing* [1] and process $S(t)$ is a *true martingale*.
3. For $\frac{1}{2} \leq \beta < 1$, a unique strong solution to SDE (3.2.2) exists, and boundary value zero is absorbing. The density function does not integrate to unity for $t > 0$ and process $S(t)$ is a *true martingale*.

For most financial applications, parameter β ranges between 0 to 1, which is included in Cases 2 and 3 in the list above. We therefore focus on these two cases, and, correspondingly, on the Items 4 and 5 in Result 3.2.1.

Based on the transition density of the squared Bessel diffusions in X-space given in Result 3.2.2, one can easily obtain the transition density for the CEV process (3.2.2) in S-space. Note first of all that

$$S(\Delta) = \left((1-\beta) \sqrt{|Z\big(\nu(\Delta)\big)|} \right)^{\frac{1}{1-\beta}}, \quad \beta \neq 1.$$

Let us define a map

$$h : s \to \left((1-\beta)\sqrt{s} \right)^{\frac{1}{1-\beta}}, \quad s \geq 0, \quad \beta \neq 1$$

with inverse

$$h^{-1} : y \to \frac{y^{2(1-\beta)}}{(1-\beta)^2}, \quad y \geq 0, \quad \beta \neq 1.$$

So, $S(\Delta) = h(Z\big(\nu(\Delta)\big))$ and $Z(0) = h^{-1}\big(S(0)\big) = S(0)^{2(1-\beta)}/(1-\beta)^2$. Then $Z\big(\nu(\Delta)\big)$ has density $q^\delta(\nu(\Delta), Z(0), y)$ and it follows that the density for $S(\Delta)$ is given by

$$p\big(S|S(0)\big) = q^\delta\big(\nu(\Delta), Z(0), h^{-1}(S)\big) \frac{dh^{-1}(S)}{dS},$$

[1] There is a degenerate part with an atom in the boundary and an absolutely continuous part over $(0, \infty)$.

where we use $p\big(S|S(0)\big)$ to denote the conditional transition density for the CEV process. By combining the two cases considered in Result 3.2.2, the related transition densities for the CEV process, $S(t)$, in Eq. (3.2.2) are of the following form:

1. For $0 < \beta < \frac{1}{2}$ with absorption at zero and for $\frac{1}{2} \leq \beta < 1$:

$$p\big(S(\Delta)|S_0\big) = \frac{1}{\nu(\Delta)}\Big(\frac{S(\Delta)}{S(0)}\Big)^{-\frac{1}{2}} \exp\Big(-\frac{S(\Delta)^{2(1-\beta)} + S(0)^{2(1-\beta)}}{2(1-\beta)^2\nu(\Delta)}\Big) \cdot I_{|\frac{\delta-2}{2}|}\Big(\frac{\big(S(0)S(\Delta)\big)^{1-\beta}}{\nu(\Delta)(1-\beta)^2}\Big)\frac{S(\Delta)^{1-2\beta}}{1-\beta},$$

   where $\nu(\Delta) = \sigma^2\Delta$ and $\delta = \frac{1-2\beta}{1-\beta}$.

2. For $0 < \beta < \frac{1}{2}$ with a reflecting boundary at $S = 0$:

$$p\big(S(\Delta)|S(0)\big) = \frac{1}{\nu(\Delta)}\Big(\frac{S(\Delta)}{S(0)}\Big)^{-\frac{1}{2}} \exp\Big(-\frac{S(\Delta)^{2(1-\beta)} + S(0)^{2(1-\beta)}}{2(1-\beta)^2\nu(\Delta)}\Big) \cdot I_{\frac{\delta-2}{2}}\Big(\frac{\big(S(0)S(\Delta)\big)^{1-\beta}}{\nu(\Delta)(1-\beta)^2}\Big)\frac{S(\Delta)^{1-2\beta}}{1-\beta}.$$

By integrating these identities, we find the cumulative distribution functions:

Result 3.2.3. *The cumulative distribution function of the CEV price process as in Eq. (3.2.2) is given by the following formulas:*

1. *For $0 < \beta < \frac{1}{2}$ with absorption at zero and for $\frac{1}{2} \leq \beta < 1$:*

$$\Pr\big(S(\Delta) \leq x|S(0)\big) = 1 - \chi^2(a; b, c). \tag{3.2.8}$$

2. *For $0 < \beta < \frac{1}{2}$ with a reflecting boundary at $S = 0$:*

$$\Pr\big(S(\Delta) \leq x|S(0)\big) = \chi^2(c; 2-b, a), \tag{3.2.9}$$

with the following parameters:

$$a = \frac{S(0)^{2(1-\beta)}}{(1-\beta)^2\nu(\Delta)}, \quad b = \frac{1}{1-\beta}, \quad c = \frac{x^{2(1-\beta)}}{(1-\beta)^2\nu(\Delta)}, \quad \nu(\Delta) = \sigma^2\Delta,$$

and $\chi^2(x; \delta, \lambda)$ is the noncentral chi-square cumulative distribution function for random variable x with non-centrality parameter λ and degree of freedom δ.

Proof. The proofs of these results can be found in Schroder [100] using classic results for Bessel processes. An alternative proof based on Green function theory can be found in Lesniewski [71]. □

As mentioned before, the density will not integrate to unity when the boundary is absorbing. The shortage in the total probability mass is the probability absorbed at $S = 0$. Following the result of Result 3.2.3, a formula for the absorption probability can be obtained:

Corollary 3.2.4. *For $0 < \beta < 1$, the probability of $S(\Delta)$, given by SDE (3.2.2) and initial condition $S(0)$, reads*

$$Pr\big(S(\Delta) = 0|S(0)\big) = 1 - \gamma\Big(\frac{1}{2(1-\beta)}, \frac{S(0)^{2(1-\beta)}}{2(1-\beta)^2\nu(\Delta)}\Big) \Big/ \Gamma\Big(\frac{1}{2(1-\beta)}\Big), \tag{3.2.10}$$

where $\gamma(\alpha, \beta)$ is the lower incomplete Gamma function and $\Gamma(\alpha)$ is the Gamma function.

Proof. Choosing x in Eq. (3.2.8) to be zero, we find that

$$\begin{aligned}\Pr\big(S(\Delta) = 0|S(0)\big) =& 1 - \chi^2\left(\frac{S(0)^{2(1-\beta)}}{(1-\beta)^2\nu(\Delta)}; \frac{1}{1-\beta}, 0\right) \\ =& 1 - \text{Chi}^2\left(\frac{S(0)^{2(1-\beta)}}{(1-\beta)^2\nu(\Delta)}; \frac{1}{1-\beta}\right),\end{aligned} \tag{3.2.11}$$

where the last equality sign is because the noncentral chi-square distribution with a zero non-centrality parameter reduces to a chi-square distribution. A chi-square distribution has an explicit cumulative function in terms of gamma functions:

$$\text{Chi}^2(x; \delta) = \frac{\gamma(\delta/2, x/2)}{\Gamma(\delta/2)}, \tag{3.2.12}$$

where $\gamma(\alpha, \beta)$ is again the lower incomplete Gamma function. We substitute Eq. (3.2.12) into Eq. (3.2.11) and prove the claim.

□

In interest rate derivative pricing, initial value, $S(0)$, is often very small, hence it is likely for $S(t)$ to reach zero. The specification of the boundary condition for process $S(t)$ at zero may have a significant impact on the distribution, as is evident from the distributions in Figure 3.1. The two plots show a comparison of the exact cumulative distribution of the CEV process versus the log-normal and normal distributions at $T = 0.25$ for $S(\Delta)$, given two different levels of $S(0)$, i.e. $S(0) = 6\%$ in left-hand plot and $S(0) = 2\%$ in the right-hand plot. The model parameters chosen were $\sigma = 0.3$ and $\beta = 0.4$. We have included the normal and log-normal distributions by matching the first two moments of $S(\Delta)$. The labels 'absorbing' and 'reflecting' distinguish the distributions of the CEV process with an absorption and a reflection boundary condition, respectively. The two matched distributions do not represent accurate approximations for the true distributions of $S(\Delta)$.

Andersen & Andreasen [4] and Rebonato [94], pp.48, argue that if the asset price follows a CEV process under a certain measure, there is only one acceptable boundary condition at zero to ensure the arbitrage-free conditions, which is the absorption condition. If there were a reflecting boundary at zero, then for an initially worthless portfolio, one would take a long position in the asset once the price zero is reached (which would happen with a strictly positive probability) and sell it immediately when the 0 boundary has reflected the price process to obtain risk-less profit.

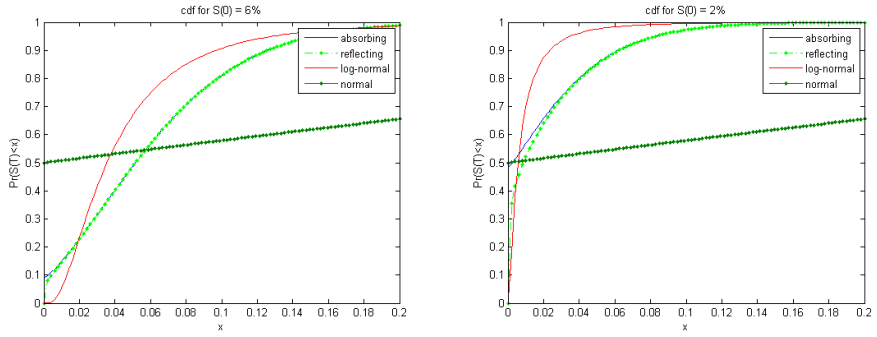

Figure 3.1: Comparison of the exact cumulative distribution of the CEV process versus the log-normal and normal distributions at $T = 0.25$ for $S(\Delta)$ given different levels of $S(0)$.

We therefore will assume that the boundary is absorbing at $S(t) = 0$ in the following sections and develop a method to sample from distribution function (3.2.8). The strategy of sampling will be discussed in full detail in the sections to follow.

3.2.2 SABR conditional distribution

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \Pr)$ be a filtered probability space generated by two Brownian motions $\{U(t), W_2(t)\}$. We denote the probability space as the product of two filtered probability spaces generated by two independent Brownian motions, i.e. $\Omega = \Omega_1 \times \Omega_2, \mathcal{F} = \mathcal{F}^1 \times \mathcal{F}^2, \{\mathcal{F}_t\} = \{\mathcal{F}_t^1 \times \mathcal{F}_t^2\}, \Pr = \Pr_1 \times \Pr_2$.

Based on the closed-form distribution function of the CEV process, Islah [53] shows that, conditional on the levels of $\sigma(\Delta)$ and $\int_0^\Delta \sigma(s)^2 ds$, the transformed asset price process $S(t)^{2-2\beta}/(1-\beta)^2$ is a shifted squared Bessel process with initial condition $\big(S(0)^{1-\beta}/(1-\beta) + \frac{\rho}{\alpha}\big(\sigma(\Delta) - \sigma(0)\big)\big)^2$. Here we recall the following results from [53]:

Result 3.2.5 (SABR Conditional Distribution [53])**.** *In the context of SABR model (3.2.1), and conditional on the level of terminal volatility, $\sigma(\Delta)$, and integrated variance, $\int_0^\Delta \sigma(s)^2 ds$, let $\omega_2 \in \Omega_2$, $t \to \sigma(t, \omega_2)$ be a volatility path. The followings statements hold:*

1. *For an invertible variable transformation, $X(S) = S^{1-\beta}/(1-\beta)$, application of Itô's lemma gives us*

$$\begin{aligned} X(\Delta) =& X(0) + \frac{\rho}{\alpha}\big(\sigma(\Delta) - \sigma(0)\big) + \sqrt{1-\rho^2}\int_0^\Delta \sigma(s) dU(s) \\ &- \int_0^\Delta \frac{\beta \sigma(s)^2}{(2-2\beta)X(s)} ds, \end{aligned} \tag{3.2.13}$$

where $U(s)$ is a standard Brownian motion, independent of $W_2(s)$ in System (3.2.1).

2. *With* $Y(S) = S^{2-2\beta}/(1-\beta)^2$ *and application of a time-change,* $\nu(t) = (1-\rho^2)\int_0^t \sigma(s)^2 ds$, $Y(t)$ *is a squared Bessel process of dimension* $\frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}$ *solving the SDE:*

$$dY\big(\nu(t)\big) = 2\sqrt{|Y\big(\nu(t)\big)|}dU\big(\nu(t)\big) + \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}d\nu(t), \quad (3.2.14)$$

with initial condition $Y(0) = \big(\frac{S(0)^{1-\beta}}{1-\beta} + \frac{\rho}{\alpha}\big(\sigma(\Delta)-\sigma(0)\big)\big)^2$.

3. *Let* τ *be the stopping time for which process* Y *hits zero, i.e.*

$$\tau = \inf\{\nu(s)|Y\big(\nu(s)\big) = 0\},$$

the 'stopped' process Y *reads*

$$\begin{aligned} Y\big(\nu(t)\wedge\tau\big) =& Y(0) + 2\int_0^{\nu(t)\wedge\tau} \sqrt{|Y\big(\nu(s)\big)|}dU\big(\nu(s)\big) \\ &+ \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}\left(\nu(t)\wedge\tau\right). \end{aligned} \quad (3.2.15)$$

Proof. In order to derive the conditional dynamics of the SABR model in Eq. (3.2.1), we first integrate the SDE for the volatility $\sigma(t)$:

$$\int_0^\Delta d\sigma(s) = \alpha\int_0^\Delta \sigma(s)dW_2(s) \quad \Rightarrow \int_0^\Delta \sigma(s)dW_2(s) = \frac{1}{\alpha}\big(\sigma(\Delta)-\sigma(0)\big). \quad (3.2.16)$$

When conditioning on the volatility level, $\sigma(\Delta)$, the above identity becomes a constant. It plays an important role in the following derivation.

Based on arguments in Section 3.2.1, the application of Itô's lemma to $X = S^{1-\beta}/(1-\beta)$ results in

$$X(\Delta) = X(0) + \int_0^\Delta \sigma(s)dW_1(s) - \int_0^\Delta \frac{\beta\sigma(s)^2}{(2-2\beta)X(s)}ds. \quad (3.2.17)$$

We now employ the Cholesky decomposition of the two correlated Brownian motions,

$$dW_1(t) = \rho dW_2(t) + \sqrt{1-\rho^2}dU(t), \;\; dW_2(t) = dW_2(t).$$

After substitution of the Cholesky decomposition into Eq. (3.2.17), we arrive at

$$X(\Delta) = X(0) + \frac{\rho}{\alpha}\big(\sigma(\Delta)-\sigma(0)\big) + \sqrt{1-\rho^2}\int_0^\Delta \sigma(s)dU(s) - \int_0^\Delta \frac{\beta\sigma(s)^2}{(2-2\beta)X(s)}ds, \quad (3.2.18)$$

where we used Identity (3.2.16).

In order to prove the second claim in Result 3.2.5, we write Eq. (3.2.18) as,

$$\begin{aligned} d\tilde{X}(t) &= \sqrt{1-\rho^2}\sigma(t)dU(t) - \frac{\beta\sigma(t)^2}{(2-2\beta)\tilde{X}(t)}dt, \quad \text{with:} \\ \tilde{X}(0) &= X(0) + \frac{\rho}{\alpha}\big(\sigma(\Delta)-\sigma(0)\big), \end{aligned}$$

where the notation $\tilde{X}(t)$ denotes process $X(t)$ with a shifted initial condition. Despite the difference in notation, $\tilde{X}(t)$ and $X(t)$ represent the same process.

We define the variable transformation $Y = \tilde{X}^2$, which, after applying Itô's lemma, gives

$$\begin{aligned} dY(t) =& 2\tilde{X}(t)d\tilde{X}(t) + d\tilde{X}(t)^2 \\ =& 2\sqrt{Y(t)}\sqrt{1-\rho^2}\sigma(t)dU(t) + \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}(1-\rho^2)\sigma(t)^2dt. \end{aligned} \tag{3.2.19}$$

Due to the independence of the Brownian motions $U(t)$ and $W_2(t)$, the integral $\int_0^t \sigma(s)dU(s)$ is a Gaussian distribution with mean zero and variance $\int_0^t \sigma(s)^2ds$.

We now consider the time-change $\nu(t) = (1-\rho^2)\int_0^t \sigma(s)^2ds$. A Brownian motion under this 'clock' will have the same distribution as $\sqrt{1-\rho^2}\int_0^t \sigma(s)dU(s)$, i.e.

$$U\big(\nu(t)\big) = \int_0^{\nu(t)} dU(s) = \sqrt{1-\rho^2}\int_0^t \sigma(s)dU(s).$$

We substitute the time-changed Brownian motion into Eq. (3.2.19), which gives us

$$dY\big(\nu(t)\big) = 2\sqrt{|Y\big(\nu(t)\big)|}dU\big(\nu(t)\big) + \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}d\nu(t),$$

i.e. a time-changed squared Bessel process of dimension $\delta = \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}$, starting at

$$Y(0) = \tilde{X}(0)^2 = \Big(X(0) + \frac{\rho}{\alpha}\big(\sigma(\Delta) - \sigma(0)\big)\Big)^2.$$

Inclusion of a stopping time, $\tau = \inf\{\nu(s)|Y\big(\nu(s)\big) = 0\}$, to the second result will prove the third claim. □

Proposition 1 (Cumulative Distribution for Conditional SABR Process)**.** *For some $S(0)$, strictly larger than 0, the conditional cumulative distribution of $S(\Delta)$ with an absorbing boundary at $S(t) = 0$ given $\sigma(\Delta)$ and $\int_0^\Delta \sigma(s)^2ds$ reads*

$$\Pr\left(S(\Delta) \le K|S(0) > 0, \sigma(\Delta), \int_0^\Delta \sigma(s)^2ds\right) = 1 - \chi^2(a; b, c), \tag{3.2.20}$$

where

$$\begin{aligned} a &= \frac{1}{\nu(\Delta)}\Big(\frac{S(0)^{1-\beta}}{(1-\beta)} + \frac{\rho}{\alpha}\big(\sigma(\Delta) - \sigma(0)\big)\Big)^2, \quad b = 2 - \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}, \\ c &= \frac{K^{2(1-\beta)}}{(1-\beta)^2\nu(\Delta)}, \qquad\qquad \nu(\Delta) = (1-\rho^2)\int_0^\Delta \sigma(s)^2ds. \end{aligned} \tag{3.2.21}$$

$\chi^2(x; \delta, \lambda)$ is again the noncentral chi-square cumulative distribution function.

Proof. Given that $Y\big(\nu(t)\big)$ is a time-changed Bessel process, we substitute parameter δ and the non-centrality parameter $Y(0)$ into the distribution function presented in Result 3.2.3. □

Note that the condition $S(0) > 0$ is crucial, because the paths that reach zero should stay in zero, due to the stopping time τ defined in Result 3.2.5.

Remark. It was argued by Andersen [7] that the continuous time process, $S(t)$, will be a martingale, i.e.

$$\mathbb{E}[S(t+\Delta)|S(t)] = S(t) < \infty.$$

The equivalent discrete-time process, $\hat{S}(t)$, generated by the low-bias simulation scheme may not satisfy the martingale condition, i.e.

$$\mathbb{E}[\hat{S}(t+\Delta)|\hat{S}(t)] \neq \hat{S}(t).$$

The net drift, away from the martingale, is visible for parameter sets with small β *and* close-to-zero rates. However, this combination of parameters does not appear in practical applications as it gives rise impractical implied volatility levels. For practical SABR parameters, the martingale bias is very small and can be controlled by reducing the size of the time step.

3.3 The discretization scheme for the SABR model

In this section we will present the low-bias simulation scheme for the SABR model. Before that, we review some existing path simulation schemes for the stochastic volatility model. We will denote time discrete approximations to $S(t)$ and $\sigma(t)$ by $\hat{S}(t)$ and $\hat{\sigma}(t)$, respectively.

3.3.1 Taylor based time discrete approximation schemes

The basic first-order Taylor approximation scheme for (3.2.1) takes the following form:

$$\begin{array}{ll} \hat{S}(\Delta) & = \hat{S}(0) + \hat{\sigma}(0)\hat{S}(0)^{\beta} Z_1 \sqrt{\Delta}, \\ \hat{\sigma}(\Delta) & = \hat{\sigma}(0) \exp\big(-\frac{1}{2}\alpha^2\Delta + \alpha Z_2 \sqrt{\Delta}\big), \end{array} \tag{3.3.1}$$

with Z_1 and Z_2 two correlated standard normal random variables, i.e. $dZ_1 dZ_2 = \rho dt$.

This Euler scheme represents an $O(\sqrt{\Delta})$-accurate Taylor approximation for the asset price process, $S(t)$. To reduce the bias introduced by the first-order approximation, Elerian [35] suggests using a transition density derived from the scheme due to Milstein [80]. Including higher-order expansion terms gives us the Milstein scheme:

$$\begin{array}{ll} \hat{S}(\Delta) & = \hat{S}(0) + \hat{\sigma}(0)\hat{S}(0)^{\beta} Z_1 \sqrt{\Delta} + \frac{1}{2}\beta\hat{\sigma}(0)^2\hat{S}(0)^{2\beta-1}(Z_1^2\Delta - \Delta), \\ \hat{\sigma}(\Delta) & = \hat{\sigma}(0) \exp\big(-\frac{1}{2}\alpha^2\Delta + \alpha Z_2 \sqrt{\Delta}\big). \end{array} \tag{3.3.2}$$

A common problem shared by the Taylor-based approximation schemes is the possibility of generating negative asset prices, $\hat{S}(t)$, that give rise to financially meaningless solutions. A remedy for the negative prices is to transform the SDE to logarithmic coordinates using Itô's lemma [6]. For the CEV asset price process the log-Euler scheme is defined as:

$$\begin{array}{ll} \hat{S}(\Delta) & = \hat{S}(0) \exp\big(-\frac{1}{2}\hat{\sigma}(0)^2 S(0)^{2\beta-2}\Delta + \hat{\sigma}(0)\hat{S}(0)^{\beta-1} Z_1 \sqrt{\Delta}\big), \\ \hat{\sigma}(\Delta) & = \hat{\sigma}(0) \exp\big(-\frac{1}{2}\alpha^2\Delta + \alpha Z_2 \sqrt{\Delta}\big). \end{array} \tag{3.3.3}$$

Whereas the log-Euler scheme preserves positivity of the asset price process, numerical experiments show that the scheme may become unstable for specific time-step sizes [6]. The instabilities occur because the diffusion terms in Eq. (3.3.3) approach infinity quickly, when $\hat{S}(t)$ reaches zero (see some results in Table 3.2).

Table 3.2: Percentages of Taylor-based simulation experiments with failure for the CEV model, with different step-sizes over a 5 years interval. The parameters are $\beta = 0.3$, $\sigma = 0.3$ and $S(0) = 4\%$.

Issue	Stepsize	Euler	Milstein	Log-Euler
Negativity	$\Delta = 0.5$	85%	49%	0%
	$\Delta = 0.25$	88%	56%	0%
	$\Delta = 0.125$	94%	59%	0%
Infinity	$\Delta = 0.5$	0%	34%	96%
	$\Delta = 0.25$	0%	31%	96%
	$\Delta = 0.125$	0%	37%	95%

3.3.2 Exact scheme of Broadie and Kaya

Broadie and Kaya [26] proposed the BK scheme, an exact simulation *for the Heston model*. Although the Heston dynamics are different from the SABR dynamics, the exact simulation concept serves as the basis to construct the discrete approximation schemes for the SABR model here.

The BK scheme is based on sampling $\sigma(\Delta)$ from its distribution function. Given $\sigma(\Delta)$ (and $\sigma(0)$), a sample from $\int_0^\Delta \sigma(s)^2 ds$ is drawn. Conditional on $\sigma(\Delta)$ and $\int_0^\Delta \sigma(s)^2 ds$, process $\ln S(\Delta)$ is from a Gaussian distribution in the Heston model.

In some more detail, with $V(t) = \sigma(t)^2$ the Heston stochastic volatility [48] process can be described as:

$$\begin{aligned} dS(t) &= \sqrt{V(t)} S(t) dW_1(t), \\ dV(t) &= \kappa\big(\theta - V(t)\big)dt + \eta\sqrt{V(t)} dW_2(t) \end{aligned} \tag{3.3.4}$$

with correlation $dW_1(t)dW_2(t) = \rho dt$, κ as speed of mean reversion, long run mean θ and volatility of volatility η. To obtain a bias-free scheme, first the SDE for the volatility is integrated, i.e.,

$$\int_0^\Delta \sqrt{V(s)} dW_2(s) = \frac{1}{\eta}\big(V(\Delta) - V(0) - \kappa\theta\Delta + \kappa \int_0^\Delta V(s) ds\big). \tag{3.3.5}$$

Application of the Cholesky decomposition, i.e.,

$$dW_1(t) = \rho dW_2(t) + \sqrt{1-\rho^2} dU(t), \quad dW_2(t) = dW_2(t),$$

gives us for $\ln S(t)$:

$$d\ln S(t) = -\frac{1}{2}V(t)dt + \rho\sqrt{V(t)} dW_2(t) + \sqrt{1-\rho^2}\sqrt{V(t)} dU(t),$$

with Brownian motion $U(t)$, independent of $W_2(t)$. In integral form, we then obtain

$$\begin{aligned}\ln S(\Delta) =& \ln S(0) - \frac{1}{2}\int_0^\Delta V(s)ds + \rho\int_0^\Delta \sqrt{V(s)}dW_2(s) \\ &+ \sqrt{1-\rho^2}\int_0^\Delta \sqrt{V(s)}dU(s) \\ =& \ln S(0) + \frac{\rho}{\eta}\big(V(\Delta) - V(0) - \kappa\theta\Delta\big) + \big(\frac{\rho\kappa}{\eta} - \frac{1}{2}\big)\int_0^\Delta V(s)ds \\ &+ \sqrt{1-\rho^2}\int_0^\Delta \sqrt{V(s)}dU(s). \end{aligned} \tag{3.3.6}$$

Due to the independence of $U(t)$ and $W_2(t)$, the Itô integral $\int_0^\Delta V(s)dU(s)$, conditional on the realized variance, is Gaussian with mean zero and variance $\int_0^\Delta V(s)ds$. It is easy to see from Eq. (3.3.6) that $\ln S(\Delta)$ is normally distributed, conditional on $V(\Delta)$ and $\int_0^\Delta V(s)ds$. By aggregation of all conditional Gaussian distributed samples, we obtain the desired distribution for the stochastic volatility model.

For the Heston model, the distribution of $V(\Delta)$ is known in closed form, but the conditional distribution of $\int_0^\Delta V(s)ds$ is not known explicitly. Broadie and Kaya [26] derive the characteristic function, which is based on two modified Bessel functions that contain infinite series expansions. A numerical Fourier-inversion step is then necessary to generate the desired conditional cumulative distribution function. The evaluation of the characteristic function for $\int_0^\Delta V(s)ds$ as well as the required Fourier inversions require significant computational effort. The implementation of these steps has to be done with great care to avoid bias from the numerical inversion.

3.3.3 A low-bias scheme for SABR simulation

By utilizing the analytic results for the conditional SABR process presented in Proposition 1, and the above mentioned technique of 'mixing conditional distributions', we present a low-bias discretization scheme for the SABR model. We start to simulate the conditional SABR process, which is a space-transformed squared Bessel process with an absorbing boundary at zero, and then mix the conditional process by the joint dynamics of the terminal volatility and integrated variance.

Despite the fact that the volatility, $\sigma(\Delta)$, can easily be sampled from a log-normal distribution, it is not straightforward to sample the integrated variance. It is also challenging to simulate a CEV process exactly. Hence, we devote the following two subsections to these issues.

Sampling the conditional CEV process

As discussed in Section 3.2.2, the CEV process is a space-transformed squared Bessel process, whose distribution function is known in closed-form as the non-central chi-square distribution. For initial asset prices far away from zero, i.e. $S(0) >> 0$, the probability of hitting zero is almost zero, i.e.

$$\Pr\left(\inf\{t|S(t)=0\} < \Delta\right) \to 0,$$

and the distribution function approaches an ordinary noncentral chi-square distribution:

$$\begin{aligned}\Pr\big(S(\Delta) \leq K\big) &= 1 - \chi^2(a; b, c) = \chi^2(c; 2-b, a) + \Pr\left(\inf\{t|S(t)=0\} < \Delta\right) \\ &\approx \chi^2(c; 2-b, a). \end{aligned} \tag{3.3.7}$$

where a, b and c are as defined in Proposition 1.

It is well-known (see Johnson et al. [58], pp. 450) that the noncentral chi-square distribution approaches a Gaussian distribution as the non-centrality parameter goes to infinity. In [3] it was stated that the noncentral chi-square distribution with a sufficiently large non-centrality parameter can be accurately approximated by a quadratic function of Gaussian variables.

Andersen's Quadratic Exponential (QE) scheme does not perform satisfactory for small values of $S(0)$, as the moment-matching method then becomes inaccurate. This is due to the fact that for small values of $S(0)$, the probability of reaching zero is high, so that the approximation in Eq. (3.3.7) does not hold and the analytic moments for the distribution $1 - \chi^2(a; b, c)$ are not known. For these small values, we therefore propose to use a Newton-type root finding method to invert the distribution function (3.2.20) directly.

More specifically, we determine a value, c^*, which solves the equation $1 - \chi^2(a; b, c^*) - U = 0$ with a high accuracy. To compute the noncentral chi-square distribution, we use Schroder's [100] recurrence formula, in which the evaluation of an infinite double sum of gamma functions is required. This computation takes only a small fraction of the costs for the computation of the original series [100].

The partial derivative of the cumulative distribution is the probability density function (PDF) which is known analytically as the transition density for a squared Bessel process, given in Eq. (3.2.6). Using this derivative information gives a substantial enhancement of the computational performance[2].

Moment-matched quadratic Gaussian approximation

The mean and the variance of the noncentral chi-square distribution, $\chi^2(x; k, \lambda)$, are $k + \lambda$ and $2(k + 2\lambda)$, respectively. Here we determine, as in Proposition 5 in [3], the values of the relevant parameters by moment matching.

Result 3.3.1 (Moment-Matched Quadratic Gaussian Approximation [3])**.** *We denote the mean and variance of a noncentral chi-square distribution, $\chi^2(x; k, \lambda)$, by $m := k + \lambda$ and $s^2 := 2(k + 2\lambda)$, and we define $\psi := s^2/m^2$. With*

$$e^2 = 2\psi^{-1} - 1 + \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1} - 1} \geq 0 \quad \textit{and} \quad d = \frac{m}{1 + e^2}.$$

A non-central chi-square distributed random variable Y is accurately approximated by

$$Y = d\,(e + Z)^2 \qquad Z \sim N(0, 1),$$

where $\mathbb{E}[Y] = m$ and $\mathbb{V}\mathrm{ar}[Y] = s^2$.

[2] When the gradient information is not supplied, the algorithm requires many extra function evaluations of the noncentral chi-square distribution function, which is generally expensive to compute.

In our case, we have

$$k = \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)},\ \lambda = \frac{1}{\nu(\Delta)}\Big(\frac{S(0)^{1-\beta}}{(1-\beta)} + \frac{\rho}{\alpha}\big(\sigma(\Delta)-\sigma(0)\big)\Big)^2,$$

with $\nu(\Delta) = (1-\rho^2)\int_0^\Delta \sigma(s)^2 ds$. We then compute $S(\Delta)^{2(1-\beta)}/(1-\beta)^2\nu(\Delta)$ by the quadratic normal approximation:

$$\frac{S(\Delta)^{2(1-\beta)}}{(1-\beta)^2\nu(\Delta)} = d(e+Z)^2 \quad \Rightarrow \quad S(\Delta) = \big((1-\beta)^2\nu(\Delta)\cdot d(e+Z)^2\big)^{\frac{1}{2(1-\beta)}}, \tag{3.3.8}$$

with $Z \sim N(0,1)$. The constants d and e are as in Result 3.3.1.

Direct inversion scheme

As mentioned earlier, the quadratic Gaussian approximation is accurate only if $S(0)$ is sufficiently large, or, equivalently, when the probability of absorption is small. For small values of $S(0)$, we invert the distribution function (3.2.20) directly by a Newton-type method. As in Andersen [3], variable $\psi = s^2/m^2$, defined in Result 3.3.1, is used as the threshold level to determine which algorithm (either the moment-matched Quadratic Gaussian or the direct inversion) is to be employed.

We first assume that the integrated variance, $\nu(\Delta) = (1-\rho^2)\int_0^\Delta \sigma(s)^2 ds$, has been determined. An algorithm to sample the conditional CEV process in terms of the transformed squared Bessel process with an absorbing boundary at zero, starting at time 0 until Δ, then reads:

1. Compute $a = \frac{1}{\nu(\Delta)}\Big(\frac{S(0)^{1-\beta}}{(1-\beta)} + \frac{\rho}{\alpha}\big(\sigma(\Delta)-\sigma(0)\big)\Big)^2$ and $b = 2 - \frac{1-2\beta-\rho^2(1-\beta)}{(1-\beta)(1-\rho^2)}$ by Result 1;
2. Draw a vector of uniform random numbers, U;
3. Compute the absorption probability $\Pr\big(S(\Delta)=0|S(0)\big)$ by Eq. (3.2.10);
   (a) **If** $S(0)=0$: $S(\Delta)=0$ and return;
   (b) **Else if** $U < \Pr\big(S(\Delta)=0|S(0)\big)$: $S(\Delta)=0$ and return;
   (c) **Otherwise**: Go to the next step;
4. Compute $\psi := s^2/m^2$ with $m := k+\lambda$ and $s^2 := 2(k+2\lambda)$;
5. Select a threshold value, $\psi^{\text{thres}} \in [1,2]$, as in [3]. Here we set $\psi^{\text{thres}} = 2$ for numerical efficiency;
6. **If** $\{0 < \psi \le \psi^{\text{thres}}\}\bigcap\{m \ge 0\}$: We sample $S(\Delta)$ by Eq. (3.3.8);
7. **Otherwise** if $\psi > \psi^{\text{thres}}$ or $\{m<0\}\bigcap\{0<\psi<=\psi^{\text{thres}}\}$: We determine the root c^* of the equation $H(a,b,c) := 1-\chi^2(a;b,c) - U = 0$ with initial guess $c_0 = a$, and repeat the Newton method until the prescribed tolerance ϵ is reached:

$$c_{n+1} = c_n - \frac{H(a,b,c_n)}{\overline{q}(a,b,c_n)},$$

where

$$\overline{q}(a,b,c) = \frac{1}{2}\left(\frac{c}{a}\right)^{\frac{b-2}{4}} \exp\left(-\frac{a+c}{2}\right) I_{|\frac{b-2}{2}|}\left(\sqrt{ac}\right).$$

The desired random number, c^*, is from a squared Bessel distribution with an absorbing boundary at zero.

8. We apply the inverse coordinate transform to recover the random numbers in asset price (or physical) space:

$$S(\Delta) = \left(c^*(1-\beta)^2\nu(\Delta)\right)^{\frac{1}{2-2\beta}}.$$

This root finding method consists of only basic operations, so that the whole procedure can be vectorized and a vector of uniform random numbers can be processed simultaneously.

Enhanced direct inversion procedure

Although the above method is accurate, it appears to be rather slow, due to a significant number of evaluations of the (expensive) noncentral chi-square CDF. In order to speed up the procedure, we determine an *accurate initial solution* by a cheap numerical procedure. With an accurate initial solution, this root finding procedure will converge in only a few iterations.

Yuan [111] gives numerical evidence for the fact that the noncentral chi-square distribution 'converges' to a normal distribution when non-centrality parameter c or random number a increase in value. A number of normal approximations to the noncentral chi-square distribution have been developed, see Johnson and Kotz [59] for a review. A particular accurate approximation is derived by Sankaran [97]:

$$\chi^2(a;b,c) \sim N\left(\frac{1-hp(1-h+\frac{1}{2}(2-h)mp) - \left(\frac{a}{b+c}\right)^h}{h\sqrt{2p(1+mp)}}\right) \tag{3.3.9}$$

where N denotes the normal cumulative density function and

$$\begin{aligned} h &= 1 - \tfrac{2(b+c)(b+3c)}{3(b+2c)^2}, \\ p &= \tfrac{b+2c}{(b+c)^2}, \\ m &= (h-1)(1-3h). \end{aligned} \tag{3.3.10}$$

This approximation consists of basic functions only, e.g. square roots, powers and normal distribution functions, that can be executed quickly on modern hardware. The approximation is sufficiently accurate for a wide range of parameters, but it varies across different sets of parameters. Especially for small values of parameter a in Eq. (3.2.20), this approximation is less accurate.

To illustrate the performance of Sankaran's approximation, we consider two different test cases with different values for parameter a (keeping the other parameters the same). Figure 3.2 presents the results. In both plots, the difference between the exact curve and Sankaran's approximation is visible, but for the larger value of a the discrepancy is substantially smaller. For the purpose of generating an initial solution for the second stage of the root finding method, however, the approximation quality is fully satisfactory.

