BERGISCHE
UNIVERSITÄT
WUPPERTAL

Fakultät für Mathematik
und Naturwissenschaften

Effective Methods for Generalized Stochastic Volatility Models

Mike Oliver Felpel

May 11, 2022

P.h.D. Thesis in Mathematics

Advisors: Dr. Jörg Kienitz and Dr. Thomas A. McWalter

Declaration

I hereby declare that this thesis has been composed by myself, that the work presented herein is the result of my own original research except where explicitly stated otherwise. All external sources of information have been acknowledged as references and I clearly marked and separately listed all of the literature and all of the other sources which I employed when producing this academic work, either literally or in content. I also declare that this thesis was never previously submitted for an academic degree or professional qualification.

Cologne, May 11, 2022, ...

Mike Oliver Felpel

Acknowledgement

First and foremost, I want to thank my supervisors Dr. Jörg Kienitz and Dr. Thomas A. McWalter for their guidance during my PhD studies. Without their valuable advice, helpful discussions and dedicated involvement in this project, this work would not exist. I would like to thank you very much for teaching me how independent mathematical research can be performed.

Next, I want to thank Dr. Axel Schmidt and Dr. Stefan-M. Heinemann for providing the opportunity for an extra-occupational doctorate at Ampega Asset Management GmbH; for supporting my PhD studies and being concerned to provide an adequate balance between workload and research time.

Most importantly, none of this could have happened without my family. I want to thank my wife for her never ending love and support, encouraging and motivating me to overcome all challenges in life together. Thank you for always being by my side. I want to thank my mother, my sisters and my grandmother for always being there for me and fully supporting all of my decisions. Without you, I would not have made it this far. I want to thank my nieces and my nephew for all the time of joyous laughter spent together, allowing me to relax and gather new energy. I am forever grateful to have you all by my side.

Finally, I also want to thank the rest of my family, my friends and all the people who supported me in writing this thesis. May it be in reading parts of this thesis or by simply spending an enjoyable evening together. Thank you.

Contents

1 Introduction

Randomness and future uncertainty can be observed in a multitude of different problems and situations in all areas of science and daily life; be it theoretical and abstract areas such as quantum mechanics, random genetic mutations observed in biology or regarding the simple question of how the weather will be tomorrow. One area that is particularly dependent on future uncertainty is the behaviour of financial markets. Consider for example the prices of a single stock; it can be observed that these prices are heavily fluctuating and it is not clear how the price may look like tomorrow. Will it continue to rise moderately or will there be a sudden drop? This uncertainty induces a risk for the participants of these markets. To be able to assess and manage these risks, it is vital to make these uncertainties tangible. The means to achieve this is provided using stochastic modelling.

The foundations of stochastic modelling began with observations regarding particle movements in pollen provided in Brown [1828]. These observations gave rise to the construction of Brownian motions, which became one of the fundamental means to describe random movements from a mathematical perspective, see Einstein [1905], Wiener [1924]. This evolved further into stochastic differential equations becoming the standard tool to model dynamics governed by randomness, Itô [1944]. Not surprisingly, this approach also found its way into the world of finance, Bachelier [1900], Samuelson [1965], Black and Scholes [1973], where the Nobel prize winning work of Black and Scholes [1973] used geometric Brownian motions to model the price evolution of a stock. Following this approach a multitude of different techniques to model financial quantities and markets were developed, see Andersen and Piterbarg [2010a,b,c] for a non-exhaustive list.

All these models were created with the purpose of capturing certain observable patterns. These may be jumps in the stock prices or observable randomness in the intensity of the fluctuations. However, the more properties that are explicitly modelled, the higher the complexity of the model and from a practical perspective it can become infeasible to work with a model. Thus, stochastic volatility models became one prominent choice to model financial quantities, such as stock

prices or interest rates. For this class of models the desired quantity as well as its volatility are governed by stochastic dynamics where both can be correlated with each other. This allows one to capture the uncertainty of the volatility observable in the markets, while still providing the mathematically simplest modelling approach available to achieve this. Nevertheless, even for these models the underlying stochastic differential equation can become quite complex and, again, it may be infeasible to compute anything meaningful in an adequate time span. We want to highlight that the computation time and complexity of a model are often a crucial factor in the process of model selection. While it may be possible to use Monte Carlo methods to simulate the modelled quantities, from a practical perspective it may be problematic to simulate e.g. one million paths in order to evaluate the price of a simple product. Instead, a simple valuation formula would be preferable.

The focus of this work is to provide a computational framework which is applicable to a large class of generalized stochastic volatility models (GSVM). This allows one to choose the model with regard to suitability and not due to the ease of implementation. The underlying idea of our approach is inspired by the work of Hagan et al. [2002, 2014, 2015, 2016] and follows the principle of using singular perturbation techniques to study the probability distribution induced by the GSVM. This is achieved by accurately approximating the marginal probability density function using a one-dimensional partial differential equation, called the effective PDE. This approximation can be used either to directly determine the induced probability distribution, or to project the original GSVM onto another, simpler, model where various computational benefits, such as explicit formulas, may exist. Here the effective PDE is constructed in such a way, that the computed marginal probability density function stays positive and therefore ensures that the computed quantities remain arbitrage-free. Intuitively, this means that it is not possible to earn money in the model without taking any risk in return and in turn avoid the possibility of exploitations.

This thesis is based on the articles Felpel et al. [2021, 2022a,b], which are my work in collaboration with Dr. Jörg Kienitz and Dr. Thomas A. McWalter during their joint supervision of my PhD studies. The three articles build the heart of this thesis and show a consecutive development of techniques.

Starting in Felpel et al. [2021] we begin to establish the foundation for our generalized arbitrage-free computational framework. We focus on parametric GSVM used to model an asset for a certain fixed time horizon, e.g. the stock price observed five years ahead in the future. This is done in a risk-neutral environment, which guarantees that the considered asset has no drift. Our chosen model framework unifies a large class of heterogeneous models, allowing to easily exchange the underlying

model in a practical application. We derive the effective PDE for a class of GSVM and derive closed form approximation formulas for the implied volatility. The implied volatility can intuitively be seen as a quotation standard for a European call option on the asset, which in turn is a certain type of financial product frequently traded at the markets. This methodology also allows us to consider new types of models and we introduce new models such as the mean-reverting ZABR model or the free ZABR model. Furthermore, the availability of explicit pricing formulas allows a very fast and simple implementation, extending one of the most valued properties only available to certain selected models to the whole class of GSVM.

Continuing to further develop our generalized model framework, we slightly shift our focus and study the pricing properties of our framework when considering more involved financial products. Having a practical application in mind, we aim to evaluate a whole portfolio of assets and therefore concentrate on prominent financial products depending on multiple quantities, called baskets or spreads. This is the focus of Felpel et al. [2022a] and we demonstrate how it is possible to price basket options using explicit closed form approximation formulas where each individual quantity is modelled using a GSVM. This is achieved by introducing a new technique called Effective Markovian Projection (EMP), which aims to project one GSVM onto another using the effective PDE in the projection procedure. The EMP also allows us to establish another projection technique by deriving explicit approximation formulas for the first moments of the induced distribution and matching those with a certain class of distributions allowing for explicit approximation formulas of the underlying probability density function. This extends our generalized model framework into a multidimensional set-up and provides additional explicit pricing formulas for a practical application.

Finally, we extend the methodology even further and consider GSVM containing a non-parametric component and introducing time dependencies in most of the coefficients. This is done in the work of Felpel et al. [2022b]. This set-up includes stochastic local volatility (SLV) models, which are characterized through an additional component, called leverage function, and allow the modelling of a complete surface. The leverage function aims to perfectly fit the model to observed market data. This is important since using simpler products to replicate a more involved financial derivative can lead to mispricing if the market data for the underlying products is not perfectly matched. The process of determining the correct leverage function is called calibration. We focus on the calibration problem of the leverage function and provide a fast and explicit calibration algorithm to match the leverage function to observable market data. This allows to transform all previously studied stochastic volatility models onto a SLV variant, making it possible to choose the dynamics in accordance to the desired model properties while still

achieving a perfect fit using the leverage function.

All three articles include explicit examples considering specific models as well as numerical results to demonstrate the applicability. The rest of this thesis is structured as follows:

In Chapter 2 we start by establishing the foundations necessary for this work. This spans the preliminary results governing stochastic differential equations in Section 2.1, an introduction to mathematical finance in Section 2.2 as well as an overview of prominent models falling into the category of GSVM in Section 2.3. We then dive deeper into the techniques necessary for our approximation methodology by considering singular perturbation techniques in Section 2.4. We finish the foundation chapter by considering numerical aspects providing the fundamentals for an actual practical implementation of our techniques in Section 2.5. In particular, we also introduce the actual numerical scheme in Section 2.5.3, which is used for our numerical experiments.

Having established the necessary foundation, we begin our research by presenting the results of Felpel et al. [2021] in Chapter 3. In Section 3.2 we introduce the methodology and derive the effective PDE for the general framework. The whole derivation— including all details—can be found in Appendix 3.5. In Subsection 3.2.2 we consider the approximating formula and then provide more detail on concrete models in Subsection 3.2.3. In Subsection 3.2.3.1 we use the method to analyse the smile dynamics of the ZABR model, while in Subsection 3.2.3.2 we investigate the extended ZABR models. The approach taken to demonstrate numerical examples in Section 3.3 is based on one-dimensional PDE solvers. We use numerical methods to compare the resulting stochastic volatility modelling approaches in Subsection 3.3.1 and the calibration behaviour in Subsection 3.3.2. Finally, Subsection 3.3.3 compares our approach to a classical Monte Carlo approach of Lord and Farebrother [2014]. Section 3.4 concludes with a summary of the chapter.

Following this path, Chapter 4 continues by presenting the results of Felpel et al. [2022a]. Section 4.2 provides the main theoretical results. After briefly summarizing the theoretical background of Markovian projection Gyöngy [1986], we introduce Effective Markovian Projection, which is based on effective probability theory Felpel et al. [2021], Hagan et al. [2014]. The numerical methods used for computing the projection are also introduced, in particular ATM-Matching, Minimal Point Matching and N-Point Matching. A brief analysis of the approximation quality is given. We also describe an Effective Markovian Projection based on moment matching for the case of the normal SABR model. Moment matching is applied to both NIG distributions and the family of Johnson distributions.

Section 4.3 focuses on examples from interest rate derivatives. Fixing notation and definitions, we consider constant maturity swap (CMS) spread and mid-curve options as our main examples. With the underlying definitions at hand we consider a ZABR-type model as our base model and the normal SABR model as our reference model in Section 4.4, where we derive the corresponding projection and basket dynamics. Here we also specifically address numéraire issues. The resulting model amalgamates all the features of the base model and parametrizes the implied volatility surface in terms of a normal SABR model for each maturity, enabling the pricing of CMS spread and mid-curve options. Section 4.5 provides numerical illustrations of all the methods considered in the previous sections. In particular, we demonstrate the additional flexibility in modelling and calibrating underlying rates in a multi-rate framework when considering base models of the ZABR type. We also illustrate the effects of changing model parameters. Section 4.6 concludes by summarizing the results presented in this chapter.

Building on these results, we arrive at Chapter 5 presenting the results of Felpel et al. [2022b]. Section 5.2 lays the theoretical groundings by extending results from Felpel et al. [2021]. In particular, we prove that the results hold true for separable time-dependent diffusion coefficients. Given the theoretical grounding, Section 5.3 provides the details of the chosen approach for calibrating the general local stochastic volatility model. With the results from Sections 5.2 and 5.3 we propose the implementation based on solving a one-dimensional PDE in Section 5.4, state the discretization scheme and shortly place our results into the known model setting in 5.4.3. In particular, we consider the approaches Bang [2019] and Muguruza [2019]. We apply our methods to several models described in Section 5.5 and give numerical examples that show the performance of our results in Section 5.6. We conclude by summarizing our results of this chapter until this point in 5.7. Extending the scope considered in Felpel et al. [2022b], we also provide additional analysis related to Felpel et al. [2022b]. These results are given in Section 5.9.

In Chapter 6 we conclude by summarizing all results and outlining prospects for future research.

2 Foundations

We start our work by establishing and recalling the underlying foundations necessary for our research. This chapter does not aim to establish new results and instead provides a condensed summary to introduce the relevant areas of research we are considering. This includes topics related to analysis, stochastics, numerics as well as finance. Providing an extensive summary of all these topics would be far above the scope of this work. Therefore, even though we establish the notation and recall the definitions and results for many preliminary structures, a general familiarity of the areas presented is assumed. In particular, in this chapter we are also exclusively interested in the definitions and results and will not consider the proofs for all of these theorems. For the proofs as well as further details, we instead refer to the literature discussed in the individual sections.

2.1 Stochastic Differential Equations

In this section, we outline the general underlying mathematical concepts to allow the study of stochastic differential equations (SDEs) and hence give a meaning to expressions of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t.$$

The goal of this section is to introduce our mathematical set-up, clarify the notation and present the main results and techniques to which we resort to in our research starting from Chapter 3 onwards. We establish results from stochastic analysis and show important connections between SDEs and partial differential equations (PDEs) using Kolmogorov equations. Again, we want to highlight that this section provides a collection of standard literature related to these topics and the results presented are a condensed summary of the work presented in Durrett [2010], Andersen and Piterbarg [2010a], Øksendal [2003], Karatzas and Shreve [1991]. For further details we refer to the respective literature.

2.1.1 Probability Theory and Brownian Motions

We begin by considering a general filtered probability space of the form

$$\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\leq T}, \mathbb{P}\right) \tag{2.1.1}$$

with a finite time horizon T. Here Ω denotes the sampling space, $\mathcal{F}$ the σ-algebra, $\{\mathcal{F}_t\}_{t\leq T}$ the filtration and $\mathbb{P}$ the probability measure. Since our focus is on financial applications, a finite time horizon is sufficient for our considerations. We do assume that the concepts of probability spaces and measurability are familiar to the reader and do not recall all underlying background. For an overview of these concepts we refer to standard literature on probability theory such as Durrett [2010] or Øksendal [2003, Chapter 2.1], where an appealing summary of these preliminaries is presented. For a more thorough study which different concepts of measurability are relevant in the area of stochastic analysis, we also refer to Karatzas and Shreve [1991]. In particular, a proper definition of the probability space can be found in e.g. Øksendal [2003, Definition 2.1.1].

The probability space of (2.1.1) provides a very general framework for our applications. Again, since we are interested in applications to finance in the spirit of Andersen and Piterbarg [2010a], we essentially only want to consider a p-dimensional vector-valued process X_t. This allows us to specify the sampling space as $\mathbb{R}^p$, consider a corresponding σ-algebra on $\mathbb{R}^p$, e.g. the Borell σ-algebra $\mathcal{B}(\mathbb{R}^p)$, see Durrett [2010], and use a filtration generated by a d-dimensional Brownian motion W_t of the form

$$\mathcal{F}_t = \sigma\left\{W_u | u \leq t\right\},$$

augmented to be complete and right-continuous to satisfy the usual conditions. The construction of this filtration can be found in e.g. Karatzas and Shreve [1991, Chapter 2.7]. Due to its importance in all applications regarding stochastics, we do, however, want to recall the definition of the Brownian motion W_t. We consider the definition given by Durrett [2010, Chapter 7.1].

Definition 2.1.1. A one-dimensional Brownian motion is a real-valued process W_t that has the following properties:

i) **Independent increments:** If $t_0 < t_1 < \ldots < t_n$, then the increments $W_{t_0}, W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent.

ii) **Normal distributed increments:** If $s, t \geq 0$ then

$$\mathbb{P}\big(W_{s+t} - W_s \in A\big) = \int_A \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}\, dx.$$

iii) **Almost sure continuous paths:** With probability one, $t \to W_t$ is continuous.

We define a d-dimensional Brownian motion as a vector-valued process where each coordinate process is given by an independent one-dimensional Brownian motion.

The definition provided by Definition 2.1.1 is only one possible way to define a Brownian motion, also often denoted as a Wiener process. There are many other ways, e.g. using the Levy characterization Karatzas and Shreve [1991, Theorem 3.3.16] or through the Wiener measure Karatzas and Shreve [1991, Chapter 2.4], to define a Brownian motion and each one has its own advantages to determine whether a process is a Brownian motion or not. For further details about Brownian motions regarding important properties or the construction of corresponding processes, we refer to Karatzas and Shreve [1991], Øksendal [2003].

This set-up provides us with the underlying probabilistic framework on which we will be working throughout this work. In the next section, we continue to formalize the p-dimensional vector-valued processes X_t which we want to study and thereafter continue to establish the important connections towards other areas such as PDEs necessary in our work.

2.1.2 Itô Processes

To properly characterize the desired stochastic processes X_t, we use a characterization in terms of an d-dimensional Itô process in accordance to Andersen and Piterbarg [2010a, Chapter 1.6].

Definition 2.1.2. We define the d-dimensional Itô process $X : [0,T] \times \Omega \to \mathbb{R}^d$, denoted as X_t, as the solution to the SDE given by

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t \tag{2.1.2}$$

with initial condition X_0. Here the functions $\mu : [0,T] \times \mathbb{R}^p \to \mathbb{R}^p$ and $\sigma : [0,T] \times \mathbb{R}^p \to \mathbb{R}^{p \times d}$ are time- and state-dependent, adapted to the filtration $\{\mathcal{F}_t\}_{t \leq T}$ and satisfy the condition

$$\int_0^T |\mu(s,x)| ds < \infty \text{ a.s.},$$
$$\int_0^T |\sigma(s,x)|^2 ds < \infty \text{ a.s.}$$

The differential notation provides a very useful abbreviation which is commonly used to express SDEs. To make the expression mathematically correct, it should

however always be understood in terms of stochastic integrals and abbreviates the expression

$$X_t = X_0 + \int_0^t \mu(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s.$$

To provide some intuition about the stochastic integral represented in the last term, it can be formally seen as the limiting process of a discrete sum over a partition with step size $h = \frac{T}{n}$ as

$$\int_0^t \sigma(s, X_s)\, dW_s = \lim_{n\to\infty} \sum_{i=1}^{n} \sigma(h(i-1), X_{h(i-1)}) \big[W_{ih} - W_{h(i-1)}\big].$$

Note that we do not use this expression as the rigorous mathematical definition, since the construction of these integrals is more involved. This expression can only be used as a definition for a certain class of elementary functions and the general integral is constructed based on an approximation of the original integrand in terms of elementary functions. Hence, this formula should only provide some intuition and for the rigorous construction of stochastic integrals we refer to Karatzas and Shreve [1991], Øksendal [2003].

Given that the Itô process is defined as a solution to the SDE (2.1.2) we need to specify some regularity conditions, which guarantee the existence and uniqueness of such solutions. Such results are provided in e.g. Karatzas and Shreve [1991, Theorem 5.2.9], Øksendal [2003, Theorem 5.2.1] or Andersen and Piterbarg [2010a, Theorem 1.6.1]. To provide a framework consistent with Definition 2.1.2, we refer to the formulation of Andersen and Piterbarg [2010a, Theorem 1.6.1].

Theorem 2.1.3. *Assume that there exists a constant $K > 0$ such that for all $t \in [0, T]$ and all $x, y \in \mathbb{R}^p$ the following conditions hold:*

i) ***Lipschitz condition:***

$$|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq K|x - y|,$$

ii) ***Growth condition:***

$$|\mu(t, x)|^2 + |\sigma(t, x)|^2 \leq K^2(1 + |x|^2).$$

Then there exists a unique strong solution to (2.1.2).

This existence and uniqueness result provides us with a very strong tool when considering SDEs and can establish the mathematical foundation of our considerations. We therefore wish to point out some important remarks related to Theorem 2.1.3.

i) **Strong and weak solutions:** When considering solutions to SDEs one usually distinguishes between a solution is a strong solution or a weak solution. When considering strong solutions, the Brownian motion is given in advance and the resulting solution X_t has to be adapted to the given filtration. Weak solutions on the other hand are constructed while also allowing to construct the corresponding Brownian motion, satisfying the same SDEs but possibly with regard to a different filtration. For more details, we refer to e.g. Øksendal [2003, Chapter 5.3]. As explained in e.g. Andersen and Piterbarg [2010a, Chapter 1.6], in a financial application often only the distribution of the process is of relevance and weak solutions are sufficient. To formalize our set-up, we are working with strong solutions throughout this thesis and, following the approach of Andersen and Piterbarg [2010a], do not explicitly refer to the type of the solution.

ii) **Non-Lipschitz coefficients:** Given how strong the existence and uniqueness results of Theorem 2.1.3 are, the theorem also imposes some considerable restrictions on the coefficients. For certain choices of coefficients, other means may be found to deduce the existence results, see e.g. Suresh Kumar [2013], Chuni [2020]. In general, it may however be possible that we consider a financial model for which no explicit existence or uniqueness result was proven. Paraphrasing Taylor [2010], one then operates under the typical, but possibly dangerous, assumption that a unique solution exists and performs the computation anyway, treating the results with a healthy dose of scepticism.

iii) **Itô diffusions:** In the case where the coefficients of (2.1.2) only depend on the state variable X_t and not on the time t, the process is called an Itô diffusion instead. Itô diffusions can simplyfy the considerations in some aspects and admit a multitude of nice properties as presented in e.g. Øksendal [2003, Chapter 7]. In our set-up, this can not always be guaranteed and we consider general Itô processes instead. However, we do use an Itô diffusion for illustration in the next section.

In summary, we assume that we are working with an Itô process X_t given as the unique solution to (2.1.2). Working in the setting of Itô processes gives us access to the famous 'Itô's Lemma', allowing us to study functionals of Itô processes. We refer to the version provided in Andersen and Piterbarg [2010a, Theorem 1.1.5].

Theorem 2.1.4. *Let $f(t,x), x = (x_1, \dots, x_p)^\top$, denote a continuous function $f : [0,T] \times \mathbb{R}^p \to \mathbb{R}$, with continuous partial derivatives $\partial_t f, \partial_x f, \partial_{x_i x_j} f$. Let X_t be given by the Itô process* (2.1.2) *and define a scalar process $Y_t = f(t, X_t)$. Then Y_t*

is an Itô process with stochastic differential

$$dY_t = \Big[\partial_t f(t, X_t) + \partial_x f(t, X_t)\mu(t, X_t)\Big]dt + \partial_x f(t, X_t)\sigma(t, X_t)dW_t + \frac{1}{2}\sum_{1\leq i,j\leq p} \partial_{x_i x_j} f(t, X_t)(\sigma(t, X_t)\sigma(t, X_t)^\top)_{i,j}dt.$$

2.1.3 Kolmogorov Equations

Next, we establish the connection between our Itô processes X_t from (2.1.2) and non-stochastic differential equations. For this we consider Kolmogorov equations. Before introducing the equations, we first construct a differential operator which aims to capture all relevant information induced by X_t. We use the second-order differential operator $\mathcal{A}_t$ in accordance with Karatzas and Shreve [1991, Equation 5.4.1].

Definition 2.1.5. For a function $f \in \mathcal{C}^2([0,\infty), \mathbb{R}^p)$, i.e. a twice differentiable function $f : [0,\infty) \to \mathbb{R}^p$ with continuous derivatives, we define the second-order differential operator $\mathcal{A}_t$, denoted as the generator of X_t, as

$$\big(\mathcal{A}_t f\big)(t,x) = \sum_{i=1}^{p} \mu_i(t,x)\partial_{x_i} f(t,x) + \frac{1}{2}\sum_{i,j=1}^{p} \Sigma_{i,j}(t,x)\partial_{x_i x_j} f(t,x)$$

with

$$\Sigma_{i,j}(t,x) = \big(\sigma(t,x)\sigma(t,x)^\top\big)_{i,j}.$$

To provide some explanation as to why we denote this operator as the generator, let us consider the simpler case of Itô diffusions instead of Itô processes, see Section 2.1.2. As depicted in Øksendal [2003, Theorem 7.3.3], in this case Definition 2.1.5 corresponds to the infinitesimal generator of the Itô diffusion given by Øksendal [2003, Definition 7.3.1]:

Definition 2.1.6. Let X_t be an Itô diffusion in $\mathbb{R}^p$. The infinitesimal generator $\mathcal{A}$ of X_t is defined as

$$\mathcal{A}f(x) = \lim_{t\downarrow 0} \frac{1}{t}\bigg(\mathbb{E}^{\mathbb{P}}\Big[f(X_t)|X_0 = x\Big] - f(x)\bigg).$$

The set of functions f for which the limit exists for all $x \in \mathbb{R}^p$ is the domain of the generator.

The infinitesimal generator, defined as in Definition 2.1.6 can be extended to a very general framework regarding Markov semigroups and the equality between Definition 2.1.5 and Definition 2.1.6 is not true for the general case. This can be seen when taking for example a Poisson process as demonstrated in John and Wu [2021]. This framework yields an interesting connection between stochastic processes and differential geometry and for more information on this topic we refer to e.g. Kac [1966], Bourgade and Croissant [2005], Henry-Labordère [2005]. When studying infinitesimal generators, it is important to always consider the generator together with its domain. For our purpose, however, Definition 2.1.5 is sufficient and we do not delve deeper into the technical details regarding generators. Instead, we state the following result of Andersen and Piterbarg [2010a, Theorem 1.8.1] based on Karatzas and Shreve [1991, Theorem 5.7.6].

Theorem 2.1.7. *Let the process X_t be given by the Itô process* (2.1.2) *where the coefficients $\mu(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are continuous in x and satisfy the Lipschitz and growth conditions of Theorem 2.1.3. Consider a continuous function $g(\cdot)$ that is either non-negative or satisfies a polynomial growth condition, meaning that for some positive constants K and q and all $x \in \mathbb{R}^p$ we have*

$$g(x) \leq K(1 + |x|^q).$$

If $u(t,x)$ solves the Kolmogorov backward equation with terminal condition $g(\cdot)$ given by

$$\begin{cases} \partial_t u(t,x) + \mathcal{A}_t u(t,x) = 0 \\ u(T,x) = g(x) \end{cases} \tag{2.1.3}$$

and in addition, $u(t,x)$ satisfies a polynomial growth condition in x, then for $t \in [0,T]$ we have

$$u(t,x) = \mathbb{E}^{\mathbb{P}}\Big[g(X_T)|X_t = x\Big].$$

This result provides a very powerful connection between the stochastic point of view expressed through the conditional expectation and the area of differential equations provided by the Kolmogorov backward equation. Using for example a terminal condition of the form

$$g(x) = e^{\mathbb{i}y^\top x}$$

with $y \in \mathbb{R}^p$, this allows to deduce the characteristic function conditioned on x by solving the corresponding Kolmogorov backward equation.

To solve Equation (2.1.3), let us for the moment neglect the boundary condition and only consider the differential equation. Following Karatzas and Shreve [1991,

Definition 5.7.9] we define the fundamental solution, also denoted as the Green's function, as follows.

Definition 2.1.8. A fundamental solution to the the second-order partial differential equation

$$\partial_t u(t,x) + \mathcal{A}_t u(t,x) = 0$$

is a non negative function $p(t,x;s,y)$ for $0 \leq t \leq s \leq T$ and $x, y \in \mathbb{R}^p$ such that for every $f \in \mathcal{C}_0(\mathbb{R}^p)$, i.e. continuous functions vanishing at infinity, and $s \in (0,T]$ the function

$$u(t,x) = \int_{\mathbb{R}^p} f(y) p(t,x;s,y)\, dy$$

is bounded, of class $\mathcal{C}^{1,2}$ and satisfies for all $x \in \mathbb{R}^p$ the condition

$$\lim_{t \to s} u(t,x) = f(x).$$

In particular, when comparing Definiton 2.1.8 with the results of Theorem 2.1.7, the fundamental solution can also be interpreted as the transition density for the process X_t defined through the equation

$$\mathbb{P}\big(X_s \in A | X_t = x\big) = \int_A p(t,x;s,y)\, dy$$

for each $A \in \mathcal{B}(\mathbb{R}^p)$. To denote this connection, we also use the notation

$$p(t,x;s,y)\, dy = \mathbb{P}\big(X_s \in [y, y+dy] | X_t = x\big).$$

As presented in Karatzas and Shreve [1991], when imposing a suitable set of regularity conditions on the coefficients of the SDE, the existence and regularity of a fundamental solution can be guaranteed. Fixing the variables (s,y), the function $p(t,x) = p(t,x;s,y)$ then satisfies the Kolmogorov backward equation in the backward variables (t,x). These regularity condition have to do with boundedness, Hölder-continuity and ellipticity and we refer to Karatzas and Shreve [1991] for more details. In particular, when also imposing regularity conditions on the derivatives of the coefficients of the SDE, the fundamental solution can also be used to create a solution to the Kolmogorov forward equation. The Kolmogorov forward equation is hereby given through the following definition.

Definition 2.1.9.

i) We define the operator $\mathcal{A}_t^*$ for a function $f \in \mathcal{C}^2([0,\infty), \mathbb{R}^p)$ as

$$\mathcal{A}_t^* f(s,y) = -\sum_{i=1}^{p} \partial_{y_i}[\mu_i(s,y) f(s,y)] + \frac{1}{2} \sum_{i,j=1}^{p} \partial_{y_i y_j}\big[\Sigma_{i,j}(s,y) f(s,y)\big].$$

The operator $\mathcal{A}_t^*$ can be viewed as the L^2-adjoint to $\mathcal{A}_t$.

ii) The Kolmogorov forward equation for a pair $(t,x) \in [0,\infty] \times \mathbb{R}^p$, also known as Fokker-Planck equation, is defined as

$$\begin{cases} -\partial_s u(s,y) + \mathcal{A}^* u(s,y) = 0 \\ u(t,y) = \delta(x-y). \end{cases}$$

Considering the fundamental solution $p(t,x;s,y)$ under the suitable regularity conditions of Karatzas and Shreve [1991], the function $p(t,x;s,y)$ with fixed variables (t,x) then also satisfies the Kolmogorov forward equation of Definition 2.1.9. This connection between SDEs and the solutions of PDEs, established through the Kolmogorov equations, provides a vital step in our further studies when analysing the transition density functions corresponding to certain Itô processes. We use both the Kolmogorov backward and forward equation to characterize these transition densities and consider approximations of the associated PDEs.

Remark 2.1.10. Note that the presented version of Theorem 2.1.7 can in fact be further extended to provide an analogue to the Feyman-Kac theorem. We refer to Karatzas and Shreve [1991, Theorem 5.7.6] for the details.

2.2 Application to Finance

Having established the underlying mathematical concepts of SDEs in Section 2.1, this section outlines the application in a financial context. In a similar manner to Section 2.1 this section should be seen as a condensed summary of standard literature on interest rate modelling. We use Andersen and Piterbarg [2010a], Brigo and Mercurio [2006], Musiela and Rutkowski [2005] as our primary references. We focus on the application towards interest rates, explaining the selection of the corresponding literature. Many of the models presented in the referenced literature, can, however, also be used to model other types of asset classes such as equities. Further literature can be found in the references therein and an application towards finance is also often included in the standard literature concerning stochastic analysis such as Øksendal [2003] or Karatzas and Shreve [1991].

2.2.1 Arbitrage

We begin by establishing a rigorous mathematical interpretation of the concept of arbitrage. As already mentioned in the introduction, arbitrage can be informally seen as the possibility to gain money without taking any risk, making arbitrage one of the fundamental concepts in mathematical finance. We consider a p-dimensional Itô process X_t given by the SDE (2.1.2). Here each coordinate process $X_{t,i}$ with $1 \leq i \leq p$ models an asset and the entirety of the p assets yields our model framework. Notice that not each asset has to be physically tradable and can for example also represent a volatility process. To create a plausible model framework based on these assets, we need to avoid the existence of arbitrage. Putting this into a rigorous mathematical context, we summarize the concepts of Andersen and Piterbarg [2010a, Chapter 1.2] in the following definition.

Definition 2.2.1.

i) **Trading strategy:** A trading strategy $\phi : [0,T] \times \Omega \to \mathbb{R}^p$ is a predictable and adapted process. Each component of a trading strategy can be interpreted as the holdings at time t in the i-th asset.

ii) **Value:** The value $\pi : [0,T] \times \Omega \to \mathbb{R}$ of a trading strategy ϕ is given by

$$\pi_t = \pi(t,\omega) = \phi(t,\omega)^\top X_t(\omega) = \sum_{i=1}^{p} \phi_i(t,\omega)^\top X_{t,i}(\omega).$$

iii) **Trading gains:** The trading gain $G : [0,T] \times \Omega \to \mathbb{R}$ of a trading strategy ϕ is given by

$$G_t = G(t,\omega) = \int_0^t \phi(u,\omega)^\top dX_u(\omega).$$

iv) **Self-financing:** A trading strategy ϕ is called self-financing, if for every $t \in [0,T]$ we have

$$\pi_t - \pi_0 = G_t.$$

v) **Arbitrage opportunity:** An arbitrage opportunity is a self-financing trading strategy ϕ for which $\pi_0 = 0$ and there exists a $t \in [0,T]$ such that

$$\pi_t \geq 0 \text{ a.s., and } \mathbb{P}(\pi_t > 0) > 0.$$

Definition 2.2.1 provides a very intuitive way to provide a mathematical definition of arbitrage using arbitrage opportunities. As depicted in e.g. Musiela and

Rutkowski [2005] this definition is adequate for discrete-time models. When considering continuous-time models, however, the definition is not sufficient to rule out arbitrage. Hence, to rigorously enforce the no-arbitrage property in the model framework, we introduce the concept of equivalent martingale measures provided by Brigo and Mercurio [2006, Definition 2.1.2].

Definition 2.2.2. An equivalent martingale measure $\mathbb{Q}^D$ is a probability measure on the space $(\Omega, \mathcal{F})$ that satisfies the following properties:

i) **Equivalence:** The measures $\mathbb{Q}^D$ and $\mathbb{P}$ are equivalent, that is for all $A \in \mathcal{F}$:

$$\mathbb{Q}^D(A) = 0 \Longleftrightarrow \mathbb{P}(A) = 0.$$

ii) **Radon-Nikodym derivative:** The Radon-Nikodym derivative

$$R = \frac{d\mathbb{Q}^D}{d\mathbb{P}}$$

is square integrable with regard to the measure $\mathbb{P}$.

iii) **Discounted assets:** The normalized asset process

$$X_t^D := \frac{X_t}{D_t}$$

with regard to a strictly positive Itô process D_t, called numéraire, is a martingale with respect to $\mathbb{Q}^D$. This means, X_t^D is an adapted process with $\mathbb{E}^{\mathbb{Q}^D}\left[|X_t^D|\right] < \infty$ for all $t \in [0,T]$ and satisfies for all $s,t \in [0,T]$ with $t \leq s$ the equation

$$\mathbb{E}^{\mathbb{Q}^D}\left[X_s^D | \mathcal{F}_t\right] = X_t^D \qquad a.s.$$

This definition is constructed in dependence on the numéraire D_t and therefore moves the considerations from the original price process X_t to the relative price process X_t^D. Thus, we also extend the definitions related to trading strategies to incorporate the relative price process. We use a slight abbreviation of Musiela and Rutkowski [2005, Definition 8.1.4] to bring the notation more in line with Andersen and Piterbarg [2010a].

Definition 2.2.3. A self-financing trading strategy ϕ is called permissible if the relative gain process

$$G_t^D = G^D(t,\omega) = \int_0^t \phi(u,\omega)^\top dX_u^D(\omega)$$

is a martingale with respect to $\mathbb{Q}^D$.

With this we have established the necessary preliminaries to formulate a first characterization to avoid arbitrage. The result is based on Andersen and Piterbarg [2010a, Theorem 1.3.2] and can also be found in Musiela and Rutkowski [2005, Proposition 8.1.1] including the full proof of the theorem.

Theorem 2.2.4. *Restrict attention to permissible trading strategies. If there is a numéraire D such that the normalized asset price process allows for an equivalent martingale measure, then there are no arbitrage opportunities.*

At this point we want to highlight that Theorem 2.2.4 only provides a sufficient condition for no-arbitrage and the inverse direction is not true in this formulation. It is possible to establish a equivalence between the existence of equivalent martingale measures and the concept of no-arbitrage and this result is known as the first fundamental theorem of asset pricing. The derivation of this result is however very technical and requires a different definition of no-arbitrage than the one provided in Definition 2.2.1. Hence we refer to Musiela and Rutkowski [2005, Section 8.1.5] and the therein mentioned work of Delbaen and Schachermayer [1994a,b, 1998] for further details. For our purposes the results of Theorem 2.2.4 are sufficient since they allow to guarantee the absence of arbitrage opportunities when pricing derivatives under a martingale measure. We will further concretize this in the next section.

2.2.2 Pricing of Contingent Claims

Having established the mathematical foundation to guarantee the no–arbitrage condition in our model framework, we continue to analyse the pricing of certain derivatives. We present concrete examples how these derivatives may look like in Section 2.2.4. For the moment, however, we will keep the notation general and consider contingent claims in accordance to Andersen and Piterbarg [2010a, Section 1.4].

Definition 2.2.5.

i) A T-maturity contingent claim is an $\mathcal{F}_T$-measurable random variable V_T with finite variance which makes no payments before time T.

ii) The contingent claim V_T is called attainable or replicatable if there exists a permissible trading strategy ϕ with value π such that

$$V_T = \pi_T \text{ a.s.}$$

This leads us to a second very important property which we impose on our model set-up, the completeness of the market.

Definition 2.2.6. The market is called complete if all finite-variance $\mathcal{F}_T$-measurable random variables can be replicated.

In essence, this means that we can span all considered contingent claims using our pricing process and a corresponding trading strategy. To impose this property, we can again consider the equivalent martingale measures of Definition 2.2.2. This connection is known as the second fundamental theorem of asset pricing and can be summarized as done in Andersen and Piterbarg [2010a, Theorem 1.4.1].

Theorem 2.2.7. *In the absence of arbitrage, a market is complete if and only if there exists a numéraire inducing a unique martingale measure.*

The focus of this theorem is placed on the uniqueness of the induced martingale measure. As stated in Musiela and Rutkowski [2005, Proposition 8.1.1] this also induces a unique replicating strategy for the contingent claim, which in turn also allows to deduce a unique price for all finite-variance $\mathcal{F}_T$-measurable random variables.

In summary, using a model set-up where it is possible to construct a unique equivalent martingale measure provides us with the desired model properties. We will now use this set-up to subsequently compute the unique price for a T-maturity contingent claim V_T. The underlying pricing formula for this is provided by e.g. Brigo and Mercurio [2006, Proposition 2.1.2].

Proposition 2.2.8. *Assume there exists an equivalent martingale measure $\mathbb{Q}^D$ induced by the numéraire D_t and let V_T be an attainable contingent claim. Then, for each time t, $0 \leq t \leq T$, there exists a unique price π_t associated with V_T, i.e.,*

$$V_t = \pi_t = D_t \mathbb{E}^{\mathbb{Q}^D}\left[\frac{V_T}{D_T}\middle|\mathcal{F}_t\right].$$

Proposition 2.2.8 basically states that under the measure $\mathbb{Q}^D$ the discounted contingent claim $\frac{V_T}{D_T}$ becomes a martingale. The formula then follows from the martingale property. For a practical application it now remains to evaluate this conditional expectation. This evaluation is heavily dependent on the chosen measure $\mathbb{Q}^D$, respectively its underlying numéraire. Thus, we continue to present some useful tools to allow a change of the underlying measure with the aim of simplifying the evaluation of the conditional expectation.

2.2.3 Change of Measures

The first step to deduce the desired toolkit is to connect the measures induced by two different numéraires with each other. This result is given in Brigo and

Mercurio [2006, Proposition 2.2.1] or equivalently Andersen and Piterbarg [2010a, Theorem 1.4.2].

Theorem 2.2.9. *Consider two numéraires N_t and M_t, inducing equivalent martingale measures $\mathbb{Q}^N$ and $\mathbb{Q}^M$, respectively. If the market is complete, then the Radon-Nykodym derivative relating the two measures is uniquely given by*

$$\mathbb{E}_t^{\mathbb{Q}^N}\left[\frac{d\mathbb{Q}^M}{d\mathbb{Q}^N}\right] = \frac{M(t)N(0)}{N(t)M(0)}.$$

Theorem 2.2.9 allows us to perform a change of measures, by changing the underlying numéraire. Let us denote the Radon-Nykodym density process by ζ_t and consider the pricing formula of Proposition 2.2.8 to get

$$V_t = N_t\mathbb{E}^{\mathbb{Q}^N}\left[\frac{V_T}{N_T}\middle|\mathcal{F}_t\right] = N_t\mathbb{E}^{\mathbb{Q}^M}\left[\frac{V_T}{N_T}\frac{\zeta_t}{\zeta_T}\middle|\mathcal{F}_t\right] = M_t\mathbb{E}^{\mathbb{Q}^M}\left[\frac{V_T}{M_T}\middle|\mathcal{F}_t\right].$$

This shows, that the computed price of the contingent claim is independent of the chosen numéraire. Hence, we can freely chose the numéraire to simplify the evaluation of the conditional expectation as much as possible. These results can also be found in Brigo and Mercurio [2006, Section 2.3] providing a nice summary of practical applicable tools.

To motivate the next step, let us for the moment assume that the relative contingent claim $V_T^D = \frac{V_T}{D_T}$ is expressible in terms of an SDE of the form

$$dV_t^D = \mu_t\,dt + \,dW_t.$$

Considering the initial price at time 0 provided by the pricing formula of Proposition 2.2.8, the computation reduces to an evaluation of the expected drift term. Depending on the function μ_t this problem can still become quite challenging and we prefer to chose a numéraire that avoids this computation. To select a suitable numéraire, or more precisely a suitable new measure, we present a very useful result known as 'Girsanov's Theorem'. We refer to the version stated in Karatzas and Shreve [1991, Theorem 3.5.1] and adjust the notation to be more in line with Andersen and Piterbarg [2010a].

Theorem 2.2.10. *Let θ_t be a d-dimensional measurable, adapted process satisfying $\forall i \leq d, 0 \leq t < \infty$*

$$\int_0^t \theta_{s,i}^2\,ds < \infty \ \ a.s.$$

and define the density process ζ_t^θ as its Dolean exponential given by

$$\zeta_t^\theta = e^{\int_0^t \theta_s^\top\,dW_s - \frac{1}{2}\int_0^t \theta_s^\top\theta_s\,ds}.$$

Suppose that ζ_t^θ is a martingale under the measure $\mathbb{P}$. Define a new process W_t^θ for $0 \leq t < \infty$ as

$$W_t^\theta = W_t - \int_0^t \theta_s \, ds.$$

Then for each fixed $T \in [0, \infty)$ the process $\{W_t^\theta, t \leq T\}$ is a d-dimensional Brownian motion under the measure $\mathbb{P}^\theta$ defined through the density process ζ_t^θ as

$$\frac{\mathbb{P}^\theta}{\mathbb{P}} = \zeta^\theta.$$

To conclude this section, we provide a series of remarks to clarify the statements and implications of Theorem 2.2.10.

Remark 2.2.11. The density process ζ_t^θ of Theorem 2.2.10 can be characterized as the solution to the SDE

$$d\zeta_t^\theta = \zeta_t^\theta \theta_t^\top \, dW_t.$$

This provides us with a representation of the measure change in terms of an SDE. This characterization is determined by the coefficient θ_t which provides the approach to steer the change of measures.

Remark 2.2.12. Girsanovs's Theorem assumes the density process to be a true martingale. This condition is not trivial and can be quite difficult to enforce. One of the most prominent assumptions to guarantee this is the 'Novikov Condition', see e.g. Andersen and Piterbarg [2010a], requiring that

$$\mathbb{E}^{\mathbb{P}}\left[e^{\frac{1}{2}\int_0^t \theta_s^\top \theta_s \, ds}\right] < \infty.$$

The Novikov condition guarantees that ζ_t^θ is a true martingale.

Remark 2.2.13. Considering the new Brownian motion W_t^θ, Girsanov's Theorem provides a handy way so simplify the drift term in an SDE. Considering e.g. an SDE under the measure $\mathbb{P}$ of the form

$$dV_t = -\sigma_t \theta_t \, dt + \sigma_t \, dW_t$$

with θ_t satisfying the conditions of Theorem 2.2.10, we can perform a change of measure and consider the equivalent SDE under the measure $\mathbb{P}^\theta$ given by

$$dV_t = \sigma_t \, dW_t.$$

Remark 2.2.14. Based on the example in Remark 2.2.13, this approach also provides a useful tool to check for the existence of equivalent martingale measures, see Andersen and Piterbarg [2010a, Corollary 1.5.2]. More precisely, if the drift term of the Itô process in (2.1.2) allows a decomposition as in Remark 2.2.13, an equivalent martingale measure exists.

2.2.4 Fixed Income Instruments

Having discussed the general mathematical set-up to characterize our financial market model, we provide more intuition on the practical applicability of this framework. For this we focus on fixed income markets, introduce some of the most prominent market instruments available and show how these can be embedded in our financial market model. The definition of these instruments can be found in all standard literature concerning interest rate models such as Brigo and Mercurio [2006], Musiela and Rutkowski [2005], Andersen and Piterbarg [2010a]. Due to the highest correspondence of notation towards our current financial market set-up, we use Andersen and Piterbarg [2010a] throughout this section for all relevant definitions.

2.2.4.1 Zero-coupon Bond

We start with one of the simplest fixed income instruments, a certificate of deposit. As the name suggests, the instrument is created by investing a unit of money for a certain time interval. At the end of this interval the investor receives the investment back in addition with some interest payment. We formalize this with the following definition.

Definition 2.2.15. (Based on Andersen and Piterbarg [2010a, Section 5.2])
A certificate of deposit is a deposit of money at a time T for a pre-specified term τ and at a pre-specified simple interest rate $L(T, T+\tau)$. The value of an investment of 1 unit of money at time T yields to a value of

$$V_{T+\tau}^{CD} = 1 + \tau L(T, T+\tau)$$

at time $T + \tau$.

In Definition 2.2.15 the value is expressed based on a unit investment at time T and the value represents a projection of the current time value to the future time $T+\tau$. For most applications, however, it is more convenient to have a projection along the other direction and be able to deduce the current value of future payments at time $T + \tau$. For this we define the zero-coupon bond.

Definition 2.2.16. (Based on Andersen and Piterbarg [2010a, Section 4.1.1])
A T-maturity zero-coupon bond, also known as discount bond, is a contract that guarantees its holder the payment of 1 unit of currency at time T, with no intermediate payments. The contract value at time $t \leq T$ is denoted by $DF(t, T)$.

Combining these two definitions, we see that investing a value of $DF(T, T+\tau)$ into a certificate of deposit has to yield a value of 1 at time $T+\tau$. If this is not the

case, arbitrage opportunities would arise. Thus, we can conclude the relationship

$$DF(T, T+\tau)\big(1+\tau L(T, T+\tau)\big) = 1. \tag{2.2.1}$$

We use this expression to characterize interest rates in a future looking set-up.

Definition 2.2.17. (Based on Andersen and Piterbarg [2010a, Section 4.1.1])

i) **Forward bonds:** The time t forward price $DF(t, T, T+\tau)$ when deciding at time t to purchase at some time T a zero-coupon bond maturing at time $T+\tau$ is defined as

$$DF(t, T, T+\tau) = \frac{DF(t, T+\tau)}{DF(t, T)}.$$

The bond is called a forward bond with expiry T, maturity $T+\tau$ and tenor τ.

ii) **Simple forward rate:** We define the simple forward rate $L(t, T, T+\tau)$ in analogy to Equation (2.2.1) as

$$L(t, T, T+\tau) = \frac{1 - DF(t, T, T+\tau)}{\tau DF(t, T, T+\tau)} = \frac{DF(t, T) - DF(t, T+\tau)}{\tau DF(t, T+\tau)}.$$

In the case where $t = T$ the forward rates are instead called spot rates and correspond to $L(T, T+\tau)$.

We now have established the necessary variables to consider a first instrument to which we can apply our results from section 2.2.2 and 2.2.3.

Definition 2.2.18. (Based on Andersen and Piterbarg [2010a, Section 5.3])
A forward rate agreement (FRA) for the period $[T, T+\tau]$ is a contract to exchange a fixed rate payment given by the rate k against the payment of the time T spot rate with tenor τ. All payments are exchanged at time T and are expressed in time $T+\tau$ units. The value at time T of the FRA is given by

$$V_T^{FRA} = DF(T, T+\tau)\tau\big(L(T, T, T+\tau) - k\big).$$

From the definition of the FRA, we can observe that the contract is formulated in a manner to specify the payments at a certain time. This is often the case for many instruments and allows the embedding into our financial model set-up. In the case of the FRA, the value at time T is explicitly specified. Assuming that the discount factor, and in turn also the spot rate, is governed by a stochastic behaviour, the value V_T^{FRA} becomes an $\mathcal{F}_T$-measurable random variable and we can apply the results of Section 2.2.2.

Lemma 2.2.19. *The time t value of the FRA is given by*

$$V_t^{FRA} = DF(t, T+\tau)\tau\big(L(t,T,T+\tau) - k\big).$$

In particular, the forward rate $L(t,T,T+\tau)$ can be characterized as the rate k that makes the FRA a fair contract, i.e. lets the time t value equal 0*.*

Proof. Assuming that we have an equivalent martingale measure induced by a numéraire N_t, we can express the time t value as

$$V_t^{FRA} = N_t \mathbb{E}_t^{\mathbb{Q}^N}\left[\frac{V_T^{FRA}}{N_T}\right] = N_t \mathbb{E}_t^{\mathbb{Q}^N}\left[\frac{DF(T,T+\tau)\tau\big(L(T,T,T+\tau)-k\big)}{N_T}\right].$$

Using the traded asset $DF(t,T+\tau)$ as our numéraire N_t, we can simplify the conditional expectation and deduce

$$V_t^{FRA} = DF(t,T+\tau)\tau\left(\mathbb{E}_t^{\mathbb{Q}^{DF(\cdot,T+\tau)}}\big[L(T,T,T+\tau)\big] - k\right).$$

Considering the definition of the forward rate $L(t,T,T+\tau)$ provided in Definition 2.2.17 we see that $L(t,T,T+\tau)$ is given as the sum of traded assets divided by $DF(\cdot,T+\tau)$. Therefore, under the measure $\mathbb{Q}^{DF(\cdot,T+\tau)}$ also $L(\cdot,T,T+\tau)$ is a martingale and we can conclude

$$\mathbb{E}_t^{\mathbb{Q}^{DF(\cdot,T+\tau)}}\big[L(T,T,T+\tau)\big] = L(t,T,T+\tau).$$

□

2.2.4.2 Swap Rates, Swaps and CMS Swaps

In the next step, we increase the complexity of the instruments considered. We start with a first natural extension and introduce coupon payments into our considerations. For these we need to refine our time-line and include multiple payment dates. Throughout this section we work on a tenor structure of the form

$$T_1 = t_0 < t_1 < \ldots < t_n \leq T_2 \tag{2.2.2}$$

where T_1 denotes the starting date, T_2 the end date and t_i the payment dates of the instrument. We denote the timespan between payments as $\delta_i = t_i - t_{i-1}$ for $i \geq 1$.

Definition 2.2.20. (Based on Andersen and Piterbarg [2010a, Section 4.1.3])

i) **Coupon-bearing bond:** A t-maturity coupon-bearing bond with coupon c is a contract that guarantees its holder the payment of 1 unit of currency at time T_2 and in addition provides intermediate payments of $c\delta_i$ at time t_i for all $i \geq 1$ in the time period $[T_1, T_2]$. The contract value at time $t \leq T_1$ is given by

$$V_t^{CB} = \sum_{T_1 < t_i \leq T_2} c\delta_i DF(t, t_i) + DF(t, T_2).$$

ii) **Annuity factor:** We define the annuity factor $A(t, T_1, T_2)$ as the value generated only by the coupon payments of a coupon-bearing bond with a coupon of $c = 1$, i.e.

$$A(t, T_1, T_2) = \sum_{T_1 < t_i \leq T_2} \delta_i DF(t, t_i).$$

This simplifies the value of the coupon-bearing bond with coupon c at time t to

$$V_t^{CB} = cA(t, T_1, T_2) + DF(t, T_2).$$

iii) **Swap rate:** We define the time t forward swap rate $R(t, T_1, T_2)$, also denoted as par rate, as

$$R(t, T_1, T_2) = \frac{DF(t, T_1) - DF(t, T_2)}{A(t, T_1, T_2)}.$$

To motivate the definition of the swap rate, we consider another instrument, the plain vanilla fixed-for-floating interest rate swap or in short simply swap.

Definition 2.2.21. (Based on Andersen and Piterbarg [2010a, Section 5.5])
A swap, is an exchange between one stream of payments, called leg, paying fixed rate payments with rates k and another leg with payments based on floating forward rates. Both legs are defined on the tenor structure of (2.2.2) and are exchanged at the end of an interval. Here the floating rates for the time interval $[t_i, t_{i+1}]$ are given by the time t_i spot rate $L(t_i, t_i, t_{i+1})$ for $i \geq 1$.

Lemma 2.2.22. *The time t value of a swap is given by*

$$V_t^{swap} = A(t, T_1, T_2)\big(R(t, T_1, T_2) - k\big)$$

with $A(t, T_1, T_2)$ denoting the annuity factor and $R(t, T_1, T_2)$ the swap rate. In particular, the swap rate $R(t, T_1, T_2)$ can be characterized as the rate k that makes the swap a fair contract, i.e. lets the time t value equal 0.

Proof. We start by analysing the floating leg. We know that for each time interval $[t_i, t_{i+1}]$ the interest rate is provided by the spot rate $L(t_i, t_i, t_{i+1})$ and paid at time t_{i+1}. Hence, the time t_i value of the payment coming from the interval $[t_i, t_{i+1}]$ is given by

$$V_{t_i}^{[t_i,t_{i+1}]} = DF(t_i, t_{i+1})\delta_{i+1}L(t_i, t_i, t_{i+1}).$$

Therefore, we can follow the proof of Lemma 2.2.19 and by choosing the numéraire $DF(t, t_{i+1})$ deduce the time t value of the payment coming from the interval $[t_i, t_{i+1}]$ as

$$V_t^{[t_i,t_{i+1}]} = DF(t, t_{i+1})\delta_{i+1}L(t, t_i, t_{i+1}).$$

Recalling Definition 2.2.17, this can further be simplified to

$$V_t^{[t_i,t_{i+1}]} = DF(t, t_i) - DF(t, t_{i+1}).$$

Thus, we can conclude that the value of the floating leg at time t is given by

$$V_t^{float} = \sum_{T_1 < t_i \leq T_2} V_t^{[t_{i-1},t_i]} = DF(t, T_2) - DF(t, T_1).$$

On the other hand, the fixed leg consists only of the coupon value of a coupon-bearing bond with coupon k and according to Definition 2.2.20 is given by

$$V_t^{fix} = kA(t, T_1, T_2).$$

Combining these legs we get

$$\begin{aligned} V_t^{swap} &= V_t^{float} - V_t^{fix} = DF(t, T_2) - DF(t, T_1) - kA(t, T_1, T_2) \\ &= A(t, T_1, T_2)\big(R(t, T_1, T_2) - k\big). \end{aligned}$$

□

Swaps belong to the most frequently traded instruments and therefore the swap marked is very liquid. This allows to consider instruments based on swaps as the underlying. One example are the constant maturity swaps (CMS) which follow the swap structure provided in Definition 2.2.21 with the slight adjustment that the floating rate is not based on a forward rate and instead on a swap rate.

Definition 2.2.23. (Based on Andersen and Piterbarg [2010a, Section 5.11])
A constant maturity swap is an exchange between one stream of payments, called leg, paying fixed rate payments with rates k and another leg with payments based on a swap rate with fixed maturity. The rate k that makes this contract fair for both sides is called the CMS rate.

In our considerations, both the swap as well as the CMS are defined containing a fixed payment leg. Both constructions can however be defined in a much greater generality allowing e.g. the exchange of variable rates on both payment legs (floating for floating), different payment dates or the averaging of rates. For a more detailed study of the various possibilities we refer to Andersen and Piterbarg [2010a], Brigo and Mercurio [2006]. For our purpose we will mainly consider European options on swaps and CMS, and the presented definitions are sufficient. We continue by discussing European options in the next section.

2.2.4.3 European Options

Regarding the contracts discussed until now, there is always an obligation included to execute the contract. From a practical perspective it is however not always desirable to be forced to perform the exchange. This motivates us to include optionality into the contract yielding so-called options. We consider optionality of the European-style, where the option can only be exercised at a pre-specified exercise date. The standard European-style options are called calls and puts. We use the characterization given in Musiela and Rutkowski [2005].

Definition 2.2.24. Let S_t denote the value of an underlying at time t. This can e.g. be a stock price or a forward rate. We define the following European-styled options.

i) **Call:** A (European) call option with strike k and maturity T is the right, but not the obligation, to buy the underlying S_t at the maturity time T for a price k.

ii) **Put:** A (European) put option with strike k and maturity T is the right, but not the obligation, to sell the underlying S_t at the maturity time T for a price k.

Since options are one of the most frequently traded financial instruments, they often build the underlying foundation to adequately calibrate a financial market model to an observable market. This means, the parameters available in the model are specified in such a way, that the option values observed in the market are reproduced in the model. Thus, it is important to be able to properly price these options. We now introduce some of the most prominent European-style options based on interest rates.