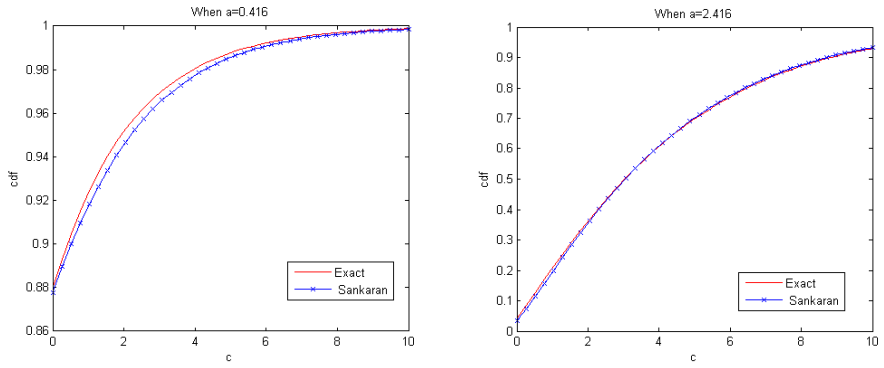

Figure 3.2: Comparison of the quality of Sankaran's approximation for two sets of parameters. Degree of freedom parameter b is set to 0.423 for both cases, but parameter a is lower in the left-side plot, $a = 0.416$ than in the right-side plot, $a = 2.416$.

After the inversion of a vector of uniform variables, U, into normal variables, $X = N^{-1}(U)$, we find a vector of roots, c, which solve the following equation

$$G(a,b,c) := \frac{1 - hp(1 - h + \frac{1}{2}(2-h)mp) - \left(\frac{a}{b+c}\right)^h}{h\sqrt{2p(1+mp)}} - X = 0, \qquad (3.3.11)$$

with h, p and m as defined in Eq. (3.3.10). This can be performed very efficiently by Newton's method with analytic derivative information.

In the next stage, we choose the c_0 in Step 7 of the *direct inversion scheme* to be the solution of Eq. (3.3.11) and execute that step. The result is a significant improvement in the number of function evaluations required for Newton's root finding algorithm to converge.

3.3.4 The integrated variance

The sampling of the conditionally integrated variance process may be an expensive procedure. There are several methods in the literature to *approximate* the integrated variance process, $\int_0^\Delta V(s)ds|V(0), V(\Delta)$, or, equivalently, $\int_0^\Delta \sigma(s)^2 ds|\sigma(0), \sigma(\Delta)$. Notable examples include the Fourier inversion technique in [26] and the drift interpolation in [3]. In the present chapter, we propose to use an approximation based on a *moment-matched drift interpolation technique*. Instead of dealing with the distribution of the quantity $\int_0^\Delta \sigma(s)^2 ds|\sigma(0), \sigma(\Delta)$, we match the first two conditional moments of $\int_0^\Delta \sigma(s)^2 ds$ given $\sigma(0)$ (and $\sigma(\Delta)$) in a log-normal distribution, and sample the distribution of $\int_0^\Delta \sigma(s)^2 ds$ from the moment-matched log-normal distribution.

Small disturbance expansion

In a first step, we derive the conditional moments of $\int_0^\Delta \sigma(s)^2 ds$, given $\sigma(\Delta)$ (and $\sigma(0)$), by the *small disturbance expansion*, proposed by Kunitomo [67]. The small disturbance expansion method is closely related to the stochastic Taylor

expansion, which is especially accurate when the quantity $\alpha^2\Delta$ is small. In order to apply the small disturbance expansion technique, we first reformulate the log-normal volatility process by introducing a small parameter $0 < \epsilon << 1$ in the diffusion coefficient, so that $\alpha = \epsilon\tilde{\alpha}$:

$$\sigma^{(\epsilon)}(t) = \sigma(0) + \epsilon\tilde{\alpha}\int_0^t \sigma^{(\epsilon)}(s)dW_2(s). \tag{3.3.12}$$

We construct an expansion of $\sigma^{(\epsilon)}(t)$ around $\sigma(0)$ by $\epsilon \to 0$.

Proposition 2. *For stochastic integral (3.3.12), we have the formal small disturbance expansion*

$$\begin{aligned}\sigma(t)^{(\epsilon)} =& \sigma(0) + \epsilon\frac{\partial\sigma(t)^{(\epsilon)}}{\partial\epsilon}\Big|_{\epsilon=0} + \frac{1}{2}\epsilon^2\frac{\partial^2\sigma(t)^{(\epsilon)}}{\partial\epsilon^2}\Big|_{\epsilon=0} + \frac{1}{6}\epsilon^3\frac{\partial^3\sigma(t)^{(\epsilon)}}{\partial\epsilon^3}\Big|_{\epsilon=0} \\ &+ \frac{1}{24}\epsilon^4\frac{\partial^4\sigma(t)^{(\epsilon)}}{\partial\epsilon^4}\Big|_{\epsilon=0} + O(\epsilon^5),\end{aligned}$$

where

$$\begin{aligned}\frac{\partial\sigma(t)^{(\epsilon)}}{\partial\epsilon}\Big|_{\epsilon=0} &= \sigma(0)\tilde{\alpha}\textstyle\int_0^t dW_2(s) = \sigma(0)\tilde{\alpha}W_2(t), \\ \frac{\partial^2\sigma(t)^{(\epsilon)}}{\partial\epsilon^2}\Big|_{\epsilon=0} &= 2\sigma(0)\tilde{\alpha}^2\textstyle\int_0^t\int_0^{s_1} dW_2(s_2)dW_2(s_1) = \sigma(0)\tilde{\alpha}^2\big(W_2(t)^2 - t\big), \\ \frac{\partial^3\sigma(t)^{(\epsilon)}}{\partial\epsilon^3}\Big|_{\epsilon=0} &= 6\sigma(0)\tilde{\alpha}^3\textstyle\int_0^t\int_0^{s_1}\int_0^{s_2} dW_2(s_3)dW_2(s_2)dW_2(s_1) \\ &= \sigma(0)\tilde{\alpha}^3\big(W_2(t)^3 - 3W_2(t)t\big), \\ \frac{\partial^4\sigma(t)^{(\epsilon)}}{\partial\epsilon^4}\Big|_{\epsilon=0} &= 24\sigma(0)\tilde{\alpha}^3\textstyle\int_0^t\int_0^{s_1}\int_0^{s_2}\int_0^{s_3} dW_2(s_4)dW_2(s_3)dW_2(s_2)dW_2(s_1) \\ &= \sigma(0)\tilde{\alpha}^3\big(W_2(t)^4 - 6W_2(t)^2t + 3t^2\big).\end{aligned} \tag{3.3.13}$$

Proof. It is easy to see from Eq. (3.3.12) that $\sigma^{(0)}(s) = \sigma(0)$. Straightforward application of the deterministic calculus rules gives us

$$\frac{\partial\sigma(t)^{(\epsilon)}}{\partial\epsilon} = \tilde{\alpha}\int_0^t \sigma^{(\epsilon)}(s)dW_2(s) + \epsilon\tilde{\alpha}\int_0^t \frac{\partial\sigma(t)^{(\epsilon)}}{\partial\epsilon}dW_2(s).$$

We take the limit $\epsilon \to 0$, and get the expression for the first-order expansion term:

$$\frac{\partial\sigma(t)^{(\epsilon)}}{\partial\epsilon}\Big|_{\epsilon=0} = \tilde{\alpha}\int_0^t \sigma^{(0)}(s)dW_2(s) = \sigma(0)\tilde{\alpha}\int_0^t dW_2(s).$$

The higher-order expansion terms follow by repeating these rules. □

Corollary 3.3.2. *The formal small disturbance expansion of the integrated volatility* $A(\Delta)^{(\epsilon)} = \int_0^\Delta \big(\sigma^{(\epsilon)}(t)\big)^2 dt$ *is now given by*

$$A^{(\epsilon)}(\Delta) = A^{(0)}(\Delta) + \epsilon A^{(1)}(\Delta) + \epsilon^2 A^{(2)}(\Delta) + \epsilon^3 A^{(3)}(\Delta) + \epsilon^4 A^{(4)}(\Delta) + O(\epsilon^5) \tag{3.3.14}$$

where

$$
\begin{aligned}
A^{(0)}(\Delta) &= \sigma(0)^2\Delta, \\
A^{(1)}(\Delta) &= \frac{\partial A^{(\epsilon)}(\Delta)}{\partial\epsilon}\Big|_{\epsilon=0} = 2\sigma(0)\int_0^\Delta \frac{\partial\sigma^{(0)}(t)}{\partial\epsilon}dt, \\
A^{(2)}(\Delta) &= \frac{\partial^2 A^{(\epsilon)}(\Delta)}{\partial\epsilon^2}\Big|_{\epsilon=0} = \int_0^\Delta \Big(\big(\frac{\partial\sigma(t)^{(0)}}{\partial\epsilon}\big)^2 + \sigma(0)\frac{\partial^2\sigma(t)^{(0)}}{\partial\epsilon^2}\Big)dt, \\
A^{(3)}(\Delta) &= \frac{\partial^3 A^{(\epsilon)}(\Delta)}{\partial\epsilon^3}\Big|_{\epsilon=0} = \int_0^\Delta \Big(\frac{\partial\sigma(t)^{(0)}}{\partial\epsilon}\frac{\partial^2\sigma(t)^{(0)}}{\partial\epsilon^2} + \frac{1}{3}\sigma(0)\frac{\partial^3\sigma(t)^{(0)}}{\partial\epsilon^3}\Big)dt, \\
A^{(4)}(\Delta) &= \frac{\partial^4 A^{(\epsilon)}(\Delta)}{\partial\epsilon^4}\Big|_{\epsilon=0} = \int_0^\Delta \Big(\frac{1}{4}\Big(\frac{\partial^2\sigma(t)^{(0)}}{\partial\epsilon^2}\Big)^2 + \frac{1}{3}\frac{\partial\sigma(t)^{(0)}}{\partial\epsilon}\frac{\partial^3\sigma(t)^{(0)}}{\partial\epsilon^3} + \frac{1}{12}\sigma(0)\frac{\partial^4\sigma(t)^{(0)}}{\partial\epsilon^4}\Big)dt.
\end{aligned}
$$

Proof. The above result is a natural extension of Proposition 2. One can easily check the validity of the above expression by a derivation from a different point-of-departure, i.e.,

$$
\begin{aligned}
A^{(\epsilon)}(\Delta) &= \int_0^\Delta \big(\sigma^{(\epsilon)}(t)\big)^2 dt \\
&= \int_0^\Delta \Bigg(\sigma(0) + \epsilon\frac{\partial\sigma(t)^{(\epsilon)}}{\partial\epsilon}\Big|_{\epsilon=0} + \frac{1}{2}\epsilon^2\frac{\partial^2\sigma(t)^{(\epsilon)}}{\partial\epsilon^2}\Big|_{\epsilon=0} + \frac{1}{6}\epsilon^3\frac{\partial^3\sigma(t)^{(\epsilon)}}{\partial\epsilon^3}\Big|_{\epsilon=0} \\
&\quad + \frac{1}{24}\epsilon^4\frac{\partial^4\sigma(t)^{(\epsilon)}}{\partial\epsilon^4}\Big|_{\epsilon=0} + O(\epsilon^5)\Bigg)^2 dt.
\end{aligned}
$$

Expand the inner quadratic expression and collect the terms up to order $O(\epsilon^4)$ gives us the expansion above. □

Based on the expressions above, the first conditional moment of $A^{(\epsilon)}(\Delta)$, given $W_2(\cdot)$ at terminal time Δ, can be computed by substituting the expansion terms (3.3.13) in Identity (3.3.14):

$$
\begin{aligned}
&\mathbb{E}\big[A^{(\epsilon)}(\Delta)|W_2(\Delta)\big] \\
=&\mathbb{E}\big[A^{(0)}(\Delta) + \epsilon A^{(1)}(\Delta) + \epsilon^2 A^{(2)}(\Delta) + \epsilon^3 A^{(3)}(\Delta) + \epsilon^4 A^{(4)}(\Delta) + O(\epsilon^5)|W_2(\Delta)\big] \\
=&\mathbb{E}\Big[\sigma(0)^2\Delta + 2\epsilon\sigma(0)^2\tilde{\alpha}\int_0^\Delta W_2(t)dt + \epsilon^2\sigma(0)^2\tilde{\alpha}^2\int_0^\Delta \big(2W_2(t)^2 - t\big)dt \\
&+ \epsilon^3\sigma(0)^2\tilde{\alpha}^3\int_0^\Delta \Big(\frac{4}{3}W_2(t)^3 - 2W_2(t)t\Big)dt \\
&+ \epsilon^4\sigma(0)^2\tilde{\alpha}^4\int_0^\Delta \Big(\frac{2}{3}W_2(t)^4 - 2W_2(t)^2 t + \frac{t^2}{2}\Big)dt + O(\epsilon^5)|W_2(\Delta)\Big] \\
=&\sigma(0)^2\Delta + 2\epsilon\sigma(0)^2\tilde{\alpha}\frac{1}{2}W_2(\Delta)\Delta + \frac{1}{3}\epsilon^2\sigma(0)^2\tilde{\alpha}^2\big(2W_2(\Delta)^2\Delta - \frac{\Delta^2}{2}\big) \\
&+ \frac{1}{3}\epsilon^3\sigma(0)^2\tilde{\alpha}^3\big(W_2(\Delta)^3\Delta - W_2(\Delta)\Delta^2\big) \\
&+ \epsilon^4\sigma(0)^2\tilde{\alpha}^4\left(\frac{2}{15}W_2(\Delta)^4\Delta - \frac{3}{10}W_2(\Delta)^2\Delta^2 + \frac{2}{5}\Delta^3\right) + O(\epsilon^5) \\
=&\sigma(0)^2\Delta\Big\{1 + \epsilon\tilde{\alpha}W_2(\Delta) + \frac{1}{3}\epsilon^2\tilde{\alpha}^2\big(2W_2(\Delta)^2 - \frac{\Delta}{2}\big) + \frac{1}{3}\epsilon^3\tilde{\alpha}^3\big(W_2(\Delta)^3 - W_2(\Delta)\Delta\big) \\
&+ \frac{1}{5}\epsilon^4\tilde{\alpha}^4\Big(\frac{2}{3}W_2(\Delta)^4 - \frac{3}{2}W_2(\Delta)^2\Delta + 2\Delta^2\Big)\Big\} + O(\epsilon^5),
\end{aligned}
\tag{3.3.15}
$$

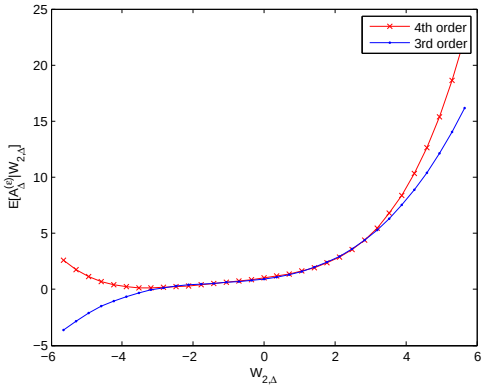

Figure 3.3: Illustration of formula (3.3.15) for the conditional mean $\mathbb{E}[A(\Delta)|W_2(\Delta)]$; including the first three expansion terms versus the first four terms.

where the derivations involving the computation of the time integral of the Wiener processes can be found in Kahl [62] (Table 4.1). The expressions for $\mathbb{E}[\int_0^\Delta W_2(t)^4 dt|W_2(\Delta)]$ and $\mathbb{E}[\int_0^\Delta W_2(t)^2 t dt|W_2(\Delta)]$ have not been provided in [62], but have been derived by ourselves as a straightforward (but tedious) extension of the derivations in [62].

Finally, we substitute $\alpha = \epsilon\tilde{\alpha}$ back in the expressions, collect the first three terms of the expansion to approximate the solution of the original model.

Remark. The expansion for the conditional mean (3.3.15) has to be performed up to fourth-order, because the third-order solution gives rise to negative values (for negative values of $W_2(\Delta)$), see Figure 3.3. However the variance should be non-negative, and the same holds for the integrated variance.

The computation of the conditional variance is however involved. Nevertheless, we identify the leading term as $\mathbb{E}\big[\epsilon^2\big(A^{(1)}(\Delta)-\mathbb{E}[A^{(1)}(\Delta)|W_2(\Delta)]\big)^2|W_2(\Delta)\big]$:

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}[A^{(\epsilon)}(\Delta)|W_2(\Delta)]\\
=&\mathbb{E}\big[\big(A^{(\epsilon)}(\Delta)-\mathbb{E}[A^{(\epsilon)}(\Delta)|W_2(\Delta)]\big)^2|W_2(\Delta)\big]\\
=&\mathbb{E}\big[\big(A^{(0)}(\Delta)+\epsilon A^{(1)}(\Delta)+\frac{1}{2}\epsilon^2 A^{(2)}(\Delta)+O(\epsilon^3)-\mathbb{E}[A^{(0)}(\Delta)+\epsilon A^{(1)}(\Delta)\\
&+\frac{1}{2}\epsilon^2 A^{(2)}(\Delta)+O(\epsilon^3)|W_2(\Delta)]\big)^2|W_2(\Delta)\big]\\
=&\mathbb{E}\Big[\Big(\epsilon\big(A^{(1)}(\Delta)-\mathbb{E}[A^{(1)}(\Delta)|W_2(\Delta)]\big)+\frac{1}{2}\epsilon^2\big(A^{(2)}(\Delta)-\mathbb{E}[A^{(2)}(\Delta)|W_2(\Delta)]\big)\\
&+O(\epsilon^3)\Big)^2|W_2(\Delta)\Big]\\
=&\mathbb{E}\big[\epsilon^2\big(A^{(1)}(\Delta)-\mathbb{E}[A^{(1)}(\Delta)|W_2(\Delta)]\big)^2|W_2(\Delta)\big]+O(\epsilon^3).
\end{aligned}
$$

So, we find:

$$\begin{aligned}
&\mathbb{E}\big[\epsilon^2\big(A^{(1)}(\Delta)-\mathbb{E}[A^{(1)}(\Delta)]\big)^2|W_2(\Delta)\big]\\
=&4\sigma(0)^4\alpha^2\mathbb{E}\big[\big(\int_0^\Delta W_2(t)dt\big)^2|W_2(\Delta)\big]-4\sigma(0)^4\alpha^2\mathbb{E}[\int_0^\Delta W_2(t)dt|W_2(\Delta)]^2\\
=&4\sigma(0)^4\alpha^2\Big(\mathbb{E}\big[\big(W_2(\Delta)\Delta-\int_0^\Delta tdW_2(t)\big)^2|W_2(\Delta)\big]-\mathbb{E}\big[\big(W_2(\Delta)\Delta\\
&-\int_0^\Delta tdW_2(t)\big)|W_2(\Delta)\big]^2\Big)\\
=&4\sigma(0)^4\alpha^2\Big(W_2(\Delta)^2\Delta^2-2W_2(\Delta)\Delta\frac{1}{2}W_2(\Delta)\Delta+\mathbb{E}\big[\big(\int_0^\Delta tdW_2(s)\big)^2|W_2(\Delta)\big]\\
&-\frac{1}{4}W_2(\Delta)^2\Delta^2\Big)\\
=&4\sigma(0)^4\alpha^2\Big(\frac{1}{4}W_2(\Delta)^2\Delta^2+\frac{1}{12}\Delta^3-\frac{1}{4}W_2(\Delta)^2\Delta^2\Big)\\
=&\frac{1}{3}\sigma(0)^4\alpha^2\Delta^3,
\end{aligned}\tag{3.3.16}$$

In the derivation above we have used the relation $d(W(t)\cdot t)=W(t)dt+tdW(t)$.

The conditional variance does not depend on $W_2(\Delta)$, which suggests that the conditional distribution of the integrated variance is 'shifted by the time Δ realization of the Brownian motion W_2', but its variance is not affected by $W_2(\Delta)$.

Conditional moment-matched log-normal sampling scheme

Under common market conditions, i.e. $\sigma(0)<1$ and $\alpha<1$, the conditional variance is a very small value, for small Δ. In other words, the randomness of the conditional distribution of $A^{(\epsilon)}(\Delta)$ is low in this situation. This suggests that one can accurately reproduce the conditional distribution of the integrated variance, $A^{(\epsilon)}(\Delta)$, by an approximate distribution having the same mean and variance. One could choose a Gaussian distribution for this purpose, however, a disadvantage is that in that case large weights are assigned to the negative part of the real axis (whereas $A(\Delta)$ cannot be negative). Therefore, we choose a (conditional) moment-matched *log-normally distributed* random variable to approximate the conditional distribution of $A^{(\epsilon)}(\Delta)$.

More precisely, we denote the conditional mean and variance of $A^{(\epsilon)}(\Delta)$ obtained from formula (3.3.15) and (3.3.16) by:

$$\begin{aligned}
m=&\ \sigma(0)^2\Delta\Big(1+\alpha W_2(\Delta)+\tfrac{1}{3}\alpha^2\big(2W_2(\Delta)^2-\tfrac{\Delta}{2}\big)+\tfrac{1}{3}\alpha^3\big(W_2(\Delta)^3-W_2(\Delta)\Delta\big)\\
&+\tfrac{1}{5}\alpha^4\Big(\tfrac{2}{3}W_2(\Delta)^4-\tfrac{3}{2}W_2(\Delta)^2\Delta+2\Delta^2\Big)\Big)\\
v=&\ \tfrac{1}{3}\sigma(0)^4\alpha^2\Delta^3,
\end{aligned}\tag{3.3.17}$$

respectively. Then, we define a log-normal random variable, $\log(X)\sim N(\mu,\sigma)$, with mean and variance m and v, respectively, i.e., $\mathbb{E}[X]=m$ and $\mathbb{V}\mathrm{ar}[X]=v$. Parameters μ and σ can be easily obtained if the values of mean and variance

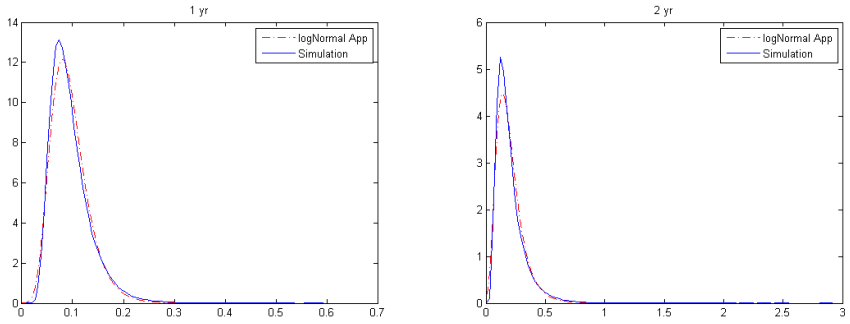

Figure 3.4: Quality of the approximation of the conditional moment-matched log-normal density compared to the true density of integrated variance, $A(\Delta)$. The parameters chosen are $\sigma(0) = 0.4$, $\alpha = 0.5$. In the LHS plot, $T = 1$; in the RHS plot $T = 2$.

are known:

$$\mu = \ln(m) - \frac{1}{2}\ln\big(1 + \frac{v}{m^2}\big), \quad \sigma^2 = \ln\big(1 + \frac{v}{m^2}\big).$$

The quality of this approximation is compared with simulation results for two time intervals, $\Delta = 1$ and $\Delta = 2$, in Figure 3.4. The shape of the density function for X is 'flatter' for larger values of Δ which reflects a higher uncertainty in the realizations of integrated variance $A(\Delta)$.

As $W_2(\Delta)$ is normally distributed with variance Δ, it is straightforward to compute the joint density of $W_2(\Delta)$ and $A^{(\epsilon)}(\Delta)$ by recalling:

$$\Pr\big(A(\Delta), W_2(\Delta)\big) \approx \Pr\big(A^{(\epsilon)}(\Delta), W_2(\Delta)\big) = \Pr\big(A^{(\epsilon)}(\Delta)|W_2(\Delta)\big)\Pr\big(W_2(\Delta)\big).$$

The above formula suggests that if we first draw normal random numbers, $W_2(\Delta)$, and then sample the integrated variance from the conditional distribution, $A(\Delta)$ given one realization of $W_2(\Delta)$, the joint realization of $A(\Delta)$ and $W_2(\Delta)$ reconstructs the desired joint density.

3.3.5 Discretization scheme for a full SABR model

We combine the two components described above and arrive at the *low-bias scheme for the SABR model* with a correlation structure. We start from the SABR model as in Eq. (3.2.1) with calibrated parameters α β and ρ. With an initial asset price and volatility at time 0, i.e. $S(0)$ and $\sigma(0)$, we simulate the discrete paths with an absorbing boundary at zero for the next time point, Δ, as follows:

1. Draw samples from a normal distribution, $W_2(\Delta) \sim N(0, \sqrt{\Delta})$, The volatility at time step Δ reads:

$$\sigma(\Delta) = \sigma(0)\exp\left(\alpha W_2(\Delta) - \frac{1}{2}\alpha^2\Delta\right).$$

2. Compute the asymptotic conditional mean, m, and variance, v, for the integrated variance, $A^{(\epsilon)}(\Delta)$, by

$$\begin{aligned} m = \quad & \sigma(0)^2\Delta\Big(1+\alpha W_2(\Delta)+\tfrac{1}{3}\alpha^2\big(2W_2(\Delta)^2-\tfrac{\Delta}{2}\big)+\tfrac{1}{3}\alpha^3\big(W_2(\Delta)^3-W_2(\Delta)\Delta\big) \\ & +\tfrac{1}{5}\alpha^4\Big(\tfrac{2}{3}W_2(\Delta)^4-\tfrac{3}{2}W_2(\Delta)^2\Delta+2\Delta^2\Big)\Big), \\ v = \quad & \tfrac{1}{3}\sigma(0)^4\alpha^2\Delta^3 \end{aligned}$$

3. Compute the parameters of the moment-matched log-normal distribution by

$$\mu = \ln(m) - \frac{1}{2}\ln\big(1+\frac{v}{m^2}\big), \quad \sigma^2 = \ln\big(1+\frac{v}{m^2}\big).$$

4. Draw (a vector of) uniform random numbers, U_1, and determine their inverse according to the log-normal distribution (defined by μ and σ):

$$A(\Delta) = exp\big(\sigma\cdot N^{-1}(U_1)+\mu\big).$$

5. Insert $A(\Delta)$ and $\sigma(\Delta)$ in the algorithm described in Sec. 3.3.3 to sample the conditional CEV process.

3.4 Numerical experiments

To analyze the validity and efficiency of our discretization schemes numerically, we price some European options based on the parameter sets in Table 3.3.

The first two test cases represent two limiting cases for the SABR model, i.e. $\beta = 1$ and $\alpha = 0$, respectively. Our aim is here to check the efficiency of two components of the scheme proposed, i.e. the moment-matched log-normal integrated variance sampling scheme and the direct inversion scheme for the conditional CEV process. In Case I, $\beta = 1$, the asset price follows a basic geometric Brownian motion and there is no complication with the absorbing boundary at zero. The main pricing bias will then be from the moment-matching algorithm to sample the integrated variance. In Case II, we set $\alpha = 0$, so that we isolate the part which is related to the simulation of the CEV process. In this second test, we give special attention to the martingale property of the simulated path.

For Test Case I, we use the option prices generated by a small time step Euler Monte Carlo simulation as reference prices, whereas for Test Case II we use an analytic option pricing formula, derived by Schroder [100], as the reference.

Next to these two tests, we also consider two practically relevant, yet challenging, parameter sets, i.e. Cases III and IV, with parameters often observed in fixed income instrument pricing. Parameter set III is representative for a low interest rate market, as observed for example in Japan. This parameter set is often embedded in the popular power-reverse dual contract. In this set, the impact of the behavior at the zero boundary on the price should be clearly visible. Parameter set IV then describes volatile market conditions as a high volatility-of-volatility parameter gives rise to a heavy tailed distribution of the asset prices.

The benchmark Monte Carlo scheme, for the low-bias scheme proposed here, is the Full Truncation Euler scheme from [75]. All Monte Carlo simulations have been performed simulating 10^5 paths.

Table 3.3: Parameters in Test Cases I to IV for the numerical experiments.

	Set I	Set II	Set III	Set IV
$S(0)$	4%	4%	0.5%	7%
$\sigma(0)$	20%	20%	20%	40%
α	0.3	0	0.3	0.8
β	1	0.4	0.5	0.5
ρ	-0.5	0	-0.3	-0.6

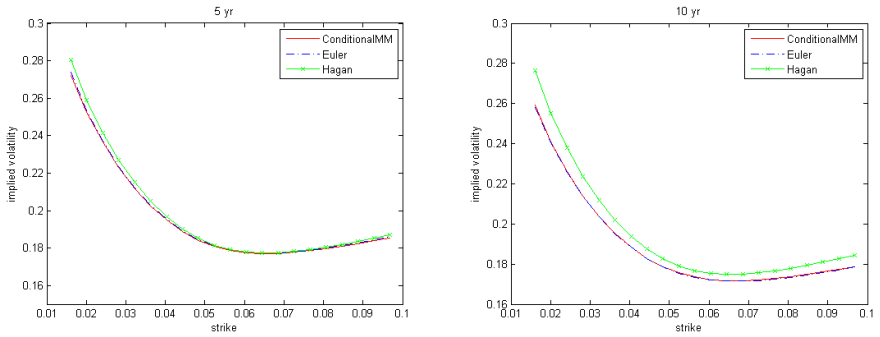

Figure 3.5: Result of the conditional moment-matching log-normal sampling scheme for the integrated variance, simulating 5 and 10 year call option prices under a double log-normal model. A comparison is made with the truncated Euler Monte Carlo scheme, and Hagan's asymptotic SABR formula. Parameters used are $\alpha = 0.3$, $\rho = -0.5$, $S(0) = 0.04$, $\sigma(0) = 0.2$.

3.4.1 Results for Test Case I

As in a double log-normal model (i.e. $\beta = 1$) $S(t) = 0$ cannot be reached, the Euler scheme performs well. It is also reasonable to expect that the Euler Monte Carlo scheme with a sufficiently large number of time steps is stable and converges to the correct solution. Here, we perform two tests, with $T = 5$ and $T = 10$, respectively.

Table 3.4 shows that the conditional moment-matched integrated variance sampling scheme produces a very small bias for a practical number of time steps, like 2 or 4 time steps per year. The accuracy of the low-bias scheme is comparable to that of an Euler scheme with 50 times more time steps. To illustrate the accuracy of the conditional moment-matching scheme, we present the implied option volatilities from the moment-matching scheme with 4 time steps a year together with the 200 time steps Euler scheme and Hagan's asymptotic formula in Figure 3.5. In the two figures, we observe that the difference between the implied volatilities from the 200 time steps Euler scheme and the low-bias SABR scheme with 4 time steps is negligible, whereas Hagan's asymptotic formula produces a visible pricing error for the 5 year maturity and is inexact for the 10 year maturity.

Table 3.4: Results of the low-bias SABR scheme with $\beta = 1$, Test Case I.

K	Δ	40%	80%	100%	120%	160%	200%
		T= 5					
Euler	1/200	0.02077	0.00889	0.00512	0.00279	0.00083	0.00029
low-bias	1/2	0.02079	0.00887	0.00510	0.00277	0.00082	0.00029
	1/4	0.02076	0.00890	0.00512	0.00279	0.00082	0.00029
Hagan		0.02083	0.00894	0.00514	0.00279	0.00082	0.00030
		T= 10					
Euler	1/200	0.02198	0.01124	0.00758	0.00503	0.00234	0.00125
low-bias	1/2	0.02196	0.01122	0.00756	0.00502	0.00233	0.00124
	1/4	0.02198	0.01124	0.00758	0.00504	0.00235	0.00126
Hagan		0.02230	0.01154	0.00781	0.00521	0.00248	0.00139

3.4.2 Results for Test Case II

With $\alpha = 0$, the stochastic volatility part vanishes and the system reduces to a plain CEV model. An option pricing formula for the CEV model is known then in closed form [100]. More recently, Lesniewski [71] provided a classification (and an explicit option pricing formula) for the CEV process with absorbing and with reflecting boundaries. Hence, we have an analytic reference value so that we can determine the accuracy of our discretization scheme and the price impact of the assumptions related to the boundary condition at zero. In this subsection, we also examine the martingale property of the discrete process in the algorithm proposed here.

In detail, we have implemented an Euler scheme with full truncation at zero, i.e. $\hat{S}(\Delta) = \max\left(\hat{S}(\Delta), 0\right)$, the direct inversion scheme for the CEV process, but with a reflecting boundary, as well as our proposed low-bias scheme. We present the implied volatilities obtained by these numerical schemes and compare them to the exact CEV option pricing formula with absorbing boundaries in [71], in Figure 3.6. We also include Hagan's asymptotic formula in the comparison by choosing a very small volatility-of-volatility parameter and a small correlation coefficient [3], i.e. $\alpha = 0.0001$ and $\rho = 0.0001$. In all experiments the Euler scheme consists of 50 times more time steps than the low-bias SABR scheme.

From Figure 3.6, we see that:

- The low-bias scheme has a low bias for all strikes and maturities. In most of the cases the implied volatilities obtained by the low-bias scheme are highly accurate when compared to the exact solution. The low-bias scheme is essentially free of bias with merely four time steps a year, whereas the Full Truncation Euler scheme requires more than 200 time steps to converge.

- The results from the low-bias scheme with the reflecting boundary agree very well with the exact solution and with our proposed low-bias scheme for strikes that are far away from zero. For small strike values there is

[3] This makes sense because the CEV model is a special case of the SABR model with a zero volatility-of-volatility parameter.

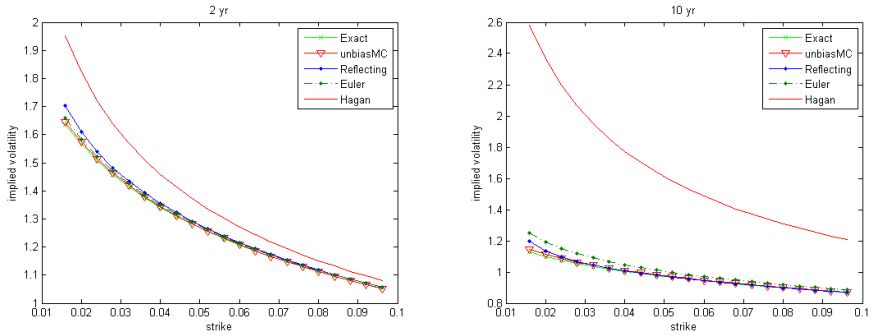

Figure 3.6: Comparison for the CEV process of the implied volatilities generated by different methods; Maturities are 2 and 10 years, and parameters $\alpha = 0.0$, $S(0) = 0.04$, $\sigma(0) = 0.2$ and $\beta = 0.4$.

a substantial miss-pricing by the direct inversion scheme with reflecting boundary and its pricing bias increases with maturity.

- Hagan's formula is not an accurate approximation for the CEV model in the parameter range of small β (i.e. $\beta \leq 0.5$).
- The small time step Full Truncation Euler scheme performs reasonably well for all maturities, in particular for short maturities. However, we observe an upward shift in the implied volatility curve (the shift is larger for long maturities). This upward 'bias' is the result of the truncation and can not be removed completely, not even by smaller time steps Δ.
- The pricing biases from the Full Truncation Euler scheme as well as from the direct inversion scheme with reflecting boundary are most significant for small values of the strike.

We focus on the martingale property of the discretized processes generated by the different simulation schemes. In Table 3.5 we see that the direct inversion scheme with reflecting boundary gives rise to a positive drift which decreases with smaller time steps. In contrast, the proposed low-bias SABR scheme does not generate any statistically significant drift, and the martingale property is preserved. The Euler scheme with Full Truncation does not preserve the martingale property, although the drift decreases (but does not disappear) with smaller time step Δ.

To show the order of convergence of the low-bias SABR scheme, we compute the root-mean-squared (RMS) errors of at-the-money (ATM) option prices obtained by the Euler and the low-bias SABR scheme for different numbers of time steps, see Figure 3.7. The convergence behavior of the low-bias scheme appears to be superior (which is also the case for other strike values, not shown).

3.4.3 Results for Cases III and IV

For Test Cases III and IV, we consider the full correlation SABR model and some practical yet challenging parameter settings, like rates almost zero or high volatility-of-volatility parameter.

Table 3.5: Test Case II. Test of martingale properties of several discretization schemes for a pure CEV process with initial asset price at 4%, i.e. $S(0) = 0.04$. The numbers shown in the table are $\frac{1}{N}\sum_{i=1}^{N}\hat{S}(T)$, with $\hat{S}(T)$ generated by different discretization schemes. The Full Truncation Euler schemes considered are implemented with 50 times more time steps than the low-bias scheme.

Δ	Low-bias	Truncated Euler	Reflecting
		$T = 5$	
1/2	4%	4.05%	4.39%
1/4	4%	4.06%	4.22%
1/8	4%	4.07%	4.13%
1/16	4%	4.04%	4.07%
		$T = 10$	
1/2	4.01%	4.09%	4.47%
1/4	3.99%	4.05%	4.25%
1/8	4%	4.02%	4.16%
		$T = 15$	
1/2	3.99%	4.06%	4.47%
1/4	4.01%	4.04%	4.31%
1/8	4%	3.99%	4.15%

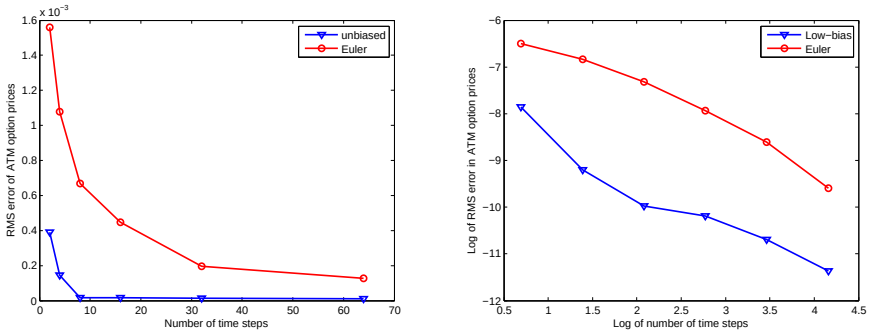

Figure 3.7: Test Case II; Convergence of the estimated RMS error for call options with decreasing time step Δ.