Definition 2.2.25. (Based on Andersen and Piterbarg [2010a, Section 5.8])

i) **Caplet:** A caplet is a scaled call option with strike k and maturity T on the

forward rate $L(t,T,T+\tau)$ such that the value at time $T+\tau$ is given by

$$V_{T+\tau}^{caplet} = \tau\big(L(T,T,T+\tau)-k\big)^+.$$

In particular, the caplet can be seen as the right, but not the obligation, to pay the fixed rate of the FRA for the period $[T,T+\tau]$ of Definition 2.2.18.

ii) **Floorlet:** A floorlet is a scaled put option with strike k and maturity T on the forward rate $L(t,T,T+\tau)$ such that the value at time $T+\tau$ is given by

$$V_{T+\tau}^{floorlet} = \tau\big(k-L(T,T,T+\tau)\big)^+.$$

In particular, the floorlet can be seen as the right, but not the obligation, to pay the floating rate of the FRA for the period $[T,T+\tau]$ of Definition 2.2.18.

iii) **Caps/Floors:** A cap/floor is a strip of caplets/floorlets on successive forward rates spanning the tenor structure of (2.2.2).

In a similar fashion we can define the right to enter into a swap or a CMS.

Definition 2.2.26. (Based on Andersen and Piterbarg [2010a, Section 5.10/5.11])

i) **Swaptions:** A European payer swaption, or simply swaption, is an option to pay the fixed leg of a fixed-for-floating swap. The value of the swaption with strike k at the time of the expiry T_1, which is the time when the swap starts, is given by

$$V_{T_1}^{swaption} = \Big(A(T_1,T_1,T_2)\big(R(T_1,T_1,T_2)-k\big)\Big)^+.$$

In a similar fashion, a receiver swaption can be defined as the right to pay the floating leg in a swap.

ii) **CMS options:** A CMS cap is an option to pay the fixed leg of a CMS. A CMS floor is an option to pay the floating leg of a CMS.

Considering the payoff functions available for these types of derivatives, we can observe that it is not possible to derive a valuation formula for some earlier time t without further knowledge of the dynamics of the underlying interest rate. Considering for example the case of a swaption, we can apply the results of Proposition 2.2.8 to conclude

$$V_t^{swaption} = A(t,T_1,T_2)\mathbb{E}_t^{\mathbb{Q}^{A(\cdot,T_1,T_2)}}\Big[\big(R(T_1,T_1,T_2)-k\big)^+\Big].$$

Since $\big(R(T_1,T_1,T_2)-k\big)^+$ is not a martingale under the measure $\mathbb{Q}^{A(\cdot,T_1,T_2)}$, the evaluation has to depend on the underlying dynamics of the swap rate. This in

turn is heavily dependent on the chosen financial market model and we will show in Section 2.3 how this can be done for specific models.

Finally, we are also interested in the pricing of options on baskets, meaning a dependence on multiple rates. Our main example for this is the CMS spread option. For this we refer to Brigo and Mercurio [2006, Section 13.16.2].

Definition 2.2.27. A CMS spread caplet with strike k is an option on the spread between two different swap rates with tenors α and β such that the payoff at a paydate T is given by

$$\big(R(T,T_1,T_1+\beta)-R(T,T_1,T_1+\alpha)-k\big)^+.$$

We will dive deeper into the pricing of these derivatives in Chapter 4 where we derive pricing formulas for these type of instruments under our chosen model set-up.

2.3 Prominent Financial Market Models

As seen in the explicit example of Section 2.2.4.3, we arrive at an expression where we have to solve a conditional expectation which is not trivial. Section 2.2.3 provides some tools to ease this computation, everything is however dependent on the explicit model dynamics chosen to determine the relevant price processes of (2.1.2). Thus, to continue our studies in this area we have to make the general set-up more explicit and introduce more structure onto the coefficients of (2.1.2). In this section we provide some overview of commonly used financial market models to explicitly specify these coefficients. We outline some important properties and the restrictions of those choices. Some of these models, such as the Heston model, Heston [1993], or the stochastic alpha beta rho (SABR) model, Hagan et al. [2002], have been intensively studied and this section should again be understood as a general summary to provide an overview of possible choices.

2.3.1 The Black-Scholes-Merton Model

We start by considering one of the most famous financial market models as an introductory example, the Black-Scholes-Merton model of Black and Scholes [1973], Merton [1973]. In the Black-Scholes-Merton model, often also only denoted as Black-Scholes model, the full price process $X_t = (S_t, \beta_t)^\top$ is characterized by the

two-dimensional Itô process

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$$

with coefficients

$$\mu(t, X_t) = \begin{pmatrix} \mu S_t \\ r\beta_t \end{pmatrix} \quad \text{and} \quad \sigma(t, X_t) = \begin{pmatrix} \sigma S_t & 0 \\ 0 & 0 \end{pmatrix}$$

and the initial value

$$X_0 = \begin{pmatrix} S_0 \\ 1 \end{pmatrix}.$$

In the original model the process S_t denotes a stock process and the process β_t the continuously compounded risk-free interest. Considering the definition of the process β_t, no stochastic behaviour is included and therefore the focus is placed on the risky stock process S_t. As demonstrated in Section 2.2.1 instead of considering the price process directly, we can consider the relative price process with regard to a numéraire. Choosing the process β_t as a numéraire, we define the relative stock process S_t^β as

$$S_t^\beta = \frac{S_t}{\beta_t}.$$

In particular, the dynamics of S_t^β are now given by the one-dimensional equation

$$dS_t^\beta = (\mu - r)S_t^\beta\, dt + \sigma S_t^\beta\, dW_t. \tag{2.3.1}$$

This equation is also known as the characterization of a geometric Brownian motion and we can now demonstrate how the change of measures of Section 2.2.3 is applicable in practice. Following Theorem 2.2.10 we define the θ-function as

$$\theta_t \equiv \frac{r - \mu}{\sigma}$$

and introduce a new Brownian motion

$$W_t^\beta = W_t - \frac{r - \mu}{\sigma} t$$

under the measure $\mathbb{Q}^\beta = \mathbb{Q}^\theta$. Considering (2.3.1) under this new measure, the dynamics can be simplified and lead to

$$dS_t^\beta = \sigma S_t^\beta\, dW_t^\beta.$$

Hence, the measure $\mathbb{Q}^\beta$ is an equivalent martingale measure and no arbitrage opportunities exist in the model. Furthermore, the SDE can now be explicitly solved. The solution can then be transformed back in terms of the original stock process S_t yielding an expression of the form

$$S_t = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma(W_t^\beta - W_0^\beta)}$$

for all $t \geq 0$. Being able to explicitly solve the underlying dynamics of the stock process provides an advantage in further calculations. Considering for example a call option on the process S_t, we can deduce a closed form expression for the price which is captured in the famous Black-Scholes pricing formula.

Theorem 2.3.1. *(Andersen and Piterbarg [2010a, Theorem 1.9.1]) Consider a call option on the process S_t with strike K and maturity T. Then the arbitrage free time t price is given by*

$$V_t^{call} = S_t\Phi(d_+) - Ke^{-r(T-t)}\Phi(d_-),$$

with

$$d_\pm = \frac{\ln\left(\frac{S_t}{K}\right) + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

and $\Phi(\cdot)$ denoting the Gaussian cumulative distribution function.

We observe that the model possesses many desirable features. This includes explicit formulas for the process and the call price or an explicitly known distribution of the underlying process. On the other hand, the model is also rather simple. Considering the volatility σ of the underlying price process, a constant value is too simplistic to match the actual observed market behaviour. Here for each strike K and maturity T a different value of σ would be required to reproduce the prices. This market behaviour often results in shapes for σ which are heavily skewed or exhibit a smile behaviour. There are a lot of further extensions to the Black-Scholes-Merton model including additional time dependence, a dividend process or a multi-dimensional set-up. We refer to Andersen and Piterbarg [2010a, Chapter 1.9] and the references therein for a continuation of this example introducing various extensions. We continue to focus further on the replicability of the observable market prices, leading to local volatility models.

2.3.2 Local Volatility Models

Given a call option with strike K, maturity T, the risk-free rate r as well as the initial value S_0, the only remaining variable in the formula of Theorem 2.3.1 is

the volatility σ. To highlight the dependence on the model we denote it as σ_B. This means, given a fixed call price, we can find a volatility σ_B which reproduces this price under the Black-Scholes-Merton model. Since the volatility is implied through the call price this is called the implied Black volatility. Using observable option prices and considering the dependence of the implied Black volatility onto the strike value K, often a skew or a smile shaped curve is observable. To capture this variety of possible shapes, local volatility models can be used. We consider two types of local volatility models, Dupire local volatility models in accordance to Dupire [1994], Derman and Kani [1994] and Bachelier local volatility models in accordance to Costeanu and Pirjol [2011].

2.3.2.1 Dupire Local Volatility Model

The Dupire local volatility model due to Dupire [1994], Derman and Kani [1994] is characterized by the one-dimensional Itô process X_t given through the expression

$$dX_t = \vartheta_{Dupire}(t, X_t) X_t \, dW_t \tag{2.3.2}$$

with initial value X_0. For our considerations we choose a definition without a drift term. For a general formulation including a drift term, we refer to Dupire [1994], Derman and Kani [1994]. As demonstrated in Dupire [1994] this framework allows to compute the price of call options $C(T, K)$ with strike K and maturity T using an equation of the form

$$\begin{cases} \partial_t C(t, K) = \dfrac{1}{2} \vartheta_{Dupire}(t, K)^2 K^2 \partial_{KK} C(t, K) \\ C(0, K) = (X_0 - K)^+. \end{cases} \tag{2.3.3}$$

The additional flexibility through the state-dependent local volatility function $\vartheta_{Dupire}(\cdot, \cdot)$ allows to perfectly reproduce the observed market smiles using (2.3.3). This can be achieved since the price $C(t, K)$ of a call option only depends on the values of the volatility function $\vartheta_{Dupire}(s, K)$ where $s \leq t$ and therefore allows for an iterative approach to match the volatility function to a discrete set of market points. The Dupire local volatility model is the natural model choice of a local volatility model when implied Black volatilities are considered. Due to the explicit inclusion of the term X_t in (2.3.2), the process stays positive as is the case in the Black-Scholes-Merton model. This implies a distribution of the process which is closer related to a log-normal distribution than a normal distribution, further strengthening the connection towards the Black-Scholes-Merton model. When considering process which can become negative, this representation is, however, not the most intuitive one and we continue by presenting the Bachelier local volatility model as an alternative to overcome this problem.

2.3.2.2 Bachelier Local Volatility Model

The Bachelier local volatility model in accordance to Costeanu and Pirjol [2011] is characterized by the one-dimensional Itô process X_t given through the expression

$$dX_t = \vartheta(t, X_t)\, dW_t \tag{2.3.4}$$

with initial value X_0. Similar to the Dupire local volatility model, the Bachelier local volatility model allows to price call options $C(T, K)$ with strike K and maturity T using the following equation, see Costeanu and Pirjol [2011],

$$\begin{cases} \partial_t C(t, K) = \dfrac{1}{2}\vartheta(t, K)^2 \partial_{KK} C(t, K) \\ C(0, K) = (X_0 - K)^+. \end{cases} \tag{2.3.5}$$

Since the Dupire local volatility model can be recovered from the Bachelier local volatility model using a parametrization of the form

$$\vartheta(t, X_t) = \vartheta_{Dupire}(t, X_t) X_t, \tag{2.3.6}$$

the properties from the Dupire local volatility model are transferred to the Bachelier local volatility model. From a practical perspective, the Bachelier local volatility model does however allow for a more natural consideration of negative processes and distributions which are more related to a normal distribution than a log-normal distribution. This is a common feature when considering interest rates and therefore also the market prices are often quoted in terms of normal implied volatilities and not implied Black volatilities. Here the normal implied volatility σ_n is the parameter which enables the Bachelier model to reproduce the observed market price using a Bachelier pricing formula of the form

$$C_\phi(T, K) = (X_0 - K)\, \Phi(d_n) + \sigma_n \sqrt{T} \phi(d_n)$$

with

$$d_n = \frac{X_0 - K}{\sigma_n \sqrt{T - t_0}},$$

and functions $\Phi(\cdot)$ and $\phi(\cdot)$ denoting the normal CDF and PDF respectively. Here $C_\phi(T, K)$ denotes the call price with strike K and maturity T under the Bachelier model. As demonstrated in Costeanu and Pirjol [2011] it is then also possible to

express (2.3.5) in terms of implied normal volatilities, enabling a direct calibration of the local volatility function using market data and an expression of the form

$$\vartheta(t,K) = \frac{\sigma_n^2(t,y) + t\partial_t\sigma_n^2(t,y)}{\left(1 - \frac{y}{\sigma_n(t,y)}\partial_y\sigma_n(t,y)\right)^2 + t\sigma_n(t,y)\partial_{yy}\sigma_n(t,y)}. \tag{2.3.7}$$

Since our focus lies in the modelling of interest rates, we use the Bachelier local volatility model when referring to local volatility models. Due to the close association between the expression 'local volatility' and the Dupire local volatility model, we also denote the local volatility function of the Bachelier local volatility model as the projected volatility function.

While local volatility models are able to perfectly replicate the observed market implied volatility smiles, they fall short when considering the dynamics implied by the models. In addition also the inter- and extrapolation qualities for strikes outside the discrete set of market points are not optimal. Since these forward implied volatilities are vital for more exotic assets, a better modelling of the dynamics is desired. Therefore, local volatility models are a common way to reference market data instead of modelling a fully dynamical behaviour. During our work we follow this practice and use the Bachelier local volatility model as a mean to characterize market data. To further investigate the modelling of financial markets, we continue with stochastic volatility models providing a first compromise between the fit to market smiles as well as the dynamics of the forward implied volatilities.

2.3.3 Stochastic Volatility Models

Stochastic volatility models arise when moving from a one-dimensional set-up to a framework where two-dimensions are modelled. In this framework the volatility is modelled using its own stochastic dynamics. This allows to better capture the observed changes in volatilities and model observed market smile dynamics. While a general consideration of two dimensional Itô process is possible, using e.g. the Kolmogorov equations of Section 2.1.3, stochastic volatility models exhibit a large heterogeneity in their design, practical applicability and available properties. To demonstrate this behaviour and the resulting advantages and challenges, we focus in this section on explicit parametrizations for the stochastic volatility models. This results in famous models such as the Heston model, the SABR model or the ZABR model. We come back to the general framework in Chapter 3 where we discuss a general unifying approach to model stochastic volatility models. To start this section, we begin with the Heston model of Heston [1993].

2.3.3.1 The Heston Model

The Heston model due to Heston [1993] can be seen as a natural extension of the Black-Scholes-Merton model into a stochastic volatility framework, respectively a stochastic variance framework. In the original Heston model we consider a price process $X_t = (S_t, v_t)^\top$ characterized by the two-dimensional Itô process

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t$$

with coefficients

$$\mu(t, X_t) = \begin{pmatrix} \mu S_t \\ \kappa(\theta - v_t) \end{pmatrix} \quad \text{and} \quad \sigma(t, X_t) = \begin{pmatrix} \sqrt{v_t} S_t & 0 \\ \rho\nu\sqrt{v_t} & \sqrt{1-\rho^2}\nu\sqrt{v_t} \end{pmatrix}$$

and initial value $X_0 = (S_0, \alpha)^\top$. We simplify the notation for these two-dimensional Itô processes and equivalently denote these dynamics through the expression

$$\begin{cases} dS_t = \mu S_t\, dt + \sqrt{v_t} S_t\, dW_t^{(1)}, & S_0 = S_0, \\ dv_t = \kappa(\theta - v_t)\, dt + \nu\sqrt{v_t}\, dW_t^{(2)}, & v_0 = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho\, dt. \end{cases} \tag{2.3.8}$$

Note that this expression is formulated under the original measure $\mathbb{Q}$ where S_t models the price process and v_t the variance. In comparison to the Black-Scholes-Merton model, no additional risk-free process β_t is modelled, allowing an application towards interest rates. As seen in (2.3.1), the inclusion of a process β_t can be performed through an adjustment of the drift coefficient μ. Having an application towards interest rates in mind, we will use this convention for all remaining models in this thesis and do not include any deterministic risk-free processes into our financial market models.

In comparison to the Black-Scholes-Merton model, under these dynamics it is not possible to explicitly deduce an expression for the price process. As demonstrated in Heston [1993] it is however possible to use the Kolmogorov equations, see Section 2.1.3, to deduce a solvable PDE for the characteristic function of the logarithm of the price process. With this, the value of a call option with strike K and maturity T can still be expressed in a semi-analytical formula of the form

$$V_t^{call} = S_t P_1(\ln(S_t), v_t, t; \ln(K)) - K e^{-\mu(T-t)} P_2(\ln(S_t), v_t, t; \ln(K))$$

with

$$P_j(x, v, T; \ln(K)) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re\left[\frac{e^{-\mathbb{i} u \ln(K)} f_j(x, v, T; u)}{\mathbb{i} u} \right] du$$

and the characteristic functions

$$f_j(x, v, T; u) = e^{\mathbb{i}ux + C_j(T-t;u) + D_j(T-t;u)v}$$

given by

$$\begin{aligned}
C_j(\tau; u) &= \mu u \mathbb{i}\tau + \frac{\kappa\theta}{\nu^2}\Big((b_j - \rho\nu u\mathbb{i} + d_j)\tau - 2\ln\Big(\frac{1 - g_j e^{d_j\tau}}{1 - g_j}\Big)\Big),\\
D_j(\tau; u) &= \frac{b_j - \rho\nu u\mathbb{i} + d_j}{\nu^2}\Big(\frac{1 - e^{d_j\tau}}{1 - g_j e^{d_j\tau}}\Big),\\
g_j &= \frac{b_j - \rho\nu u\mathbb{i} + d_j}{b_j - \rho\nu u\mathbb{i} - d_j},\\
d_j &= \sqrt{(\rho\nu u\mathbb{i} - b_j)^2 - \nu^2(2a_j u\mathbb{i} - u^2)},\\
a_j &= \mathbb{1}_{\{j=1\}}\frac{1}{2} - \mathbb{1}_{\{j=2\}}\frac{1}{2},\\
b_j &= \kappa - \mathbb{1}_{\{j=1\}}\rho\nu.
\end{aligned}$$

We observe that the valuation of call options is much more advanced compared to Section 2.3.1. It is however still possible to derive a semi-analytical expression. Considering the complex integration necessary to evaluate the integrals, additional challenges in a practical numerical implementation may appear, see Albrecher et al. [2006]. The more realistic modelling of the volatility through its own dynamics still out-weighs these drawbacks and being able to deduce semi-analytical expressions has to be seen as an advantage compared to other approaches. Therefore, the Heston model has become one of the most prominent financial market models and a lot of research has been done to improve and extend this model. This includes the numerical implementation using various methods to evaluate the complex integral, Lord and Kahl [2010], solving the underlying PDEs directly, In 't Hout and Foulon [2010], or the simulation of paths, Staunton [2007], Chan and Joshi [2010]. For further references, different approaches and possible numerical implementations we also refer to the book of Rouah [2013], where an extensive study of the Heston model is performed. Furthermore, based on the Heston model, various model extensions have been researched. To mention a few extensions include stochastic interest rates, Fang and Janssens [2007], Grzelak and Oosterlee [2011], Kammeyer and Kienitz [2012a,b,c], or additional jump processes resulting in the Bates model, Bates [1996]. We do not dive deeper into the various model extensions and instead introduce a model often used as a reference model throughout our further studies.

2.3.3.2 The SABR Model

The SABR model of Hagan et al. [2002] is defined through the two-dimensional Itô process given by the dynamics

$$\begin{cases} dF_t = v_t F_t^{\beta}\, dW_t^{(1)}, & F_0 = f, \\ dv_t = \nu v_t\, dW_t^{(2)}, & v_0 = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\, dt. \end{cases} \tag{2.3.9}$$

The model belongs to the class of stochastic volatility models and, in comparison to the Heston model, the SABR model is often used as a reference model when considering interest rates such as forward or swap rates, under a measure where the rate is a local martingale. Thus, no drift term is included in the forward process F_t of (2.3.9). One of the reasons why the SABR model became famous are its implied volatility formulas, see Hagan et al. [2002], allowing to directly approximate the implied Black volatilities as

$$\begin{aligned} \sigma_B(t,K) \approx \alpha \Bigg((fK)^{\frac{1-\beta}{2}} \Big(1 + \frac{(1-\beta)^2}{24} \ln\Big(\frac{f}{K}\Big)^2 + \frac{(1-\beta)^4}{1920} \ln\Big(\frac{f}{K}\Big)^4 \Big) \Bigg)^{-1} \frac{z}{x(z)} \\ * \Bigg(1 + \Big(\frac{(1-\beta)^2}{24} \frac{\alpha^2}{(fK)^{1-\beta}} + \frac{1}{4} \frac{\rho\beta\nu\alpha}{(fK)^{\frac{1-\beta}{2}}} + \frac{2-3\rho^2}{24}\nu^2 \Big) T \Bigg) \end{aligned}$$

with

$$\begin{aligned} z &= \frac{\nu}{\alpha} (fK)^{\frac{1-\beta}{2}} \ln\Big(\frac{f}{K}\Big) \\ x(z) &= \ln\left(\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho} \right). \end{aligned}$$

This provides a fast and easy to implement methodology to price call options in the SABR model while at the same time having modelled an underlying stochastic behaviour of the volatility. One of the drawbacks of the original implied volatility formula was the possibility of arbitrage opportunities, visible through a negative probability density function implied through the implied volatility formula. This has been addressed by Hagan et al. [2014] and various adjustments to this formula have been applied. As in the case of the Heston model, the SABR model has been extensively studied and further developed. Including possible negative rates in Antonov et al. [2015b,a] or different approaches for evaluation and the numerical implementation, see Balland and Tran [2013], Le Floc'h and Kennedy [2017],

Kienitz et al. [2017], McGhee [2021]. In particular, the methodology applied in Hagan et al. [2014] was also applied to certain extensions of the SABR model as well as the Heston model, see Hagan et al. [2018a, 2020b]. These extensions motivate us to derive similar results regarding the implied volatility formulas for a general framework of financial market models, allowing more flexibility when choosing the model. One class of models we are particularly interested in, are models of the ZABR-type.

2.3.3.3 The ZABR Model

The ZABR model is an extension of the SABR model introduced by Andreasen and Huge [2011b] and allows for more flexibility in the volatility dynamics by introducing an additional parameter γ. The dynamics of the ZABR model are given through the SDE

$$\begin{cases} dF_t = v_t F_t^{\beta}\, dW_t^{(1)}, & F_0 = f, \\ dv_t = \nu v_t^{\gamma}\, dW_t^{(2)}, & v_0 = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho\, dt. \end{cases} \tag{2.3.10}$$

As demonstrated in Andreasen and Huge [2011b] it is possible to derive a local volatility function ϑ defining a local volatility model in accordance to Section 2.3.2 to approximate the ZABR model. This in turn allows the pricing of call options $C(T, K)$ with strike K and maturity T using the methodologies described in Section 2.3.2

In comparison to the SABR model this provides an advantage when calibrating the model parameters to observed market data and therefore allowing a better pricing of more advanced products such as CMS spreads. On the other hand, the additional flexibility also makes the model set-up more advanced and there are no explicit formulas available as in the case of the SABR model. This will be one of the driving factors motivating our research and we demonstrate in Chapter 3 how it is possible to derive implied volatility formulas similar to that of Section 2.3.3.2. Furthermore, in Chapter 3 we introduce new model variants closely related to the ZABR model. Therefore, we do not dive deeper into the existing model specifications at this point and instead move on to introduce a model class combining the properties of local volatility and stochastic volatility models.

2.3.4 Stochastic Local Volatility Models

Considering the previous model examples we have introduced two different classes of models providing their own advantages and properties. Local volatility models provide the means to perfectly replicate the observed market implied volatility smiles, do however fall short in modelling the dynamics of forward implied volatilities. On the other hand, stochastic volatility models provide an adequate approach to model these dynamics, are however, due to their parametric form, not able to capture the observed market implied volatilities perfectly. To provide a model approach which satisfies both properties, stochastic local volatility models are often the selected choice. The idea behind this class of models is to combine both approaches and use the parametric stochastic volatility model to drive the dynamics and include a non-parametric local volatility function to perfectly replicate the observed market implied volatility smiles. The non-parametric component is captured in a leverage function $\sigma_{\text{slv}}(\cdot,\cdot)$ which is introduced into the equation characterizing the stock or forward process. Taking for example the Heston model of Section 2.3.3.1 without a drift term, the extension towards a stochastic local volatility model would yield a model of the form

$$\begin{cases} dS_t = \sigma_{\text{slv}}(t,S_t)\sqrt{v_t}S_t\,dW_t^{(1)}, & S_0 = S_0, \\ dv_t = \kappa(\theta - v_t)\,dt + \nu\sqrt{v_t}\,dW_t^{(2)}, & v_0 = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\,dt. \end{cases} \tag{2.3.11}$$

Due to the additional dependencies on the process S_t, introduced in the leverage function, the pricing formulas of Section 2.3.3.1 are not applicable any more and more computational demanding methodologies have to be applied. Such methods are based on PDE or Monte Carlo methods, see Section 2.5. In particular, one of the major challenges when working with stochastic local volatility models is the calibration of the leverage function. While in a classical local volatility model, the local volatility function may be derived directly from market data, as seen in e.g. (2.3.7), this is not applicable any more when the additional stochastic component is included. Hence, the computation of prices and the calibration of the leverage function are very often intervened and performed simultaneously. Using PDE or Monte Carlo methods this results in an approach where in each computation step a price computation is performed and the leverage function is then specified to fit the prices to the observed market data. These techniques can be found in Ren et al. [2007], Saporito et al. [2019], Van der Stoep et al. [2014], Guyon and Henry-Labordère [2012] or Muguruza [2019]. We dive deeper into the theory of stochastic local volatility models in Chapter 5 where we derive a new model approach allowing to separate the calibration procedure from the price computation.

This approach builds and extends the results derived in Chapters 3 and 4, following the natural model extensions presented in this section. This separation of the calibration algorithm then allows to ease the computational demand coming from a simultaneous procedure. While our work is concerned with models which fit into the general context of stochastic local volatility models, including stochastic volatility models as a special case, there are also other possibilities available. To conclude this section, we provide a short outline of possible alternatives which are commonly found in practice.

2.3.5 Alternative Model Types

To provide a general overview we shortly outline a few alternative approaches. This list is non exhaustive and should only provide a general idea how alternatives may look like. We cover jump-diffusion models, rough volatility models as well as market models.

2.3.5.1 Jump-Diffusion Models

Jump-diffusion models provide a natural extension when the observed trajectories include discontinuities. Instead of Itô processes as in Definition 2.1.2, processes with additional jump components are considered. The jumps are modelled through a compound Poisson process. The results are stochastic paths including discontinuities. For the general theory regarding these types of models we refer to Kou [2002], Matsuda [2004], Tankov and Voltchkova [2009], Wang [2013]. For an explicit example we also refer to the Bates model of Bates [1996] which allows to extend the Heston model of Section 2.3.3.1 to the class of jump-diffusion models while still allowing the semi-analytical pricing using characteristic functions.

2.3.5.2 Rough Volatility Models

The second class of models we want to mention are rough volatility models, Gatheral et al. [2018], which have recently gained a lot of popularity. Instead of Brownian motions being the main source of randomness in the underlying stochastic behaviour, fractional Brownian motions are considered. This introduces trajectories which exhibit a smaller degree of continuity compared to the trajectories of Brownian motions, see Decreusefond and Üstünel [1999], Alòs et al. [2000]. These new trajectories look in some sense 'rougher', explaining the name of this model class. These rougher trajectories are able to better explain the observed behaviour

for the volatility and therefore allow to construct models which are capable to simultaneously fit options on the stock price as well as options on the volatility, see Gatheral et al. [2020]. On the other hand, however, a fractional Brownian motion is not Markovian any more and many classical computational techniques are not applicable. Therefore, rough volatility models are being extensively studied and we refer to Guerreiro and Guerra [2021], Gatheral [2021] for further details.

2.3.5.3 Market Models

Finally, we also want to mention the class of market models. While these types of models do not directly use different mathematical foundations compared to e.g. stochastic volatility models, the perspective in modelling differs. In market models a multitude of different rates or processes are modelled simultaneously. While each individual process may be modelled using a stochastic volatility model, additional challenges appear when considering a multitude of different processes simultaneous, which may all be correlated with each other. Due to the large amount of available degrees of freedom, these type of models are very flexible and accurate. On the other hand however, the calibration and parametrization is also more demanding, since much more data is necessary to determine all relevant parameters. For an extensive discussion of these models we refer to Brace et al. [1997], Brigo and Mercurio [2006], Andersen and Piterbarg [2010b]. This concludes our summary of prominent financial market models at this point and we continue by introducing the techniques used in our analysis.

2.4 Perturbation Theory

In this section, we provide an introduction into singular perturbation techniques and outline the applicability to our research. Since a full discussion of this topic is far beyond the scope of this thesis, we refer to Kevorkian and Cole [1996], Hunter [2004], Verhulst [2005], Holmes [2013], for a more thorough discussion. To illustrate the relevant principles, we focus on explicit examples and restrict the general framework to the one-dimensional real space. We start by discussing general perturbation theory, consider singular perturbation theory afterwards and conclude by outlining the application to our work.

2.4.1 General Perturbation Problems

As described in Hunter [2004] perturbation problems can be very generally characterized as problems of the form

$$P^{\varepsilon}(x) = 0 \tag{2.4.1}$$

which depend on a small, real-valued parameter ε. This parameter ε can be seen as a perturbation of the unperturbated problem where $\varepsilon = 0$. The unperturbated problem usually provides a simplified version of the original problem (2.4.1) and may even be explicitly solvable. As summarized in Hunter [2004] 'the aim of perturbation theory is to determine the behaviour of the solution $x = x^{\varepsilon}$ as $\varepsilon \to 0$'. To achieve this goal, we derive a solution which solves the problem up to a small error where the accuracy is dependent on the parameter ε. To properly define this solution, called the asymptotic solution, we recall the fundamental definitions necessary for the consideration. We are referring to the definitions provided in Hunter [2004, Chapter 2].

Definition 2.4.1.

i) **Order symbols:** Let $f, g : \mathbb{R}\backslash 0 \to \mathbb{R}$ be real valued functions. We say that $f = \mathcal{O}(g)$ as $\varepsilon \to 0$ if there exist constants $C > 0$ and $r > 0$ such that for all $0 < \varepsilon < r$ we have

$$|f(\varepsilon)| \leq C|g(\varepsilon)|.$$

Furthermore, we say that $f = o(g)$ as $\varepsilon \to 0$ if for every $\delta > 0$ there is an $\varepsilon_0 > 0$ such that for all $0 < \varepsilon \leq \varepsilon_0$ we have

$$|f(\varepsilon)| \leq \delta|g(\varepsilon)|.$$

ii) **Asymptotic sequence:** A sequence of functions $\{\varphi_n\}_{n\geq 0} : \mathbb{R}\backslash 0 \to \mathbb{R}$ is an asymptotic sequence as $\varepsilon \to 0$ if for each n we have

$$\varphi_{n+1} = o(\varphi_n) \quad \text{as } \varepsilon \to 0.$$

iii) **Asymptotic expansion:** If $f : \mathbb{R}\backslash 0 \to \mathbb{R}$ is a function, $\{a_n\}_{n\geq 0}$ a real-valued sequence of coefficients and $\{\varphi_n\}_{n\geq 0}$ an asymptotic sequence we call

$$\sum_{n=0}^{\infty} a_n \varphi_n$$

the asymptotic expansion of f with respect to $\{\varphi_n\}_{n\geq 0}$ as $\varepsilon \to 0$ if for all $N \geq 0$ we have

$$f(\varepsilon) - \sum_{n=0}^{N} a_n \varphi_n(\varepsilon) = o(\varphi_N) \quad \text{as } \varepsilon \to 0.$$

Given these definitions, we can come back to the problem (2.4.1) and slightly modify it to derive an approximating problem of the form

$$P^\varepsilon(x_N^\varepsilon) = \mathcal{O}(\varepsilon^{N+1})$$

with $N \in \mathbb{N}$. This new problem is an approximation of the original problem up to a small error and provides us with an asymptotic solution which may be more accessible to derive. If we can further show that this asymptotic solution x_N^ε is close to the original solution x^ε of (2.4.1), in the sense that

$$x^\varepsilon = x_N^\varepsilon + \mathcal{O}(\varepsilon^{N+1}),$$

we can use the asymptotic solution as an admissible approximation to the original problem.

These kind of perturbation problems can be observed in a multitude of different mathematical and practical areas. These may be algebraic equations, approximations of integral functions, or the application towards differential equations. We refer to Kevorkian and Cole [1996], Hunter [2004], Verhulst [2005], Holmes [2013] for a detailed discussion of these examples and the various techniques applicable in these areas. As nicely summarized by Verhulst [2005], perturbation theory can be seen as 'the theory of approximating solutions of mathematical problems'. For our work we are mainly interested in an application towards differential equations in the category of singular perturbation problems.

2.4.2 Singular Perturbation Problems

In our framework we consider the behaviour when $\varepsilon \to 0$. Roughly speaking, we can now differentiate our problem by considering the unperturbated problem where $\varepsilon = 0$. As stated in Hunter [2004] a 'regular perturbation problem is one for which the perturbed problem for small, non-zero values of ε is qualitatively the same as the unperturbed problem for $\varepsilon = 0$'. This provides a rather imprecise distinction between regular and singular perturbation problems, but should already provide some intuition on the differences. To make the distinction precise, let us consider the definition of regular given in Verhulst [2005, Chapter 4.1].

Definition 2.4.2. Consider a function $\phi_\varepsilon(\cdot)$ defined on $D \subset \mathbb{R}$; an asymptotic expansion for $\phi_\varepsilon(\cdot)$ will be called regular if it takes the form

$$\phi_\varepsilon(x) = \sum_{n=0}^{m} \phi_n(x)\delta_n(\varepsilon) + \mathcal{O}(\delta_{m+1})$$

with $\{\delta_n(\varepsilon)\}_{n\geq 0}$ being an asymptotic sequence and $\{\phi_n(x)\}_{n\geq 0}$ some coefficient functions on D.

We will call a problem singular, if it is not regular. To demonstrate from where this definition originates, we illustrate two examples from Verhulst [2005]. The following examples correspond to Example 1.2 and Example 1.4 of Verhulst [2005].

Example 2.4.3. For our first example, we consider a function $\phi_\varepsilon(\cdot)$ on $[0, 1]$ which is given as the solution to the initial value problem

$$\begin{cases} \partial_x \phi(x) + \varepsilon\phi(x) = \cos(x) \\ \phi(0) = 0 \end{cases} \tag{2.4.2}$$

To solve this problem, we start with the unperturbated problem when $\varepsilon = 0$. In this case the differential equation can be explicitly solved and we arrive at the solution

$$\phi_0(x) = \sin(x).$$

To consider the case where $\varepsilon > 0$, let us assume that the solution is given by a formal asymptotic expansion, see Definition 2.4.1, of the form

$$\phi_\varepsilon(x) = \sum_{n=0}^{\infty} \phi_n(x)\varepsilon^n.$$

Substituting this expansion into the differential equation (2.4.2), we can set up differential equations for the individual terms. In the case of $n = 1$, we can again explicitly solve the equation to derive an expression of the form

$$\phi_\varepsilon(x) = \sin(x) + \varepsilon(\cos(x) - 1) + \mathcal{O}(\varepsilon^2).$$

This provides us with an asymptotic solution to order $\mathcal{O}(\varepsilon^2)$ which is constructed as the solution of the unperturbated problem including higher order correction terms to take the existing perturbation into account. This asymptotic expansion satisfies the conditions of Definition 2.4.2 and is therefore an example for a regular perturbation problem.

Example 2.4.4. For the next example, let us slightly modify the differential equation of (2.4.2) and move the ε dependence to the other term. In this case we consider a function $\phi_\varepsilon(\cdot)$ on $[0,1]$ which is given as the solution to the initial value problem

$$\begin{cases} \varepsilon\partial_x\phi(x) + \phi(x) = \cos(x) \\ \phi(0) = 0 \end{cases} \tag{2.4.3}$$

Even though these problems look similar at first glance, they are governed by generally different behaviours. To demonstrate this, let us again start by considering the unperturbated problem where $\varepsilon = 0$. This reduces the problem to

$$\begin{cases} \phi_0(x) = \cos(x) \\ \phi_0(0) = 0. \end{cases} \tag{2.4.4}$$

As we can observe, this problem does not permit a solution since the initial value of $\phi_0(0) = 0$ is not compatible with the function $\phi_0(x) = \cos(x)$. In the informal definition of regular perturbation problems, this would provide a qualitatively different behaviour and gives the first impression that this problem is not regular anymore. Therefore, let us proceed and again consider a formal expansion of the form

$$\phi_\varepsilon(x) = \sum_{n=0}^{\infty} \phi_n(x)\varepsilon^n.$$

Inserting this into the problem yields

$$\phi_0(x) - \cos(x) + \sum_{n=1}^{\infty} \varepsilon^n \Big(\phi_n(x) + \partial_x\phi_{n-1}(x) \Big) = 0,$$

which gives an asymptotic expression of the form

$$\phi_\varepsilon(x) = \cos(x) + \varepsilon\sin(x) - \varepsilon^2\cos(x) + \mathcal{O}(\varepsilon^3).$$

Here we observe that this asymptotic expansion is again given by the solution of the unperturbated problem, corrected through higher order terms. However, as in the case of the unperturbated problem, this solution is again not able to satisfy the initial condition. The problem in this example is that the order of the unperturbated problem is reduced by one in comparison to the original problem. Therefore, it is not able to capture the initial condition and properly approximate the solution in a neighbourhood of $x = 0$, where the desired solution is quickly varying to satisfy the initial condition. This provides, in some sense, a singularity at this

point which is not captured by the formal expansion. Outside the neighbourhood of $x = 0$ the formal expansion does, however, provide a valid approximation. As shown in Verhulst [2005] to deduce a proper asymptotic solution, we would need to consider an expansion of the form

$$\phi_\varepsilon(x) = \sum_{n=0}^{\infty} \phi_n(x, \varepsilon)\delta_n(\varepsilon).$$

Therefore, this problem does not satisfy the conditions of Definition 2.4.2 and falls into the category of singular perturbation problems.

2.4.3 Singular Perturbation Techniques

As demonstrated in Example (2.4.3), an ansatz using a regular formal expansion is not sufficient in all situations. Therefore, we have to consider how it is possible to derive asymptotic solutions to perturbation problems which are singular. Many different techniques to derive such solutions were developed, including matched asymptotic expansions or multiple scales to mention a few prominent techniques. Again a discussion of these approaches is far above the scope of this work and we refer to the literature provided by Kevorkian and Cole [1996], Hunter [2004], Verhulst [2005], Holmes [2013] for an in-depth discussion in full detail.

Instead, we provide a general understanding of how it is possible to handle the singularity using local analysis and derive an asymptotic solution in the domain of this singularity. For this we introduce the concept of distinguished limits, also called significant degenerations. For the purpose of illustration, we restrict our definition to the one-dimensional case provided in Verhulst [2005, Chapter 4.3] and regard a boundary layer point $x = x_0$ as the area of interest. We are well aware that this definition may be too simple in some cases, it is however sufficient to present the general concept.

Definition 2.4.5. Consider an operator L_ε expressed in the variable x, near the boundary layer point x_0 and the operator L_ε^* resulting from expressing L_ε in terms of local variables of the form

$$\xi = \frac{x - x_0}{\delta(\varepsilon)}$$

near x_0 translated through a function $\delta(\cdot)$. L_0^* is called a significant degeneration of L_ε if, by taking the formal limit as $\varepsilon \to 0$ of L_ε^*, the resulting operator contains the formal limits for all other choices of local variables. Put in a different way, a significant degeneration implies a choice of the local variable ξ such that the corresponding operator contains as much information as possible when $\varepsilon \to 0$.

This definition is quite abstract and we will provide explicit examples in this section to make it tangible. For this let us consider the local variable ξ a little bit closer. Suppose that near the point x_0 the boundary area is characterized in size by an order function $\delta(\varepsilon)$ characterizing the area of interest. In the limit when $\varepsilon \to 0$ this area becomes smaller and smaller. Therefore we 'stretch' the variable x by a rescaling to the local variable ξ. To continue the local analysis, we can now proceed and apply a regular expansion in terms of the new variable to deduce an analytic solution valid in the neighbourhood of x_0. To better illustrate this approach, we consider the example presented in Holmes [2013, Section 2.2].

Example 2.4.6. Let us consider a function $\phi_\varepsilon(\cdot)$ on $[0,1]$ which is given as the solution to the initial value problem

$$\begin{cases} \varepsilon\partial_{xx}\phi(x) + 2\partial_x\phi(x) + 2\phi(x) = 0 \\ \phi(0) = 0 \\ \phi(1) = 1. \end{cases} \tag{2.4.5}$$

This induces an operator of the form

$$L_\varepsilon(y(x)) = \Big(\varepsilon\partial_{xx} + 2\partial_x + 2\Big)y(x)$$

to characterize the problem. Applying a regular expansion of the form

$$\phi_\varepsilon(x) = \sum_{n=0}^{\infty} \phi_n(x)\varepsilon^n,$$

we can deduce that the term $\phi_0(x)$ has to solve

$$\begin{cases} \partial_x\phi_0(x) + \phi_0(x) = 0 \\ \phi_0(0) = 0 \\ \phi_0(1) = 1. \end{cases}$$

A general solution valid in the inner domain can be characterized as

$$\phi_0(x) = ae^{-x} \tag{2.4.6}$$

with some general constant a. We observe that this solution is not capable of fitting both initial conditions as the boundary value at 0 would impose a trivial solution. Thus, the problem falls into the category of singular perturbation problems. Therefore, to adequately approximate the behaviour at the boundary layer

placed around $x_0 = 0$, we require a different modelling approach. For this let us consider the local variable

$$\xi = \frac{x}{\delta(\varepsilon)} = \frac{x}{\varepsilon^\alpha},$$

where the constant $\alpha > 0$ still has to be determined. This changes the operator L_ε in terms of the variable x into the operator L_ε^* in terms of the variable ξ given by

$$L_\varepsilon^*(\phi^*(\xi)) = \left(\frac{1}{\varepsilon^{2\alpha-1}}\partial_{\xi\xi} + 2\frac{1}{\varepsilon^\alpha}\partial_\xi + 2\right)\phi^*(\xi).$$

Depending on the value of α we have the following formal limits when $\varepsilon \to 0$:

i) For $\alpha < 1$ we have $L_0^*(\phi^*(\xi)) = 2\partial_\xi\phi^*(\xi)$.

ii) For $\alpha = 1$ we have $L_0^*(\phi^*(\xi)) = \partial_{\xi\xi}\phi^*(\xi) + 2\partial_\xi\phi^*(\xi)$.

iii) For $\alpha > 1$ we have $L_0^*(\phi^*(\xi)) = \partial_{\xi\xi}\phi^*(\xi)$.

To achieve the significant degeneration this yields to the choice of $\alpha = 1$. Fixing this value of $\alpha = 1$, we can proceed and apply a regular extension in the new local variable ξ of the form

$$\phi_\varepsilon^*(\xi) = \sum_{n=0}^{\infty} \phi_n^*(\xi)\varepsilon^n,$$

and deduce that the term $\phi_0^*(\cdot)$ has to solve

$$\begin{cases} \partial_{\xi\xi}\phi_0^*(\xi) + 2\partial_\xi\phi_0^*(\xi) = 0 \\ \phi_0^*(0) = 0. \end{cases}$$

A general solution valid in the inner domain can now be characterized as

$$\phi_0^*(\xi) = A(1 - e^{-2\xi}) \tag{2.4.7}$$

with some general constant A. This provides us with a second asymptotic solution which is valid in the neighbourhood of $x = 0$. In the language of Holmes [2013] this is called the inner solution. In contrast to the outer solution (2.4.6), this solution does not claim to be a good approximation outside the neighbourhood of $x = 0$. To derive an asymptotic solution which is valid over the whole domain, it is possible to match both solutions as further demonstrated in Holmes [2013]. For our purpose the derivation of the inner solution using the distinguished limit is sufficient at this point.

2.4.4 Application to our Research

Having established a general foundation of singular perturbation theory, we provide an outline of its application to our work. In our set-up we investigate stochastic volatility models such as the SABR or ZABR models, see section 2.3.3.2 and 2.3.3.3. From a practical perspective we observe that the volatility as well as the volatility of volatility are in most applications rather small, consider e.g. the numerical applications presented in Andreasen and Huge [2011b] or In 't Hout and Foulon [2010]. Therefore, it is reasonable to introduce the parameter ε into the equations, capturing the small volatility regime and deduce an asymptotic solution in this domain. For this we use the tools of singular perturbation theory presented in Section 2.4.3. This includes various non-trivial changes of variables to analyse a distinguished limit. Expressed in these new variables we can consider a formal expansion and perform a leading order analysis to arrive at an explicit solution for the first term of the expansion, a similar approach is performed in the example of (2.4.5). Even though this explicit solution is not our desired approximation, it allows us to deduce further information about the other coefficients and in the end allows us to simplify our original differential equation. The details of these derivations are presented in the sections 3.5 and 5.8.

2.5 Numerical Schemes

Having established all theoretical background, we now focus on practical applicability and study how these concepts can be implemented using numerical schemes. To motivate our approach, let us consider a one dimension local volatility model characterized by its projected volatility function $\vartheta(\cdot,\cdot)$ of the form

$$\left\{ dF_t = \vartheta(t, F_t)\, dW_t, \qquad F_0 = f. \right. \tag{2.5.1}$$

As outlined in the previous sections, one of the most important features for our financial market models is the capacity to price instruments. Let us denote the price of a European call option with maturity T and strike K by $C(T,K)$. Using the results of Section 2.1.3 we can directly derive the following partial differential equations to compute model prices.

(i) **Kolmogorov backward equation:** Using the classical Kolmogorov backward equation, see Section 2.1.3, we can deduce a natural representation of

the price through a PDE of the form

$$\begin{cases} \partial_t C(t,F) = -\frac{1}{2}\vartheta(t,F)^2 \partial_{FF} C(t,F) \\ C(T,F) = (F-K)^+ \end{cases} \tag{2.5.2}$$

The terminal condition is often found to be placed at time 0 and instead of the time itself, the time to maturity is considered then. The corresponding PDE formulation is given by

$$\begin{cases} \partial_t C(t,F) = \frac{1}{2}\vartheta(t,F)^2 \partial_{FF} C(t,F) \\ C(0,F) = (F-K)^+ \end{cases} \tag{2.5.3}$$

(ii) **Kolmogorov forward equation:** Using the Kolmogorov forward equation, see Section 2.1.3, we arrive at another natural representation of the SDE through a PDE. In contrast to the backward equation this approach does not solve the price of the instrument directly, but instead evolves the underlying probability density function $Q(t,F)$ forward in time. This yields a PDE of the form

$$\begin{cases} \partial_t Q(t,F) = \frac{1}{2}\partial_{FF}\left[\vartheta(t,F)^2 Q(t,F)\right] \\ Q(0,F) = \delta(F-f) \end{cases} \tag{2.5.4}$$

While these two differential equations differ a lot on the first glance, they do fall into a common model set-up, allowing for a generalized characterization of the form

$$\partial_t U(t,F) = \mathcal{L} U(t,F), \tag{2.5.5}$$

where U_t denotes the function, e.g. the call price or the density function, and $\mathcal{L}$ denotes the corresponding operator, e.g. $\mathcal{L} = \frac{1}{2}\vartheta(t,F)^2\partial_{FF}$ in the case of (2.5.3).

To solve such a PDE, finite difference methods are a commonly used strategy. Again there is a multitude of different literature available on this topic and this section should provide a condensed summary of the topic to introduced the relevant numerical schemes used for our computations. We use Thomas [1995], Evans et al. [2000], Duffy [2006], Andersen and Piterbarg [2010a] as the underlying literature and start by considering the one-dimensional set-up with a special focus on certain forms for the projected volatility function $\vartheta(\cdot,\cdot)$. These techniques will be the favoured approach for our studies. To allow a comparison to other techniques, we also briefly introduce the concept of two-dimensional finite difference methods as well as Monte Carlo techniques.

2.5.1 One-Dimensional Finite Difference Schemes

The core idea behind finite difference schemes is to approximate the derivatives using simpler differences on a finite numerical grid. In a practical application there may be many possible approximations which can be made to achieve this goal, and to derive the best possible results, the numerical set-up should always consider the underlying problem at hand. To demonstrate the general principle we rely on Andersen and Piterbarg [2010a, Chapter 2] in this section and we start by considering a second-order operator of the form

$$\mathcal{L}U(t,x) = \mu(t,x)\partial_x U(t,x) + \frac{1}{2}\sigma^2(t,x)\partial_{xx}U(t,x). \tag{2.5.6}$$

This operator represents a special case of the convection-diffusion equation considered in Duffy [2006, Chapter 8] and fits into our considerations by providing a natural connection towards the Kolmogorov backward equation through the operator given in Definition 2.1.5.

Furthermore, we now define a numerical grid $\{t_i\}_{0\leq i\leq n} \times \{x_j\}_{0\leq j\leq m+1}$ as a discretization for a domain $[0,T]\times[F_{min},F_{max}]$. The choice of this grid has a major impact on the quality of the scheme and has to be carefully chosen and possibly tailored to the problem at hand. In particular, the truncation of the domain, i.e. an adequate choice of the bounds F_{min} and F_{max}, as well as the construction of the grid points has to be taken into account. For simplicity, let us assume a uniform grid in both dimensions with a spacing given by h_x respectively h_t. For a general theory based on more advanced grids we refer to Andersen and Piterbarg [2010a, Chapter 2.4] and the literature therein. We do not discuss the truncation of the domain in general at this point, an explicit example of how a possible truncation is applied is however provided in the numerical algorithm presented in Section 2.5.3. Having defined the underlying grid, we now proceed to approximate the derivatives appearing in (2.5.6) using finite differences. We use the definitions provided in Duffy [2006].

Definition 2.5.1. Given a grid $\{x_j\}_{0\leq j\leq m+1}$ we define the following finite differences δ_x and δ_{xx} on the inner grid points $\{x_j\}_{1\leq j\leq m}$ as approximations for the derivatives of a function $f:\{x_j\}_{0\leq j\leq m+1}\to\mathbb{R}$.

i) **Centred differences:**

$$\delta_x f(x_j) = \frac{1}{2h_x}\big(f(x_{j+1}) - f(x_{j-1})\big)$$

ii) **Forward differences:**

$$\delta_x f(x_j) = \frac{1}{h_x}\big(f(x_{j+1}) - f(x_j)\big)$$

iii) **Backward differences:**

$$\delta_x f(x_j) = \frac{1}{h_x}\big(f(x_j) - f(x_{j-1})\big)$$

iv) **Centred second-order differences:**

$$\delta_{xx} f(x_j) = \frac{1}{h_x^2}\big(f(x_{j+1}) - 2f(x_j) + f(x_{j-1})\big)$$

Again, we observe that multiple approximations are possible and the choice of the right definition can yield to a major improvement in performance. Using a Taylor expansion, it is possible to show that the centred differences are second-order approximations while the other two are only of first-order, see Duffy [2006] or Andersen and Piterbarg [2010a, Lemma 2.2]. The better convergence property would suggest to only use centred differences. Depending on the regularity of the underlying function it may however be better to use other variants or even apply different variants in different domains using e.g. upwinding techniques, see Andersen and Piterbarg [2010a, Chapter 2.6.1], In 't Hout and Foulon [2010], Andreasen and Huge [2011a]. Using centred finite differences, this allows us to construct the following second-order approximation of the operator $\mathcal{L}$, see e.g. Andersen and Piterbarg [2010a].

Definition 2.5.2. Given an operator $\mathcal{L}$ as in (2.5.6) as well as a discretization grid $\{x_j\}_{0\leq j\leq m+1}$, we define a tri-diagonal matrix A_t with $1 \leq i, j \leq m$ as

$$\big(A_t\big)_{i,j} = \begin{cases} l_i(t) \text{ if } j = i-1 \\ c_i(t) \text{ if } j = i \\ u_i(t) \text{ if } j = i+1 \\ 0 \text{ otherwise} \end{cases},$$

with diagonal functions

$$\begin{aligned} l_i(t) &= -\frac{1}{2h_x}\mu(t, x_i) + \frac{1}{2h_x^2}\sigma^2(t, x_i) \\ c_i(t) &= -\frac{1}{h_x^2}\sigma^2(t, x_i) \\ u_i(t) &= \frac{1}{2h_x}\mu(t, x_i) + \frac{1}{2h_x^2}\sigma^2(t, x_i). \end{aligned}$$

The matrix A_t provides a second-order approximation for the operator $\mathcal{L}$ in the sense that, restricted to the inner grid $\{x_j\}_{1\leq j\leq m}$, we have

$$\mathcal{L}U(t,x_j) = A_t U(t,x_j) + \mathcal{O}(h_x^2).$$

The outer points x_0 and x_{m+1} represent the boundaries and are governed by boundary conditions captured in an additional term Ω_t. These conditions can be explicitly given through the underlying problem in the form of a payoff function, or can be derived from the inner grid using additional modelling assumptions. As demonstrated in e.g. Andersen and Piterbarg [2010a, Chapter 2.2.2] using the assumption that the function is linear in x at the boundary, the value at the point x_{m+1} can be deduced as

$$U(t,x_{m+1}) = 2U(t,x_m) - U(t,x_{m-1}).$$

In particular, it is then possible to incorporate the boundary conditions into the matrix A_t by extending it adequately to be valid on the whole grid $\{x_j\}_{0\leq j\leq m+1}$. We do not go into further details on boundary conditions at this point, we do, however, stress that proper boundary conditions are crucial for a good numerical scheme and we refer to Thomas [1995], Evans et al. [2000], Duffy [2006], Andersen and Piterbarg [2010a]. Restricted to the full grid $\{x_j\}_{0\leq j\leq m+1}$, this yields an approximation of the operator $\mathcal{L}$ of the form

$$\mathcal{L}U(t,x) \approx A_t U(t,x) + \Omega_t. \tag{2.5.7}$$

In our considerations, we mostly integrate the boundary conditions into the matrix A_t resulting in $\Omega_t = 0$. Including a time discretization, again using finite differences, we can now state a full finite difference scheme. One of the most prominent finite difference schemes is the θ-scheme. We present the version given in Andersen and Piterbarg [2010a, Proposition 2.2.2].

Proposition 2.5.3. *On the grid $\{x_j\}_{0\leq j\leq m+1}$ the solution $U(t_i)$ respectively $U(t_{i+1})$ of (2.5.5) at time t_i and t_{i+1} is characterized through the equation*

$$(\mathbb{1} + \theta h_t A_{t_i^{i+1}(\theta)})U(t_i) + \theta h_t \Omega_{t_i} = (\mathbb{1} - (1-\theta)h_t A_{t_i^{i+1}})U(t_{i+1}) - (1-\theta)h_t \Omega_{t_{i+1}} + e_i^{i+1}$$

with time

$$t_i^{i+1}(\theta) = (1-\theta)t_{i+1} + \theta t_i$$

and error term

$$e_i^{i+1} = h_t \mathcal{O}(h_x^2) + \mathbb{1}_{\theta\neq\frac{1}{2}}\mathcal{O}(h_t^2) + \mathcal{O}(h_t^3).$$

This scheme allows us to deduce the values of the function $U(\cdot,\cdot)$ at time t_{i+1} based on the values at time t_i. Given an initial condition, this allows to solve the PDE iteratively and in turn solve the corresponding initial value problem.

The scheme provides an additional degree of freedom through the parameter θ. Some of the most prominent choices are the fully implicit scheme with $\theta = 1$, the fully explicit scheme with $\theta = 0$ or the Crank-Nicolson scheme with $\theta = \frac{1}{2}$. To conclude this excursion on the basics of finite difference schemes, we address which properties can be used to analyse the quality of the numerical scheme. We focus on consistency, convergence and stability as defined in Thomas [1995].

2.5.1.1 Consistency

The first property we discuss is the consistency of the numerical scheme. This can be seen as a property to describe if the chosen discretization is consistent with the original problem. Considering e.g. the θ-scheme of Proposition 2.5.3, we observe that the scheme includes some truncation error e_i^{i+1}. This error appears when discretizing the derivatives to apply a single time step. We call the numerical scheme consistent, if this truncation error vanishes as the grid sizes become smaller. This ensures that applying our discretization to the true solution of the underlying problem is consistent with the PDE formulation and the truncation error is only induced by the chosen discretization and not through a structural discrepancy.

2.5.1.2 Convergence

Convergence captures one of the key properties desired in a numerical scheme, namely that the scheme is in fact approximating the true solution of the problem. Using Thomas [1995, Definition 2.2.2] as the underlying reference, consider a time t at which we want to deduce a solution to the underlying initial value problem. We deduce by $A_t^{n,m}U(t,\cdot)^{n,m}$ the scheme approximating $\mathcal{L}U(t,\cdot)$ on a grid with n time steps and $m+1$ space steps. The scheme is called convergent at time t if

$$\|U(t,\cdot)^{n,m} - U(t,\cdot)\| \to 0$$

as $h_x \to 0$ and $h_t \to 0$ for some norm $\|\cdot\|$ and a grid spacing such that $(n+1)h_t \to t$. This states that if our approximating grid-spacing becomes smaller, the approximating solution in fact converges to the correct solution. In comparison to consistency, convergence considers the evolution of the solution along the time grid and the corresponding iterative application of a single time step as described e.g. in Proposition 2.5.3. Furthermore, it is also possible to capture the speed at which

the scheme converges. Considering the θ-scheme as an example Andersen and Piterbarg [2010a, Proposition 2.2.3] states that the error of the scheme compared to the exact solution is of order

$$\mathcal{O}(h_x^2) + \mathbb{1}_{\theta \neq \frac{1}{2}} \mathcal{O}(h_t) + \mathcal{O}(h_t^2).$$

This error can be computed based on the truncation error available at each time step and the corresponding iterative accumulation along the time evolution. Hence, we can observe that all θ-schemes are second-order convergent in space and at least first order-order convergent in time. In the special case of the Crank-Nicolson scheme even second-order time convergence is achieved.

2.5.1.3 Stability

Lastly, we also consider the stability of the numerical scheme. For illustration, stability can be seen as a way to control the errors along the progression of time. Again, considering the θ-scheme as an example, we express a single time step from time t_{i+1} to t_i, starting from a terminal condition, as

$$U(t_i) = B_i^{i+1} U(t_{i+1}).$$

Iterating this multiple times, we arrive at an expression of the form

$$U(t_k) = B_k^n U(t_n)$$

to evolve a solution at time t_n to t_k. Here B_k^n denotes the composition of the functions available in each individual time step. We call the numerical scheme stable if the initial respectively terminal condition is bounded and there exists a constant $C > 0$ such that for all $0 \leq k < n$ we have

$$\|B_k^n\| \leq C.$$

In most cases, this condition is not easy to prove and other techniques have to be used to derive equivalent results. One prominent approach is the von Neumann analysis based on Fourier series. For further details we refer to Thomas [1995], Evans et al. [2000], Duffy [2006], Andersen and Piterbarg [2010a]. Using these techniques as done in e.g. Andersen and Piterbarg [2010a], it is possible to show that the θ-scheme is always stable if $\theta \in [\frac{1}{2}, 1]$. This includes the implicit as well as the Crank-Nicolson scheme. The explicit scheme, however, is only conditionally stable, meaning that we need to impose additional conditions to guarantee the

stability. This may impose some severe restriction on the grid size, enforcing unnecessary large grids and increasing the computational time.

To conclude, we present a powerful result connecting all these properties with each other. The result is called the *Lax equivalence theorem* and we refer to the the version of Thomas [1995, Theorem 2.5.1]. The statement is formulated for well-posed problems, meaning that the PDE has a unique solution that is continuously dependent on the input data.

Theorem 2.5.4. *A consistent, two level difference scheme for a well-posed linear initial-value problem is convergent if and only if it is stable.*

2.5.2 Special Local Volatility Models

Having established the foundation to the numerics behind the one-dimensional finite difference schemes, we come back to the initial example considered in Section 2.5. As we will see in our studies performed in chapters 3-5 , it is often possible to reduce the underlying problems we study to a form as in (2.5.1). Here the function $\vartheta(\cdot,\cdot)$ is governed by a specific underlying structure of the form

$$\vartheta(t,F) = D(F)E(t,F). \tag{2.5.8}$$

Hence, we continue by analysing the corresponding differential equations derived for the models of Section 2.3 and establish the necessary expressions to apply a one-dimensional finite difference scheme. Considering the SABR model of Section 2.3.3.2 it was shown in Hagan et al. [2014] that it is possible to characterize the model using an approximation through (2.5.8) with an explicit representation of the form

$$\begin{aligned} D(F) &= \sqrt{\alpha^2 C(F)^2\big(1 + 2bz(F) + cz(F)^2\big)} \\ E(t,F) &= \sqrt{e^{G(t)}} \end{aligned}$$

where the coefficients are given by

$$b = \rho\nu\alpha^{-1}, \quad c = \nu^2\alpha^{-2}, \quad z(F) = \int_f^F \frac{1}{C(u)}\,du,$$

and the function

$$G(t) = t\rho\nu\alpha C'(f).$$

The representation of the projected volatility function through an expression of the form (2.5.8) has the advantage that a separation between the major part of the space-dependence and the time-dependence is performed. In this set-up it is rather natural to consider a transformation of the form

$$y(F) = \int_f^F \frac{1}{D(u)}\, du \tag{2.5.9}$$

to create a more space-independent variable. From a numerical perspective, this makes it easier to generate an adequate grid in the space dimension. As demonstrated in Hagan [2015] in the case of the SABR model, the transformation between the variables $y(\cdot)$ and $z(\cdot)$ can be explicitly performed through the functions [1]

$$Y_{SABR}(z) = \frac{1}{\nu} \log\left(\frac{\sqrt{1-\rho^2+(\rho+\nu z\alpha^{-1})^2}+\rho+\nu z\alpha^{-1}}{1+\rho} \right) \tag{2.5.10}$$

$$Z_{SABR}(y) = \frac{\alpha}{\nu}\left(\sinh(\nu y) + \rho\left(\cosh(\nu y)-1\right)\right). \tag{2.5.11}$$

Remark 2.5.5. Notice that the characterization of the SABR model is derived as an approximation and therefore other parametrizations of the form (2.5.8) are possible. Considering e.g. the ZABR model of 2.3.3.3 as an example, it was shown in Andreasen and Huge [2011b] that it is possible to characterize the ZABR model using an approximation of the form (2.5.8) where the function $E(\cdot,\cdot)$ equals 1. In the special case of the SABR model the representation can then again be explicitly given by

$$D(F) = \sqrt{\alpha^2 C(F)^2\big(1+2bz(F)+cz(F)^2\big)}.$$

This characterization is in line with the one of Balland and Tran [2013], but loses accuracy due to the missing time dependencies caused by the vanishing function $G(\cdot)$.

2.5.3 Efficient One-Dimensional SABR Scheme

To conclude our discussion of one-dimensional finite differences schemes, we provide the complete numerical scheme which we use when solving the SABR model using partial differential equations. As will be demonstrated in Chapter 3, it is possible to adequately approximate a multitude of different models using the SABR model and hence an accurate solution to the SABR model is fundamental. We apply the numerical scheme provided in Hagan [2015] using an implementation in accordance to Le Floc'h and Kennedy [2017] presented in the following.

[1] To align the notation to the rest of this thesis, the notation for the functions is switched compared to the original notation in Hagan [2015].

2.5.3.1 PDE Formulation

We consider solving the forward equation (2.5.4) to compute the underlying probability density function $Q(t, F)$ of the SABR model. Here the function $\vartheta(\cdot, \cdot)$ appearing in (2.5.4) is characterized by (2.5.8) and the explicit functions of Section 2.5.2. As suggested by Hagan [2015], it is more advantageous to not directly solve (2.5.4) and instead consider the transformation (2.5.9) and express the PDE in terms of the variable y. We define the new density function $\theta(\cdot, \cdot)$ in the variable y as

$$\theta(t, y) = \frac{1}{h} Q(t, F_{SABR}(y)) D(F_{SABR}(y)),$$

with $F_{SABR}(\cdot)$ denoting the inverse transformation from y back to the original variable F. We then arrive at the transformed PDE

$$\begin{cases} \partial_t \theta(t, y) = \frac{1}{2} \partial_y \left[\frac{1}{D(F_{SABR}(y))} \partial_y \left[\vartheta(t, F_{SABR}(y)) \theta(t, y) \right] \right] \\ \theta(0, y) = \delta(y). \end{cases} \tag{2.5.12}$$

This is the initial value problem to which we apply the one-dimensional finite difference scheme.

2.5.3.2 Grid Construction

The first step in the application of a finite difference scheme, lies in the construction of a discretization grid which allows to adequately capture the important features of the initial value problem. Considering the discretization along the time dimension, this can usually be achieved using an equidistant grid starting at the initial time and running to the largest time of interest. Using this methodology, we generate the corresponding equidistant time grid $\{t_i\}_{0 \leq i \leq n}$ as a discretization of the domain $[0, T]$.

For the space domain, our grid construction becomes slightly more advanced. Considering the initial variable F of (2.5.4), the choice of an adequate domain to discretize can become difficult since the domain can heavily depend on the underlying model parameters. Therefore, we construct our grid on the transformed variable y which is much more space-independent then the original variable F. We construct an equidistant grid where the boundaries of the domain are derived using a given number of standard deviations. We apply some adjustments to this methodology to incorporate specified model barriers and extend the domain depending on the considered strikes. Further, the grid is constructed in a way that the initial forward

value f is included as the middle point between two grid points. Following the approach of Hagan [2015] we continue to work with the middle points of the grid to conserve the first moments of the distribution. The details of the algorithm are presented in Algorithm 1 and follow the implementation of Le Floc'h and Kennedy [2017]. In particular, the grid construction in the variable y also has the advantage, that the transformed grid in the variable F is not equidistant any more and includes more points in the areas of interest where more accuracy is needed. The result of this construction is the original grid $\{\tilde{y}_j\}_{1\leq j\leq M}$, the grid $\{y_j\}_{1\leq j\leq M}$ denoting the middle points and the transformed grid middle points $\{F_j\}_{1\leq j\leq M}$ in the original variable F.

Algorithm 1: Adapted grid generation in accordance to Le Floc'h and Kennedy [2017].

Data: Strikes $\{K_i\}_{1\leq i\leq \tilde{M}}$, Barrier d, Maturity T, Number of standard deviations n_{std}

Result: Space grid $\{F_j\}_{1\leq j\leq M}$

Set $y_{min} = -n_{std}\sqrt{T}$, $y_{max} = -y_{min}$; `// Initial domain`

Adapt $y_{min} = \min\{y_{min}, Y_{SABR}(z(-d))\}$; `// Adapt domain to the barrier`

Adapt $y_{max} = \max\{y_{max}, Y_{SABR}(z(\max_{i\leq \tilde{M}} K_i))\}$; `// Adapt domain to the strikes`

Set $h_0 = \frac{1}{M-2}(y_{max} - y_{min})$; `// Initial grid spacing`

Set $j_0 = \left\lfloor \frac{-y_{min}(M-2)}{y_{max}-y_{min}} \right\rfloor$; `// Grid index related to the initial forward` f

Set $h = \frac{-y_{min}}{j_0-\frac{1}{2}}$; `// Grid spacing to place` f `in the middle of the interval`

Generate $\{y_j\}_{1\leq j\leq M}$; `// Equidistant grid from` y_{min} `with step size` h

Adapt $y_j = y_j - \frac{1}{2}h$; `// Shift the grid to consider the middle points`

Transform $\{F_j\}_{1\leq j\leq M}$; `// Transform to original variable`

Adapt $F_1 = 2F_{SABR}(y_{min}) - F_2$, $F_M = 2F_{SABR}(y_{max}) - F_{M-1}$; `// Adapt boundaries`

2.5.3.3 Space Discretization

Following the moment preserving scheme of Hagan [2015], we define the densities in an interval $[\tilde{y}_j, \tilde{y}_{j+1}]$ of the original grid as

$$\theta_{j+1}(t) = \frac{1}{h}\int_{\tilde{y}_j}^{\tilde{y}_{j+1}} \theta(t,y)\,dy.$$

Furthermore, we define a tri-diagonal coefficient matrix A_t similar to that of Definition 2.5.2 through its diagonal functions given by

$$\begin{aligned} l_i(t) &= \frac{1}{2h}\frac{1}{F_i - F_{i-1}}\vartheta(t, F_{i-1}), \\ c_i(t) &= -\frac{1}{2h}\left(\frac{1}{F_{i+1} - F_i} + \frac{1}{F_i - F_{i-1}}\right)\vartheta(t, F_i), \\ u_i(t) &= \frac{1}{2h}\frac{1}{F_{i+1} - F_i}\vartheta(t, F_{i+1}). \end{aligned}$$

Using this coefficient matrix evaluated on the grid $\{F_j\}_{1\leq j\leq M}$ along with the interval densities $\theta(t) = \{\theta_j(t)\}_{1\leq j\leq M}$, we arrive at a space discretization, see Hagan [2015], Le Floc'h and Kennedy [2017], of the form

$$\partial_t\theta(t) = A_t\theta(t). \tag{2.5.13}$$

2.5.3.4 Time Discretization

Given the space discretization of (2.5.13) only a time discretization is missing to establish a full finite difference scheme. For this we apply the Lawson-Swayne scheme of Lawson and Swayne [1977]. The scheme uses two fully implicit discretization steps and an additional extrapolation of those. This makes the scheme second order and unconditionally stable. The algorithm for a single time step at a fixed grid point is provided in Algorithm 2. For possible alternatives as well as a discussion of the advantages and disadvantages, we refer to Le Floc'h and Kennedy [2017].

Algorithm 2: Lawson Swayne scheme in accordance to Lawson and Swayne [1977], Le Floc'h and Kennedy [2017].

Data: Initial value $\theta_j(t_n, y_j)$, step size δ, space discretization matrix $A_{t,j}$ for the grid point y_j

Result: Value $\theta_j(t_{n+1}, y_j)$ at time t_{n+1}

Set $b = 1 - \frac{\sqrt{2}}{2}$; `// Scheme specific coefficient`

Set $\delta_b = \delta b$; `// Step size for the impicit steps`

Set $\theta_j(t_n + \delta_b) = \left(1 - \delta_b A_{t_n+\delta_b,j}\right)^{-1}\theta_j(t_n)$; `// First implicit step`

Set $\theta_j(t_n + 2\delta_b) = \left(1 - \delta_b A_{t_n+2\delta_b,j}\right)^{-1}\theta_j(t_n + \delta_b)$; `// Second implicit step`

Set $\theta_j(t_{n+1}) = (\sqrt{2}+1)\theta_j(t_n + 2\delta_b) - \sqrt{2}\theta_j(t_n + \delta_b)$; `// Extrapolation step`

2.5.3.5 Boundaries

Given the results above, we have established a one-dimensional finite difference scheme for the interior of the domain. What is missing is the behaviour at the boundaries. The initial condition is given, hence we focus on the boundaries induced by the space domain. Following Hagan [2015], we consider accumulating boundaries with a focus on the conservation of the total mass. This has the advantage that, together with the accumulated boundaries, the computed probability density function always equals 1 over the selected domain. The characterization of these accumulated probabilities, denoted by Q^L and Q^R, is given by

$$\partial_t Q^L(t) = \lim_{F \downarrow F_{min}} \partial_F \left[\frac{1}{2} \vartheta(t,F)^2 Q(t,F) \right], Q^L(0) = 0$$

and

$$\partial_t Q^R(t) = -\lim_{F \uparrow F_{max}} \partial_F \left[\frac{1}{2} \vartheta(t,F)^2 Q(t,F) \right], Q^R(0) = 0.$$

Placing the boundaries at the grid points y_1 and y_M we impose the condition that the probability density function approaches 0 when it approaches the boundary. Hence following Hagan [2015], Le Floc'h and Kennedy [2017] we arrive at the boundary conditions

$$\begin{aligned}
\vartheta(t,F_1)\theta_1(t) &= -\vartheta(t,F_2)\theta_2(t), \\
\vartheta(t,F_M)\theta_M(t) &= -\vartheta(t,F_{M-1})\theta_{M-1}(t), \\
\partial_t Q^L(t) &= \frac{1}{F_2 - F_1}\vartheta(t,F_2)\theta_2(t), \\
\partial_t Q^R(t) &= \frac{1}{F_M - F_{M-1}}\vartheta(t,F_{M-1})\theta_{M-1}(t).
\end{aligned}$$

Saving the values of $Q^L(t)$ and $Q^R(t)$ alongside those of $\theta(t)$, the matrix A_t can be adjusted to incorporate all boundary conditions. The time discretization using the Lawson-Swayne scheme can then be performed as before.

2.5.3.6 Call Prices

Using the outlined scheme, we can compute the values $Q^L(T)$, $Q^R(T)$ and $\theta(T)$ to describe our probability distribution for a given time T. We highlight that the

values $\theta(T)$ correspond to a kind of average probability density observed on the intervals and to derive the whole distribution, the different interval sizes have to be considered as well. To conclude this section, we briefly discuss how we can use these computed values to evaluate option prices such as a European call option with maturity T and strike K. Denoting by k the index such that

$$y_{min} + \left(k - 1 - \frac{1}{2}\right)h < y(K) \leq y_{min} + \left(k - \frac{1}{2}\right)h$$

the value of the call is given by

$$\begin{aligned} V_{call} = &\int_{y(K)}^{y_{min}+(k-\frac{1}{2})h} \big(F_{SABR}(y) - K\big)^+ \theta(T, y)\, dy \\ &+ \sum_{j=k+1}^{M-1} \left(F_{SABR}(y_{min} + jh - \frac{1}{2}h) - K\right)^+ h\theta_j(T) \\ &+ \big(F_{SABR}(y_{max}) - K\big)^+ Q^R(T). \end{aligned}$$

The first term needs some additional approximation to evaluate the distribution within a grid interval. This can be done in various ways and we choose the subgridscale model given in Hagan [2015, Section 2.4.1].

2.5.4 Other Numerical Methods

While we mostly aim to derive an approximating PDE formulation which can be solved using a one-dimensional finite difference scheme, there are a multitude of other numerical methods available to consider the models of Section 2.3. In particular, Monte Carlo methods are a flexible and commonly used technique to evaluate SDEs. We use these techniques to put our results into context by generating a benchmark using a Monte Carlo approach. Before introducing the general concept of these techniques, we provide further insight on more advanced numerical PDE methods.

2.5.4.1 Two-Dimensional Finite Difference Schemes

Considering the models of Section 2.3, most are two-dimensional. Hence, applying the Kolmogorov equations of Section 2.1.3 yields a two-dimensional PDE. This two-dimensional PDE can also be solved directly using two-dimensional finite difference schemes. To introduce the general idea we consider a finite difference

approximation similar to that of (2.5.7) of the form

$$\mathcal{L}U(t,x) \approx A_t U(t,x)$$

where the operator $\mathcal{L}$ consists of derivatives based on two variables and their mixed terms. Assuming no mixed terms for the moment, which corresponds to a correlation of 0 in the models of Section 2.3, the so called Alternating Direction Implicit (ADI) method can be used to solve the problem. Using Andersen and Piterbarg [2010a] as the underlying reference the idea behind the technique is to split the original operator into two separate operators depending only on one variable each. For the matrix A_t this yields a decomposition of the form

$$A_t = A_t^{(1)} + A_t^{(2)},$$

where the matrices $A_t^{(i)}$ only depend on a single variable. Considering a Crank-Nicolson scheme this then yields to an approximation step of the form

$$\left(\mathbb{1} - \frac{1}{2}h_t\left(A_t^{(1)} + A_t^{(2)}\right)\right) \approx \left(\mathbb{1} - \frac{1}{2}h_t A_t^{(1)}\right)\left(\mathbb{1} - \frac{1}{2}h_t A_t^{(2)}\right).$$

Since the variables are now applied successively, each single application can be handled as in the one-dimensional case. In the case where mixed terms are available, such a simple decomposition is not possible any more and advanced ADI schemes have to be applied. For further details we refer to In 't Hout and Foulon [2010] where an extensive study of ADI schemes is performed in the context of the Heston model.

2.5.4.2 Monte Carlo Methods

While the finite difference schemes produce valid and accurate results, they are also governed by some restrictions. First of all, starting with an SDE, we have to derive a suitable PDE to be able to apply the proposed finite difference schemes. Secondly, as demonstrated in e.g. Andersen and Piterbarg [2010a] the computational complexity of an ADI scheme is exponentially increasing depending on the dimension of the underlying problem. To overcome these restrictions, Monte Carlo methods provide a possible solution. In this section we provide a general intuition how the techniques work and introduce one of the simplest numerical Monte Carlo schemes available, the Euler scheme. For an extensive discussion we refer to Jäckel [2002], Glasserman [2004], Andersen and Piterbarg [2010a].

The principle behind Monte Carlo methods, is to combine the strong law of large numbers, see e.g. Durrett [2010, Theorem 2.4.1], with a discretization procedure to sample variables from the underlying distribution based on the underlying SDE. To formalize this principle, we consider a general Itô process X_t given by an SDE as defined in Definition 2.1.2. Considering the application in a financial framework, Proposition 2.2.8 provides us with the pricing formula for contingent claims. Here the price is expressed as the expectation of the discounted contingent claim, which is dependent on the underlying Itô process, and we can use the strong law of large numbers to approximate the expectation as

$$\mathbb{E}^{\mathbb{Q}^D}\left[\frac{V_T}{D_T}\middle|\mathcal{F}_t\right] \approx \frac{1}{M}\sum_{i=1}^{M}\frac{V_{T,i}}{D_{T,i}}.$$

Here M denotes the number of paths used for the approximation. By the strong law of large numbers we know that the right side of the equation converges to the left side and with a suitable large M, the approximation error becomes as small as desired. The challenges of the technique now lie in an adequate generation of the samples $\frac{V_{T,i}}{D_{T,i}}$ under the measure $\mathbb{Q}^D$ conditioned on $\mathcal{F}_t$. Without loss of generality we can assume that these samples are given through a functional form depending on X_T. This transforms the problem to generate samples of X_T. In some cases the distribution may be explicitly computable and a direct sampling approach can be considered. In the general formulation however, we only have the underlying SDE as a source of information to estimate the process behaviour. To generate these random samples, we apply the Euler scheme. The idea of the scheme is to consider the SDE of (2.1.2) for a small time step from time t_i to $t_i + \delta$ and approximate the coefficients using the time t_i values. This yields to an approximation of the form

$$\begin{aligned} X_{t_i+\delta} &= X_{t_i} + \int_{t_i}^{t_i+\delta} \mu(s, X_s)\, ds + \int_{t_i}^{t_i+\delta} \sigma(s, X_s)\, dW_s \\ &\approx X_{t_i} + \mu(t_i, X_{t_i})\delta + \sigma(t_i, X_{t_i})(W_{t_i+\delta} - W_{t_i}). \end{aligned}$$

The last term can now be explicitly sampled using Gaussian random variables. For the sampling of random numbers we refer to Glasserman [2004, Chapter 2] where a detailed exposition is provided. Given a time discretization, the scheme allows to sample the steps until time T and in turn provides the sample values for X_T. Since the scheme is one of the simplest possible Monte Carlo approximation schemes, there are many other possibilities and improvements available. We refer to Jäckel [2002], Glasserman [2004], Andersen and Piterbarg [2010a] for a complete exposition.

This concludes our summary of the necessary foundations and we continue with our research on generalized stochastic volatility models in the next chapter.

3 Effective Stochastic Volatility

In this chapter we study generalized stochastic volatility models and derive a one-dimensional arbitrage free approximation technique which yields to implied volatility formulas in correspondence to the results presented in Section 2.3.3.2.

To provide some insight into our motivation, let us reconsider the financial market model from a practical perspective. As outlined in Section 2.3.3.2, the implied Black volatility provides a standard way to quote prices of European call options observed on the market. Hence, if we would chose the Black-Scholes-Merton model of Section 2.3.1 as our underlying financial market model, we would observe a constant implied volatility independent of strike or maturity. Looking at actual observable market quotes, however, we can observe that the implied volatility exhibits much more involved structures resulting in various smile and skew behaviours. This discrepancy shows, that more involved financial market models have to be used to adequately capture the observed market behaviour, motivating the introduction of the stochastic volatility (SV) models presented in sections 2.3.3.1-2.3.3.3.

Indeed, we can observe that through the explicit stochastic modelling of the volatility process, also the implied volatility under SV models can take a variety of different shapes allowing to capture the desired smile and skew behaviour. This explains the popularity of SV models in practice. Considering the available possibilities for SV models, a large heterogeneity is present. This can already be observed for the three presented models in Section 2.3. Using an SABR model, we can directly use an explicit implied volatility formula resulting in a very easy implementation from a practical side. When calibrating the model to market data, we would, however, prefer as much control as possible to increase the accuracy of our model. Hence, the ZABR model is preferred, resulting in a numerical PDE evaluation to compute the desired prices. Thus, to switch from an SABR to a ZABR model, we also have to consider a completely different implementation. Taking the Heston model into the consideration, we get another additional implementation through Fourier methods.

Therefore, our motivation for this piece of research is twofold. First, we aim to provide a generalized, arbitrage free, computational framework applicable to a large class of SV models, covering the presented examples. To achieve this, we work with a generalized stochastic volatility model to deduce our approximation techniques. Second, we aim to derive explicit approximation formulas for this generalized framework which are similar to that in Section 2.3.3.2. This allows to easily exchange the desired financial market model. In addition this framework also provides the possibility to actively choose between accuracy and ease of implementation due to the availability of explicit approximation formulas.

The approximation technique is set up to cover a large class of stochastic volatility models which, in particular, also includes the ZABR model, providing a reference example throughout this research. This extends the results of Section 2.3.3.3 by establishing an explicit approximation formula for the implied volatility and in turn allowing a very easy way to implement the model. Due to the generality in which the technique is derived, this also gives rise to new types of stochastic volatility models such as the free-ZABR model or the mean-reverting ZABR model. The technique is based on results of Hagan et al. [2002, 2014, 2015, 2016] and uses singular perturbation techniques to analyse the induced Kolmogorov equations. The core idea behind this analysis can be summarized through the following steps.

1) Assuming a small volatility regime expressed through small model parameter values, we introduce a perturbation parameter ε and aim to approximate the system up to order $\mathcal{O}(\varepsilon^2)$.

2) To achieve this, we analyse certain Kolmogorov backward equations using singular perturbation techniques involving the following steps.

    i) We apply various variable transformations to simplify the equation in order to make the leading order term emerge.

    ii) We perform a leading order analysis which yields to an explicitly solvable heat equation.

    iii) Using the additional information from the leading order analysis, we can approximate the system by dropping terms of order $\mathcal{O}(\varepsilon^3)$ and higher.

3) Considering the Kolmogorov forward equation together with the already derived approximations gives rise to a one-dimensional differential equation to characterize the marginal probability density induced by the model under consideration.

Using this approximating equation for the marginal probability density function to characterize the generalized stochastic volatility model, the model can further

be projected to an SABR model preserving the approximation order of $\mathcal{O}(\varepsilon^2)$. The projected SABR model can then be solved using the algorithm presented in Section 2.5.3 or using the various implied volatility formulas available for the model.

The rest of this chapter presents the publication Felpel et al. [2021] starting from its introduction. This is an Accepted/Original Manuscript of an article published by Taylor & Francis Group in Quantitative Finance on 23.10.2020, available online: https://wwww.tandfonline.com/doi/abs/10.1080/14697688.2020.1814396.

Before starting with the presentation of Felpel et al. [2021], we provide some additional remarks to embed the results into the context of this thesis.

Remark 3.0.1. The publication is slightly modified to fit into a unified presentation in the context of this thesis. The content remains unchanged, the format and the notation are, however, slightly adjusted. Most notably, the notation of the forward process is changed compared to Felpel et al. [2021]. In Felpel et al. [2021] we denote the forward process as $\tilde{F}_t$ whereas here, we use the notation F_t. This adjustment is performed to provide a consistent notation towards the models presented in chapters 4 and 5.

Remark 3.0.2. In the context of the paper Felpel et al. [2021], we sometimes refer to the local volatility function, see e.g Section 3.2. To stay as close as possible to the original publication, we keep this notation. In the context of this thesis, however, this should be understood as the projected volatility function. A local volatility function suggests a local volatility setting in accordance to Dupire [1994], whereas we use a local volatility function for a Bachelier model as depicted in Section 2.3.2.

Remark 3.0.3. In Figure 3.1 as well as Figure 3.6, we present computed distributions. For a better understanding, we want to highlight, that these graphs have to be understood in terms of the scheme presented in Section 2.5.3 depicting the transformed average densities observed on the corresponding intervals.

3.1 Introduction

This paper is concerned with modelling the implied volatility surface for an underlying asset. In our exposition, we consider the dynamics of the forward rate, which simplifies the SDEs considered by ensuring that the asset has no drift. The implied volatility surface is defined in relation to a reference valuation model with an analytic solution for vanilla European call and put options. For the underlying

asset, F, strike value, K, and maturity time, T, the contract pays

$$V_T = \max(F_T - K, 0) \text{ for a call,} \qquad \text{and}$$
$$V_T = \max(K - F_T, 0) \text{ for a put.}$$

Given the maturity, strike and current (time-t_0) price of the underlying, f, the reference formula used to compute the current price, V_{t_0}, must have only one additional degree of freedom, the *implied volatility*, which is specified in order to match the quoted market price, V_{Market}. The standard reference models used are the Black and Bachelier models. Since we shall only consider examples for interest rates, we use the Bachelier model, also known as the *Gauss* or *normal* model.

The Bachelier pricing formulas for calls and puts are

$$C_\phi(T, K, \sigma) = (f - K)\,\Phi(d) + \sigma\sqrt{T}\phi(d) \qquad \text{and}$$
$$P_\phi(T, K, \sigma) = (K - f)\,\Phi(-d) + \sigma\sqrt{T}\phi(d)$$

with

$$d = \frac{f - K}{\sigma\sqrt{T - t_0}},$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the normal CDF and PDF respectively. Here, σ is called the *Bachelier*, *normal* or *bp* implied volatility. The choice of the reference model depends on the given market. For instance the Bachelier model can be applied to negative asset values (applicable to interest rates) and does not require the values to be bounded.

Often we are not only interested in a single option price, but require prices for a set of maturities and strikes. Consider a set $\mathcal{T} := \{T_1, T_2, \ldots, T_N\}$ of option maturities and let $\mathcal{K} := \{\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_N\}$, $\mathcal{K}_i := \{K_{i,1}, K_{i,2}, \ldots, K_{i,M_i}\}$ be sets of strike values indexed by the number of maturities considered. Usually $\mathcal{K}_i = \mathcal{K}_j$ for all $1 \leq i, j \leq N$. Now, consider the implied volatility for each quoted option with respect to $T_i \in \mathcal{T}$, $K_j \in \mathcal{K}$:

$$\Sigma_d : \mathcal{T} \times \mathcal{K} \to \mathbb{R}^+$$
$$(T, K) \mapsto \sigma_d.$$

The map Σ_d is called the discrete implied volatility surface. From this set, values for strikes $K \notin \mathcal{K}$ may be inferred by interpolation and extrapolation. The latter techniques must respect arbitrage relationships, requiring practitioners to use interpolation methods consistent with an arbitrage-free model. Furthermore, starting with the current levels of volatility as an input, it is relevant to consider

the dynamics of the implied volatilities in order to manage exotic options. Such options may, for instance, include payoffs that strongly depend on the forward volatility. Examples include forward-starting options and cliquet options. Practitioners refer to the implied volatility surface as the skew or smile. These names originate from the shapes that the surfaces exhibit in typical market environments.

While there are many approaches for modelling the dynamics of instantaneous volatility, including Levy and jump-diffusion models, see, e.g., Merton [1976], Madan and Seneta [1990], Madan et al. [1998] and Schoutens [2003] for a non-exhaustive list, practitioners most often use stochastic volatility models to manage this type of risk. Selecting a stochastic volatility model determines the continuous implied volatility surface. The process of matching to the observed discrete implied volatility surface is called calibration, and, once a model is calibrated, the continuous implied volatility surface may be used for interpolation and extrapolation. We consider the continuous implied volatility surface given by the map

$$\begin{aligned} \Sigma_{c,0} : [0,T] \times [K_l, K_u] &\to \mathbb{R}^+ \\ (T,K) &\mapsto \sigma_c \end{aligned}$$

and its dynamics, $\Sigma_{c,t}(T,K)$, $t \in \mathbb{R}^+$. By choosing a stochastic volatility model these dynamics are implicitly determined by the model. In particular, we consider the coupled SDEs given by

$$\begin{cases} dF_t = C(F_t) v_t \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(v_t)\, dt + \nu(v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases} \tag{3.1.1}$$

The choice of the model and parameters should ensure the best fit to the current (discrete) market implied volatility surface and that the dynamics are suitable for risk management and hedging of exotic contracts. Often ease of implementation determines the choice of the model, rather than model suitability. We provide a general modelling approach with a tractable computational framework that does not require this compromise.

To achieve numerical tractability, we use singular perturbation methods to derive an approximate PDE, called the *effective PDE*, for the marginal probability density of the asset. Here, this probability density should be understood as

$$\mathbb{P}\left[F < F_t < F + dF \,|\, F_{t_0} = f, v_{t_0} = \alpha \right].$$

This technique was originally introduced by Hagan et al. [2002, 2014, 2015, 2016] for SABR models. Our general framework provides an extension to this approach

and covers most well-known stochastic volatility models, including these and other SABR models, as well as the Stein and Stein [1991], and Schöbel and Zhu [1999] models. Having derived the general effective equation, we use it to analyse all the SABR and ZABR model variants.

The parameters of the classic SABR model have a well-known interpretation. The CEV coefficient, β, specified in the backbone of the implied volatility, $C(F_t) = F_t^\beta$, blends between two extreme cases, being the Gaussian ($\beta = 0$) and lognormal ($\beta = 1$) stochastic volatility models. The initial value of the volatility is determined by the level of the initial at-the-money volatility, v_{atm}, the initial forward rate, f, and β using $\alpha = f^{1-\beta} v_{\text{atm}}$. The volatility has no drift, i.e., $\mu(v_t) = 0$, and the volatility of volatility, $\nu(v_t) = \nu v_t$, controls the level of convexity in the smile through the constant ν. The correlation parameter ρ controls the influence of the stochastic volatility component. For $\rho = 0$ this influence is high, while for values of $|\rho|$ close to 1 we essentially recover local volatility dynamics, with the stochastic behaviour resulting mostly from the Brownian motion driving the forward value.

Since its introduction, a critique of the classic SABR model has been that it cannot model mean reversion of the volatility process and that it only has limited control over the wings of the implied volatility smile. These features are especially important given the prevailing regimes in interest rate markets, with negative rates and very high prices for caps and swaptions. Further modelling flexibility is also necessary when pricing instruments that are sensitive to the shape of the implied volatility smile, such as constant maturity products or cash-settled swaptions.

To model negative rates, two approaches are used. The first is to augment the backbone volatility function of the classic SABR model with a displacement parameter d, in which case $C(F_t) = (F_t + d)^\beta$. Although strictly speaking this is a displaced SABR model, we shall refer to it as the SABR model. The second approach is to use free boundary models. These models have a volatility function given by $C(F_t) = |F_t|^\beta$, and were introduced by Antonov et al. [2015b].

To provide more control of the smile, in particular the wings, we shall consider two further extensions. The first is the ZABR model, originally introduced by Andreasen and Huge [2013]. This model specifies the volatility as a CEV process, using $\nu(v_t) = \nu v_t^\gamma$, in terms of the CEV coefficient γ. We further propose a second extension by allowing the drift of the volatility to be mean reverting using $\mu(v_t) = \kappa(\theta - v_t)$, specified in terms of mean-reversion level θ and mean-reversion speed κ. These two methods may be used separately or in conjunction with one another.

While we shall focus almost exclusively on the extensions of the ZABR model, we emphasize that other modelling choices for the implied volatility backbone

are possible, corresponding to different choices of the function $C(\cdot)$. We consider the displaced ZABR, free boundary ZABR (fZABR) and mean-reverting ZABR (mrZABR) models. As with the SABR model, we refer to the displaced ZABR model as just the ZABR model. For these cases, we consider the stylized effect of the parameters on the implied volatility smile and how they may be set with regard to given market data. In particular, these models incorporate desired model dynamics and observable market behaviour, allowing the volatility to remain range-bound and enabling better flexibility for controlling the wings of the volatility surface, all while achieving the same at-the-money calibration quality as the SABR model.

Our methodology relies on efficiently solving the one-dimensional effective PDE. The alternative would be to solve the corresponding two-dimensional problem. From a numerical perspective, a non-zero correlation between the asset and the volatility driver makes it impossible to apply alternating direction implicit (ADI) methods directly. For an approach that uses transforms and an application to the Heston model, see In 't Hout and Foulon [2010]. Transformations may be problematic when it comes to boundary conditions. The boundaries of the transformed dynamics may be more complicated than the original ones. Another approach is to handle the terms involving both quantities F_T and v_T by an explicit step, but this requires small time steps in order to avoid numerical instability. When considering examples to illustrate the new models, we use the numerical methods described in Hagan [2015], Le Floc'h and Kennedy [2017], and Kienitz et al. [2017]. Rather than delving into the intricacies of two-dimensional finite difference techniques, we provide evidence of numerical accuracy by comparing our method with Monte Carlo simulation.

In summary, we present a methodology useful for the analysis and application of a variety of stochastic volatility models. Our approach is to derive a deterministic one-dimensional local volatility model for implementation using standard numerical techniques. The new methodology allows us to consider several parametrisations and generalizations of the classic SABR model, thus incorporating stylized features of the implied volatility smile observed in the market. In particular, the approach enables the incorporation of instruments other than vanilla European call and put options when performing calibrations. The added flexibility is essential, for instance, when accurately pricing constant maturity swaps. Furthermore, such models may be useful in a combined local stochastic volatility setting where stochastic volatility dynamics are enhanced by a leverage function computed using a local volatility model.

The rest of the paper proceeds as follows: in Section 3.2 we introduce the methodology and derive the effective PDE for the general framework. The whole derivation—

including all details—can be found in Appendix 3.5. In Section 3.2.2 we consider the approximating formula and then provide more detail on concrete models in Section 3.2.3. In Section 3.2.3.1 we use the method to analyse the smile dynamics of the ZABR model, while in Section 3.2.3.2 we investigate the extended ZABR models.

The approach taken to demonstrate numerical examples in Section 3.3 is based on one-dimensional PDE solvers. We use numerical methods to compare the resulting stochastic volatility modelling approaches in Section 3.3.1 and the calibration behaviour in Section 3.3.2. Finally, Section 3.3.3 compares our approach to a classical Monte Carlo approach of Lord and Farebrother [2014]. Section 3.4 concludes with a summary, draws conclusions and outlines directions for future research.

3.2 Main Results

To provide a tractable way to compute solutions for the general two-dimensional SDE in (3.1.1), our main goal is to approximate the dynamics by a suitable one-dimensional local version. For this we consider the marginal probability density of the asset, also called the reduced density or effective probability, as our main object of interest. The reduced density Q of F at time t, starting from time t_0, is defined as

$$Q(t,F)\,dF = \mathbb{P}\Big[F < F_t < F + dF \,\Big|\, F_{t_0} = f, v_{t_0} = \alpha\Big]. \tag{3.2.1}$$

Given the reduced density for a specified exercise time T, we can then recover the corresponding call and put prices for all strikes by an evaluation of

$$C_Q(T,K) = \int_K^\infty (F-K)Q(T,F)\,dF \qquad \text{and}$$
$$P_Q(T,K) = \int_{-\infty}^K (K-F)Q(T,F)\,dF.$$

To compute the reduced density, we derive a PDE of the form

$$\partial_t Q(t,F) = \partial_{FF}\left[D(t,F)Q(t,F)\right], \quad Q(t_0,f) = \delta(F-f), \tag{3.2.2}$$

where $D(\cdot,\cdot)$ is a function that involves the model parameters and depends on t and the asset value F. This PDE describes the pricing equation in the local volatility setting, originally due to Dupire [1994]. More precisely, the local volatility function corresponds to $\sqrt{2D(\cdot,\cdot)}$.

To derive this PDE, we use singular perturbation methods to systematically analyse the system

$$\begin{cases} dF_t = \varepsilon C(F_t) v_t \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(v_t)\, dt + \varepsilon \nu(v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases} \tag{3.2.3}$$

Once the result is derived we set ε back to 1.

The effective PDE, also called the effective forward equation, is accurate to order $\mathcal{O}(\varepsilon^2)$. To ensure a stable and efficient numerical implementation for solving this PDE, we must specify the boundary behaviour. We consider two PDEs for accumulating probability. The lower boundary b_l is either explicitly specified when the model under consideration is not admissible for values below b_l, or artificially when setting up a grid for numerical computation. The upper boundary is set by specifying the highest level b_u to be considered. The corresponding values of the probability densities are denoted by Q^L and Q^R. In our setting the PDEs are given by

$$\begin{aligned} & \partial_t Q^L(t) = \lim_{F \downarrow b_l} \partial_F \left[D(t,F) Q(t,F) \right], && Q^L(t_0) = 0 \\ \text{and} \qquad & \partial_t Q^R(t) = -\lim_{F \uparrow b_u} \partial_F \left[D(t,F) Q(t,F) \right], && Q^R(t_0) = 0. \end{aligned}$$

For the SABR model the derivation can be found in Hagan [2015], and Le Floc'h and Kennedy [2017], while the fSABR model is considered in Kienitz et al. [2017].

Next, we derive the effective equation for the general setting (3.1.1), then we restrict ourselves to a special class of dynamics, called the ZABR class. This includes the standard ZABR model of Andreasen and Huge [2013], and also the free boundary version of ZABR, which generalizes the results from Antonov et al. [2015b], and the mean-reverting ZABR.

Remark 3.2.1. To put our result in a broader context, we mention a result by Gyöngy [1986], which applies to the general stochastic process satisfying

$$dY_t = \alpha(t, Y_t)\, dt + \beta(t, Y_t)\, dW_t,$$

where α and β are bounded[1] adapted stochastic processes. He shows that there exists an SDE

$$dx_t = a(t, x_t), dt + b(t, x_t)\, dW_t,$$

[1]This assumption is not a severe constraint. When working with a stochastic volatility function that may not be bounded, it is always possible to bound it at a large enough level that the probability of reaching that level is tiny and of no concern.

in terms of non-random coefficients, with solution x_t having the same one-dimensional probability distribution as Y_t. The coefficients a and b satisfy

$$a(t,x) = \mathbb{E}[\alpha(t,Y_t)|Y_t = x] \qquad \text{and} \qquad b(t,x)^2 = \mathbb{E}[\beta^2(t,Y_t)|Y_t = x].$$

In particular this holds true if β is a stochastic volatility process, which may be dependent on Y_t and another Brownian motion correlated with W_t. This existence result makes it possible to price European options by considering a one-dimensional local volatility model [Derman and Kani, 1994, Dupire, 1994]. Using the results and notation of the latter reference and considering an SDE of a forward, i.e., $\alpha_t = 0$, the result states that

$$\sigma^2_{\text{local}}(x,t) = \mathbb{E}[\beta(t,Y_t)^2|Y_t = x].$$

Crucially, if the local volatility function remains positive it implies that the one-dimensional model is arbitrage-free. Furthermore, it is possible to discretize the corresponding Fokker-Planck equation in an arbitrage-free way. This is essentially what we achieve with our method. We derive a one-dimensional deterministic representation of the local volatility from a stochastic volatility model by singular perturbation methods. For our numerical results we apply Hagan [2015], which directly solves the PDE for the probability density. Another method, introduced by Andreasen and Huge [2011a], starts by considering the backward equation with corresponding boundary conditions. Within a discrete time and space setting, the backward equation is discretized and multiplied by the transpose of the probability to obtain adjoint equations, with the solution yielding the probability vector. This approach overcomes the necessity of specifying and handling boundary conditions for the forward equation, and leads to an arbitrage-free discrete density (positive and summing to one).

3.2.1 The Effective PDE

To derive the effective PDE we make the following assumptions related to (3.1.1):

Assumption I. The drift term, $\mu(\cdot)$, is differentiable, with derivative $\mu'(\cdot)$, and a solution $Y(t,t_0,\alpha)$ to the following PDE exists:

$$\begin{cases} \partial_t Y(t,t_0,\alpha) = \mu(Y(t,t_0,\alpha)) \\ Y(t,t,\alpha) = \alpha \\ Y(t_0,t_0,\alpha) = \alpha. \end{cases}$$

Assumption II. The function Y is differentiable and has an inverse function $y(t_0,t,a)$ such that

$$Y(t,t_0,\alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0,t,a).$$

Remark 3.2.2. Functions $\mu(\cdot)$ allowing a closed-form solution include

(i) for $\mu(x) = \mu$ the solution is $Y(t, t_0, \alpha) = \alpha + \mu(t - t_0)$;

(ii) for $\mu(x) = \kappa(\theta - x)$ the solution is $Y(t, t_0, \alpha) = \alpha e^{-\kappa(t-t_0)} + \theta(1 - e^{-\kappa(t-t_0)})$.

Assumption III. The functions

$$X(t, t_0, \alpha) = \partial_\alpha Y(t, t_0, \alpha), \qquad Z(t, u) = Z(t, u, t_0, \alpha) = y(u, t, Y(t, t_0, \alpha)),$$
$$z(F) = \int_f^F \frac{1}{C(u)}\, du, \qquad s(t) = S(t_0, t, \alpha) = \int_{t_0}^t Z(t, u, t_0, \alpha)^2\, du$$

and

$$\psi(t, u, Z) = \nu\big(Z(t, u)\big) Z(t, u) X\big(t, u, Z(t, u)\big)$$

are well defined[2], $X\big(t, u, Z(t, u)\big)^{-1}$ exists, and the following integral functions are defined:

$$I_1(t) = \rho \int_{t_0}^t \psi(t, u, Z)\, du,$$
$$I_2(t) = 2 \int_{t_0}^t \nu\big(Z(t, u)\big)^2 X\big(t, u, Z(t, u)\big)^2 \int_u^t Z(t, v) X\big(t, v, Z(t, v)\big)^{-1}\, dv\, du,$$
$$I_3(t) = \rho \int_{t_0}^t \psi(t, u, Z) \int_u^t Z(t, v) X\big(t, v, Z(t, v)\big)^{-1}\, dv\, du,$$
$$I_4(t) = \rho^2 \int_{t_0}^t \psi(t, u, Z) \int_u^t \partial_Z\Big(\psi(t, v, Z)\Big) X\big(t, v, Z(t, v)\big)^{-1}\, dv\, du,$$
$$I_5(t) = \int_{t_0}^t \nu\big(Z(t, u)\big)^2 X\big(t, u, Z(t, u)\big)^2\, du.$$

Assumption IV. The function $C(\cdot)$ is differentiable at f, with derivative denoted by $C'(\cdot)$.

Theorem 3.2.3. *Given that the general stochastic volatility model* (3.1.1) *obeys Assumptions* I–IV*, an effective PDE for the effective probability* (3.2.1)*, of the form* (3.2.2)*, can be derived, with function D given by*

$$D(t, F) = \frac{1}{2} a(t)^2 C(F)^2 e^{G(t)} \big(1 + 2b(t) z(F) + c(t) z(F)^2\big),$$

[2]Note that this definition of $z(F)$ differs by a factor of $\frac{1}{\varepsilon}$ compared to the definition in (3.5.4).

where the coefficients are specified as

$$a(t) = Y(t, t_0, \alpha), \qquad c(t) = b(t)^2 + \frac{1}{a(t)s(t)^2} I_2(t) - \frac{6b(t)}{s(t)^2} I_3(t) + \frac{2}{a(t)s(t)^2} I_4(t),$$
$$b(t) = \frac{1}{a(t)s(t)} I_1(t), \quad G(t) = -s(t)c(t) - s(t)b(t)\Gamma_0 + \frac{1}{a^2} I_5(t)$$

and

$$\Gamma_0 = -C'(f).$$

Proof. In Appendix 3.5 we show that the effective PDE of order $\mathcal{O}(\varepsilon^2)$ is given by

$$\begin{cases} \partial_t Q(t,F) = \frac{1}{2}\varepsilon^2 a(t)^2 \partial_{FF}\Big[C(F)^2 Q(t,F) e^{\varepsilon^2 G(t)}\big(1 + 2\varepsilon b(t) z(F) + \varepsilon^2 c(t) z(F)^2\big)\Big] \\ Q(t,F) \to \delta(F - f) \text{ as } t \to t_0^+. \end{cases} \tag{3.2.4}$$

This comes from combining (3.5.3), (3.5.5), (3.5.10) and (3.5.1). Setting ε to 1 yields the desired form. □

Remark 3.2.4. The class of models is not restricted to the above choices of the function Γ_0. Choosing a different form could in turn impose a new version of Assumption IV.

Remark 3.2.5. The derivation of the effective PDE, as shown in Appendix 3.5, may be extended to include time dependent parameters. The proof stays the same—only the functions specified in Assumptions *I–III* must be adjusted to incorporate the extra dependency.

Figure 3.1 shows the output obtained by numerically solving the effective PDE. It is the density of the asset at maturity and depends on all the input parameters.

3.2.2 Implied Volatility Formula

As seen in Theorem 3.2.3, the effective PDE of the general model is of the same form as the classical or mean-reverting SABR Hagan et al. [2014, 2020b]. In particular, both these models fit into the same framework, which allows for a direct approximation of the implied normal or Black volatility [Hagan et al., 2016].

To show that our model also fits into this framework, we use Effective Media theory [Hagan et al., 2018b] to approximate the effective PDE of Equation (3.2.4) by a

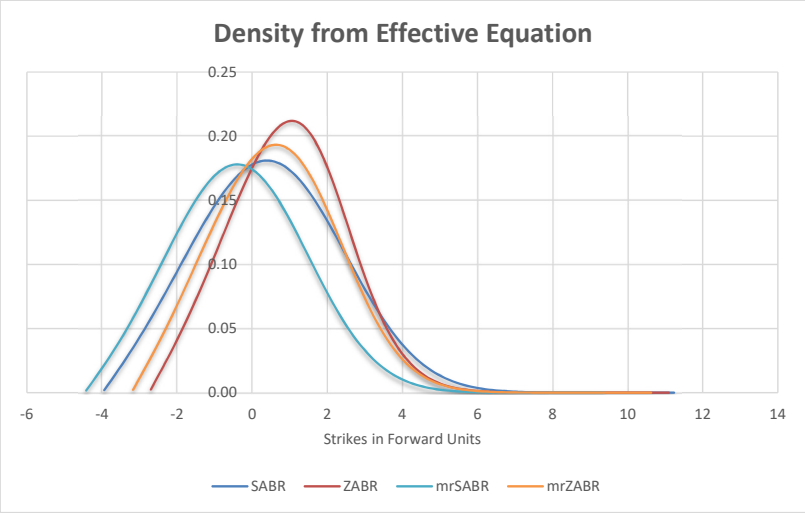


Figure 3.1: Numerical solution of the effective PDE for the SABR, ZABR, mrSABR and mrZABR models, with parameters $\beta = 0.5$, $\nu = 0.3$, $\rho = -0.8$, $f = 0.005$, $T = 5$, $d = 0.001$, $\gamma = 0.8$, $\kappa = 0.2$, with $\alpha = f^{1-\beta} v_{\text{atm}}$ and $v_{\text{atm}} = 0.3$ representing the initial at-the-money implied volatility.

suitable SABR effective PDE. To facilitate this we consider a fixed maturity T and set the initial time $t_0 = 0$, from now on. We start by defining time-independent parameters $\bar{b}$, $\bar{c}$ and $\bar{G}$, which allow for an approximation of Equation (3.2.4) to order $\mathcal{O}(\varepsilon^2)$, at time T. These parameters are generally given by Hagan et al. [2018b, Equation 2.4]:

$$\begin{aligned}
\bar{b} &= \frac{2}{T^2} \int_0^T u b(u)\, du \\
\bar{c} &= \frac{3}{T^3} \int_0^T u^2 c(u)\, du + \frac{18}{T^3} \int_0^T b(u) \int_0^u v b(v)\, dv\, du - 3\bar{b}^2 \\
\bar{G} &= \frac{1}{T} \int_0^T G(u)\, du + \frac{1}{T} \int_0^T u(c(u) - \bar{c})\, du.
\end{aligned}$$

Note that before using these equations, the functions b and c of Theorem 3.2.3 must be modified to fit into the setting of Hagan et al. [2018b, Equation 2.2].

With the constant effective parameters we can define the so called effective SABR parameters:

$$\nu_{\text{eff}} = \sqrt{\bar{c}}, \qquad \rho_{\text{eff}} = \frac{\bar{b}}{\sqrt{\bar{c}}}, \qquad \alpha_{\text{eff}} = \alpha \left(1 + \frac{1}{2} \bar{G} + \frac{1}{4} \alpha \bar{b} \Gamma_0 T \right). \tag{3.2.5}$$

These allow us to approximate our model to order $\mathcal{O}(\varepsilon^2)$ with an SABR model, and in turn provide us with various available implied volatility formulae. One prominent example would be the formula provided in Hagan et al. [2016]:

$$\sigma(T,K) = \varepsilon \frac{\nu_{\text{eff}}(K-f)}{R(\zeta)} \begin{cases} 1+\varepsilon^2\Theta(\zeta)T & \text{if } \Theta \geq 0 \\ \frac{1}{1-\varepsilon^2\Theta(\zeta)T} & \text{if } \Theta < 0, \end{cases}$$

with

$$R(\zeta) = \log\left(\frac{\rho_{\text{eff}}+\zeta+E(\zeta)}{1+\rho_{\text{eff}}}\right),$$
$$\Theta(\zeta) = \frac{\nu_{\text{eff}}^2}{24}\left(3\frac{\rho_{\text{eff}}+\zeta-\rho_{\text{eff}}E(\zeta)}{R(\zeta)E(\zeta)}-1\right) + \frac{\Delta_0\alpha_{\text{new}}^2}{6}\left(1-\rho_{\text{eff}}^2+\frac{(\rho_{\text{eff}}+\zeta)E(\zeta)-\rho_{\text{eff}}}{R(\zeta)}\right),$$

where

$$\zeta = \frac{\nu_{\text{eff}}}{\alpha_{\text{new}}}\int_f^K \frac{1}{C(u)}\,du, \qquad E(\zeta) = \sqrt{1+2\rho\zeta+\zeta^2},$$
$$\alpha_{\text{new}} = \alpha_{\text{eff}}\big(1-\frac{1}{4}\rho_{\text{eff}}\nu_{\text{eff}}\alpha_{\text{eff}}\Gamma_0 T\big), \qquad \Delta_0 = \frac{1}{4}C(f)C''(f)-\frac{1}{8}C'(f)^2.$$

In the next section we consider special cases to illustrate the applicability of our approach.

3.2.3 ZABR-type Models

We now illustrate our approximation by considering various extensions to the ZABR model. In particular, we provide a flexible framework that incorporates several novel parametrisations allowing control of features of the implied volatility smile, like the outer wings. We compare these extensions to existing approaches and mention how various other parametrisations of the local volatility backbone of the stochastic volatility model may be handled.

Consider SABR and ZABR-type stochastic volatility models of the form (3.2.3), where the functions $C(\cdot)$, $\mu(\cdot)$ and $\nu(\cdot)$ are given in Table 3.1. The SABR-type models were considered by Hagan et al. [2002, 2014], Antonov et al. [2015b], and Kienitz et al. [2017] and the ZABR model, without displacement ($d = 0$), was introduced by Andreasen and Huge [2013]. We consider the displaced ZABR, free ZABR and mean-reverting ZABR models. It should be noted that the mrZABR model generalizes all the other models except the free-boundary models.

$C(F_t)$	$\mu(v_t)$	$\nu(v_t)$	Model
$(F_t+d)^\beta$	0	νv_t	SABR
1	0	νv_t	Normal SABR (nSABR)
$\lvert F_t\rvert^\beta$	0	νv_t	Free SABR (fSABR)
$(F_t+d)^\beta$	$\kappa(\theta - v_t)$	νv_t	Mean reverting SABR (mrSABR)
$(F_t+d)^\beta$	0	νv_t^γ	ZABR
$\lvert F_t\rvert^\beta$	0	νv_t^γ	Free ZABR (fZABR)
$(F_t+d)^\beta$	$\kappa(\theta - v_t)$	νv_t^γ	Mean reverting ZABR (mrZABR)

Table 3.1: Parametrizations of the SABR and ZABR-type models, in terms of constants β, d, κ, θ, ν, and γ.

Remark 3.2.6. Note that the modulus function appearing in $C(\cdot)$ for fSABR and fZABR models is not differentiable at the point 0 and thus does not strictly satisfy Assumption IV. Consequently it does not fit into the framework of Theorem 3.2.3. To overcome this problem we follow the setting of the original free SABR model [Antonov et al., 2015b] and only consider the case where $f \neq 0$.

The models presented in Table 3.1 are a non-exhaustive list, with many other choices possible. As mentioned previously, the models of Stein and Stein [1991], and Schöbel and Zhu [1999] may be accommodated. Other examples include the models of Jäckel and Kahl [2008], Karlsmark [2013] or Balland and Tran [2013], which suggest non-CEV forms of the function $C(\cdot)$ including:

(i) $C(F_t) = F_t^{\beta(F_t)}$, where $\beta(F_t) = \beta_0 + (\beta_\infty - \beta_0)\left(1 - e^{-\frac{F_t}{F_{\max}}}\right)$ with $F_{\max} \gg F_{t_0}$,

(ii) a double-beta backbone $C(F_t) = F_t^\beta \dfrac{\left(\frac{F_t}{F_1}\right)^{\beta_1} + 1}{\left(\frac{F_t}{F_2}\right)^{\beta_2} + 1}$, or

(iii) a hyperbolic backbone

$$C(F_t) = \frac{f^\beta}{\beta}\left[(1-\beta+\beta^2)\frac{F_t}{f} + (\beta-1)\left(\sqrt{\left(\frac{F_t}{f}\right)^2 + \beta^2\left(1-\frac{F_t}{f}\right)^2} - \beta\right)\right].$$

While our method may be used directly on (ii) and (iii), since the integral $\int C(u)^{-1}\,du$ can be computed in closed form, the corresponding integral for (i) must be evaluated numerically leading to possible performance issues. It should be noted that the closed-form solution of (ii) involves hypergeometric functions, which may not be numerically stable.