Due to the absence of reference values (by a semi-closed form solution or an accurate approximation), we make use of the following result to determine the convergence order of our algorithm:

Result 3.4.1 (Weak Convergence Order without an Exact Solution [98]). *1) If a discrete approximation $\hat{X}$ to a continuous process X with time step Δ has a weak convergence order γ for some positive constant K_1, i.e.:*

$$\left|\mathbb{E}\big[g\big(X(T)\big)\big] - \mathbb{E}\big[g\big(\hat{X}(T,\Delta)\big)\big]\right| \leq K_1\Delta^{\gamma} \tag{3.4.1}$$

Then, there exists a positive constant, K_2, so that:

$$\left|\mathbb{E}\big[g\big(\hat{X}(T,\Delta)\big)\big] - \mathbb{E}\big[g\big(\hat{X}(T,\frac{\Delta}{2})\big)\big]\right| \leq K_2\Delta^{\gamma}. \tag{3.4.2}$$

2) Conversely, if it is known that the discretization is weakly convergent and Eq. (3.4.2) holds for some positive constant K_2, then the weak convergence order is γ. The proof can be found in Schmitz-Abe & Shaw [98].

Parameter set III is particularly challenging for the Euler discretization scheme and for Hagan's SABR formula, because the initial rates are close to zero. It is known that when initial asset prices are close to zero, many paths may reach negative values. The Full Truncation Euler scheme will project the negative values to zero. The drawback is that the truncated Gaussian process is not a martingale anymore, and an increasing number of time steps has to be employed to reduce the resulting bias. On the other hand, the asymptotic SABR formula by Hagan is not valid for strikes $K \to 0$. The formula is a result of keeping $\log\big(f/K\big)$ constant and taking $T \to 0$. However, when $K \to 0$, it follows that $\log\big(f/K\big) \to \infty$, which is an incorrect way of approaching the asymptotic limit. As a result, Hagan's formula is not accurate for very low strike prices.

In the low-bias SABR scheme, option pricing at low strike values does not pose any problem. We apply Formula (3.4.2) to the Monte Carlo prices of the ATM options and define the relative error to be

$$\epsilon = \left|\hat{C}\big(\hat{S}(T,\Delta)\big) - \hat{C}\big(\hat{S}(T,\frac{\Delta}{2})\big)\right|.$$

Here, $\hat{C}$ denotes the Monte Carlo estimate of the call option price for underlying discrete process $\hat{S}$. Clearly, the low-bias scheme produces convergent Monte Carlo prices (see Figures 3.8 and 3.9) and smaller errors than the truncated Euler scheme. Again, the relative error of the low-bias scheme with only 4 time steps per year is comparable to that of the Euler scheme with more than 32 time steps per year.

For Test Case IV the resulting call option prices are presented in Table 3.6.

3.4.4 Computational time

From the numerical results presented above, the accuracy of the low-bias scheme has been confirmed. In addition, the CPU time to compute one sample path is of great importance for practical application. The CPU time required for

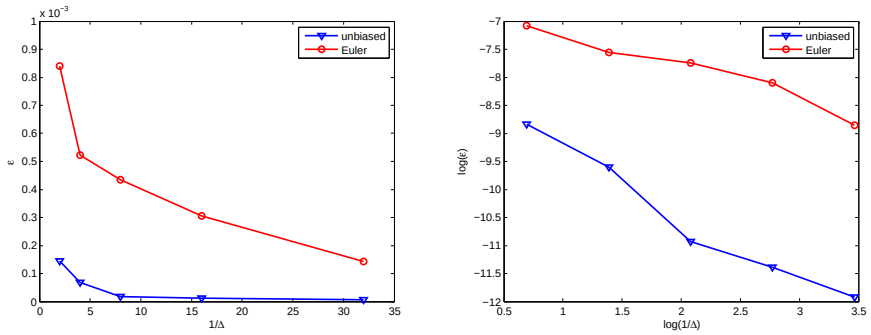

Figure 3.8: Convergence of relative errors; Test Case III; left: error versus time step size, right: same picture in log-log scale.

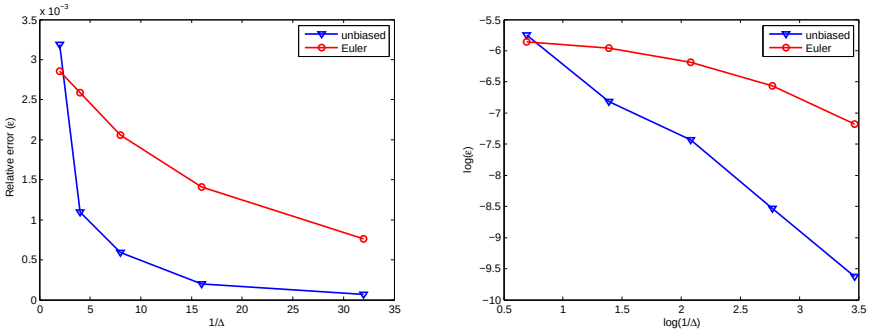

Figure 3.9: Convergence of relative errors, $T = 5$; Test Case IV; error versus time step size, right: same picture in log-log scale.

Table 3.6: Estimated call option prices for Test Case IV.

	$K = 40\%$		$K = 100\%$		$K = 160\%$	
Δ	Low-bias	Euler	Low-bias	Euler	Low-bias	Euler
	$T = 2$					
1	0.0803	0.0922	0.0645	0.0714	0.0510	0.0538
1/2	0.0688	0.0838	0.0535	0.0649	0.0405	0.0489
1/4	0.0642	0.0771	0.0492	0.0596	0.0366	0.0449
1/8	0.0619	0.0708	0.0474	0.0546	0.0352	0.0409
1/16	0.0610	0.0673	0.0468	0.0518	0.0347	0.0388
1/32	0.0604	0.0643	0.0463	0.0494	0.0343	0.0368
	$T = 5$					
1	0.0795	0.1074	0.0672	0.0895	0.0564	0.0735
1/2	0.0693	0.0928	0.0576	0.0767	0.0454	0.0624
1/4	0.0667	0.0827	0.0540	0.0680	0.0429	0.0550
1/8	0.0643	0.0759	0.0523	0.0621	0.0412	0.0499
1/16	0.0632	0.0702	0.0512	0.0572	0.0406	0.0457
1/32	0.0625	0.0674	0.0506	0.0548	0.0400	0.0437
	$T = 10$					
1	0.0765	0.1133	0.0665	0.0966	0.0575	0.0815
1/2	0.0678	0.0975	0.0566	0.0823	0.0466	0.0687
1/4	0.0669	0.0850	0.0550	0.0713	0.0451	0.0591
1/8	0.0652	0.0774	0.0537	0.0645	0.0434	0.0530
1/16	0.0635	0.0705	0.0523	0.0584	0.0422	0.0476
1/32	0.0630	0.0678	0.0520	0.0561	0.0421	0.0456

the low-bias scheme is largely dependent on the value of parameter β. Computations are faster for $\beta \approx 1$ and somewhat slower for $\beta \approx 0$. This is due to the space transform employed in the derivation of the low-bias SABR scheme, $Y = S^{2-2\beta}/(1-\beta)^2$. With $\beta \approx 1$, Y tends to infinity and the distribution approaches a Gaussian distribution. Most of the sample paths will then be drawn from the 'cheap to evaluate' quadratic Gaussian approximation. On the contrary, when $\beta \approx 0$, most of the draws will be from the direct inversion scheme, which takes more CPU time.

Table 3.7: Computational time (in seconds) for a 1 year option with parameters given in Test Case III and IV.

Δ	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$
	Test Case III				
Low-bias	8.59	10.97	13.75	23.61	40.48
Euler FT	0.26	0.43	0.83	1.59	3.17
	Test Case IV				
Low-bias	4.75	7.38	12.61	23.06	42.13
Euler FT	0.26	0.43	0.83	1.59	3.17

Table 3.7 presents CPU times for the Full Truncation Euler and the low-bias schemes used in Cases III and IV for a small number of time steps.

For Test Case III, we now analyze the martingale property, $\mathbb{E}[\hat{S}(T)] = \hat{S}(0)$, of the discrete process to analyze the accuracy (or bias) of the low-bias scheme. For comparison, we also simulate with the Full Truncation Euler scheme with the same number of time steps (see Figure 3.10).

The results in Figure 3.10 indicate that the low-bias scheme with the absorbing boundary generates a very small bias, which rapidly disappears with a larger number of time steps. For the truncated Euler scheme the martingale bias decreases slowly in the number of time steps and the bias disappears after using 512 steps per year.

In Case III, due to the size of the β parameter, the computational time per step of the low-bias scheme is approximately 10 times that of the Euler scheme. The truncated Euler scheme suffers however from a non-negligible martingale (first moment) bias, even with 256 time steps per year (see Figure 3.10). In contrast, the low-bias scheme is almost free of bias with 2 steps per year. With 4 time steps per year, the low-bias scheme requires approximately 10 seconds per year, whereas 50.34 seconds are used by the truncated Euler scheme with 512 time steps.

Case IV is challenging because a relatively small β-value is combined with a large α-parameter ($\alpha = 80\%$). The low-bias scheme uses on average more than 10 times the computational effort per time step compared to the truncated Euler scheme in this case. Despite this, we argue that it is preferable to choose the low-bias scheme also for this parameter set because the truncated Euler scheme tends to be highly biased [4] and this bias is still significant for 2048 time steps per year (see Figure 3.11), costing more than 200 seconds computational time

[4] The probability of hitting zero is very high, so that a large number of truncations of the negative paths is expected.

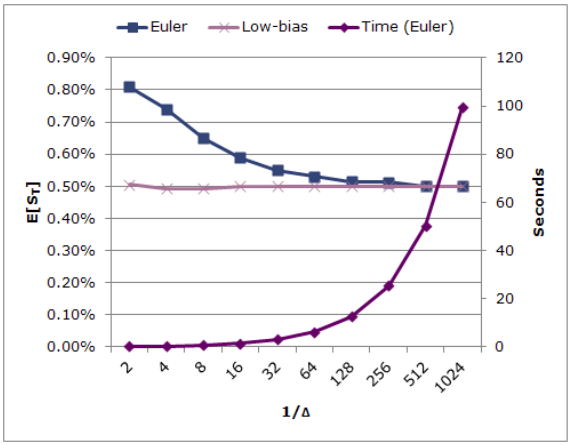

Figure 3.10: Comparison of martingale biases of the low-bias scheme and the truncated Euler scheme for a simulation up to one year in Test Case III. The $E[\hat{S}(T)]$ curves are the discrete approximations of $\frac{1}{N}\sum_{i=1}^{N}\hat{S}(T)$, with $\hat{S}(T)$ generated by different discretization schemes. The computational time of the truncated Euler scheme is plotted against the number of time steps in the secondary y-axis.

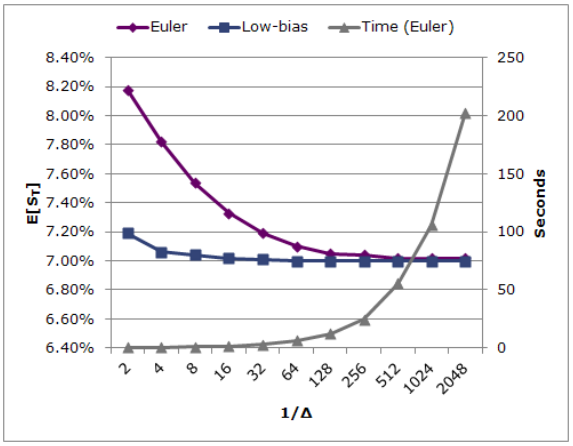

Figure 3.11: Comparison of martingale bias of the low-bias scheme and the truncated Euler scheme for a simulation up to one year in Test Case IV. The computational time of the truncated Euler scheme is plotted against the number of time steps in the secondary y-axis.

per year. The low-bias scheme is then clearly more efficient to reach the same level of accuracy.

3.5 Conclusion

In this chapter we have presented a low-bias SABR simulation scheme. We firstly reviewed some analytic properties of the CEV process, which is a space transformed squared Bessel process, and discussed the classification of boundary conditions and the associated probability density and distribution functions. As the conditional SABR process, given the terminal volatility level and the integrated variance, is also a squared Bessel process, we can find an explicit distribution function for the conditional SABR process.

Based on the idea of mixing conditional distributions and a direct inversion of the noncentral chi-square distributions, we have proposed a low-bias SABR Monte Carlo scheme. The scheme proposed can deal with the – often problematic – behavior of the CEV process in the vicinity of the zero boundary. The low-bias scheme is stable and exhibits a highly satisfactory convergence behavior compared to the truncated Euler scheme. The scheme is an alternative when a truncated Euler scheme gives rise to significant bias, even with a very large number of time steps, which is the case, for example, when $S(0) \approx 0$ or when the skewness parameter, β, is less than $\frac{1}{2}$.

CHAPTER 4

Calibration and Monte Carlo Pricing of the SABR-Hull-White Model for Long-Maturity Equity Derivatives

This chapter is adapted from the article
"Calibration and Monte Carlo Pricing of the SABR-Hull-White Model for Long-Maturity Equity Derivatives"
accepted by *Journal of Computational Finance* in Volume 15, Issue 4, pp. 79-113 (2012) [30].

We model the joint dynamics of stock and interest rate by a hybrid SABR-Hull-White model, in which the asset price dynamics are modelled by the SABR model [46] and the interest rate dynamics by the Hull-White short-rate model [49]. We propose a projection formula, mapping the SABR-HW model parameters onto the parameters of the *nearest* SABR model. Further a time-dependent parameter extension of this SABR-HW model is introduced to make the calibration of the model consistent across maturities. The inverse of the projection formula enables a rapid calibration of the model. As the calibration quality is subject to the approximation errors of the projection formula, we subsequently apply a non-parametric numerical calibration technique based on the non-uniformly weighted Monte Carlo technique [13] to improve the calibration. In this step, the Monte Carlo weights are not uniform and chosen in such a way that the calibration market instruments are perfectly replicated.

4.1 Introduction

Equity derivative models and yield curve models have been developed independently of each other for a long time. Whereas the equity derivative models focused on the implied volatility skew/smile by local or stochastic volatility features [37], short-rate models improved the accuracy of the yield curve dynamics. With an increasing interest in long-maturity equity derivatives, as well as in equity-interest rate hybrid products from retail and long term institutional investors (see [51]), the industrial practice also demands models that are capable of describing joint dynamics of interest rates and equity. Indeed, one would intuitively assert that the interest rate is stochastic and that there is non-zero correlation between the interest rate and the equity.

In the present chapter, we propose a hybrid extension of the Stochastic Alpha Beta Rho (SABR) model [46] to model the joint equity-interest rate dynamics. We construct a hybrid model, called the SABR-Hull-White (SABR-HW) model, in which the equity process is driven by the SABR model and the interest rate by the short-rate model of [49]. In this framework the equity process is assumed to be correlated with the interest rate process.

The SABR model, which is not often used in the equity derivative literature, has several attractive features for the modelling of long term equity-linked products. It generalizes stochastic volatility models such as Heston's model by introducing an explicit stock price dependence in a *power law* local volatility term, $\phi(x) = x^\beta$ [77]. Further, the SABR model admits a closed-form approximation formula ("Hagan's formula") for the Black implied volatilities which greatly simplifies calibration. Thirdly, the SABR process is, for certain parameter values, an absorbing process at the zero asset price boundary [32], which models the fact that companies may default in time ([77]). Last but not least, the parameters in the SABR model have a direct connection to market instruments or market price features, in contrast to, for example, the speed-of-mean-reversion parameter in the Heston model.

One contribution in this chapter is an invertible projection formula of the constant parameter SABR-HW model onto the *plain* SABR model. This formula enables a highly efficient calibration of the constant parameter SABR-HW model based on the established calibration procedure for the SABR model. The resulting calibration parameters remain however only valid for a single maturity time and cannot provide a consistent dynamic description of the underlying asset prices across multiple maturity times. We deal with this issue by adopting a dynamic SABR(-HW) extension in the spirit of Rebonato [93].

Moreover, we will use the well-known weighted Monte Carlo (WMC) technique, proposed by [13], as another stage of calibration of the SABR-HW model. By this we can deal with the inconsistency between the true model dynamics and those implied by Hagan's asymptotic approximation formula (by which the calibration instruments are quoted).

For the Monte Carlo simulation, we adopt the low-bias discretization from the previous chapter for the SABR-HW dynamics, which has some advantages over a basic Euler scheme as it gives a low bias (ie, stable and accurate) when large time steps are used (say 4 time steps per year), see [32].

This chapter is organized as follows. In Section 2, we define the dynamical SABR-HW model and discuss the building blocks. We show how to project the constant parameter SABR-HW dynamics onto a plain SABR model in Section 3.

In Section 4, we utilize this projection formula to calibrate the constant parameter version of the model. We also show how to determine the time-dependent functions in the dynamical SABR-HW model and use the WMC technique [13]. In Section 5, the low-bias Monte Carlo simulation for the dynamical SABR-HW model is presented. Numerical experiments for validation and calibration are discussed throughout this chapter.

4.2 The Dynamical SABR-HW Model

This section describes the construction of the dynamical Stochastic Alpha Beta Rho Hull-White (SABR-HW) equity-interest rate model. We assume efficient markets and the existence of an equivalent martingale measure (EMM) $\mathbb{Q}$ when appropriate numéraires need to be chosen.

4.2.1 Model Definition

We define the dynamical SABR-HW model in a similar fashion as Rebonato's SABR-LMM model [93] for forward rates. The full-scale dynamical SABR-HW model for equity-interest rate products, *under the* $\mathbb{Q}$*-measure* associated with $B(t)$, a money-saving account, is given by:

$$\begin{aligned} dS(t)/S(t) &= r(t)dt + \Sigma(t)S^{\beta-1}(t)dW_x(t), \quad S(0) > 0, \\ dr(t) &= \lambda(\theta(t) - r(t))dt + \eta dW_r(t), \quad r(0) > 0, \\ \Sigma(t) &= g(t) \cdot k(t), \\ dk(t) &= h(t)k(t)dW_\Sigma(t), \quad k(0) = 1, \\ dW_x(t)dW_\sigma(t) &= \rho_{x,\sigma}dt, \quad -1 \leq \rho_{x,\sigma} \leq 1, \\ dW_x(t)dW_r(t) &= \rho_{x,r}dt, \quad -1 \leq \rho_{x,r} \leq 1 \end{aligned} \tag{4.2.1}$$

with constant parameters β, λ, η, $\rho_{x,r}$ and $\rho_{x,\sigma}$, and appropriately chosen time-dependent functions $\theta(t)$, $g(t)$ and $h(t)$. For simplicity, we assume here that the interest rates are independent of the stochastic volatility, $dW_r(t)dW_\sigma(t) = 0$. The parameters will be discussed in the subsections to follow, and details of the functional form of $g(t)$ and $h(t)$ are given in Section 4.4.2.

Remark. For some additional insight in the functions $g(t)$ and $h(t)$, we derive the dynamics of the time-dependent volatility $\Sigma(t)$, by applying Itô's lemma,

$$\begin{aligned} d\Sigma(t) \equiv \quad d(g(t)k(t)) &= k(t)dg(t) + g(t)dk(t) + dk(t)dg(t) \\ &= \left(\frac{1}{g(t)}\frac{dg(t)}{dt}\right)\Sigma(t)dt + h(t)\Sigma(t)dW_\Sigma(t) \\ &= \frac{d\log g(t)}{dt}\Sigma(t)dt + h(t)\Sigma(t)dW_\Sigma(t). \end{aligned}$$

Overall this implies that the dynamics for the volatility can be defined as:

$$d\Sigma(t) = \hat{g}(t)\Sigma(t)dt + h(t)\Sigma(t)dW_\Sigma(t),$$

with

$$\hat{g}(t) = \frac{d\log g(t)}{dt},$$

which means that we deal with a lognormal process with time-dependent drift and volatility terms. It is obvious that the function $h(t)$ plays the role of volatility coefficient for this volatility process and the function $g(t)$, appearing in the drift term, shifts the volatility up and down deterministically.

The Hull-White Model

One of the building blocks of hybrid model (4.2.1) is the Hull-White, single-factor, no-arbitrage yield curve model in which the short-term interest rate, $r(t)$, is driven by an Ornstein-Uhlenbeck (OU) mean reverting process, with $\theta(t) > 0$, $t \in \mathbb{R}^+$ a time-dependent drift term, to fit theoretical bond prices to the yield curve observed in the market. Parameter η determines the overall level of volatility and the reversion rate parameter, λ, determines the relative volatilities.

Under the Hull-White model the dynamics of the zero-coupon bond, paying €1 at maturity T, are given by:

$$\frac{dP(t,T)}{P(t,T)} = r(t)dt + \frac{\eta}{\lambda}\left(\mathrm{e}^{-\lambda(T-t)} - 1\right) dW_r(t). \tag{4.2.2}$$

Since the Hull-White model belongs to the class of affine diffusion processes, the solution of (4.2.2) is known analytically and reads:

$$P(t,T) = \exp\left(A(t,T) + B_r(t,T)r(t)\right), \tag{4.2.3}$$

with

$$B_r(t,T) = \frac{1}{\lambda}\left(\mathrm{e}^{-\lambda(T-t)} - 1\right), \tag{4.2.4}$$

$$A(t,T) = \exp\left(\log\left(\frac{P(0,T)}{P(0,t)}\right) - B_r(t,T)f(0,t) - \frac{\eta^2}{4\lambda}\left(1 - \mathrm{e}^{-2\lambda t}\right)\left(B_r(t,T)\right)^2\right), \tag{4.2.5}$$

where $f(0,t) := -\partial P(0,t)/\partial t$, with $P(0,t)$ the market discount factor for maturity t.

By the Radon-Nikodým derivative [38],

$$\frac{d\mathbb{Q}^T}{d\mathbb{Q}} = \frac{P(t,T)}{P(0,T)B(t)}, \tag{4.2.6}$$

we find the following change of measure: $dW_r^T(t) = dW_r(t) - \eta B_r(t,T)dt$.

Short rate, $r(t)$, under the T-forward measure is governed by the following dynamics:

$$dr(t) = \left(\lambda(\theta(t) - r(t)) + \eta^2 B_r(t,T)\right)dt + \eta dW_r^T(t),$$

which can be written as:

$$dr(t) = \lambda\left(\hat{\theta}(t) - r(t)\right)dt + \eta dW_r^T(t), \tag{4.2.7}$$

with $\hat{\theta}(t) = \theta(t) + \frac{\eta^2}{\lambda} B_r(t,T)$, and $B_r(t,T)$ in (4.2.4). Since the process under the T-forward measure in (4.2.7) 1 is of 'Hull-White form', it is normally distributed [25] with expectation and variance given by:

$$\mathbb{E}^T(r(t)) = r_0 \mathrm{e}^{-\lambda t} + \lambda \int_0^t \hat{\theta}(s) \mathrm{e}^{-\lambda(t-s)} ds, \tag{4.2.8}$$

$$\mathbb{V}\mathrm{ar}^T(r(t)) = \frac{\eta^2}{2\lambda}\left(1 - \mathrm{e}^{-2\lambda t}\right). \tag{4.2.9}$$

A disadvantage of the Hull-White model is that it may give rise to negative interest rates. The negative interest rate, however, may be present in the real market [1]. An alternative to the Hull-White model is the Cox-Ingersoll-Ross (CIR) model. A hybrid SABR-CIR model is, however, not tractable if there is a non-zero correlation between the interest rate and the SABR equity process. The choice between a CIR or Hull-White model within the hybrid process is a trade-off between non-zero correlation and non-negative rates.

For hybrid structured products, a non-zero correlation is a crucial feature that should be incorporated into a model (see [41] for analysis and further arguments), whereas the appearance of negative interest rates in a Hull-White process is an inherent feature of the model and is known by practitioners for quite some time [25]. There are practical fixes to this problem, eg, choosing parameters which give rise to lower probabilities for negative rates. We therefore prefer the HW process over the CIR process as part of our equity-interest rate hybrid model.

Constant Parameter SABR-HW model

The second building block of Model (4.2.1) is the Stochastic Alpha Beta Rho (SABR) stochastic volatility model by [46].

The SABR SDE system with constant parameters was originally defined under the T-forward measure, as:

$$\begin{aligned} dS(t) &= \sigma(t) S(t)^{\beta} dW_x^T(t), \\ d\sigma(t) &= \alpha\sigma(t) dW_\sigma^T(t), \end{aligned} \tag{4.2.10}$$

with $dW_x^T(t) dW_\sigma^T(t) = \rho_{x,\sigma} dt$.

One of the reasons why the original SABR model is not applied to equity derivatives is that a drift term is lacking. Risk-neutral equity price processes are defined with a drift term, and are assumed to be arbitrage-free under the risk-neutral measure associated with the money-savings account. For long-maturity equity options and equity interest rate hybrids, however, industrial practice is to treat the interest rate as a stochastic process as well. As shown below, when combining the Hull-White interest rate model with the SABR equity model the drift term appears naturally in the SABR equity dynamics under the risk neutral $\mathbb{Q}$-measure:

$$\begin{aligned} dS(t)/S(t) &= r(t)dt + \sigma(t) S^{\beta-1}(t) dW_x(t), \quad S(0) > 0, \\ d\sigma(t) &= \alpha\sigma(t) dW_\sigma(t), \quad \sigma(0) > 0, \\ dr(t) &= \lambda(\theta(t) - r(t))dt + \eta dW_r(t), \quad r(0) > 0, \end{aligned} \tag{4.2.11}$$

[1] http://en.wikipedia.org/wiki/Interest_rate

with constant model parameters $0 < \beta < 1$, $\alpha > 0$, $\lambda > 0$ and $\eta > 0$. As in System (4.2.1), we assume non-zero correlations $dW_x(t)dW_\sigma(t) = \rho_{x,\sigma}dt$, $dW_x(t)dW_r(t) = \rho_{x,r}dt$ and $dW_r(t)dW_\sigma(t) = 0$.

Since the interest rate diffusion coefficient in (4.2.11) is not explicitly dependent on the interest rate, it is convenient to move from the spot measure, generated by the money-savings account, $B(t)$, to the *forward measure* for which the numéraire is the zero-coupon bond, $P(t,T)$:

$$F(t) := \frac{S(t)}{P(t,T)} \tag{4.2.12}$$

(details of $P(t,T)$ are given in (4.2.3)).

By Itô's lemma the dynamics of forward price, $F(t)$, in (4.2.12) are given by:

$$\begin{aligned} dF(t)/F(t) &= \left(\eta^2 B_r^2(t,T) - \rho_{x,r}\eta B_r(t,T)\sigma(t)S^{\beta-1}(t)\right)dt \\ &\quad +\sigma(t)S^{\beta-1}(t)dW_x(t) - \eta B_r(t,T)dW_r^T(t), \end{aligned}$$

combined with the volatility process for $\sigma(t)$ in System (4.2.11). Since the forward, $F(t)$, is a martingale under the T-forward measure, the forward dynamics should not contain a drift term. This implies that "dt"-terms will not appear in the (reformulated) dynamics of $dF(t)$, ie,

$$\begin{aligned} dF(t) &= \sigma(t)F^\beta(t)\left(P^{\beta-1}(t,T)dW_x^T(t) - \frac{\eta B_r(t,T)F(t)}{\sigma(t)F^\beta(t)}dW_r^T(t)\right), \\ d\sigma(t) &= \alpha\sigma(t)dW_\sigma^T(t). \end{aligned} \tag{4.2.13}$$

We assume that the interest rate is independent of the volatility process, so that a change of measure won't affect the dynamics of the variance process, $\sigma(t)$.

By factorization, Model (4.2.13) can be expressed as:

$$\begin{aligned} dF(t) &= \sigma(t)v(t)F^\beta(t)dW_F^T(t), \\ d\sigma(t) &= \alpha\sigma(t)dW_\sigma^T(t), \end{aligned} \tag{4.2.14}$$

with

$$v^2(t) := P^{2(\beta-1)}(t,T) + \left(\frac{\eta B_r(t,T)}{\sigma(t)F^{\beta-1}(t)}\right)^2 - 2\rho_{x,r}\frac{\eta B_r(t,T)P^{\beta-1}(t,T)}{\sigma(t)F^{\beta-1}(t)}. \tag{4.2.15}$$

Now, the instantaneous correlation coefficient, $\rho_{F,\sigma}$, must be determined, which, by definition, is defined as:

$$\rho_{F,\sigma} = \frac{\mathbb{C}\text{ov}(dF(t), d\sigma(t))}{\sqrt{\mathbb{V}\text{ar}(dF(t))\mathbb{V}\text{ar}(d\sigma(t))}} = \rho_{x,\sigma}\Psi(t,\sigma(t),F(t),P(t,T)), \tag{4.2.16}$$

with [2]

$$\Psi(t,\sigma(t),F(t),P(t,T)) := \frac{F^\beta P^{\beta-1}\sigma}{\sqrt{\sigma^2F^{2\beta}P^{2(\beta-1)} + \eta^2B_r^2F^2 - \rho_{x,r}\sigma F^{\beta+1}P^{\beta-1}\eta B_r}}. \tag{4.2.17}$$

[2] To simplify notation we suppress the arguments t and T here.

Model (4.2.14) with (4.2.15) and (4.2.16) is not in the well-known *plain SABR model form*, because the local volatility is not expressed only by $\sigma(t)F^{\beta}(t)$ but contains additional terms, like $v(t)$. Moreover, the instantaneous correlation between forward and volatility processes, $\rho_{F,\sigma}$, is a state-dependent function of time. In order to make use of Hagan's asymptotic formulas [46] for the plain SABR model in the current setting, we propose a projection formula in the next section.

4.3 Projection Formula for the Constant Parameter SABR-HW Model

In this section we describe the model approximations that bring the SABR-HW model in the desired SABR model form. The approximations enable us to carry out efficient calibration based on the analytic implied volatility formulas for the SABR model.

4.3.1 Projection Step for the Constant Parameter SABR-HW Model

In order to present Model (4.2.14) in *SABR form*, we need to *approximate* the additional terms from the local volatility for the forward process, $F(t)$, and simplify the associated correlation structure. In a plain SABR model the volatility process, $\sigma(t)$, is lognormal, which suggests that a *projection* of the volatility term $\sigma(t)v(t)$ in (4.2.14) on a lognormal distribution may give the desired SABR form, which is:

$$\begin{aligned} dF(t) &= \hat{\sigma}(t)F^{\beta}(t)dW_F^T(t), \quad F(0) > 0, \\ d\hat{\sigma}(t) &= \hat{\alpha}\hat{\sigma}(t)dW_{\sigma}^T(t), \quad \hat{\sigma}(0) > 0, \end{aligned} \tag{4.3.1}$$

with constant parameters $\hat{\sigma}(0)$ and $\hat{\alpha}$, and constant correlation $\hat{\rho}_{F,\sigma}$.

The term $v(t)$ in (4.2.15) depends on forward $F^{\beta-1}(t)$, volatility $\sigma(t)$ and on zero-coupon bond $P(t,T)$. With a function $v(t)$ which is independent of these state variables the expression simplifies. This can be achieved by freezing the forward and variance, $F(t)$ and $\sigma(t)$, at their initial values, ie, $F(t) \approx F(0)$, $\sigma(t) \approx \sigma(0)$, respectively, and by projecting $P(t,T)$, on its expectation, i.e. $P(t,T) \approx \mathbb{E}^T[P(t,T)|\mathcal{F}_t] =: \xi(t)$.

By this, function $v^2(t)$ is approximated by:

$$v^2(t) \approx \xi^{2(\beta-1)}(t) + \left(\frac{\eta B_r(t,T)}{\sigma(0)F^{\beta-1}(0)}\right)^2 - 2\rho_{x,r}\frac{\eta B_r(t,T)\xi^{\beta-1}(t)}{\sigma(0)F^{\beta-1}(0)}. \tag{4.3.2}$$

With help of the well-known formulas (4.2.2) and (4.2.7), we obtain the following closed-form solution for $\xi(t)$:

$$\xi(t) \quad = \quad \exp\left(A(t,T) + B_r(t,T)\mathbb{E}^T[r(t)] + \frac{1}{2}B_r^2(t,T)\mathbb{V}\mathrm{ar}^T(r(t))\right) \tag{4.3.3}$$

with $\mathbb{E}^T[r(t)]$ and $\mathbb{V}\mathrm{ar}^T(r(t))$ given by (4.2.8) and (4.2.9), respectively, and $A(t,T)$ defined in (4.2.5).

Note that $F(t)$ and $\sigma(t)$ are both martingales, due to (4.2.14), which implies that the values of their expectations *oscillate* around their initial values, $F^{\beta-1}(0)$ and $\sigma(0)$. Function $v(t)$ has become deterministic by the approximations made.

We then determine the dynamics for the *linearized* volatility structure, $\bar{\sigma}(t) := \sigma(t)v(t)$. By applying the Itô product rule we find

$$d\bar{\sigma}(t)/\bar{\sigma}(t) = v'(t)dt + \alpha v(t)dW_\sigma^T(t).$$

The $\bar{\sigma}(t)$-dynamics are thus governed by a state-dependent drift term. They are therefore not yet in standard SABR volatility form, which should not contain any drift term.

However, since $v(t)$ is approximated by a deterministic time-dependent function, the process $\bar{\sigma}(t) := v(t)\sigma(t)$ remains lognormal. The idea is now to determine the first two moments of process $\bar{\sigma}(t)$ and to project them onto the moments of the SABR volatility process in (4.3.1), defined as $d\hat{\sigma}(t) = \hat{\alpha}\hat{\sigma}(t)dW_\sigma^T(t)$, $\hat{\sigma}(0) > 0$, with parameters $\hat{\alpha}$ and $\hat{\sigma}(0)$.

The expectation and variance of process $\hat{\sigma}(t)$ in (4.3.1) are given by:

$$\mathbb{E}^T[\hat{\sigma}(t)] = \hat{\sigma}(0), \quad \mathbb{V}\mathrm{ar}^T(\hat{\sigma}(t)) = \hat{\sigma}^2(0)\left(\mathrm{e}^{\hat{\alpha}^2 t} - 1\right). \tag{4.3.4}$$

On the other hand, the expectation and the variance of $\bar{\sigma}(t) = v(t)\sigma(t)$ are given by:

$$\mathbb{E}^T[\bar{\sigma}(t)] = v(t)\sigma(0), \quad \mathbb{V}\mathrm{ar}^T(\bar{\sigma}(t)) = v^2(t)\sigma^2(0)\left(\mathrm{e}^{\alpha^2 t} - 1\right). \tag{4.3.5}$$

The main objective is to find the *effective parameters* $\hat{\sigma}(0)$ and $\hat{\alpha}$, so that the expectations and variances in (4.3.4) and (4.3.5) match.

By matching the expectations and variances, we arrive at the following optimization problem:

$$\underset{\hat{\alpha},\hat{\sigma}(0)}{\mathrm{argmin}} \begin{cases} \displaystyle\int_0^T \left(\mathbb{E}^T[\hat{\sigma}(t)] - \mathbb{E}^T[\bar{\sigma}(t)]\right) dt, \\ \displaystyle\int_0^T \left(\mathbb{V}\mathrm{ar}^T(\hat{\sigma}(t)) - \mathbb{V}\mathrm{ar}^T(\bar{\sigma}(t))\right) dt. \end{cases} \tag{4.3.6}$$

Typically the optimization problem in (4.3.6) is easy since the expectations and variances are analytic deterministic functions. In Result 4.3.1 a straightforward approach for parameter estimation is presented.

Result 4.3.1. *A simple approximation for $\hat{\sigma}(0)$ is the averaged parameter estimate, given by:*

$$\int_0^T \mathbb{E}^T[\hat{\sigma}(t)]dt = \int_0^T \mathbb{E}^T[\bar{\sigma}(t)]dt \quad \Longrightarrow \quad \hat{\sigma}(0) = \frac{\sigma(0)}{T}\int_0^T v(s)ds. \tag{4.3.7}$$

By matching the variances we obtain:

$$\int_0^T \mathbb{V}\mathrm{ar}^T(\hat{\sigma}(t))dt = \int_0^T \mathbb{V}\mathrm{ar}^T(\bar{\sigma}(t))dt \quad \Longrightarrow \quad \hat{\alpha} = \frac{1}{T}\int_0^T \sqrt{\frac{1}{s}\log\left(\frac{\mathbb{V}\mathrm{ar}^T(\bar{\sigma}(s))}{\hat{\sigma}^2(0)} + 1\right)}ds, \tag{4.3.8}$$

with $\mathbb{V}\mathrm{ar}^T(\bar{\sigma}(t))$ given in (4.3.5).

In the search for the optimal parameters $\hat{\sigma}(0)$ and $\hat{\alpha}$, so that the constant SABR-HW model in (4.2.11) *is connected* to the SABR dynamics given by (4.3.1), the correlation $\rho_{F,\sigma}$ has not yet been included. This allows us to determine the *effective correlation*, $\hat{\rho}_{F,\sigma}$, independent of the other approximations. Since the equation for correlation $\rho_{F,\sigma}$ is involved and state-dependent, we seek for a simplification also here. By freezing the volatility, and forward to their initial values, and by projection of zero-coupon bonds on their expectations, the correlation in Equation (4.2.16) can be approximated by:

$$\rho_{F,\sigma} \quad \approx \quad \rho_{x,\sigma}\Psi\left(t,\sigma(0),F(0),\mathbb{E}^T[P(t,T)|\mathcal{F}_t]\right) =: \rho_{x,\sigma}\psi(t). \tag{4.3.9}$$

In order to use Hagan's implied volatility SABR formula, the correlation must be constant, so we need to determine an *averaged* correlation, defined as:

$$\hat{\rho}_{F,\sigma} = \frac{\rho_{x,\sigma}}{T}\int_0^T \psi(s)ds, \tag{4.3.10}$$

with $\psi(t)$ given by (4.3.9).

The estimates obtained for $\hat{\alpha}$, $\hat{\sigma}(0)$ and $\hat{\rho}_{F,\sigma}$ allow us to use the Hagan implied volatility formula for the plain SABR model, as a first approximation in the calibration procedure.

Remark. Our approximations in (4.3.7) and (4.3.8) perform well for relatively short maturity times, like $T \leq 10y$. In the case of larger maturity times, ($T \geq 10y$), we prefer to solve Problem (4.3.6) by an optimization routine, for example, the *Nelder-Mead Simplex* algorithm. Furthermore, the weighted Monte Carlo method, to be discussed in Section 4, will be used to improve the calibration in those cases.

In the next subsection we check the accuracy of the approximations developed for a few parameter sets. The SABR model in (4.3.1) with the parameters $\hat{\alpha}$, $\hat{\sigma}(0)$ and $\hat{\rho}_{F,\sigma}$ will be called the *SABR-HW1* model here.

4.3.2 Numerical Validation of the SABR-HW Projection Method

We check the performance of our approximation model, SABR-HW1, in comparison with the constant parameter SABR-HW model.

The numerical experiment is set up as follows. We first prescribe a set of parameters for the constant parameter SABR-HW model in (4.2.11) for which, by means of an Euler-based Monte Carlo scheme, the European option prices are simulated. Secondly, we compute the effective parameters $\hat{\alpha}$, $\hat{\sigma}(0)$ and $\hat{\rho}_{F,\sigma}$ by solving (4.3.6) and calculating (4.3.10). These parameters are inserted in the plain SABR model (4.3.1). For the resulting SABR-HW1 model we then calculate the corresponding implied volatilities by Hagan's asymptotic formula. We compare these results and, in addition, we determine the error in the case that α, $\sigma(0)$ and $\rho_{x,\sigma}$ were used, instead of $\hat{\alpha}, \hat{\sigma}(0), \hat{\rho}_{x,\sigma}$.

The simulations have been performed with 100.000 paths and $20T$ steps. The initial stock price is set to $S(0) = 0.8$, and the zero-coupon bonds, $P(0,T)$, have been generated by the Hull-White model with constant long-term mean, $\theta = 0.03$. We also define the strikes, as in [88], with expiry times given by

$T \in \{1, 5, 10, 15\}$ years; the strikes are computed by the formula:

$$\begin{aligned} K_n(T) &= F(0) \exp\left(0.1\delta_n \sqrt{T}\right), \quad \text{with} \\ \delta_n &= \{-1.5,\ -1.0,\ -0.5,\ 0,\ 0.5,\ 1.0,\ 1.5\}, \end{aligned} \tag{4.3.11}$$

and $F(0)$ is as in (4.2.12). This formula for the strikes is convenient, since for $n = 4$, the strikes $K_4(\cdot)$ are equal to the forward prices.

par.	β	α	λ	η	$\sigma(0)$	$\rho_{x,\sigma}$	$\rho_{x,r}$
set 1	30%	30%	20%	1%	20%	-30%	20%
set 2	50%	40%	1%	0.5%	20%	-10%	40%
set 3	40%	10%	60%	0.1%	30%	-30%	-30%

Table 4.1: Sets of parameters used in the simulations.

In Table 4.1 we present three different sets of parameters. For those sets we determine the estimators $\hat{\alpha}$, $\hat{\sigma}(0)$ and $\hat{\rho}_{F,\sigma}$. They are tabulated in Table 4.2.

	estimators	T=1y	T=5y	T=10y	T=15y
set 1	$\hat{\alpha}$	29.76%	29.02%	28.34%	27.85%
	$\hat{\sigma}(0)$	20.30%	21.51%	23.01%	24.56%
	$\hat{\rho}_{F,\sigma}$	-30.08%	-30.25%	-30.30%	-30.29%
set 2	$\hat{\alpha}$	39.75%	38.93%	38.16%	37.50%
	$\hat{\sigma}(0)$	20.24%	21.28%	22.73%	24.52%
	$\hat{\rho}_{F,\sigma}$	-10.00%	-10.01%	-9.97%	-9.90%
set 3	$\hat{\alpha}$	9.97%	9.78%	9.56%	9.35%
	$\hat{\sigma}(0)$	30.26%	31.36%	32.83%	34.39%
	$\hat{\rho}_{F,\sigma}$	-30.01%	-30.02%	-30.02%	-30.02%

Table 4.2: Effective constant parameters $\hat{\alpha}$, $\hat{\sigma}(0)$ and $\hat{\rho}_{F,\sigma}$, defined in (4.3.6), and determined by solving the non-linear least squares problem (by Matlab function *lsqcurvefit*).

We measure the maximum absolute difference in the implied volatilities for Model (4.3.1) with the estimates in (4.3.7), (4.3.8), and the constant parameter SABR-HW model (4.2.11).

Two errors are defined: *error 1* is the error when the naive approach is used, ie, $\hat{\alpha} = \alpha$ and $\hat{\sigma}(0) = \sigma(0)$; *error 2* corresponds to the bias obtained using the adjusted parameters $\hat{\alpha}$ and $\hat{\sigma}(0)$ (the SABR-HW1 model). Table 4.3 presents these results.

Our approach for $\hat{\alpha}$, $\hat{\sigma}(0)$ and $\hat{\rho}_{F,\sigma}$ provides a significantly better fit to the constant parameter SABR-HW model than the model with the naively chosen parameters. For the maturity times of 1, 5, and 15 years, Figure 4.1 presents the corresponding implied volatilities.

The performance of the SABR-HW1 approximation is most accurate when the volatility for the short rate, determined by η, is not too large, ie, $\eta < 1.5\%$. Fortunately, this is very often the case in the calibration of the Hull-White

	T=1y		T=5y		T=10y		T=15y	
	error 1	error 2	error 1	error 2	error 1	error 2	error 1	error 2
set 1	0.36%	0.02%	1.63%	0.19%	2.93%	0.43%	3.94%	0.61%
set 2	0.27%	0.01%	1.38%	0.08%	2.65%	0.32%	3.74%	0.51%
set 3	0.31%	0.03%	1.63%	0.04%	3.12%	0.07%	4.31%	0.11%

Table 4.3: The absolute maximum percentage difference between implied volatilities from two different models. Both errors relate to the constant parameter SABR-HW model and respective approximations.

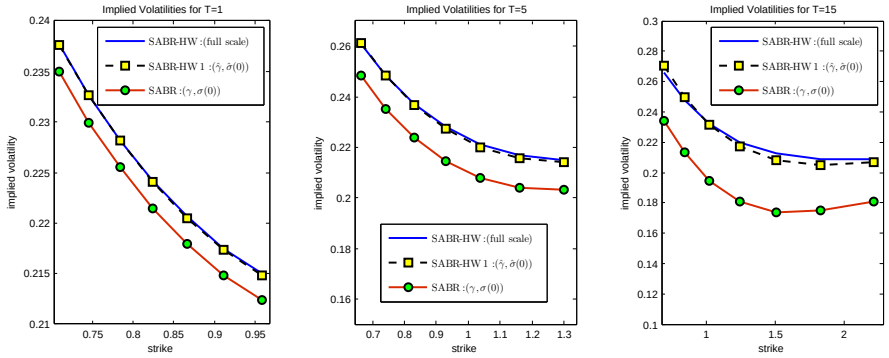

Figure 4.1: Comparison of implied Black-Scholes volatilities for European equity options and parameter set 2 in Table 4.1; For the SABR-HW model, Euler Monte Carlo was used with 100.000 paths and 20T intermediate steps.

model to market data. In the experiments we have chosen parameter $\beta \leq 50\%$ (as in the case $\beta > 50\%$ an even better accuracy is expected, because then the model behavior is closer to that of a lognormal model [31]). We also see that the correlation approximation, $\hat{\rho}_{F,\sigma}$, is close to the initial correlation $\rho_{F,\sigma}$. This is because the function $\Psi(t, \sigma(t), F(t), P(t,T))$ in (4.2.17) converges to 1 as $t \to T$, implying that $\rho_{F,\sigma} \to \hat{\rho}_{F,\sigma}$.

4.4 The Calibration Procedure

We present a calibration procedure for the SABR-HW model in three stages, and start by applying the inverse projection formulas from the previous section to calibrate the constant parameter SABR-HW model for every single maturity. In the second stage, we determine the parameters of the time-dependent functions in the dynamical SABR model in order to produce coherent model dynamics across the different maturities. In the final stage, the calibration is refined by means of a weighted Monte Carlo simulation. These stages are discussed in subsequent subsections.

4.4.1 Stage I: Parameter Projection for the SABR-HW Model

In the calibration of the SABR-HW model, the Hull-White part, which is connected to the function $\theta(t)$, is calibrated to the yield curve, whereas the parameters λ and η are calibrated to swaption prices separately. This is well-known and we refer to [25] for further information on this topic. The asset-interest rate correlation will be prescribed *a priori* based on historical data.

After the calibration of the HW model, we consider the determination of the parameters of the stochastic volatility SABR part.

One of the consequences of the projection of the constant parameter SABR-HW model onto a plain SABR model is the rapid calibration by means of Hagan's formula [107]. The projection formula, described in Section 4.3, can also be inverted numerically to retain the constant parameter SABR-HW parameters, $\sigma(0), \alpha, \rho_{x,\sigma}, \rho_{x,r}, \lambda, \theta(t), \eta$, from those of a plain SABR model, $\hat{\sigma}(0), \hat{\alpha}, \hat{\rho}_{F,\sigma}$. Since two parameters, β and $\rho_{x,\sigma}$, control the skewness of the implied volatility curve, one of them (parameter β in our case) is fixed *a-priori*, as in [94].

We briefly recall the calibration of the plain SABR model, in which for β different values are prescribed, like $\beta \in \{0.25, 0.5, 0.75, 1\}$, see for example [107], [94]. By numerical experiments we observe that different combinations of β and ρ give rise to parameter fits of very similar quality. This is especially true for short maturity implied volatilities (see Figure 4.2). The specific β which gives the best fit for both short and long maturities will be determined in the second calibration stage.

Parameter $\hat{\sigma}(0)$ is determined with the help of the at-the money (ATM) implied volatility. West [107] showed that when the forward in the plain SABR model is equal to the strike price, $F = K$, the ATM implied volatility in Hagan's formula simplifies to:

$$\sigma_{ATM} = \frac{\hat{\sigma}(0)\left(1 + \left(\frac{(1-\beta)^2}{24}\frac{\hat{\sigma}(0)^2}{F^{2-2\beta}} + \frac{1}{4}\frac{\hat{\rho}_{F,\sigma}\beta\hat{\alpha}\hat{\sigma}(0)}{F^{1-\beta}} + \frac{2-3\hat{\rho}_{F,\sigma}^2}{24}\hat{\alpha}^2\right)T\right)}{F^{1-\beta}}.$$

This equation is inverted, as in [107], to calculate $\hat{\sigma}(0)$ as a root of the cubic equation:

$$\frac{(1-\beta)^2 T}{24F^{2-2\beta}}\hat{\sigma}(0)^3 + \frac{\hat{\rho}_{F,\sigma}\beta\hat{\alpha}T}{4F^{1-\beta}}\hat{\sigma}(0)^2 + \left(1 + \frac{2-3\hat{\rho}_{F,\sigma}^2}{24}\hat{\alpha}^2 T\right)\hat{\sigma}(0) - \sigma_{ATM}F^{1-\beta} = 0.$$

For typical parameters the above cubic equation has only one real-valued root (and two imaginary roots), but it is in general possible to have three real-valued roots. In such cases, the smallest positive root should be chosen [107]. As a result of the cubic equation formulation, $\hat{\sigma}(0)$ is not a free variable anymore, but a function of the parameters $\hat{\rho}_{F,\sigma}$, $\hat{\alpha}$ and the market ATM implied volatility, σ_{ATM}. Subsequently, the calibration only has to be performed over the parameters $\hat{\rho}_{F,\sigma}$ and $\hat{\alpha}$, which can be done very efficiently.

From Eq. (4.3.7), we know that $\hat{\sigma}(0) = \sigma(0)/T\int_0^T v(s)ds$, where

$$v(t) \approx \sqrt{\xi^{2(\beta-1)}(t) + \left(\frac{\eta B_r(t,T)}{\sigma(0)F^{\beta-1}(0)}\right)^2 - 2\rho_{x,r}\frac{\eta B_r(t,T)\xi^{\beta-1}(t)}{\sigma(0)F^{\beta-1}(0)}}.$$

Table 4.4: Calibrated parameters for the SABR (parameters with hat) and SABR-HW (without hat) models, to 5 and 15 years DAX options.

β	$\hat{\sigma}(0)$	$\hat{\alpha}$	$\hat{\rho}_{F,\sigma}$	SSE	$\sigma(0)$	α	$\rho_{F,\sigma}$
				T = 5 years			
0.25	0.280	0.033	-1	3.82×10^{-4}	0.274	0.034	-0.964
0.5	0.276	0.113	-0.955	2.22×10^{-6}	0.269	0.117	-0.922
0.75	0.274	0.224	-0.824	9.94×10^{-6}	0.267	0.234	-0.800
1	0.279	0.319	-0.840	6.53×10^{-5}	0.272	0.333	-0.815
				T = 15 years			
0.25	0.402	10^{-7}	-1	1.16×10^{-4}	0.369	1.0134×10^{-7}	-0.942
0.5	0.356	10^{-7}	-1	1.75×10^{-5}	0.327	1.0134×10^{-7}	-0.937
0.75	0.328	10^{-7}	-1	1.73×10^{-6}	0.301	1.0134×10^{-7}	-0.933
1	0.319	0.033	-0.758	3.89×10^{-8}	0.292	0.033	-0.722

SSE: squared sum errors

So, $\hat{\sigma}(0)$ itself is a function of $\sigma(0)$ (since the other parameters and functions in the equation for $v(t)$ have been determined in earlier steps). Applying a numerical root-finding routine provides us with a value for $\sigma(0)$. Similarly, we can find the solution for α via Eqs. (4.3.8) and (4.3.5). With Formula (4.2.16) we can rewrite the correlation $\hat{\rho}_{F,\sigma}$ as a function of $\rho_{x,\sigma}$. The numerical inversion of this expression gives us parameter $\rho_{x,\sigma}$. After this, all parameters of the SABR-HW system, $\sigma(0), \alpha, \rho_{x,\sigma}, \rho_{x,r}, \lambda, \theta(t), \eta$, have been determined.

This stage of the calibration procedure is highly efficient as most of the evaluations are based on analytic expressions. The numerical root-finding procedure is used four times. The overall CPU time is less than a second.

We present an example of the calibration of the parameters $\hat{\rho}_{F,\sigma}$ and $\hat{\alpha}$ to the 5- and 15-years DAX options with equally spaced strike values from 40% to 220% with 10% intervals (in total 19 strikes), from 27-09-2010, based on the procedure described. The calibration has been performed with 4 sets of parameters with different a priori chosen values for β. The whole procedure (for the four sets of parameters) takes approximately 0.3 seconds CPU time on a desktop computer. The resulting parameters and squared sum errors (SSE) are presented in Table 4.4. Different values of β result in a qualitatively similar fit to the market implied volatilities. The fit of the SABR-HW model (based on Hagan's formula) to the market implied volatilities is presented in Figure 4.2.

The last three columns in Table 4.4, the columns $\sigma(0)$, α and $\rho_{F,\sigma}$, present the constant parameter SABR-HW model parameters obtained from the inversion of the projection formulas described above. We see that constant parameters produce a very good fit for individual maturity times, but the resulting parameters differ for different maturities.

4.4.2 Stage II: Calibration of the Dynamical SABR-HW Model

Calibration of the constant parameter SABR-HW model results in a series of independent implied volatility smiles across several maturities, which do not show coherent dynamics over a longer time period. We therefore describe here

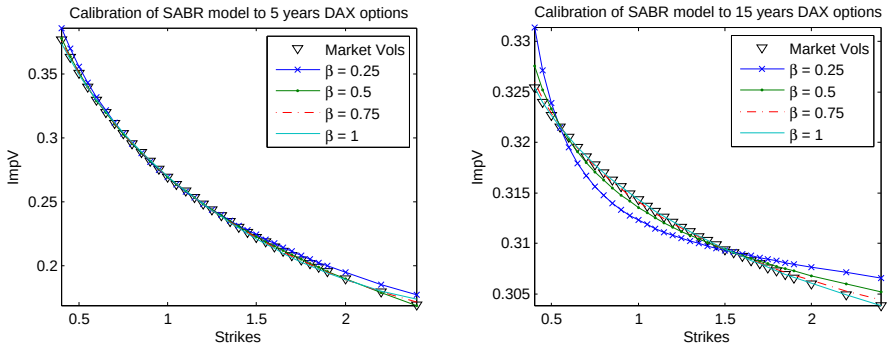

Figure 4.2: Calibration results for the SABR model with different *a priori* chosen β parameters to the implied volatilities of 5 and 15 years maturity.

the calibration of the *time-dependent* functions in the dynamical SABR-HW system (4.2.1), and start with the parameters β and $\rho_{x,\sigma}$ for this dynamical SDE system. The value for β which fits optimally for all maturities (eg, the optimal value from Table 4.4) is chosen, and simultaneously we average the calibrated correlation parameters, $\rho_{x,\sigma}$, for the corresponding β-value, over the different maturity times.

Then, the time-dependent function $h(t)$ in System (4.2.1) is parametrized in the form proposed by [93]:

$$h(t) = \big(a_1 + b_1 t\big)\exp(-c_1 t) + d_1. \tag{4.4.1}$$

The parameters a_1, b_1, c_1 and d_1 are determined, as in [95], by solving the following system of equations for all maturities T_i included in the calibration instruments:

$$\alpha^{T_i} - \frac{1}{\sigma^{T_i}(0)T_i}\sqrt{2\int_0^T g(t)^2\hat{h}(t)^2 t dt} = 0. \tag{4.4.2}$$

Here superscript T_i denotes the maturity for which the parameter is determined, and $\hat{h}(t)$ denotes the mean value of $h(\cdot)$ up to time t, ie,

$$\hat{h}(t) = \sqrt{\frac{1}{t}\int_0^t (h(s))^2 ds}.$$

Eq. (4.4.2) can be best dealt with by a numerical root finding technique.

For the time-dependent function $g(t)$, a common parametrization is:

$$g(t) = \big(a_2 + b_2 t\big)\exp(-c_2 t) + d_2, \tag{4.4.3}$$

which can also be found in [25] or [92]. We obtain a_2, b_2, c_2 and d_2 by minimizing the sum of squared errors:

$$\min_{a_2,b_2,c_2,d_2}\sum_{i=1}^{M}\left[\sigma^{T_i}(0) - \hat{g}(t)\right]^2,\ \hat{g}(t) = \sqrt{\frac{1}{T_i}\int_0^{T_i}\left[(a_2 + b_2 t)\exp(-c_2 t) + d_2\right]^2 dt},$$

Table 4.5: Parameters a_1, b_1, c_1 and d_1 for calibrated function $h(\cdot)$ and a_2, b_2, c_2 and d_2 for function $g(\cdot)$ for DAX option implied volatilities on 27-09-2010.

Parameters	$a_1(a_2)$	$b_1(b_2)$	$c_1(c_2)$	$d_1(d_2)$
$h(\cdot)$	0.5928	−0.1943	1.1936	0.1080
$g(\cdot)$	0.0949	0.0673	0.1297	0.0858

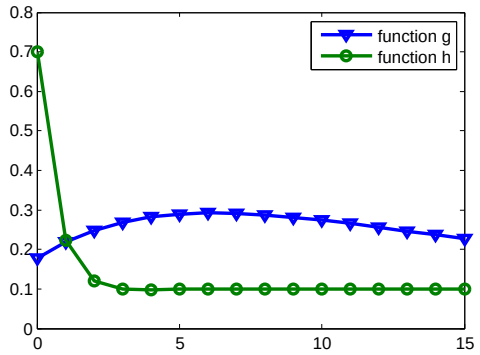

Figure 4.3: The calibrated $g(\cdot)$ and $h(\cdot)$ functions.

with M the number of option maturity times.

The time-dependent functions, $g(t)$ (4.4.3) and $h(t)$ (4.4.1), are then fitted to the parameters $\sigma(0)$ and α of all maturities obtained from the constant parameter calibration. The resulting parameters are presented in Table 4.5. The functions are illustrated in Figure 4.3. We fix the parameter d_1 in function $h(t)$ to the value of the volatility-of-volatility parameter of the longest maturity to prevent it from attaining negative values.

Remark: The Hagan implied volatility function [46] is based on asymptotic expansions that have a limited range of applicability. The formula is not exact, for example, for deep out-of-the-money strikes, especially for strikes close to zero, and for long maturities. Thus, the model dynamics that are simulated by the Monte Carlo technique may not resemble the parameters determined during calibration. In the next section, we will propose a method to eliminate such approximation error induced calibration error.

4.4.3 The Weighted Monte Carlo Technique

We employ a *non-parametric* approach to further improve the SABR-HW model calibration. The general idea is to *perturb* the weights of the individual Monte Carlo paths so that calibration instruments such as options, forwards and bonds, resemble the corresponding market prices in a better way.

Most often, one deals with ordinary Monte Carlo methods, that are governed by the fact that the same *weight* (ie, $1/N$, with N the total number of paths) is assigned to each sampled path. For a claim with a payoff, ϕ, the derivative

value, at $t = 0$, is then determined as:

$$V = \frac{1}{N} \sum_{i=1}^{N} \phi(\omega_i),$$

where ω_i denotes the i-th Monte Carlo path.

In addition, *weighted Monte Carlo methods* have been developed by [14, 12, 13], in which different 'probability' weights, $p_1, p_2, \ldots, p_N$, are assigned to the individual Monte Carlo paths. The value of the claim then reads:

$$V = \sum_{i=1}^{N} \phi(\omega_i) p_i.$$

The weights are determined so that the model values of the calibration instruments match well with the market prices and these weights should be *kept as close as possible* to the uniform weights ($p_i = \frac{1}{N}$).

We denote the market prices of M calibration instruments by $C_1, \ldots, C_M$ and represent the present values of the derivative products of the j-th calibration instruments along path ω_i by $\phi_{ij}, j = 1, 2, \ldots, M$. The first index represents the Monte Carlo path number and the expression is short notation for $\phi_{ij} \equiv \phi_j(\omega_i)$.

The path weights, or probabilities $p = (p_1, p_2, \ldots, p_N)$, have to be determined, so that

$$\sum_{i=1}^{N} p_i \phi_{ij} = C_j, \tag{4.4.4}$$

or so that the difference between the left- and right-hand side is minimized. A criterion (which is adopted here) to find these weights is the minimization of the *relative entropy* of a non-uniformly sampled probability with respect to a uniform distribution.

The concept of relative entropy is not new in computational finance. Buchen & Kelly [27] proposed the use of the minimization of relative entropy to determine the Arrow-Debreu probability in a single-period model. This method was generalized to dynamical models by [14, 12, 13].

Based on two sets of N discrete probabilities, $p = (p_1, p_2, \ldots, p_N)$ and $q = (q_1, q_2, \ldots, q_N)$, the relative entropy of p, with respect to q, is defined as

$$D(p||q) := \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right).$$

In the case of a Monte Carlo simulation, in which $q_i = 1/N, \forall i$, we have

$$D(p||q) = \sum_{i=1}^{N} p_i \log(p_i) - \sum_{i=1}^{N} p_i \log\left(\frac{1}{N}\right) = \sum_{i=1}^{N} p_i \log(p_i) + \log(N), \tag{4.4.5}$$

where we used that $\sum_{i=1}^{N} p_i = 1$. The objective is to minimize Eq. (4.4.5) under the linear constraints implied by Eq. (4.4.4). A true advantage of the relative

entropy objective function lies in the fact that Eq. (4.4.5) is convex in all p_i-values [3]. It is well-known in optimization theory that the above minimization problem has a unique global minimum solution, if it exists [22], and that the Lagrange multiplier technique determines this solution in an efficient way.

Here we present results obtained by the WMC approach (see Avellaneda et al. [13] for more details).

In probability and in information theory, the relative entropy, or Kullback-Leibler divergence, is a so-called non-symmetric measure of the difference between two probability distributions p and q. The relative entropy measure is an indication for the difference between any two models. In our case, it quantifies the consistency, or inconsistency, between the calibrated true SABR-HW model and the SABR-HW1 model obtained from the first stages of the calibration. The relative entropy distance is defined as $D(p||u)$ in Eq. (4.4.5), in which u denotes a *uniform probability of N Monte Carlo samples*. Since the term $\sum_{i=1}^{N} p_i \log p_i$ in Eq. (4.4.5) is negative (as $p_i \leq 1$), $D(p||u) \in [0, \log N]$. The minimum value, $D(p||u) = 0$, corresponds to $p_i = 1/N$, ie, the calibrated vector p equals the prior u. The maximum value, $D(p||u) = \log N$, is realized when the probability is concentrated at a single path, ie, $p_i = 1$. Consider a probability distribution which is uniformly distributed on a *subset* of paths of size N^α, with $0 < \alpha < 1$. Substitution of the corresponding probabilities, gives [13]:

$$D(p||u) = \log N + \log\left(\frac{1}{N^\alpha}\right) = (1-\alpha)\log N, \tag{4.4.6}$$

using $\sum_{i=1}^{N^\alpha} N^{-\alpha} = 1$.

The relative entropy distance can thus be connected to *the effective number of paths*, N^α, supported by the prior probability measure. The effective number of paths can be obtained, as $\alpha = 1 - D(p||u)/\log N$, with $D(p||u)/\log N \in [0, 1]$. If $D(p||u)/\log N << 1$ the number of significant paths is close to N, whereas $D(p||u)/\log N \approx 1$ is connected to a measure with *'thin support'* [13]. Thin support implies that in the calibration a large number of paths is discarded, which is inefficient from a computational point of view.

Remark. It is possible that a solution to the minimum entropy problem does not exist, when the initial problem parameters result in prices of the calibration instruments that are very different from the market prices. In such a case, the minimum entropy algorithm won't work but one may use a quadratic difference function,

$$D^Q(p||u) = \sum_{i=1}^{N}\left(p_i - \frac{1}{N}\right)^2,$$

instead of the relative entropy distance, which guarantees a solution, see [15].

4.4.4 Stage III: Calibration by Weighted Monte Carlo Method

Here, we use the DAX 1-, 5- and 10-years implied volatilities from 27-09-2010 with equally spaced strike values from 40% to 220% with 10% intervals. Af-

[3] It is straightforward to show that

$$\frac{\partial D(p||q)}{\partial p_i} = \log(p_i) + 1, \quad \frac{\partial^2 D(p||q)}{\partial p_i^2} = \frac{1}{p_i} > 0.$$

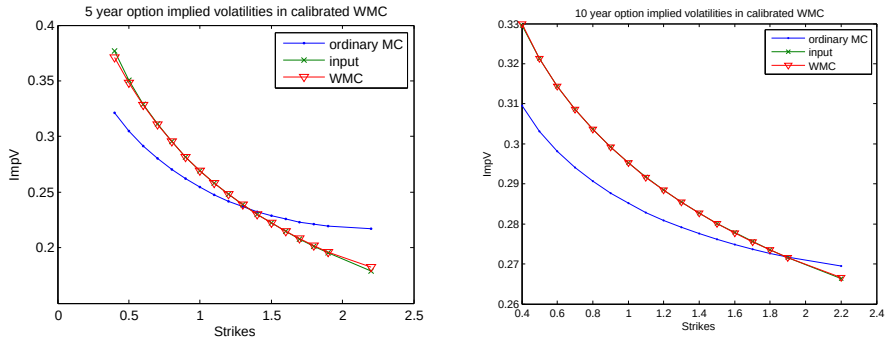

Figure 4.4: The 5 and 10 year option implied volatilities produced by WMC compared against the input market implied volatilities.

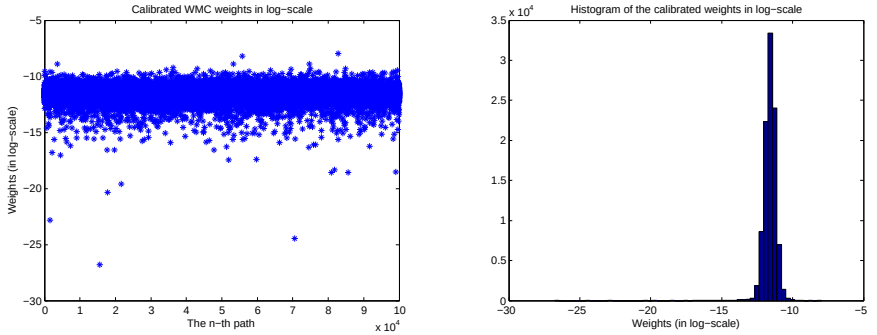

Figure 4.5: Left: a cloud plot of the 10^5 weights (in log-scale) of the WMC paths obtained after the calibration. Right: histogram of the values of the calibrated weights in log-scale.

ter the computation of the WMC weights for these financial derivatives, the weighted Monte Carlo method perfectly replicates the prices of these calibration instruments (see Figure 4.4).

We plot the resulting WMC weights in log-scale at the left-hand graph of Figure 4.5. The weights seem to be randomly distributed around their mean value of 10^{-5}. Certain paths are given a small weight, which means that these paths are effectively discarded. The histogram of the weights at the right-hand side of Figure 4.5 indicates that only a small fraction of the weights is in the left tail and most of them are distributed around the mean. The resulting effective number of paths, obtained by Eq. (4.4.6), is 9.3517×10^4, so that 6.48% of the paths are discarded. This is efficient from a computational point of view, given the excellent WMC calibration results.

Remark: The WMC calibration procedure is highly efficient, but it is non-parametric, and this may hamper its practical application. If either the model or the model parameters are not carefully chosen, too many paths will be discarded and the WMC efficiency would be lost. The WMC technique can however also be used as an *a posteriori check* of the quality of parameters obtained from calibration to Hagan's implied volatilities. If, after the first stage of calibration,

the Monte Carlo weights are equally distributed and very close to $p_i = 1/N$, this may give confidence in the quality of the first stage of calibration.

4.5 Pricing Options under the SABR-HW Model

On the basis of the calibrated SABR-HW model, we are now ready to apply a (weighted) Monte Carlo simulation for the pricing and hedging of exotic derivatives. We present an advanced time stepping scheme for the Monte Carlo simulation, leading to a *low-bias Monte Carlo simulation*. This scheme is accurate also when only a few time steps are employed. It has also been used within the WMC part of the calibration procedure, described previously.

4.5.1 Low-Bias Time Discrete Scheme

Applications of the SABR-HW model include the pricing of long-maturity equity options, equity-linked structured notes, like cliquet options, and equity-linked hybrid derivatives. Structured products usually have a long time horizon and a complicated payoff. It is difficult to find analytic approximations for these product prices, and often one has to rely on Monte Carlo methods to obtain prices and hedge ratios (eg, price sensitivities, the Greeks). If we apply an Euler discretization scheme to the SABR-HW system, the discrete bias has to be analyzed with care. For example, in Eq. (4.2.11) the drift term is stochastic and driven by two stochastic factors. In this case, an Euler approximation for the drift term,

$$\int_0^\Delta r(s)S(s)ds \approx r(0)S(0)\Delta,$$

is biased in general and a large number of time steps is required to reach an acceptable level of accuracy. An Euler scheme may therefore be inefficient for pricing long-term equity-linked structured products.

Here we adapt the low-bias MC scheme proposed for the SABR model in [32] to discretize the SABR-HW model. The approach is to map the asset price process onto a square root process by a series of spatial and time transformations. In the Monte Carlo simulation we will draw samples from the analytic distribution function of the square root process (ie, the non-central chi square distribution), as described in Section 4.5.2, and, in full detail, in [32].

The SABR model considered in [32] was developed for a system without drift term. A stochastic interest rate can be incorporated by a technique described in [40] (pp. 28), which was introduced for a Constant Elasticity of Variance (CEV) process with drift term, $rS(t)dt$:

$$dS^{(r)}(t) = rS^{(r)}(t)dt + \sigma S^{(r)}(t)^\beta dW(t). \tag{4.5.1}$$

We use the superscript (r) to distinguish the process with drift from the process without drift, which does not have a superscript. The distribution of the CEV process in (4.5.1) can be obtained from its sister without drift via a time change:

$$S^{(r)}(t) = e^{rt}S\big(\tau(t)\big), \quad \tau(t) = \frac{1}{2r(\beta-1)}\big(e^{2r(\beta-1)t} - 1\big). \tag{4.5.2}$$

The validity of this transformation can easily be explained as the limit $r \to 0$ recovers the original clock, ie,

$$\lim_{r\to 0} \tau(t) = \lim_{r\to 0} \frac{1}{2r(\beta-1)}\big(e^{2r(\beta-1)t} - 1\big) = t.$$

This result is not restricted to the constant interest rate case. The time transformation in (4.5.2) also applies to stochastic interest rates [40].

In an SDE system with a stochastic interest rate, the time transformation is different for each path, due to the randomness of the rates:

$$\tau(t,\omega) = \frac{\Delta}{2(\beta-1)\int_0^\Delta r(s)ds}\big(e^{2(\beta-1)\int_0^\Delta r(s)ds} - 1\big), \tag{4.5.3}$$

where $\omega \in \Omega$ denotes a random scenario. Expression (4.5.3) suggests that the pathwise time transformation, $\tau(t,\omega)$, can be determined without all details of the interest rate path, ω, as long as we have knowledge of $\int_0^\Delta r(s)ds$ at each path.

Following the arguments by [3], we focus on the evolution of the system over a small time interval $[0,\Delta]$ and repeat the one-period Δ-scheme to produce a complete time discrete path. Note that we consider the SDE system in the τ-time scale, so that the time interval for the system equals $[0,\tau(\Delta,\omega)]$.

The CEV system in (4.5.2) is then simulated on a time scale $\tau(t,\omega)$, induced by a stochastic interest rate. Subsequently, the result is multiplied by an exponentially integrated interest rate:

$$\begin{aligned} S\big(\tau(\Delta,\omega)\big) &= S(0) + \int_0^{\tau(\Delta,\omega)} \sigma\big(\tau(s,\omega)\big)S\big(\tau(s,\omega)\big)^\beta dW_x\big(\tau(s,\omega)\big), \\ S^{(r)}(\Delta) &= \exp\big(\int_0^\Delta r(s)ds\big)S\big(\tau(\Delta,\omega)\big). \end{aligned}$$

Although a transformed model based on time changes may not be intuitively clear, it is easy to implement numerically. We just replace the uniform time interval Δ by τ, based on the computation on each interest rate path.

4.5.2 Discretization of the SDE System

For the SABR-HW system, we also consider first the system without drift with time scale $\tau(t,\omega)$, and a low-bias Monte Carlo simulation scheme (Section 4.5.2, [32]). We then multiply the result by the exponentially integrated interest rate.

Let us first consider a system of three Brownian motions, for stochastic interest rate, equity price and stochastic volatility respectively, correlated with each other by the following correlation matrix:

$$\begin{pmatrix} 1 & \rho_{x,\sigma} & \rho_{x,r} \\ \rho_{x,\sigma} & 1 & 0 \\ \rho_{x,r} & 0 & 1 \end{pmatrix}.$$

Applying the Cholesky decomposition gives

$$dW_x(\tau) = \rho_{x,\sigma} dW_1(\tau) + \rho_{x,r} dW_2(\tau) + \sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\, dU(\tau), \tag{4.5.4}$$

$$dW_\sigma(\tau) \equiv dW_1(\tau), \tag{4.5.5}$$

$$dW_r(\tau) \equiv dW_2(\tau), \tag{4.5.6}$$

where the Brownian motions $W_1(\tau), W_2(\tau)$ and $U(\tau)$ are mutually independent.

In this subsection, we describe the low-bias time discretization scheme to simulate the plain SABR system running on the changed time scale $\tau(t,\omega)$ dependent on the stochastic interest rate:

$$\begin{aligned} dS(\tau) &= \sigma(\tau) S(\tau)^\beta dW_x(\tau), \\ d\sigma(\tau) &= \alpha\sigma(\tau) dW_\sigma(\tau), \end{aligned}$$

where we denote $\tau(t,\omega)$ simply by τ.