3.2.3.1 Effective PDE for ZABR

Consider the ZABR model as specified in Table 3.1. The corresponding coefficients for the effective PDE in (3.2.4) may be greatly simplified.

First of all, the function Y has an explicit form that equals the initial value:

$$Y(t, t_0, \alpha) = \alpha.$$

The evaluation of the coefficients needed for the effective PDE yields

$$\begin{aligned} b &= \rho\nu\alpha^{\gamma-2}, \\ c &= \nu^2\alpha^{2(\gamma-2)}(1 + (\gamma-1)\rho^2), \\ G(t) &= -\rho^2\nu^2\alpha^{2(\gamma-1)}(t-t_0)(\gamma-1) - (t-t_0)\rho\nu\alpha^{\gamma}\Gamma_0 \end{aligned} \tag{3.2.6}$$

and gives the ZABR effective PDE characterized by the function

$$D(t, F) = \frac{1}{2}\alpha^2 C(F)^2(1 + 2bz(F) + cz(F)^2)e^{G(t)}.$$

For the numerical implementation of the ZABR model, we use the effective SABR parameters presented in Section 3.2.2. Given the explicit form of the coefficients in (3.2.6), the effective SABR parameters are given by

$$\begin{aligned} \nu_{\text{eff}} &= \nu\alpha^{\gamma-1}\sqrt{1 + (\gamma-1)\rho^2}, \\ \rho_{\text{eff}} &= \frac{\rho}{\sqrt{1 + (\gamma-1)\rho^2}}, \\ \alpha_{\text{eff}} &= \alpha\left(1 + \frac{1}{4}\rho^2\nu^2\alpha^{2(\gamma-1)}(1-\gamma)T\right). \end{aligned} \tag{3.2.7}$$

Note that these are the coefficients used for the actual implementation. In particular, we have already set $\varepsilon = 1$ and approximated the original exponential function. To guarantee that the term $\sqrt{D(t, F)}$ remains real, we further impose the condition

$$\gamma > 1 + \frac{\rho^2 - 1}{\rho^2}.$$

Furthermore, we emphasize that our local volatility approximation is of higher order than those provided by Andreasen and Huge [2013] or Balland and Tran [2013]. The difference is an exponential term that incorporates time dependence, i.e. the term $e^{G(t)}$ from Theorem 3.2.3. To illustrate the differences stemming from the different approximations of local volatility we consider the local volatility

function of the SABR model approximated by the methods in Andreasen and Huge [2013] or Balland and Tran [2013] compared to our approach, which recovers the form derived by Hagan et al. [2014]. The results are depicted in Figure 3.2 and demonstrate the additional dependence of the approximation on the maturity under consideration.

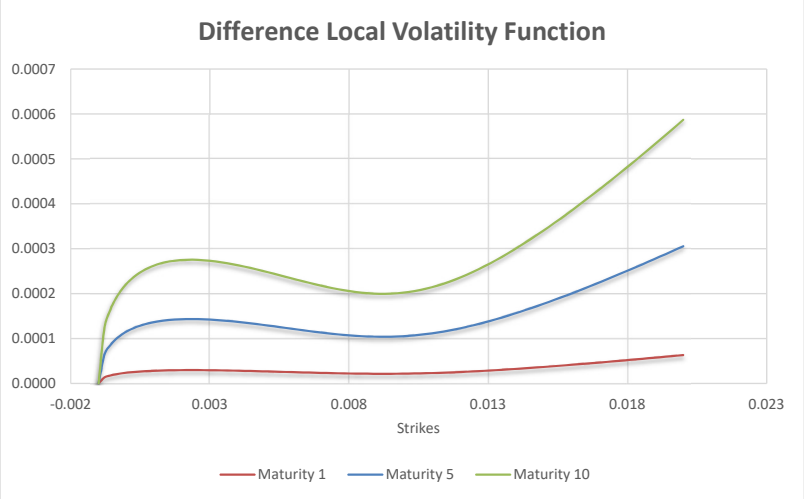


Figure 3.2: Illustrating the difference in the local volatility approximation in the SABR Model for different maturities and given model parameters $\beta = 0.5$, $\nu = 0.3$, $\rho = -0.8$, $f = 0.005$, $T \in \{1, 5, 10\}$, $d = 0.001$, with $\alpha = f_0^{1-\beta} v_{\text{atm}}$ and $v_{\text{atm}} = 0.3$ representing the initial at-the-money implied volatility.

3.2.3.2 Effective PDE for Free ZABR and Mean-Reverting ZABR

Having examined the ZABR model, we can now consider the natural extensions of the fZABR and mrZABR models.

For the fZABR model, the effective SABR parameters are the same as those for the ZABR model and are given in (3.2.7). The difference between the models lies exclusively in the function $C(F)$, as was the case in the extension of the SABR to the fSABR model, see Kienitz et al. [2017].

Considering the mrZABR model with a reversion back to the initial state, i.e. with $\theta = \alpha$, as specified in Table 3.1, the corresponding parameters for the effective

forward equation change and are given by

$$
\begin{aligned}
b(t) &= \frac{\rho\nu\alpha^{\gamma-2}}{\kappa(t-t_0)}\big(1-e^{-\kappa(t-t_0)}\big),\\
c(t) &= \frac{(1+\rho^2)\nu^2\alpha^{2(\gamma-2)}}{\kappa^2(t-t_0)^2}\big(1-e^{-\kappa(t-t_0)}\big)^2\\
&\quad+\frac{6\rho^2\nu^2\alpha^{2(\gamma-2)}}{\kappa^3(t-t_0)^3}\big(1-e^{-\kappa(t-t_0)}\big)\big(1-\kappa(t-t_0)-e^{-\kappa(t-t_0)}\big)\\
&\quad+(1+\gamma)\frac{2\rho^2\nu^2\alpha^{2(\gamma-2)}}{\kappa^2(t-t_0)^2}\big(1-(1+\kappa(t-t_0))e^{-\kappa(t-t_0)}\big),\\
G(t) &= -\alpha^2(t-t_0)c-\frac{\rho\nu\alpha^{\gamma}}{\kappa}\big(1-e^{-\kappa(t-t_0)}\big)\Gamma_0+\frac{\nu^2\alpha^{2(\gamma-1)}}{2\kappa}\big(1-e^{-2\kappa(t-t_0)}\big).
\end{aligned}
$$

Fixing a specified maturity T, the corresponding constant effective parameters of Section 3.2.2 are given by

$$
\begin{aligned}
\bar{b} &= \frac{2\rho\nu\alpha^{\gamma-1}}{\kappa^2T^2}\big(\kappa T-1+e^{-\kappa T}\big),\\
\bar{c} &= \frac{3(1+\rho^2)\nu^2\alpha^{2(\gamma-1)}}{2(\kappa T)^3}\big(2\kappa T+4e^{-\kappa T}-3-e^{-2\kappa T}\big)\\
&\quad+6(1+\gamma)\frac{\rho^2\nu^2\alpha^{2(\gamma-1)}}{(\kappa T)^3}\big(\kappa T+2e^{-\kappa T}-2+\kappa Te^{-\kappa T}\big)\\
&\quad-12\rho^2\nu^2\alpha^{2(\gamma-1)}\left(\frac{\kappa T-1+e^{-\kappa T}}{(\kappa T)^2}\right)\\
\bar{G} &= \frac{\nu^2\alpha^{2(\gamma-1)}}{4\kappa^2T}\big(2\kappa T+e^{-2\kappa T}-1\big)-\frac{1}{2}\bar{c}T-\frac{\rho\nu\alpha^{\gamma}}{\kappa^2T}\big(\kappa T-1+e^{-\kappa T}\big)\Gamma_0.
\end{aligned}
$$

Here we can now use (3.2.5) to derive the effective SABR parameters. Since we are using an SABR model to approximate the ZABR type models, it is not surprising that we observe similar behaviour of the models. For example, if we shift the underlying forward rate f, as shown in Figure 3.3, we see that in both cases the implied volatility moves in the same direction as the forward.

3.2.3.3 Stylized Features of Stochastic Volatility Models

In the introduction we described the effect of the parameters on the SABR model. We now provide some stylized features of the parameters for the extended models.

First, consider the CEV parameter, γ, in the ZABR model, which introduces different dynamics for the volatility, as opposed to the classical SABR model. The

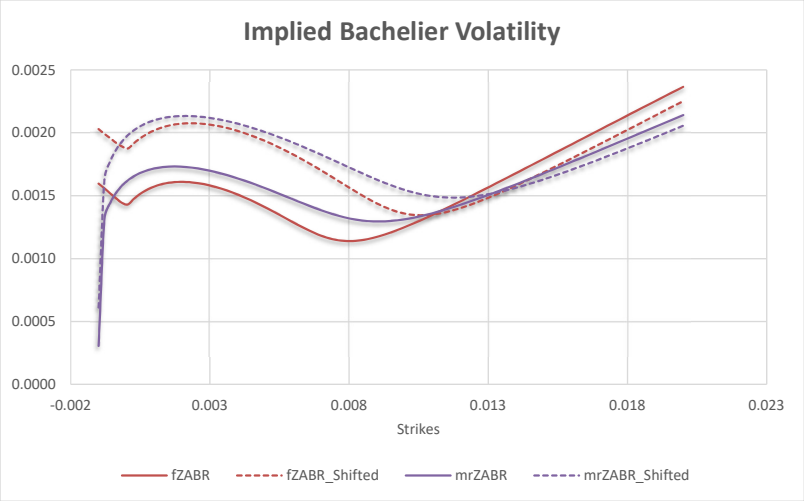


Figure 3.3: Free ZABR and mean reverting ZABR implied volatility with parameters $\beta = 0.5$, $\nu = 0.3$, $\rho = -0.8$, $f = 0.005$, $T = 5$, $d = 0.001$, $\gamma = 0.8$, $\kappa = 0.2$, with $\alpha = f^{1-\beta} v_{\text{atm}}$ and $v_{\text{atm}} = 0.3$ representing the initial at-the-money implied volatility.

action of this parameter was analysed by Andreasen and Huge [2013], where it is shown that it affects the wing of the implied volatility smile. Figure 3.4 illustrates this feature.

Second, consider the parameter κ, which is the speed of mean reversion of the volatility process in the mrSABR and mrZABR models. This parameter controls the speed at which the stochastic volatility reverts to its long-term mean, given by θ. This results in a mechanism used to control the effect of the volatility of volatility. The phrase 'κ kills the skew' describes this effect. We expect a subtle interplay between the parameters governing the volatility of volatility and the mean-reversion speed. Thus calibration and estimation of these parameters may be difficult in practice. The effect of the mean reversion in the mrZABR model, keeping all the other parameters constant, is shown in Figure 3.5.

Third, we consider different backbones of the implied volatility by changing the functional form of $C(\cdot)$. This function can either be chosen beforehand by the modeller, taking into account historical market data as suggested by Hagan et al. [2002], or with the intention of modelling certain features of the underlying asset. We consider two functions, being the displaced CEV function and the free boundary CEV function, which involves a modulus function. As mentioned previously, both methods allow the modelling of negative rates, but the left tails are qualitatively different, as seen in Figure 3.3.

Using a model with a rich parameter set may make it more difficult to perform a successful calibration. To ease this process, we suggest either choosing some of the parameters upfront before calibrating to European call and put option prices or including other financial instruments. If other instruments are available, it is preferable to use those that are directly affected by features controlled by the parameters introduced. For instance, if one introduces a parameter that controls the wings of the implied volatility smile then one should include instruments that depend on that part of the smile. As mentioned previously, when considering interest rates, constant maturity products are ideal for this purpose.

3.3 Numerical Examples

For our numerical implementation we use the effective SABR parameters derived for each model and the methods described in Hagan [2015], Le Floc'h and Kennedy [2017] and Kienitz et al. [2017] adapted to our setting. If not otherwise specified, we use $\beta = 0.5$, $\nu = 0.3$, $\rho = -0.8$, $f = 0.005$, $T = 5$, $d = 0.001$, $\gamma = 0.8$ and $\kappa = 0.2$ as our standard set of parameters. Here $\alpha = f^{1-\beta} v_{\text{atm}}$ with $v_{\text{atm}} = 0.3$ representing the initial at-the-money implied volatility.

3.3.1 Comparison of SABR and ZABR-type Models

To see the impact of our new models, we compare the implied volatilities generated by each model. For this we modify the values of the new parameters γ and κ and compute the implied volatility curve. The results are shown in Figure 3.4 and Figure 3.5. As we can see, the change from an SABR to a ZABR model heavily influences the smile. In particular the OTM end shows much steeper behaviour. The inclusion of mean reversion again works in the contrary direction and brings the model closer to the SABR case.

In Figure 3.6 we considered the density of the ZABR model. Here we can see that the model exhibits a much higher and steeper peak in the density function, compared with the SABR model.

Finally, Figure 3.7 shows that the implied volatility formulas yield a good approximation to the models for moderate to high strikes. As in the case of the SABR model, the approximation becomes worse for very low strikes.

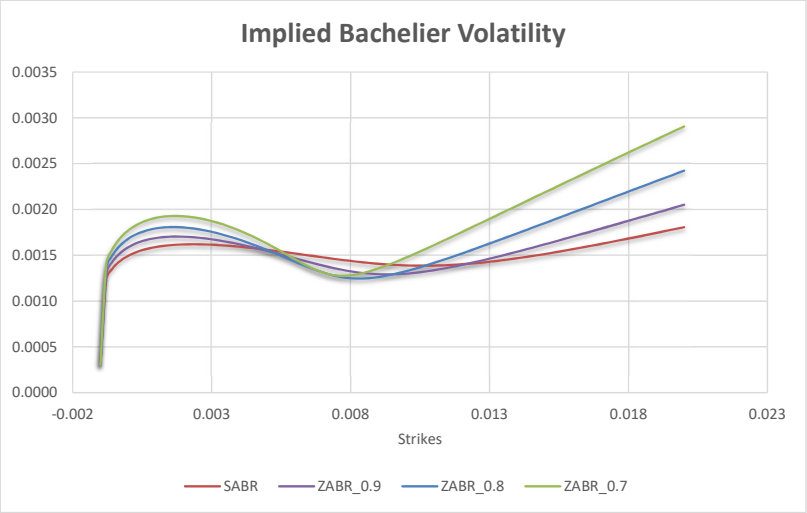


Figure 3.4: Implied volatility for the ZABR model when γ changes.

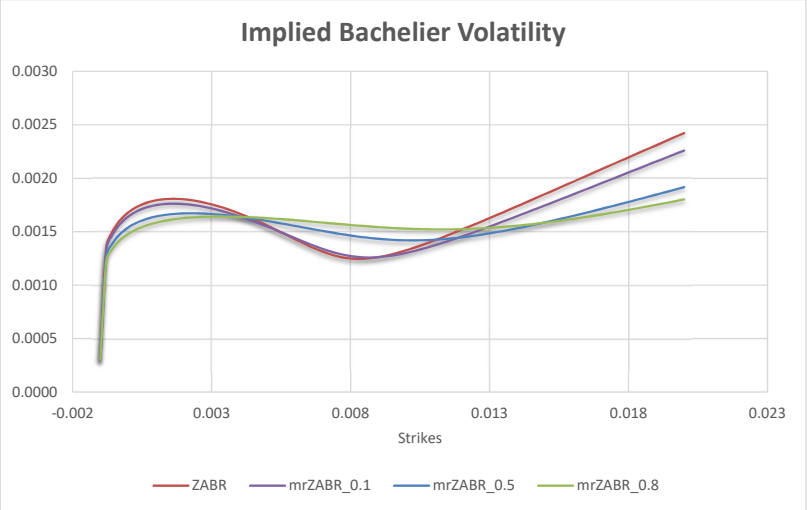


Figure 3.5: Implied volatility for the mean reversion ZABR when κ changes.

3.3.2 Calibration

One crucial aspect of a model is its ability to fit real market data. One common way to calibrate a model is to consider a set of implied volatilities and try to adjust the model parameters to minimize the error between market and model volatilities. For our numerical experiment, we calibrated the models on a set of

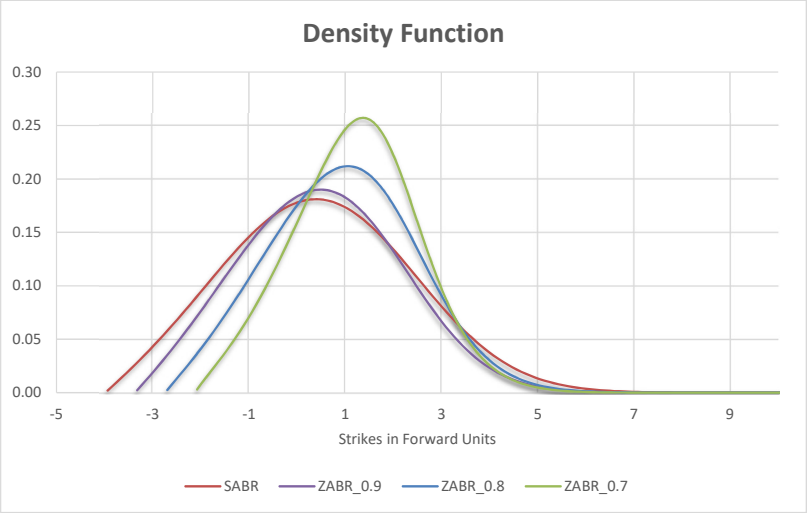


Figure 3.6: Density for the ZABR model when γ changes.

normal swaption volatilities with maturity and tenor of 5 years. The volatility data is for the EUR currency and corresponds to the dates 2 Sept 2019 and 1 Nov 2019.

For our first example, we consider the calibration of the SABR, ZABR and mrZABR models to the volatility data. For all three models, the forward was taken from the market data and the displacement was set at $d = 2\%$. For the ZABR and mrZABR models, we also fixed two additional parameters, $\gamma = 0.9$ and $\kappa = 0.3$, in advance. The remaining parameters were calibrated to the swaption data. Table 3.2 shows the quality of the SABR fit, which deviates very little from the market data. For the strikes shown and to the stated precision, the ZABR and mrZABR calibration gave the same fit as the SABR model, so we have not repeated the numbers. Overall, we observe a very good fit of the models to the market data.

Since the quality of the calibration for the ZABR and mrZABR models is so similar to that of the SABR model, one may ask: why not just use the SABR model? Here it is relevant to mention that in most applications, it is not enough merely to match current market data. It is also necessary to provide scenarios for future prices and obligations with more complicated, and even path-dependent, dynamics. In Antonov et al. [2015a], it is shown that models with the same calibration quality may reveal their flexibility only when pricing exotic derivatives. For instance, CMS index-related derivatives illustrate this. Furthermore, from a regulatory perspective financial institutions are required to perform a prudent val-

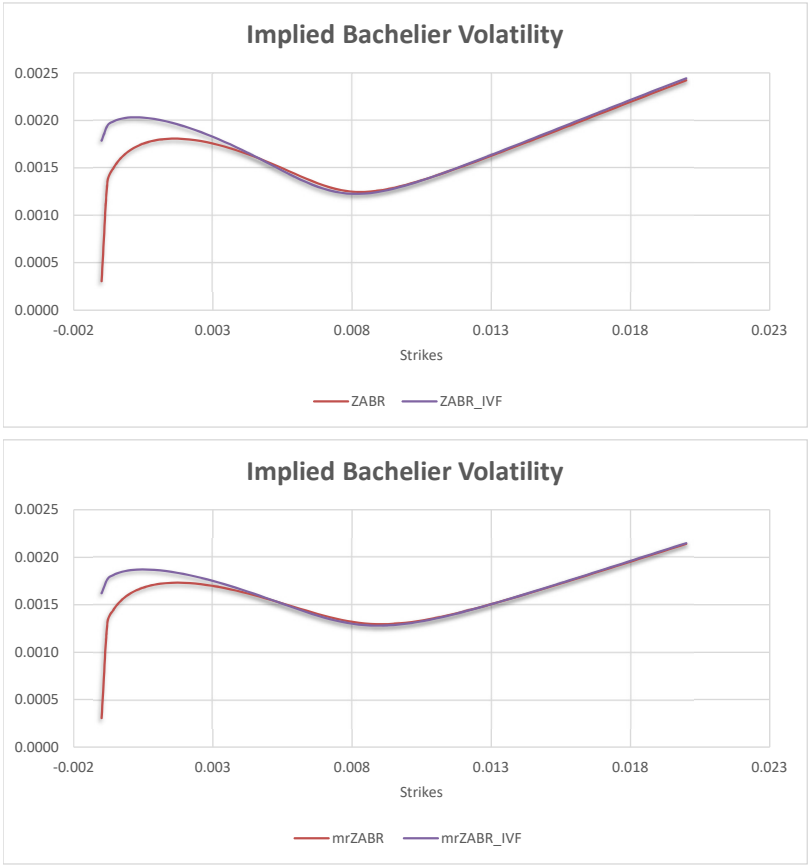


Figure 3.7: Comparison of implied Bachelier volatility functions for the ZABR and mrZABR models.

uation. To satisfy this requirement they often use models that include parameters that allow control of features that cannot be modelled by their production model, and then quantify the difference.

To illustrate this behaviour, consider a reduced set of swaption volatilities, where only the data in the region of the ATM strike is given. This restriction allows for more flexibility in modelling the wings of the smile. In our case, we consider the strike values between -0.36% and 0.64% in Table 3.2 for 1 Nov 2019, and calibrate the models to this reduced data set. Figure 3.8 shows the resulting curves. As can be seen, we achieve good control over the wings, while at the same time keeping the same order of accuracy for the swaption prices.

This increased control over the wings allows for better fits when pricing more exotic products. Consider, for example, CMS index products. We use the results

2 Sept 2019			1 Nov 2019		
Strike	Market Vol	SABR Vol	Strike	Market Vol	SABR Vol
−1.24	0.498	0.482	−0.86	0.473	0.465
−0.74	0.501	0.500	−0.36	0.468	0.468
−0.49	0.505	0.506	−0.11	0.469	0.470
−0.24	0.511	0.511	0.14	0.474	0.474
0.01	0.519	0.516	0.39	0.482	0.480
0.26	0.527	0.523	0.64	0.491	0.488
0.76	0.546	0.539	1.14	0.515	0.510
1.26	0.569	0.561	1.64	0.544	0.540
1.76	0.594	0.589	2.14	0.577	0.577
2.76	0.648	0.656	3.14	0.651	0.656

Table 3.2: Swaption market data for 2 Sept 2019 and 1 Nov 2019. All strikes and implied volatilities are quoted as percentages. The ATM strikes are −0.24% and 0.14%, respectively.

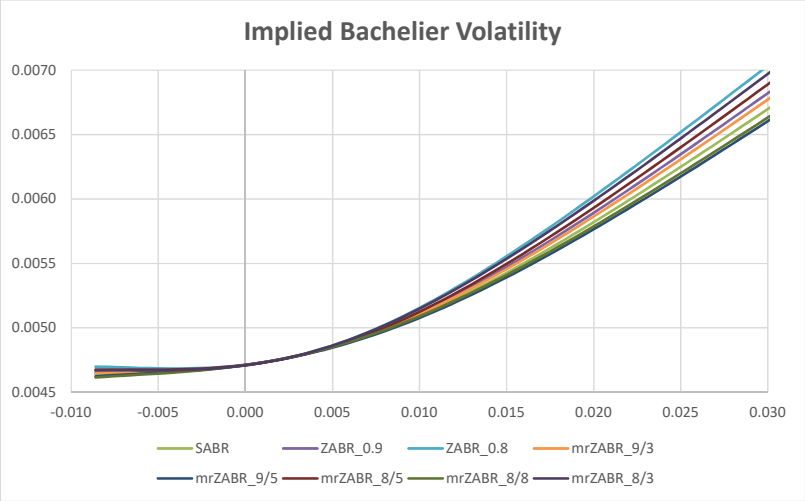


Figure 3.8: The ZABR type models with different values of γ and κ calibrated to swaption data from the 1 Nov 2019.

of Hagan et al. [2020a] to directly approximate CMS swap prices using the parameters of the SABR model. Using the effective SABR parameters derived in Section 3.2.3, these approximations are also valid for the ZABR models. We compute the quadratic swap volatility as given in Hagan et al. [2020a] for each curve of Figure 3.8. The results of the calibration and the quadratic swap volatilities are

Strike (%)	SABR	ZABR (0.9)	ZABR (0.8)	mrZABR (0.9; 0.3)	mrZABR (0.9; 0.5)	mrZABR (0.8; 0.3)	mrZABR (0.8; 0.5)	mrZABR (0.8; 0.8)
-0.36	0.06	0.03	0.10	0.05	0.12	0.03	0.17	0.02
-0.11	0.07	0.06	0.11	0.07	0.05	0.09	0.03	0.05
0.14	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.01
0.39	0.08	0.07	0.02	0.06	0.10	0.07	0.06	0.02
0.64	0.02	0.06	0.21	0.04	0.08	0.08	0.01	0.20
Swap	7.3321	7.3825	7.4854	7.3598	7.2870	7.4435	7.4108	7.2888

Table 3.3: Absolute errors and quadratic swap volatilities in basis points for all calibrated models.

given in Table 3.3. In particular, the quadratic swap volatilities show a variety of different, evenly spaced values, allowing one to choose the most suitable model to fit the desired volatility.

Aside from CMS products, this approach would be suitable for more computationally demanding products. In theory, we could reproduce the experiments using an SABR model; in practice, however, the desired control is much harder to achieve.

3.3.3 Comparison to MC

For the comparison with a Monte Carlo experiment, we chose the same parameter set as above and simulated 1 million paths and 240 time steps for all the models, except for the free SABR and ZABR models where we used 10 million paths and 480 time steps. Table 3.4 shows the observed Bachelier implied volatility. The results using both approaches are very similar. Here we observe an average relative error of about 5%, with the most significant contribution in the tails. Figure 3.9 shows the relative error as a function of strike for all the models considered.

Strike	SABR	SABR(MC)	ZABR	ZABR(MC)	fSABR	fSABR(MC)	fZABR	fZABR(MC)	mrSABR	mrSABR(MC)	mrZABR	mrZABR(MC)
0.00	0.1086	0.1084	0.1175	0.1042	0.1147	0.1180	0.1461	0.1631	0.1049	0.0929	0.1272	0.1180
0.10	0.1189	0.1179	0.1301	0.1218	0.1222	0.1220	0.1499	0.1626	0.1145	0.1136	0.1389	0.1363
0.20	0.1253	0.1239	0.1380	0.1310	0.1274	0.1262	0.1529	0.1628	0.1205	0.1217	0.1462	0.1454
0.30	0.1299	0.1282	0.1436	0.1371	0.1313	0.1296	0.1554	0.1630	0.1248	0.1269	0.1513	0.1514
0.40	0.1333	0.1315	0.1478	0.1415	0.1343	0.1324	0.1573	0.1633	0.1281	0.1308	0.1551	0.1557
0.42	0.1360	0.1341	0.1509	0.1447	0.1367	0.1347	0.1587	0.1634	0.1308	0.1337	0.1580	0.1590
0.44	0.1381	0.1361	0.1533	0.1471	0.1386	0.1366	0.1598	0.1634	0.1329	0.1362	0.1603	0.1615
0.46	0.1398	0.1378	0.1551	0.1489	0.1402	0.1382	0.1605	0.1632	0.1347	0.1381	0.1621	0.1633
0.48	0.1411	0.1391	0.1563	0.1501	0.1414	0.1395	0.1609	0.1628	0.1362	0.1398	0.1634	0.1648
0.50	0.1421	0.1402	0.1571	0.1509	0.1424	0.1406	0.1610	0.1623	0.1375	0.1412	0.1644	0.1658
0.52	0.1430	0.1410	0.1576	0.1514	0.1431	0.1414	0.1608	0.1616	0.1386	0.1424	0.1651	0.1665
0.54	0.1436	0.1416	0.1577	0.1515	0.1437	0.1421	0.1605	0.1608	0.1395	0.1434	0.1656	0.1669
0.56	0.1440	0.1421	0.1575	0.1513	0.1441	0.1426	0.1599	0.1598	0.1403	0.1443	0.1658	0.1671
0.58	0.1443	0.1424	0.1571	0.1509	0.1444	0.1429	0.1591	0.1586	0.1409	0.1450	0.1659	0.1670
0.60	0.1445	0.1426	0.1564	0.1502	0.1445	0.1432	0.1581	0.1574	0.1415	0.1456	0.1658	0.1668
0.70	0.1445	0.1427	0.1556	0.1494	0.1446	0.1433	0.1570	0.1560	0.1419	0.1461	0.1655	0.1664
0.80	0.1445	0.1427	0.1545	0.1483	0.1445	0.1433	0.1557	0.1544	0.1423	0.1466	0.1651	0.1659
0.90	0.1443	0.1426	0.1532	0.1470	0.1444	0.1433	0.1542	0.1527	0.1426	0.1469	0.1646	0.1652
1.00	0.1441	0.1424	0.1518	0.1456	0.1441	0.1431	0.1526	0.1509	0.1428	0.1472	0.1640	0.1644
1.20	0.1438	0.1421	0.1502	0.1440	0.1438	0.1429	0.1509	0.1490	0.1429	0.1474	0.1633	0.1634
1.40	0.1434	0.1418	0.1485	0.1422	0.1434	0.1426	0.1490	0.1470	0.1430	0.1476	0.1625	0.1624
1.60	0.1430	0.1414	0.1466	0.1404	0.1430	0.1422	0.1471	0.1448	0.1431	0.1477	0.1616	0.1613
1.80	0.1425	0.1410	0.1446	0.1383	0.1425	0.1418	0.1450	0.1426	0.1431	0.1478	0.1607	0.1601

Table 3.4: Implied Bachelier volatility computed from the call option prices obtained from the effective equation and Monte Carlo simulation for the SABR, ZABR, fSABR, fZABR, mrSABR and mrZABR models. All strikes and volatilities are quoted as percentages.

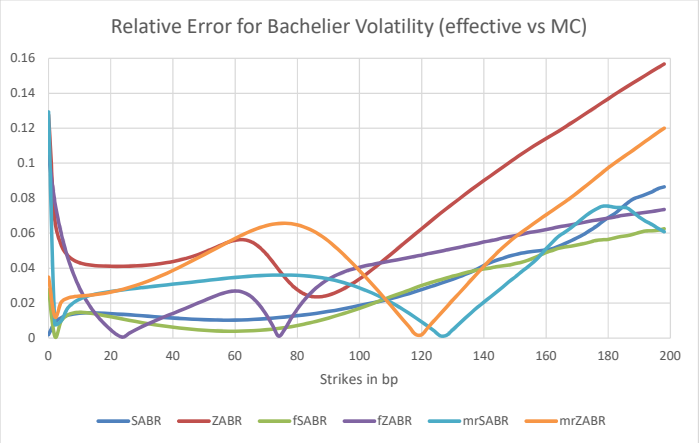


Figure 3.9: Relative error for the implied Bachelier volatility computed from the call option prices obtained from the effective equation and Monte Carlo simulation for the SABR, ZABR, fSABR, fZABR, mrSABR and mrZABR models.

3.4 Conclusion and Summary

We have outlined an approach applicable to a large class of stochastic volatility models. The approach is based on an effective PDE associated with the stochastic volatility model. In particular we derived a (one-dimensional) local volatility representation of the (two-dimensional) model. Then, we applied accurate and efficient numerical schemes to calculate option prices. General conditions for applicability were derived. We extended the modelling to include the ZABR model and introduced the free ZABR and mean-reverting ZABR models. The flexibility of the new models was explored and a numerical comparison with existing methods was given.

To derive our results we applied singular perturbation theory. There are, however, other techniques that may be used to derive a local volatility model associated with a stochastic volatility model, for instance the method of Markovian Projection, see, e.g. Piterbarg [2006], Antonov et al. [2008], and Antonov and Misirpashaev [2009], which is based on the result of Gyöngy [1986]. Further comparison and investigation of the applicability of these methods and application to other models is envisaged. From a numerical methods perspective, we applied the method considered by Hagan [2015], Le Floc'h and Kennedy [2017], and Kienitz et al. [2017]. While we have included a basic comparison to Monte Carlo simulation results, further comparison using numerical schemes for a two-dimensional PDE, as in

Sheppard [2007], as well as to the existing numerical scheme for the ZABR model, provided by Andreasen and Huge [2013], is possible. The generalization of the numerical scheme derived in the context of analysing the Heston model seems to be a good starting point for further analysing the quality of our approximation using effective equations.

3.5 Appendix: Derivations

3.5.1 Deriving the Effective Forward Equation

As described in Section 3.2, we analyse a stochastic volatility model of the form

$$\begin{cases} dF_t = \varepsilon C(F_t) v_t \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(v_t)\, dt + \varepsilon \nu(v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases}$$

Following Hagan et al. [2014, 2020b, 2018a], we define the probability density $p(t_0, f, \alpha, t, F, A)$ that $F_t = F$ and $v_t = A$ at time t, given that $F_{t_0} = f$ and $v_{t_0} = \alpha$ at time t_0. Furthermore, we define the moments

$$Q^{(k)}(t_0, f, \alpha, t, F) = \int_0^\infty A^k p(t_0, f, \alpha, t, F, A)\, dA$$

and set

$$Q(t, F) = Q^{(0)}(t_0, f, \alpha, t, F) \tag{3.5.1}$$

to be the reduced probability density of the model given t_0, f and α.

3.5.2 Volatility Drift

Before analysing the corresponding PDE equations, let us first take a look at the drift term of the volatility function. For this, we consider the following PDE given by the drift term

$$\begin{cases} \partial_t Y(t, t_0, \alpha) = \mu(Y(t, t_0, \alpha)) \\ Y(t, t, \alpha) = \alpha \\ Y(t_0, t_0, \alpha) = \alpha. \end{cases} \tag{3.5.2}$$

By Assumptions I and II we know that there exists a solution $Y(t, t_0, \alpha)$ and an inverse function $y(t_0, t, a)$, such that

$$Y(t, t_0, \alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0, t, a).$$

Considering the integrated form of the PDE (3.5.2), the derivatives of $Y(t, t_0, \alpha)$ satisfy

$$\partial_{t_0} Y(t, t_0, \alpha) = -\mu(\alpha) + \int_{t_0}^{t} \mu'(Y(s, t_0, \alpha)) \partial_{t_0} Y(s, t_0, \alpha)\, ds$$

$$\text{and} \qquad \partial_{\alpha} Y(t, t_0, \alpha) = 1 + \int_{t_0}^{t} \mu'(Y(s, t_0, \alpha)) \partial_{\alpha} Y(s, t_0, \alpha)\, ds.$$

This, in turn, implies that

$$\partial_{t_0} Y(t, t_0, \alpha) = -\mu(\alpha) \partial_{\alpha} Y(t, t_0, \alpha).$$

Remark 3.5.1. Note that in both cases of Remark 3.2.2 we have that Y coincides with the expected volatility:

$$Y(t, t_0, \alpha) = \mathbb{E}\big[v_t \,|\, v_{t_0} = \alpha\big].$$

If μ is a non-linear function, this would not be true any more, since we would have to exchange the expectation with the function μ, which in general cannot be done.

3.5.3 The Forward Equation

Now, we start by considering the Kolmogorov forward equation to get

$$\begin{aligned} \partial_t p(t, F, A) = &-\partial_A \Big[\mu(A) p(t, F, A)\Big] + \frac{1}{2}\varepsilon^2 \partial_{FF} \Big[C(F)^2 A^2 p(t, F, A)\Big] \\ &+\varepsilon^2 \rho \partial_{FA} \Big[C(F) A \nu(A) p(t, F, A)\Big] + \frac{1}{2}\varepsilon^2 \partial_{AA} \Big[\nu(A)^2 p(t, F, A)\Big], \end{aligned}$$

where we have abbreviated $p(t_0, f, \alpha, t, F, A)$ as $p(t, F, A)$. Integrating over A and considering reflecting boundary conditions to conserve the probability, as done for example in Hagan et al. [2014], we get

$$\int_0^{\infty} \partial_A \Big[\mu(A) p(t, F, A)\Big] dA = \Big[\mu(A) p(t, F, A)\Big]\Big|_0^{\infty} = 0$$

$$\int_0^{\infty} \partial_{FA} \Big[C(F) A \nu(A) p(t, F, A)\Big] dA = \partial_F \Big[C(F) A \nu(A) p(t, F, A)\Big]\Big|_0^{\infty} = 0$$

$$\int_0^{\infty} \partial_{AA} \Big[\nu(A)^2 p(t, F, A)\Big] dA = \partial_A \Big[\nu(A)^2 p(t, F, A)\Big]\Big|_0^{\infty} = 0.$$

With this, we get the forward equation

$$\begin{cases} \partial_t Q^{(0)}(t,F) = \frac{1}{2}\varepsilon^2 \partial_{FF}\Big[C(F)^2 Q^{(2)}(t,F)\Big] \\ Q^{(0)}(t,F) \to \delta(F-f) \text{ as } t \to t_0^+, \end{cases} \tag{3.5.3}$$

for $t > t_0$

3.5.4 The Backward Equation

Next, let us consider the Kolmogorov backward equation for $Q^{(k)}$ given by

$$\begin{cases} -\partial_{t_0} Q^{(k)} = \mu(\alpha)\partial_\alpha Q^{(k)} + \frac{1}{2}\varepsilon^2\alpha^2\Big[C(f)^2\partial_{ff}Q^{(k)} + 2\rho\frac{\nu(\alpha)}{\alpha}C(f)\partial_{f\alpha}Q^{(k)} \\ \qquad\qquad + \left(\frac{\nu(\alpha)}{\alpha}\right)^2 \partial_{\alpha\alpha}Q^{(k)}\Big] \\ Q^{(k)} \to \alpha^k \delta(F-f) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $Q^{(k)}(t_0, f, \alpha, t, F)$ as $Q^{(k)}$. To cancel the drift term, we change variables from α to $a = Y(t, t_0, \alpha)$. As seen in Section 3.5.2 the change of variables is provided by

$$\begin{aligned} &\partial_\alpha \to \partial_\alpha Y(t,t_0,\alpha)\partial_a = X(t,t_0,y(t_0,t,a))\partial_a, \\ &\partial_{t_0} \to \partial_{t_0} - \mu(\alpha)\partial_\alpha Y(t,t_0,\alpha)\partial_a = \partial_{t_0} - \mu(y(t_0,t,a))X(t,t_0,y(t_0,t,a))\partial_a. \end{aligned}$$

Here, we use the function X as an abbreviation of $\partial_\alpha Y(t,t_0,\alpha)$. With this the drift term vanishes and we get

$$\begin{cases} -\partial_{t_0} Q^{(k)} = \frac{1}{2}\varepsilon^2 y(t_0,t,a)^2\Big[C(f)^2\partial_{ff}Q^{(k)} + 2\rho\tilde{\nu}C(f)\partial_{f\alpha}Q^{(k)} \\ \qquad\qquad + \tilde{\nu}^2\partial_{\alpha\alpha}Q^{(k)}\Big] \\ Q^{(k)} \to a^k \delta(F-f) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $Q^{(k)}(t_0, f, a, t, F)$ as $Q^{(k)}$, and

$$\tilde{\nu} = \tilde{\nu}(t_0,t,a) = \frac{\nu(y(t_0,t,a))}{y(t_0,t,a)} X(t,t_0,y(t_0,t,a)).$$

Note that this equation corresponds to the one in Hagan et al. [2020b, A.19]. The only difference is the form of $\tilde{\nu}(t_0,t,a)$ where we have an additional dependence on

$y(t_0, t, a)$. Thus, until we reach the point where the explicit form of $\tilde{\nu}$ is needed, our reasoning is the same as that used by Hagan et al. [2018a, 2020b]. Nevertheless, for completeness, we briefly present the main steps.

We start by changing the variable f to

$$z = \frac{1}{\varepsilon} \int_f^F \frac{1}{C(u)}\, du \tag{3.5.4}$$

and re-scale $Q^{(k)}$ to be

$$Q^{(k)}(t_0, f, a, t, F) = \frac{a^k}{\varepsilon B(0)} \tilde{Q}^{(k)}(t_0, z, a, t, F) \tag{3.5.5}$$

with variables

$$B(\varepsilon z) = C(f) \qquad \text{and} \qquad \Gamma(\varepsilon z) = \frac{B'(\varepsilon z)}{B(\varepsilon z)}.$$

This yields the equation

$$\begin{cases} -\partial_{t_0} \tilde{Q}^{(k)} = y(t_0, t, a)^2 \Big[\frac{1}{2} \partial_{zz} \tilde{Q}^{(k)} - \frac{1}{2} \varepsilon \Gamma(\varepsilon z) \partial_z \tilde{Q}^{(k)} - \varepsilon k \rho \frac{\tilde{\nu}}{a} \partial_z \tilde{Q}^{(k)} \\ \qquad\qquad + \frac{1}{2} \varepsilon^2 k(k-1) \frac{\tilde{\nu}^2}{a^2} \tilde{Q}^{(k)} - \varepsilon \rho \tilde{\nu} \partial_{za} \tilde{Q}^{(k)} \\ \qquad\qquad + \varepsilon^2 k \frac{\tilde{\nu}^2}{a} \partial_a \tilde{Q}^{(k)} + \frac{1}{2} \varepsilon^2 \tilde{\nu}^2 \partial_{aa} \tilde{Q}^{(k)} \Big] \\ \tilde{Q}^{(k)} \to \delta(z) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $\tilde{Q}^{(k)}(t_0, z, a, t, F)$ as $\tilde{Q}^{(k)}$.

Now, to cancel the a dependent term $y(t_0, t, a)^2$ in front of every term, we change the time scaling from t_0 to s given by

$$s = S(t_0, t, a) = \int_{t_0}^t y(u, t, a)^2\, du.$$

Again, we consider the inverse denoted by $\tilde{t_0}$ such that $t_0 = \tilde{t_0}(s, t, a)$, where

$s = S(t_0, t, a)$. Then we get the new PDE

$$\begin{cases} \partial_s \tilde{Q}^{(k)} = \frac{1}{2}\partial_{zz}\tilde{Q}^{(k)} - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z\tilde{Q}^{(k)} - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\,\partial_{zs}\tilde{Q}^{(k)} + \partial_{za}\tilde{Q}^{(k)}\Big) \\ \qquad\qquad + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\,\partial_{ss}\tilde{Q}^{(k)} + \partial_{aa}s\,\partial_s\tilde{Q}^{(k)}\Big) - \varepsilon k\rho\frac{\tilde{\nu}}{a}\partial_z\tilde{Q}^{(k)} \\ \qquad\qquad + \varepsilon^2 k\frac{\tilde{\nu}^2}{a}\partial_a s\,\partial_s\tilde{Q}^{(k)} + \frac{1}{2}\varepsilon^2 k(k-1)\frac{\tilde{\nu}^2}{a^2}\tilde{Q}^{(k)} \\ \qquad\qquad + \varepsilon^2 k\frac{\tilde{\nu}^2}{a}\partial_a\tilde{Q}^{(k)} + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(2\partial_a s\,\partial_{sa}\tilde{Q}^{(k)} + \partial_{aa}\tilde{Q}^{(k)}\Big) \\ \tilde{Q}^{(k)} \to \delta(z) \text{ as } s \to 0^+, \end{cases} \tag{3.5.6}$$

where we have abbreviated $\tilde{Q}^{(k)}(s, z, a, t, F)$ as $\tilde{Q}^{(k)}$. Note that this equation corresponds to the ones in, for example, Hagan et al. [2020b, A.29] or Hagan et al. [2018a, A.28]. Now, we argue that to leading order we only have the heat equation

$$\begin{cases} \partial_s\tilde{Q}^{(k)} = \frac{1}{2}\partial_{zz}\tilde{Q}^{(k)} \\ \tilde{Q}^{(k)} \to \delta(z) \text{ as } s \to 0^+. \end{cases} \tag{3.5.7}$$

Thus if we expand $\tilde{Q}^{(k)} = \tilde{Q}_0^{(k)} + \varepsilon\tilde{Q}_1^{(k)} + \varepsilon^2\tilde{Q}_2^{(k)} + \cdots$ we can conclude that the first term $\tilde{Q}_0^{(k)}$ is given by the solution of (3.5.7) as

$$\tilde{Q}_0^{(k)}(s, z, a) = \tilde{Q}_0^{(k)}(s, z) = \frac{1}{\sqrt{2\pi s}}e^{-\frac{z^2}{2s}}.$$

In particular, $\tilde{Q}_0^{(k)}$ does not depend on a and thus $\partial_a\tilde{Q}^{(k)}$ is actually of order $\mathcal{O}(\varepsilon)$ and the last two terms of (3.5.6) are of order $\mathcal{O}(\varepsilon^3)$ and can be neglected.

3.5.5 Computing $\tilde{Q}^{(2)}$

With (3.5.6) we can set up the two PDEs for $\tilde{Q}^{(0)}$ and $\tilde{Q}^{(2)}$. Without explicitly writing them down, we can still see that for $k = 2$ the PDE contains all the terms of $k = 0$ and the additional terms of the last lines. Since our goal is to express $\tilde{Q}^{(2)}$ in terms of $\tilde{Q}^{(0)}$ and to insert this back into the forward equation, we use the ansatz

$$\tilde{Q}^{(2)}(s, z, a, t, F) = H(s, z, a, t, F)e^{2\varepsilon b(s,a)z + \varepsilon^2\tilde{c}(s,a)z^2 + \varepsilon^2 G(s,\alpha)}$$

and determine the coefficients in such a way that the PDE for H corresponds to the one of $\tilde{Q}^{(0)}$.

This yields the PDE given by

$$\begin{aligned}\partial_s H = &\frac{1}{2}\partial_{zz}H - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z H - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\,\partial_{zs}H + \partial_{za}H\Big)\\ &+\frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\,\partial_{ss}H + \partial_{aa}s\,\partial_s H\Big) - \Big(s\partial_s b + b - \frac{\rho\tilde{\nu}}{a}\Big)\frac{2\varepsilon z}{s}H\\ &-\Big(s\partial_s\tilde{c} + 2\tilde{c} - 2\rho\tilde{\nu}\Big(\partial_a b + \partial_s b\,\partial_a s\Big)\Big)\frac{\varepsilon^2 z^2}{s}H\\ &+\Big(-\partial_s G + \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}}{a}b + \frac{\tilde{\nu}^2}{a^2} - b\Gamma_0 - 2\rho\tilde{\nu}\Big(\partial_a b + \partial_s b\,\partial_a s\Big)\Big)\varepsilon^2 H\\ &+2\Big(b - \frac{\rho\tilde{\nu}}{a}\Big)\varepsilon\Big(\partial_z H + \frac{z}{s}H\Big) - 2\Big(\rho\tilde{\nu}b - \frac{\tilde{\nu}^2}{a}\Big)\varepsilon^2\partial_a s\,\partial_s H,\end{aligned} \tag{3.5.8}$$

where we have abbreviated $H(s,z,a,t,F)$ as H. Here $\Gamma_0 = \Gamma(0)$ is a suitable approximation of $\Gamma(\varepsilon z)$ with order $\mathcal{O}(\varepsilon^2)$. To bring the equation of H closer to the one for $\tilde{Q}^{(0)}$ we set b such that

$$s\partial_s b + b - \frac{\rho\tilde{\nu}}{a} = 0,$$

in which case b has the form

$$b(s,a) = \frac{1}{as}I_1(s,t,a)$$
$$\text{with} \qquad I_1 = I_1(s,t,a) = \rho\int_0^s \tilde{\nu}\big(\tilde{t_0}(x,t,a),t,a\big)\,dx.$$

Switching the integration variable from x to $u = \tilde{t_0}(x,t,a)$, $I_1(s,t,a)$ may be written as

$$\begin{aligned}I_1(s,t,a) &= \rho\int_{\tilde{t_0}(s,t,a)}^t \tilde{\nu}\big(u,t,a\big)y^2(u,t,a)\,du\\ &= \rho\int_{\tilde{t_0}(s,t,a)}^t \nu\big(y(u,t,a)\big)y(u,t,a)X(t,u,y(u,t,a))\,du.\end{aligned}$$

Note that here the only s dependence of the integral I_1 comes from the boundary $\tilde{t_0}(s,t,a)$. To handle the term containing $\varepsilon(\partial_z H + \frac{z}{s}H)$ we recall that

$$H_0(s,z) = \frac{1}{\sqrt{2\pi s}}e^{-\frac{z^2}{2s}}$$

and that $\partial_z H^0(s,z) + \frac{z}{s}H^0(s,z) = 0$. With this the PDE of order $\mathcal{O}(\varepsilon)$ is

$$\partial_s H = \frac{1}{2}\partial_{zz}H - \frac{1}{2}\varepsilon\Gamma_0\partial_z H - \varepsilon\rho\tilde{\nu}\partial_a s\,\partial_{zs}H.$$

As in Hagan et al. [2020b] we can show through the concrete form of H^0 that H^1 has the form

$$H^1(s,z) = -\frac{1}{2}s\Gamma_0\partial_z H^0 - I_3(s,t,a)\partial_{zzz}H^0$$
$$\text{with} \qquad I_3 = I_3(s,t,a) = \frac{1}{2}\rho\int_0^s \tilde{\nu}(\tilde{t_0}(x,t,a),t,a)\partial_a S(\tilde{t_0}(x,t,a),t,a)\,dx.$$

Using our new form of $\tilde{\nu}$ and the explicit form of $\partial_a S(\tilde{t_0}(x,t,a),t,a)$ given by

$$\begin{aligned}\partial_a S(\tilde{t_0},t,a) &= 2\int_{\tilde{t_0}}^t y(u,t,a)\partial_a y(u,t,a)\,du \\ &= 2\int_{\tilde{t_0}}^t y(u,t,a)X(t,u,y(u,t,a))^{-1}\,du,\end{aligned}$$

we get

$$\begin{aligned}I_3(s,t,a) = \rho\int_{\tilde{t_0}(s,t,a)}^t &\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a)) \\ &\times\int_u^t y(v,t,a)X(t,v,y(v,t,a))^{-1}\,dv\,du.\end{aligned}$$

Moreover, we can conclude as in Hagan et al. [2020b] that

$$\varepsilon\left(\partial_z H + \frac{z}{s}H\right) = \frac{3I_3}{s^2}\frac{\varepsilon^2 z^2}{s}H - \left(\frac{3I_3}{s^2} - \frac{1}{2}\Gamma_0\right)\varepsilon^2 H + \mathcal{O}(\varepsilon^3).$$

Before determining the parameters $\tilde{c}$ and G in (3.5.8) we rewrite the last term in terms of H. For this note that to order $\mathcal{O}(\varepsilon^2)$ we have

$$\varepsilon^2\partial_s H = \frac{1}{2}\varepsilon^2 s\partial_{zz}H = \frac{1}{2}\varepsilon^2\left(\frac{z^2}{s^2} - \frac{1}{s}\right)H.$$

With this, the PDE in (3.5.8) simplifies to

$$\begin{aligned}\partial_s H = &\frac{1}{2}\partial_{zz}H - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z H - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\,\partial_{zs}H + \partial_{za}H\Big)\\ &+\frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\,\partial_{ss}H + \partial_{aa}s\,\partial_s H\Big)\\ &-\Big[s\partial_s\tilde{c} + 2\tilde{c} - 2\rho\tilde{\nu}\partial_a b + 6\frac{I_3}{s}\partial_s b\\ &\quad - \Big(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\Big)\frac{\partial_a s}{s}\Big]\frac{\varepsilon^2 z^2}{s}H\\ &+\Big[-\partial_s G + \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}}{a}b + \frac{\tilde{\nu}^2}{a^2} - \frac{\rho\tilde{\nu}}{a}\Gamma_0 - 2\rho\tilde{\nu}\partial_a b\\ &\quad + 6\frac{I_3}{s}\partial_s b - \Big(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\Big)\frac{\partial_a s}{s}\Big]\varepsilon^2 H.\end{aligned}$$

3.5.5.1 Computing $\tilde{c}$

From the above equation it is clear that we should choose $\tilde{c}$ in order to cancel the $\frac{\varepsilon^2 z^2}{s}H$ term. This means we set

$$s\partial_s\tilde{c} + 2\tilde{c} - 2\rho\tilde{\nu}\partial_a b + 6\frac{I_3}{s}\partial_s b - \Big(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\Big)\frac{\partial_a s}{s} = 0.$$

Multiplying by s and summarizing the terms yields

$$\begin{aligned}\partial_s\big(s^2\tilde{c}\big) + 6I_3\partial_s b &= 2s\rho\tilde{\nu}\partial_a b + \partial_a s\frac{\tilde{\nu}^2}{a} + \Big(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b\Big)\partial_a s\\ &= 2s\rho\tilde{\nu}\partial_a b + \partial_a s\frac{\tilde{\nu}^2}{a} + \Big(2\rho\tilde{\nu}\partial_s\big(sb\big) - 3\rho\tilde{\nu}b\Big)\partial_a s.\end{aligned}$$

Taking into consideration that I_3 was chosen to satisfy

$$\partial_s I_3 = \frac{1}{2}\rho\tilde{\nu}\partial_a s,$$

we can conclude that c must solve

$$\partial_s\big(s^2\tilde{c}\big) = -6\partial_s\big(I_3 b\big) + \frac{\tilde{\nu}^2}{a}\partial_a s + 2\rho\tilde{\nu}\Big(\partial_s\big(sb\big)\,\partial_a s + s\partial_a b\Big).$$

To explicitly solve this, let us first consider the last term of this equation. We get

$$\begin{aligned}\partial_s\big(sb\big)\,\partial_a s + s\partial_a b &= \partial_s\Big(\frac{1}{a}I_1\Big)\,\partial_a s + \partial_a\Big(\frac{1}{a}I_1\Big)\\ &= \frac{1}{a}\partial_a S\partial_s I_1 - \frac{1}{a^2}I_1 + \frac{1}{a}\partial_a I_1. \qquad (3.5.9)\end{aligned}$$

Recall that $I_1(s,t,a)$ is defined as

$$I_1(s,t,a) = \rho \int_{\tilde{t_0}(s,t,a)}^{t} \nu\big(y(u,t,a)\big) y(u,t,a) X(t,u,y(u,t,a))\, du.$$

In particular, this means for the a-derivative that we have

$$\begin{aligned}\partial_a I_1(s,t,a) = &-\partial_a \tilde{t_0}(s,t,a)\rho\tilde{\nu}(\tilde{t_0}(s,t,a),t,a)y^2(\tilde{t_0}(s,t,a),t,a) \\ &+\rho \int_{\tilde{t_0}(s,t,a)}^{t} \partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\, du.\end{aligned}$$

Following Hagan et al. [2020b] we can see that we have

$$\partial_a S(\tilde{t_0}(s,t,a),t,a) = y^2(\tilde{t_0}(s,t,a),t,a)\partial_a \tilde{t_0}(s,t,a).$$

Thus (3.5.9) reduces to

$$\begin{aligned}&\partial_s\big(sb\big)\,\partial_a s + s\partial_a b \\ &\quad= -\frac{1}{a^2} I_1(s,t,a) + \frac{1}{a}\rho \int_{\tilde{t_0}(s,t,a)}^{t} \partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\, du \\ &\quad= -\frac{1}{a} sb + \frac{1}{a}\rho \int_{\tilde{t_0}(s,t,a)}^{t} \partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\, du,\end{aligned}$$

and yields

$$\begin{aligned}&2\rho\tilde{\nu}\Big[\partial_s\big(sb\big)\,\partial_a s + s\partial_a b\Big] = \\ &\quad-\partial_s\big(s^2b^2\big) + \frac{2\rho\tilde{\nu}}{a}\rho \int_{\tilde{t_0}(s,t,a)}^{t} \partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\, du.\end{aligned}$$

Switching from $\tilde{c}$ to $c = \tilde{c} + 2b^2$ we finally get

$$\begin{aligned}\partial_s\big(s^2c\big) = &-6\partial_s\big(I_3 b\big) + \frac{\tilde{\nu}^2}{a}\partial_a s + \partial_s\big(s^2b^2\big) \\ &+\frac{2\rho\tilde{\nu}}{a}\rho \int_{\tilde{t_0}(s,t,a)}^{t} \partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\, du.\end{aligned}$$

Now, we only need to integrate over s to conclude that

$$s^2c(s,a) = s^2b^2 + \frac{1}{a} I_2(s,t,a) - 6I_3(s,t,a)b + \frac{2}{a} I_4(s,t,a).$$

Here the integral functions are given by

$$
\begin{aligned}
&I_2(s,t,a)\\
&\quad= \int_0^s \partial_a S\big(\tilde{t_0}(x),t,a\big)\tilde{\nu}\big(\tilde{t_0}(x),t,a\big)^2\,dx\\
&\quad= \int_{\tilde{t_0}(s,t,a)}^t \partial_a S\big(u,t,a\big)\nu(y(u,t,a))^2 X(t,u,y(u,t,a))^2\,du\\
&\quad= 2\int_{\tilde{t_0}(s,t,a)}^t \nu(y(u,t,a))^2 X(t,u,y(u,t,a))^2 \int_u^t y(v,t,a)\partial_a y(v,t,a)\,dv\,du\\
&\quad= 2\int_{\tilde{t_0}(s,t,a)}^t \nu(y(u,t,a))^2 X(t,u,y(u,t,a))^2 \int_u^t y(v,t,a) X(t,v,y(v,t,a))^{-1}\,dv\,du
\end{aligned}
$$

and

$$
\begin{aligned}
&I_4(s,t,a)\\
&\quad= \int_0^s \rho\tilde{\nu}(\tilde{t_0}(x,t,a),t,a)\rho \int_{\tilde{t_0}(x,t,a)}^t \partial_a\Big(\nu(y(v,t,a))y(v,t,a)X(t,v,y(v,t,a))\Big)\,dv\,dx\\
&\quad= \rho^2\int_{\tilde{t_0}(s,t,a)}^t \psi(u,t,a)\int_u^t \partial_a\psi(v,t,a)\,dv\,du.
\end{aligned}
$$

Here the function ψ is defined as

$$\psi(u,t,a) = \nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a)).$$

3.5.5.2 Computing G

Analogously to determining $\tilde{c}$, we set G to cancel the remaining $\varepsilon^2 H$ term. This means

$$
\begin{aligned}
\partial_s G &= \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}b}{a} + \frac{\tilde{\nu}^2}{a^2} - \frac{\rho\tilde{\nu}}{a}\Gamma_0 - 2\rho\tilde{\nu}\partial_a b + 6\frac{I_3}{s}\partial_s b\\
&\quad -\left(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\right)\frac{\partial_a s}{s}\\
&= \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}b}{a} + \frac{\tilde{\nu}^2}{a^2} - \frac{\rho\tilde{\nu}}{a}\Gamma_0 - s\partial_s\tilde{c} - 2\tilde{c}\\
&= -\partial_s(sc) - \partial_s(sb)\Gamma_0 + \frac{\tilde{\nu}^2}{a^2}.
\end{aligned}
$$

Now we can integrate over s and conclude that

$$G(s,a) = -sc - sb\Gamma_0 + \frac{I_5(s,t,a)}{a^2},$$

with

$$I_5(s,t,a) = \int_0^s \tilde{\nu}(\tilde{t_0}(x,t,a),t,a)^2\,dx = \int_{\tilde{t_0}(s,t,a)}^t \nu(y(v,t,a))^2 X(t,u,y(u,t,a))^2\,du.$$

3.5.6 The Effective Forward Equation

With these choices we have found the following PDE for H:

$$\begin{cases} \partial_s H = \frac{1}{2}\partial_{zz}H - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z H - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\,\partial_{zs}H + \partial_{za}H\Big) \\ \qquad\qquad +\frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\,\partial_{ss}H + \partial_{aa}s\,\partial_s H\Big) \\ H \to \delta(z) \text{ as } s \to 0^+, \end{cases}$$

where we have abbreviated $H(s,z,a,t,F)$ as H. Since this is the same equation as for $\tilde{Q}^{(0)}(s,z,a,t,F)$, we can identify $H(s,z,a,t,F)$ as $\tilde{Q}^{(0)}(s,z,a,t,F)$ to order $\mathcal{O}(\varepsilon^2)$. So, we have

$$H(s,z,a,t,F) = \tilde{Q}^{(0)}(s,z,a,t,F).$$

Thus we have shown that

$$\tilde{Q}^{(2)}(s,z,a,t,F) = \tilde{Q}^{(0)}(s,z,a,t,F)e^{2\varepsilon b(s,a)z+\varepsilon^2\tilde{c}(s,a)z^2+\varepsilon^2 G(s,a)},$$

and by further approximating the exponential function we can conclude that

$$\tilde{Q}^{(2)}(s,z,a,t,F) = \tilde{Q}^{(0)}(s,z,a,t,F)e^{\varepsilon^2 G(s,a)}(1+2\varepsilon b(s,a)z+\varepsilon^2 c(s,a)^2 z^2). \quad (3.5.10)$$

Now, to express this in terms of the original variables α and t_0, simply recall our previous changes of variable

$$\begin{aligned} a &= Y(t,t_0,\alpha) \\ \text{and} \qquad s &= \int_{t_0}^t y(u,t,a)^2\,du. \end{aligned}$$

Furthermore, recall that y was defined as the inverse of Y. Thus, we can express $y(u,t,a)$ as

$$y(u,t,a) = y(u,t,Y(t,t_0,\alpha)) = Z(t,u,t_0,\alpha).$$

With this we can express s in terms of α as

$$s = \int_{t_0}^t Z(t,u,t_0,\alpha)^2\,du.$$

Making these changes in the integral functions I_1–I_5 yields those found in Assumption III.