Based on an argument by [100], we consider the invertible transformation of variables, $X(\tau) = S(\tau)^{1-\beta}/(1-\beta)$, $\beta \neq 1$, so that

$$dX(\tau) = \sigma(\tau) dW_x(\tau) - \frac{\beta\sigma(\tau)^2}{(2-2\beta)X(\tau)} d\tau. \tag{4.5.7}$$

We substitute Eqn. (4.5.4) into Eqn. (4.5.7) and integrate from 0 to $\tau(\Delta,\omega)$, which gives

$$\begin{aligned} X(\tau(\Delta,\omega)) =& X(0,\omega) + \rho_{x,\sigma}\int_0^{\tau(\Delta,\omega)} \sigma(s) dW_1(s) + \rho_{x,r}\int_0^{\tau(\Delta,\omega)} \sigma(s) dW_2(s) - \\ & \int_0^{\tau(\Delta,\omega)} \frac{\beta\sigma(s)^2}{(2-2\beta)X(s)} ds + \sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\int_0^{\tau(\Delta,\omega)} \sigma(s) dU(s). \end{aligned}$$

In [32] it was shown that $\int_0^{\tau(\Delta,\omega)} \sigma(s) dW_1(s) = \big(\sigma\big(\tau(\Delta,\omega)\big) - \sigma(0)\big)/\alpha$. It is also not difficult to show that $\int_0^{\tau(\Delta,\omega)} \sigma(s) dW_2(s)$ is a Gaussian integral with variance $\int_0^{\tau(\Delta,\omega)} \sigma(s)^2 ds$, because of the independence of W_1 and W_2. Hence, we can replace the Gaussian integral:

$$\int_0^{\tau(\Delta,\omega)} \sigma(s) dW_2(s) = \zeta\big(\tau(\Delta,\omega)\big) W_2\big(\tau(\Delta,\omega)\big), \tag{4.5.8}$$

where we have defined $\zeta\big(\tau(\Delta,\omega)\big) := \sqrt{\int_0^{\tau(\Delta,\omega)} \sigma(s)^2 ds / \tau(\Delta,\omega)}$.

Based on these results, we can sample the SABR system without drift term and use the conditional scheme proposed in [32], conditional on the terminal volatility, the integrated volatility and the realization of W_2:

$$\begin{aligned} X(\tau(\Delta,\omega)) =& X(0) + \frac{\rho_{x,\sigma}}{\alpha}\big\{\sigma\big(\tau(\Delta,\omega)\big) - \sigma(0)\big\} + \rho_{x,r}\zeta\big(\tau(\Delta,\omega)\big) W_2\big(\tau(\Delta,\omega)\big) - \\ & \int_0^{\tau(\Delta,\omega)} \frac{\beta\sigma(s)^2}{(2-2\beta)X(s)} ds + \sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\int_0^{\tau(\Delta,\omega)} \sigma(s) dU(s), \end{aligned}$$

where we have used (4.5.8).

Conditional on the volatility, the integrated variance and Brownian motion W_2, process X is a *shifted Bessel process*, $\tilde{X}$, with dynamics:

$$d\tilde{X}(\tau) := \sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\sigma(\tau)dU(\tau) - \frac{\beta\sigma(\tau)^2}{(2-2\beta)\tilde{X}(\tau)}d\tau,$$
$$\tilde{X}(0) = X(0) + \frac{\rho_{x,\sigma}}{\alpha}\left\{\sigma\big(\tau(\Delta)\big) - \sigma(0)\right\} + \rho_{x,r}\zeta\big(\tau(\Delta)\big)W_2\big(\tau(\Delta)\big).$$

We define another change of variables, $Y(\tau) := \tilde{X}(\tau)^2$, and apply Itô's lemma:

$$\begin{aligned}
dY(\tau) =& 2\tilde{X}(\tau)d\tilde{X}(\tau) + d\tilde{X}(\tau)^2 \\
=& 2\tilde{X}(\tau)\left(\sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\sigma(\tau)dU(\tau) - \frac{\beta\sigma(\tau)^2}{(2-2\beta)\tilde{X}(\tau)}d\tau\right) \\
&+ \left(1-\rho_{x,\sigma}^2-\rho_{x,r}^2\right)\sigma(\tau)^2 d\tau \\
=& 2\sqrt{Y(\tau)}\sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\sigma(\tau)dU(\tau) + \\
&\left(\frac{1-2\beta-(\rho_{x,\sigma}^2+\rho_{x,r}^2)(1-\beta)}{(1-\beta)(1-\rho_{x,\sigma}^2-\rho_{x,r}^2)}\right)\left(1-\rho_{x,\sigma}^2-\rho_{x,r}^2\right)\sigma(\tau)^2 d\tau. \quad (4.5.9)
\end{aligned}$$

Let us define the *time change*, $\nu\left(\tau(\Delta,\omega)\right) = \left(1-\rho_{x,\sigma}^2-\rho_{x,r}^2\right)\int_0^{\tau(\Delta,\omega)}\sigma(s)^2 ds$.

Due to the independence of the Brownian motion, U, and the volatility process, a Brownian motion under clock $\nu(\tau(\cdot,\cdot))$ has the same distribution as $\sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\int_0^{\tau(\Delta,\omega)}\sigma(s)dU(s)$, ie,

$$U\Big(\nu\big(\tau(\Delta,\omega)\big)\Big) \equiv \int_0^{\nu(\tau(\Delta,\omega))} dU(s) = \sqrt{1-\rho_{x,\sigma}^2-\rho_{x,r}^2}\int_0^{\tau(\Delta,\omega)}\sigma(s)dU(s).$$

We substitute this time change into Eq. (4.5.9), and obtain:

$$dY(\nu(\tau)) = 2\sqrt{Y(\nu(\tau))}dU\big(\nu(\tau)\big) + \left(\frac{1-2\beta-(\rho_{x,\sigma}^2+\rho_{x,r}^2)(1-\beta)}{(1-\beta)(1-\rho_{x,\sigma}^2-\rho_{x,r}^2)}\right)d\nu(\tau), \quad (4.5.10)$$

which gives us a time changed squared Bessel process of dimension $\delta = \frac{1-2\beta-(\rho_{x,\sigma}^2+\rho_{x,r}^2)(1-\beta)}{(1-\beta)(1-\rho_{x,\sigma}^2-\rho_{x,r}^2)}$ with initial value $Y(0) = \tilde{X}(0)^2$. The time change depends on the interest rate path via $\tau(\cdot,\cdot)$ and depends on the volatility path as a result of the time change function $\nu(\cdot)$. Each change of time is conditioned on specific path information of the volatility and the interest rate paths.

The stochastic volatility induced time change, $\nu(\cdot)$, can be computed by an asymptotic expansion, see [32] (Sect. 3.4). In order to determine the change of time scale $\tau(\cdot,\cdot)$, we have to compute integral $\int_0^\Delta r(s)ds$ in Eq. (4.5.3). There are several ways to approximate this integral. A straightforward approach is the following discrete approximation:

$$\int_0^\Delta r(s)ds \approx \Delta\left[w_1 r(0) + w_2 r(\Delta)\right].$$

The constants w_1, w_2 can be chosen in different ways, for example, in an Euler scheme we have $w_1 = 1, w_2 = 0$. A central discretization employs $w_1 = w_2 = \frac{1}{2}$. This latter scheme is computationally efficient and sufficiently accurate if the underlying stochastic process is slowly varying. This is the case for these interest rate processes. The calibrated interest rate process usually includes a volatility parameter $\eta < 1\%$.

The CDF of process Y can be obtained by the properties of squared Bessel processes. By a mapping:

$$h : s \to \frac{s^{2-2\beta}}{(1-\beta)^2}, \quad s \geq 0,$$

with its inverse

$$h^{-1} : y \to \frac{y^{2-2\beta}}{(1-\beta)^2}, \quad y \geq 0,$$

we can define $S(\tau(\Delta,\omega)) = h(Y(\nu(\tau(\Delta,\omega))))$ and $Y(0) = h^{-1}(S(0)) = S(0)^{2(1-\beta)}/(1-\beta)^2$.

So, we have the following proposition:

Proposition 3 (Cumulative distribution for the conditional SABR process)**.** *The cumulative distribution for $S(\tau(\Delta,\omega))$, conditional on $\sigma\big(\tau(\Delta,\omega)\big)$ and $\int_0^{\tau(\Delta,\omega)} \sigma_s^2 ds$ with an absorbing boundary at $S = 0$ reads*

$$\Pr\big[S\big(\tau(\Delta,\omega)\big) \leq x | S(0)\big] = 1 - \chi^2(a; b, c), \tag{4.5.11}$$

where

$$\begin{aligned}
a &= \frac{1}{\nu\big(\tau(\Delta,\omega)\big)}\Big[\frac{S(0)^{1-\beta}}{(1-\beta)} + \frac{\rho}{\alpha}\big\{\sigma\big(\tau(\Delta,\omega)\big) - \sigma(0)\big\} \\
&\quad + \rho_{x,r} \cdot \zeta\big(\tau(\Delta,\omega)\big) W_2\big(\tau(\Delta,\omega)\big)\Big]^2, \\
b &= 2 - \frac{1 - 2\beta - (\rho_{x,\sigma}^2 + \rho_{x,r}^2)(1-\beta)}{(1-\beta)(1-\rho_{x,\sigma}^2 - \rho_{x,r}^2)}, \\
c &= \frac{S\big(\tau(\Delta,\omega)\big)^{2(1-\beta)}}{(1-\beta)^2 \nu\big(\tau(\Delta,\omega)\big)}, \\
\nu\big(\tau(\Delta,\omega)\big) &= (1-\rho^2)\int_0^{\tau(\Delta,\omega)} \sigma(s)^2 ds.
\end{aligned} \tag{4.5.12}$$

and $\chi^2(x; \delta, \lambda)$ is the non-central chi square cumulative distribution function for random variable x with non-centrality parameter λ and degree of freedom δ.

Proof: The proof proceeds along the lines of the work in [53] and [32].

Summary of the algorithm The algorithm to sample the SABR-HW system for a time interval $[0, \Delta]$ consists of the following steps:

1. Draw samples of $r(\Delta)$ from a normal distribution with mean and variance defined by Eq. (4.2.8) and (4.2.9), respectively;

2. Apply a drift interpolation $\int_0^\Delta r(s)ds \approx \Delta\left(\frac{1}{2}r(0) + \frac{1}{2}r(\Delta)\right)$, for the quantity in Formula (4.5.3) for the pathwise (stochastic interest rate induced) rescaled time step $\tau(\Delta,\omega)$;

Table 4.6: Estimated 10 year ATM call option prices for Cases I, II and III.

	Case I		Case II		Case III	
	T = 10					
Δ	Low-bias	Euler	Low-bias	Euler	Low-bias	Euler
1	0.4653	0.4699	0.4731	0.4771	0.5485	0.5374
1/2	0.4680	0.4697	0.4680	0.4703	0.5473	0.5382
1/4	0.4664	0.4662	0.4665	0.4691	0.5420	0.5479
1/8	0.4666	0.4664	0.4659	0.4681	0.5440	0.5451
1/16	0.4669	0.4675	0.4644	0.4676	0.5430	0.5449
1/32	0.4669	0.4673	0.4633	0.4667	0.5431	0.5432
Reference	0.4671		0.4634		0.5431	

3. Conditional on $\sigma\big(\tau(\Delta,\omega)\big)$, draw samples of $\int_0^{\tau(\Delta,\omega)}\sigma(s)^2ds$ by the method proposed in Section 3.3.4;

4. Conditional on $\sigma\big(\tau(\Delta,\omega)\big)$ and $\int_0^{\tau(\Delta,\omega)}\sigma(s)^2ds$, draw samples of the dynamics without drift, S, in the time scale τ from the non-central chi square distribution, described in Proposition 3;

5. Find the asset price process with drift, $S^{(r)}\big(\tau(\Delta,\omega)\big)$, by multiplying the estimated $\int_0^{\Delta}r(s)ds$ (Step 2) with $S\big(\tau(\Delta,\omega)\big)$.

Numerical Experiment with European Options

We consider the pricing of European options in the SABR-HW model by the low-bias Monte Carlo method. We focus on a call option, maturing at time T with strike price K, and denote the exact option price at initial time by $C(K,0)$. It is approximated by:

$$\hat{C}(K,0) = P(t,T)\mathbb{E}^T\big[\big(S(T)-K\big)^+\big].$$

where $\hat{C}(K,0)$ is typically not equal to $C(K,0)$, and we define the bias, e, of the simulation as a function of the time step, Δ, and analyze its behaviour.

We will use the parameter sets I, II and III in Table 4.1 to carry out the numerical experiments for the call options with $T = 10$ and ATM strikes. As the benchmark, we apply a Monte Carlo simulation based on the Euler discretization with a large number of time steps, 400 time steps per year, and report the average of three runs as the reference prices. For Sets I, II and III, the option prices obtained by the low-bias Monte Carlo method are listed in Table 4.6 for time steps Δ ranging from 1 to 32 steps a year.

We plot the differences between the MC estimate and the reference values for different step sizes in Figure 4.6, from which we see that the low-bias scheme is advantageous to the Euler scheme with respect to its low bias. By increasing the number of time steps the low-bias scheme produces smaller errors than the Euler scheme for all parameter settings.

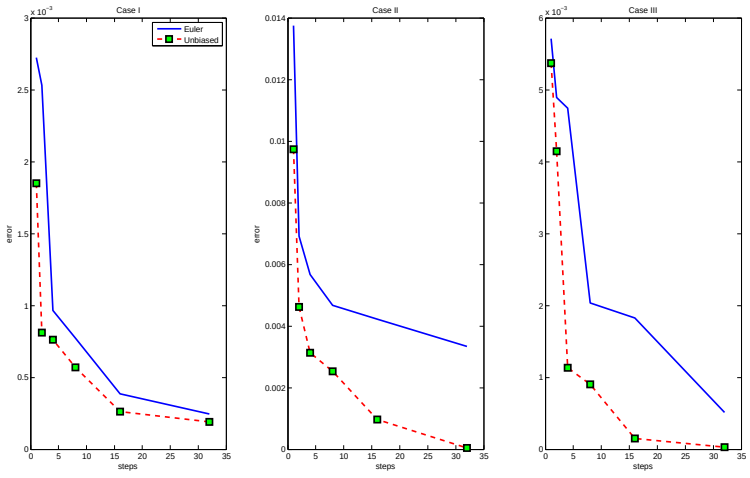

Figure 4.6: The MC error as a function of the number of time steps. Note the different scaling of the three figures.

Time-Dependent Parameters

We can also discretize System (4.2.1) with the low-bias scheme by assuming that the functions can be approximated by piecewise constant functions on $[t, t+\Delta]$ with value $\bar{\alpha}$. According to the arguments by [3] and [39] (pp. 130) we then use $\bar{\alpha} = (\alpha(t) + \alpha(t+\Delta))/2$, which leads to a modification of System (4.2.1) that can easily be simulated by the low-bias scheme. More precisely, the volatility process in the dynamical SABR-HW system is sampled by the following formula

$$\Sigma(t+\Delta) = \bar{g} \cdot k(t) \cdot \exp\left(-\frac{1}{2}\bar{h}^2\Delta + \bar{h}Z\sqrt{\Delta}\right), \; k(0) = 1, \tag{4.5.13}$$

where

$$\bar{g} = \frac{g(0,t) + g(0,t+\Delta)}{2}, \; \bar{h} = \frac{h(0,t) + h(0,t+\Delta)}{2}.$$

As a result, the formula for the integrated variance has to be adapted as well: it has to be multiplied by the factor $\bar{g}^2$.

4.6 Conclusion

In this chapter we have presented the dynamical SABR-HW model as an alternative model for pricing long-maturity equity options and equity-interest rate hybrid products. We have defined the model, introduced its building blocks and described several issues for the practical application of the SABR-HW model, like model calibration and option pricing. At several places we have presented numerical techniques that are not commonly used by the financial industry, like the low-bias discretization scheme and a weighted Monte Carlo technique to enhance the calibration.

In particular, we have proposed an invertible projection formula for the constant parameter SABR-HW model connecting it to the plain SABR model. The basis for this projection was a change of measure, to the T-forward measure, and a linearization. The projection formula greatly simplifies the calibration of the SABR-HW model.

The inversion of the projection formula serves as a first step in the calibration procedure, ie, it gives a rapid and fairly accurate approximation of the constant SABR-HW parameters at each maturity. Based on these parameters, we have defined time-dependent functions in the dynamical SABR-HW model, that are consistent with the market implied volatilities for all maturities.

In the final calibration step, non-uniform Monte Carlo weights have been determined in such a way that the implied volatilities from the market and those generated by the Monte Carlo paths of SABR-HW model match optimally. The overall calibration procedure is highly efficient and accurate.

Exotic contracts have then been priced using the weighted Monte Carlo paths generated by a low-bias time discretization scheme of the dynamical SABR-HW model. An advantage of the low-bias scheme is that accurate Monte Carlo results can be obtained for large time steps. This is particularly useful when long-maturity options are considered.

CHAPTER 5

On a Calibration Technique for the Term Structure of Skew Stochastic Volatility Libor Market Model

This chapter is adapted from the article
"On a Calibration Technique for the Term Structure of Skew Stochastic Volatility Libor Market Model"
submitted for publication.

In the contemporary swaption market, especially after the 2008 crisis, we observe a higher level of time-inhomogeneity in the skews of the implied swaption volatilities. To model this feature, we choose the *Term Structure of Skew Libor Market Model* (TSS-LMM), proposed by Piterbarg [86], for the forward rate dynamics. We investigate the accuracy and numerical efficiency of the Markov projection technique and parameter averaging approximations in the context of TSS-LMM. The 'freezing projection' formula (which can be seen as Markov projection in its simplest form) often provides satisfactory accuracy at reasonable computational costs. Our contribution here is a calibration technique based on a *convex optimization* formulation to calibrate the time-dependent skew parameters in the TSS-LMM.

5.1 Introduction and Motivation

The valuation of exotic interest rate (IR) derivatives requires flexible model specifications and *richer IR dynamics* than those that are present in low-dimensional Markov models, like the one- or two-factor Hull-White or Cox-Ingersoll-Ross (CIR) models. Certain exotic derivatives may be sensitive to the *joint evolution*

of multiple points on the yield curve [9], whereas other exotics may be sensitive to the term structure of the swaption implied volatilities across different tenors. Besides, many exotic products have an implicit exposure to the vanilla option market, which is not always directly obvious. In these cases, it is difficult, if not impossible, to identify a small, representative set of plain vanilla options to which a low-dimensional model should be calibrated. As a direct consequence, these models may often perform unsatisfactorily for market risk management, as they do not adequately model the underlying risk factors relevant to the exotic options. Usually models that are sufficiently rich to explain the dynamics of an entire yield curve and allow for volatility calibration to a sufficiently large set of European options which can describe the volatility exposure of these exotic options accurately [9]. On the other hand, low-dimensional models can fit volatility information for selected LIBOR and swap rates, but this can lead to extreme parameter values, which implies an unrealistic evolution of the volatility structure.

The 2008 crisis seems to have brought two essential changes to the interest rate swaption volatilities, for which a flexible modelling framework seems to be even more favourable compared to low-dimensional models:

- Previously, it was seen that the term structure of the swaption volatilities possessed a downward sloping shape – the ATM swaption volatilities decreased in expiry and tenor. However, in the current market, this does not seem to be the case. The ATM swaption volatilities decrease first for short to medium expiry dates, and then increase for longer expiry times, e.g. 10, 15 and 20 years (see LHS graph of Figure 5.1). This may be due to strong demand from long term swaptions by asset liability management (ALM) investors to hedge their long-dated exposures in the wake of the financial crisis.

- Forward rates have decreased while the volatility skew has increased significantly [60], and the increase in the skew of *long tenor, short expiry swaptions* is especially pronounced. The reason for this phenomenon is unclear, but it may be related to the widening of the tenor basis spread (see RHS graph of Figure 5.1).

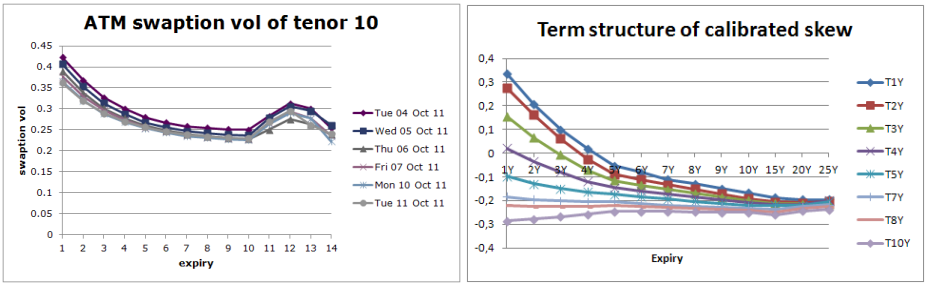

Figure 5.1: LHS: The term structure of ATM swaption volatility with tenor 10 years observed at different spot dates. RHS: The term structure of skew obtained by calibrating a constant parameter displaced Heston model to swaption prices with different combinations of expiry/tenor ($T1Y$ stands for tenor 1 year).

Flexible models are usually of multi-factor form. Notable examples are multi-factor short rate models [25], multi-factor Quasi-Gaussian models [2] or LIBOR market models [24] [57].

This paper focuses on the LIBOR market model (LMM) framework and its extensions. Since its introduction by Brace, Gatarek, and Musiela [24] and Jamshidian [57], the LMM framework enjoys popularity among practitioners, mainly due to the fact that the model primitives can be directly related to market observables, e.g. forward LIBOR rates and caplet implied Black volatilities. In this framework, closed form solutions for caps and European swaptions (although not in the same formulation) can be obtained as the LMM is based on the assumption that the discrete forward LIBOR rate follows a lognormal distribution under its own numeraire.

Since the mid-1990s, the volatility skew and smile have become more apparent in market quotes, and extensions have been introduced to model those features. Notable examples are the skew consistent LMM introduced by Andersen & Andreasen [4] and the stochastic volatility LMM from Andersen & Brotherton-Ratcliffe [6], Joshi & Rebonato [61], Piterbarg [86] and Rebonato & White [93]. Piterbarg [86] proposed a flexible modelling framework introducing time-dependent volatility and skew parameters to fit the market quotes across an entire swaption cube.

In contrast to the theoretical improvements in interest rate modelling, the development of calibration procedures suited for increasingly complicated interest rate models does not attract much research attention. In [86] a three step calibration procedure, based on a least squares minimization for TSS-LMM, was introduced. Andersen & Piterbarg [9] added details and gave a systemic introduction to the calibration procedure for the TSS-LMM, but the main idea was the same as in the original article [86]. Amin [1] proposed a procedure based on an analytically derived Jacobian matrix to speed up the computations. These approaches are able to produce calibration results, however, they give rise to highly non-linear optimization problems that are not always efficiently solved (although [9] claims otherwise). A consequence of this non-linear inverse problem is the non-uniqueness of the solution. With the industrial practice of daily re-calibration of a model, a small change in the calibration instruments could result in significant changes in the model parameters, leading to unstable hedges and high hedging costs not due to market changes but because of numerical issues.

Here, we propose an alternative formulation of the TSS-LMM calibration problem and impose stability of the solution. For the LIBOR volatilities, we assume a parametric time-homogeneous function for the time-dependent volatility. Our contribution is a convex optimization formulation of the skew calibration problem: We set up the high-dimensional model calibration problem to determine a parameter vector which minimizes a positive semi-definite system and satisfies a number of linear constraints. The advances in convex optimization (see Boyd and Vandenberghe [23]) have led to algorithms that solve such problems highly efficiently with a complexity comparable to that for linear programs [84]. A consequence of the convex problem formulation is that the optimal solution is global and unique [23], which may lead to fewer fluctuations in the daily calibration results.

This paper is organized around the effective approximation of swaption prices in the TSS-LMM and the calibration technique for the model. In Section 2, we

briefly introduce the LMM framework and define the TSS-LMM. In Section 3, we focus on the model mapping techniques, which relate the time-dependent parameters in TSS-LMM to the swaption prices. Two approximation techniques will be discussed and are compared in the section that follows, based on their accuracy and efficiency. In Section 4, we describe the calibration method for the TSS-LMM in detail, formulate the skew calibration problem as a convex optimization problem and present corresponding calibration results.

5.2 Term Structure of Interest Rate and LIBOR Market Models

Before discussing the finer details of the LMM framework, we clarify some important concepts and define necessary notation; for more details, see [18] or [8].

We start by defining a tenor structure

$$t < T_0 < T_1 < T_2 < \cdots < T_N, \quad \delta_i = T_{i+1} - T_i, \quad i = 0, 1, \ldots, N-1.$$

Interval δ_i is set to be either 3 or 6 months. For simplicity, we assume that all day-count fractions are equal. Note that the model's tenor structure, $\{T_i\}_{i=1}^N$, can in principle be specified freely, and can be different from the tenor structure of the contract. Most of the time, we will choose the tenor structure as close as possible to the tenor structure of the product that we aim to price.

On this tenor structure, we define the price of a zero coupon bond, paying one dollar at time T_i observed at time t, $t \leq T_i$, to be $P(t, T_i)$. The forward LIBOR rate, $F(t, T_{i-1}, T_i)$, is the pre-determined fixed interest rate on a forward rate agreement (FRA) contract at time t with maturity T_i. With δ_i the day-count factor for the period spanned by the FRA contract, we have

$$F(t, T_i, T_{i+1}) = \frac{P(t, T_i) - P(t, T_{i+1})}{\delta_{i+1} P(t, T_{i+1})}, \quad 0 \leq i < N. \tag{5.2.1}$$

We use the notation $F_i(t)$ for $F(t, T_i, T_{i+1})$, and the set of discrete forward rates $\{F_i(t)\}_{i=1}^N$ is the set of state variables describing the state of the economy. The forward rates are the primitive variables for which we wish to specify dynamics in the LMM framework.

An *interest rate swap* is an agreement between two counterparties to exchange cash flows over a certain tenor structure at fixed payment dates, $T_1, T_2, \ldots, T_N$. These parties exchange fixed cash flows for floating ones, where the floating legs typically depend on LIBOR rates. The fixed payment, which is the swap rate, is determined in such a way that the net present value of the contract is equal to zero at the inception of the contract, i.e.

$$\sum_{i=n}^{m-1} P(t, T_{i+1}) \delta_i \big(F_i(t) - S\big) = 0$$

with some expiry n and maturity m. The *swap rate*, $\mathrm{SR}_{n,m}(t)$, is the S-value which solves the above equation, i.e.

$$\mathrm{SR}_{n,m} = \frac{\sum_{i=n}^{m-1} P(t, T_{i+1}) \delta_i F_i(t)}{\sum_{i=n}^{m-1} \delta_i P(t, T_{i+1})}.$$

By substituting Eq. (5.2.1) into the above identity, the sum in the denominator cancels out as a telescope sum, resulting in

$$\mathrm{SR}_{n,m} = \frac{P(t,T_n) - P(t,T_m)}{\sum_{i=n+1}^{m} \delta_i P(t,T_i)}. \tag{5.2.2}$$

5.2.1 LMM Framework and Extensions

We outline the general LMM framework and also a series of extensions to the original lognormal framework by Brace et al. [24] and Jamshidian [57], to model the observed market volatility skew and smile patterns accurately.

Given the definition of the i-th forward LIBOR rate in Eq. (5.2.1), $F_i(t)$ is a martingale under the forward measure, associated with bond $P(t,T_{i+1})$. In other words, the stochastic process of $F_i(t)$ does not have a drift term under this measure.

Brace et al. [24] and Jamshidian [57] specify the dynamics of the forward LIBOR rate, $F_i(t),\ i = 1,\ldots,N$, under its natural martingale measure as a lognormal process:

$$\frac{dF_i(t)}{F_i(t)} = \sigma_i(t)dW_i(t), \quad dW_i(t)dW_j(t) = \rho_{i,j}dt,\ i,j = 1,2,\ldots,N-1, \tag{5.2.3}$$

where $W_i(t)$ is a Wiener process under the forward measure $\mathbb{Q}^i$, associated with the numeraire asset $P(t,T_{i+1})$, and $\sigma_i(t)$ represents the volatility of the forward rate process.

The lognormal forward rate assumption is consistent with the widely used Black formula for caplets.

Since the mid-nineties, a volatility skew has appeared in the implied volatilities of caplets, as a monotonically decreasing function of the strike. This dependency of the forward volatility on the level of the forward rate is not modelled well by the lognormal forward rate assumption [82].

Several extensions to the LMM framework to incorporate the skew feature in the model have been proposed in the literature. Notable examples include the constant elasticity of variance (CEV) [4] and displaced diffusion [96] extensions of the LMM.

The skew-consistent forward LIBOR model, introduced by Andersen & Andreasen [4] specifies a stochastic process, without drift, for the forward LIBOR rate, $F_i(t)$, under the forward measure $\mathbb{Q}^i$, associated with numeraire asset $P(t,T_{i+1})$, as

$$dF_i(t) = \sigma_i(t)F_i(t)^{\beta}dW_i(t), \quad 0 \le \beta \le 1, \quad i = 1,2,\ldots,N-1, \tag{5.2.4}$$

where $W_i(t)$ is the Wiener process under the forward measure $\mathbb{Q}^i$, and $\sigma_i(t)$ represents the volatility of the forward rate process. By construction, negative rates are not allowed in the CEV model. However, for many parameter sets calibrated to the market, the CEV model exhibits a large probability of absorption at zero, especially when the long-dated forward rates are considered [56].

In Andersen & Andreasen's model (5.2.4), a local volatility function, $\phi(x)$, was introduced to each of the LIBOR rates, $\phi(x) = x^{\beta}$. Several other choices for $\phi(x)$ have been presented in the literature. A popular choice is the so-called *displaced diffusion*, i.e. $\phi(x) = a(x-b)$.

The two models mentioned above are related, as shown by Marris [76]. With a suitable parametrization, e.g. β close to 1 and the forward not too close to zero, satisfactory correspondence exists between the prices of caplets based on the CEV and the displaced diffusion dynamics over a range of strike prices. Since it is easier to deal numerically with a displaced diffusion model, this approach can be used as an approximation of the CEV approach, even if a trader has confidence in the CEV model.

The displaced diffusion model and its use as the *local volatility function* of the LMM are frequently discussed in the literature, e.g., Joshi & Rebonato [61], Piterbarg [86], Christopher & Joshi [17]. We skip details of the properties of the displaced diffusion model here, but discuss some relevant aspects in Appendix 5.5.

Both the CEV and displaced-diffusion extensions of the LMM accurately model *monotonically decreasing smile functions*. However, since approximately 1998, the caplet and swaption implied volatility smiles have shown more complex shapes, like a decreasing function for a wide range of strikes but then increasing for a short range of strikes, resembling a *hockey-stick shape*. A number of stochastic volatility models have been proposed to describe this hockey-stick shaped smile, like, for example, Andersen & Brotherton-Ratcliffe [6], Joshi & Rebonato [61], Piterbarg [86] and Rebonato & White [93].

Among the proposed models, the *Term-Structure of Skew (TSS) LMM* [86] seems promising due to its capability of capturing a flexible, seemingly arbitrary, term-structure of the volatility, and volatility skews or smiles across an entire grid of swaptions. In this paper, we consider the TSS-LMM model.

In this model, the forward rates follow a displaced diffusion process, with a stochastic variance process and a set of time-dependent functions (instantaneous forward LIBOR skews) $\{b_i(t),\ t \geq 0\}_{i=1}^{N-1}$, so that

$$dF_i(t) = \big(b_i(t)F_i(t) + \big(1 - b_i(t)\big)F_i(0)\big)\,\sigma_i(t)\sqrt{z(t)}\left(\sqrt{z(t)}\mu(t)dt + dW_i(t)\right), \tag{5.2.5}$$
$$i = 1, 2, \ldots, N-1, \qquad b_i(t) \in [-1, 1],$$

where $\mu(t)$ is the arbitrage-free drift associated with a certain chosen numeraire and $\sigma_i(t)$ are deterministic volatility functions. Jäckel [56] suggests that the skew parameter $b_{(\cdot)}$ should be within the range from 0 to 1. The argument for such a choice is that 'the maximum attainable (negative) skew of the displaced diffusion model when $b_{(\cdot)} \to 0$ is usually more than sufficient for caplets' and it predicts that the forward rate 'may take on any negative value with potentially quite large probability'. However, the skew in the current market is much stronger than those in the old days; we have to consider the model which generates skew stronger than 'normal skew limit' (see Appendix 5.5). And the issue of negative rates will lead to a debate based on economic grounds that is of no particular relevance here.

The stochastic variance process, $z(t)$, is defined by the mean reverting SDE

$$dz(t) = \kappa\big(z(0) - z(t)\big)dt + \eta\sqrt{z(t)}dU(t), \quad z(0) = 1, \tag{5.2.6}$$

where $U(t)$ is a Brownian motion, independent of $W_i(t)$. Here the process $z(t)$ is in fact a stochastic perturbation factor which moves the volatilities at any given moment away from (but near, which is guaranteed by $z(0) = 1$),

its deterministic levels. By this construction, the deterministic volatility level and the stochastic volatility component are effectively separated to facilitate calibration. Notice that all forward LIBOR rates are driven by a common stochastic variance process. This is because a single stochastic volatility process can describe the curvature of the volatility smile for different swaptions well [86].

Due to the large number of free parameters, $\{b_i(t),\ t \geq 0\}_{i=1}^{N-1}$ and $\{\sigma_i(t),\ t \geq 0\}_{i=1}^{N-1}$, the TSS-LMM can model seemingly *arbitrary forms* of the volatility term structure and skews/smiles across an entire swaption grid. However, it is also due to this large number of free parameters that the calibration procedure gives rise to a highly non-linear inverse problem with a large set of feasible solutions.

5.3 Model Mapping for TSS-LMM

Swap rates are not directly modelled by the LMM, but their dynamics are implied by the dynamics of all underlying forward LIBOR rates. To derive swap rate dynamics, we recall that

$$\mathrm{SR}_{n,m}(t) = \frac{P(t,T_n) - P(t,T_m)}{\sum_{j=n+1}^{m} \delta_j P(t,T_j)}.$$

D'Aspremont [34] explains that the drift term due to the change of measure between the forward and the swap martingale measures can be neglected in the computation of the swaption price. By applying Itô's lemma and given the dynamics of the forward LIBOR rates in Eq. (5.2.5) under the definition of the TSS-LMM, we find the swap rate dynamics implied by the TSS-LMM, under swap measure $\mathbb{Q}^{n,m}$:

$$\begin{aligned} d\mathrm{SR}_{n,m}(t) &= \sum_{i=n}^{m-1} \frac{\partial \mathrm{SR}_{n,m}(t)}{\partial F_i(t)} dF_i(t) \\ &= \sum_{i=n}^{m-1} \frac{\partial \mathrm{SR}_{n,m}(t)}{\partial F_i(t)} \left(b_i(t)F_i(t) + \left(1 - b_i(t)\right)F_i(0)\right) \sigma_i(t) \sqrt{z(t)} dW_i^{n,m}(t). \end{aligned} \tag{5.3.1}$$

The partial derivatives, $\frac{\partial \mathrm{SR}_{n,m}(t)}{\partial F_j(t)}$, are functions of the underlying LIBOR rates at time t, i.e.

$$\frac{\partial \mathrm{SR}_{n,m}(t)}{\partial F_j(t)} = f\left(F_n(t), \ldots, F_{m-1}(t)\right).$$

A direct computation of the above expression for the swap rate dynamics (5.3.1) is not possible by any analytic or efficient numerical technique. One commonly used strategy is to employ a *model mapping* strategy, which relates an involved model, as in (5.3.1), to a simpler, tractable model, so that plain vanilla option prices are efficiently computable.

Here we adopt the model mapping strategy introduced in [86] to obtain an accurate yet numerically efficient approximation to the swap rate process (5.3.1), suitable for calibration applications. In a first step, we *project* the full swap rate dynamics (5.3.1) onto a low-dimensional model with time-dependent parameters. More precisely, we approximate the swap rate dynamics under the swap

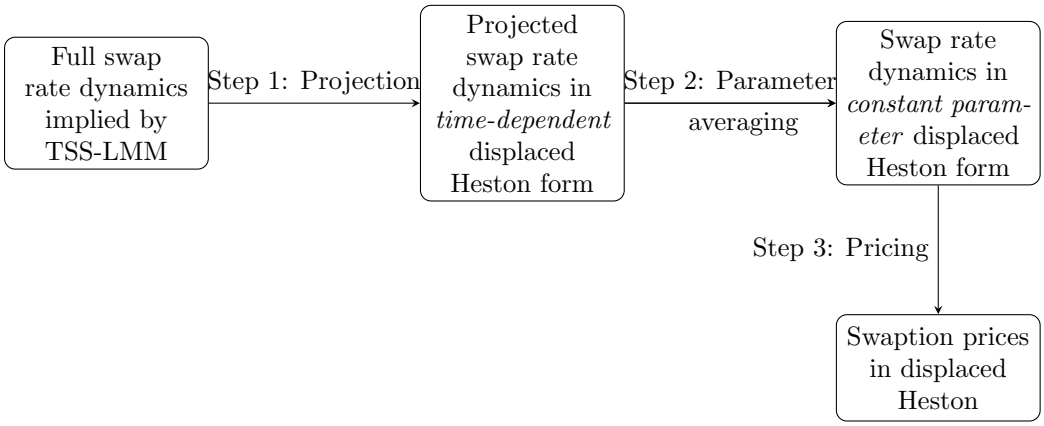

Figure 5.2: Model mapping procedure for the TSS-LMM

measure, $\mathbb{Q}^{n,m}$, by the following form

$$d\hat{\mathrm{SR}}_{n,m}(t) = \left(\hat{b}_{n,m}(t)\hat{\mathrm{SR}}_{n,m}(t) + \left(1-\hat{b}_{n,m}(t)\right)s_{n,m}\right)\hat{\sigma}_{n,m}(t)\sqrt{z(t)}dW^{n,m}(t), \tag{5.3.2}$$

where $\hat{b}_{n,m}(t)$ is the time-dependent skew parameter for the swap rate $\hat{\mathrm{SR}}_{n,m}(t)$, $\hat{\sigma}_{n,m}(t)$ are volatilities of the swap rate and we define $s_{n,m} := \mathrm{SR}_{n,m}(0)$. Details of the approximation procedure will be given in Section 5.3.1.

Model (5.3.2) is simpler than the full swap rate dynamics in Eq. (5.3.1), but European option prices are still not easily computed in this model, because the coefficients are time-dependent. Additional steps have to be made to *project* the time-dependent approximate model (5.3.2) onto a further simplified model by averaging the time-dependent parameters, i.e.

$$d\bar{\mathrm{SR}}_{n,m}(t) = \left(\bar{b}_{n,m}\bar{\mathrm{SR}}_{n,m}(t) + (1-\bar{b}_{n,m})s_{n,m}\right)\bar{\lambda}_{n,m}\sqrt{z(t)}dW^{n,m}(t), \tag{5.3.3}$$

with constant skew parameters, $\bar{b}_{n,m}$, time-independent volatility, $\bar{\lambda}_{n,m}$, and $z(t)$ is the square root process in Eq. (5.2.6), see Section 5.3.2. Model (5.3.3) resembles a displaced Heston model, whose European option prices can be obtained semi-analytically via the available characteristic function. Various efficient methods are available for European option valuation, like FFT-based methods [29], or the COS method [36], so that we can effectively calibrate the displaced Heston model to market quotes of European interest rate options, e.g. caplets and swaptions.

Schematically, we illustrate the model mapping procedure in Figure 5.2. The details of the different steps are discussed in the following subsections.

5.3.1 Step 1: Model Projection onto a Time-Dependent Displaced Diffusion

Piterbarg [86] proposed a strategy to project the full swap dynamics (5.3.1) under the TSS-LMM onto the displaced diffusion model with time-dependent

coefficients. In [87] and [11] the *Markov projection* technique was presented to improve the accuracy of the approximate model. Here, we briefly describe these techniques, and we compare, by numerical experiments, the accuracy and computational costs of these two methods in Section 5.4.2.

The First Projection: Freezing Approximation

The full swap dynamics under the TSS-LMM are given by Equation (5.3.1), where the partial derivatives $\frac{\partial \mathrm{SR}_{n,m}(t)}{\partial F_j(t)}$ are functions of the underlying LIBOR rates. The approximation in [86] was to project the SDEs (5.3.1) onto Model (5.3.2). In order to perform this projection step, the *freezing condition* was imposed, so that Eqs. (5.3.1) and (5.3.2) should give the same results 'along the forward path', i.e. $\hat{\mathrm{SR}}_{n,m}(t) = s_{n,m}$, $F_j(t) = F_j(0)$, and their gradients should also be the same along the forward path. From these two conditions, we obtain

$$\hat{\sigma}_{n,m}(t) = \sum_{j=n}^{m-1} \frac{F_j(0)}{s_{n,m}} \frac{\partial s_{n,m}}{\partial F_j(0)} \sigma_j(t). \tag{5.3.4}$$

The skew parameters, $\hat{b}_{n,m}(\cdot)$, have to satisfy the condition,

$$\hat{b}_{n,m}(t)\hat{\sigma}_{n,m}(t) = b_j(t)\sigma_j(t), \forall j = n, \ldots, m-1.$$

In [87] the problem was reformulated in least-squares form:

$$\hat{b}_{n,m}(t) = \arg\min_{\gamma} \sum_{j=n}^{m-1} \left(\gamma \cdot \hat{\sigma}_{n,m}(t) - b_j(t)\sigma_j(t)\right)^2, \tag{5.3.5}$$

which leads to the following result.

Result 5.3.1 (The first projection of the swap rate dynamics [86])**.** *Given the forward LIBOR rate dynamics, as defined by Eqs. (5.2.5) and (5.2.6) in the TSS-LM model, the projected dynamics of the swap rate, $\hat{SR}_{n,m}(t)$, under the swap measure $Q^{n,m}$, satisfy the following SDE:*

$$d\hat{SR}_{n,m}(t) = \left(\hat{b}_{n,m}(t)\hat{SR}_{n,m}(t) + \left(1 - \hat{b}_{n,m}(t)\right)s_{n,m}\right)\sqrt{z(t)} \sum_{k=1}^{K} \hat{\sigma}_{n,m}(t;k) dZ_k^{n,m}(t),$$

where

$$\hat{\sigma}_{n,m}(t;k) := \sum_{i=n}^{m-1} q_i(n,m)\sigma_i(t;k), \quad \hat{b}_{n,m}(t) := \sum_{i=n}^{m-1} p_i(n,m)b_i(t), \tag{5.3.6}$$

and

$$q_i(n,m) = \frac{F_i(0)}{s_{n,m}} \frac{\partial s_{n,m}}{\partial F_i(0)}, \quad p_j(n,m) = \frac{\sum_{k=1}^{K} \sigma_j(t;k)\hat{\sigma}_{n,m}(t;k)}{(m-n)\sum_{k=1}^{K} \hat{\sigma}_{n,m}(t;k)^2}, \quad i,j = n, \ldots, m-1.$$

Remark. In formulation (5.2.5) the volatility coefficients, $\sigma_i(t)$, are the volatilities for individual forward rates. The formulation in (5.3.6), however, has been chosen so that the volatility coefficients, $\sigma_i(t;k)$, are not directly associated

to any individual forward rate volatility, but specify the volatilities of a K-dimensional (K usually smaller than the number of underlying forward rates) independent Brownian motion vector. Thus, when applying the freezing formulas in our case, we have to transform the volatility coefficients into coefficients for independent Brownian motions. The same holds for the volatility coefficients of the swaption, $\hat{\sigma}_{n,m}(t;k)$, defined as the volatility vector of a K-dimensional independent Brownian motion.

Second Model Projection: Markov Projection

In the method described in the previous subsection, the approximate swap rate dynamics were obtained by 'freezing' all state-dependent variables in Eq. (5.3.1). This approach is rather straightforward and possibly prone to error if the underlying state variables exhibit large variations. Therefore, in [87] a refined approximation was proposed for the swap rate, based on the idea of *Markov projection*, from Gyöngy's theorem [42]:

Result 5.3.2 (Gyöngy [42])**.** *Consider a stochastic process defined by*

$$dX(t) = b(t,\omega)ds + \sigma(t,\omega)dW(t). \tag{5.3.7}$$

Let a diffusion matrix be defined by $\alpha := \sigma(s,\omega)\sigma(s,\omega)^T$*. If* α *satisfies the condition*

$$\sum_{i,j} \alpha_{i,j} z_i z_j \geq \xi \sum_i |z_i|^2,$$

in which ξ *is a fixed positive constant, for every* $(t,\omega) \in [0,\infty) \times \Omega$ *and* $z \in \mathbb{R}^n$*, then bounded measurable functions,* $b^P : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ *and* $\sigma^P : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$*, exist, so that*

$$\begin{aligned} b^P(t,x) &= \mathbb{E}\left[b(t,\omega)|X(t)=x\right], \\ \sigma^P(t,x)\sigma^P(t,x)^T &= \mathbb{E}\left[\sigma(t,x)\sigma(t,x)^T|X(t)=x\right], \end{aligned} \tag{5.3.8}$$

almost certainly, for all $(t,x) \in [0,\infty)\times\mathbb{R}^n$*. Moreover, the stochastic differential equation*

$$X^P(t) = X(0) + \int_0^t b^P\big(s, X^P(s)\big)ds + \int_0^t \sigma^P\big(s, X^P(s)\big)dW(s),$$

admits a weak solution, $X^P(t)$*, which has the same marginal distribution as the original process,* $X(t)$*.*

The Markov projection projects an involved underlying multi-dimensional process on a simplified, low-dimensional process with the same marginal distributions as the target process. As a result, the projected process results in the same European option prices as the original process. Theorem 5.3.2 states that in this projected process the original drift and diffusion coefficients are approximated by their conditional expectations, conditional on the current state variables.

More precisely, we write Eq. (5.3.1) in the following form

$$d\mathrm{SR}_{n,m}(t) = \sqrt{z(t)} \sum_{k=1}^{K} \pi_{n,m,k}(t) dZ_k^{n,m}(t), \tag{5.3.9}$$

with K-dimensional independent Brownian motions, $Z^{n,m}$, under the annuity measure, and *total volatility terms*, $\pi_{n,m,k}(t)$, defined by

$$\pi_{n,m,k}(t) := \sum_{i=n}^{m-1} \frac{\partial \mathrm{SR}_{n,m}(t)}{\partial F_j(t)} \left(b_j(t)F_j(t) + \big(1-b_j(t)\big)F_j(0)\right)\sigma_j(t)r_{i,k}, \quad (5.3.10)$$

where $r_{i,k}$ is an element of the *principal matrix square root* of the forward LIBOR correlation matrix.

By Result 5.3.2, the Markov projection of the swap rate dynamics (5.3.1) reads

$$d\hat{\mathrm{SR}}_{n,m}(t) = \left(\mathbb{E}\left[||\boldsymbol{\pi}_{n,m}(t)||^2|\hat{\mathrm{SR}}_{n,m}(t)\right]\right)^{\frac{1}{2}} \sqrt{z(t)}dW^{n,m}(t), \quad (5.3.11)$$

with

$$||\boldsymbol{\pi}_{n,m}(t)||^2 = \left(\sum_{k=1}^{K} \pi_{n,m,k}(t)\right)^2.$$

Variance process $z(t)$ has not been projected, because of the independence of the stochastic variance and the forward LIBOR processes, which implies that $z(t)$ does not have an impact on the conditional expectation in Eq. (5.3.11). An approximate process of the form (5.3.11) may pose difficulties when computing the conditional expectation. We therefore consider an approximate process of the following form, which seems more natural for the computation of the Markov projection of the swap rate dynamics:

$$d\hat{\mathrm{SR}}_{n,m}(t) = \hat{\sigma}_{n,m}(t)\left(\hat{b}_{n,m}(t)\hat{\mathrm{SR}}_{n,m}(t) + \big(1-\hat{b}_{n,m}(t)\big)s_{n,m}\right)\sqrt{z(t)}dW^{n,m}(t).$$

Result 5.3.3 (Swap rate Markovian projection [9])**.** *The swap rate volatility is approximated by*

$$\hat{\sigma}_{n,m}(t) = \frac{1}{s_{n,m}}\left(\sum_{k=1}^{K}\left(\pi_{n,m,k}(t)\Big|_{\mathbb{E}^{n,m}[\boldsymbol{F}_{n,m-1}(t)]}\right)^2\right)^{\frac{1}{2}}, \quad (5.3.12)$$

where we use the notation

$$F_{i,j}(t) := \{F_i(t),\dots,F_j(t)\}, \text{ and } \mathbb{E}\left[F_{i,j}(t)\right] = \{\mathbb{E}[F_i(t)],\dots,\mathbb{E}[F_j(t)]\}. \quad (5.3.13)$$

Further, the notation $|_{\mathbb{E}\dots}$ means conditional on the realization of the forward rate vector being equal to the expectation of the forward rates.

The swap rate skew is then approximated by

$$\hat{b}_{n,m}(t) = \frac{s_{n,m}}{m-n}\sum_{i=n}^{m-1}\frac{\partial \ln||\pi_{n,m}||}{\partial F_i(t)}\Bigg|_{F_{n,m-1}(0)} \cdot \frac{\sum_{k=1}^{K}\int_0^t \pi_{n,m,k}\big(u,F_{n,m-1}(0)\big)F_k(0)\sigma_i(u)r_{i,k}(u)du}{\sum_{k=1}^{K}\int_0^t \pi_{n,m,k}\big(u,F_{n,m-1}(0)\big)^2 du}. \quad (5.3.14)$$

The partial derivatives are evaluated at their initial values to facilitate the computation:

$$\begin{aligned}
&\frac{\partial \ln ||\boldsymbol{\pi}_{n,m}||}{\partial F_i(t)}\Big|_{F_{n,m-1}(0)} \\
=&\frac{1}{||\boldsymbol{\pi}_{n,m}\big(t, F_{n,m-1}(0)\big)||^2}\frac{\partial SR_{n,m}(t)}{\partial F_i(t)}\Big|_{F_{n,m-1}(0)} \cdot b_i(t) \\
&\cdot \sum_{k=1}^{K} \pi_{n,m,k}\big(t, F_{n,m-1}(0)\big) F_k(0)\sigma_i(t) r_{i,k}(t) + \frac{1}{||\boldsymbol{\pi}_{n,m}\big(t, F_{n,m-1}(0)\big)||^2} \\
&\cdot \sum_{p=n}^{m-1} \frac{\partial^2 SR_{n,m}(t)}{\partial F_i(t)\partial F_p(t)}\Big|_{F_{n,m-1}(0)} \cdot F_p(0) \sum_{k=1}^{K} \pi_{n,m,k}\big(t, F_{n,m-1}(0)\big)\sigma_p(t) r_{p,k}(t).
\end{aligned}$$

This Markov projection formula looks complicated, but there are only a few differences between the Markov projection and the freezing projection in Result 5.3.1. The most important difference is that in the Markov projection the volatility of the swap rate (5.3.12) is approximated by the expected LIBOR rates at each specific time point t, instead of projecting them onto their initial values. If we had evaluated the volatility term (5.3.12), not by means of the expected LIBOR rates but by their initial values, we would recover the projection from Result 5.3.1:

$$\hat{\sigma}_{n,m}(t) = \frac{1}{s_{n,m}}\left(\sum_{k=1}^{K}\big(\pi_{n,m,k}(t)|_{F_{n,m-1}(0)}\big)^2\right)^{\frac{1}{2}} = \left(\sum_{k=1}^{K}\left(\sum_{i=n}^{m-1} q_i(n,m)\sigma_i(t;k)\right)^2\right)^{\frac{1}{2}}.$$

The challenge to applying the Markov projection technique to swap rate dynamics is the efficient calculation of the conditional expectation of the swap rate volatility. Piterbarg & Andersen [9] provide approximation formulas for the conditional expectation, based on the so-called *Gaussian approximation* [87]:

Result 5.3.4 (Conditional expected forward LIBOR rates [9])**.** *For $n \leq i \leq m-1$, the expected value of the i-th forward LIBOR rate under the annuity measure, $\mathbb{Q}^{n,m}$, can be accurately approximated by*

$$\mathbb{E}\left[F_i(t)\right] \approx F_i(0)\big(1 + c_i(t)\big),$$

with

$$\begin{aligned}
c_i(t) =& \frac{1}{F_i(0)A_{n,m}(0)}\sum_{j=n}^{m-1}\left(\frac{\delta_j P_{j+1}(0)}{P_j(0)}\big(\mathbf{1}_{j\leq i}A_{n,m}(0) - A_{j,m}(0)\big)\right) \\
&\cdot \int_0^t \sigma_i(s)\sigma_j(s)\sum_{k=1}^{K} r_{i,k}(s)r_{j,k}(s)ds.
\end{aligned}$$

It is useful to analyze whether or not the accuracy of the approximation improves, and whether there is a balance between the accuracy and computation time. This issue is treated in Section 5.4.2, where a series of numerical experiments detail the accuracy and computation time of the two swap rate projection formulas presented.

5.3.2 Step 2: Parameter Averaging

The second step of the model mapping is the so-called *parameter averaging*, to determine constant parameters $\bar{\lambda}_{n,m}$ and $\bar{b}_{n,m}$, so that SDE (5.3.2) can be best approximated by a constant parameter swap rate process also under swap measure $\mathbb{Q}^{n,m}$,

$$d\bar{\mathrm{SR}}_{n,m}(t) = \big(\bar{b}_{n,m}\bar{\mathrm{SR}}_{n,m}(t) + (1-\bar{b}_{n,m})s_{n,m}\big)\,\bar{\lambda}_{n,m}\sqrt{z(t)}dW^{n,m}(t), \quad (5.3.15)$$

with

$$dz(t) = \kappa\big(z(0) - z(t)\big)dt + \eta\sqrt{z(t)}dU(t), \quad z(0) = z_0.$$

In [86] it was found that the effective skew parameter, $\bar{b}_{n,m}$, can be written as a *volatility weighted average* of the skew function $\hat{b}_{n,m}(t)$ in (5.3.14):

$$\bar{b}_{n,m} := \int_0^{T_n} \hat{b}_{n,m}(t)\omega_{n,m}(t)dt, \quad (5.3.16)$$

where

$$\omega_{n,m}(t) = \frac{\nu(t)^2\hat{\sigma}_{n,m}(t)^2}{\int_0^{T_n}\nu(t)^2\hat{\sigma}_{n,m}(t)^2dt}, \quad (5.3.17)$$

and

$$\nu(t)^2 = z_0^2\int_0^t \hat{\sigma}_{n,m}(s)^2ds + z_0\eta^2e^{-\kappa t}\int_0^t \hat{\sigma}_{n,m}(s)^2\frac{e^{\kappa s}-e^{-\kappa s}}{2\kappa}ds. \quad (5.3.18)$$

The effective (constant) volatility is given by parameter $\bar{\lambda}_{n,m}$, which solves the following equations involving two Laplace transforms:

$$\mathcal{L}_0\big(c;\bar{\lambda}_{n,m}^2\int_0^T z(t)dt\big) = \mathcal{L}\big(c;\int_0^T \hat{\sigma}_{n,m}(t)^2z(t)dt\big) \quad (5.3.19)$$

with

$$c = \left(2\int_0^T \hat{\sigma}_{n,m}(t)^2dt\right)^{-1} + \frac{\bar{b}_{n,m}^2}{8}.$$

The Laplace transform $\mathcal{L}$ is defined by

$$\mathcal{L}\left(\mu;\{\zeta(t):0\le t\le T\}\right) := \exp\left(A(T) + B(T)z_0\right),$$

where the functions $A(\tau)$ and $B(\tau)$ are functions of μ and $\zeta(t)$ and can be obtained by solving the following Riccati-type ordinary differential equations:

$$\begin{aligned} \frac{dA(\tau)}{d\tau} &= \kappa z_0 B(\tau), & A(0) &= 0, \\ \frac{dB(\tau)}{d\tau} &= -\kappa B(\tau) + \tfrac{1}{2}\eta^2B(\tau)^2 - \mu\zeta(T-\tau), & B(0) &= 0. \end{aligned} \quad (5.3.20)$$

The Laplace transforms can be seen as *moment generating functions of the integrated volatilities*, i.e.

$$\mathcal{L}_0\big(c;\bar{\lambda}^2_{n,m}\int_0^T z(t)dt\big)=\mathbb{E}\big[\exp\big(-c\cdot\bar{\lambda}^2_{n,m}\int_0^T z(t)dt\big)\big],$$
$$\mathcal{L}\big(c;\int_0^T\hat{\sigma}_{n,m}(t)^2z(t)dt\big)=\mathbb{E}\big[\exp\big(-c\cdot\int_0^T\hat{\sigma}_{n,m}(t)^2z(t)dt\big)\big].$$

For the constant parameter displaced Heston model, we have $\zeta(t)=\bar{\lambda}^2_{n,m}$, so that $A(T)$ and $B(T)$ can be determined as

$$A(T)=\frac{\kappa z_0}{\eta^2}\left((\kappa+\bar{\phi})T-2\log\left(\frac{1-\nu\cdot e^{\bar{\phi}T}}{1-\nu}\right)\right), \tag{5.3.21}$$

$$B(T)=\frac{2\mu\lambda_T^2(1-e^{-\bar{\phi}T})}{(\kappa+\bar{\phi})(1-e^{-\bar{\phi}T})+2\bar{\phi}e^{-\bar{\phi}T}}, \tag{5.3.22}$$

with

$$\bar{\phi}:=\sqrt{\kappa^2+2\mu\eta^2\bar{\lambda}^2_{n,m}},\ \nu=\frac{\kappa+\bar{\phi}}{\kappa-\bar{\phi}}.$$

The Riccati equations associated with the time-dependent volatility, $\hat{\sigma}_{m,n}(t)$, can only be solved numerically by applying a root finding method.

In general, there is one averaged volatility, $\bar{\lambda}_{n,m}$, each expiry/tenor date. For a large swaption matrix with many expiry dates and tenors, it may result in a large number of averaged volatilities to be determined. Fortunately, the averaged volatilities, $\bar{\lambda}_{n,m}$ for each combination of expiry/tenor are independent of each other, so that it does not result in a high-dimensional search problem. We merely need to deal with a collection of one-dimensional problems.

5.3.3 Step 3: Option Pricing in the Displaced Heston Model

For notational convenience, we omit the index of the swap expiry/tenor in this subsection. The effective swap rate dynamics (5.3.15), driven by the stochastic variance process (5.2.6), resemble the displaced Heston dynamics:

$$\begin{array}{lll} d\bar{\mathrm{SR}}(t)= & \big(\bar{b}\bar{\mathrm{SR}}(t)+(1-\bar{b})\bar{\mathrm{SR}}(0)\big)\,\bar{\lambda}\sqrt{z(t)}dW(t), & \bar{\mathrm{SR}}(0)=s_{n,m},\\ dz(t)= & \kappa\,(z(0)-z(t))\,dt+\eta\sqrt{z(t)}dU(t), & z(0)=1, \end{array} \tag{5.3.23}$$

where $W(t)$ and $U(t)$ are two independent Brownian motions. At this stage, the displaced Heston model is cast into the well-known Heston model, so that European option prices can be obtained directly by applying Fourier techniques.

Let us define a displacement parameter, $\theta:=\frac{\bar{b}-1}{\bar{b}}\bar{\mathrm{SR}}(0)$. The adjusted forward rate process, $\bar{\mathrm{SR}}_\theta:=\bar{\mathrm{SR}}-\theta$, in the displaced Heston model will then be of log-normal form, i.e.

$$\begin{aligned} d\bar{\mathrm{SR}}_\theta(t)=&d\left(\bar{\mathrm{SR}}(t)-\theta\right)=\bar{\lambda}\bar{b}\left(\bar{\mathrm{SR}}(t)-\theta\right)\sqrt{z(t)}dW(t)\\ =&\bar{\mathrm{SR}}_\theta(t)\sqrt{\bar{\lambda}^2\bar{b}^2z(t)}dW(t),\quad \bar{\mathrm{SR}}_\theta(0)=\frac{\bar{\mathrm{SR}}(0)}{\bar{b}}. \end{aligned}$$

When evaluating option prices, the strike price thus has to be adjusted in the shifted model, i.e. $\bar{K}_\theta = \bar{K} - \theta$.

Subsequently, the stochastic variance process, $z(t)$, has to be adjusted too:

$$\begin{aligned} dz_\theta(t) =& d\left(\bar{\lambda}^2\bar{b}^2 z(t)\right) = \bar{\lambda}^2\bar{b}^2\kappa\left(z(0) - z(t)\right)dt + \bar{\lambda}^2\bar{b}^2\eta\sqrt{z(t)}dU(t) \\ =& \kappa\left(z_\theta(0) - z_\theta(t)\right)dt + \eta\bar{\lambda}^2\bar{b}^2\sqrt{\frac{z_\theta(t)}{\bar{\lambda}^2\bar{b}^2}}dU(t) \\ =& \kappa\left(z_\theta(0) - z_\theta(t)\right)dt + \eta_\theta\sqrt{z_\theta(t)}dU(t), \end{aligned}$$

where $z_\theta(0) = \bar{\lambda}^2\bar{b}^2 z(0)$, $\eta_\theta = \bar{\lambda}\bar{b}\eta$.

By the shifting technique, we can compute European option prices of the displaced Heston model (5.3.23), C^{DHes}, by a well-known Heston pricing method, denoted by C^{Hes},

$$\begin{aligned} & C^{\text{DHes}}\left(\bar{\text{SR}}(0), \bar{K}, \bar{\lambda}, \bar{b}, \eta, \kappa, T\right) \\ =& C^{\text{Hes}}\left(\bar{\text{SR}}_\theta(0), \bar{K}_\theta, \mu_\theta = \bar{\lambda}^2\bar{b}^2, z_\theta(0) = \bar{\lambda}^2\bar{b}^2, \eta_\theta = \bar{\lambda}\bar{b}\eta, \rho = 0, \kappa, T\right), \end{aligned} \quad (5.3.24)$$

where the long-term mean parameter, μ_θ, is set to be equal to the initial variance, $z_\theta(0)$.

The implied Black-Scholes volatility [20], obtained by a displaced Heston model, is then obtained by inverting the following equation,

$$C^{\text{DHes}}\left(\bar{\text{SR}}(0), \bar{K}, \bar{\lambda}, \bar{b}, \eta, \kappa, T\right) = C^{\text{BS}}(\bar{\text{SR}}(0), \bar{K}, T, \sigma_{\text{impv}}),$$

with σ_{impv} the implied Black-Scholes volatility, within the Black-Scholes pricing formula, denoted by C^{BS}. The above relation connects the option price to the implied volatility, according to the market convention to quote options in terms of implied volatilities.

We illustrate schematically the effect of parameter $\bar{b}$ on the shape of the implied volatility skew and term structure of the ATM implied volatility in Figure 5.3. Notice that we allow negative values for parameter $\bar{b}$ which exhibit a more pronounced implied skew. Some details on negative $\bar{b}$-values are given in Appendix 5.5. Interestingly, parameter $\bar{b}$ not only affects the level of the skew of the implied volatility, but also the ATM volatility term structure. This suggests that the parameters $\bar{\lambda}$ and $\bar{b}$ should be *calibrated jointly* to compensate for the impact of $\bar{b}$ on the term structure.

5.4 Calibration of the TSS-LMM

This section describes the method to calibrate the TSS-LMM to an entire swaption cube. The purpose of the calibration is to find optimal time-dependent volatility and skew parameters for the TSS-LMM, in such a way that a whole market swaption cube is matched well. The proposed calibration method relies on the model mapping results introduced above, i.e. Results 5.3.1 and Eqs. (5.3.16–5.3.19).

Generally, calibration proceeds in the reverse direction to model mapping. We first determine the market implied parameters, $\bar{\lambda}^{\text{mkt}}_{n,m}$ and $\bar{b}^{\text{mkt}}_{n,m}$, by calibrating the constant parameter displaced Heston model to European swaptions of expiry n and maturity m; then, at the same time, we consider the full TSS-LMM

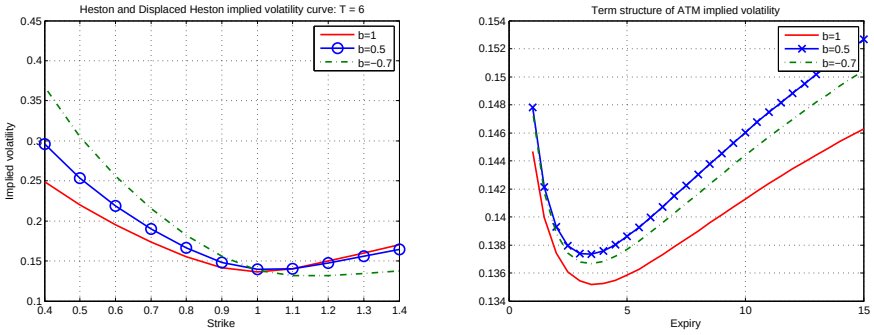

Figure 5.3: Implied volatility curve of one expiry slice and term structure of ATM implied volatility.

model's implied effective model parameters $\bar{\lambda}_{n,m}$ and $\bar{b}_{n,m}$ as functions of the time-dependent parameters $\{\sigma_i(t), t \geq 0\}_i$ and $\{b_i(t), t \geq 0\}_i$ by means of the model mapping techniques; optimize the time-dependent parameters so that the difference between the market implied parameters and the TSS-LMM implied effective model parameters (obtained from the model mapping) is the least.

More specifically, we employ a *three-stage calibration method*, similar to the method proposed by Andersen & Piterbarg [9]:

1. We first determine the global parameters κ (speed of mean reversion) and η (volatility of variance), so that the curvature of the swaption implied volatilities across different expiration dates and tenors is fitted well. Then we determine the market implied parameters $\bar{b}^{\text{mkt}}_{n,m}$ and $\bar{\lambda}^{\text{mkt}}_{n,m}$ by solving an optimization problem, locally for each expiry date n and maturity date m,

$$\{\bar{b}^{\text{mkt}}_{n,m}, \bar{\lambda}^{\text{mkt}}_{n,m}\} = \underset{\bar{b}_{n,m}, \bar{\lambda}_{n,m}}{\arg\min} \sum \Big(\text{Swpt}^{\text{mkt}}_{n,m}(\bar{K}, T) - \text{Swpt}^{\text{mod}}_{n,m}(\bar{K}, T, \bar{b}_{n,m}, \bar{\lambda}_{n,m}) \Big)^2, \tag{5.4.1}$$

   where $\text{Swpt}(\cdot)$ denotes the swaption price for a given strike, expiry date and parameters; *mkt* and *mod* stand for market and model, respectively.

2. Secondly, we assume that the time-dependent volatility $\{\sigma_i(t), t \geq 0\}_i$ is time-homogeneous and of parametric form, according to Rebonato [90],

$$\sigma_i(t) = g(T_i - t) = \big(a_1 + a_2(T_i - t)\big)e^{-a_3(T_i - t)} + a_4, \qquad a_1, a_2, a_3, a_4 \in \mathbb{R}^+, \tag{5.4.2}$$

   and optimize over set $\{a_1, a_2, a_3, a_4\}$ to match the TSS-LMM implied effective volatility parameters $\{\bar{\lambda}_{n,m}\}$ to the market implied values $\{\bar{\lambda}^{\text{mkt}}_{n,m}\}$ for all n and m;

3. Thirdly, we introduce *piecewise constant* TSS parameters, $\{b_i(t), t \geq 0\}_i$, to match the TSS-LMM implied swaption skews $\{\bar{b}_{n,m}\}$ to the market implied skew $\{\bar{b}^{\text{mkt}}_{n,m}\}$ for all n and m.

The individual stages are described in greater detail in the following subsections.

5.4.1 Data Description

We will apply the proposed calibration method to a set of swaption volatilities from 04-Oct-2011 to 11-Oct-2011, covering 6 business days [1].

We consider all quotes from the swaptions of expiry $[1, 2, 5, 10, 15]$ years and tenors $[2, 5, 10, 20]$ years. As an example, we show two slices of the market swaption volatility skew in Figure 5.4.

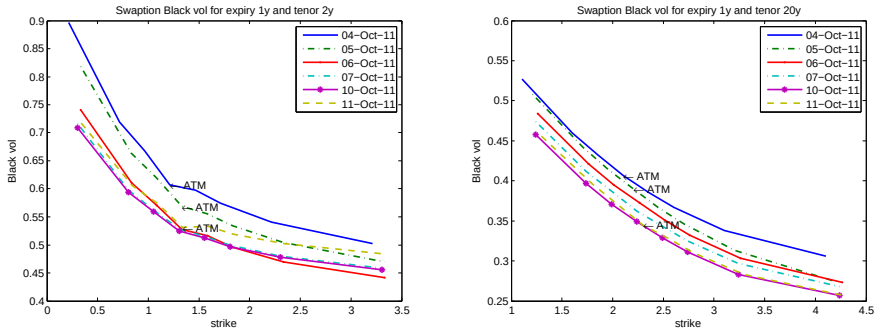

Figure 5.4: Plot of market swaption Black's volatility skew for different expiries and tenors.

Here we choose a forward correlation in the so-called *modified exponential form*, suggested by Rebonato [91],

$$\rho_{i,j} \quad = \rho_\infty + (1 - \rho_\infty) e^{-\delta(T_i,T_j)|T_i - T_j|}, \quad \text{with} \quad \delta(T_i, T_j) := \delta_0 e^{\epsilon \min(T_i,T_j)}. \tag{5.4.3}$$

In reality, for example, the forward rates fixing in nine and ten years will have a lower correlation than the rates fixing in one and two years. The advantage of the *modified exponential* correlation function is that it models an increasing de-correlation among LIBOR rates with similar distance for longer fixing rates than the shorter fixing rates. In this research we do not calibrate the correlation structure jointly with the volatility, because the swaption prices depend only weakly on the correlation structure. A joint calibration procedure may even give rise to unnaturally high correlations among forward rates. Here we just choose some pre-defined parameters for the correlation function, i.e. $\rho_\infty = 0.30, \delta_0 = 0.05$ and $\epsilon = 0.05$ (see the correlation surface plotted in Figure 5.5). Alternatively, we recommend a calibration of the correlation function to historical correlation matrices.

5.4.2 Accuracy of the Model Mapping

By the model mapping described previously, we can relate the TSS-LMM parameters to swaption prices. We analyze here which model projection method to employ for the first mapping step, the *freezing* approximation or the *Markov projection*. We define numerical experiments to assess the accuracy of the two

[1] The data set is courtesy of the Risk Management department of Rabobank International, London Office.

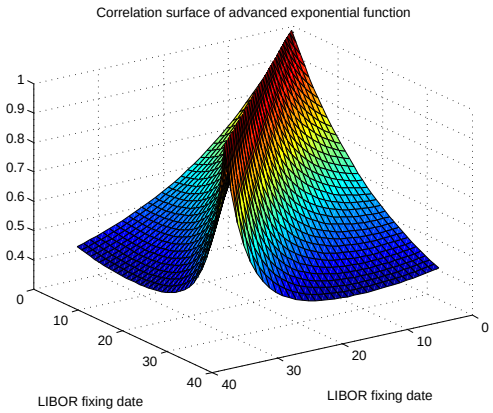

Figure 5.5: Plot of the correlation surface generated by the modified exponential correlation function (5.4.3).

model mapping techniques and also compare the efficiency of the model projection methods. Hereby, we keep the second step approximation, i.e. the parameter averaging step, unchanged.

We compare the approximate swaption prices obtained by the Heston pricing technique as a result of the model mapping with benchmark values that are obtained by Monte Carlo simulation of the full-scale TSS-LMM. The approximate swaption prices due to the model mapping technique are evaluated in three steps:

1 We apply the two model projection methods from Results 5.3.1 and 5.3.3 to obtain the projected time-dependent model parameters;

2 We then perform parameter averaging to obtain constant parameters for the displaced Heston model from the time-dependent projected parameters, and

3 We employ a Heston option pricing technique, e.g. the COS method [36], to evaluate swaption prices in the constant parameter displaced diffusion Heston model.

We perform numerical experiments for three parameter configurations corresponding to three different market conditions, i.e., *usual market conditions*, *stressed conditions* with higher volatilities (also a higher volatility-of-variance and higher skew parameters), and *flat market conditions*, in which all volatilities are constant and volatility skews and smiles do not appear.

We consider a semi-annual tenor structure, up to $T_N = 8$ years, with the final LIBOR resetting period 7.5 years from now. Note that the LIBOR from year 0 to 0.5 is already fixed. Hence, we consider fifteen forward LIBOR rates in total. Ten swaptions over this tenor structure are considered (see Table 5.1). We will not compare the swaption prices, but the implied Black volatilities of the swaptions.

Benchmark prices for the swaptions are computed by a Monte Carlo simulation of the fifteen forward LIBOR rates, where $2 \cdot 10^6$ paths are simulated

Tenor	1y	2y	4y	7y
Expiry 1y	$\mathrm{SR}_{1,2}$	$\mathrm{SR}_{1,3}$	$\mathrm{SR}_{1,5}$	$\mathrm{SR}_{1,8}$
Expiry 2Y	$\mathrm{SR}_{2,3}$	$\mathrm{SR}_{2,4}$	$\mathrm{SR}_{2,6}$	
Expiry 4Y	$\mathrm{SR}_{4,5}$	$\mathrm{SR}_{4,6}$		
Expiry 7y	$\mathrm{SR}_{7,8}$			

Table 5.1: Swaption grid considered in the numerical experiments.

with a time step $\Delta t = 0.05$. The basic Euler discretization scheme is used for the simulation of the forward rates, whereas the Quadratic Exponential (QE) scheme [3] is employed for the stochastic variance process.

Regarding the parameters and initial setting of the TSS-LMM in the experiments, we assume an upward sloping yield curve, which is typical under common market conditions (see the LHS graph of Figure 5.6). The forward rate correlations are prescribed by the 'modified exponential' formula (5.4.3), with $\rho_\infty = 0.30, \alpha_0 = 0.05$ and $\epsilon = 0.05$. The instantaneous volatility is modelled as a time-homogeneous parametric function, as in Eq. (5.4.2) with parameters $a_1 = 0.05, a_2 = 0.10, a_3 = 0.35, a_4 = 0.10$.

The instantaneous LIBOR skews are time-dependent, prescribed as piecewise-constant functions, as shown in the RHS graph of Figure 5.6. For the stochastic variance process we assume that the speed of mean reversion and the volatility of variance parameters are given by $\kappa = 0.15,\ \eta = 1.3$.