4 Effective Markovian Projection

Building on the results of Chapter 3, this chapter aims to further deepen our understanding of the possible applications which may be derived using the approximation techniques outlined in Chapter 3. Special focus is placed on the pricing of basket options, which are derivatives depending on multiple rates. This extends the applicability of our techniques by allowing the consideration of multidimensional financial derivatives.

To motivate our research, we recall the current set-up resulting from Chapter 3. We deduced a general arbitrage free computational framework which allows explicit pricing formulas for European call options. Using these European call options, we calibrate our model to observable market data. This provides us with a fully parametrized financial market model. From a practical perspective we now progress and apply this model to evaluate an existing portfolio of assets. Even though explicit formulas are available in the case of European call options, this does not guarantee that the valuation of other financial instruments can be done with ease. One class of instruments which are very common in practice are options on baskets respectively the spread of individual rates. The motivation behind this is very intuitive since an option on the difference between two interest rates allows to control the risk coming from varying interest rate term structures.

From a practical perspective, the pricing of derivatives depending on baskets is not trivial. Having a financial market model available for each individual rate, it is in general not clear which dynamics govern the basket. For special types of models such as Black baskets, see Antonov [2020] or the basket of nSABR models, see Hagan et al. [2021a,b] it is possible to show that the baskets are governed by the same model structure as the individual rates which in turn allows for explicit pricing formulas for basket options.

Therefore, our motivation for this piece of research is to extend our existing computational framework to a multidimensional set-up where each individual rate is governed by a financial market model in accordance to Chapter 3. As observed in

Section 3.3.2, the possibility to model a single rate with a ZABR model compared to an SABR model provides more control in the calibration procedure and can be used to achieve a better fit to these more advanced products such as CMS spreads. We analyse these calibration advantages in more detail and show how it is possible to exploit these in a multidimensional set-up, applying explicit pricing formulas for basket products.

Our work is inspired by the work of Hagan et al. [2021a,b] where explicit formulas for CMS spread options are derived using normal SABR models for the individual rates. The core idea behind this work is to consider the basket of individual rates and perform an approximation using singular perturbation techniques of the basket to characterize it through a suitable model. The techniques presented therein are, however, not directly applicable to other stochastic volatility (SV) models using e.g. ZABR models for the individual rates. The reason behind this are additional drift terms. In Chapter 3 the model is formulated using a martingale measure. When considering a single rate, this is no restriction and can be achieved using a change of measure. When multiple rates are involved, however, in most cases it becomes impossible to find a measure which becomes a martingale measure for all considered rates at the same time. Hence, additional drift terms appear and the model does not satisfy the conditions for the techniques in Chapter 3 any more.

To overcome this problem, we do not consider a direct approximation of the basket itself, and instead focus on the individual rates. For this we project the models individually to another SV model. In Chapter 3 we already derived the projection to certain types of SABR models and we extend this projection to any desired SV model. For our application we have a projection onto a normal SABR in mind. To achieve this, we combine the techniques outlined in Chapter 3 with the Markovian projection technique. We call the resulting technique the Effective Markovian Projection (EMP). We present multiple algorithms to implement the EMP. This allows a fast and accurate approximation from one SV model onto another and, using the results based on baskets of normal SABR models, we can explicitly price basket options. We demonstrate this using CMS spread options as well as call options on mid-curves. In addition we present explicit numerical results to show how the additional flexibility in the calibration of the individual rates can be transferred to a better control on the prices of basket options. The projection to other SV models also provides other benefits depending on the chosen model and we use this to derive explicit approximation formulas for the probability density function of a model using the Johnson's SU distribution and a moment matching algorithm.

The rest of this chapter presents the publication Felpel et al. [2022a] starting from its introduction. This is an Accepted/Original Manuscript of an article published

by Taylor & Francis Group in Quantitative Finance on 18.03.2022, available online: https://doi.org/10.1080/14697688.2022.2043558.

4.1 Introduction

Markovian projection was introduced to the area of quantitative finance by Piterbarg [2006], and Andersen and Piterbarg [2010a], where it was originally applied to produce closed-form approximations for European option prices on basket models for stocks, multi-factor interest rate models and hybrid interest rate/FX models. Since then the method has been extended and refined. The Heston stochastic volatility model has been considered by Antonov et al. [2008] under the assumption of zero correlation between the Brownian motions driving the asset and the volatility processes. The displaced diffusion model [Antonov and Misirpashaev, 2009] and the SABR-LIBOR model [Tsuchiya, 2015] have also been investigated. Other works on the standard SABR model include Karlsmark [2013], and Kienitz and Wetterau [2013].

In the present paper we use results from our recent work on general stochastic volatility models [Felpel et al., 2021] to extend the scope of application for Markovian projection.

In particular, our novel approach allows pricing (and calibration) of multi-asset options using a large range of time-homogeneous stochastic volatility models with asset dynamics specified by

$$\begin{cases} dF_t = v_t C(F_t)\, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(v_t)\, dt + \nu(v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho\, dt. \end{cases} \tag{4.1.1}$$

This class of stochastic differential equations (SDEs) covers all of the standard stochastic volatility models widely applied in practice. The model and parameters, as specified by the functions $C(\cdot)$, $\mu(\cdot)$ and $\nu(\cdot)$, are chosen to ensure the best fit to the current (discrete) market implied volatility surface and provide dynamics that are suitable for risk management and hedging of exotic contracts. The generality of the specification allows the flexibility and control necessary to ensure good calibration under varying market conditions. In our applications, we focus on multi-rate interest rate derivatives and consider constant maturity swap (CMS) spread and mid-curve options as primary examples.

In summary, our research objectives are threefold: We provide a new Markovian projection technique, showcase applications on interest rate derivatives with multiple underlying rates and provide new results on moment-matching techniques outlining their application on general stochastic volatility models.

Our approach, to which we give the name Effective Markovian Projection, allows the projection of any general stochastic volatility model onto another model within the class specified by (4.1.1). This entails identifying the stochastic volatility model that, in a sense yet to be specified, best approximates the original dynamics. In Felpel et al. [2021] we derived an effective partial differential equation (PDE) for the dynamics specified by (4.1.1). This PDE may be interpreted in terms of a parametric (Bachelier) local volatility model [Dupire, 1994] using Markovian projection [Gyöngy, 1986]. Owing to a common underlying structure of the effective PDE, it is possible to find parametrizations for specific stochastic volatility models that provide good approximations to the general model.

When considering the application of Markovian projection to multi-factor baskets, we shall choose the reference model in such a way that it allows the effective modelling of general basket dynamics. To this end, we focus on the normal SABR model as our reference model since it can be shown that under a suitable numéraire the basket dynamics of normal SABR models is again a normal SABR model [Hagan et al., 2021b,a]. Having specified how to proceed in terms of the projection and the choice of the reference model, we provide numerical recipes for applying the method.

The remainder of the paper consists of five sections. Section 4.2 provides the main theoretical results. After briefly summarizing the theoretical background of Markovian projection [Gyöngy, 1986], we introduce Effective Markovian Projection, which is based on effective probability theory [Felpel et al., 2021, Hagan et al., 2014]. The numerical methods used for computing the projection are also introduced, in particular ATM-Matching, Minimal Point Matching and N-Point Matching. It should be emphasized that Effective Markovian Projection is directly applicable in the full general stochastic volatility setting. This is in contrast to classical Markovian projection, which must be tailored to the specific process, as seen for the general SABR model. It is known that this may produce approximations for volatility dynamics that are crude for long-dated instruments [Kienitz and Wetterau, 2013]. A brief analysis of the approximation quality is given. We also describe an Effective Markovian Projection based on moment matching for the case of the normal SABR model. Moment matching is applied to both NIG distributions and the family of Johnson distributions. The former approach has been considered previously by Charvet and Ticot [2011], and Eriksson et al. [2009], and we show that there are some limitations in its application. The approach based

on Johnson distributions provides a more stable and accurate result. The close connection of certain Johnson distributions to distributions arising from a SABR model has previously been recognized by Choi et al. [2019]. Section 4.3, focuses on examples from interest rate derivatives. Fixing notation and definitions, we consider CMS spread and mid-curve options as our main examples. With the underlying definitions at hand we consider a ZABR-type model as our base model and the normal SABR model as our reference model in Section 4.4, where we derive the corresponding projection and basket (spread) dynamics. Here we also specifically address numéraire issues. The resulting model amalgamates all the features of the base model and parametrizes the implied volatility surface in terms of a normal SABR model for each maturity, enabling the pricing of CMS spread and mid-curve options. Section 4.5 provides numerical illustrations of all the methods considered in the previous sections. In particular, we demonstrate the additional flexibility in modelling and calibrating underlying rates in a multi-rate framework when considering base models of the ZABR type. We also illustrate the effects of changing model parameters. Section 4.6 concludes by summarizing the results and giving prospects for future research.

4.2 Effective Markovian Projection

In this section we present our main result and introduce a new technique called Effective Markovian Projection (EMP). The idea underlying this technique is to combine the results of singular perturbation techniques (see e.g. Felpel et al. [2021], Hagan et al. [2018a]) with the technique of Markovian projection (see e.g. Gyöngy [1986], Andersen and Piterbarg [2010c]). To begin with we present the underlying mathematical framework necessary to apply the technique.

4.2.1 Mathematical Framework

Since our focus in this paper is on applications related to interest rate derivatives, we specify our modelling framework in a manner consistent with standard literature in the area, such as Andersen and Piterbarg [2010a]. Consider the abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ together with an augmented filtration $\{\mathcal{F}_t\}$ over a finite time horizon $[0, T]$.[1] For our applications we consider the n-dimensional real space $\mathbb{R}^n$ as our sample space, Ω, and assume the filtration to be generated by a

[1] Here T may be interpreted as the maturity of a claim. If several maturities are considered, T is set to the maximum of these times.

d-dimensional Brownian motion W_t, i.e., of the form

$$\mathcal{F}_t = \sigma\left\{W_u | u \leq t\right\},$$

augmented to be complete and right-continuous.

On this probability space we consider general Itô processes, X_t, specified by a system of SDEs of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t. \tag{4.2.1}$$

with adapted functions $\mu : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times d}$ such that

$$\begin{aligned} &\int_0^T |\mu(s,x)|\, ds < \infty, \\ &\int_0^T |\sigma(s,x)|^2\, ds < \infty \qquad a.s. \end{aligned}$$

Furthermore, we assume that there exists a unique strong solution in the sense of Øksendal [2003] to this SDE.

Remark 4.2.1. The topic of existence and uniqueness of solutions to SDE is of vast complexity and can heavily depend on the specific coefficients included in the equations. There are some general results available, e.g., assuming that the coefficients are Lipschitz and follow a growth condition of the form

$$\begin{aligned} |\mu(t,x)| + |\sigma(t,x)| &\leq K(1+|x|) \\ |\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| &\leq L|x-y|, \end{aligned}$$

for some constants constants K, $L > 0$. Then (4.2.1), with $X_0 = x_0$, admits a unique and almost sure continuous solution such that X_t is $\mathcal{F}_t$-adapted, see Øksendal [2003, Theorem 5.2.1].

Remark 4.2.2. In most cases one makes the assumptions of Remark 4.2.1 to guarantee the existence and uniqueness of the SDEs. This is, however, not universally applicable to all stochastic volatility models. For instance, in the SABR model, where the function $C(\cdot)$ of (4.1.1) is given by $C(x) = x^\beta$ for $\beta \in (0,1)$, it is apparent that the global Lipschitz condition is not satisfied. The SABR model is closely related to the CEV process, for which exact distributions are available, see Lindsay and Brecher [2012], which allows consideration of the SABR model, see Horvath and Reichmann [2018]. Given the broad range of SDEs used as financial models, see e.g., Andersen and Piterbarg [2010a,b,c], there are cases that may not be covered by these considerations. The area of degenerate SDEs with non-Lipschitz coefficients is being actively researched, see Suresh Kumar [2013], Chuni [2020].

There is, however, the possibility that none of these results is applicable to a chosen model. Then, of course, one should use the techniques we explore in this paper with caution. Paraphrasing Taylor [2010], one then operates under the typical, but possibly dangerous, assumption that a unique solution exists and performs the computation anyway, treating the results with a healthy dose of scepticism.

4.2.2 Effective Markovian Projection for Generalized Stochastic Volatility

Before introducing EMP, let us first recall classical Markovian projection, which is based on the results in Gyöngy [1986], and Andersen and Piterbarg [2010c]. We consider a one-dimensional process Y_t on a probability space equipped with a filtration constructed in accordance with Section 4.2.1 where the process is governed by the SDE

$$dY_t = \beta(t,\omega)^\top \, dW_t. \tag{4.2.2}$$

Here $\beta(\cdot,\cdot)$ is a very general d-dimensional $\mathcal{F}_t$-adapted process (hence the dependence on $\omega \in \Omega$) whose norm is uniformly bounded away from 0 and (4.2.2) admits a unique and strong solution. Then by Andersen and Piterbarg [2010c, Corollary A.1.3] there exists an SDE

$$dx_t = b(t,x_t)\, d\tilde{W}_t,$$

in terms of a one-dimensional Brownian motion $\tilde{W}_t$, and non-random coefficient $b(\cdot,\cdot)$ satisfying

$$b(t,x)^2 = \mathbb{E}[\beta(t,\omega)^\top \beta(t,\omega) | Y_t = x],$$

such that x_t admits a weak solution having the same one-dimensional probability distribution as Y_t. This mapping of one model onto another is known as Markovian projection.

Putting this result in the context of financial applications, the existence of the weak solution makes it possible to price European options. We shall call the function, $b(t,x)$, the projected volatility. In a certain sense it is the local volatility function for a Bachelier model in analogue to the results of Dupire [1994], and Derman and Kani [1994], who define the local volatility function in the context of a log-normal model. Throughout the remainder of the text, we define the projected volatility as

$$\sigma_{\text{proj}}(t,x) := b(t,x), \tag{4.2.3}$$

and, when referring to the squared quantity, call it the projected variance.

The main challenge for practical applications then lies in the proper evaluation of the conditional expectation. Possible approaches are based on Gaussian or least-square approximations, see Andersen and Piterbarg [2010c]. We call the approach using these approximation techniques classical Markovian projection (CMP). Here we propose an alternative approach based on effective probability theory, see e.g. Felpel et al. [2021], Hagan et al. [2014], which allows fast and accurate approximation. To motivate our new technique, we consider general time-homogeneous stochastic volatility models of the form

$$\begin{cases} dF_t = v_t C(F_t)\, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(v_t)\, dt + \nu(v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho\, dt, \end{cases} \tag{4.2.4}$$

which satisfy the assumptions of Section 4.2.1 and, in addition, the constraints imposed by Theorem 1 of Felpel et al. [2021]. For convenience these assumptions can be found in Appendix 4.7.1. With this setup, we provide the central result of effective Markovian projection.

Proposition 4.2.3. *Given a general stochastic volatility model* (4.2.4)*, satisfying the assumptions of Theorem 1 of Felpel et al. [2021], the square of the projected volatility function* (4.2.3) *is approximated as*

$$\sigma_{\text{proj}}^2(t,x) \approx C(x)^2 a(t)^2 e^{G(t)} \big(1 + 2b(t)z(x) + c(t)z(x)^2\big), \tag{4.2.5}$$

where the coefficients a, b, c, z and G are specified in Appendix 4.7.1.

We call this projection from a general stochastic volatility model onto a (Bachelier) local volatility model the ***Effective Markovian Projection*** *(EMP).*

Proof. Let $p(t_0, f, \alpha, t, F, A)$ be the probability density function that $F_t = F$ and $v_t = A$ at time t, given that $F_{t_0} = f$ and $v_{t_0} = \alpha$ at time t_0. With this we can describe the conditional probability

$$\begin{aligned} \sigma_{\text{proj}}^2(t,x) = C(x)^2 \mathbb{E}\big[v_t^2 \big| F_t = x\big] &= C(x)^2 \frac{\mathbb{E}\big[v_t^2 \mathbb{I}_{\{F_t = x\}}\big]}{\mathbb{E}\big[\mathbb{I}_{\{F_t = x\}}\big]} \\ &= C(x)^2 \frac{\int_0^\infty A^2 p(t_0, f, \alpha, t, x, A)\, dA}{\int_0^\infty p(t_0, f, \alpha, t, x, A)\, dA} \\ &= C(x)^2 \frac{Q^{(2)}(t_0, f, \alpha, t, x)}{Q^{(0)}(t_0, f, \alpha, t, x)}, \end{aligned}$$

where the function $Q^{(k)}$ is defined by

$$Q^{(k)}(t_0, f, \alpha, t, F) := \int_0^\infty A^k p(t_0, f, \alpha, t, F, A)\, dA.$$

Applying Theorem 1 of Felpel et al. [2021] we have

$$Q^{(2)}(t_0, f, \alpha, t, x) \approx P(t, x) Q^{(0)}(t_0, f, \alpha, t, x),$$

with

$$P(t, x) = a(t)^2 e^{G(t)} \big(1 + 2b(t)z(x) + c(t)z(x)^2\big), \tag{4.2.6}$$

where the coefficients a, b, c, z and G are specified in Felpel et al. [2021] and provided in Appendix 4.7.1. □

In the context of singular perturbation theory, this approximation has a truncation error of $\mathcal{O}(\varepsilon^2)$, which is the same order of accuracy provided by the SABR implied volatility formula (see Hagan et al. [2014]). To conclude this section, we provide further remarks on the applicability of the technique.

Remark 4.2.4. Although we have formulated EMP in terms of a stochastic volatility model, it may also be applied to a stochastic variance model (e.g. the Heston model). In this case the projected variance is expressed as

$$\sigma^2_{\text{proj}}(t, x) = C(x)^2 \frac{Q^{(1)}(t_0, f, \alpha, t, x)}{Q^{(0)}(t_0, f, \alpha, t, x)}.$$

As shown by Hagan et al. [2018a], in this setting an analogous representation of

$$Q^{(1)} = P(t, x) Q^{(0)}(t_0, f, \alpha, t, x)$$

is applicable, yielding the representation of the projected volatility function as in Proposition 4.2.3.

Remark 4.2.5. We stress that the function $C(\cdot)$ need not be restricted to be a parametric function. Choosing a non-parametric function allows the application of EMP to stochastic local volatility models (see Alexander and Nogueira [2008], Göttker-Schnetmann and Spanderen [2015], Saporito et al. [2019] for a non-exhaustive list).

Remark 4.2.6. To guarantee that the resulting model is arbitrage-free, it is enough to ensure that the projected volatility function remains positive. Assuming that $C(\cdot) > 0$, one then requires the additional constraint that $P(\cdot, \cdot) > 0$. The function $C(\cdot)$ determines the movement of the at-the-money volatility point when the forward moves, this is the case for all models used in practical applications. Market practitioners call this function the backbone.

4.2.3 Projection onto another Model

Using EMP we wish to project one model, which we call the *base model*, onto another which we call the *projection model*. This is useful because some models are more tractable than others, e.g., possessing analytical solutions for various contingent claims, or because modelling becomes simpler, e.g., in the case of basket options using Gaussian models.

The idea is to consider matching the corresponding projected volatility functions of both models. To this end, let us consider two projected variances σ^2_{proj} and $\tilde{\sigma}^2_{\text{proj}}$ of the form (4.2.5)

$$\sigma^2_{\text{proj}}(t,x) = C(x)^2 a(t)^2 e^{G(t)}\big(1 + 2b(t)z(x) + c(t)z(x)^2\big) \tag{4.2.7}$$

$$\tilde{\sigma}^2_{\text{proj}}(t,x) = \tilde{C}(x)^2 \tilde{a}(t)^2 e^{\tilde{G}(t)}\big(1 + 2\tilde{b}(t)\tilde{z}(x) + \tilde{c}(t)\tilde{z}(x)^2\big), \tag{4.2.8}$$

where the relationship between the coefficients in the underlying mathematical models is to be determined using EMP. In most applications, the projection model, specified by $\tilde{\sigma}^2_{\text{proj}}(\cdot,\cdot)$, will, in some sense, allow simplification of the original (base) problem, specified by $\sigma^2_{\text{proj}}(\cdot,\cdot)$.

We shall explore the case where the projection model is the displaced SABR (dSABR) model, which has the form

$$\begin{cases} dF_t = v_t(F_t + d)^\beta \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \nu v_t \, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases} \tag{4.2.9}$$

We assume that the parameters for the function $\tilde{C}(\cdot)$, i.e., d and β, are specified prior to projection. Later, most of our applications will use normal SABR (nSABR) as the projection model, which is just the special case of the above system of SDEs where $\beta = 0$ and d is irrelevant.

At this stage of the analysis, the coefficients and parameters of the base model, with projected variance given by (4.2.7), are left unspecified. Our goal is to find values for the remaining three parameters of the dSABR model (ν, α and ρ) in such a way that the resulting variance in (4.2.8), matches (4.2.7) as closely as possible. This may be achieved by proceeding in two steps:

1) We start by determining the relationships between the coefficient functions of the volatility function in (4.2.8) and those specifying (4.2.7), so as to ensure the best possible fit between the two models. For reasons that will become apparent when we implement step two, we do not work directly on

the expression (4.2.8) but instead use a slightly modified specification of the variance given by

$$\tilde{\sigma}^2_{\text{proj}}(t,x) = \tilde{C}(x)^2\hat{a}(t)^2\big(1 + 2\tilde{b}(t)\tilde{z}(x) + \tilde{c}(t)\tilde{z}(x)^2\big). \tag{4.2.10}$$

Since we have assumed that $\tilde{C}(\cdot)$ is pre-specified, $\tilde{z}(\cdot)$ is also known, and we need only match the remaining coefficient functions $\hat{a}(\cdot)$, $\tilde{b}(\cdot)$ and $\tilde{c}(\cdot)$. We describe three methods for specifying these coefficient functions. The first two methods consider the well posed problem of solving for the three parameters by specifying three simultaneous equations. The third approach is a numerical fit, which minimizes the error between the projected variances.

2) In the second step we consider the problem of determining the underlying model parameters based on the relationships of the coefficients $\hat{a}(\cdot)$, $\tilde{b}(\cdot)$ and $\tilde{c}(\cdot)$ found in step 1.

In what follows we present three different methodologies that may be used to match the projected variances in accordance with Step 1. Thereafter, in Section 4.2.3.4, we explore the problem of determining the parameters in Step 2.

4.2.3.1 ATM Matching

To match the three coefficients $\hat{a}(\cdot)$, $\tilde{b}(\cdot)$ and $\tilde{c}(\cdot)$ of (4.2.10), we need at least three distinct equations for the problem to be well posed. There is some freedom in choosing these equations and the choice may be individually tailored to the corresponding problem at hand. The first approach we present is based exclusively on matching the at-the-money (ATM) initial point, $x = f$. Considering only this one point we choose to match the value of the volatility functions in addition to the value of the first two derivatives at x. This corresponds to

$$\begin{aligned}
\sigma^2_{\text{proj}}(t,f) &= \tilde{\sigma}^2_{\text{proj}}(t,f) \\
\partial_x\sigma^2_{\text{proj}}(t,f) &= \partial_x\tilde{\sigma}^2_{\text{proj}}(t,f) \\
\partial_{xx}\sigma^2_{\text{proj}}(t,f) &= \partial_{xx}\tilde{\sigma}^2_{\text{proj}}(t,f).
\end{aligned} \tag{4.2.11}$$

Evaluating the first equation and noting that $z(f) = \tilde{z}(f) = 0$ we deduce that

$$\hat{a}(t)^2 = \frac{C(f)^2}{\tilde{C}(f)^2}a(t)^2e^{G(t)}. \tag{4.2.12}$$

Evaluating the derivatives at $x = f$ then yields

$$\begin{aligned}\partial_x \sigma^2_{\text{proj}}(t,f) &= 2\sigma^2_{\text{proj}}(t,f)\left[\frac{C'(f)}{C(f)} + b(t)z'(f)\right] \\ \partial_{xx}\sigma^2_{\text{proj}}(t,f) &= 2\sigma^2_{\text{proj}}(t,f)\Bigg[\frac{C'(f)^2}{C(f)^2} + \frac{C''(f)}{C(f)} + 4\frac{C'(f)}{C(f)}b(t)z'(f) \\ &\qquad + b(t)z''(f) + c(t)z'(f)^2\Bigg].\end{aligned}$$

Taking into account that $z'(x) = \frac{1}{C(x)}$ we further simplify this to

$$\begin{aligned}\partial_x \sigma^2_{\text{proj}}(t,f) &= 2\sigma^2_{\text{proj}}(t,f)\left[\frac{1}{C(f)}\Big(C'(f) + b(t)\Big)\right] \\ \partial_{xx}\sigma^2_{\text{proj}}(t,f) &= 2\sigma^2_{\text{proj}}(t,f)\left[\frac{1}{C(f)^2}\Big(C'(f)^2 + C''(f)C(f) + 3C'(f)b(t) + c(t)\Big)\right].\end{aligned} \tag{4.2.13}$$

Thus, matching (4.2.11) with the derivatives provided in (4.2.13), requires a matching of the bracketed terms. In addition to (4.2.12), we then have

$$\begin{aligned}\tilde{b}(t) &= \frac{\tilde{C}(f)}{C(f)}\left(C'(f) + b(t)\right) - \tilde{C}'(f) \\ \tilde{c}(t) &= \frac{\tilde{C}(f)^2}{C(f)^2}\Big(C'(f)^2 + C''(f)C(f) + 3C'(f)b(t) + c(t)\Big) \\ &\qquad - \tilde{C}'(f)^2 - \tilde{C}''(f)\tilde{C}(f) - 3\tilde{C}'(f)\tilde{b}(t).\end{aligned}$$

We shall refer to this algorithm as EMP-ATM matching.

4.2.3.2 Minimal Point Matching

The second matching approach we propose is the minimal point (EMP-MP) matching algorithm and is based on matching the projected variances of the two models at three distinct points. These three points are the minimal requirement for a well-posed specification for deducing the three coefficients $\hat{a}(\cdot)$, $\tilde{b}(\cdot)$ and $\tilde{c}(\cdot)$ of (4.2.10). When including in-the-money (ITM) and/or out-the-money (OTM) values, the method allows better fitting of the wings. As one of the matching points we take the ATM value, $x = f$, corresponding to the condition (4.2.12) of the form

$$\hat{a}(t)^2 = \frac{C(f)^2}{\tilde{C}(f)^2}a(t)^2 e^{G(t)}.$$

Inserting this into the expressions for the variances, for each x we get the condition

$$2\tilde{b}(t)\tilde{z}(x) + \tilde{c}(t)\tilde{z}(x)^2 = \frac{C(x)^2\tilde{C}(f)^2}{\tilde{C}(x)^2C(f)^2}\big(1 + 2b(t)z(x) + c(t)z(x)^2\big) - 1.$$

By selecting two additional points (x_1 and x_2) these equations allow an explicit solution where the new coefficients are given by

$$\tilde{b}(t) = \frac{1}{2}\frac{r(t,x_1)\tilde{z}(x_2)^2 - r(t,x_2)\tilde{z}(x_1)^2}{\tilde{z}(x_1)\tilde{z}(x_2)^2 - \tilde{z}(x_1)^2\tilde{z}(x_2)},$$
$$\tilde{c}(t) = \frac{r(t,x_2)\tilde{z}(x_1) - r(t,x_1)\tilde{z}(x_2)}{\tilde{z}(x_1)\tilde{z}(x_2)^2 - \tilde{z}(x_1)^2\tilde{z}(x_2)},$$

with

$$r(t,x) = \frac{C(x)^2\tilde{C}(f)^2}{\tilde{C}(x)^2C(f)^2}\big(1 + 2b(t)z(x) + c(t)z(x)^2\big) - 1.$$

4.2.3.3 N-Point Matching

The last matching algorithm we propose is based on numerical minimization. Specifying a set of more than three points, it is always possible to minimize an error function that quantifies the difference between the projected variances evaluated at these points. To mention a few possibilities, this includes functions based on the absolute error, relative error or mean squared error. This procedure may yield a better fit along the entire curve, but comes at the cost of requiring a minimization problem to be solved, since the coefficients are no longer explicit. Compared to a direct calibration of the underlying models, based, for example, on the implied volatility curve, this approach will still provide computational advantages owing to the simpler form of the volatility functions. We denote this approach the EMP-NP matching algorithm.

This method is the most general in its application since it is possible to formulate the numerical minimization problem for any model considered. In particular, when the underlying model parameters are the objective variables in the minimization, steps one and two are performed simultaneously. However, to guarantee realistic and stable parameters, additional constraints may be required while performing the minimization.

4.2.3.4 Parameters of the Projection Model

Applying one of the two matching algorithms presented in Section 4.2.3.1 or Section 4.2.3.2, we derived functions $\hat{a}$, $\tilde{b}$ and $\tilde{c}$, which characterize the variance

(4.2.10). We now consider the problem of deducing the underlying model parameters in the dSABR projection model.

From Felpel et al. [2021], the corresponding coefficients for the dSABR model are given by

$$\tilde{a}(t) = \alpha, \qquad \tilde{b}(t) = \frac{\rho\nu}{\alpha}, \qquad \tilde{c}(t) = \frac{\nu^2}{\alpha^2}, \qquad \text{and}$$
$$\tilde{G}(t) = \rho\nu\alpha\beta(f+d)^{\beta-1}t = \tilde{b}(t)\alpha^2\beta(f+d)^{\beta-1}t, \tag{4.2.14}$$

with the transformed variable given by

$$\tilde{z}(x) = \frac{1}{1-\beta}\left((x+d)^{1-\beta} - (f+d)^{1-\beta}\right),$$

for $\beta < 1$. To determine the underlying model parameters we follow a recipe that relies on the projection model being the dSABR model. It may, however, be possible to generalize this approach for more complicated models. Given that f, β and d are known, our goal is to determine the parameters ρ, ν and α.

i) Start by specifying an appropriate fitting maturity t. Using $\tilde{C}(x) = (x+d)^\beta$ and the coefficient specification of the original problem ($C(x)$, $a(t)$, $b(t)$, $c(t)$ and $G(t)$), compute values for $\hat{a}(t)^2$, $\tilde{b}(t)$ and $\tilde{c}(t)$ using the equations in Section 4.2.3.1 or Section 4.2.3.2.

ii) Using the definition of $\hat{\alpha}(t)^2$, and (4.2.14) we can now solve for α. In terms of the quantities found in Step i), find α^2 as the zero of the equation

$$0 = \alpha^2 e^{\tilde{b}(t)\alpha^2\beta(f+d)^{\beta-1}t} - \hat{a}(t)^2. \tag{4.2.15}$$

A good initial guess for α^2 is just the first order Taylor series approximation of the expression, and it may even be possible to provide a solution using the Lambert W function [Lambert, 1758]. Here we see why $\hat{a}(\cdot)$ was identified for the analysis—it allows the inversion of a single equation, given in terms of α and constants (using coefficient values of the base model).

iii) Finally, determine ρ and ν as

$$\rho = \frac{\tilde{b}(t)}{\sqrt{\tilde{c}(t)}} \qquad \text{and} \qquad \nu^2 = \tilde{c}(t)\alpha^2.$$

Following this approach we observe that determination of the parameters is quite straightforward, but may require imposing the restrictions that $\tilde{c}(t) > 0$ and $|\tilde{b}(t)/\sqrt{\tilde{c}(t)}| \leq 1$.

Notice that in the case of the nSABR model where $\beta = 0$ the expression (4.2.15) simplifies to

$$\alpha^2 = \hat{a}(t)^2,$$

which leads to

$$\tilde{\sigma}^2_{\text{proj}}(t,x) = \alpha^2\left(1 + 2\frac{\rho\nu}{\alpha}\tilde{z}(x) + \frac{\nu^2}{\alpha^2}\tilde{z}(x)^2\right)$$

with $\tilde{z}(x) = x - f$. This allows for fast and accurate solutions.

4.2.4 Approximation Quality

We now briefly consider the quality of approximation made by our method. In essence our technique consists of three steps. The first step uses singular perturbation techniques to derive an expression of the form (4.2.7), which has a truncation error of order $\mathcal{O}(\varepsilon^2)$—see Felpel et al. [2021] for full justification. For other papers with approximations of this order of accuracy see Hagan et al. [2014, 2018a, 2021a].

The second and third steps correspond to the recipe described in Section 4.2.3, i.e., when matching (4.2.7) with (4.2.8), and determining the underlying model parameters. As seen, for example in Felpel et al. [2021] and in the flexibility of (4.2.7), even simple SV models can produce volatility functions with complicated shapes. Since, these coefficients may be very general and have a varying number of underlying parameters (see (4.1.1)), we are not able to provide explicit theoretical error bounds for these approximation steps. We can, however, provide convincing numerical evidence showing that the approximation error is not egregious when one projects onto a model with enough flexibility.

To this end we conduct four types of numerical experiment. In the first, given in Section 4.5.1, we explore the error associated with the volatility when projecting the base model onto a projection model using a large range of realistic model parameters appropriate for the model. This shows, in the specific case of projecting dZABR onto nSABR, that there is enough modelling flexibility in the projection model to produce good results. In using other model combinations similar analyses would need to be conducted to ensure that there is not an egregious loss of fidelity.

Given the promising results from the first set of numerical experiments, we further look at the impact of approximation on the implied volatility in Section 4.5.2. This shows that results are accurate and realistic, with the ability to control left and right tail dynamics. In the third (Section 4.5.3), we consider the pricing of various derivative instruments using the projection techniques, again with favourable results around the at-the-money levels. Finally, we consider the explicit densities of

the reference and projection models in Section 4.5.4. The latter results are based on the approach of moment matching.

4.2.5 Moment Matching

We now explore a second application of EMP in which we derive explicit approximating formulas for the density and characteristic functions of a general stochastic volatility model. In this case we produce an explicit projection of the base model onto an nSABR model. Thereafter we use techniques based on a moment matching algorithm.

Concentrating on the nSABR for the moment, we fix a maturity and use explicit formulas for the first four moments available. As was demonstrated by Charvet and Ticot [2011] or Choi et al. [2019], these moments can be characterized in their standardized forms, i.e., by the mean $\mathcal{M}$, variance $\mathcal{V}$, skew $\mathcal{S}$ and excess kurtosis $\mathcal{K}$ as

$$\begin{aligned}
\mathcal{M} &= \mathbb{E}\big[F_t\big] = f \\
\mathcal{V} &= \mu_2 = \frac{\alpha^2}{\nu^2}(x-1) \\
\mathcal{S} &= \frac{\mu_3}{\mu_2^{3/2}} = \rho(x+2)\sqrt{x-1} \\
\mathcal{K} &= \frac{\mu_4}{\mu_2^2} - 3 = (x-1)\left(\frac{1+4\rho^2}{5}(x^3+3x^2+6x+5)+1\right),
\end{aligned}$$

where

$$\mu_i = \mathbb{E}\big[(F_t - f)^i\big] \qquad \text{and} \qquad x = e^{\nu^2 t}.$$

Having these moments available, we proceed in the spirit of Charvet and Ticot [2011] and Tavin [2012], and consider an approximation of this distribution using a parametric distribution of a specific form. Charvet and Ticot [2011] demonstrated this in the case where the approximating parametric distribution was the normal-inverse Gaussian (NIG) distribution. This distribution has four underlying parameters and explicit formulas for its moments. It is, therefore, possible to set up a well posed system of equations to determine the underlying parameters by matching the moments of the distributions. Corresponding algorithms to determine the parameters of the NIG distribution based on the nSABR moments can be found in Charvet and Ticot [2011] or Eriksson et al. [2009].

Instead of considering the NIG distribution, we use the Johnson's-S_U distribution (see Johnson [1949]). Johnson's-S_U distribution defines a random variable, X,

through a transformation of a standard normal random variable, Z, given by

$$X = \xi + \lambda \sinh\left(\frac{Z-\gamma}{\delta}\right).$$

The probability density function of this random variable, ϕ_S, is specified in terms of the standard normal density function, ϕ_N, as

$$\phi_S(x) = \frac{\delta}{\lambda\sqrt{\left(\frac{x-\xi}{\lambda}\right)^2+1}}\phi_N\left(\gamma + \delta \sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)\right).$$

As outlined by Choi et al. [2019], both the distributions generated by the nSABR model, as well as Johnson's-S_U distribution can be seen as special cases of the hyperbolic normal stochastic volatility model. This underscores the similarity of the distributions and motivates our choice. As in the case of the NIG distribution, the Johnson's-S_U distribution has four underlying parameters and explicit formulas for the first four moments, see, e.g., Choi et al. [2019] or Tuenter [2001]. This allows the use of the matching algorithm proposed by Tuenter [2001] to determine the underlying parameters of the distribution. An outline of the algorithm is given in Algorithm 3, for all technical details we refer to the original paper.

Algorithm 3: Matching algorithm proposed by Tuenter [2001].

Data: Moments $\mathcal{M}$, $\mathcal{V}$, $\mathcal{S}$, $\mathcal{K}$
Result: Johnson's-S_U distribution parameters δ, γ, λ, ξ
set $\beta_1 = \mathcal{S}^2$ and $\beta_2 = \mathcal{K} + 3$;
define $m(\omega) = -2 + \sqrt{4 + 2(\omega^2 - \frac{\beta_2+3}{\omega^2+2\omega+3})}$;
define $f(\omega) = (\omega - 1 - m(\omega))(\omega + 2 + 0.5m(\omega))^2$;
compute boundaries (ω_1, ω_2) using procedure in Tuenter [2001];
if $f(\omega_1) \leq \beta_1$ **then**
  break;
end
solve for ω^* s.t. $f(\omega^*) = \beta_1$;
set $m = m(\omega^*)$ and $\Omega = -\text{sgn}(\mathcal{S}) \sinh^{-1}\left(\sqrt{\frac{\omega^*+1}{2\omega^*}\left(\frac{\omega^*-1}{m} - 1\right)}\right)$;
set $\delta = \frac{1}{\log(\omega^*)}$, $\gamma = \frac{\Omega}{\log(\omega^*)}$, $\lambda = \frac{\sqrt{\mathcal{V}}}{\omega^*-1}\sqrt{\frac{2m}{\omega^*+1}}$ and
$\xi = \mathcal{M} - \text{sgn}(\mathcal{S})\frac{\sqrt{\mathcal{V}}}{\omega^*-1}\sqrt{\omega^* - 1 - m}$;

Having determined the underlying parameters, we now have a completely specified probability distribution with explicit formulas for the density function. In the case

of the NIG distribution, explicit formulas for the characteristic function are also available and allow the application of popular computation techniques using the fast Fourier transform (FFT) [Carr and Madan, 1999].

To extend these results to more complex base models the following steps may be used:

i) Using EMP, project the base model onto the nSABR model.

ii) Using the moment matching algorithm, fit a Johnson's-S_U or NIG distribution to the nSABR model.

iii) Using the Johnson's-S_U or NIG distribution, explicit formulas for the density function and/or the characteristic function are available.

This procedure allows us to describe the distribution of the base model at a fixed maturity by an explicit analytical formula.

4.3 Interest Rate Derivatives

Having derived the general methodology we now demonstrate the applicability of the new approach by considering the pricing of interest rate products, where the underlying swap rates are governed by more complex models of the ZABR type. We consider the displaced diffusion ZABR model (dZABR) as our primary example in this setting. The dZABR is specified by the following system of SDEs:

$$\begin{cases} dF_t = v_t(F_t + d)^\beta \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \nu v_t^\gamma \, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases}$$

We focus our attention on options on constant maturity swap (CMS) spreads and mid-curves. Constant maturity swaps are interest rate swaps where one party pays a money market rate, e.g., the 3M Libor rate every quarter and the other party pays a swap rate, e.g., the 10y swap rate on an annual basis. For this type of swap, a corresponding fair constant maturity swap rate can be determined, and it turns out that it is different from the swap rate of liquid fixed-for-floating interest rate swaps, see Andersen and Piterbarg [2010c]. The reason for this is that the CMS can be equivalently described by a replicating portfolio of cash-settled swaptions over the full strike range. This replicating portfolio depends on the full implied swaption volatility surface. The CMS spread is now the difference of two such CMS rates, e.g., the 10y and 2y rates.

The other instrument we consider is the mid-curve swaption. This instrument is a swaption for which the underlying swap starting date is not at the expiry date but later. An example of such a swaption is one that expires in 1 year on a swap starting in 2 years after expiry of the trade and maturing 3 years later, thus in 5 years after expiry of the option. In the special case where the underlying rates are modelled using nSABR, options on CMS spreads have been extensively studied (see e.g. Hagan et al. [2021b]). When the rates are governed by Black basket dynamics the pricing of options on CMS spreads and mid-curves can be found in Antonov [2020]. With our new approach we extend these pricing techniques to allow a general stochastic volatility model as the underlying model for the single rates. We demonstrate this in the case where each rate follows dZABR dynamics under its own measure.

We start by introducing some notation for the interest rate derivatives and payoffs that we consider in the remainder of the paper.

4.3.1 Interest Rate Notation

To introduce the previously mentioned pricing techniques, let us first specify the setting and corresponding notation. For the general tenor structure we consider the starting date T_1 and the end date T_2. The payment dates in between are denoted by $T_1 < t_1, \ldots, t_n \leq T_2$. Furthermore, the exercise date is denoted T and we assume that payment happens at time T_0. We use the notation $DF(T, T_0)$ to denote the discount factor from the payment date to the exercise date and define the annuity factor $A(t, T_1, T_2)$ and the forward level $L(t, T_0, T_1, T_2)$ by

$$\begin{aligned} A(t, T_1, T_2) &= \sum_{T_1 < t_n \leq T_2} \delta_n DF(t, t_n) \qquad \text{and} \\ L(t, T_0, T_1, T_2) &= \sum_{T_1 < t_n \leq T_2} \delta_n DF(t, T_0, t_n) = \frac{A(t, T_1, T_2)}{DF(t, T_0)}, \end{aligned}$$

respectively. Here $\delta_n = t_n - t_{n-1}$ denotes the day count fraction. Finally, the swap rates $R(t, T_1, T_2)$ are defined as

$$R(t, T_1, T_2) = \frac{DF(t, T_1) - DF(t, T_2)}{A(t, T_1, T_2)}. \tag{4.3.1}$$

Notice that the definition of the swap rates can be further generalized as in Hagan et al. [2021b]. For our situation, however, this generalization is not necessary and we retain the notation used in Antonov [2020].

4.3.2 Options on CMS Spreads

The first interest rate derivatives we consider are options on CMS spreads. To this end, we provide a short summary of the results of Hagan et al. [2021b], where the goal is the evaluation of caplets, floorlets or swaplets on CMS spreads. We denote by $(R_i(t))_{i=1,2}$ the two swap rates $R(t, T_0, T_i)$. Moreover, we define the spread, $S(t)$, on the two rates as

$$S(t) = R_2(t) - R_1(t)$$

with initial value $s = r_2 - r_1$. Using a caplet as an example, the payoff function at the expiry date T is given by

$$V_{spread}(T; T, K) = DF(T, T_0)(S(T) - K)^+.$$

In turn, the value of the caplet at an earlier date is given as the conditional expectation under a suitable martingale measure. Under the forward measure, i.e., with the discount factor $DF(\cdot, T_0)$ as the numéraire, the value is expressed as

$$V_{spread}(t; T, K) = DF(t, T_0)\mathbb{E}^{T_0}\big[(S(T) - K)^+ \big| t\big].$$

Without loss of generality we consider the value at the initial time, $t = 0$. To evaluate this expression, Hagan *et al.* introduced a suitable measure, H, under which the spread process becomes a martingale[2]. By further analysing the convexity correction the value of the caplet can be evaluated as (see e.g. Hagan et al. [2021b], Equation 3.22)

$$V_{spread}(0; T, K) = DF(0, T_0)\big((1-(s-K)\lambda_s)C^S(0; T, K)+\lambda_s Q_C^S(0; T, K)\big). \quad (4.3.2)$$

Here the expressions C^S and Q_C^S denote the vanilla and quadratic calls of the spread under the martingale measure H, i.e.,

$$C^S(t; T, K) = \mathbb{E}^H\Big[(S(T) - K)^+ \big| t\Big]$$
$$Q_C^S(t; T, K) = \mathbb{E}^H\Big[\big((S(T) - K)^+\big)^2 \big| t\Big].$$

The coefficient λ_s denotes the convexity coefficient which, following the suggestion of Hagan et al. [2021b], is set in a manner consistent with the underlying swap rates as

$$\lambda_s = \frac{\lambda_2 Q_S^2(0; T, r_2) - \lambda_1 Q_S^1(0; T, r_1)}{Q_S^S(0; T, s)}.$$

[2]In the following applications the explicit form of the measure H is not of importance—only the dynamics of the spread measure expressed in this measure is relevant.

Here the functions Q_S^S and Q_S^i express the quadratic swaps

$$Q_S^S(t;T,K) = \mathbb{E}^H\Big[(S(T)-K)^2\Big|t\Big]$$
$$Q_S^i(t;T,K) = \mathbb{E}^i\Big[(R_i(T)-K)^2\Big|t\Big],$$

and the coefficients λ_i are the convexity coefficients of the underlying rates determined by

$$\frac{L(0,T_0,T_0,T_i)}{L(T,T_0,T_0,T_i)} = \frac{L_i(0)}{L_i(T)} = 1 + \lambda_i(R_i(T) - R_i(0)) + \cdots . \tag{4.3.3}$$

Given the convexity coefficients and the explicit values of the vanilla and quadratic calls this provides a closed form pricing formula.

By construction, see Hagan et al. [2021b], these formulae satisfy put-call parity.

4.3.3 Options on Mid-curves

Another popular semi-vanilla product is the mid-curve call option. This option provides the holder with the right to enter into a swap with starting time T_1 and maturity T_2 at some exercise date T. As shown by Antonov [2020], and Feldman [2020], the value of the mid-curve call option can be expressed as a call option on the mid-curve rate, $R_{mc}(T) = R(T,T_1,T_2)$, under the annuity measure, which uses numéraire $A_{mc}(T) = A(T,T_1,T_2)$. At time 0, this yields the value

$$V_{mc}(0,T,K) = A_{mc}(0)\mathbb{E}^{A_{mc}}\Big[(R_{mc}(T)-K)^+\Big]. \tag{4.3.4}$$

Note that by definition the mid-curve rate, $R_{mc}(T)$, is a martingale under the chosen annuity measure. To evaluate the expectation we do not assume an additional model for the mid-curve rate, but instead focus on the choice of the underlying swap rates. Therefore, in the rest of this section, we follow the approach presented by Antonov [2020] and express the mid-curve rate as a weighted spread on swap rates. We start by considering the rates $R_i(T)$ and deduce that

$$R_{mc}(T) = R_2(T)\frac{A_2(T)}{A_{mc}(T)} - R_1(T)\frac{A_1(T)}{A_{mc}(T)}.$$

Defining new modified rates $\hat{R}_i(T)$ as

$$\hat{R}_i(T) = \frac{R_i(T)}{\hat{M}_i(T)} = \frac{R_i(T)}{\frac{A_{mc}(T)A_i(0)}{A_i(T)A_{mc}(0)}},$$

the mid-curve rate becomes

$$R_{mc}(T) = \hat{R}_2(T)\frac{A_2(0)}{A_{mc}(0)} - \hat{R}_1(T)\frac{A_1(0)}{A_{mc}(0)}.$$

At this point we highlight an important structural difference in comparison to the CMS spread options presented previously in Section 4.3.2, namely that under the chosen annuity measure the mid-curve rate, $R_{mc}(T)$, as well as the rates, $\hat{R}_i(T)$, are martingales at the same time. To evaluate the CMS spread options we work under a measure where only the spread is a martingale, with the result that the weighted spread and all the single rates are simultaneously martingales.

To characterize these new modified rates, $\hat{R}_i(T)$, we assume that the martingales $\hat{M}_i(T)$ are approximated by

$$\hat{M}_i(T) = \frac{A_{mc}(T)A_i(0)}{A_i(T)A_{mc}(0)} \approx 1 + \hat{\lambda}_i(R_i(T) - R_i(0)) + \cdots,$$

and, as demonstrated by Antonov [2020], this allows us to recover the distributions of the modified rates from the original ones by considering approximate option prices using the relation

$$\mathbb{E}^{A_{mc}}\left[(\hat{R}_i(T) - K)^+\right] \approx \mathbb{E}^{A_i}\left[\left(R_i(T)(1 - \hat{\lambda}_i K) - K(1 - \hat{\lambda}_i r_i)\right)^+\right]. \tag{4.3.5}$$

4.4 Projection of ZABR-type Models

In this section we choose the base model to be displaced ZABR (dZABR). To enable the efficient application of EMP, we first provide explicit expressions for the coefficient functions. Given that dZABR is specified as (4.1.1) with $C(F_t) = (F_t + d)^\beta$, $\mu(v_t) = 0$ and $\nu(v_t) = \nu v_t^\gamma$, the coefficients are

$$a(t) = \alpha, \qquad b(t) = \rho\nu\alpha^{\gamma-2}, \qquad c(t) = \nu^2\alpha^{2(\gamma-2)}\left(1 + (\gamma-1)\rho^2\right), \qquad \text{and}$$

$$G(t) = -\rho^2\nu^2\alpha^{2(\gamma-1)}(\gamma-1)t + \rho\nu\alpha^\gamma\beta(f+d)^{\beta-1}t, \tag{4.4.1}$$

with the transformed variable given by

$$z(x) = \frac{1}{1-\beta}\left((x+d)^{1-\beta} - (f+d)^{1-\beta}\right),$$

for $\beta < 1$.

Note that we have chosen the dZABR base model as our primary example to ensure that notation remains relatively simple. More complicated examples such as the mean-reverting ZABR or Heston models may be considered using the corresponding coefficients given, for example, in Felpel et al. [2021] and Hagan et al. [2018a].

As our initial applications of EMP, we project the dZABR model onto the dSABR and nSABR models. These examples illustrate how the method works with more extensive numerical analyses performed in Section 4.5.

Having described the general framework for pricing interest rate derivatives in the previous section, we next show how to apply EMP to a general stochastic volatility model for the underlying swap rates, with applications to the aforementioned CMS spreads and mid-curves.

4.4.1 Projection onto Displaced SABR and Normal SABR Models

As our first example, we illustrate the projection of dZABR onto dSABR at time $T = 1$. Parameters for the base model are given in column one of Table 4.1. In column two we list the parameters for the projection onto a dSABR model. Since this model uses the same $C(\cdot)$ function as the base model, which leads to the same $z(\cdot)$ functions, the projection volatility functions for the base and projection models take the identical functional form in terms of x. Thus the three projection constants are uniquely determined by any of the EMP-ATM, EMP-MP or EMP-NP approaches. This exact fit is shown in the top left panel of Figure 4.1.

To show an example where there is not an exact fit, we change the displacement and β parameters in the projection model, thus ensuring that the $C(\cdot)$ functions are no longer identical. Columns three, four and five show the fits for the three approaches, while the three remaining panels of Figure 4.1 show the results graphically, with the corresponding points used for matching represented by crosses.

As a second example, we now project the same dZABR base model onto the nSABR model. Columns two to five of Table 4.2 show the projection parameters obtained using the three approaches. The corresponding graphs are shown in the four panels of Figure 4.2. As can be seen, even though the projection model has three less degrees of freedom (d, β, and γ), the projection model still yields relatively good results in each case. Larger deviations are observed in the wings, when compared to the projections obtained using the dSABR projection model.

	dZABR base	dSABR EMP (any)	dSABR EMP-ATM	dSABR EMP-MP	dSABR EMP-NP
f	0.005	0.005	0.005	0.005	0.005
d	0.002	0.002	0.003	0.003	0.003
α	0.012488	0.01254811	0.00730453	0.00731934	0.00717598
β	0.4	0.4	0.3	0.3	0.3
ν	0.3	0.71771391	0.69462229	0.75552662	0.72011663
ρ	-0.3	-0.30273701	-0.26276722	-0.32425693	-0.22915447
γ	0.8	-	-	-	-

Table 4.1: Parameters used for the dZABR base model (column one). The other columns display the parameters for the dSABR projection models that best match the base dZABR projected volatility for the various projection methods used. The right three columns used a projection model with a different displacement parameter, d, so that $\tilde{C}(\cdot)$ is not identical to the base model.