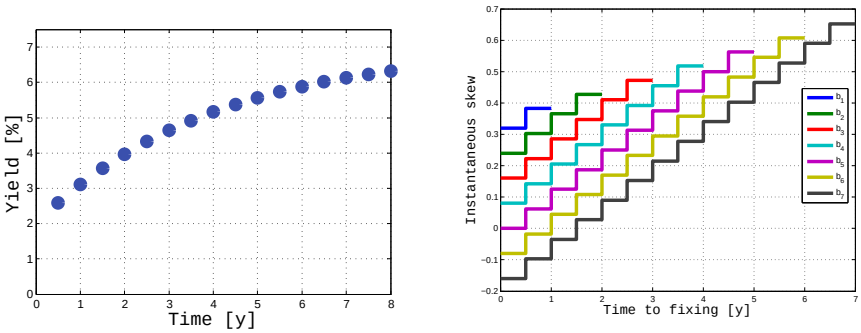

Figure 5.6: LHS: Initial yield curve for the experiment. RHS: The instantaneous LIBOR skews considered in the experiment.

The differences between the benchmark MC swaption implied volatilities and the swaption volatilities obtained by the model mapping with the basic freezing projection and the Markov projection, *under common market conditions*, are presented in Tables 5.2 and 5.3, respectively. Shown is the difference in basis points (0.01%) between the approximate and the full model implied volatilities, for each combination of strike price and swaption.

By comparing Tables 5.2 and 5.3, it is not apparent which projection method is superior. For some of the entries, the Markov projection out-performs while it does not perform well for other entries. Overall, Markov projection gives a somewhat higher approximation quality. Furthermore, several observations can be made based on this comparison under common market conditions: Considering swaptions with a fixed expiry, the errors tend to decrease in the longer

Strikes (% a.t.m)	60	70	80	90	100	110	120	130	140
Swp. $1y \times 2y$	64	66	66	75	59	68	83	101	115
Swp. $1y \times 3y$	−20	−12	−2	23	13	12	11	15	20
Swp. $1y \times 5y$	−65	−44	−18	28	27	20	8	4	3
Swp. $1y \times 8y$	−110	−79	−42	17	33	27	12	6	5
Swp. $2y \times 3y$	168	146	117	92	57	68	93	124	154
Swp. $2y \times 4y$	103	98	91	88	59	59	63	74	89
Swp. $2y \times 6y$	−7	6	19	38	17	8	−5	−8	−2
Swp. $4y \times 5y$	87	88	89	87	47	24	6	0	1
Swp. $4y \times 6y$	13	25	37	46	19	−10	−38	−55	−62
Swp. $7y \times 8y$	−43	−7	23	35	−1	−56	−122	−179	−224

Table 5.2: Swaption implied volatility differences with the MC benchmark volatilities for the freezing approximation (deviation in basis points).

Strikes (% a.t.m)	60	70	80	90	100	110	120	130	140
Swp. $1y \times 2y$	11	11	7	11	−8	4	23	45	61
Swp. $1y \times 3y$	−22	−13	−3	23	14	13	13	18	23
Swp. $1y \times 5y$	−65	−43	−17	29	28	21	10	6	4
Swp. $1y \times 8y$	−76	−53	−25	27	33	19	−4	−14	−21
Swp. $2y \times 3y$	170	148	119	94	58	69	94	124	154
Swp. $2y \times 4y$	103	99	92	89	59	60	64	74	90
Swp. $2y \times 6y$	−1	11	22	41	18	7	−8	−12	−7
Swp. $4y \times 5y$	89	90	90	87	47	24	6	−1	0
Swp. $4y \times 6y$	16	27	39	47	19	−11	−40	−57	−65
Swp. $7y \times 8y$	−42	−6	23	34	−2	−58	−124	−182	−227

Table 5.3: Swaption implied volatility errors for the Markov projection (deviation in basis points).

Mean absolute error	normal	flat	stressed
Freezing	55.81	46.17	55.34
Markov projection	46.19	36.39	48.49

Table 5.4: Mean absolute errors (in basis points) by different swap rate approximation methods, under three different market conditions.

tenors, even though this is not true for low strike swaptions. For example, by considering only the swaptions with an expiry of one year and focusing on the first projection (see Table 5.2), we see that the out-the-money (OTM) errors decrease from 115 to 5 basis points for tenors increasing from 2y to 8y. A similar pattern is observed for implied volatilities of other expiration dates. We can also conclude that, under common market conditions, the far out-the-money swaptions are more sensitive to the choice of projection than those with strike values close to the at-the-money value. This makes sense because the non-ATM strike computations are sensitive to both the skew and volatility projections, whereas the swaptions close to at-the-money strike values are more sensitive to volatility than to skew projection.

For both projection techniques, the largest errors are obtained for far in- and far out-the-money swaptions with an expiry date of two years, and for high strike swaptions with an expiry date of seven years. In both cases, the errors are largest for the swaptions with a tenor of one year. In the numerical experiments of caplet approximation, we encounter the same error pattern and the error estimates are higher than those when only parameter averaging is considered. This is mainly due to the lower number of Monte Carlo simulations, the larger time step taken in these simulations and the discretization errors in the Libor drifts. Since swaptions with a tenor of one year are structurally similar to caplets [2], we reason that these relatively higher inaccuracies are not related to the swap rate projections but are more likely due to numerical errors in the Monte Carlo simulation. Moreover, for longer tenor swaptions, errors in various approximation steps may partly cancel each other out. This also explains why the approximation quality for longer tenor swaptions appears to be better than the shorter tenor swaptions.

The error patterns for the two approximation methods are structurally similar under the common market conditions, despite different results for individual entries in the tables – for some entries the Markov projection results are more accurate, while for other entries the freezing projection performs best.

We also compare the errors on an aggregate level, in Table 5.4, by means of the mean absolute errors over all swaptions considered *under three market conditions*, i.e. the common, flat and stressed conditions. The results in this table indicate that the Markov projection outperforms, on the aggregate level, the freezing projection under different market conditions. However, the improvement comes at substantial computational cost. Table 5.5 presents the normalized computation time of the Markov projection, relative to the freezing approximation. The computation time required for Markov projection is substantially higher and increases in expiry times and tenors. For a realistic swaption matrix, with tenors and expiry dates ranging up to thirty years, the

[2] because the swap rate underlying these swaptions is based on only two underlying LIBORs, one of which is the same as for the corresponding caplet.

Swaption:	1×2	1×3	1×5	1×8	2×3	2×4	2×6	4×5	4×6	7×8
Markov time:	42.4	91.5	133.1	162.4	183.5	233.8	325.1	289.1	448.6	449.3

Table 5.5: The ratio between the computation time of Markov projection and the computation time of the freezing approximation.

Markov projection, the way we use it, may become inefficient for calibration applications, despite the improved accuracy.

Based on these experimental results, we conclude that the freezing approximation for model projection gives satisfactory and acceptable accuracy at relatively low computational costs. It is therefore best suited for the model calibration in the present paper.

5.4.3 Stage 1: Determine the Market Implied Effective Parameters

The first stage of the calibration involves the fitting of the constant parameter displaced Heston model to the market quotes of the swaptions for each expiry and tenor by the pricing formula given in Eq. (5.3.23). Four parameters, $\bar{b}_{n,m}$, $\bar{\lambda}_{n,m}$, η and κ, are involved in this stage of calibration. For calibration problems with a small number of free parameters, the Levenberg-Marquardt method performs sufficiently well.

We first fit one set of parameters, $\bar{b}$, $\bar{\lambda}$, η, κ, for all swaptions of different expiration dates but with the same tenor date, because η and κ also have an impact on the term structure of the curvature, i.e. on the speed at which the curvature flattens in time. By calibrating to swaptions with different expiry dates, we include volatility term structure information in the calibration, which is highly relevant information for η and κ. We repeat this procedure for swaptions of different tenors, set the initial parameter set for a next tenor date equal to the optimal solution of the previous tenor, to ensure that the optimization will reach a minimum in the vicinity of the optimal parameters of the previous tenor.

We obtain the global values for κ and η by averaging the individual $\kappa_{(\cdot)}$- and $\eta_{(\cdot)}$-values across different tenor dates. The resulting parameters are stable across adjacent spot dates (see Table 5.6).

Given the global parameters η and κ, we can determine the local parameters [3], $\bar{b}^{\text{mkt}}_{n,m}, \bar{\lambda}^{\text{mkt}}_{n,m}$, by solving, for all strike values of swaptions, for each combination of expiry n and maturity m, the optimization problem in Eq. (5.4.1).

The resulting skew and volatility parameters, $\bar{b}_{n,m}$ and $\bar{\lambda}_{n,m}$, are shown in Figure 5.7. The local calibration of $\bar{b}_{n,m}$ and $\bar{\lambda}_{n,m}$ for each combination of expiry/maturity results, in general, in a very satisfactory fit to the market swaption quotes. The quality of fit is shown in Figure 5.8.

The calibrated effective parameters $\bar{b}_{n,m}$ and $\bar{\lambda}_{n,m}$ are however *not* always smooth in tenor and expiry dates. If the effective parameters are not smooth, then the time-dependent parameters, which are calibrated to the effective parameters, are expected to be fluctuating too, which may result in calibration

[3] They are local in the sense that there is one pair of parameters, $\bar{b}_{n,m}, \bar{\lambda}_{n,m}$, for a set of swaptions for each combination of expiry n and maturity m.

Table 5.6: Calibrated global parameters.

Date	04-Oct-11	05-Oct-11	06-Oct-11	07-Oct-11	10-Oct-11	11-Oct-11
κ	0.0641	0.0504	0.0569	0.0546	0.0370	0.0523
η	1.5911	1.5600	1.8042	1.4725	1.2539	1.4042

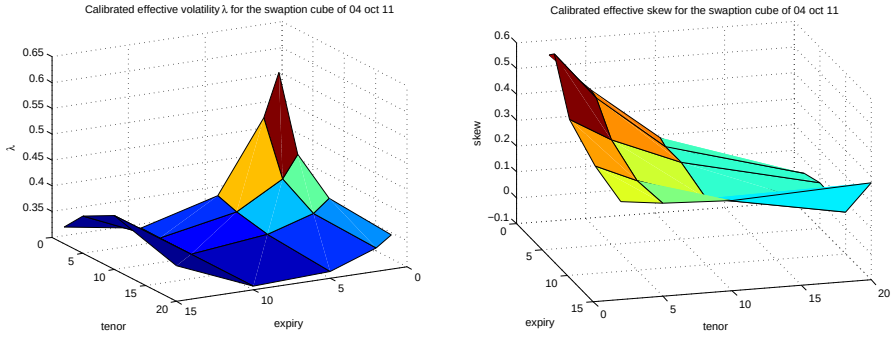

Figure 5.7: Calibrated parameters λ (left) and b (right) for the swaption cube of 04-oct-2011.

problems as the time-dependent parameters are not within certain bounds of feasibility. Therefore, we pre-process the calibrated effective parameters by *a curve smoothing procedure* for the second stage of calibration. We again choose the exponential function in Eq. (5.4.2), as well as a robust least squares method with the Huber loss function,

$$L_\epsilon(x) = \begin{cases} \left(\frac{1}{2}\right) x^2 & \forall |x| \leq \epsilon, \\ \epsilon \left(|x| - \frac{\epsilon}{2}\right) & |x| > \epsilon, \end{cases}$$

to fit the function to existing data points. The advantage of using the Huber loss function compared to a standard sum-of-squares loss function is that it is not very sensitive with respect to undesired outliers. It may capture an overall trend in data points rather than fitting individual erratic data points.

5.4.4 Stage 2: Calibrate the Term-Structure of Volatility

By inserting the model mapping results into Eqs. (5.3.6) and (5.3.19), we establish a connection between the parameters of the TSS-LMM model, $\{\sigma_i(t), t \geq 0\}_i$, and the TSS-LMM implied effective parameters, $\{\bar{\lambda}_{n,m}\}_{n,m}$. More precisely, we describe effective swaption volatility parameters, $\bar{\lambda}_{n,m}$, as functions of the TSS-LMM parameters, $\{\sigma_i(t), t \geq 0\}_i$.

The instantaneous volatility is modelled by a parametric time-homogeneous function, as in Eq. (5.4.2), meaning that all forward LIBOR rates with different fixing dates share the same volatility function which is stationary in the time to expiry. By numerical experiments, we show that this parametric form provides a satisfactory fit to the market volatility structure.

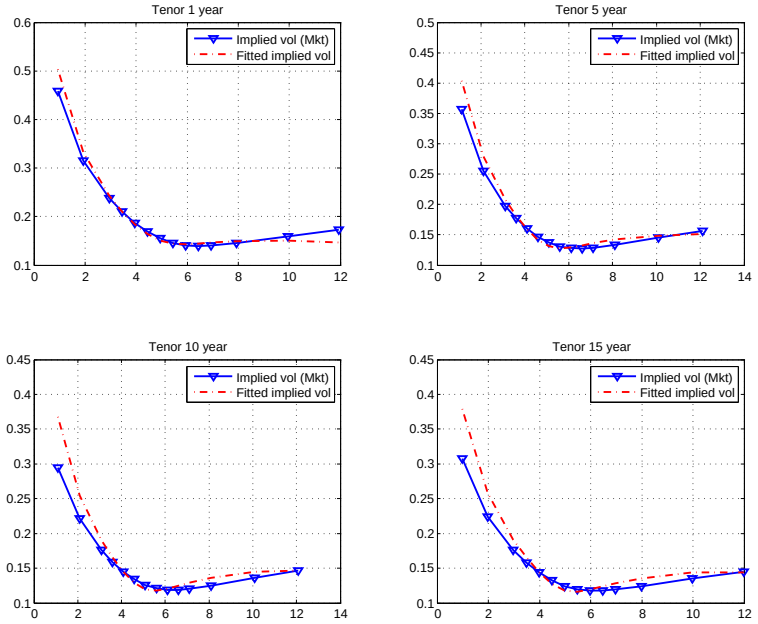

Figure 5.8: Quality of fit of the displaced Heston model with locally calibrated b, λ parameters, and globally calibrated κ, η to 04-oct-2011 swaption implied volatilities of 5 year expiry and various tenors.

With this volatility function, only four parameters, $\{a_1, a_2, a_3, a_4\}$, need to be adjusted to change the effective swaption volatility parameters, $\{\bar{\lambda}_{n,m}\}$, that are functions of the forward LIBOR volatilities, $\{\sigma_i(t), t \geq 0\}_i$. We compare the model parameters, $\{\bar{\lambda}_{n,m}\}$, with the market-implied parameters, $\{\bar{\lambda}^{\text{mkt}}_{n,m}\}$, by minimizing the following objective function:

$$\underset{a_1,a_2,a_3,a_4}{\arg\min} \sum_{n,m} \left(\bar{\lambda}_{n,m}\left(\{\sigma_i(t), t \geq 0\}_i\right) - \bar{\lambda}^{\text{mkt}}_{n,m}\right)^2. \tag{5.4.4}$$

This least-squares minimization problem with parameters $\{a_1, a_2, a_3, a_4\}$ can be solved efficiently by the Levenberg-Marquardt method. In our market data, we consider only 4 pivot tenors, $[2, 5, 10, 20]$ (years), and 5 pivot expiry dates, $[1, 2, 5, 10, 15]$ (years). In total, there are 4×5 effective volatility parameters, $\bar{\lambda}_{n,m}$, and the overall computation time is less than 8 seconds.

Even though the time-dependent volatility function (5.4.2) is meant to match a normal, humped ATM volatility term structure by construction, the new market reality (i.e. the 'reversed humped' volatility term structure as shown in the left-hand side graph of Figure 5.1) does not pose serious difficulties for the modelling of the target volatility – one only has to include negative values for some of the parameters in function (5.4.2) to replicate the reversed humped volatility term structure (see Table 5.7).

The quality of fit of the calibrated effective volatilities and the market implied effective volatilities is illustrated in Figure 5.9. The overall quality is sat-

Table 5.7: Calibrated time-homogeneous volatility function parameters. The calibrated functions are plotted in Figure 5.10.

Date	04-Oct-11	05-Oct-11	06-Oct-11	07-Oct-11	10-Oct-11	11-Oct-11
a_1	0.2200	0.1744	0.1075	0.1915	0.0642	0.0506
a_2	−0.1633	−0.1340	−0.1138	−0.1215	−0.1350	−0.1459
a_3	0.2145	0.2037	0.1815	0.2204	0.1644	0.1584
a_4	0.5596	0.5283	0.5501	0.4674	0.5690	0.6156

isfactory, given the fact that only a small number of parameters (in total four) are used to fit the term structure of the market implied average volatilities, $\{\bar{\lambda}^{\text{mkt}}_{n,m}\}_{n,m}$. The part that is not fitted well is due to the time-inhomogeneity in the forward rate correlation, which we assume to be stationary here.

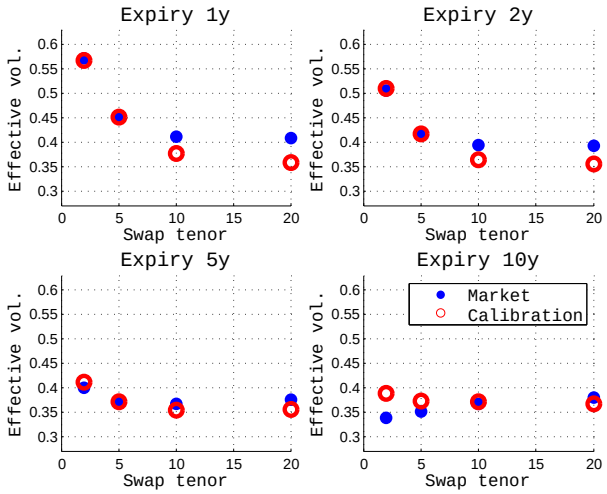

Figure 5.9: Calibrated effective volatility and the market implied effective volatility.

The calibrated parameters of the time-homogeneous volatility function in Table 5.7 exhibit a satisfactory stability over time. As a result, in the time-homogeneous function in Figure 5.10, except for a few small parallel shifts, the shape of the volatility function remains similar across different spot dates.

5.4.5 Stage 3: Calibrate the Term-Structure of Skews

The complexity of volatility calibration is simplified by the fact that the volatility structure is modelled by the four-parameter exponential function. In the case of the skew calibration, however, we cannot use a similar approximation, because the pattern in the swaption skew does not appear to be time-homogeneous [86]. We, therefore, model the market swaption skew pattern by a flexible skew structure, assuming that the skew parameters, $\{b_i(t), t \geq 0\}_i$, are *annually piecewise*

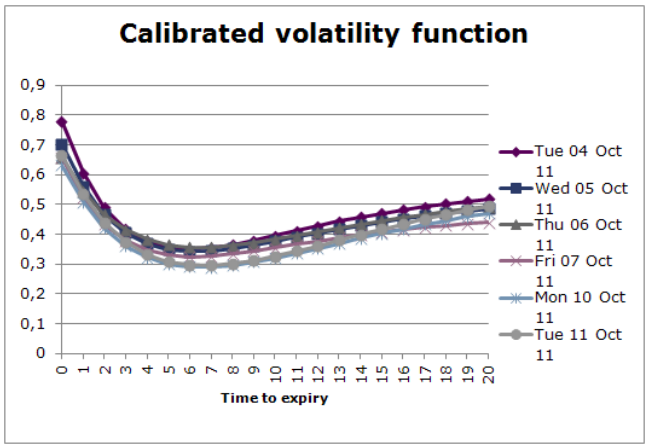

Figure 5.10: The plot of calibrated time-homogeneous volatility functions across several spot dates.

constant, i.e. $\{b_i(t), 0 \leq t \leq 1\} \equiv b_i(0:1)$, $\{b_i(t), 1 \leq t \leq 2\} \equiv b_i(1:2)$, $\ldots, i = 1, 2, \ldots, N-1$.

In principle, we can formulate the calibration problem as a least squares problem, as in the previous calibration stage, i.e.

$$\underset{\{b_i(t), t\geq 0\}_i}{\arg\min} \sum_{n,m} \left(\bar{b}_{n,m}\left(\{b_i(t), t \geq 0\}_i\right) - \bar{b}^{\text{mkt}}_{n,m}\right)^2. \tag{5.4.5}$$

However, minimization of this objective function by standard methods may give rise to numerical problems, as a time-dependent skew leads essentially to too many free variables. For example, with a realistic market swaption matrix of 30 expiry dates and 30 tenor dates, 60 annually forward rates need to be considered. With the assumption of annually piecewise constant skew parameters, a total of 1350 free parameters should be calibrated. Straightforward application of non-linear optimization techniques will not lead to stable parameters, and the convergence will typically be slow.

Here, we therefore employ a stable and numerically efficient optimization method to calibrate the time-dependent skew parameters (with a large number of free variables), based on a quadratic programming (QP) problem formulation, i.e. we find a vector, $\mathbf{b}$, containing the piecewise constant skew parameters, in

$$\min \quad \mathbf{b}^T M \mathbf{b}, \text{ subject to } A\mathbf{b} = \bar{\mathbf{b}}^{\text{mkt}}, \quad -1 \leq \mathbf{b} \leq 1, \tag{5.4.6}$$

In this formulation, the correspondence between the model and market skew is not in the objective function, as in Eq. (5.4.5), but a constraint in the optimisation problem, $A\mathbf{b} = \bar{\mathbf{b}}^{\text{mkt}}$. We write the effective model skew, $\bar{b}_{n,m}$, as a linear combination of the time-dependent forward rate skews, $\mathbf{b}$, so that the left-hand side of the constraint equality contains the value of a vector of the model effective skew (see Section 5.4.5). The smoothness of the solution is enhanced by smoothing operators in the objective function of the optimization problem. As a result, by solving the optimization problem, we obtain a smooth solution in the set of feasible solutions which satisfies the constraints. The details are presented in the following subsections.

Integration Weights

Next to the dimensionality of the optimization problem, the computation time is also influenced negatively by the integration of the time-dependent skew parameters over the projection weights, $p_{(\cdot)}(\cdot,\cdot)$, and the parameter averaging weights, $\omega_{(\cdot)}(\cdot)$ in Eq. (5.3.17), during the computation of $\bar{b}_{n,m}$ by Eq. (5.3.16) for each combination of expiration and tenor date. In each optimization step, the time-dependent skew is updated, so that the integrals for all intermediate weights would have to be re-computed for all combinations of expiry and maturity dates.The computation time can be significantly reduced when the computation of the projection weights can be performed independently of the update of the time-dependent skew, $b_{(\cdot)}(\cdot)$, and when the projection weights can be stored.

This is possible if we model the time-dependent skew by a *piecewise constant function*, i.e. constant for each year interval. As an example, let us compute the effective skew of a swaption of expiry 2 years with a 3 year tenor (thus a maturity of 5 years). The effective skew, given by formula (5.3.16), reads

$$\bar{b}_{2,5} = \int_0^{T_2=2} b_2(t)p_2(2,5)\omega_{25}(t)dt + \int_0^{T_2=2} b_3(t)p_3(2,5)\omega_{25}(t)dt + \int_0^{T_2=2} b_4(t)p_4(2,5)\omega_{25}(t)dt.$$

By choosing annually piecewise constant skew parameters, we can rewrite the skew formula as

$$\begin{aligned}\bar{b}_{2,5} =& b_2(0:1)\int_0^1 p_2(2,5)\omega_{2,5}(t)dt + b_3(0:1)\int_0^1 p_3(2,5)\omega_{2,5}(t)dt \\ &+ b_4(0:1)\int_0^1 p_4(2,5)\omega_{2,5}(t)dt + b_2(1:2)\int_1^2 p_2(2,5)\omega_{2,5}(t)dt \\ &+ b_3(1:2)\int_1^2 p_3(2,5)\omega_{2,5}(t)dt + b_4(1:2)\int_1^2 p_4(2,5)\omega_{2,5}(t)dt,\end{aligned}$$

so that the integral is independent of the skew parameters, and the number of terms to be stored is equal to the number of tenor dates times the expiry dates, which increases in the tenor and expiry dates. Let's consider an example of a 3×3 swaption matrix, and for notational convenience denote the integrals by $\zeta^j_{(2,5)}(t_1 : t_2) := \int_{t_1}^{t_2} p_j(2,5)\omega_{2,5}(t)dt$ with $j = [1,2,3]$, in which the subscript denotes the payment term starting from swaption expiry and spanned by the swaption tenor, the superscript indicates the associated underlying forward rate and within the brackets we have the time interval of integration. We group the cached integrals by tenors and expiries in Table 5.8.

Tenor / *Expiry*	$1y$	$2y$	$3y$
$1y$	$\zeta^1_{1,2}(0:1)$	$\zeta^1_{1,3}(0:1), \zeta^2_{1,3}(0:1)$	$\zeta^1_{1,4}(0:1), \zeta^2_{1,4}(0:1), \zeta^3_{1,4}(0:1)$
$2y$	$\zeta^2_{2,3}(0:1)$ $\zeta^2_{2,3}(1:2)$	$\zeta^2_{2,4}(0:1), \zeta^3_{2,4}(0:1)$ $\zeta^2_{2,4}(1:2), \zeta^3_{2,4}(1:2)$	$\zeta^2_{2,5}(0:1), \zeta^3_{2,5}(0:1), \zeta^4_{2,5}(0:1)$ $\zeta^2_{2,5}(1:2), \zeta^3_{2,5}(1:2), \zeta^4_{2,5}(1:2)$
$3y$	$\zeta^3_{3,4}(0:1)$ $\zeta^3_{3,4}(1:2)$ $\zeta^3_{3,4}(2:3)$	$\zeta^3_{3,5}(0:1), \zeta^4_{3,5}(0:1)$ $\zeta^3_{3,5}(1:2), \zeta^4_{3,5}(1:2)$ $\zeta^3_{3,5}(2:3), \zeta^4_{3,5}(2:3)$	$\zeta^3_{3,6}(0:1), \zeta^4_{3,6}(0:1), \zeta^5_{3,6}(0:1)$ $\zeta^3_{3,6}(1:2), \zeta^4_{3,6}(1:2), \zeta^5_{3,6}(1:2)$ $\zeta^3_{3,6}(2:3), \zeta^4_{3,6}(2:3), \zeta^5_{3,6}(2:3)$

Table 5.8: Integration weights for the yearly piecewise constant skew in order to obtain the effective skew $\bar{b}_{(\cdot)}$.

System of Equations and the Matrix Representation

By using the weights given in Table 5.8, we can write

$$\begin{aligned}
\bar{b}_{1,2} &= b_1(0:1)\zeta^1_{1,2}(0:1),\\
\bar{b}_{1,3} &= b_1(0:1)\zeta^1_{1,3}(0:1) + b_2(0:1)\zeta^2_{1,3}(0:1),\\
\bar{b}_{1,4} &= b_1(0:1)\zeta^1_{1,4}(0:1) + b_2(0:1)\zeta^2_{1,4}(0:1) + b_3(0:1)\zeta^3_{1,4}(0:1),\\
\bar{b}_{2,3} &= b_2(0:1)\zeta^1_{2,3}(0:1) + b_2(1:2)\zeta^1_{2,3}(1:2),\\
\bar{b}_{2,4} &= b_2(0:1)\zeta^1_{2,4}(0:1) + b_3(0:1)\zeta^2_{2,4}(0:1)\\
&+ b_2(1:2)\zeta^1_{2,4}(1:2) + b_3(1:2)\zeta^2_{2,4}(1:2),\\
\bar{b}_{2,5} &= b_2(0:1)\zeta^1_{2,5}(0:1) + b_3(0:1)\zeta^2_{2,5}(0:1) + b_4(0:1)\zeta^3_{2,5}(0:1),\\
&+ b_2(1:2)\zeta^1_{2,5}(1:2) + b_3(1:2)\zeta^2_{2,5}(1:2) + b_4(1:2)\zeta^3_{2,5}(1:2),\\
\vdots &\quad \vdots
\end{aligned}$$

Equating $\bar{b}_{n,m}$ and $\bar{b}^{\text{mkt}}_{n,m}$, for all n, m, gives us the following system of equations:

$$\begin{aligned}
\bar{b}^{\text{mkt}}_{1,2} &= b_1(0:1)\zeta^1_{1,2}(0:1),\\
\bar{b}^{\text{mkt}}_{1,3} &= b_1(0:1)\zeta^1_{1,3}(0:1) + b_2(0:1)\zeta^2_{1,3}(0:1),\\
\bar{b}^{\text{mkt}}_{1,4} &= b_1(0:1)\zeta^1_{1,4}(0:1) + b_2(0:1)\zeta^2_{1,4}(0:1) + b_3(0:1)\zeta^3_{1,4}(0:1),\\
\bar{b}^{\text{mkt}}_{2,3} &= b_2(0:1)\zeta^1_{2,3}(0:1) + b_2(1:2)\zeta^1_{2,3}(1:2),\\
\bar{b}^{\text{mkt}}_{2,4} &= b_2(0:1)\zeta^1_{2,4}(0:1) + b_3(0:1)\zeta^2_{2,4}(0:1) + b_2(1:2)\zeta^1_{2,4}(1:2) + b_3(1:2)\zeta^2_{2,4}(1:2),\\
\bar{b}^{\text{mkt}}_{2,5} &= b_2(0:1)\zeta^1_{2,5}(0:1) + b_3(0:1)\zeta^2_{2,5}(0:1) + b_4(0:1)\zeta^3_{2,5}(0:1) + b_2(1:2)\zeta^1_{2,5}(1:2) +\\
&\quad b_3(1:2)\zeta^2_{2,5}(1:2) + b_4(1:2)\zeta^3_{2,5}(1:2),\\
\vdots &\quad \vdots
\end{aligned}$$

We can write the system of linear equality constraints in matrix form, $A\mathbf{b} = \bar{\mathbf{b}}^{\text{mkt}}$, where

$$\bar{\mathbf{b}}^{\text{mkt}} = \left[\bar{b}^{\text{mkt}}_{1,2}, \bar{b}^{\text{mkt}}_{13}, \bar{b}^{\text{mkt}}_{1,4}, \bar{b}^{\text{mkt}}_{2,3}, \bar{b}^{\text{mkt}}_{2,4}, \bar{b}^{\text{mkt}}_{2,5}, \bar{b}^{\text{mkt}}_{3,4}, \bar{b}^{\text{mkt}}_{3,5}, \bar{b}^{\text{mkt}}_{3,6}\right]^T,$$

and

$$\mathbf{b} = \big[b_1(0:1), b_2(0:1), b_3(0:1), b_4(0:1), b_5(0:1), b_2(1:2), \\ b_3(1:2), b_4(1:2), b_5(1:2), b_3(2:3), b_4(2:3), b_5(2:3)\big]^T.$$

Note that $b_1(1:2)$, $b_1(2:3)$ and $b_2(2:3)$ do not appear in the vector, because the fixing dates of the corresponding forward rates have been exceeded, and these forward rates are thus known.

Matrix A is a matrix with as its elements the weights in Table 5.8. The matrix consists of three blocks, each of which represents the weights associated with an interval, $A = [A_1, A_2, A_3]$, where

$$A_1 = \begin{pmatrix} \zeta_{12}^1(0:1) & 0 & 0 & 0 & 0 \\ \zeta_{13}^1(0:1) & \zeta_{13}^2(0:1) & 0 & 0 & 0 \\ \zeta_{14}^1(0:1) & \zeta_{14}^2(0:1) & \zeta_{14}^3(0:1) & 0 & 0 \\ 0 & \zeta_{23}^1(0:1) & 0 & 0 & 0 \\ 0 & \zeta_{24}^1(0:1) & \zeta_{24}^2(0:1) & 0 & 0 \\ 0 & \zeta_{25}^1(0:1) & \zeta_{25}^2(0:1) & \zeta_{25}^3(0:1) & 0 \\ 0 & 0 & \zeta_{34}^1(0:1) & 0 & 0 \\ 0 & 0 & \zeta_{35}^1(0:1) & \zeta_{35}^2(0:1) & 0 \\ 0 & 0 & \zeta_{36}^1(0:1) & \zeta_{36}^2(0:1) & \zeta_{36}^3(0:1) \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \zeta_{23}^1(1:2) & 0 & 0 & 0 \\ \zeta_{24}^1(1:2) & \zeta_{24}^2(1:2) & 0 & 0 \\ \zeta_{25}^1(1:2) & \zeta_{25}^2(1:2) & \zeta_{25}^3(1:2) & 0 \\ 0 & \zeta_{34}^1(1:2) & 0 & 0 \\ 0 & \zeta_{35}^1(1:2) & \zeta_{35}^2(1:2) & 0 \\ 0 & \zeta_{36}^1(1:2) & \zeta_{36}^2(1:2) & \zeta_{36}^3(1:2) \end{pmatrix}$$

and

$$A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \zeta_{34}^1(2:3) & 0 & 0 \\ \zeta_{35}^1(2:3) & \zeta_{35}^2(2:3) & 0 \\ \zeta_{36}^1(2:3) & \zeta_{36}^2(2:3) & \zeta_{36}^3(2:3) \end{pmatrix}.$$

The dimension of matrix A is, of course, related to the problem set-up. The number of rows depends on the number of swaptions considered. Thus we are dealing here with nine rows in matrix A. The block matrix A_1 corresponds to the evolution of the five underlying forward rates in the time interval $[0, t_1]$. Block matrices A_2 and A_3 are connected with the evolution of the five underlying forward rates in the time intervals $[t_1, t_2]$ and $[t_2, t_3]$, respectively. Since forward

rate F_1 is known at t_1, the first column is removed from matrix A_2. Similarly, first two columns are not present in A_3.

The size of solution vector $\mathbf{b}$, i.e. the number of unknowns, is larger than the number of equality constraints, which makes the problem *under-determined*. In general, an infinite number of solutions exist to the system, and these solutions form a convex set in the real domain.

In order to obtain a numerically stable solution, one has to add penalty functions or smoothness regularisation terms to restrict the solution space. One of the choices is to prescribe the time-homogeneity of the solution, as suggested in [9], i.e.

$$\min \sum_{i,j} \left(b_i(T_j) - b_{i-1}(T_{j-1})\right)^2,$$

which can be described by $\min_{\mathbf{b}} \mathbf{b}^T (R^T R)\mathbf{b}$, with R a sparse matrix,

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Since the quadratic terms will be larger or equal to zero, it follows that the matrix $M = R^T R$ is a positive semi-definite matrix by definition, i.e. $\mathbf{b}^T M \mathbf{b} \geq 0$. The shape of matrix M is shown in the left-hand side graph of Figure 5.11.

Additional Smoothness Conditions

We add additional desirable features of the skew pattern in the above optimisation problem, as a set of additional calibration objectives, next to the time-homogeneity. For example, we require

1. Smoothness in calendar time, i.e. $\min \sum \left(b_i(T_j) - b_i(T_{j-1})\right)^2$,

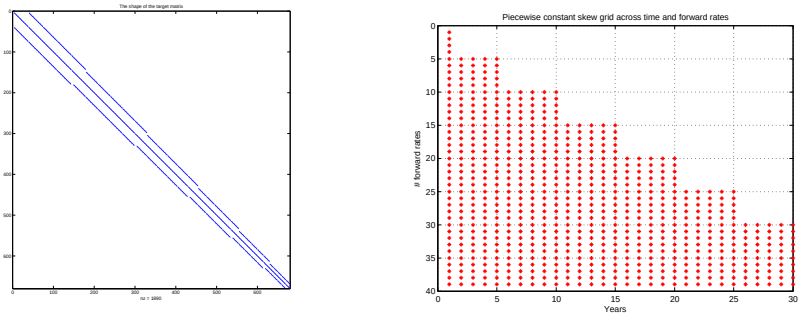

Figure 5.11: LHS: The target matrix M for a 30 by 30 swaption matrix with the time-homogeneity constraint. RHS: A 'skew pattern' for a (30×30) swaption matrix with expiry dates at $1, 5, 10, 15, 20, 25, 30$ years.

2. Homogeneity of neighbouring forward rates: $\min \sum \left(b_i(T_j) - b_{i-1}(T_j)\right)^2$.

Both objectives can be cast into matrix form and they can easily be added to matrix R by adding additional rows. The matrix corresponding to the smoothness among neighbouring forward rates can be written as

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

Matrix M' is now defined by means of an updated matrix R', i.e. $M' = R'^T R'$, where matrix R' contains the nonzero elements of matrix R, augmented with additional nonzero elements representing the additional smoothing requirements..

Solutions via Convex Optimization

Now we are ready to solve the calibration problem by determining the solution from the following quadratic programming (QP) problem: Find vector $\mathbf{b}$ in

$$\min \quad \mathbf{b}^T M \mathbf{b}, \quad \text{subject to} \quad A\mathbf{b} = \bar{\mathbf{b}}^{\text{mkt}}, \quad -1 \leq \mathbf{b} \leq 1.$$

The advantage of dealing an optimization problem in SQP form is the convexity of the problem which brings beneficial convergence properties, like the solution to the problem is a global optimum, and Newton-type algorithms can be employed with polynomial complexity. The use of Newton's method leads to highly efficient solution methods.

We have used the convex problem solution package, SeDuMi, by Sturm [102] to solve the quadratic programming problem. The solution method is efficient for problems with approximately 1000 parameters. The problem is then solved within half a second.

We first test the quality of the SPQ formulation by solving a sample problem, which was presented in Section 9 of the paper [86].

We show the array of the skew parameters that are determined by the dots in the right-hand side plot of Figure 5.11, where each row represents skew parameters for a forward rate at a different time point, and a column shows the number of "alive" forward rates at a time point.