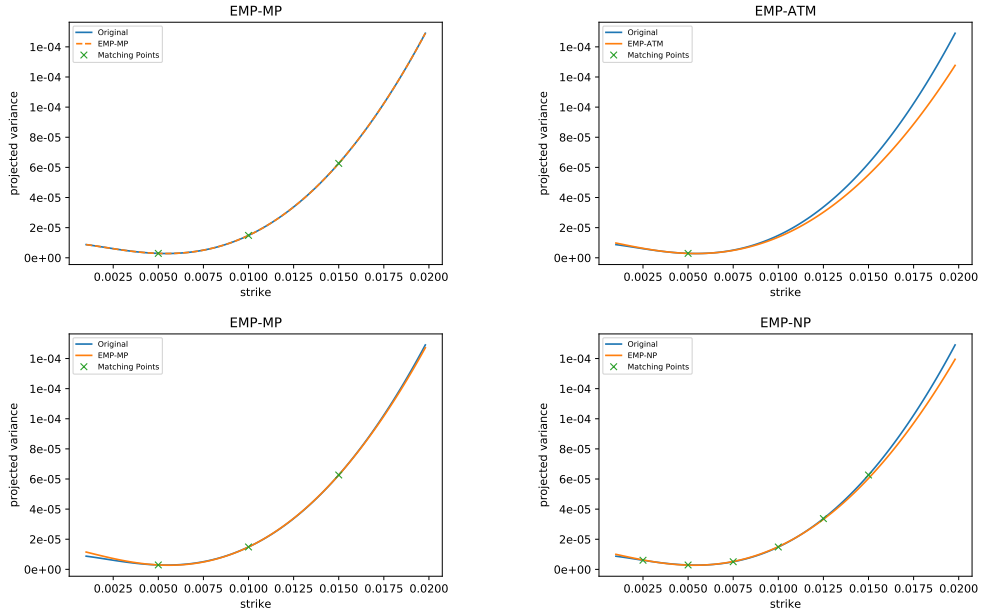


Figure 4.1: Projected variance, $\sigma^2_{\text{proj}}(T, \cdot)$, for the parameter sets given in Table 4.1. Green crosses indicate matching points used.

	dZABR base	nSABR EMP-ATM	nSABR EMP-MP	nSABR EMP-MP	nSABR EMP-NP
f	0.005	0.005	0.005	0.005	0.005
d	0.002	-	-	-	-
α	0.012488	0.00170596	0.00170596	0.00170596	0.001644852
β	0.4	-	-	-	-
ν	0.3	0.66074529	0.84698470	0.69725701	0.73344524
ρ	-0.3	-0.17780185	-0.41495911	-0.01811182	-0.073130682
γ	0.8	-	-	-	-

Table 4.2: Parameters used for the dZABR base model. The right four columns display the parameters for the nSABR projection models that best match the base dZABR projected variance. We applied direct matching using all EMP matching approaches with two different sets of matching points for the EMP-MP approach.

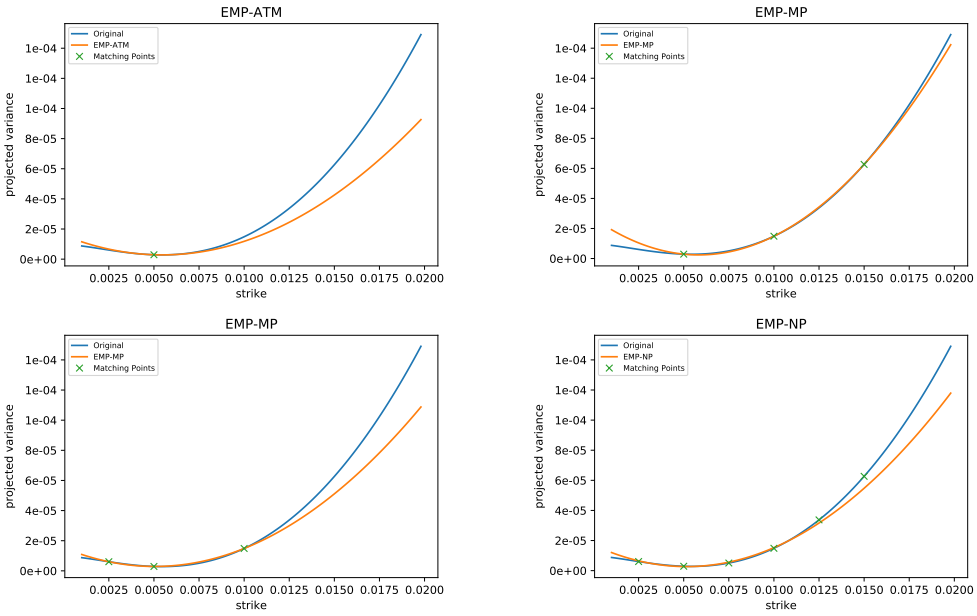


Figure 4.2: Projected variance, $\sigma^2_{\text{proj}}(T, \cdot)$, for the parameter sets given in Table 4.2. Green crosses indicate matching points used.

4.4.2 Options on CMS Spreads

Given the pricing formula for options on CMS spreads in (4.3.2), we now consider valuation of vanilla and quadratic calls on the spread and swap rates. This depends heavily on the model chosen for the single rates. Assuming that each swap rate is modelled using dZABR, the call options can be valued using the methods described by Andreasen and Huge [2011b], and Felpel et al. [2021]. However, it is not clear how to value options on the spread process since the underlying model is not known. Therefore, in the explicit case, where each rate follows an nSABR model under its respective measure, Hagan et al. [2021b,a] showed that it is possible to approximate the dynamics of the spread using an nSABR model.[3] To achieve this, singular perturbation techniques were used and the resulting approximation has the same order of accuracy as the classical SABR implied volatility formulas. Applying these results and the following steps, we can evaluate CMS spread options using dZABR models for each of the rates:

i) Given that each swap rate $R_i(t)$ is modelled using a dZABR model, we use EMP to deduce approximate rates $\tilde{R}_i(t)$ characterized in terms of nSABR models.

ii) We approximate the spread, $S(t)$, on the swap rates as

$$S(t) \approx \tilde{S}(t) = \tilde{R}_2(t) - \tilde{R}_1(t).$$

Using the results in Hagan et al. [2021b,a] this allows for an approximation of the spread in terms of an nSABR model.

iii) We evaluate the corresponding vanilla and quadratic calls using the explicit formulas presented in Hagan et al. [2021b] and can, in turn, evaluate the price of the option on the CMS spread.

This approach can be extended to the general SV models of (4.2.4). We do, however, stress that this approach introduces multiple approximations at various steps, making it numerically very challenging to provide proper error estimation. For the case where the swap rates are modelled using a dZABR model, as assumed above, our numerical examples look very promising, see Section 4.5. This suggests that the new technique has the potential to provide a valid computational framework for a vast class of stochastic volatility models.

[3] This actually also holds true when we replace the spread by a basket of nSABR models

4.4.3 Options on Mid-Curves

To deduce the price of a call option on a mid-curve, we proceed in a fashion similar to that as above. In Section 4.3.3 it was shown that the underlying mid-curve rate can be expressed as a weighted spread of two modified rates, $\hat{R}_i(T)$. From an analytical perspective we cannot deduce a general dynamic of the rates $\hat{R}_i(T)$. We can, however, resort to numerical means in the spirit of Antonov [2020]. Given that the underlying rates $R_i(T)$ follow a dZABR model under their own measure, we assume that the modified rates $\hat{R}_i(T)$ are governed by a dZABR model under the annuity measure A_{mc}. Using the parameters of the rates $R_i(T)$ we can then calibrate the model parameters of the rates $\hat{R}_i(T)$ using (4.3.5). Since we know that the rates $\hat{R}_i(T)$ are martingales we approximate them using EMP onto corresponding nSABR models. Given that the original weighted sum of the rates, $\hat{R}_i(T)$, is also a martingale under the A_{mc}-measure, we can use the more general basket result of Hagan et al. [2021b,a] to deduce that the mid-curve rate, $R_{mc}(T)$, is approximated by an nSABR model. This reduces the evaluation of the option on mid-curves to a vanilla call option under an nSABR model.

4.5 Numerical Results

In order to evaluate the quality of our new approximation technique we perform a series of numerical experiments. We start by considering the various matching algorithms described in Section 4.2.3 and investigate their ability to fit the projected volatility function. Thereafter, we explore the implications on the implied volatility surface and demonstrate the flexibility obtained when pricing various interest rate derivatives. We conclude by analysing the accuracy of the distribution matching approach proposed in Section 4.2.5.

For most of our examples we shall consider dZABR models specified using the parameter sets listed in Table 4.3. In the case where $\gamma = 1$ these models reduce to dSABR models. In this simpler setting, the main difference, when compared with the nSABR model, is the additional parameter β. We, therefore, highlight the dependence on this parameter and consider various values for it.

For most of our examples we desire time independent parameter values of the projected nSABR model. We select a universal fitting maturity of $T = 1$. To provide some intuition behind this choice, consider the dZABR coefficients given in Section 4.4. We observe that only $G(\cdot)$ has dependence on the maturity time. Considering, for example, the EMP-MP approach in Section 4.2.3, this dependence is translated into a scaling of the initial volatility. Thus, the fitting maturity $T = 1$

Parameter	Set 1	Set 2	Set 3	Set 4
f	0.005	0.005	0.005	0.005
α	$0.3f^{1-\beta}$	$0.3f^{1-\beta}$	$0.3f^{1-\beta}$	$0.3f^{1-\beta}$
d	0.002	0.002	0.002	0.002
ν	0.3	0.3	0.3	0.5
ρ	-0.3	-0.7	0	-0.3
T	5	5	10	1
γ	0.8	0.9	0.8	0.9

Table 4.3: Parameter values for dZABR models used in numerical experiments.

provides a compromise between capturing the initial behaviour and the dynamical behaviour of the $G(\cdot)$ coefficient. If not mentioned otherwise, we apply EMP using this fitting maturity.

4.5.1 Approximation Quality of the dZABR

For our first numerical experiment, we analyse the approximation quality of our technique. As discussed in Section 4.2.4, an exact derivation of the approximation error is not viable, we therefore consider an illustrative numerical experiment using the dZABR model.

4.5.1.1 Projected Volatility

To apply the different matching algorithms described in Section 4.2.3 we require additional parameters to be specified. In particular, other than the at-the-money value, two additional matching points are required to implement the EMP-MP algorithm. To demonstrate the influence of these points on the projection, we evaluate the projection volatility function for the dZABR model using parameter set 1 with $\beta = 0.2$ and $\beta = 0.8$, and generate the curve using the EMP-MP algorithm for different values of the matching points. The errors observed between the original model and the approximating curves are presented in Tables 4.4 and 4.5.

For this example we observe a reasonably good fit in all cases. Here the match including the point $x_1 = 0.5f$ provides the best fit for the left wing, whereas the match including the additional points $x_1 = 2f$ and $x_2 = 3f$ provides an excellent fit for the longer right wing. The latter also provides a reasonably good fit for the

x	dZABR	x_1=0.5f x_2=2f	x_1=0.5f x_2=3f	x_1=2f x_2=3f
$0.5f$	$8.48e^{-6}$	0	0	$3.08e^{-6}$
$0.75f$	$4.57e^{-6}$	$-2.45e^{-7}$	$-3.48e^{-7}$	$1.04e^{-6}$
f	$2.58e^{-6}$	0	0	0
$1.5f$	$5.73e^{-6}$	$5.92e^{-7}$	$1.41e^{-6}$	$-4.36e^{-7}$
$2f$	$1.97e^{-5}$	0	$2.47e^{-6}$	0
$3f$	$8.37e^{-5}$	$-8.22e^{-6}$	0	0
$4f$	$1.99e^{-4}$	$-2.90e^{-5}$	$-1.17e^{-5}$	$-4.32e^{-6}$

Table 4.4: Values of the projected variance, $\sigma^2_{\text{proj}}(T, \cdot)$, for the dZABR model using parameter set 1 with $\beta = 0.2$, and the error in the EMP-MP projection variance for different values of the matching points, x_1 and x_2.

x	dZABR	x_1=0.5f x_2=2f	x_1=0.5f x_2=3f	x_1=2f x_2=3f
$0.5f$	$3.56e^{-6}$	0	0	$4.10e^{-6}$
$0.75f$	$3.55e^{-6}$	$-2.18e^{-7}$	$-3.55e^{-7}$	$1.49e^{-6}$
f	$3.75e^{-6}$	0	0	0
$1.5f$	$5.85e^{-6}$	$6.35e^{-7}$	$1.73e^{-6}$	$-7.33e^{-7}$
$2f$	$1.18e^{-5}$	0	$3.28e^{-6}$	0
$3f$	$4.09e^{-5}$	$-1.09e^{-5}$	0	0
$4f$	$1.01e^{-4}$	$-4.24e^{-5}$	$-1.94e^{-5}$	$-9.58e^{-6}$

Table 4.5: Values of the projected variance, $\sigma^2_{\text{proj}}(T, \cdot)$, for the dZABR model using parameter set 1 with $\beta = 0.8$, and the error in the EMP-MP projection variance for different values of the matching points, x_1 and x_2.

left wing. Thus, from now on, when we refer to the EMP-MP algorithm we shall use the additional points $x_1 = 2f$ and $x_2 = 3f$.

In Figure 4.3 we graph the projected variance generated by the original dZABR model in comparison with the projected variance generated by the EMP-MP, EMP-ATM and EMP-NP algorithms. When implementing the EMP-NP algorithm we minimized the relative mean squared error evaluated at equally spaced grid points from $0.5f$ to $3f$, using steps of length $0.5f$. It is observed that the EMP-MP and EMP-NP approaches seem to outperform the EMP-ATM matching procedure. With our choice of $x_1 = 2f$ and $x_2 = 3f$ as the additional matching points in the EMP-MP algorithm, we observe a better fit along the right wing, whereas the EMP-NP algorithm provides better results along the left wing.

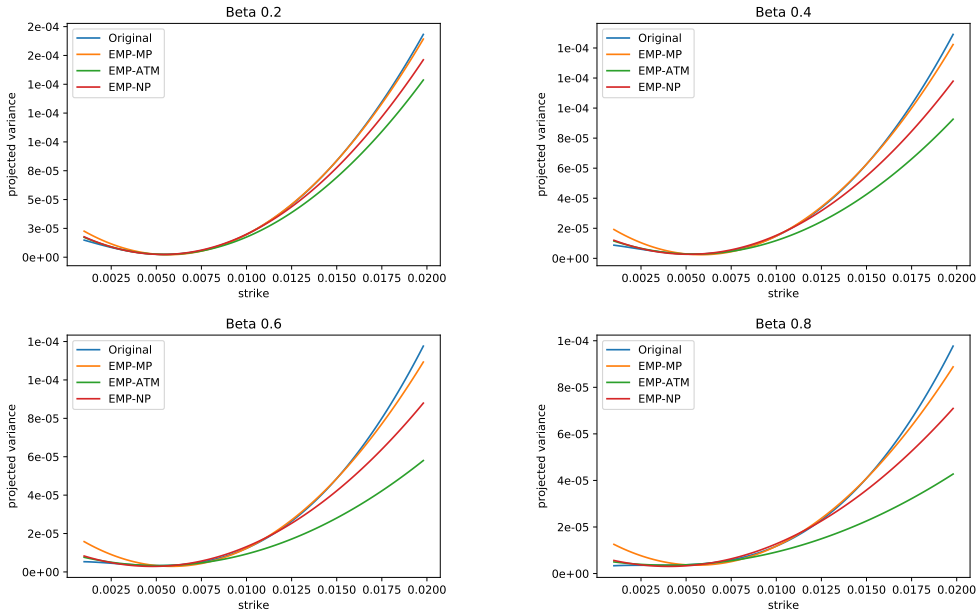


Figure 4.3: Projected variance, $\sigma^2_{\text{proj}}(T, \cdot)$, for the different EMP algorithms using parameter set 1.

In conclusion we observe good fits for the projected dZABR model using EMP. Comparing the matching algorithms, the multiple point algorithms seem to outperform the ATM matching algorithm in terms of stability and quality of approximation.

4.5.1.2 Error Distribution

To further illustrate approximation quality, in particular when using the EMP-ATM and EMP-MP approaches, we consider projections using a broad range of parameters, which cover a large selection of values. We consider market-realistic strikes in the range $[0.7f, 1.3f]$, and compute the mean absolute error using the two methods, where the error is computed as the difference between projected variance of the base and projected models sampled at 50 equal intervals over this range. Error statistics are provided for 100 000 scenarios; ensuring that $e^G(t) < 150$ and $c(t) < 10000$. Correspondingly, the first factor of the projected variance, (4.2.7), is given by $(F_t + d)^{2\beta}\alpha^2 e^{G(t)}$. If the values of $G(\cdot)$, given in (4.4.1), are too large (in our case bigger than 5) this leads to unreasonably large values of the projected volatility. The range of parameters, displayed in Table 4.6, are chosen to prevent this. Here, as previously mentioned, we fixed $T = 1$. Since time only appears as a factor in $G(\cdot)$, changing maturity will require changes to the parameter bounds

parameter
$f \in [0, 0.05]$
$d \in [0, 0.01]$
$\alpha \in [0.0001, 0.1]$
$\beta \in [0.001, 0.95]$
$\nu \in [0.01, 0.51]$
$\rho \in [-0.99, 0]$
$\gamma \in [0.05, 0.95]$
$T = 1$

Table 4.6: Parameter ranges used for sampling.

to maintain realistic volatility ranges.

To present results, consider the following error levels $c_0 = 0$, $c_i = 10^{-i}$, $i = 1, \ldots, 7$. Using these levels, we produce a discrete distribution of the mean absolute error (less than the levels) and its corresponding cumulative distribution. These are graphed in Figure 4.4.

Figure 4.5 shows the scenario that had the largest mean absolute error. In this figure we also show the fit produced by the N-point matching approach with $N = 6$.

Even in the worst scenario the approximation quality is not entirely egregious, with the EMP-MP method performing reasonably well around the ATM price. However, owing to the restriction of the parabolic shape of the projected variance curve, which results from using the nSABR as the projection model, it is not possible to accurately recover the shape of the left wing. This numerical experiment suggests that the results lead to reasonable approximations over a wide parameter range, representing a large range of realistic modelling cases, covering realistic intervals of the forward f. While this analysis has been restricted to the case of projecting dZABR onto nSABR, a similar exercise may be carried out for other projections.

4.5.2 Implied Volatility

The above results for the projected variance are promising. From a practical perspective it is, however, not clear how differences in these projected variance curves translate into pricing differences. We, therefore, generate implied volatility curves for call prices using each algorithm. To generate the curves, we use the numerical PDE framework described by Felpel et al. [2021] and Hagan [2015] to

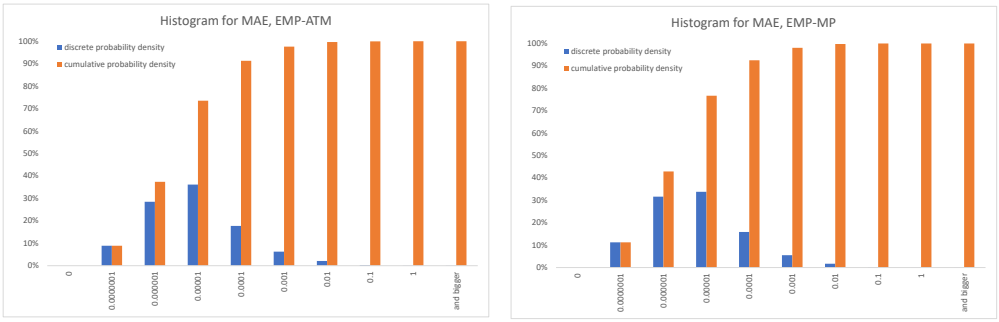


Figure 4.4: Histogram of the mean absolute differences for the projected variances of the base and projected models for the EMP-ATM (left) and EMP-MP matching methods (right).

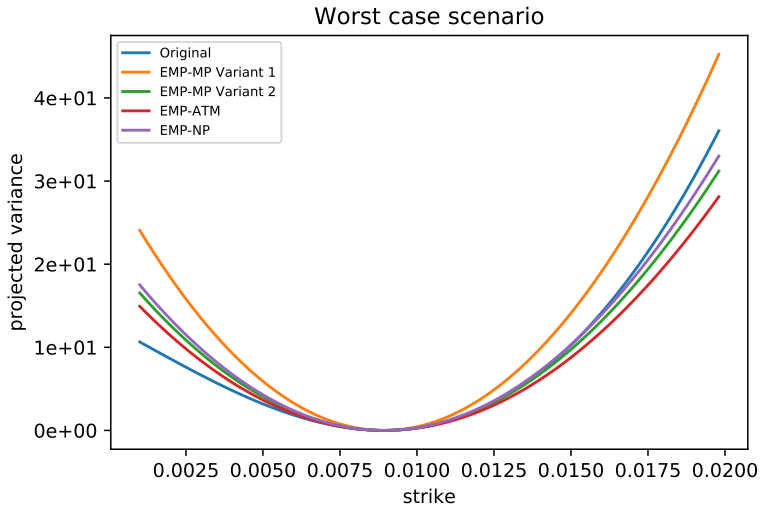


Figure 4.5: Worst case scenario corresponding to the nSABR projection parameter values (rounded to 4 places) of $f = 0.0089$, $d = 0.0096$, $\alpha = 0.0020$, $\beta = 0.5697$, $\rho = -0.0526$, $\nu = 0.5092$ and $\gamma = 0.2678$.

evaluate the corresponding probability density functions. Using the EMP-MP and EMP-NP algorithms, and again projecting the dZABR model onto nSABR, this results in a PDE characterized by

$$\begin{cases} \partial_t Q(t,F) = \partial_{FF}\left[\frac{1}{2}\sigma^2_{\text{nSABR}}(t,F)Q(t,F)\right] \\ Q(t,F) \to \delta(F-f) \text{ as } t \to t_0^-. \end{cases}$$

To solve this PDE, we must specify appropriate boundary conditions. Using EMP we have two valid options available:

i) The first approach is to solve the PDE for the density function in the setting of an nSABR model. This implies a probability distribution on the whole space and yields the boundary conditions

$$\begin{cases} Q(t,F) \to 0 \text{ as } F \to -\infty \\ Q(t,F) \to 0 \text{ as } F \to \infty. \end{cases}$$

ii) The second approach is to solve the PDE in the setting of the original model, using EMP to only approximate the projected volatility function. In the case of the dZABR model, this approach respects the limits imposed on the model at the displacement d, providing a probability distribution on the restricted space. This is expressed using the boundary conditions

$$\begin{cases} Q(t,F) \to 0 \text{ as } F \to -d^- \\ Q(t,F) \to 0 \text{ as } F \to \infty. \end{cases}$$

To provide some intuition on the choice of the boundary condition, the first approach yields a more consistent setup when considering the pricing of baskets, whereas the latter allows for a more accurate approximation and maintains the imposed restrictions on the domain of the model. In Figures 4.6 and 4.7 the normal implied volatility curves for the original model, using the framework described by Felpel et al. [2021] and Hagan [2015], are compared with the curves generated by the nSABR model using the EMP-MP and EMP-NP algorithms, under the same framework and parameters under both boundary conditions described above. The curves are computed using parameter set one. In Figures 4.8 and 4.9 a similar comparison is provided using parameter set two. We observe a good fit of the implied volatility curves for both approaches. In particular, the right wing is nearly perfectly matched in most cases. The greatest difference is observed on the left wing and for the largest value of $\beta = 0.8$. Similar results are observed in Tables 4.7 and 4.8, which show implied volatilities for parameter sets 3 and 4 computed under the boundary condition (ii). Once again, the largest deviation in values is observed in the left wing, which is still reasonable given that we are using an arbitrage free approximation.

Note that the matching algorithms do not necessarily guarantee that the resulting value of ρ is bounded by -1 and 1. To ensure that the nSABR models are arbitrage free we imposed an additional bound of 0.999 on the absolute value of ρ.

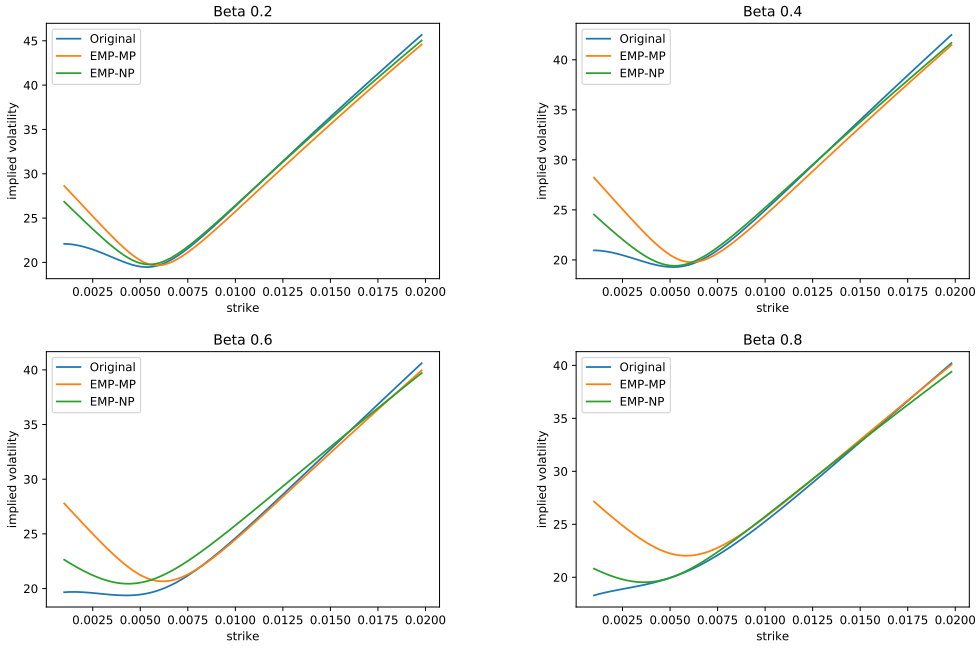


Figure 4.6: Normal implied volatility curves in bp computed using parameter set 1 and the full nSABR PDE.

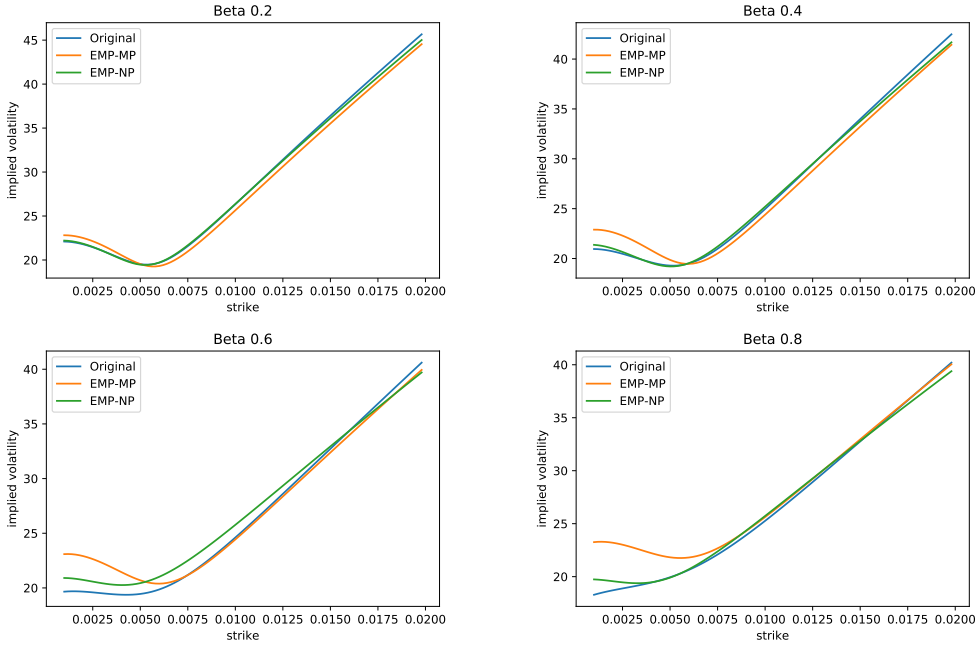


Figure 4.7: Normal implied volatility curves in bp computed using parameter set 1 with the PDE considering the barrier.

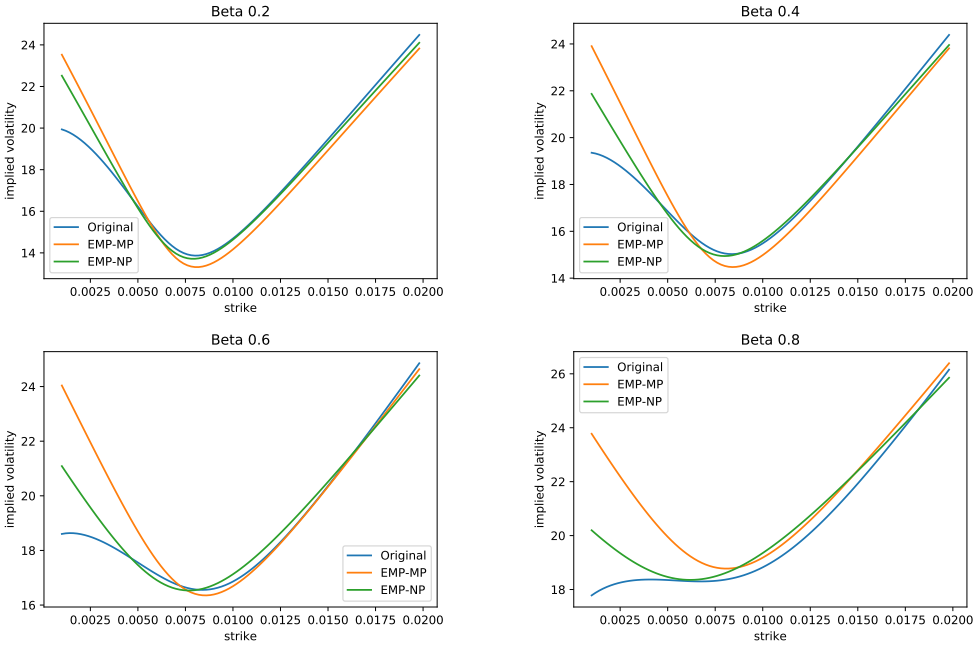


Figure 4.8: Normal implied volatility curves in bp computed using parameter set 2 and the full nSABR PDE.

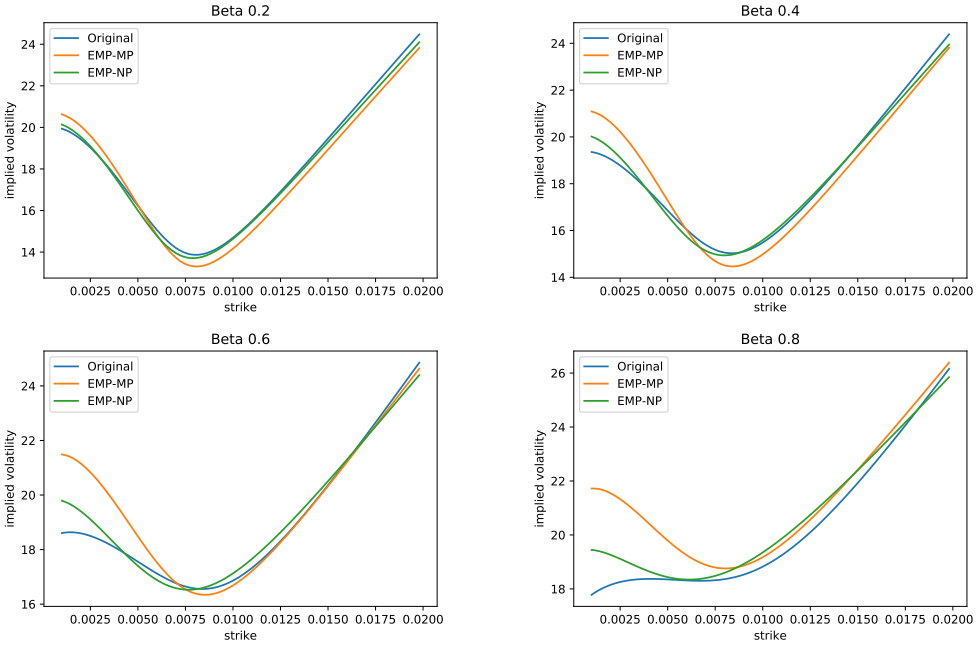


Figure 4.9: Normal implied volatility curves in bp computed using parameter set 2 with the PDE considering the barrier.

	$\beta = 0.2$			$\beta = 0.8$		
x	dZABR	EMP-MP	EMP-NP	dZABR	EMP-MP	EMP-NP
$0.5f$	21.14	21.94	21.13	19.14	22.95	19.56
$0.75f$	21.49	22.24	21.45	20.37	23.61	20.42
f	22.19	22.71	22.17	21.85	24.38	21.76
$1.5f$	25.98	26.20	26.05	25.57	27.21	25.46
$2f$	30.96	31.13	30.95	29.83	31.12	29.53
$3f$	40.90	40.97	40.64	38.62	39.44	37.56
$4f$	50.30	50.27	49.77	47.23	47.52	45.19

Table 4.7: Normal implied volatility in bp for the dZABR model computed using parameter set 3 in comparison with the normal implied volatilities generated by EMP.

	$\beta = 0.2$			$\beta = 0.8$		
x	dZABR	EMP-MP	EMP-NP	dZABR	EMP-MP	EMP-NP
$0.5f$	21.20	22.95	21.37	20.28	25.70	20.03
$0.75f$	18.69	19.65	18.75	19.89	22.69	19.56
f	16.83	16.87	17.03	19.91	20.17	20.05
$1.5f$	18.37	17.55	18.99	22.10	19.95	23.25
$2f$	23.33	22.48	23.94	26.34	24.13	27.53
$3f$	33.80	32.85	34.15	36.28	33.92	36.29
$4f$	43.80	42.66	43.80	46.35	43.42	44.66

Table 4.8: Normal implied volatility in bp for the dZABR model computed using parameter set 4 in comparison with the normal implied volatilities generated by EMP.

4.5.3 Pricing of Interest Rate Derivatives

Having shown that the approximation techniques yield satisfactory results, we go a step further and demonstrate the possible benefits for interest rate derivative pricing. To demonstrate the advantages consider the following toy model:

4.5.3.1 Setup

For our concrete application we shall consider the $1y2y$- and the $1y5y$-swap rates as underlying rates and model them using dZABR models. We assume that the

payment date corresponds to the exercise date $T = 1$. To show the additional flexibility of using dZABR models in comparison to, for example, nSABR models, we construct some sample data around the ATM values for each of the swap rates. These sample points were constructed using the nSABR model with the parameters presented in Table 4.9.

	$1y2y$	$1y5y$
f	0.003	0.005
α	0.0009	0.0015
ν	0.3	0.3
ρ	-0.5	-0.7

Table 4.9: Parameters of the nSABR models used to generate sample data for swap rates.

Next, we specify a few variations of the dZABR model for consideration. The values for the additional parameters in each of the versions are presented in Table 4.10. For each rate and each version the parameters α, ν and ρ are calibrated using the ATM samples. To be more precise we use sample points from 70%-ATM to 130%-ATM in steps of 5%. In Figure 4.10 and Figure 4.11 we graph the corresponding implied volatility curves for the rates. Call prices are computed using the explicit formulas presented in Hagan et al. [2021b]. In these figures the additional flexibility available for controlling the wings is clearly visible.

	V1	V2	V3
β	0.4	0.4	0.5
γ	0.8	0.9	0.8
d	0.002	0.002	0.002

Table 4.10: Additional parameters for three versions of the dZABR model.

To price the interest rate derivatives presented in Section 4.3 we must specify the correlation structure between swap rates. Here we assume a single driving Brownian motion for the stochastic volatility and impose a correlation structure between the rates of the form

$$\begin{bmatrix} 1 & \omega & \rho_{1y2y} \\ \omega & 1 & \rho_{1y5y} \\ \rho_{1y2y} & \rho_{1y5y} & 1 \end{bmatrix}.$$

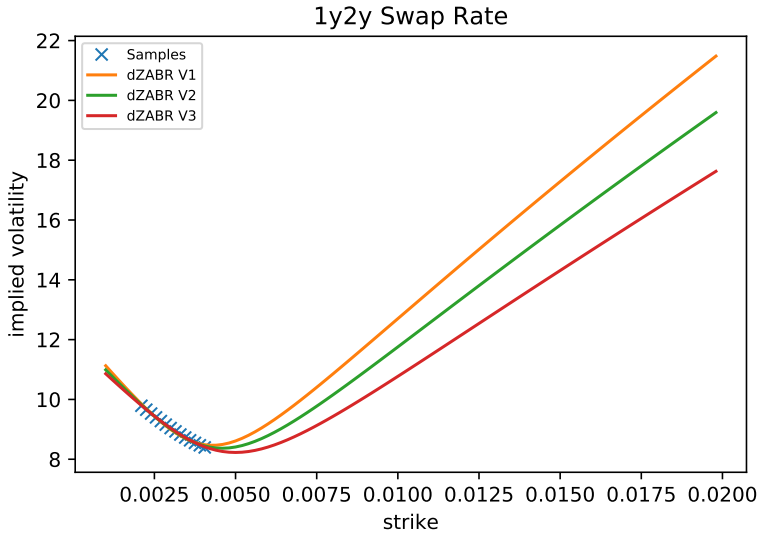


Figure 4.10: Normal implied volatility curves in bp for the dZABR models calibrated to the $1y2y$ swap rate.

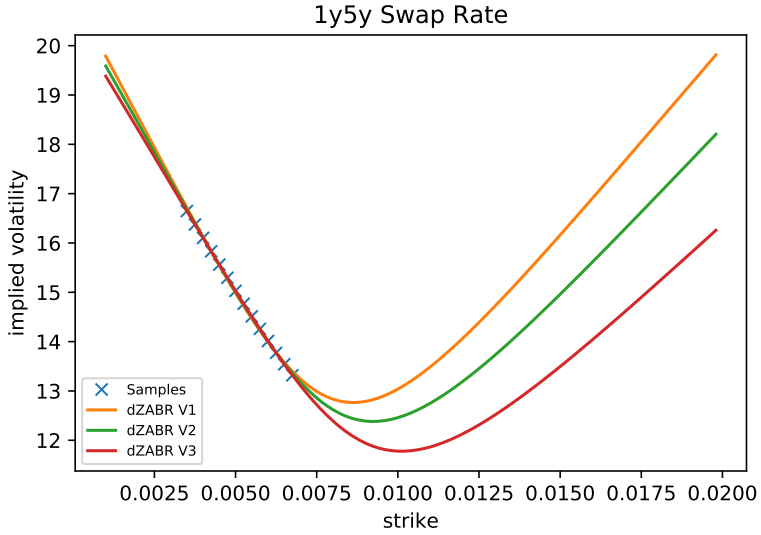


Figure 4.11: Normal implied volatility curves in bp for the dZABR models calibrated to the $1y5y$ swap rate.

Using the results described in Section 4.4, we evaluate vanilla and quadratic call options on a basket of nSABR models. To determine the dynamics of the spread we use the results of Hagan et al. [2021b] to express the spread, $S(T)$, of two nSABR models with parameters $(f_1, \alpha_1, \nu_1, \rho_1)$ and $(f_2, \alpha_2, \nu_2, \rho_2)$ as an nSABR

model with parameters given by

$$\begin{aligned}
f_s &= f_2 - f_1, \\
\alpha_s &= \sqrt{\alpha_2^2 - 2\omega\alpha_1\alpha_2 + \alpha_1^2}, \\
\nu_s &= \frac{1}{\alpha_s^2}(\alpha_2^2\nu_2 - \omega\alpha_1\alpha_2(\nu_1 + \nu_2) + \alpha_1^2\nu_1), \\
\rho_s &= \frac{1}{\alpha_s}(\alpha_2\rho_2 - \alpha_1\rho_1).
\end{aligned}$$

These parameters are derived under the measure H at the exercise time T.

We note that these parameters are a rough approximation. For a more precise approximation the additional terms Γ and κ of Hagan et al. [2021b] must be considered. Moreover, similar parameters may be computed for the weighted spread. For both generalizations we refer to Hagan et al. [2021b].

4.5.3.2 CMS Spreads

Using the method described in Section 4.4, we can price call options on the CMS spread if we correctly specify the convexity coefficient. To determine the convexity coefficient we follow the approach of Antonov [2020] and assume a flat yield curve. The convexity coefficients are then given by

$$\lambda_i = \frac{T_i - T}{2},$$

with a complete derivation of this expression provided in Appendix 4.7.3.

Having defined all necessary terms we use the parametrization presented in Table 4.9 to compute the implied volatility of a call on a $5y - 2y$ spread with a maturity of 1 year. The correlation parameter ω was set to 0.5, and, in order to evaluate the expectations, we used the explicit formulas of Hagan et al. [2021b]. In Figure 4.12 the implied volatility curves for all possible combinations of the dZ-ABR rates are presented. The additional correlation parameter ω was calibrated to match the area from 90%-ATM to 110%-ATM with steps of size 2%. Here, we clearly see how the additional control of the wing of the swap rates allows good control of the wing of the CMS spread. In particular, this is useful when the quality of the data in the wings is questionable but the entire curve is required for risk management.

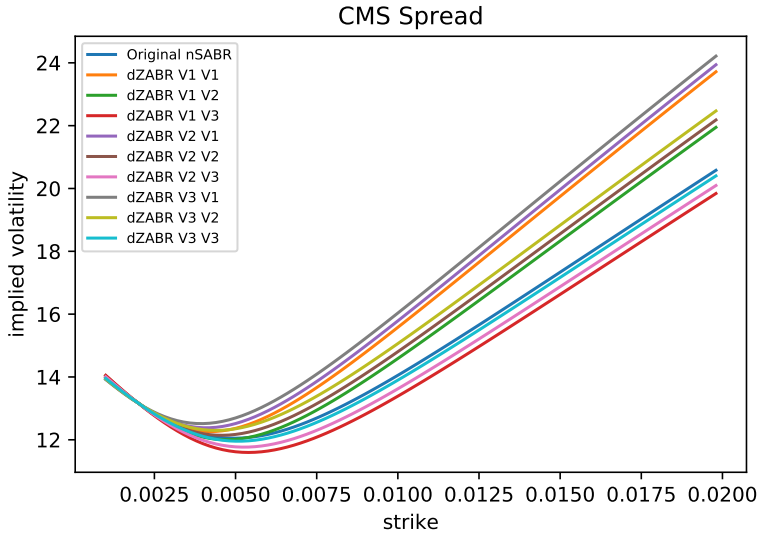


Figure 4.12: Normal implied volatility curves in bp for a caplet on the CMS spread. All possible combinations of the dZABR models specified in Table 4.10 are presented.

4.5.3.3 Mid-Curves

In a manner similar to that used above we approximate the convexity coefficients as

$$\hat{\lambda_i} = \frac{T + T_i - T_1 - T_2}{2}.$$

Using the approach described in Section 4.4.3 we evaluate the call prices on mid-curves. For the annuity factors we use further approximations and set the initial values to be

$$\begin{aligned} A_{mc}(0) &\approx T_2 - T_1 \\ A_1(0) &\approx T_1 - T \\ A_2(0) &\approx T_2 - T. \end{aligned}$$

In Table 4.11 implied volatilities are shown for selected combinations of the dZABR models. Again, better control of the right wing is achieved. To visualize the effects, we computed the implied volatility of the expectation of (4.3.4), i.e., without the additional factor coming from $A_{mc}(0)$. The results are shown in Figure 4.13.

Strikes	dZABR V1 V1	dZABR V2 V2	dZABR V3 V3
20	231.0	230.8	230.7
40	121.8	121.7	121.6
60	48.0	48.1	48.2
80	23.4	23.3	23.2
100	18.6	18.2	17.6
150	19.3	18.1	16.6
190	22.0	20.4	18.4

Table 4.11: Normal implied volatility of a call on the mid-curve for selected combinations of the dZABR models. All values are in bp.

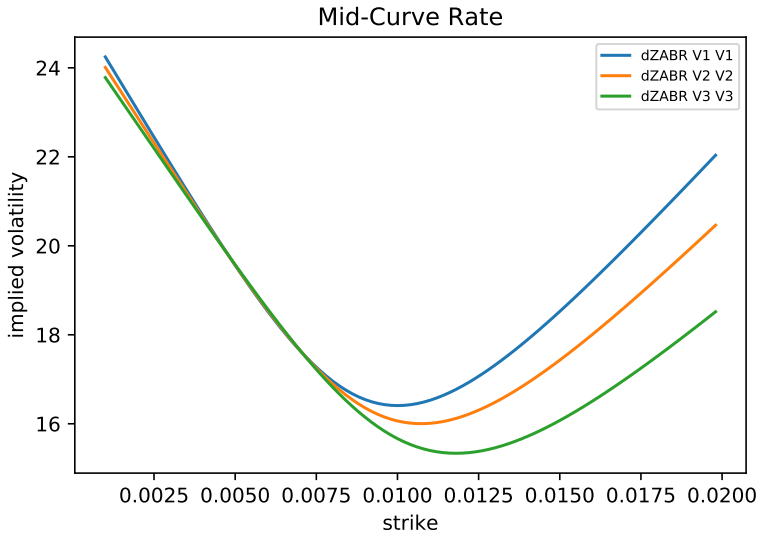


Figure 4.13: Normal implied volatility curves in bp for the caplet on the mid-curve without scaling.

4.5.4 Explicit Probability Density Functions

For our last numerical demonstration we investigate the accuracy of various approximating distributions using the methodology described in Section 4.2.5. To be more precise, we compute the grid according to the framework described in Hagan [2015] and Felpel et al. [2021] and deduce an average probability density over these intervals. In the notation of Hagan [2015] this corresponds to the function $(\theta_j)_j$. For Johnson's-S_U distribution and the NIG distribution we evaluated the corresponding mass over the interval. The parameters of Johnson's-S_U distribu-

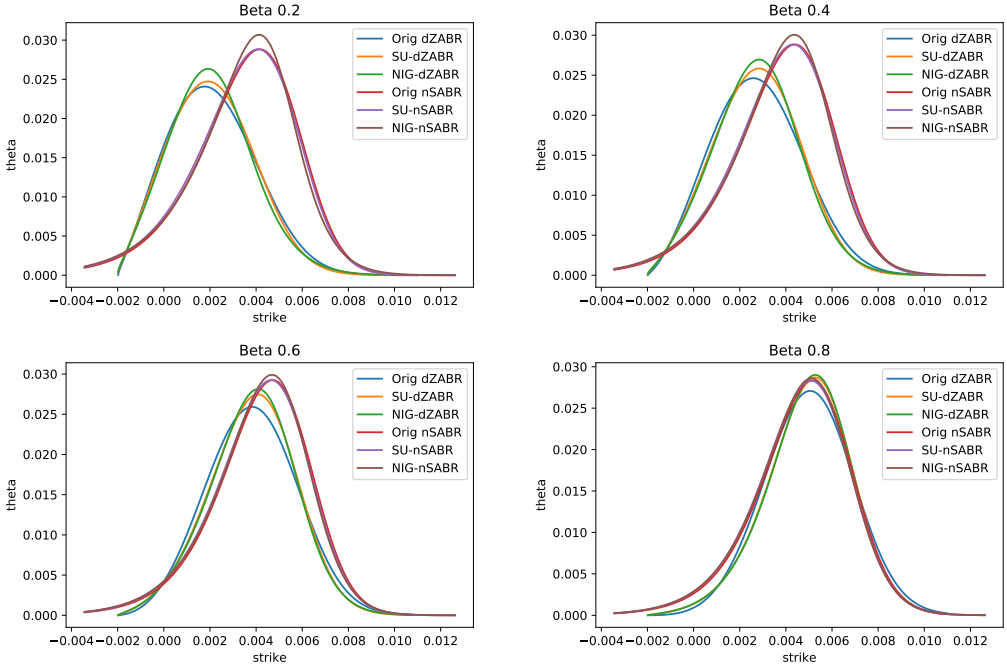


Figure 4.14: Probability over grid intervals for parameter set 2 and maturity $T = 1$.

tion were computed using the algorithm described in Section 4.2.5, and the NIG parameters were determined using the method provided by Eriksson et al. [2009]. In Figure 4.14 the corresponding averaged probabilities are shown in comparison with the original dZABR distribution and the approximated nSABR distribution using the EMP-MP algorithm. The figures were generated using parameter set 2 of Table 4.9 with a slight modification of the maturity, which was set to $T = 1$.

As can be seen the approximation yields very good results and manages to capture the shape of the original distribution. In particular, both approximating distributions produce a nearly perfect fit to the nSABR model. The quality of dZABR approximation is of the same order as the EMP nSABR approximation of dZABR. From a practical perspective, however, this approach must be carefully applied when considering larger maturities. In Figures 4.15 and 4.16 the results of the approximations are shown for maturities $T = 5$ and $T = 7$, with all other parameters remaining the same.

Here, we observe that the Johnson distribution provides a better fit to the original distribution than the NIG distribution. For larger maturities this approximation is far from perfect—the corresponding moments are, however, perfectly matched. This behaviour can only be explained if the underling distributions cannot be characterized completely using only four moments. In particular, this implies

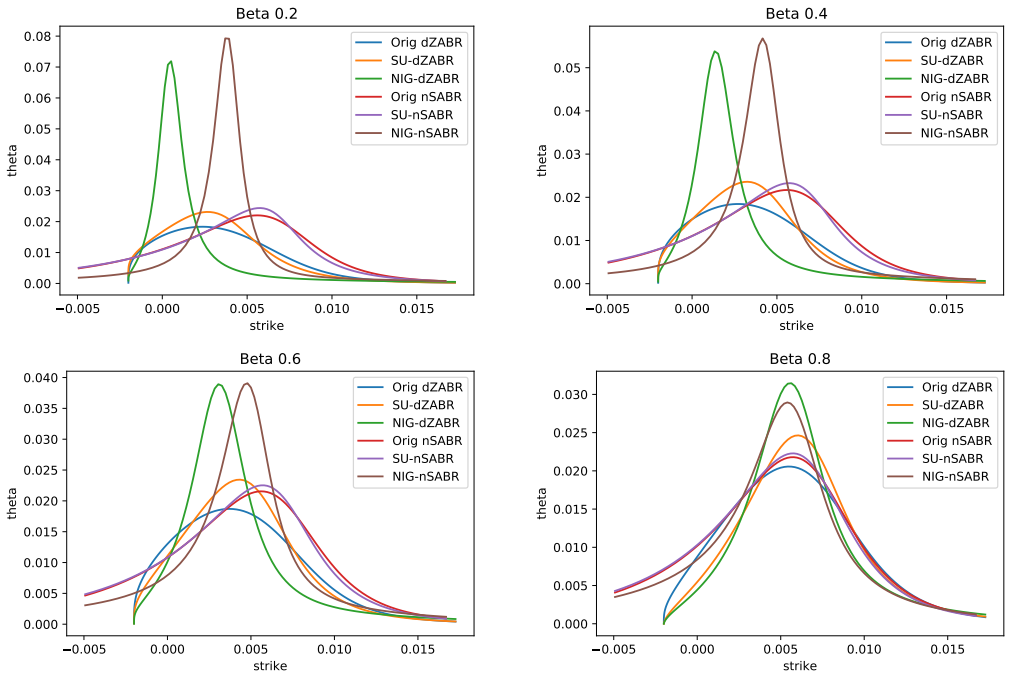


Figure 4.15: Probability over grid intervals for parameter set 2 and maturity $T = 5$.

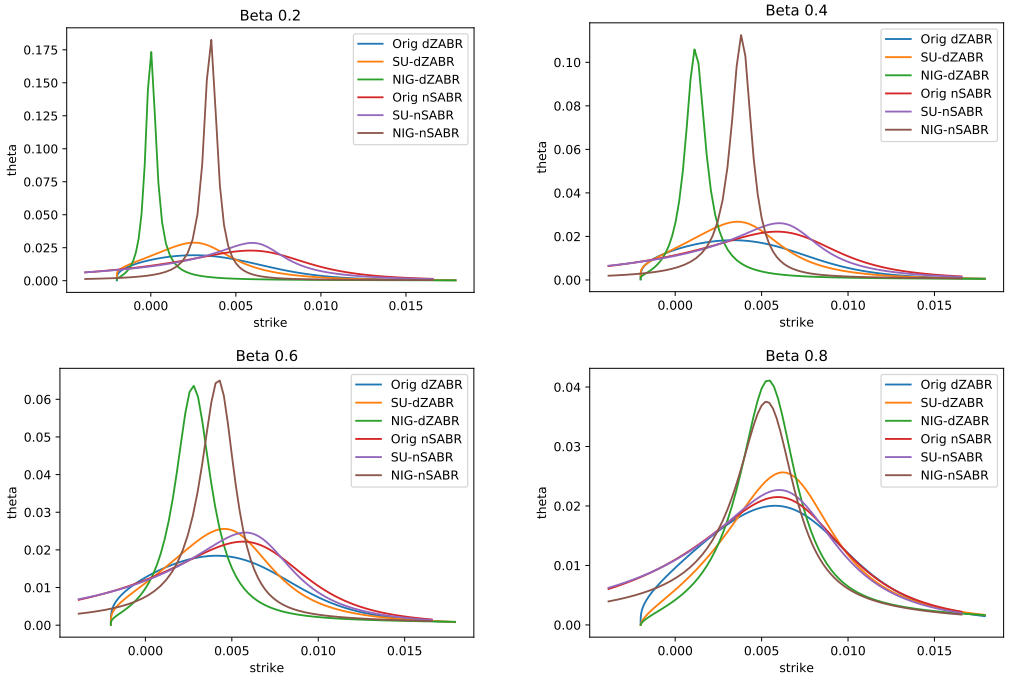


Figure 4.16: Probability over grid intervals for parameter set 2 and maturity $T = 7$.

one of two possibilities. The first possibility is that the higher moments of the distribution, which are not matched, become more important and should be taken into account when matching the parameters. The other possibility is that the underlying distribution may fall foul of the 'moment problem'. To provide some intuition, consider the first example provided in Durrett [2010, Chapter 3.3.5] concerning the standard log-normal distribution, of a random variable X, which has probability density function

$$\phi_{logN}(x) = \frac{1}{\sqrt{2\pi}x} e^{-\frac{\log(x)^2}{2}} \text{ for } x \geq 0.$$

It was shown that the standard lognormal distribution, as well as its modified versions given by

$$\phi_a(x) = \phi_{logN}(x)(1 + a\sin(2\pi\log(x)))$$

with $a \in [-1, 1]$, all share the same moments, for all k, given by

$$\mathbb{E}[X^k] = e^{\frac{k^2}{2}}.$$

The problem is that the moments of the log-normal distribution grow too quickly, see Durrett [2010, Theorem 3.3.25], where a sufficient condition for good behaviour is given by

$$\limsup_k \frac{\mathbb{E}[X^{2k}]^{\frac{1}{2k}}}{2k} = r < \infty. \tag{4.5.1}$$

Considering the leading order of the first four moments of the nSABR model, provided in Section 4.2.5, we assume the moments are of the form

$$\mathbb{E}[X^k] = \mathcal{O}(x^{\frac{(k-1)k}{2}}) = \mathcal{O}(e^{\tilde{\nu}^2 t \frac{(k-1)k}{2}}),$$

which does not obey the condition in (4.5.1). In particular, the growth of the moments increases with increasing maturity, influencing the speed of divergence. This may be translated into the quality of the matching algorithm, explaining the good fit for small maturities and the increasing discrepancies for larger maturities.

In conclusion, this suggests that our approach may not be universally applicable. In particular, when considering larger maturities the results must be analysed carefully, requiring a possible change of the matching algorithm to ensure the best fit. For small maturities, however, this approach yields very good results, which allows for closed-form formulas describing the distribution and characteristic functions.

4.6 Conclusion

In this paper we have considered the setting of general stochastic volatility models as introduced by Felpel et al. [2021]. We have extended this work to allow pricing and calibration of multi-rate interest rate derivatives. Effective Markovian Projection was used with the normal SABR model acting as a reference model. This enabled us to apply the general setting to CMS spread and mid-curve options. Using a moment matching method based on Johnson distributions we were also able to make use of closed form solutions for pricing.

Future research will entail applying the Effective Markovian Projection method to non-parametric stochastic local volatility models with further application of these techniques for pricing derivatives.

4.7 Appendix

4.7.1 Coefficients for General Stochastic Volatility Models

Here we recall the necessary assumptions that apply to Theorem 1 of Felpel et al. [2021].