All entries in the lower triangular matrix can in principle be determined within the optimization. However, only swaption prices at pivot points are available (expiry 1y, 5y, 10y and so on). Information regarding forward rates 2, 3, and 4 at the years 2, 3, and 4 is not available, and those values are determined only by means of the smoothing operators, interpolating between the nodes. This is the reason why those points are not shown as node points in the array with the skew parameters.

The smoothing operator in the objective function prevents arbitrary values within $-1 \leq \mathbf{b} \leq 1$. The calibrated skew surface is presented in the left-hand

side plot of Figure 5.12. It resembles the solution surface presented in Figure 2 of [86] very well.

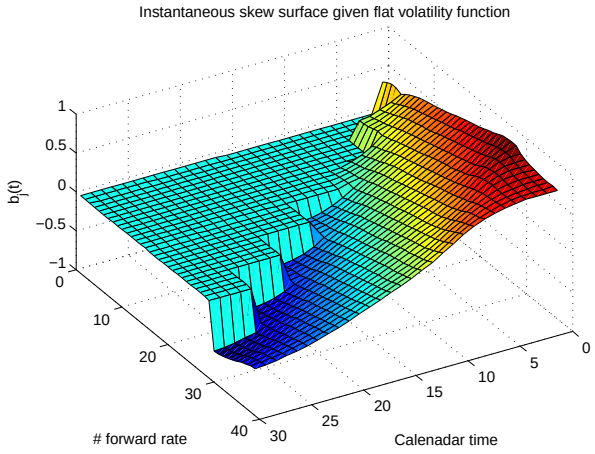

Figure 5.12: The calibrated instantaneous skew surface for the sample problem in [86].

We now present the calibration results based on the application of the SQP method for the market data under consideration. In the formulation of the optimization problem, the correspondence between the model and the market implied skews is prescribed in the constraints. When a feasible solution is determined the correspondence between the two skews is obtained. With the smoothing matrices in the objective function, a smooth solution in terms of 'time-homogeneity' and 'smoothness in calendar time' criteria (see Section 5.4.5) within the set of feasible solutions is prescribed - see the result in the left-hand side graph of Figure 5.13. The impact of the smoothing matrices can be observed, since we also present the skew surface for an SQP formulation without smoothing operators, see the right-hand graph of Figure 5.13. A comparison shows that the solution without any smoothing operator includes jumps in the calibrated model parameters, which is undesirable for a practical model. The computation time of this stage is less than half a second.

5.5 Conclusion

In this chapter, we have discussed the calibration issues arising in practical applications of the Term Structure of the Skew Libor Market Model (TSS-LMM). Our major contribution is a convex optimization formulation of the time-dependent skew parameter calibration problem. Due to the large number of free variables, the traditional least square formulation of the problem gives rise to a highly non-linear optimization problem, which is difficult to solve. But in our formulation, the problem has been translated into a simple quadratic programming form with matrix-vector formulas. By this formulation, we are able to effectively locate the global optimal parameter set consisting of a large number of free skew parameters within a very short computation time.

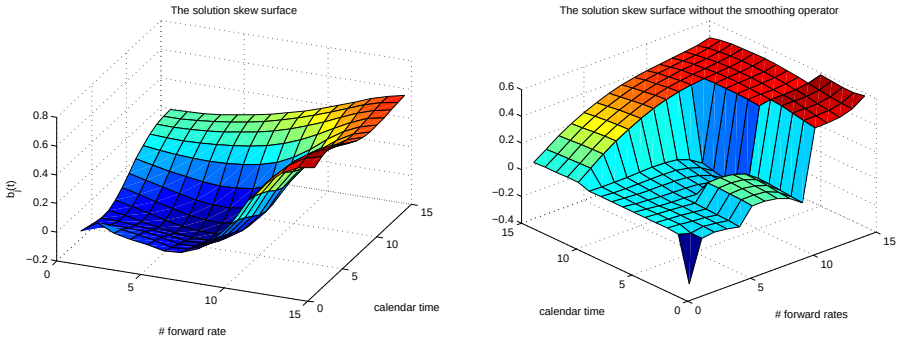

Figure 5.13: The calibrated instantaneous skew surface for the current market data set.

Due to the complicated joint forward LIBOR dynamics implied by TSS-LMM, the calibration of the model to the swaption quotes has to rely on a model-mapping procedure, which relates the time-dependent model parameters in a TSS-LMM model to the resulting swaption prices. In the first step of model mapping procedure, the high-dimensional swap rate dynamics implied by the model are mapped onto a one-dimensional displaced diffusion process with time-dependent coefficients. In the next step, the effective constant parameters are obtained from the time-dependent parameters of the projected model via a parameter averaging technique. Two known projection methods available in the literature, the freezing projection and the more involved Markov projection, have been compared within the calibration process.

Although the Markovian projection is generally more accurate than the basic freezing technique, the difference is merely marginal, especially when a single swaption is considered. The Markov projection is, however, computationally much more intensive than the freezing projection. Thus we have concluded that the basic freezing projection achieves an acceptable accuracy at significantly less computational cost, and that it is thus suitable within the calibration purpose.

In addition to the numerical performance, the second advantage of the freezing projection is that we can express the effective skew implied by the TSS-LMM as a linear combination of time-dependent forward LIBOR rate skew parameters. We then resolve the problem of determining the high-dimensional time-dependent skew as a convex optimization problem. The numerical stability of the calibrated parameters should enable efficient daily re-calibration and re-hedging.

The other parts of the model are calibrated in different calibration stages. As a result, we have defined an efficient black-box calibration tool for the TSS-LMM.

To improve the quality or the smoothness of the solution, one can add more intermediate nodes so that there are more free variables and possibly a combination with increased smoothness can be found, fitting the same set of calibration instruments.

An interesting future research challenge is to also place the calibration of the volatility for the TSS-LMM into convex optimization form, so that the volatility term structure can also be matched with high quality. At present, an apparent

difficulty in the calibration stages is the parameter averaging procedure for the volatility, which involves the solution of a series of Riccati equations.

Appendix A: Displaced Diffusion and Anti-Displaced Diffusion

In a displaced diffusion framework, the forward rates are defined as follows:

Definition 5.5.1. *A process, SR, is based on displaced diffusion dynamics, with displacement $\theta \in \mathbb{R}$, if*

$$dSR(t) = \alpha(SR(t) - \theta)dW_t, \qquad SR(0) > \theta, \alpha > 0, \tag{5.5.1}$$

where W_t is a Brownian motion.

Here, $\mathrm{SR} - \theta$ is a geometric Brownian motion ($\in \theta, \infty$), without drift, with volatility α. We use an alternative formulation, as follows

$$d\mathrm{SR}(t) = \sigma \left(b\mathrm{SR}(t) + (1-b)\mathrm{SR}(0)\right) dW_t. \tag{5.5.2}$$

Despite the difference in presentation, the processes are equivalent, if

$$d\mathrm{SR}(t) = \sigma b \left(\mathrm{SR}(t) - \frac{b-1}{b}\mathrm{SR}(0)\right) dW_t \Rightarrow \begin{cases} \theta = \frac{b-1}{b}\mathrm{SR}(0), \\ \alpha = \sigma b. \end{cases}$$

The advantage of formulation (5.5.2) is the ease of interpretation. The displaced diffusion process is a combination of a normal model and a lognormal process depending on parameter b.

5.5.1 Implied Volatility in Displaced Diffusion

For $\bar{K} > \theta$, a European call option on a displaced lognormal SR-process has the value

$$\mathbb{E}\big[(\mathrm{SR}(T) - \bar{K})^+\big] = \mathbb{E}\big[\left(\big(\mathrm{SR}(T) - \theta\big) - (\bar{K} - \theta)\right)^+\big]. \tag{5.5.3}$$

Plain vanilla options can still be priced using the Black-Scholes model with displaced parameters.

We can compute the option price under displaced lognormal dynamics by application of the Black-Scholes's formula:

$$C^{\mathrm{DD}}(\mathrm{SR}(0), \bar{K}, \sigma_{\mathrm{impv}}, T) = C^{\mathrm{BS}}(\mathrm{SR}(0) - \theta, \bar{K} - \theta, \alpha, T),$$

where σ_{impv} denotes the implied Black-Scholes volatility of the option price of the given strike and maturity.

The left-hand graph of Figure 5.14 shows that the implied volatility skew is highest for $b = 0$. This is the so-called 'normal skew limit' of a displaced diffusion model. In practice, the skews observed in the equity options [70] and plain vanilla interest rate caplets and swaptions often exceed the 'normal skew limit'. The displaced *anti-lognormal* model can deal with the steepness constraint.

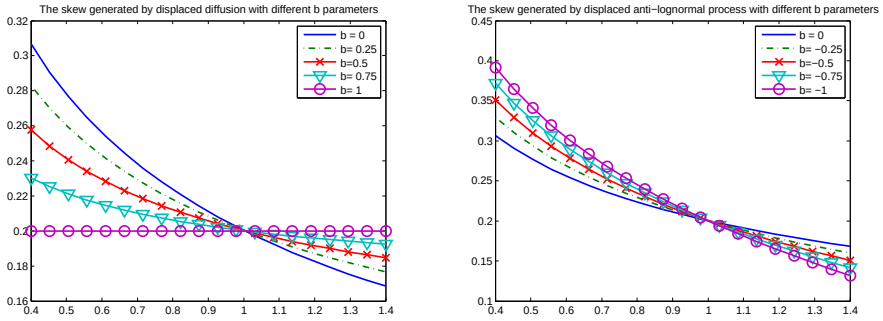

Figure 5.14: LHS: The steepness in skews given parameters b in Eq. (5.5.2). RHS: The steepness in skews given parameters b in Eq. (5.5.2).

Definition 5.5.2 (Displaced Anti-lognormal process)**.** *A process, SR, follows a displaced anti-lognormal process, if*

$$dSR(t) = \alpha(SR(t) - \theta)dW_t, \quad 0 < SR(0) < \theta, \alpha < 0, \tag{5.5.4}$$

where W is a Brownian motion.

This process is connected to negative b-values in Eq. (5.5.2), i.e.

$$d\mathrm{SR}(t) = \sigma b \left(\mathrm{SR}(t) - \frac{b-1}{b}\mathrm{SR}(0) \right) dW_t \Rightarrow \begin{cases} \theta = \frac{b-1}{b}\mathrm{SR}(0) > \mathrm{SR}(0), \\ \alpha = \sigma b < 0. \end{cases}$$

As a result of the above definition, $\theta - \mathrm{SR}$ is a geometric Brownian motion ($\in (-\infty, \theta)$), without drift, with volatility $-\alpha > 0$.

To price a call with $\bar{K} < \theta$, under a displaced anti-lognormal SR process, we have

$$\begin{aligned} \mathbb{E}(\mathrm{SR}(T) - \bar{K})^+ = \mathbb{E}\big(\theta - \bar{K} - (\theta - \mathrm{SR}(T))\big)^+ &= P^{\mathrm{BS}}(\theta - \mathrm{SR}(0), \theta - \bar{K}, -\alpha, T) \\ &= C^{\mathrm{BS}}(\mathrm{SR}(0) - \theta, \bar{K} - \theta, \alpha, T). \end{aligned}$$

This formula for call options in the anti-displaced diffusion model [70] has the same form as a displaced diffusion process. Here the first three arguments are negative, but this is no problem for the Black-Scholes formula.

CHAPTER 6

Conclusions

In this thesis we have discussed several problems arising from the application of the Stochastic-Alpha-Beta-Rho (SABR) model when pricing either plain vanilla options or exotic options. We have discussed the modelling as well as the numerical treatment.

In Chapter 2, we have presented a convexity correction for Constant Maturity Swap (CMS) products under a two-factor SABR model in the form of a variance and covariance formulation of the two driving factors. We derived an approximation for the convexity correction by applying the *small time asymptotic expansion* technique to the Wiener processes involved. As a result, we obtained an efficient and easy-to-implement approximation formula for the CMS convexity correction.

In Chapter 3, we have again applied the *small time asymptotic expansion*, but now to the problem of approximating the first and second moments of the integrated variance of the log-normal volatility process. This has been done in the context of defining a low-bias discretization scheme for the SABR model. With the approximated moment information, we can approximate the density of the integrated variance by means of a log-normal distribution with its first two moments matched to that information. We have shown that the conditional SABR process, given the terminal volatility level and the integrated variance, is a squared Bessel process, for which we found an explicit distribution function in the form of noncentral chi-square distributions. Based on the idea of mixing conditional distributions and a direct inversion of the noncentral chi-square distributions, we have proposed the low-bias SABR Monte Carlo scheme. The scheme proposed can deal with the – often problematic – behaviour of the CEV process in the vicinity of the zero boundary. The low-bias scheme is stable and exhibits a highly satisfactory convergence behaviour compared to the truncated Euler scheme.

In Chapter 4, we extended the discretization scheme proposed in Chapter 3 towards a SABR model with stochastic interest rate in the form of a Hull-White short rate model, the SABR-HW model. The hybrid model is meant for pricing long-dated equity-interest-rate linked exotic options with exposure to both the interest rate and the equity price risk. For the calibration of the

SABR-HW model, we proposed a projection formula, mapping the SABR-HW model parameters onto the parameters of the *nearest* SABR model. The inverse of the projection formula enables a rapid calibration of the model.

In Chapter 5, we have considered a version of stochastic volatility LIBOR Market Model with time-dependent skew and volatility parameters. As a result of the time-dependent parameters, the model has the flexibility to match to the market quotes of an entire swaption cube (in terms of various combinations of expiry, tenor and strike), as observed in the current interest rate market. Compared to the interest rate models proposed in the previous chapters of this thesis, the former ones can be seen as so-called local models, meant to fit a selective part of the market information to accurately price *a small subset of target instruments*. The stochastic volatility LIBOR Market Model considered in Chapter 4, more specifically the TSS-LMM, however, is a global model, which can be fitted to many different forms occurring in the current market quotes of swaptions and caps. Thus, this model is in principle well-suited for managing the risk of the complete exotic option trading book in a financial institution, consisting of both exotic options and its plain vanilla hedge instruments.

The calibration of the model to the swaption quotes has to rely on a model-mapping procedure, which relates the time-dependent parameters in a TSS-LMM model to the resulting swaption prices. In the first step of model mapping procedure the high-dimensional swap rate dynamics implied by the model are mapped onto a one-dimensional displaced diffusion process with time-dependent coefficients. After that the effective constant parameters are obtained by averaging the projected model via a parameter averaging technique. Two known projection methods, available in the literature, the freezing projection and the more involved Markov projection have been compared within the calibration process. We concluded that the basic freezing projection achieves an acceptable accuracy at significantly lower computational cost, and that it is thus suitable within the calibration purpose.

A second advantage of the freezing projection formula is that it enabled us to find a convex optimization formulation to determine the time-dependent skew parameters. Our contribution in this chapter was the convex optimization formulation of the skew calibration problem. With the convex formulation, we were able to cast the calibration problem, with a large number of free variables, into a well-known quadratic programming form, which can be efficiently solved by specialized algorithms in only a few seconds. Furthermore the obtained solution is a global optimum by the convexity of the problem. The stability of the procedure may be beneficial for application in the day-to-day derivative trading practice, i.e. the daily re-calibration and hedging.

6.1 Outlook

> *Ten years ago if you had suggested that a sophisticated investment bank did not know how to value a plain vanilla interest rate swap, people would have laughed at you. But that isn't too far from the case today.*
>
> – Deus Ex Machiatto, June 23, 2010

The 'back-to-basic' and various other industrial trends mentioned in the introduction of this thesis, keep the derivative trading business interesting. Nowadays, even the pricing of basic derivatives, like interest rate forward agreements, swaps, or plain vanilla swaptions, have become complicated tasks that require careful analysis and sophisticated models.

For example, previously derivatives pricing theory was based on the assumption that one can replicate the value of a derivative by borrowing and lending money and other securities at a unique risk-free rate. The reality we are faced with at a derivatives trading desk today is rather different. After the 2008 crisis, the historically stable relationship between a bank's funding rate, the government rates, and the LIBOR rates shows inconsistencies [89]. For example, the LIBOR rate, at which the international banks (AA rated) borrow at short maturities (say 3 months) is now a rather risky rate, which is significantly higher (up to 50 basis points) than the true risk free rate, obtained by compounding the overnight interest rate.

This inconsistency between previously-equivalent market rates can be explained by credit risk, liquidity risk and other effects [78]. Previously, it was assumed that the international bank short horizon default rates were effectively zero. However, the crisis of 2008 showed that the bank balance sheets are opaque. Banks have valued illiquid assets not by marking-to-market, but by marking to models. The unreliability of the accounting information given by the banks created a 'jump to default' risk and non-negligible credit risk at the short time horizon [106].

Moreover, there is also a difference between the six month LIBOR deposit rate and the three month LIBOR rated deposit rate that is rolled over two consecutive 3 month LIBOR rated deposits. Before the crisis, by compounding two consecutive 3-month forward LIBOR rates, one could recover or, at least, closely approximate the corresponding 6-month forward LIBOR rate. After the crisis, we see that the forward LIBOR rates implied by compounding two consecutive forward LIBOR rates started to deviate significantly from the 6 month rates (see Figure 6.1). This implies that an interest rate swap, whose floating leg is three month LIBOR is not of the same value as a swap whose floating leg is the six month LIBOR.

The phenomena outlined above will affect the funding cost of a derivative trading desk and hence affect the valuation of all trades. This is because the capital to fund the operations of a derivative desk comes from multiple sources. The desk can borrow/lend money from the money market based on the bank's credit rating, or it can fund a trade based on collateral as in a credit support annex (CSA) agreement, mentioned in the introduction to this thesis. As collateral is used as a compensation for liabilities in case of a default, it can be thought of as a risk-free investment. In general, a collateralized funding will have a favourable interest/discount rate compared to an unsecured funding. Moreover, an unsecured funding at different payment frequencies will also give rise to different interest rates reflecting the differences in the embedded credit and liquidity risks. A funding with a smaller payment frequency, e.g. three months, will be based on a lower interest rate than that of a six month or even longer term funding. As a result, different funding structures adopted in each individual contract/trade, e.g. the payment frequency, CSA or not CSA, even which currency is to be settled on, will give rise to different derivative valuations.

Because of this new reality in the interest rate derivative market, practi-

Figure 6.1: Basis spread between 5y 3 month tenor swap rates and 5 year 6 month tenor swap rates, from 02-01-2006 to 02-01-2010, EUR market [78]

tioners came up with several empirical or, in other words, pragmatic solutions. For example, for a given contract, a specific discount curve is selected, consistent with the contractual features and the counterparty under consideration, to calculate the net present value (NPV) of the contract's future payments.

The assumption of the co-existence of various discounting and forward curves for a single currency is to some extent contradictory to the classic pricing theory, which assumes a unique zero-coupon curve containing all relevant information about the risk neutral estimation of future rates and the NPV calculation of associated payoffs [78]. In order to resolve the inconsistency in the empirical approaches, a new model paradigm is required to model the new market reality of multiple discounting and forward curves for different tenors. One popular approach is to extend the current stochastic volatility LIBOR market model to the so-called basis spread LIBOR market model setting, by modelling the forward rates and the basis spreads associated with different yield curves jointly, treating the dynamics of the basis spread as an additional risk factor. The additional factors add flexibility to the model, but, at the same time, they complicate the calibration of the model. One would have to find a calibration procedure to appropriately determine parameters for the forward curve and the basis spread, respectively.

Moreover, the model's capability to calibrate to the deterministic multi-curve information embedded in liquid (heavily traded) market instruments is not sufficient, since large volumes of bespoke are traded without collateral agreement [1] and the future credit worthiness of these clients are rather random than deterministic. For these trades, it is useful to have a model with stochastic funding spread. Piterbarg [89] proposed a model which incorporates the stochastic funding spread factor between CSA and non-CSA trades in the model. It should be possible to derive a mathematical relation between a stochastic funding model

[1] The collateral agreements usually regulate highly standardized contracts, but do not cover the non-standardised contracts.

and a stochastic interest rate model, so that the SABR-Hull-White model introduced in Chapter 4 can be applied to compute the option price convexity adjustment induced by the stochastic funding spread. With many new rules and new pricing practices emerging in the industry, there is ample room for new models and corresponding pricing techniques for future PhD students.

Curriculum vitae

Bin Chen was born in Hangzhou, China, on May 29, 1981. After finishing high school and the ensuing Chinese national entrance exam, he succeeded in entering one of the prestigious universities in China, Zhejiang University, studying Thermal Physics and Power Engineering. Shortly after graduation, he decided to pursue graduate studies abroad. In 2005, he moved to the Netherlands and enrolled in the Master of Science programme of the Engineering Policy Analysis faculty at Delft University of Technology. In 2007, he finished his Masters thesis project entitled 'Calibration of the Heston Model with Application in Derivative Pricing and Hedging' in Rabobank International, Utrecht, under the supervision of Cornelis Oosterlee. He continued with the PhD project research in the Modelling Analysis and Computing group 'Scientific Computing and Control Theory' (MAC-2) at CWI - the Dutch National Research Centre for Mathematics and Computer Science, under the supervision of Cornelis Oosterlee. The research project was supported by Rabobank International.

List of publications

- B. Chen and C. Oosterlee and S. van Weeren, Analytical Approximation to Constant Maturity Swap Convexity Correction in a Multi-Factor SABR Model, *Journal of Theoretical and Applied Finance*,**13**(7), 2010, 1019-1046.

- B. Chen and C.W. Oosterlee and Hans van der Weide, A low-bias simulation scheme for the SABR stochastic volatility model, *International Journal of Theoretical and Applied Finance*, **15(2)** 1-37 (2012).

- B. Chen, L. Grzelak, and C.W. Oosterlee, Calibration and Monte Carlo Pricing of the SABR-Hull-White Model for Long-Maturity Equity Derivatives"
Journal of Computational Finance, **15(4)** 79-113 (2012).

- B. Chen, R. van der Werf, S. van Weeren and C.W. Oosterlee, On a Calibration Technique for the Term Structure of Skew Stochastic Volatility Libor Market Model, submitted for publication.

BIBLIOGRAPHY

[1] A. Amin. Calibration, Simulation and Hedging in a Heston Libor Market Model with Stochastic Basis. *SSRN eLibrary*, 2010.

[2] L. Andersen. Yield curve constructuion with tension splines. *Review of Derivatives Research*, 10(3):227 – 267, 2005.

[3] L. Andersen. Efficient simulation of the heston stochastic volatility model. *Journal of Computational Finance*, 11(3), 2008.

[4] L. Andersen and J. Andreasen. Volatility skews and extensions of the libor market model. *Applied Mathematical Finance*, 7:1–32, 2000.

[5] L. Andersen and J. Andreasen. Volatile volatilities. *Risk Magazine*, December, 2002.

[6] L. Andersen and R. Brotherton-Ratcliffe. Extended libor market models with stochastic volatility. *Journal of Computational Finance*, Vol. 9 No. 1, 2005.

[7] L. Andersen and V. Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11:1, 2007.

[8] L. Andersen and V. Piterbarg. *Interest Rate Modeling. Volume I: Foundations and Vanilla Models*. Atlantic Financial Press, 2010.

[9] L. Andersen and V. Piterbarg. *Interest Rate Modeling. Volume II: Term Structure Models*. Atlantic Financial Press, 2010.

[10] L. Andersen and V. Piterbarg. *Interest Rate Modeling. Volume III: Products and Risk Management*. Atlantic Financial Press, 2010.

[11] A. Antonov and T. Misirpashaev. Markovian Projection onto a Displaced Diffusion: Generic Formulas with Applications. *SSRN eLibrary*, 2006.

[12] M. Avellaneda. Minimun-relative-entropy calibration of asset-pricing models. *International Journal of Theoretical and Applied Finance*, Vol. 1, Issue 4:447–472, 1998.

[13] M. Avellaneda, R. Buff, C. Friedman, N. Grandechamp, L. Kruk, and J. Newman. Weighted monte carlo: A new technique for calibrating asset-pricing models. *International Journal of Theoretical and Applied Finance*, 4:91–119, 2000.

[14] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi. Calibrating volatility surfaces via relative-entropy minimization. *Applied Mathematical Finance*, Vol. 4, Issue 1:37 – 64, 1997.

[15] M. Avellaneda and P. Jäckel. *Encyclopedia of Quantitative Finance*. John Wiley and Sons, first edition, 2010.

[16] Eric Benhamou. Swaps: Constant maturity swaps (cms) and constant maturity treasury (cmt) swaps. Swap Strategy Note, FICC, Goldman Sachs International, 2003.

[17] Christopher Beveridge and Mark Joshi. Practical Policy Iteration: Generic Methods for Obtaining Rapid and Tight Bounds for Bermudan Exotic Derivatives Using Monte Carlo Simulation. *SSRN eLibrary*, 2009.

[18] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 1998.

[19] F. Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3:167 – 179, 1976.

[20] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, Vol. 81, no. 3, 1973.

[21] S. Borodin. *Handbook of Brownian Motion*. Birkhäuser, second edition, 2002.

[22] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, first edition, 2004.

[23] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambirdge University Press, first edition, 2004.

[24] A. Brace, D. Gatarek, and M. Muisiela. The market model of interest rate dynamics. *Mathematical Finance*, 7:127 – 154, 1997.

[25] D. Brigo and F. Mercurio. *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*. Springer Finance, second edition, 2007.

[26] M. Broadie and Ö. Kaya. Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, Vol. 54, No. 2, 2006.

[27] P. W. Buchen and M. Kelly. The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis*, Vol. 31:143–159, 1996.

[28] G. Campolieti and R. Makarov. Pricing path-dependent options on state dependent volatility models with a bessel bridge. *International Journal of Theoretical and Applied Finance*, 10:1:51 – 88, 2007.

[29] P. Carr and D. Madan. Nature of option. Courant Institute Working Paper, NYU, Apirl 2000.

[30] B. Chen, L.A. Grzelak, and C.W. Oosterlee. Calibration and monte carlo pricing of the sabr-hull-white model for long-maturity equity derivatives. *Journal of Computational Finance*, 15(4):79–113, 2012.

[31] B. Chen, C. Oosterlee, and S. van Weeren. Analytical approximation to constant maturity swap convexity correction in a multi-factor sabr model. *Journal of Theoretical and Applied Finance*, 13(7):1019–1046, 2010.

[32] B. Chen, C.W. Oosterlee, and Hans van der Weide. A low-bias simulation scheme for the sabr stochastic volatility model. *International Journal of Theoretical and Applied Finance*, 15(2):1–37, 2012.

[33] J.C. Cox. Note on option pricing i: Constant elasticity of variance diffusions. *Stanford University Working Paper*, pages 229 – 263, 1975.

[34] A. D'Aspreont. Interest rate model calibration using semidefinite programming. *Applied Mathematical Finance*, 10(3), 2003.

[35] O. Elerian. A note on the existence of a closed form conditional transition density for the milstein scheme. Nuffield College Working Paper, Oxford University, 1998.

[36] F. Fang and C.W. Oosterlee. A novel option pricing method based on fourier-cosine series expansions. *SIAM J. Scientific Computing*, 31(2):826–848, 2008.

[37] J. Gatheral. *The Volatility Surface: A practitioner's Guide*. Wiley New York, 2006.

[38] H. Geman, N. El Karoui, and J.C. Rochet. Changes of numéraire, changes of probability measures and pricing of options. *Journal of Applied Probability*, 32:443 – 548, 1996.

[39] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004.

[40] D. Goldenberg. A unified method for pricing options on diffusion processes. *Journal of Financial Economics*, 29, 1991.

[41] L.A. Grzelak and C.W. Oosterlee. On the heston model with stochastic interest rates. *SIAM Journal of Financial Mathematics*, 2:255–286, 2011.

[42] I. Gyöngy. Mimicking the one-dimensional marginal distribution of process having an itô differential. *Probability Theory and Related Fields*, 71:501–516, 1986.

[43] A. Van Haastrecht and A.A. Pelsser. Efficient, Almost Exact Simulation of the Heston Stochastic Volatility Model. *International Journal of Theoretical and Applied Finance*, 13:1:1–43, 2010.

[44] P. Hagan. Convexity conundrums: Pricing cms swaps, caps and floors. Wroking Paper, Gorila Science, 2000.

[45] P. Hagan and A. Lesniewski. Libor market model with sabr style stochastic volatility. Working Paper, 2008.

[46] P.S. Hagan, D. Kumar, A.S. Lesniewski, and D.E. Woodward. Managing smile risk. *Wilmott Magazine*, 3:84–108, 2002.

[47] M. Henrard. Cms swaps in separable one-factor gaussian llm and hjm model. MPRA Paper from University Library of Munich, Germany, May 2007.

[48] S. L. Heston. A closed form solution for options with stochastic volatility. *Review of financial studies*, Vol. 6:327–343, 1993.

[49] J. Hull and A. White. Using hull-white interest rate trees. *Journal of Derivatives*, 3(3):26–36, 1996.

[50] J. Hull and A. White. Forward rate volatilities, swap rate volatilities and the implementation of the libor market model. *Journal of Fixed Income*, Volume 10, Issue 2:46–62, 2000.

[51] C. Hunt. *The Euromoney Derivatives and Risk Management Handbook*. BNP Paribas, 2005.

[52] P.J. Hunt and J.E. Kennedy. *Financial Derivatives in Theory and Practice*. WILEY, 2000.

[53] Othmane Islah. Solving SABR in Exact Form and Unifying it with LIBOR Market Model. *SSRN eLibrary*, 2009.

[54] K. Itô. Multiple wiener integral. *The Mathematical Society of Japan*, Vol. 3:157–169, 1951.

[55] P. Jäckel and R. Rebonato. Linking caplet and swaption volatilities in a bgm/j framework: Approximate solutions. *Journal of Computational Finance*, Vol. 6, No. 4, 2003.

[56] P. Jaeckel. Mind the cap. *Wilmott*, pages 54–68, 2003.

[57] F. Jamshidian. Libor and swap market models and measures. *Finance & Stochastic*, Vol. 1:293–330, 1997.

[58] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions Vol. 2*. Wiley Interscience, first edition, 1995.

[59] N.L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions Vol. 2*. Boston: houghton Miffin Company, first edition, 1970.

[60] S. Johnson and B. Nonas. Arbitrage-free construction of the swaption cube. *Wilmott*, 1(3):137–143, 2009.

[61] M.S. Joshi and R. Rebonato. A stochastic volatility displaced-diffusion extension for the libor market model. *Quantitative Finance*, 3(6), 2003.

[62] C. Kahl. *Modelling and simulation of stochastic volatility in finance*. Dissertation.com, Boca Raton, first edition, 2007.

[63] C. Kahl and P.Jäckel. Fast strong approximation monte-carlo schemes for stochastic volatility models. *Journal of Quantitative Finance*, Vol. 6, No. 6:513–536, 2006.

[64] A. Kawai. A new approximate swaption formula in the libor market model: An asymptotic expansion approach. *Applied Mathematical Finance*, Vol. 10, 2003.

[65] A. Kawai and P. Jackël. An asymptotic fx option formula in the cross currency libor market model. *Wilmott*, 2007.

[66] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992, 1995, 1999.

[67] N. Kunitomo and A. Takahashi. The asymptotic expansion approach to the valuation of interest rate contingent claims. *Mathematical Finance*, Vol. 11, 2001.

[68] N. Kunitomo and A. Takahashi. On validity of the asymptotic expansion approach in contingent claim analysis. *Annals of Applied Probabilities*, Vol. 13, No. 3, 2003.

[69] H. Labordere. Combining the sabr and lmm models. *Risk*, October, 2007.

[70] R. Lee and D. Wang. Displaced lognormal volatility skews: analysis and applications to stochastic volatility simulations. *Annals of Finance*, pages 1–23, 2009. 10.1007/s10436-009-0145-7.

[71] A. Lesniewski. Notes on the cev model. NYU working paper, 2009.

[72] A.L. Lewis. The mixing approach to stochastic volatility and jump models. *Wilmott*, March, 2002.

[73] M. Lewis. *The Big Short: Inside the Doomsday Machine*. W.W. Norton & Company, 2010.

[74] F.A. Longstaff, P. Santa-Clara, and E.S. Schwartz. Throwing away a billion dollars: the cost of suboptimal exercise strategies in the swaptions market. *Journal of Financial Economics*, 62:39–66(28), October 2001.

[75] R. Lord, R. Koekkoek, and D. van Dijk. A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10:2:177–194, 2010.

[76] D. Marris. Financial option pricing and skewed volatility. MPhil Thesis, Statistical Laboratory, University Cambridge, 1999.

[77] R. Mendoza, P. Carr, and V. Linetsky. Time changed markov processes in unified credit-equity modelling. *Mathematical Finance*, 20(4), 2010.

[78] F. Mercurio. Modern libor market models: Using different curves for projecting rates and for discounting. *International Journal of Theoretical and Applied Finance*, 13(1):113–137, 2010.

[79] F. Mercurio and A. Pallavicini. Smiling at convexity: Bridging swaption skews and cms adjustments. Available at SSRN: http://ssrn.com/abstract=892287, 2006.

[80] G.N. Milstein. A method of second order accuracy integration of stochastic differential equations. *SIAM: Theory of Probability and Its Applications*, 23:396–401, 1978.

[81] M. Morini and F. Mercurio. No-Arbitrage Dynamics for a Tractable SABR Term Structure Libor Model. *SSRN eLibrary*, 2007.

[82] S. Nawalkha. The LIBOR/SABR Market Models: A Critical Review. *SSRN eLibrary*, 2009.

[83] Sanjay K. Nawalkha and Riccardo Rebonato. What Interest Rate Models to Use? Buy Side Versus Sell Side. *SSRN eLibrary*, 2011.

[84] I.E. Nesterov and M. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimizatin*, 8:324–364, 1998.

[85] A. Pelsser. Mathematical foundation of convexity correction. *Quantitative Finance*, Vol. 3, Issue 1:59–65, 2003.

[86] V. Piterbarg. Stochastic volatility model with time dependent skew. *Applied Mathematical Finance*, 12(2):147–185, 2005.

[87] V. Piterbarg. Markovian Projection Method for Volatility Calibration. *SSRN eLibrary*, 2006.

[88] V. Piterbarg. Smiling hybrids. *Risk*, 19, 2006.

[89] V. Piterbarg. Funding beyond discounting: collateral agreements and derivatives pricing. *Risk*, February:97–102, 2010.

[90] R. Rebonato. *Interest-Rate Option Models : Understanding, Analyzing and Using Models for Exotic Interest-Rate Options*. Wiley, New York, second edition, 1998.

[91] R. Rebonato. Calibrating the bgm model. *Risk*, March:88–94, 1999.

[92] R. Rebonato. *Modern Pricing of Interest-Rate Derivatives*. Princeton University Press, first edition, 2002.

[93] R. Rebonato. A time-homogeneous, sabr-consistent extension of the lmm. QUARC Working Paper, Royal Bank of Scotland, 2007.

[94] R. Rebonato. *The SABR/LIBOR Market Model: Pricing, Calibration and Hedging for Complex Interest-Rate Derivatives*. John Wiley and Sons, first edition, 2009.

[95] R. Rebonato and R. White. Linking caplet and swaption prices in the lmm-sabr model. QUARC Working Paper, Royal Bank of Scotland, 2007.

[96] M. Rubinstein. Displaced diffusion option pricing. *Journal of Finance*, 38:213–217, 1983.

[97] M. Sankaran. Approximations to the non-central chi-square distribution. *Biometrika*, 50:199 – 204, 1963.

[98] K. Schmitz-Abe and W.T. Shaw. Measure order of convergence without an exact solution, euler vs. milstein scheme. *International Journal of Pure and Applied Mathematics*, Vol. 24, No. 3, 2005.

[99] M. Scholes. Market-based mechanisms to reduce systemic risk. *The Road Ahead for the Fed*, pages 103 – 121, 2009.

[100] M. Schroder. Computing the constant elasticity of variance option pricing formula. *Journal of Finance*, Vol. 1, No. 44, 1989.

[101] Will Spinney. Managing interest rate risk. Working Paper QFINANCE, www.qfinance.com.

[102] J.F. Sturm. Using sedumi 1.0x, a matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653, 1999.

[103] N.N. Taleb. *Taleb on Risk Dynamic Hedging*. John Wiley & Sons, 1997.

[104] N.N. Taleb. *The Black Swan: The Impact of the Highly Improbable*. Allen Lane, 2007.

[105] S. van Weeren and E. van Dijk. Convexity correction: a general approach. Rabobank Derivative Research & Validation Working Paper, 1999.

[106] J.R. Varma. An introductory note on two curve discounting. Indian Institute of Management Working paper, 2010.

[107] G. West. Calibration of the sabr model in illiquid markets. *Applied Mathematical Finance*, 12(4):371–385, 2005.

[108] G.A. Willard. Calculating prices and sensitvities for path-independent derivative securities in multifactor models. *Journal of Derivatives*, 5, 1997.

[109] L. Yinqiu and N. Salih. Convexity adjustments and forward libor model: Case of constant maturity swaps. National Centre of Competence in Research Financial Valuation and Risk Management. Working Paper No. 115, 2003.

[110] N. Yoshida. Asymptotic expansion for small diffusion via the theory of malliavin-watanabe. *Probability Theory and Related Fields*, Vol. 92, 1992.

[111] L. Yuan and J.D. Kalbfleisch. On the bessel distribution and related problems. *Annals of the Institute of Statistical Mathematics*, 53(3):438 – 477, 2000.