Assumption V. The drift term, $\mu(\cdot)$, is differentiable, with derivative $\mu'(\cdot)$, and a solution $Y(t, t_0, \alpha)$ to the following PDE exists:

$$\begin{cases} \partial_t Y(t, t_0, \alpha) = \mu(Y(t, t_0, \alpha)) \\ Y(t, t, \alpha) = \alpha \\ Y(t_0, t_0, \alpha) = \alpha. \end{cases}$$

Assumption VI. The function Y is differentiable and has an inverse function $y(t_0, t, a)$ such that

$$Y(t, t_0, \alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0, t, a).$$

Assumption VII. The functions

$$X(t, t_0, \alpha) = \partial_\alpha Y(t, t_0, \alpha), \qquad Z(t, u) = Z(t, u, t_0, \alpha) = y(u, t, Y(t, t_0, \alpha)),$$
$$z(F) = \int_f^F \frac{1}{C(u)}\, du, \qquad s(t) = S(t_0, t, \alpha) = \int_{t_0}^t Z(t, u, t_0, \alpha)^2\, du$$

and

$$\psi(t,u,Z) = \nu\big(Z(t,u)\big)Z(t,u)X\big(t,u,Z(t,u)\big)$$

are well defined, $X\big(t,u,Z(t,u)\big)^{-1}$ exists, and the following integral functions are defined:

$$\begin{aligned}
I_1(t) &= \rho\int_{t_0}^{t}\psi(t,u,Z)\,du,\\
I_2(t) &= 2\int_{t_0}^{t}\nu\big(Z(t,u)\big)^2X\big(t,u,Z(t,u)\big)^2\int_{u}^{t}Z(t,v)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,\\
I_3(t) &= \rho\int_{t_0}^{t}\psi(t,u,Z)\int_{u}^{t}Z(t,v)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,\\
I_4(t) &= \rho^2\int_{t_0}^{t}\psi(t,u,Z)\int_{u}^{t}\partial_Z\Big(\psi(t,v,Z)\Big)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,\\
I_5(t) &= \int_{t_0}^{t}\nu\big(Z(t,u)\big)^2X\big(t,u,Z(t,u)\big)^2\,du.
\end{aligned}$$

Assumption VIII. The function $C(\cdot)$ is differentiable at f, with derivative denoted by $C'(\cdot)$.

Given these assumptions the coefficients of Equation (4.2.6) are given by

$$\begin{aligned}
&a(t) = Y(t,t_0,\alpha), \qquad c(t) = b(t)^2 + \frac{1}{a(t)s(t)^2}I_2(t) - \frac{6b(t)}{s(t)^2}I_3(t) + \frac{2}{a(t)s(t)^2}I_4(t),\\
&b(t) = \frac{1}{a(t)s(t)}I_1(t), \quad G(t) = -s(t)c(t) - s(t)b(t)\Gamma_0 + \frac{1}{a^2}I_5(t)
\end{aligned}$$

and

$$\Gamma_0 = -C'(f).$$

4.7.2 Basket Dynamics under suitable Numéraire

Let us consider the dynamics of the forward rate F_t given by a general SABR model

$$\begin{cases}
dF_t = u_tC(F_t)\,dW_t, & F_{t_0} = f,\\
du_t = \nu u_t\,dZ_t, & u_{t_0} = 1,\\
\text{with} \quad d\langle W_t, Z_t\rangle_t = \gamma\,dt.
\end{cases}$$

Let θ denote the corresponding model parameters, for instance after calibrating the model to given market data. When we consider forward swap rates we must determine each convexity-adjusted forward swap rate $\tilde{F}_t$. Given payment date T_p in the future, the convexity adjusted rate at T_p is determined by the conditional expectation of F_t with regard to the T_p-forward measure, thus,

$$\tilde{F}_t = \mathbb{E}^{T_p}[F_t].$$

The forward rate F_t is not a martingale with respect to the T_p-forward measure but the convexity adjusted rate $\tilde{F}_t$ is. Thus, given a parametrized model for F_t with model parameters θ we calibrate the respective model for $\tilde{F}_t$ and get model parameters $\tilde{\theta}$. This is similar to the method applied for the consideration in Section 4.3.3

Now, if we consider basket options we can consider each forward rate $F_{i,t}$ to be the convexity adjusted forward rate $\tilde{F}_{i,t}$ and safely assume that the corresponding dynamics are martingale, i.e., driftless with regard to the T_p-forward measure. Then, we proceed to price options on the basket by considering the basket for the driftless rates $\tilde{F}_{i,t}$.

4.7.3 Convexity Coefficients

To apply the pricing formulas of Section 4.3 we need to deduce the interdependence between the rates and the annuity expressed in terms of the convexity coefficients. Due to the currently low interest regime, for our examples we do this by imposing the simplest form of a flat yield curve structure as in Antonov [2020], i.e.,

$$DF(t,T) = e^{-(T-t)x(t)}.$$

Furthermore, we assume linear dependence of the functions on the factor x, which means that

$$DF(t,T) \approx 1 - (T-t)x(t) + \mathcal{O}(x^2).$$

Using the continuous representation of the annuity $A(t,T_1,T_2)$ we can approximate

$$\begin{aligned} A(t,T_1,T_2) &\approx \int_{T_1}^{T_2} DF(t,u)\,du \approx \int_{T_1}^{T_2} 1-(u-t)x(t)+\mathcal{O}(x^2)\,du \\ &= (T_2-T_1) - x(t)\int_{T_1}^{T_2}(u-t)\,du + \mathcal{O}(x^2), \end{aligned}$$

and can further deduce the linear dependence of the swap rates $R(t, T_1, T_2)$ on the factor x as

$$R(t, T_1, T_2) = \frac{x(t)(T_2 - T_1)}{(T_2 - T_1) - x(t) \int_{T_1}^{T_2} (u - t)\, du} + \mathcal{O}(x^2) = x(t) + \mathcal{O}(x^2). \tag{4.7.1}$$

With this we can express the dependence of the annuity on the rates as

$$A(t, T_1, T_2) \approx (T_2 - T_1) - R(t, T_1, T_2) \int_{T_1}^{T_2} (u - t)\, du + \mathcal{O}(x^2).$$

To evaluate the convexity coefficients let us note that the desired terms are generally given by the form

$$\begin{aligned} M(x(t), x(s)) &= \frac{(a + b(t)x(t))(c + d(s)x(s))}{(c + d(t)x(t))(a + b(s)x(s))} \\ &\approx 1 + \frac{cb(t) - ad(t)}{ac} x(t) + \frac{ad(s) - cb(s)}{ac} x(s) + \mathcal{O}(x^2). \end{aligned} \tag{4.7.2}$$

For our concrete application this yields

$$\begin{aligned} M_i(x(T), x(t)) = \frac{L_i(t)}{L_i(T)} &= \frac{A_i(t) DF(T, T_0)}{A_i(T) DF(t, T_0)} \\ &\approx \frac{(1 - (T_0 - T)x(T))((T_i - T_0) - x(t) \int_{T_0}^{T_i} (u - t)\, du)}{((T_i - T_0) - x(T) \int_{T_0}^{T_i} (u - T)\, du)(1 - (T_0 - t)x(t))} \\ &\approx 1 + \frac{-\tau_i (T_0 - T) + \int_{T_0}^{T_i} (u - T)\, du}{\tau_i} x(T) \\ &\quad + \frac{-\int_{T_0}^{T_i} (u - t)\, du + \tau_i (T_0 - t)}{\tau_i} x(t) \\ &\approx 1 + \frac{\int_{T_0}^{T_i} u\, du - \tau_i T_0}{\tau_i} (x(T) - x(t)). \end{aligned}$$

where we set $\tau_i = T_i - T_0$. This means we can approximate the convexity coefficients by

$$\lambda_i = \frac{1}{\tau_i} \int_{T_0}^{T_i} u\, du - T_0.$$

For the mid-curve options the convexity coefficient becomes

$$\begin{aligned}\hat{M_T^i}(x(T), x(0)) &= \frac{A^{mc}(T)A^i(0)}{A^i(T)A^{mc}(0)} \\ &\approx \frac{((T_2 - T_1) - x(T)\int_{T_1}^{T_2}(u-T)\,du)((T_i - T_0) - x(0)\int_{T_0}^{T_i} u\,du)}{((T_i - T_0) - x(T)\int_{T_0}^{T_i}(u-T)\,du)((T_2 - T_1) - x(0)\int_{T_1}^{T_2} u\,du)} \\ &\approx 1 + \frac{\tau_{21}\int_{T_0}^{T_i} u\,du - \tau_i\int_{T_1}^{T_2} u\,du}{\tau_{21}\tau_i}(x(T) - x(0)).\end{aligned}$$

This means we have

$$\hat{\lambda_i} = \frac{\tau_{21}\int_{T_0}^{T_i} u\,du - \tau_i\int_{T_1}^{T_2} u\,du}{\tau_{21}\tau_i}.$$

5 Effective Stochastic Local Volatility Models

Considering the results of Chapter 3 and Chapter 4 we have established approximation techniques to allow the explicit pricing of call options in a one- as well as multidimensional financial market framework. The pricing and modelling in these chapters is mostly performed with regard to a fixed maturity. In this chapter we further increase the scope of our methodology and consider the modelling of a full time and space dependent surface. To achieve this we focus our research on stochastic local volatility (SLV) models in this chapter.

Recalling our initial motivation in Chapter 3 to adequately model the smile and skew behaviour in the implied volatility curve, we considered generalized stochastic volatility models (SV). While this class of models allows to generate various different curve structures, our examples are all governed by a parametric model. Having only a finite number of model parameters, it is not always possible to perfectly reproduce all observed market values in the calibration. In practice this is not optimal when a replication of more involved financial instruments is done using these vanilla options. Here the small inaccuracy in the pricing of call options can yield to a mispricing of the desired instrument. To overcome this problem, SLV models are introduced as a mixture between SV models and local volatility models.

The mixture is performed by taking the dynamics of the SV model and introducing an additional non-parametric component into the model. This non-parametric component is concentrated in a leverage function and provides additional degrees of freedom in the model. This allows to perfectly match the prices of all observed vanilla options using the SLV model. To derive approximation results similar to that in Chapter 3, we consider leverage functions which can be decomposed into a non-parametric time-dependent component, used to model the ATM evolution in time, and a non-parametric space-dependent component to perfectly match the state at a fixed maturity. Under this assumption we prove that an analysis using

singular perturbation techniques is possible and derive an extension of Theorem 3.2.3 which covers these types of SLV models. This allows to transform all models covered in Chapter 3 into a corresponding SLV model.

Having established the mathematical foundation to consider SLV models using our existing framework, we discuss the practical applicability. Special consideration is placed on the calibration of the leverage function. We demonstrate how it is possible to derive an explicit and fast algorithm to calibrate the leverage function to a given market surface. Due to our explicit model approximations this can be performed independently of the model simulation and evaluation of prices. Furthermore, having established the model characterization as well as the calibration procedure, we also reconsider the model implementation. Recalling Chapter 3, we have a general model framework in mind which can be applied to a large class of heterogeneous generalized stochastic volatility models. Thus, we demonstrate how it is possible to incorporate our results in the computational framework already available in Chapter 3 and Chapter 4. To achieve this we show how it is possible to modify the algorithm presented in Section 2.5.3 to fit into the new model set-up.

The rest of this chapter presents the preprint Felpel et al. [2022a] starting from its introduction. In addition, some additional, unpublished material is presented to provide more insight into selected topics. This content is presented in Section 5.9.

5.1 Introduction

When pricing options there is often the need to exactly fit observed implied volatilities for market quoted prices. For instance, using the technique of static replication in terms of vanilla instruments to price and hedge exotic options results in mispricing if some of the vanillas are mispriced to start with. It is possible to exactly calibrate to the observed implied volatility surface using a local volatility model Dupire [1994], Derman and Kani [1994]. This approach is, however, non-parametric, and while, in principle, it has infinite degrees of freedom for fitting, it has the drawback that it does not provide suitable dynamics for the forward implied volatility surface. This means that such a model generates flattening forward implied volatility surfaces, with the result that it is not possible to model smile dynamics with significant convexity. This makes it unsuitable for modelling financial instruments that are sensitive to a forward implied volatility smile. Forward starting options are an example of such instruments, requiring techniques such as the sticky skew approach, based on forward volatilities as observed today, to be used to project the skew/smile for future dates. Stochastic volatility models provide better forward implied volatility surface dynamics. But, owing to their

parsimonious parametrization, they are unable to price calibration instruments exactly, making it impossible to fit the quoted option implied volatilities exactly.

Combining the best features of both of the aforementioned approaches leads to the class of local stochastic volatility models. We shall consider local stochastic volatility models given by

$$\begin{cases} dF_t = \sigma_{\text{slv}}(t, F_t)C(F_t, v_t)\, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(t, v_t)\, dt + \nu(t, v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\, dt. \end{cases}$$

The function $\sigma_{\text{slv}}(\cdot,\cdot)$ is called the leverage and $C(\cdot)$ the stochastic volatility backbone. Thus, such models are essentially a mix of a stochastic and a local volatility component. In most cases the leverage will not only depend on the process F_t but also on its law. From a mathematical perspective the model then falls into the category of a McKean SDE and the discussion of the existence of solutions becomes very challenging and is for most cases still an open issue. For our analysis we progress assuming the existence of a solution and refer to Guyon and Henry-Labordère [2011], Jourdain and Zhou [2020] for further details on this topic.

In the Markovian projection literature the Heston Heston [1993], with backbone $C(s, v) = \sqrt{v}s$, as well as the SABR Hagan et al. [2002], with backbone $C(s, v) = vs^{\beta}$, stochastic volatility models have been considered, see, e.g., Henry-Labordère [2009], Guyon and Henry-Labordère [2012], Van der Stoep et al. [2014], Ren et al. [2007]. More recently, Muguruza [2019] extended the framework to cover rough stochastic volatility models. The main difficulty in calibration is the derivation of the leverage function, $\sigma_{\text{slv}}(\cdot,\cdot)$. Considering a given local volatility function $\sigma_{\text{mkt}}(\cdot,\cdot)$ characterizing the underlying market, it is possible to apply the Markovian projection technique Gyöngy [1986], Andersen and Piterbarg [2010c] and project the stochastic local volatility model onto a local volatility model resulting in an expression of the form

$$\sigma_{\text{mkt}}^2(t, x) = \sigma_{\text{slv}}(t, x)^2 \mathbb{E}^{\text{slv}}\big[C(x, v_t)^2 | F_t = x\big]. \tag{5.1.1}$$

Hence, considering the Heston and the SABR stochastic local volatility models, the calibration of the leverage function can be performed if the conditional expectation of the variance (resp. volatility), given an asset price for the chosen model can be evaluated. The valuation of the conditional expectation is one of the main challenges when considering stochastic local volatility models and there have been many approaches suggested using, for example, a binning approach to estimate the conditional expectation during a Monte Carlo simulation Van der Stoep et al.

[2014], particle methods Guyon and Henry-Labordère [2012], Muguruza [2019] or direct PDE solutions Ren et al. [2007], Saporito et al. [2019].

In this paper we shall extend the class of general stochastic volatility models explored by Felpel *et al* Felpel et al. [2021], to incorporate a local volatility component and demonstrate that it is possible to efficiently calibrate to observed option prices using a direct approximation of the above conditional expectation through an explicit formula. This widens the scope of applicable models beyond the Heston or SABR models and can enrich the dynamics for modelling the evolution of the implied volatility surface. We focus on the calibration of the leverage function using a direct approximation of the conditional expectation under the stochastic local volatility model. Our proposed calibration methodology does not rely on solving a two dimensional PDE solution or a time consuming Monte Carlo approach, see Henry-Labordère [2009], Guyon and Henry-Labordère [2012], Van der Stoep et al. [2014], Ren et al. [2007]. For the direct approximation of the stochastic local volatility model, we instead propose a simplified version of the general framework used in Felpel et al. [2021], approximating $\sigma_{\text{slv}}(t, F_t)$ by a piecewise constant function that is separable, i.e., $\sigma_{\text{slv}}(t, F_t) = \sigma(t)L(F_t)$. While at first glance this separability may seem restrictive, we show that in practice this assumption is not egregious and that it allows fast and robust calibration.

In summary, our contributions in this paper are as follows. We start by extending Theorem 1 of Felpel et al. [2021] to include the general separable SLV model with $\sigma_{\text{slv}}(t, F_t) = \sigma(t)L(F_t)$. This allows fast explicit calibration. To enable this, we let $\sigma(\cdot)$ and $L(\cdot)$ be piecewise constant functions. We then calibrate $\sigma(\cdot)$ to the at-the-money (ATM) values along a maturity grid and $L(\cdot)$ along the strike grid for a fixed maturity T. The base parameters of the model are calibrated to additional points to provide a good fit to the whole volatility surface or directly to market implied volatilities. To ensure computational efficiency we extend the one-dimensional PDE considered in Felpel et al. [2021], Kienitz et al. [2017], Hagan [2013] to our setting. To demonstrate the approach, we consider many examples using the SABR, Heston and ZABR models.

The paper is structured as follows. Section 5.2 lays the theoretical groundwork by extending results from Felpel et al. [2021]. In particular we prove that these results hold true for separable time-dependent diffusion coefficients. Given these preliminaries, Section 5.3 provides the details of the chosen approach for calibrating the general local stochastic volatility model. With the results from Sections 5.2 and 5.3 we propose an implementation based on solving a one-dimensional PDE in Section 5.4, state the discretization scheme and briefly contrast our approach to other known methods. In particular we consider the approaches Bang [2019] and Muguruza [2019]. We apply our methods to several models described in Sec-

tion 5.5 and give numerical examples that show the performance of our results in Section 5.6. We conclude with a summary of our results and consider topics for future research in Section 5.7.

5.2 Theoretical Foundation

In this section we extend Theorem 1 of Felpel et al. [2021] to allow additional time dependence of the coefficients of the underlying SDE. This allows the application of an effective partial differential equation to the more general class of stochastic local volatility models. To be precise, we shall consider general time-dependent stochastic volatility models of the form

$$\begin{cases} dF_t = \sigma(t)C(F_t)v_t\, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(t, v_t)\, dt + \nu(t, v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\, dt. \end{cases}$$

In comparison with the theorem as stated in Felpel et al. [2021], this set-up introduces time dependence for the functions $\mu(\cdot,\cdot)$ and $\nu(\cdot,\cdot)$, as well as the freedom of an additional function $\sigma(\cdot)$. The latter provides an additional time dependence for the forward dynamics. To apply our methodology we impose some regularity assumptions on the model. In particular, the following assumptions are required:

Assumption IX. The function $\sigma(\cdot) \neq 0$, and the function

$$\zeta(t) = \int_{t_0}^{t} \sigma(u)^2\, du$$

is well defined. Furthermore, the inverse function $\tilde{t}(\zeta)$ exists such that

$$\zeta(t) = \zeta \qquad \Leftrightarrow \qquad t = \tilde{t}(\zeta).$$

Assumption X. The drift term, $\mu(\cdot,\cdot)$, is differentiable with regard to its second argument and a solution $Y(t, t_0, \alpha)$ to the following PDE exists:

$$\begin{cases} \partial_t Y(t, t_0, \alpha) = \mu(t, Y(t, t_0, \alpha)), \\ Y(t, t, \alpha) = \alpha, \\ Y(t_0, t_0, \alpha) = \alpha. \end{cases}$$

Assumption XI. The function Y is differentiable and has an inverse function $y(t_0, t, a)$ such that

$$Y(t, t_0, \alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0, t, a).$$

Assumption XII. The functions

$$X(t,t_0,\alpha)=\partial_\alpha Y(t,t_0,\alpha),\qquad Z(t,u)=Z(t,u,t_0,\alpha)=y(u,t,Y(t,t_0,\alpha)),$$
$$z(F)=\int_f^F\frac{1}{C(u)}\,du,\qquad s(t)=S(t_0,t,\alpha)=\int_{t_0}^t\sigma(u)^2Z(t,u)^2\,du$$

and

$$\psi(t,u,Z)=\nu\big(u,Z(t,u)\big)Z(t,u)X\big(t,u,Z(t,u)\big)$$

are well defined, $X\big(t,u,Z(t,u)\big)^{-1}$ exists, and the following integral functions are defined:

$$I_1(t)=\rho\int_{t_0}^t\sigma(x)\psi(t,x,Z)\,dx,$$
$$I_2(t)=2\int_{t_0}^t\nu\big(x,Z(t,x)\big)^2X\big(t,x,Z(t,x)\big)^2\int_x^t\sigma(w)^2Z(t,w)X\big(t,w,Z(t,w)\big)^{-1}\,dw\,dx,$$
$$I_3(t)=\rho\int_{t_0}^t\sigma(x)\psi(t,x,Z)\int_x^t\sigma(w)^2Z(t,w)X\big(t,w,Z(t,w)\big)^{-1}\,dw\,dx,$$
$$I_4(t)=\rho^2\int_{t_0}^t\sigma(x)\psi(t,x,Z)\int_x^t\sigma(w)\partial_Z\Big(\psi(t,w,Z)\Big)X\big(t,w,Z(t,w)\big)^{-1}\,dw\,dx,$$
$$I_5(t)=\int_{t_0}^t\nu\big(x,Z(t,x)\big)^2X\big(t,x,Z(t,x)\big)^2\,dx.$$

Assumption XIII. The function $C(\cdot)$ is differentiable at f, with its derivative denoted by $C'(\cdot)$.

With these assumptions we now provide a theorem for general time-dependent stochastic volatility models.

Theorem 5.2.1. *Given that the general time dependent stochastic volatility model* (5.2) *obeys Assumptions* IX–XIII*, an effective PDE for the effective probability,* $Q(\cdot,\cdot)$*, of the form*

$$\begin{cases}\partial_tQ(t,F)=\partial_{FF}\big[D(t,F)Q(t,F)\big],\\ Q(t,F)\to\delta(F-f)\ \text{ as } t\to t_0^-,\end{cases}$$

can be derived, characterized by the function $D(\cdot,\cdot)$ *given by*

$$D(t,F)=\frac{1}{2}\sigma(t)^2C(F)^2a(t)^2e^{G(t)}\big(1+2b(t)z(F)+c(t)z(F)^2\big),$$

where the coefficients are specified as

$$a(t) = Y(t, t_0, \alpha), \qquad c(t) = b(t)^2 + \frac{1}{a(t)s(t)^2} I_2(t) - \frac{6b(t)}{s(t)^2} I_3(t) + \frac{2}{a(t)s(t)^2} I_4(t),$$
$$b(t) = \frac{1}{a(t)s(t)} I_1(t), \quad G(t) = -s(t)c(t) - s(t)b(t)\Gamma_0 + \frac{1}{a(t)^2} I_5(t)$$

and

$$\Gamma_0 = -C'(f).$$

In terms of the effective Markovian Projection (EMP), Felpel et al. [2022a], this result is expressed through a projected variance of the form

$$\sigma^2_{\text{proj}}(t, F) = \sigma(t)^2 C(F)^2 a(t)^2 e^{G(t)} \big(1 + 2b(t)z(F) + c(t)z(F)^2\big).$$

Proof. An outline of the proof is provided here, with complete detail given in Appendix 5.8. We start with a time change from t to $\zeta(t)$. This moves the dependence on the function $\sigma(\cdot)$ from the forward dynamics to the dynamics of the volatility. This enables us to deduce a simpler extension with only time-dependent functions $\mu(\cdot, \cdot)$ and $\nu(\cdot, \cdot)$. Given these results we perform a change of variables from $\zeta(t)$ back to t, allowing all results to be expressed in terms of t. □

Remark 5.2.2. To put this in the context of Equation (5.1.1), the result provides us with an approximation of the form

$$\mathbb{E}^{\text{slv}}\big[v_t^2 | F_t = x\big] \approx a(t)^2 e^{G(t)} \big(1 + 2b(t)z(x) + c(t)z(x)^2\big).$$

In particular, we highlight that the dependence of the leverage function on F_t is included in the transformed variable $z(x)$.

5.3 Calibration Algorithm

Having laid the mathematical foundations in Theorem 5.2.1, we apply the results to a general stochastic local volatility model (SLV) of the form

$$\begin{cases} dF_t = \sigma_{\text{slv}}(t, F_t) C(F_t) v_t \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(t, v_t)\, dt + \nu(t, v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases}$$

When the leverage function $\sigma_{\text{slv}}(\cdot,\cdot)$ is correctly specified, this allows a perfect match to the observed market surface. Note that this definition of an SLV model is formulated in a general framework in the style of Cui et al. [2018] and does not assume a Dupire local volatility term in accordance with Dupire [1994], Derman and Kani [1994], which can be found, for example, in Ren et al. [2007], Muguruza [2019].

To apply Theorem 5.2.1 to this class of models, we impose the assumption that the leverage function be separable:

Assumption XIV. The function $\sigma_{slv}(\cdot,\cdot)$ is separable and may be expressed as

$$\sigma_{slv}(t, F_t) = \sigma(t)L(F_t).$$

We construct a fast and accurate scheme to calibrate the leverage function to a given market surface. We assume a full characterization in terms of a reference local volatility model, where the market prices can be reconstructed using a model of the form

$$dF_t = \sigma_{\text{mkt}}(t, F_t)\, dW_t^{(1)}, \qquad F_{t_0} = f. \tag{5.3.1}$$

This means that for each maturity, T, and strike, K, there exists a projected market variance, $\sigma^2_{\text{mkt}}(T, K)$. In practice, the market is usually characterized in terms of an implied volatility surface. To deduce the corresponding local volatility surface, a calibration is required. We assume a representation in terms of (5.3.1) and not, as is also commonly observed, in terms of a Dupire local volatility, Dupire [1994], Derman and Kani [1994]. This allows us to capture negative values for the forward, and in the case where the forward stays positive the transformation from the projected volatility to the local volatility, $\sigma_{\text{Dupire}}(t, F_t)$, is provided by the expression

$$\sigma_{\text{mkt}}(t, F_t) = F_t \sigma_{\text{Dupire}}(t, F_t).$$

In practice, this curve is usually not given in functional form but is available as a set of discrete values on a specific grid. We assume that such a grid exists with maturities $\mathcal{T} = \{t_0, \ldots, t_M\}$ and strikes, $\mathcal{K} = \{K_0, \ldots, K_N\}$. For this grid the market local volatility function is defined. We also assume that the ATM value is always included in the grid of strikes and denote its corresponding index by i, i.e., $K_i = f$. One of the maturities, denoted by $T \in \mathcal{T}$, is selected to be the foundation of the calibration. The specific value or index of T is chosen by the modeller and tailored to the problem at hand. In particular, this allows calibration of the leverage function without any previous evaluation of certain model components.

5.3.1 Specifications of the Leverage Function

To calibrate the leverage function to market data, two components, the leverage $L(\cdot)$ and the time leverage $\sigma(\cdot)$, must be calibrated. To accomplish this we apply the EMP formulation of Theorem 5.2.1 and deduce an expression of the form

$$\sigma^2_{\text{mkt}}(t,x) = \sigma(t)^2 L(x)^2 C(x)^2 a(t)^2 e^{G(t)}\big(1 + 2b(t)z(x) + c(t)z(x)^2\big), \tag{5.3.2}$$

where the coefficients a, b, c and G are specified in Theorem 5.2.1 and the function z is given by

$$z(x) = \int_f^x \frac{1}{L(u)C(u)}\, du.$$

In order to apply Theorem 5.2.1, we need the condition that $C_{\text{slv}}(\cdot) \coloneqq L(\cdot)C(\cdot)$ is differentiable at the point f. To incorporate this condition together with the initial market data being defined only on a discrete set of strikes, we assume that the leverage $L(\cdot)$ as well as the time leverage $\sigma(\cdot)$ are piecewise constant. To be more precise, let us define the following interval system I_n depending on the ATM-index i and the underlying discrete market grid:

$$I_n = \begin{cases} (-\infty, K_0] & \text{if } n = 0, \\ (K_{n-1}, K_n] & \text{if } 1 \le n < i, \\ (K_{i-1}, K_{i+1}) & \text{if } n = i, \\ [K_n, K_{n+1}) & \text{if } i < n \le N-1, \\ [K_N, \infty) & \text{if } n = N. \end{cases}$$

On this system of intervals we define the leverage $L(\cdot)$ as

$$L(u) = l_n \text{ if } u \in I_n.$$

For the time leverage $\sigma(\cdot)$, the corresponding discrete grid is defined as

$$J_m = \begin{cases} [t_m, t_{m+1}) & \text{if } 0 \le m \le M-1, \\ [t_M, \infty) & \text{if } m = M, \end{cases}$$

with

$$\sigma(u) = \sigma_m \text{ if } u \in J_m.$$

The assumption of a piecewise constant leverage function is common in practice, see, e.g., Muguruza [2019], and the leverage function is typically defined on a discrete space structure.

5.3.2 Calibration Procedure

To calibrate the leverage function, we first assume that the underlying parameters of the base model are already determined, i.e., $C(\cdot)$, $\mu(\cdot,\cdot)$, $\nu(\cdot,\cdot)$ and ρ are fully specified. Given this set of parameters, we demonstrate how to calibrate the leverage function explicitly. In a second step, the underlying parameters of the base model themselves can be calibrated. Since the calibration of the leverage function, given a set of parameters, is explicit, it can be performed in each step of the calibration procedure to deduce the base parameters. For our calibration scheme we apply the following two step procedure to deduce the leverage function, given a set of base parameters:

i) In the first step we calibrate the time leverage $\sigma(\cdot)$ to the ATM value $f = K_i$ along all maturities of the market surface. This ensures a perfect fit to the values $\sigma^2_{\text{mkt}}(t_m, f)$ for all m.

ii) Given $\sigma(\cdot)$, we calibrate the leverage function, $L(\cdot)$, to all strike values for the pre-specified fitting maturity T. This ensures a perfect fit to the values $\sigma^2_{\text{mkt}}(T, K_n)$ for all n.

5.3.2.1 Calibrating $\sigma(\cdot)$

To calibrate the time leverage to the ATM value $f = K_i$ along the maturity grid, we use (5.3.2) which is further simplified in the ATM case to

$$\sigma^2_{\text{mkt}}(t, f) = \sigma(t)^2 L(f)^2 C(f)^2 a(t)^2 e^{G(t)}.$$

Notice that in our scheme, $\sigma^2_{\text{mkt}}(T, f)$ is characterized by the function $\sigma(T)$ as well as $L(f)$. We therefore decide to assign the whole dependence to $\sigma(T)$ and set

$$L(f) = L(K_i) = l_i = 1. \tag{5.3.3}$$

With this we may evaluate the value of $\sigma(t_m)^2$ for each m as

$$\sigma(t_m)^2 = \frac{\sigma^2_{\text{mkt}}(t_m, f)}{C(f)^2 a(t_m)^2 e^{G(t_m)}}. \tag{5.3.4}$$

5.3.2.2 Calibrating $L(\cdot)$

Having specified the function $\sigma(t_m)^2$ for each m in (5.3.4), the value for $\sigma(T)^2$ is known since T is assumed to be included in the time grid. Considering (5.3.2) at

time T and inserting expression (5.3.4), yields

$$\begin{aligned} L(x)^2 &= \frac{\sigma^2_{\text{mkt}}(T,x)}{\sigma(T)^2 C(x)^2 a(T)^2 e^{G(T)}\big(1+2b(T)z(x)+c(T)z(x)^2\big)} \\ &= \frac{\tilde{\sigma}^2_{\text{mkt}}(T,x)}{\tilde{C}(x)^2\big(1+2b(T)z(x)+c(T)z(x)^2\big)}, \end{aligned}$$

where

$$\tilde{\sigma}^2_{\text{mkt}}(T,x) = \frac{\sigma^2_{\text{mkt}}(T,x)}{\sigma^2_{\text{mkt}}(T,f)} \qquad \text{and} \qquad \tilde{C}(x) = \frac{C(x)}{C(f)}.$$

Considering the interval system I_n, the value of the function $L(\cdot)$ on this interval, l_n, may be characterized as

$$l_n^2 = L(K_n)^2 = \frac{\tilde{\sigma}^2_{\text{mkt}}(T,K_n)}{\tilde{C}(K_n)^2\big(1+2b(T)z(K_n)+c(T)z(K_n)^2\big)}.$$

Thus, only the evaluation of the function $z(K_n)$ is necessary. This is performed recursively for the starting value K_i, yielding

$$z_n = z(K_n) = \begin{cases} z_{n+1} - \frac{1}{l_{n+1}} \int_{K_n}^{K_{n+1}} \frac{1}{C(u)}\, du & \text{if } n < i, \\ 0 & \text{if } n = i, \\ z_{n-1} + \frac{1}{l_{n-1}} \int_{K_{n-1}}^{K_n} \frac{1}{C(u)}\, du & \text{if } n > i. \end{cases} \tag{5.3.5}$$

In particular we already know the value of l_i in (5.3.3). Since the value of z_n, for $n > i$, only depends on the values up to l_{n-1}, we successively evaluate all functions and compute the values l_n. Respectively, the value of z_n, for $n < i$, only depends on the values from l_{n+1} and, thus, we can evaluate all values l_n recursively.

Remark 5.3.1. Notice that the values of l_n are positive by definition. This is important to guarantee an arbitrage free model.

5.3.2.3 Calibrating additional parameters

Considering stochastic local volatility models, calibration is achieved in two distinct phases. The underlying base model is calibrated to market data in order to provide its parameters. Based on these parameters, the leverage function is then calibrated to further improve the fit of the model. Since our approach allows fast calibration of the leverage function without the need to directly solve the

full model, a combined calibration of the base model and the leverage function is possible. For example, to calibrate the underlying base parameters we specify an additional maturity $\tilde{T} \in \mathcal{T}$ and determine these parameters to further increase the fit along this maturity by minimizing the mean squared error

$$\mathsf{mse} = \sum_{n \geq 0} \left(\sigma^2_{\mathrm{mkt}}(\tilde{T}, K_n) - \sigma^2_{\mathrm{loc}}(\tilde{T}, K_n) \right)^2.$$

Alternatively, we directly consider the implied volatility function and calibrate the underlying base parameters to minimize the differences between the implied volatilities. Here, the calibration of the leverage function can be performed depending on the changing base parameters to provide the optimal fit of the complete stochastic local volatility model to the market data.

5.4 PDE Implementation

Having calibrated the leverage functions as described in Section 5.3, all necessary functions for the evaluation of the projected variance function required by Theorem 5.2.1 are given. We proceed to compute the effective probability density function, $Q(\cdot, \cdot)$, by solving the effective PDE

$$\begin{cases} \partial_t Q(t,F) = \partial_{FF}\left[\frac{1}{2} \sigma(t)^2 L(F)^2 C(F)^2 a(t)^2 e^{G(t)} \big(1 + 2b(t)z(F) + c(t)z(F)^2\big) Q(t,F) \right], \\ Q(t,F) \to \delta(F - f) \text{ as } t \to t_0^-. \end{cases}$$

This partial differential equation can be solved using a classical PDE method such as Duffy [2006], Tavella and Randall [2000], Smith [1985]. For our further studies, however, we wish to embed the stochastic local volatility models into the same numerical framework used in Felpel et al. [2021, 2022a]. More precisely, this means that we apply the conservative one-dimensional PDE scheme introduced by Hagan [2015] to solve the effective PDE. To apply this scheme, the effective PDE should be expressed in the form

$$\begin{cases} \partial_t Q(t,F) = \partial_{FF}\left[\frac{1}{2} D(F)^2 E(t) Q(t,F) \right], \\ Q(t,F) \to \delta(F - f) \text{ as } t \to t_0^-. \end{cases}$$

Since the coefficients $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are truly time dependent, the term $D(\cdot)^2$ becomes trivial and the PDE scheme is not applicable in full strength. To overcome

this problem, we apply an additional approximation to the projected variance function and approximate the coefficients $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ by piecewise constant functions. This is achieved by freezing the coefficients over each increment of the underlying discrete time grid, so that on the time interval $J_i = [t_i, t_{i+1})$ the projected variance function is given by

$$\sigma^2_{\text{proj}}(t, F) = \sigma(t_i)^2 L(F)^2 C(F)^2 a(t_i)^2 e^{G(t)} \big(1 + 2b(t_i)z(F) + c(t_i)z(F)^2\big).$$

Given this form, the function $D(\cdot)$ on the interval J_i is given by

$$D(F) = D_i(F) \coloneqq \sigma(t_i)L(F)C(F)a(t_i)\sqrt{1 + 2b(t_i)z(F) + c(t_i)z(F)^2},$$

and we apply the conservative one-dimensional PDE scheme to evolve the effective density function from t_i to t_{i+1}. We further demonstrate how the necessary transformations embedded in the scheme are applicable to our new set-up and how to transition between time intervals J_i and J_{i+1}.

Remark 5.4.1. The freezing of the coefficients is applied at the initial time of each interval. The only exception to this rule is at the initial time t_0. Here the coefficients are instead taken as the right limit when $t \to t_0^-$.

5.4.1 Functions for the Conservative PDE Scheme

For the application of the conservative one-dimensional PDE scheme, the underlying PDE is reformulated by the space transformation through the Lamperti transform

$$y_{\text{slv}}(F) = \int_f^F \frac{1}{D(z_{\text{slv}}(u))}\, du,$$

where

$$z_{\text{slv}}(F) = \int_f^F \frac{1}{C_{\text{slv}}(u)}\, du$$

and $C_{\text{slv}}(u) = L(u)C(u) = L(u)C_{\text{base}}(u)$. To apply the scheme, we evaluate these functions as well as their inverse functions, $F_{\text{slv}}(\cdot)$ and $Z_{\text{slv}}(\cdot)$. This is achieved by expressing these functions in terms of the corresponding functions of the underlying base models. In the following we assume that we are working on the time interval J_i.

5.4.1.1 The Function $z_{\text{slv}}(\cdot)$

To evaluate the function $z_{\text{slv}}(\cdot)$ in terms of the base function $z_{\text{base}}(\cdot)$, we recall that

$$z_{\text{slv}}(F) = \int_f^F \frac{1}{C_{\text{slv}}(u)}\,du = \int_f^F \frac{1}{L(u)C_{\text{base}}(u)}\,du.$$

Considering the interval system I_n, we identify in which interval the value of F is contained. Consequently, we define the index n_F such that $F \in I_{n_F}$. This allows three possible outcomes: $n_F = i$, $n_F > i$ or $n_F < i$. Evaluating each case, leads to the expression

$$z_{\text{slv}}(F) = \frac{1}{l_{n_F}} z_{\text{base}}(F) + M(n_F), \tag{5.4.1}$$

where the function $M(\cdot)$ is dependent on the index n_F given by

$$M(n_F) = \begin{cases} \sum_{j=n_F}^{i-1} \left(\frac{1}{l_{j+1}} - \frac{1}{l_j}\right) z_{\text{base}}(K_j) & \text{if } n_F < i, \\ 0 & \text{if } n_F = i, \\ \sum_{j=i+1}^{n_F} \left(\frac{1}{l_{j-1}} - \frac{1}{l_j}\right) z_{\text{base}}(K_j) & \text{if } n_F > i. \end{cases}$$

5.4.1.2 The Function $F_{\text{slv}}(\cdot)$

We investigate the inverse function $F_{\text{slv}}(\cdot)$ of $z_{\text{slv}}(\cdot)$ and begin with a transformation of the intervals I_n using the z_{slv} variable and define the transformed system of intervals I_n^z by

$$I_n^z = \begin{cases} (-\infty, z_{\text{slv}}(K_0)] & \text{if } n = 0, \\ (z_{\text{slv}}(K_{n-1}), z_{\text{slv}}(K_n)] & \text{if } 1 \le n < i, \\ (z_{\text{slv}}(K_{i-1}), z_{\text{slv}}(K_{i+1})) & \text{if } n = i, \\ [z_{\text{slv}}(K_n), z_{\text{slv}}(K_{n+1})) & \text{if } i < n \le N-1, \\ [z_{\text{slv}}(K_N), \infty) & \text{if } n = N. \end{cases}$$

By construction, we know that the values of l_n are positive as well as $C_{\text{slv}}(\cdot)$, the function $z_{\text{slv}}(\cdot)$ is increasing and, for $F_2 > F_1$, we have

$$z_{\text{slv}}(F_2) - z_{\text{slv}}(F_1) = \int_{F_1}^{F_2} \frac{1}{L(u)C_{\text{base}}(u)}\,du > 0.$$

This implies a bijection between intervals I_n and I_n^z, and we have the equivalence relation

$$n_F = m \quad \Longleftrightarrow \quad F \in I_m \quad \Longleftrightarrow \quad z_{\text{slv}}(F) \in I_m^z \quad \Longleftrightarrow \quad n_z = m.$$

In other words, this allows us to replace the index n_F, associated with F, with index n_z, associated with $z_{\text{slv}}(F)$. This yields

$$z = z_{\text{slv}}(F) = \frac{1}{l_{n_F}} z_{\text{base}}(F) + M(n_F) = \frac{1}{l_{n_z}} z_{\text{base}}(F) + M(n_z),$$

and for the inverse function

$$F_{\text{slv}}(z) = F_{\text{slv}}(l_{n_z}(z - M(n_z))). \tag{5.4.2}$$

5.4.1.3 The Function $y_{\text{slv}}(\cdot)$

To evaluate $y_{\text{slv}}(\cdot)$ we note that $D(\cdot)$ is expressed as

$$D(u) = C_{\text{slv}}(u) h_{\text{slv}}(z_{\text{slv}}(u)),$$

with

$$h_{\text{slv}}(u) = \sigma(t_i) a(t_i) \sqrt{1 + 2b(t_i)u + c(t_i)u^2}.$$

To compute the Lamperti transform we note that the derivatives with respect to F are given by

$$\begin{aligned}
\partial_F \big[z_{\text{slv}}(F) \big] &= \frac{1}{C_{\text{slv}}(F)}, \\
\partial_F \big[y_{\text{slv}}(F) \big] &= \partial_F \big[z_{\text{slv}}(F) \big] h^{-1}(z_{\text{slv}}(F)).
\end{aligned}$$

In particular, a solution of the last equation is expressed as $y_{\text{slv}}(F) = H_{\text{slv}}(z_{\text{slv}}(F))$, provided that H_{slv} solves

$$H'_{\text{slv}}(x) = h_{\text{slv}}^{-1}(x).$$

With the given form for h_{slv} the function H_{slv} is given by

$$H_{\text{slv}}(x) = \frac{1}{\sigma(t_i) a(t_i) \sqrt{c(t_i)}} \log \left(\frac{\sqrt{1 - \frac{b(t_i)^2}{c(t_i)} + M(x)^2} + M(x)}{1 + \frac{b(t_i)}{\sqrt{c(t_i)}}} \right),$$

with $M(x) = \sqrt{c(t_i)} x + \frac{b(t_i)}{\sqrt{c(t_i)}}$. In particular, an alternative formulation of the function H_{slv} is provided by

$$H_{\text{slv}}(x) = \frac{-1}{\sigma(t_i) a(t_i) \sqrt{c(t_i)}} \log \left(\frac{\sqrt{1 - \frac{b(t_i)^2}{c(t_i)} + M(x)^2} - M(x)}{1 - \frac{b(t_i)}{\sqrt{c(t_i)}}} \right).$$

5.4.1.4 The Function $Z_{\text{slv}}(\cdot)$

To deduce the inverse function of $y_{\text{slv}}(\cdot)$ we consider the inverse function, $Z_{\text{slv}}(\cdot)$, of $H_{\text{slv}}(\cdot)$. For this function we know that

$$Z'_{\text{slv}}(y) = h_{\text{slv}}(Z_{\text{slv}}(y)) = \sigma(t_i)a(t_i)\sqrt{1 + 2b(t_i)Z_{\text{slv}}(y) + c(t_i)Z_{\text{slv}}(y)^2}.$$

This leads to

$$Z''_{\text{slv}}(y) = (\sigma(t_i)a(t_i))^2 b(t_i) + (\sigma(t_i)a(t_i))^2 c(t_i) Z_{\text{slv}}(y).$$

To solve the last equation we use the ansatz:

$$Z_{\text{slv}}(y) = A\sinh(\sigma(t_i)a(t_i)\sqrt{c(t_i)}y) + B\cosh(\sigma(t_i)a(t_i)\sqrt{c(t_i)}y) - C.$$

Differentiating twice we have that $C = \frac{b(t_i)}{c(t_i)}$. Since $Z_{\text{slv}}(0) = 0$, we conclude that $B = \frac{b(t_i)}{c(t_i)}$. Finally, inserting $H_{\text{slv}}(z)$ into Z_{slv} yields z_{slv}, and we have $A = \frac{1}{\sqrt{c(t_i)}}$. Combining the latter, this leads to the function

$$Z_{\text{slv}}(y) = \frac{1}{\sqrt{c(t_i)}}\sinh(\sigma(t_i)a(t_i)\sqrt{c(t_i)}y) + \frac{b(t_i)}{c(t_i)}(\cosh(\sigma(t_i)a(t_i)\sqrt{c(t_i)}y) - 1).$$

5.4.2 Switching Time Regimes

Given the functions in Section 5.4.1, we have established the necessary framework to apply the conservative one-dimensional PDE scheme to the time interval $J_i = [t_i, t_{i+1})$. We apply the same procedure to any other interval, e.g., $J_{i+1} = [t_{i+1}, t_{i+2})$. It remains to show how the output generated on the interval J_i is used to construct the initial state for the interval J_{i+1}. To accomplish this, we recall the approach used in the PDE scheme of Hagan [2015]. First, the Lamperti transform $y_{\text{slv}}(\cdot)$ is used to transform the space. For the new variable, denoted by y, a uniform grid with step size h is constructed, and in each grid interval the average probability density is computed as

$$\theta_j = \frac{1}{h}\int_{F(y_{j-1})}^{F(y_j)} Q(t, F)\, dF.$$

The term $F(y_{j-1})$ denotes the abbreviation of $F_{\text{slv}}(Z_{\text{slv}}(y_{j-1}))$. For these values of θ_j a PDE scheme is constructed in the new space. Applying this procedure to the time interval J_i yields the following variables at time t_{i+1}:

i) A grid $(y_j^i)_j$ with boundaries y_0^i and y_N^i as well as the step size h^i.

ii) The inner grid values $\theta_j^i(t_{i+1})$ at the time t_{i+1} corresponding to the grid $(y_j^i)_j$.

iii) The accumulating boundary values $\theta_0^i(t_{i+1})$ and $\theta_N^i(t_{i+1})$ at time t_{i+1}.

For the next time interval, J_{i+1}, the values $\theta_j^i(t_{i+1})$, expressed on the grid y^i, must be transformed to values $\theta_j^{i+1}(t_{i+1})$, which are expressed in the grid y^{i+1} induced by the interval J_{i+1}. For this, we approximate the density function $Q(t,F)$ using a piecewise constant approximation based on the average probability functions of the grid y^i as

$$Q(t,F) \approx \sum_{n=1}^{N-1} \frac{h^i \theta_n^i(t)}{F^i(y_n^i) - F^i(y_{n-1}^i)} \mathbb{1}_{[F^i(y_{n-1}^i), F^i(y_n^i))}(F).$$

With the above, the values $\theta_j^{i+1}(t_{i+1})$ on the space y^{i+1} are given by

$$\begin{aligned}
&\theta_j^{i+1}(t_{i+1}) \\
&\quad = \frac{1}{h^{i+1}} \int_{F^{i+1}(y_{j-1}^{i+1})}^{F^{i+1}(y_j^{i+1})} Q(t_{i+1}, F)\, dF \\
&\quad \approx \frac{1}{h^{i+1}} \sum_{n=1}^{N-1} \frac{h^i \theta_n^i(t_{i+1})}{F^i(y_n^i) - F^i(y_{n-1}^i)} \int_{F^{i+1}(y_{j-1}^{i+1})}^{F^{i+1}(y_j^{i+1})} \mathbb{1}_{[F^i(y_{n-1}^i), F^i(y_n^i))}(F)\, dF \\
&\quad = \frac{h^i}{h^{i+1}} \sum_{n=1}^{N-1} \theta_n^i(t_{i+1}) \frac{\left(\min(F^{i+1}(y_j^{i+1}), F^i(y_n^i)) - \max(F^{i+1}(y_{j-1}^{i+1}), F^i(y_{n-1}^i))\right)^+}{F^i(y_n^i) - F^i(y_{n-1}^i)}.
\end{aligned}$$

To allow accumulation of probability in the boundaries, we consider a grid configuration where the boundary values of the F-grids do not change when switching between time intervals. In this set-up the accumulated probability is transferred to the new boundary values y_0^{i+1} and y_N^{i+1}. For our numerical implementation we take the F-boundaries generated on the first time interval as the corresponding boundaries for all other time intervals.

Remark 5.4.2. Notice that the transformation from grid $(y_j^i)_j$ to $(y_j^{i+1})_j$ is consistent in the sense that, if the underlying grids are the same, no modification to the density values is performed.

5.4.3 Embedding into Existing Methods

To place our approach into the existing frameworks for calibrating and computing stochastic local volatility models, we differentiate between three different ap-

proaches. These approaches are based on the use of PDE methods, Monte Carlo methods, and analytical approximations.

5.4.3.1 PDE based methods

In essence, our method clearly falls into the category of PDE-based methods. There are, however, some structural differences compared to the standard approaches provided by, e.g., Ren et al. [2007], Saporito et al. [2019]. In these approaches, a general leverage function of the form $\sigma_{\text{slv}}(t, x)$ is assumed with no additional separability property. On the one hand, this provides more degrees of freedom and the possibility of a better fit to the complete surface. On the other hand, however, this makes the calibration procedure much more complicated, since the conditional expectation necessary to determine $\sigma_{\text{slv}}(\cdot, \cdot)$ may not be as easily computed (c.f. Saporito et al. [2019]). Consequently, the calibration of the leverage function, and in turn the evaluation of the conditional expectation, is performed while solving the PDE equation for the probability density function. This results in an iterative procedure of the following form to update the leverage function based on the results available at time t_n:

i) Update the density function using an approximation depending only on $\sigma_{\text{slv}}(t_n, \cdot)$. This yields the new density function $p_{t_{n+1}}$.

ii) Using the density $p_{t_{n+1}}$ update the leverage function $L(t_{n+1}, \cdot)$.

Thus, in comparison, our approach allows the separation of calibration and PDE evaluation. This results in a much faster calibration scheme and allows one to combine the calibration of the underlying base model and the leverage function into one optimization step. This, however, comes at the cost of fewer degrees of freedom, since the leverage function has the additional constraint of separability.

5.4.3.2 Monte Carlo based methods

The second category are methods based on Monte Carlo. Since these follow a completely different numerical approach, it would be inappropriate to provide a detailed exposition. We do, however, use the approach provided by Muguruza [2019] as a benchmark for validating our results in the numerical examples of Section 5.6.3. Hence, we provide the general idea of some prominent methodologies such as Van der Stoep et al. [2014], Guyon and Henry-Labordère [2012], Muguruza [2019]. For this let us consider a framework where we have simulated M Monte Carlo paths $(F_{t_m,u}, v_{t_m,u})_{u \leq M}$ at some time t_m under the chosen stochastic local volatility model. To proceed to the next time step t_{m+1} the corresponding

leverage function $\sigma_{\text{slv}}(t_m, F_t)$ needs to be calibrated. Assuming, for illustration, a specification of the form

$$C(F_t, v_t) = \sqrt{v_t} F_t$$

the calibration boils down to an evaluation of the conditional expectation $\mathbb{E}^{\text{slv}}\big[v_{t_m} | F_{t_m} = x\big]$. Prominent methods to do this are the following:

i) Following Van der Stoep et al. [2014] a binning methodology can be used to estimate the conditional expectation. Extending the condition from $F_{t_m} = x$ to a condition of the form $F_{t_m} \in B$ where B is some chosen interval containing x, multiple paths become valid to compute the expression.

ii) Another prominent method is the particle method of Guyon and Henry-Labordère [2012]. Here the conditional expectation is estimated using a Nadaraya-Watson kernel K_h with a bandwidth h resulting in an expression of the form

$$\mathbb{E}^{\text{slv}}\big[v_{t_m} | F_{t_m} = x\big] \approx \sum_{i=1}^{M} v_{t_m,u} K_h(F_{t_m,u} - x).$$

iii) Instead of using a general Nadaraya-Watson kernel, a direct estimation of the conditional distribution can be performed as done in Muguruza [2019]. This allows the computation of the conditional expectation using the sampled paths and distribution estimation along these paths.

At this point we also highlight that in our comparison we are assuming a setting where both approaches, i.e., our methodology and the one of Muguruza [2019], are applicable. Since many approaches, including Muguruza [2019], consider stochastic local volatility models with regard to a Dupire local volatility, this imposes the restriction that the forward stays positive. Therefore a direct application is not available when considering models which admit negative forward values. These models can be considered using our approach. On the contrary, the approach of Muguruza [2019] allows the inclusion of drift terms into the forward dynamics, which in turn is not included in our considerations.

5.4.3.3 Analytical expressions

The last category includes methods based on analytical expressions. One prominent example is the approach of Bang [2019]. This method provides accurate approximation of stochastic volatility models using closed form expressions. The main difference when compared with our approach is the underlying class of applicable models and the purpose of the approximation. Bang only focuses on pricing

options for a fixed maturity. This means that the leverage function only has dependence on the rate considered and no additional time dependence. Consequently, each relevant maturity may have a different parametrization. In comparison, our approach models dynamical behaviour and incorporates time dependence in the leverage function, allowing all relevant maturities to be modelled using a single parametrization.

5.5 Model Examples

To demonstrate the applicability of our new approach, we consider stochastic local volatility models of the ZABR-type and the Heston-type as our primary examples.

5.5.1 The dZABR Model

For our first example we consider the displaced ZABR, dZABR, model of Andreasen and Huge [2011b] as the underlying base model. The slv-dZABR model is specified by the system of SDEs

$$\begin{cases} dF_t = \sigma_{\text{slv}}(t, F_t)(F_t + d)^{\beta} v_t \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \nu v_t^{\gamma} \, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases}$$

Special cases are provided by the slv-dSABR model where $\gamma = 1$ and the slv-nSABR, slv-normal-SABR, model where $\gamma = 1$ and $\beta = 0$. For the slv-dZABR model the coefficients of Theorem 5.2.1 are given by

$$\begin{aligned} a(t) &\equiv \alpha, \\ b(t) &= \rho\nu\alpha^{\gamma-2}\frac{\theta_1(t_0,t)}{\theta_2(t_0,t)}, \\ c(t) &= 2\nu^2\alpha^{2(\gamma-2)}\frac{\Omega_{0,2}(t_0,t)}{\theta_2(t_0,t)^2} \\ &\quad + \rho^2\nu^2\alpha^{2(\gamma-2)}\left[\frac{\theta_1(t_0,t)^2}{\theta_2(t_0,t)^2} - 6\frac{\theta_1(t_0,t)\Omega_{1,2}(t_0,t)}{\theta_2(t_0,t)^3} + 2(\gamma+1)\frac{\Omega_{1,1}(t_0,t)}{\theta_2(t_0,t)^2}\right], \\ G(t) &= -\alpha^2\theta_2(t_0,t)c(t) + \rho\nu\alpha^{\gamma}\beta(f+d)^{\beta-1}\theta_1(t_0,t) + \nu^2\alpha^{2(\gamma-1)}\theta_0(t_0,t), \end{aligned}$$

where the additional time dependent functions are defined as

$$\theta_i(x,t) = \int_x^t \sigma(u)^i \, du,$$
$$\Omega_{i,j}(t_0,t) = \int_{t_0}^t \sigma(x)^i \theta_j(x,t) \, dx.$$

Following the setting of Section 5.3 and Section 5.4, these functions are explicitly evaluated along the discrete-time grid of maturities, yielding

$$\theta_i(t_0,t_j) = \begin{cases} \sum_{m=0}^{j-1} \sigma_m^i \delta_m & \text{for } j \geq 1, \\ \theta_i(t_0,t_1) & \text{for } j = 0, \end{cases}$$

and

$$\Omega_{i,k}(t_0,t_j) = \begin{cases} \frac{1}{2}\sum_{n=0}^{j-1} \sigma_n^{i+k}\delta_n^2 + \sum_{n=0}^{j-1}\sum_{m=n+1}^{j-1} \sigma_m^k \sigma_n^i \delta_m \delta_n & \text{for } j \geq 2, \\ \frac{1}{2}\sum_{n=0}^{j-1} \sigma_n^{i+k}\delta_n^2 & \text{for } j = 1, \\ \Omega_{i,k}(t_0,t_1) & \text{for } j = 0, \end{cases}$$

where $\delta_n = t_{n+1} - t_n$. This in turn allows the evaluation of the coefficients along the time grid and an application of the numerical scheme presented in Section 5.4.

5.5.2 The Heston Model

A well known and popular model among practitioners in stochastic local volatility models is the Heston model, Van der Stoep et al. [2014], Muguruza [2019], Saporito et al. [2019]. Although our setting focuses on the application to stochastic volatility and not stochastic variance models we nevertheless demonstrate an application using the Heston model being the base model. In this special case, the corresponding approximations can be found in Hagan et al. [2018a]. For general stochastic variance models, the proof of Theorem 5.2.1 must be modified to incorporate the square root function of the volatility in the forward dynamics. This is possible, however, since we only focus on the Heston model, we refer to the special case that was demonstrated in Hagan et al. [2018a]. In particular, once the coefficients of Theorem 5.2.1 are derived, the procedure to solve the PDE and compute prices for a stochastic variance model is the same as the approach used for stochastic volatility models. For the slv-Heston model with a mean reversion back to initial

volatility, the corresponding coefficients are given by

$$
\begin{aligned}
a(t) &\equiv \sqrt{\alpha},\\
b(t) &= \frac{\rho\nu}{2\alpha}\frac{\varphi_1(t_0,t)}{\theta_2(t_0,t)},\\
c(t) &= \frac{\nu^2}{2\alpha^2}\frac{\Sigma_{0,2,2}(t_0,t)}{\theta_2(t_0,t)^2} - \frac{3\rho^2\nu^2}{2\alpha^2}\frac{\varphi_1(t_0,t)\Sigma_{1,1,2}(t_0,t)}{\theta_2(t_0,t)^3} + \frac{\rho^2\nu^2}{\alpha^2}\frac{\Omega_{1,1,1}(t_0,t)}{\theta_2(t_0,t)^2},\\
G(t) &= -\alpha\theta_2(t_0,t)(c(t)-b(t)),
\end{aligned}
$$

where the additional time dependent functions are defined as

$$
\begin{aligned}
\theta_i(x,t) &= \int_x^t \sigma(u)^i\,du,\\
\varphi_i(t_0,t) &= \int_{t_0}^t \sigma(u)^i e^{-\kappa(t-u)}\,du,\\
\tilde{\varphi}_i(t_0,t) &= \int_{t_0}^t \sigma(u)^i e^{\kappa(t-u)}\,du,\\
\Sigma_{i,j,k}(t_0,t) &= \int_{t_0}^t \sigma(u)^i e^{-j\kappa(t-u)}\tilde{\varphi}_k(u,t)\,du,\\
\Omega_{i,j,k}(t_0,t) &= \int_{t_0}^t \sigma(u)^i e^{-j\kappa(t-u)}\theta_k(u,t)\,du.
\end{aligned}
$$

In the setting of Section 5.3 and Section 5.4, these function are evaluated along the discrete time grid of maturities as

$$
\begin{aligned}
\theta_i(t_0,t) &= \sum_{n=0}^{j}\sigma_n^i\delta_n && \text{for } j\geq 1,\\
\varphi_i(t_0,t_j) &= \frac{1}{\kappa}e^{-\kappa t_j}\sum_{n=0}^{j-1}\sigma_n^i\tilde{\delta}_n(\kappa) && \text{for } j\geq 1,
\end{aligned}
$$

$$
\Sigma_{i,l,k}(t_0,t_j) = \begin{cases}
-\frac{e^{(1-l)\kappa t_j}}{l\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}e^{-\kappa t_{n+1}}\tilde{\delta}_n(l\kappa) + \frac{e^{(1-l)\kappa t_j}}{(l-1)\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}\tilde{\delta}_n((l-1)\kappa) & \\
\quad -\frac{e^{(1-l)\kappa t_j}}{l\kappa^2}\sum_{n=0}^{j-1}\sum_{m=n+1}^{j-1}\sigma_n^i\sigma_m^i\tilde{\delta}_n(l\kappa)\tilde{\delta}_m(-\kappa) & \text{for } j\geq 2, l\neq 1,\\
-\frac{e^{(1-l)\kappa t_j}}{l\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}e^{-\kappa t_{n+1}}\tilde{\delta}_n(l\kappa) + \frac{e^{(1-l)\kappa t_j}}{\kappa}\sum_{n=0}^{j-1}\sigma_n^{i+k}\delta_n & \\
\quad -\frac{e^{(1-l)\kappa t_j}}{l\kappa^2}\sum_{n=0}^{j-1}\sum_{m=n+1}^{j-1}\sigma_n^i\sigma_m^i\tilde{\delta}_n(l\kappa)\tilde{\delta}_m(-\kappa) & \text{for } j\geq 2, l=1,\\
-\frac{e^{(1-l)\kappa t_j}}{l\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}e^{-\kappa t_{n+1}}\tilde{\delta}_n(l\kappa) + \frac{e^{(1-l)\kappa t_j}}{(l-1)\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}\tilde{\delta}_n((l-1)\kappa) & \text{for } j=1, l\neq 1,\\
-\frac{e^{(1-l)\kappa t_j}}{l\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}e^{-\kappa t_{n+1}}\tilde{\delta}_n(l\kappa) + \frac{e^{(1-l)\kappa t_j}}{\kappa}\sum_{n=0}^{j-1}\sigma_n^{i+k}\delta_n & \text{for } j=1, l=1,
\end{cases}
$$

and

$$\Omega_{i,l,k}(t_0,t_j) = \begin{cases} -\frac{e^{-l\kappa t_j}}{l\kappa}\sum_{n=0}^{j-1}\sigma_n^{i+k}e^{l\kappa t_n}\delta_n + \frac{e^{-l\kappa t_j}}{l^2\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}\tilde{\delta}_n(l\kappa) \\ \qquad\qquad \frac{e^{-l\kappa t_j}}{l\kappa}\sum_{n=0}^{j-1}\sum_{m=n+1}^{j-1}\sigma_n^i\sigma_m^i\tilde{\delta}_n(l\kappa)\delta_m & \text{for } j \geq 2, \\ -\frac{e^{-l\kappa t_j}}{l\kappa}\sum_{n=0}^{j-1}\sigma_n^{i+k}e^{l\kappa t_n}\delta_n + \frac{e^{-l\kappa t_j}}{l^2\kappa^2}\sum_{n=0}^{j-1}\sigma_n^{i+k}\tilde{\delta}_n(l\kappa) & \text{for } j = 1. \end{cases}$$

For all cases, we approximate the value for $j = 0$ with the value for $j = 1$.

5.6 Numerics

Having established the necessary framework to calibrate stochastic local volatility models and evaluate the relevant density functions, we now perform a series of numerical experiments to demonstrate the functionality.

5.6.1 The slv-nSABR Model

We start by considering the slv-nSABR. We simulate a reference set of market projected variances using a dZABR model with the parameters in Table 5.1 and the EMP of Felpel et al. [2022a].

Parameter	dZABR	nSABR_1	nSABR_2	nSABR_3
f	50 bp	50 bp	50 bp	50 bp
β	0.4	0	0	0
γ	0.9	1	1	1
α	$0.3f^{1-\beta}$	200 bp	200 bp	200 bp
ν	0.3	0.2	0.4	0.1
ρ	-0.7	-0.5	-0.7	-0.5
d	11 bp	11 bp	11 bp	11 bp

Table 5.1: Parametrization for the first example.

The "market data" is generated on a strike grid from $-10bp$ to $200bp$ with step size of $2bp$. The time grid starts at time zero and ends at a maturity of ten years with quarterly intervals. On these grids the leverage functions $\sigma(\cdot)$ and $L(\cdot)$ are calibrated as described in Section 5.3 with a fitting maturity of $T_{\text{fit}} = 1$.

Figure 5.1 shows the resulting projected variance surfaces evaluated on the calibration grid using the slv-nSABR model where the base model is parametrized using the nSABR_1 parametrization of Table 5.1. Using a finer grid for evaluation, the piecewise constant structure of the leverage functions becomes visible. Figure 5.2 shows the projected variance and the implied volatility for the fitting maturity and the largest maturity on the grid for the three different parametrizations of Table 5.1. We observe an excellent fit. In particular, for the fitting maturity it is shown that the leverage function $L(\cdot)$ is calibrated to exactly match the projected variance along the calibration grid.

It is observed that the underlying base parameters provide some additional control over the model behaviour. The different parametrizations yield different projected and implied volatility curves. In particular, this becomes visible for the large maturities. Considering the fitting maturity, the different parametrizations only impact the behaviour in between the calibration grid. On the calibration grid, all curves are the same. It is observed that the nSABR_2 parametrization provides the best fit to the implied volatility curve, providing an almost perfect match. At first glance this is surprising, since the other parametrizations yield a better fit along the projected variance curve for the same maturity. This can be explained by the fact that a much larger strike range is necessary to solve the PDEs for large maturities. Consequently, the projected variance function follows the underlying base model once the last valid calibration grid point is reached. Since the nSABR_3 parametrization has a rather flat extrapolation, the results of the nSABR_2 parametrization are better, with smaller error in the area beyond the calibration grid.

Remark 5.6.1. All the variants of the nSABR model with parameters given in Table 5.1 include a displacement parameter d. Owing to the specification $\beta = 0$, this value has no direct impact on the characterization of the stochastic differential equation. In the implementation of the PDE scheme of Section 5.4 this parameter is, however, used to characterize the boundary where the accumulation of the density takes place.

5.6.2 Error Distribution

To further illustrate calibration quality, we use a broad range of parameters, which cover a large selection of possible values. For this we randomly sample the parameters of a dZABR model from the range of parameters displayed in Table 5.2. These parameters are used to generate a reference set of market projected variances using the EMP of Felpel et al. [2022a] as done in Section 5.6.1. The "market data" is generated on a time grid from time zero to a maturity of ten years with

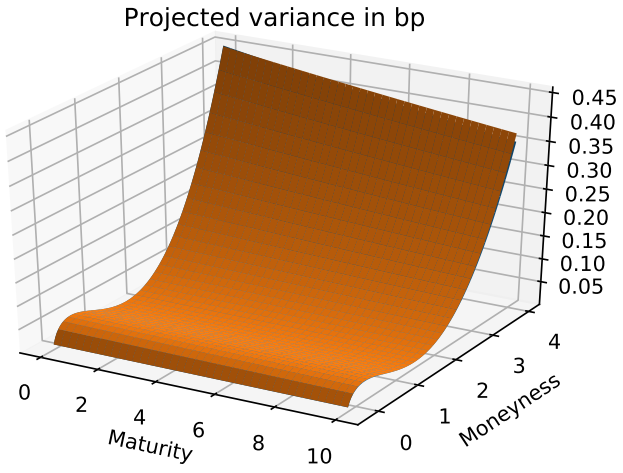
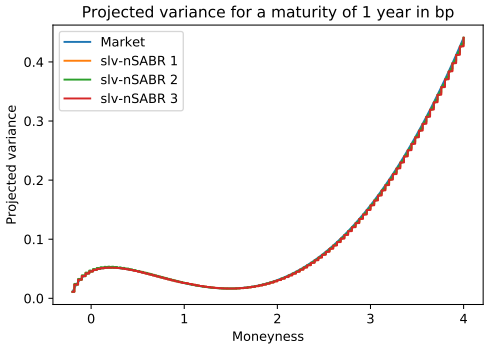
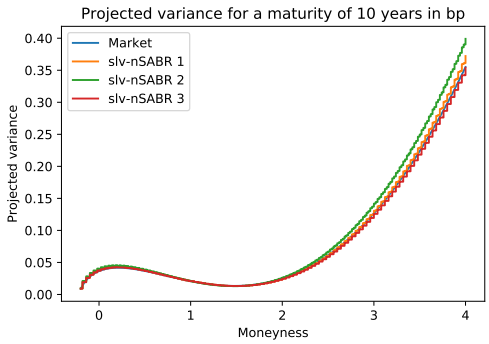


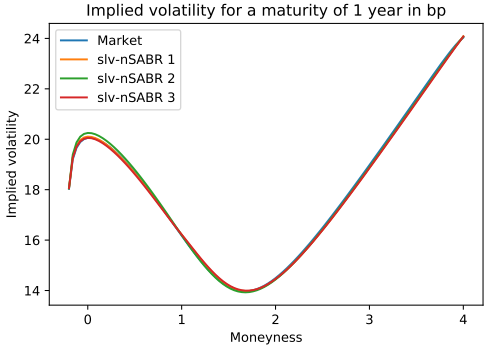
Figure 5.1: Projected variance surface on the grid points.



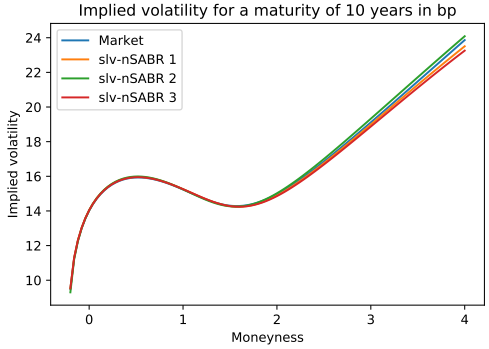
(a) Projected variance for maturity 1.



(b) Projected variance for maturity 10.



(c) Implied volatility for maturity 1.



(d) Implied volatility for maturity 10.

Figure 5.2: Projected variance and implied volatility functions for different model parametrizations.

yearly intervals. The strike grid is set in dependence on the sampled parameters and starts at a value of $-0.99d$ and spans a total of 40 grid points where the grid size is determined in such a way that the sampled forward value is the tenth grid point.

parameter
$f \in [0, 0.05]$
$d \in [0, 0.01]$
$\alpha \in [0.0001, 0.1]$
$\beta \in [0.001, 0.95]$
$\nu \in [0.01, 0.51]$
$\rho \in [-0.99, 0]$
$\gamma \in [0.05, 0.95]$

Table 5.2: Parameter ranges used for sampling.

We sample a total of 15 000 different parameter sets and consider the paths where the projected variance for the full surface stays within a reasonable range, defined as $[0, 10^{100}]$. For each of these admissible paths we calibrate the slv-nSABR model to the corresponding surface. The calibration is performed over the three parameters $\alpha_{\text{nSABR}} \in [10^{-6}, 0.1]$, $\nu_{\text{nSABR}} \in [0.01, 1]$ and $\rho_{\text{nSABR}} \in [-0.9999, 0.9999]$ with an initial parametrization provided in Table 5.3. Here in each calibration step the leverage function is calibrated as described in Section 5.3 with a fitting maturity of $T_{\text{fit}} = 1$. The minimization is performed using a relative mean squared error function of the form

$$\mathsf{mse} = \sum_{T \in \{1,5,10\}} \frac{1}{3} \Bigg(\sum_{n \geq 0} \left(\frac{\sigma^2_{\text{mkt}}(\tilde{T}, K_n) - \sigma^2_{\text{slv-nSABR}}(\tilde{T}, K_n)}{\sigma^2_{\text{slv-nSABR}}(\tilde{T}, K_n)} \right)^2 \Bigg).$$

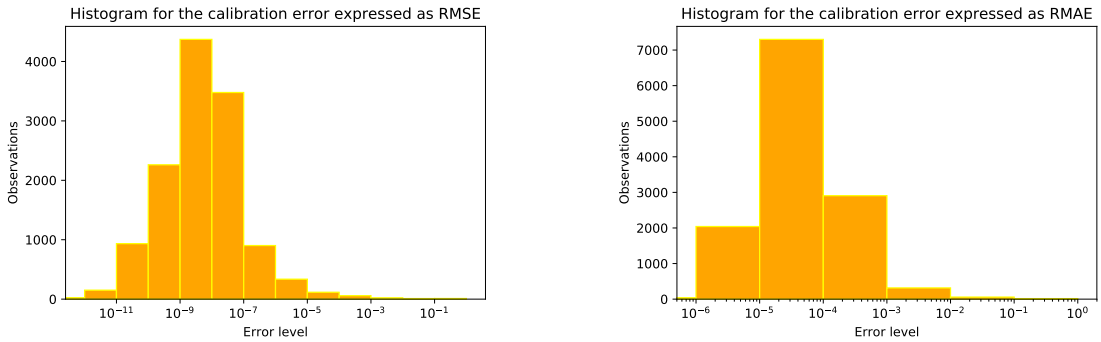
To present results, consider the following error levels $c_0 = 0$, $c_i = 10^{-i}$, $i = 1, \ldots, 12$. Using these levels, we produce a discrete distribution of the relative mean absolute and squared error (less than the levels) which are graphed in Figure 5.3. All errors larger than 10^{-1} are accumulated into a single bucket.

5.6.3 The Heston Model

We demonstrate the applicability of our approach to the slv-Heston model. Before considering the calibration of the leverage function, we provide some intuition on the accuracy of the approximation. This is demonstrated by considering the original Heston model Heston [1993] parametrized as in Table 5.4.

Parameter	slv-nSABR
f_{nSABR}	f_{mkt}
β_{nSABR}	0
γ_{nSABR}	1
α_{nSABR}	0.02
ν_{nSABR}	0.2
ρ_{nSABR}	-0.5
d_{nSABR}	d_{mkt}

Table 5.3: Parametrization for the initial slv-nSABR model.



(a) Relative mean squarred error.

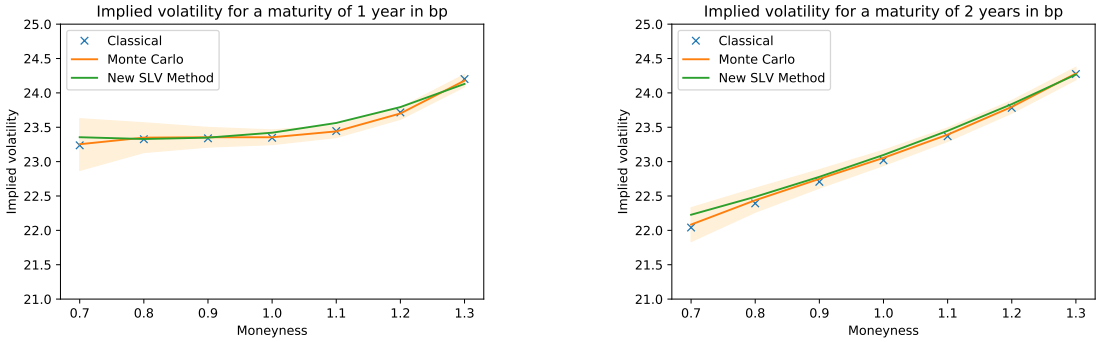
(b) Relative mean absolute error.

Figure 5.3: Histogram for the calibration error.

We compare the accuracy by computing the call prices under the Heston model using the classical semi-analytical formulas with a Black Scholes control variate, Andersen and Piterbarg [2010a], and using an slv-Heston model where the leverage functions are constant and set to one. For the slv-Heston model the time grid starts at zero and ends at two years with a step size of 1/48 years. The strike grid starts at zero and moves to a strike of 200 with a step size of ten. Using this model we compute the prices with our new approach and a PDE grid size of 1280 points. To provide a comparison we simulate the prices using a classical Euler Monte Carlo scheme with one million sample paths. Figure 5.4 depicts the resulting implied volatilities. Here the shaded area represents the three-standard-deviation area induced by the central limit theorem Durrett [2010]. As observed, our new method yields very good results and for almost all observed points lies well inside the error bounds of the reference curve.

Parameter	Heston
f	100
α	0.06
ν	0.3
ρ	-0.5
d	0
κ	1

Table 5.4: Parametrization for the first Heston example.



(a) Implied volatility for maturity 1.

(b) Implied volatility for maturity 2.

Figure 5.4: Implied volatility functions for the Heston model in comparison to Monte Carlo samples and the slv-Heston model.

5.6.4 The slv-Heston Model

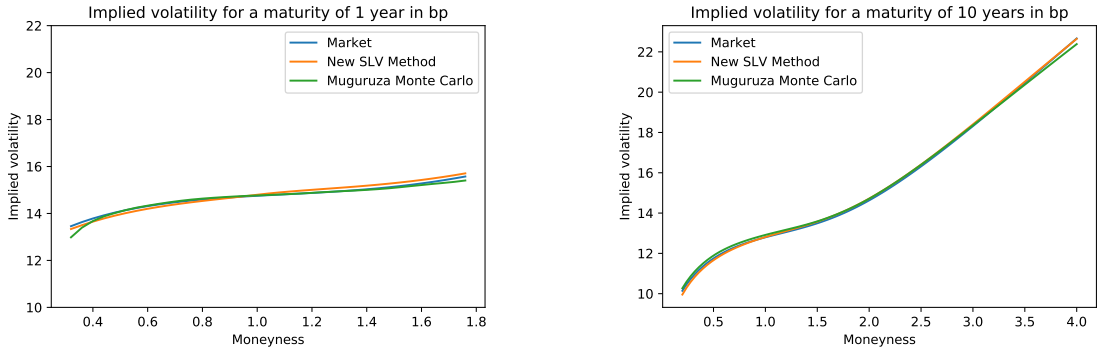
In a manner similar to Section 5.6.1, we use a dZABR model to simulate a market projected variance surface and calibrate the corresponding slv-Heston model to this surface. The parametrization of the underlying base models is provided in Table 5.5.

The "market data" is generated on a strike grid from $10bp$ to $200bp$ in steps of $2bp$. The time grid starts at zero and increases to a maturity of ten years on a quarterly basis. Using this parametrization, we calibrate the slv-Heston model to the projected variance surface using our new approach and the Monte Carlo scheme from Section 5.4.3.2. For our PDE scheme we use 640 space steps. For the Monte Carlo approach we considered a curve generated using 100 000 sample paths and time spacing of 24 steps per year. The corresponding implied volatilities are shown in Figure 5.5. Both approaches yield a very good fit to the market curves. For the smaller maturity, the sampling curve does not reproduce the observed

Parameter	ZABR	Heston
f	50 bp	50 bp
β	0.9	/
γ	0.9	/
α	$0.3f^{1-\beta}$	0.08
ν	0.3	0.1
ρ	-0.7	-0.5
d	0	0
κ	/	0.01

Table 5.5: Parametrization of the simulated market curves and the underlying base models.

market values for the deep ITM and deep OTM area. This is a common sampling problem and can be addressed by increasing the time spacing and considering additional sampling techniques such as importance sampling Glasserman [2004], Jäckel [2002]. Since our focus is, however, on demonstrating our new approach, we do not delve deeper into such sampling techniques and consider the generated curve as satisfactory to provide a benchmark for the ATM area and for larger maturities. This shows that our new approach yields a stable methodology for generating prices for small maturities and strikes far from the ATM value.



(a) Implied volatility for maturity 1.

(b) Implied volatility for maturity 10.

Figure 5.5: Implied volatility functions for the slv-Heston model in comparison to a Monte Carlo approach.

5.7 Conclusion

In this paper we have considered the setting of general stochastic local volatility models and have extended the work of Felpel et al. [2021] to incorporate this new class of models. Under the assumption of a piecewise constant and separable leverage function we have derived a fast and robust calibration scheme for the leverage function allowing the calibration and computation to be performed separately. A numerical PDE scheme was presented to allow the pricing of stochastic local volatility models in our setting. Future research will entail analyzing a relaxation of the assumptions on the leverage function with further application of these techniques for pricing derivatives.

5.8 Appendix

5.8.1 Intermediate Theorem

The goal of this section is to derive a simpler version of Theorem 5.2.1 where the time dependence in the forward dynamics is constant and set to $\sigma(\cdot) \equiv 1$. This yields a system of the form

$$\begin{cases} dF_t = C(F_t)v_t\, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(t, v_t)\, dt + \nu(t, v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\, dt. \end{cases}$$

The corresponding assumptions and the resulting theorem are provided as follows:

Assumption XV. The drift term, $\mu(\cdot, \cdot)$, is differentiable with regard to its second argument and a solution $Y(t, t_0, \alpha)$ to the following PDE exists:

$$\begin{cases} \partial_t Y(t, t_0, \alpha) = \mu(t, Y(t, t_0, \alpha)) \\ Y(t, t, \alpha) = \alpha \\ Y(t_0, t_0, \alpha) = \alpha. \end{cases}$$

Assumption XVI. The function Y is differentiable and has an inverse function $y(t_0, t, a)$ such that

$$Y(t, t_0, \alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0, t, a).$$

Assumption XVII. The functions

$$X(t,t_0,\alpha)=\partial_\alpha Y(t,t_0,\alpha),\qquad Z(t,u)=Z(t,u,t_0,\alpha)=y(u,t,Y(t,t_0,\alpha)),$$
$$z(F)=\int_f^F\frac{1}{C(u)}\,du,\qquad s(t)=S(t_0,t,\alpha)=\int_{t_0}^t Z(t,u,t_0,\alpha)^2\,du$$

and

$$\psi(t,u,Z)=\nu\big(u,Z(t,u)\big)Z(t,u)X\big(t,u,Z(t,u)\big)$$

are well defined, $X\big(t,u,Z(t,u)\big)^{-1}$ exists, and the following integral functions are defined:

$$I_1(t)=\rho\int_{t_0}^t\psi(t,u,Z)\,du,$$
$$I_2(t)=2\int_{t_0}^t\nu\big(u,Z(t,u)\big)^2X\big(t,u,Z(t,u)\big)^2\int_u^t Z(t,v)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,$$
$$I_3(t)=\rho\int_{t_0}^t\psi(t,u,Z)\int_u^t Z(t,v)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,$$
$$I_4(t)=\rho^2\int_{t_0}^t\psi(t,u,Z)\int_u^t\partial_Z\Big(\psi(t,v,Z)\Big)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,$$
$$I_5(t)=\int_{t_0}^t\nu\big(u,Z(t,u)\big)^2X\big(t,u,Z(t,u)\big)^2\,du.$$

Assumption XVIII. The function $C(\cdot)$ is differentiable at f, with derivative denoted by $C'(\cdot)$.

Theorem 5.8.1. *Given that the general stochastic volatility model* (5.8.1) *obeys Assumptions* XV–XVIII*, an effective PDE for the effective probability, $Q(\cdot,\cdot)$, of the form*

$$\begin{cases}\partial_tQ(t,F)=\partial_{FF}\big[D(t,F)Q(t,F)\big]\\ Q(t,F)\to\delta(F-f)\ \text{ as } t\to t_0^-,\end{cases}$$

can be derived, characterized by the function $D(\cdot,\cdot)$ given by

$$D(t,F)=\frac{1}{2}a(t)^2C(F)^2e^{G(t)}\big(1+2b(t)z(F)+c(t)z(F)^2\big),$$

where the coefficients are specified as

$$a(t)=Y(t,t_0,\alpha),\qquad c(t)=b(t)^2+\frac{1}{a(t)s(t)^2}I_2(t)-\frac{6b(t)}{s(t)^2}I_3(t)+\frac{2}{a(t)s(t)^2}I_4(t),$$
$$b(t)=\frac{1}{a(t)s(t)}I_1(t),\quad G(t)=-s(t)c(t)-s(t)b(t)\Gamma_0+\frac{1}{a^2}I_5(t)$$

and

$$\Gamma_0 = -C'(f).$$

Below, the proof of this intermediate theorem is presented in five stages.

5.8.1.1 Deriving the Effective Forward Equation

To prove the intermediate theorem, we perform the known singular perturbation techniques of Felpel et al. [2021], Hagan et al. [2014, 2020b, 2018a] to analyze the system

$$\begin{cases} dF_t = \varepsilon C(F_t) v_t \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \mu(t, v_t)\, dt + \varepsilon \nu(t, v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases}$$

Following Felpel et al. [2021], Hagan et al. [2014, 2020b, 2018a], we define the probability density $p(t_0, f, \alpha, t, F, A)$ that $F_t = F$ and $v_t = A$ at time t, given that $F_{t_0} = f$ and $v_{t_0} = \alpha$ at time t_0. Furthermore, we define the moments

$$Q^{(k)}(t_0, f, \alpha, t, F) = \int_0^\infty A^k p(t_0, f, \alpha, t, F, A)\, dA$$

and set

$$Q(t, F) = Q^{(0)}(t_0, f, \alpha, t, F) \tag{5.8.1}$$

to be the reduced probability density of the model given t_0, f and α.

5.8.1.2 Volatility Drift

Before analysing the corresponding PDE equations, let us first take a look at the drift term of the volatility function. For this, we consider the PDE induced by the drift term

$$\begin{cases} \partial_t Y(t, t_0, \alpha) = \mu(t, Y(t, t_0, \alpha)) \\ Y(t, t, \alpha) = \alpha \\ Y(t_0, t_0, \alpha) = \alpha. \end{cases} \tag{5.8.2}$$

By Assumptions XV and XVI we know that there exists a solution $Y(t, t_0, \alpha)$ and an inverse function $y(t_0, t, a)$, such that

$$Y(t, t_0, \alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0, t, a).$$

Considering the integrated form of the PDE (5.8.2), the derivatives of $Y(t, t_0, \alpha)$ satisfy

$$\partial_{t_0} Y(t, t_0, \alpha) = -\mu(t_0, \alpha) + \int_{t_0}^{t} \partial_Y \mu(s, Y(s, t_0, \alpha)) \partial_{t_0} Y(s, t_0, \alpha)\, ds$$

$$\text{and} \qquad \partial_\alpha Y(t, t_0, \alpha) = 1 + \int_{t_0}^{t} \partial_Y \mu(s, Y(s, t_0, \alpha)) \partial_\alpha Y(s, t_0, \alpha)\, ds.$$

This, in turn, implies that

$$\partial_{t_0} Y(t, t_0, \alpha) = -\mu(t, \alpha) \partial_\alpha Y(t, t_0, \alpha).$$

Remark 5.8.2. In integrated form the PDE corresponds to the integral equation

$$Y(t, t_0, \alpha) = \alpha + \int_{t_0}^{t} \mu(s, Y(s, t_0, \alpha))\, ds.$$

5.8.1.3 The Forward Equation

Now, we start by considering the Kolmogorov forward equation to get

$$\begin{aligned} \partial_t p(t, F, A) = &-\partial_A \Big[\mu(t, A) p(t, F, A)\Big] + \frac{1}{2}\varepsilon^2 \partial_{FF}\Big[C(F)^2 A^2 p(t, F, A)\Big] \\ &+\varepsilon^2 \rho \partial_{FA}\Big[C(F) A \nu(t, A) p(t, F, A)\Big] + \frac{1}{2}\varepsilon^2 \partial_{AA}\Big[\nu(t, A)^2 p(t, F, A)\Big], \end{aligned}$$

where we have abbreviated $p(t_0, f, \alpha, t, F, A)$ as $p(t, F, A)$. Integrating over A and considering reflecting boundary conditions to conserve the probability, as done for example in Hagan et al. [2014], we get

$$\begin{aligned} \int_0^\infty \partial_A \Big[\mu(t, A) p(t, F, A)\Big] dA &= \Big[\mu(t, A) p(t, F, A)\Big]\Big|_0^\infty = 0 \\ \int_0^\infty \partial_{FA}\Big[C(F) A \nu(t, A) p(t, F, A)\Big] dA &= \partial_F \Big[C(F) A \nu(t, A) p(t, F, A)\Big]\Big|_0^\infty = 0 \\ \int_0^\infty \partial_{AA}\Big[\nu(t, A)^2 p(t, F, A)\Big] dA &= \partial_A \Big[\nu(t, A)^2 p(t, F, A)\Big]\Big|_0^\infty = 0. \end{aligned}$$

With this, we get the forward equation

$$\begin{cases} \partial_t Q^{(0)}(t, F) = \dfrac{1}{2}\varepsilon^2 \partial_{FF}\Big[C(F)^2 Q^{(2)}(t, F)\Big] \\ Q^{(0)}(t, F) \to \delta(F - f) \text{ as } t \to t_0^+, \end{cases} \tag{5.8.3}$$

for $t > t_0$. Next we want to express $Q^{(2)}(t, F)$ in terms of $Q^{(0)}(t, F)$. This is done by considering the Kolmogorov backward equation.

5.8.1.4 The Backward Equation

We consider the Kolmogorov backward equation for $Q^{(k)}$ given by

$$\begin{cases} -\partial_{t_0} Q^{(k)} = \mu(t_0,\alpha)\partial_\alpha Q^{(k)} + \frac{1}{2}\varepsilon^2\alpha^2\Big[C(f)^2\partial_{ff}Q^{(k)} + 2\rho\frac{\nu(t_0,\alpha)}{\alpha}C(f)\partial_{f\alpha}Q^{(k)} \\ \qquad\qquad + \left(\frac{\nu(t_0,\alpha)}{\alpha}\right)^2\partial_{\alpha\alpha}Q^{(k)}\Big] \\ Q^{(k)} \to \alpha^k\delta(F-f) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $Q^{(k)}(t_0, f, \alpha, t, F)$ as $Q^{(k)}$. To cancel the drift term, we change variables from α to $a = Y(t, t_0, \alpha)$. As seen in Section 5.8.1.2 the change of variables is provided by

$$\begin{aligned} &\partial_\alpha \to \partial_\alpha Y(t,t_0,\alpha)\partial_a = X(t,t_0,y(t_0,t,a))\partial_a, \\ &\partial_{t_0} \to \partial_{t_0} - \mu(t_0,\alpha)\partial_\alpha Y(t,t_0,\alpha)\partial_a = \partial_{t_0} - \mu(t_0,y(t_0,t,a))X(t,t_0,y(t_0,t,a))\partial_a. \end{aligned}$$

Here, we use the function X as an abbreviation of $\partial_\alpha Y(t, t_0, \alpha)$. With this the drift term vanishes and we get

$$\begin{cases} -\partial_{t_0} Q^{(k)} = \frac{1}{2}\varepsilon^2 y(t_0,t,a)^2\Big[C(f)^2\partial_{ff}Q^{(k)} + 2\rho\tilde{\nu}C(f)\partial_{f\alpha}Q^{(k)} + \tilde{\nu}^2\partial_{\alpha\alpha}Q^{(k)}\Big] \\ Q^{(k)} \to a^k\delta(F-f) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $Q^{(k)}(t_0, f, a, t, F)$ as $Q^{(k)}$, and

$$\tilde{\nu} = \tilde{\nu}(t_0,t,a) = \frac{\nu(t_0,y(t_0,t,a))}{y(t_0,t,a)}X(t,t_0,y(t_0,t,a)).$$

Note that this equation is the same one as in the proof of Felpel et al. [2021, Theorem 1]. The only difference is the form of $\tilde{\nu}(t_0, t, a)$. Thus, until we reach the point where the explicit form of $\tilde{\nu}$ is needed, our reasoning is the same as before and we refer to Felpel et al. [2021] for all the remaining details until then.

5.8.1.5 New Integral Expressions

To complete the proof, the explicit form of $\tilde{\nu}$ is first needed in the representation of the integral terms. We demonstrate these steps only for the first expression I_1, the other expressions follow in an analogous manner based on the proof Felpel

et al. [2021]. For the integral expression I_1 we arrive at the representation

$$\begin{aligned} I_1(s,t,a) &= \rho \int_{\tilde{t_0}(s,t,a)}^{t} \tilde{\nu}\big(u,t,a\big) y^2(u,t,a)\, du \\ &= \rho \int_{\tilde{t_0}(s,t,a)}^{t} \nu\big(u, y(u,t,a)\big) y(u,t,a) X(t,u,y(u,t,a))\, du. \end{aligned}$$

Here the variable $\tilde{t_0}(s,t,a)$ comes from an additional change of variables which can be found in the proof of Felpel et al. [2021]. Expressing this term in the original variables, we arrive exactly at the expression provided in Assumption XVII. In a similar manner we can derive all the integral expressions of Assumption XVII and the proof can be concluded following the remaining steps of the proof in Felpel et al. [2021].

5.8.2 Proof of Theorem 1

Given the results for the intermediate Theorem, we are now ready to prove Theorem 5.2.1.

5.8.2.1 Time Change

The key step to prove Theorem 5.2.1 is a time change to allow the application of Theorem 5.8.1. For this, we start by considering the Kolmogorov forward equation to deduce that

$$\begin{aligned} \partial_t p(t,F,A) = &-\partial_A\Big[\mu(t,A)p(t,F,A)\Big] + \frac{1}{2}\varepsilon^2\sigma(t)^2\partial_{FF}\Big[C(F)^2A^2p(t,F,A)\Big] \\ &+\varepsilon^2\rho\sigma(t)\partial_{FA}\Big[C(F)A\nu(t,A)p(t,F,A)\Big] + \frac{1}{2}\varepsilon^2\partial_{AA}\Big[\nu(t,A)^2p(t,F,A)\Big], \end{aligned}$$

where again we have abbreviated the density function $p(t_0, f, \alpha, t, F, A)$ as $p(t,F,A)$. By Assumption IX we can define a time change from t to $\zeta = \zeta(t)$ defined by

$$\zeta(t) = \int_{t_0}^{t} \sigma(u)^2 du.$$

Under this time change the forward equation becomes

$$\begin{aligned} \partial_\zeta p(\zeta,F,A) = &-\partial_A\Big[\tilde{\mu}(\zeta,A)p(\zeta,F,A)\Big] + \frac{1}{2}\varepsilon^2\partial_{FF}\Big[C(F)^2A^2p(\zeta,F,A)\Big] \\ &+\varepsilon^2\rho\partial_{FA}\Big[C(F)A\tilde{\nu}(\zeta,A)p(\zeta,F,A)\Big] + \frac{1}{2}\varepsilon^2\partial_{AA}\Big[\tilde{\nu}(\zeta,A)^2p(\zeta,F,A)\Big], \end{aligned}$$

where

$$\tilde{\mu}(\zeta, A) = \frac{\mu(\tilde{t}(\zeta), A)}{\sigma(\tilde{t}(\zeta))^2}, \quad \tilde{\nu}(\zeta, A) = \frac{\nu(\tilde{t}(\zeta), A)}{\sigma(\tilde{t}(\zeta))},$$

and $p(\zeta, F, A)$ denotes the abbreviated density function $p(\zeta_0, f, \alpha, \zeta, F, A)$ with $\zeta_0 = \zeta(t_0)$. This essentially allows us to deduce an equivalent system of stochastic differential equations as seem under the new time space induced by $\zeta(\cdot)$ given by

$$\begin{cases} dF_\zeta = C(F_\zeta) v_\zeta \, dW_\zeta^{(1)}, & F_{\zeta_0} = f, \\ dv_\zeta = \tilde{\mu}(\zeta, v_\zeta) \, d\zeta + \tilde{\nu}(\zeta, v_\zeta) \, dW_\zeta^{(2)}, & v_{\zeta_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_\zeta = \rho \, d\zeta. \end{cases} \tag{5.8.4}$$

5.8.2.2 Theorem under the time space ζ

Given the system (5.8.4) we can apply the intermediate theorem version presented in Section 5.8.1. To do this we must impose the following assumptions on the system (5.8.4):

Assumption XIX. The drift term, $\tilde{\mu}(\cdot, \cdot)$, is differentiable with regard to its second argument and a solution $\tilde{Y}(\zeta, \zeta_0, \alpha)$ to the following PDE exists:

$$\begin{cases} \partial_\zeta \tilde{Y}(\zeta, \zeta_0, \alpha) = \tilde{\mu}(\zeta, \tilde{Y}(\zeta, \zeta_0, \alpha)) \\ \tilde{Y}(\zeta, \zeta, \alpha) = \alpha \\ \tilde{Y}(\zeta_0, \zeta_0, \alpha) = \alpha. \end{cases}$$

Assumption XX. The function $\tilde{Y}$ is differentiable and has an inverse function $\tilde{y}(\zeta_0, \zeta, a)$ such that

$$\tilde{Y}(\zeta, \zeta_0, \alpha) = a \qquad \Leftrightarrow \qquad \alpha = \tilde{y}(\zeta_0, \zeta, a).$$

Assumption XXI. The functions

$$\tilde{X}(\zeta, \zeta_0, \alpha) = \partial_\alpha \tilde{Y}(\zeta, \zeta_0, \alpha), \qquad \tilde{Z}(\zeta, u) = \tilde{Z}(\zeta, u, \zeta_0, \alpha) = \tilde{y}(u, \zeta, \tilde{Y}(\zeta, \zeta_0, \alpha)),$$
$$z(F) = \int_f^F \frac{1}{C(u)} \, du, \qquad \tilde{s}(\zeta) = \tilde{S}(\zeta_0, \zeta, \alpha) = \int_{\zeta_0}^{\zeta} \tilde{Z}(\zeta, u, \zeta_0, \alpha)^2 \, du$$

and

$$\tilde{\psi}(\zeta, u, \tilde{Z}) = \tilde{\nu}\big(u, \tilde{Z}(\zeta, u)\big) \tilde{Z}(\zeta, u) \tilde{X}\big(\zeta, u, \tilde{Z}(\zeta, u)\big)$$

are well defined, $\tilde{X}\big(\zeta, u, \tilde{Z}(\zeta,u)\big)^{-1}$ exists, and the following integral functions are defined:

$$\tilde{I}_1(\zeta) = \rho \int_{\zeta_0}^{\zeta} \tilde{\psi}(\zeta, u, \tilde{Z})\, du,$$

$$\tilde{I}_2(\zeta) = 2\int_{\zeta_0}^{\zeta} \tilde{\nu}\big(u, \tilde{Z}(\zeta,u)\big)^2 \tilde{X}\big(\zeta, u, \tilde{Z}(\zeta,u)\big)^2 \int_u^{\zeta} \tilde{Z}(\zeta, v)\tilde{X}\big(\zeta, v, \tilde{Z}(\zeta,v)\big)^{-1}\, dv\, du,$$

$$\tilde{I}_3(\zeta) = \rho \int_{\zeta_0}^{\zeta} \tilde{\psi}(\zeta, u, \tilde{Z}) \int_u^{\zeta} \tilde{Z}(\zeta, v)\tilde{X}\big(\zeta, v, \tilde{Z}(\zeta,v)\big)^{-1}\, dv\, du,$$

$$\tilde{I}_4(\zeta) = \rho^2 \int_{\zeta_0}^{\zeta} \tilde{\psi}(\zeta, u, \tilde{Z}) \int_u^{\zeta} \partial_{\tilde{Z}}\Big(\tilde{\psi}(\zeta, v, \tilde{Z})\Big)\tilde{X}\big(\zeta, v, \tilde{Z}(\zeta,v)\big)^{-1}\, dv\, du,$$

$$\tilde{I}_5(\zeta) = \int_{\zeta_0}^{\zeta} \tilde{\nu}\big(u, \tilde{Z}(\zeta,u)\big)^2 \tilde{X}\big(\zeta, u, \tilde{Z}(\zeta,u)\big)^2\, du.$$

Assumption XXII. The function $C(\cdot)$ is differentiable at f, with derivative denoted by $C'(\cdot)$.

Given these assumptions, we can finally apply Theorem 5.8.1 and deduce an effective PDE in the time space ζ given by

$$\begin{cases} \partial_\zeta \tilde{Q}(\zeta, F) = \partial_{FF}\big[\tilde{D}(\zeta, F)\tilde{Q}(\zeta, F)\big] \\ \tilde{Q}(\zeta, F) \to \delta(F - f) \text{ as } \zeta \to \zeta_0^-, \end{cases}$$

characterized by the function

$$\tilde{D}(\zeta, F) = \frac{1}{2}\tilde{a}(\zeta)^2 C(F)^2 e^{\tilde{G}(\zeta)}\big(1 + 2\tilde{b}(\zeta)z(F) + \tilde{c}(\zeta)z(F)^2\big), \tag{5.8.5}$$

where the coefficients are specified as

$$\tilde{a}(\zeta) = \tilde{Y}(\zeta, \zeta_0, \alpha), \qquad \tilde{c}(\zeta) = \tilde{b}(\zeta)^2 + \frac{1}{\tilde{a}(\zeta)\tilde{s}(\zeta)^2}\tilde{I}_2(\zeta) - \frac{6\tilde{b}(\zeta)}{\tilde{s}(\zeta)^2}\tilde{I}_3(\zeta) + \frac{2}{\tilde{a}(\zeta)\tilde{s}(\zeta)^2}\tilde{I}_4(\zeta),$$

$$\tilde{b}(\zeta) = \frac{1}{\tilde{a}(\zeta)\tilde{s}(\zeta)}\tilde{I}_1(\zeta), \quad \tilde{G}(\zeta) = -\tilde{s}(\zeta)\tilde{c}(\zeta) - \tilde{s}(\zeta)\tilde{b}(\zeta)\Gamma_0 + \frac{1}{\tilde{a}(\zeta)^2}\tilde{I}_5(\zeta)$$

and

$$\Gamma_0 = -C'(f).$$

5.8.2.3 Reverse Time Change

Given the results expressed in the time space ζ it remains to express these results in the original variables and show that the Assumptions XIX–XXII are induced by

the Assumptions IX–XIII. Now, applying the reverse time change from $\zeta(t)$ back to time t, the effective PDE becomes

$$\begin{cases} \partial_t Q(t,F) = \partial_{FF}\big[\sigma(t)^2 \tilde{D}(\zeta(t),F) Q(t,F)\big] \\ Q(t,F) \to \delta(F-f) \text{ as } t \to t_0^-, \end{cases}$$

where we have used the abbreviation $Q(t,F) = \tilde{Q}(\zeta(t),F)$. This expression yields the characterization

$$\begin{aligned} D(t,F) &= \sigma(t)^2 \tilde{D}(\zeta(t),F) \\ &= \frac{1}{2}\sigma(t)^2 C(F)^2 \tilde{a}(\zeta(t))^2 e^{\tilde{G}(\zeta(t))}\big(1 + 2\tilde{b}(\zeta(t))z(F) + \tilde{c}(\zeta(t))z(F)^2\big). \end{aligned}$$

Similar to the notation of $Q(\cdot,\cdot)$, we define the coefficients

$$a(t) = \tilde{a}\big(\zeta(t)\big),\ b(t) = \tilde{b}\big(\zeta(t)\big),\ c(t) = \tilde{c}\big(\zeta(t)\big),\ G(t) = \tilde{G}\big(\zeta(t)\big),$$

to get the desired expression

$$D(t,F) = \frac{1}{2}\sigma(t)^2 C(F)^2 a(t)^2 e^{G(t)}\big(1 + 2b(t)z(F) + c(t)z(F)^2\big).$$

5.8.2.3.1 Coefficient a

Considering the definition of $a(t)$ as

$$a(t) = \tilde{a}\big(\zeta(t)\big) = \tilde{Y}(\zeta(t),\zeta(t_0),\alpha),$$

we use the last expression to define a new function

$$\hat{Y}(t,t_0,\alpha) = \tilde{Y}(\zeta(t),\zeta(t_0),\alpha).$$

Knowing that Assumption XIX is satisfied, we can deduce that $\hat{Y}$ in turn has to be a solution to

$$\begin{cases} \partial_t \hat{Y}(t,t_0,\alpha) = \sigma(\tilde{t}(\zeta))^2 \partial_\zeta \tilde{Y}(\zeta,\zeta(t_0),\alpha) = \mu(t,\tilde{Y}(\zeta(t),\zeta(t_0),\alpha)) = \mu(t,\hat{Y}(t,t_0,\alpha)) \\ \hat{Y}(t,t,\alpha) = \alpha \\ \hat{Y}(t_0,t_0,\alpha) = \alpha. \end{cases}$$

By Assumption X this in turn implies that $\hat{Y}$ is exactly the desired function Y and we have

$$Y(t,t_0,\alpha) = \hat{Y}(t,t_0,\alpha) = \tilde{Y}(\zeta(t),\zeta(t_0),\alpha).$$

This leads to

$$a(t) = Y(t, t_0, \alpha)$$

and shows that

$$\begin{aligned} \text{Assumption X} &\Rightarrow \text{Assumption XIX} \\ \text{Assumption XI} &\Rightarrow \text{Assumption XX}. \end{aligned}$$

5.8.2.3.2 Remaining coefficients

Similar to the case of $a(t)$ it remains to show that the remaining coefficients can be expressed exactly by the formulas of Assumption XII. For this, we demonstrate the necessary steps for the coefficient $s(t)$. In this case we have

$$\begin{aligned} s(t) = \tilde{s}(\zeta(t)) &= \int_{\zeta(t_0)}^{\zeta(t)} \tilde{Z}(\zeta(t), u, \zeta(t_0), \alpha)^2 \, du \\ &= \int_{\zeta(t_0)}^{\zeta(t)} \tilde{y}(u, \zeta(t), \tilde{Y}(\zeta(t), \zeta(t_0), \alpha))^2 \, du \\ &= \int_{\zeta(t_0)}^{\zeta(t)} \tilde{y}(u, \zeta(t), Y(t, t_0, \alpha))^2 \, du. \end{aligned}$$

Proceeding with a change of variables from u to $x = \tilde{t}(u)$ we get

$$\begin{aligned} s(t) &= \int_{t_0}^{t} \sigma(x)^2 \tilde{y}(\zeta(x), \zeta(t), Y(t, t_0, \alpha))^2 \, dx \\ &= \int_{t_0}^{t} \sigma(x)^2 y(x, t, Y(t, t_0, \alpha))^2 \, dx \\ &= \int_{t_0}^{t} \sigma(x)^2 Z(t, x)^2 \, dx, \end{aligned}$$

where we have used the inverse function

$$y(t_0, t, a) = \tilde{y}(\zeta(t_0), \zeta(t), a).$$

Since all the transformations are done while keeping the equality, the expression $\tilde{s}(\zeta(t))$ is well defined by Assumption XII. Similar transformations are made to evaluate all the integral terms I_1 to I_5, which allows us to compute the remaining coefficients. This provides us with exactly the terms defined in Assumption XII and we can deduce that

$$\text{Assumption XII} \Rightarrow \text{Assumption XXI}.$$

This completes the proof.

5.9 Extended Research

This section is not part of the original publication Felpel et al. [2022b] and aims to provide additional insight by outlining model properties in more detail, providing further numerical analysis and explicitly presenting the derivation of the model coefficients.

5.9.1 User Guide for Model Construction

We start by providing a very helpful guide for the derivation of model coefficients induced by Theorem 5.2.1. This guide aims to outline all necessary steps to evaluate the induced model coefficients. To provide a better visualization we take the slv-dZABR model as an example to present the application of the individual steps. This results in the coefficients given in Section 5.5. For completeness we recall that the slv-dZABR model is characterized as

$$\begin{cases} dF_t = \sigma(t)L(F_t)(F_t+d)^{\beta} v_t \, dW_t^{(1)}, & F_{t_0} = f, \\ dv_t = \nu v_t^{\gamma} \, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases}$$

To evaluate the model coefficients induced by Theorem 5.2.1, the following steps are applied with an application towards the slv-ZABR presented in the examples.

1. **Solve for Y:** Solve the PDE for the function $Y(t, t_0, \alpha)$.
   Example: The PDE becomes trivial and we can conclude $Y(t, t_0, \alpha) \equiv \alpha$.
2. **Compute X:** Compute the derivative of $Y(t, t_0, \alpha)$ w.r.t. the variable α.
   Example: We compute $X(t, t_0, \alpha) \equiv 1$.
3. **Compute Z:** Compute the inverse function of $Y(t, t_0, \alpha)$.
   Example: We compute $Z(t, u) \equiv \alpha$.
4. **Specify functions:** Specify all relevant model functions necessary for the integral equations.
   Example: We define

$$\nu(t,x) = \nu x^{\gamma}, \quad \psi(t,u,x) = \nu x^{\gamma+1}, \quad \theta_i(x,t) = \int_x^t \sigma(u)^i \, du$$

$$\text{and} \quad \Omega_{i,j}(t_0,t) = \int_{t_0}^t \sigma(x)^i \theta_j(x,t) \, dx.$$

5. **Evaluate s:** Evaluate the integral function $s(t)$.
   Example: We evaluate
$$s(t) = \int_{t_0}^{t} \sigma(u)^2 Z(t,u)^2 \, du = \alpha^2 \int_{t_0}^{t} \sigma(u)^2 \, du = \alpha^2 \theta_2(t_0, t).$$

6. **Evaluate $\mathbf{I}_1$:** Evaluate the integral function $I_1(t)$.
   Example: We evaluate
$$I_1(t) = \rho \int_{t_0}^{t} \sigma(x)\psi(t,x,Z) \, dx = \rho\nu\alpha^{\gamma+1} \int_{t_0}^{t} \sigma(x) \, dx = \rho\nu\alpha^{\gamma+1}\theta_1(t_0, t).$$

7. **Evaluate $\mathbf{I}_2$:** Evaluate the integral function $I_2(t)$.
   Example: We evaluate
$$\begin{aligned} I_2(t) &= 2 \int_{t_0}^{t} \nu\big(x, Z(t,x)\big)^2 X\big(t, x, Z(t,x)\big)^2 \\ &\qquad \times \int_{x}^{t} \sigma(w)^2 Z(t,w) X\big(t, w, Z(t,w)\big)^{-1} \, dw \, dx \\ &= 2 \int_{t_0}^{t} \nu^2 \alpha^{2\gamma} \int_{x}^{t} \sigma(w)^2 \alpha \, dw \, dx = 2\nu^2\alpha^{2\gamma+1} \int_{t_0}^{t} \theta_2(x,t) \, dx \\ &= 2\nu^2\alpha^{2\gamma+1}\Omega_{0,2}(t_0, t). \end{aligned}$$

8. **Evaluate $\mathbf{I}_3$:** Evaluate the integral function $I_3(t)$.
   Example: We evaluate
$$\begin{aligned} I_3(t) &= \rho \int_{t_0}^{t} \sigma(x)\psi(t,x,Z) \int_{x}^{t} \sigma(w)^2 Z(t,w) X\big(t, w, Z(t,w)\big)^{-1} \, dw \, dx \\ &= \rho \int_{t_0}^{t} \sigma(x)\nu\alpha^{\gamma+1} \int_{x}^{t} \sigma(w)^2 \alpha \, dw \, dx = \rho\nu\alpha^{\gamma+2} \int_{t_0}^{t} \sigma(x)\theta_2(x,t) \, dx \\ &= \rho\nu\alpha^{\gamma+2}\Omega_{1,2}(t_0, t). \end{aligned}$$

9. **Evaluate $\mathbf{I}_4$:** Evaluate the integral function $I_4(t)$.
   Example: We evaluate
$$\begin{aligned} I_4(t) &= \rho^2 \int_{t_0}^{t} \sigma(x)\psi(t,x,Z) \int_{x}^{t} \sigma(w)\partial_Z\Big(\psi(t,w,Z)\Big) X\big(t, w, Z(t,w)\big)^{-1} \, dw \, dx \\ &= \rho^2 \int_{t_0}^{t} \sigma(x)\nu\alpha^{\gamma+1} \int_{x}^{t} \sigma(w)\partial_Z\Big(\nu Z^{\gamma+1}\Big) \, dw \, dx \\ &= \rho^2\nu^2\alpha^{2\gamma+1}(\gamma+1) \int_{t_0}^{t} \sigma(x)\theta_1(x,t) \, dx \\ &= (\gamma+1)\rho^2\nu^2\alpha^{2\gamma+1}\Omega_{1,1}(t_0, t). \end{aligned}$$

10. **Evaluate $\mathbf{I}_5$:** Evaluate the integral function $I_5(t)$.
Example: We evaluate

$$\begin{aligned} I_5(t) &= \int_{t_0}^{t} \nu\big(x, Z(t,x)\big)^2 X\big(t, x, Z(t,x)\big)^2 \, dx \\ &= \int_{t_0}^{t} \nu^2 \alpha^{2\gamma} \, dx = \nu^2 \alpha^{2\gamma} (t - t_0) \\ &= \nu^2 \alpha^{2\gamma} \theta_0(t_0, t). \end{aligned}$$

11. **Compute the coefficient Gamma:** Compute the derivative of $C(\cdot)$ evaluated at the point f.
Example: We compute

$$\Gamma_0 = -\beta(f + d)^{\beta - 1}.$$

12. **Compute the coefficient a:** Compute the coefficient using the previously computed expressions.
Example: We compute

$$a(t) \equiv \alpha.$$

13. **Compute the coefficient b:** Compute the coefficient using the previously computed expressions.
Example: We compute

$$b(t) = \frac{1}{a(t)s(t)} I_1(t) = \frac{1}{\alpha \alpha^2 \theta_2(t_0, t)} \rho \nu \alpha^{\gamma+1} \theta_1(t_0, t) = \rho \nu \alpha^{\gamma - 2} \frac{\theta_1(t_0, t)}{\theta_2(t_0, t)}.$$

14. **Compute the coefficient c:** Compute the coefficient using the previously computed expressions.
Example: We compute

$$\begin{aligned} c(t) &= b(t)^2 + \frac{1}{a(t)s(t)^2} I_2(t) - \frac{6b(t)}{s(t)^2} I_3(t) + \frac{2}{a(t)s(t)^2} I_4(t) \\ &= \rho^2 \nu^2 \alpha^{2(\gamma-2)} \frac{\theta_1(t_0, t)^2}{\theta_2(t_0, t)^2} + 2\nu^2 \alpha^{2(\gamma-2)} \frac{\Omega_{0,2}(t_0, t)}{\theta_2(t_0, t)^2} \\ &\quad - 6\rho^2 \nu^2 \alpha^{2(\gamma-2)} \frac{\theta_1(t_0, t) \Omega_{1,2}(t_0, t)}{\theta_2(t_0, t)^3} + 2(\gamma + 1) \rho^2 \nu^2 \alpha^{2(\gamma-2)} \frac{\Omega_{1,1}(t_0, t)}{\theta_2(t_0, t)^2} \\ &= 2\nu^2 \alpha^{2(\gamma-2)} \frac{\Omega_{0,2}(t_0, t)}{\theta_2(t_0, t)^2} \\ &\quad + \rho^2 \nu^2 \alpha^{2(\gamma-2)} \left[\frac{\theta_1(t_0, t)^2}{\theta_2(t_0, t)^2} - 6 \frac{\theta_1(t_0, t) \Omega_{1,2}(t_0, t)}{\theta_2(t_0, t)^3} + 2(\gamma + 1) \frac{\Omega_{1,1}(t_0, t)}{\theta_2(t_0, t)^2} \right]. \end{aligned}$$

15. **Compute the coefficient G:** Compute the coefficient using the previously computed expressions.
Example: We compute

$$\begin{aligned} G(t) &= -s(t)c(t) - s(t)b(t)\Gamma_0 + \frac{1}{a(t)^2} I_5(t) \\ &= -\alpha^2\theta_2(t_0,t)c(t) + \rho\nu\alpha^\gamma\beta(f+d)^{\beta-1}\theta_1(t_0,t) + \nu^2\alpha^{2(\gamma-1)}\theta_0(t_0,t). \end{aligned}$$

Following this guide, we have demonstrated how the coefficients of Section 5.5 are derived. A similar computation is performed to derive the coefficients of the slv-Heston model.

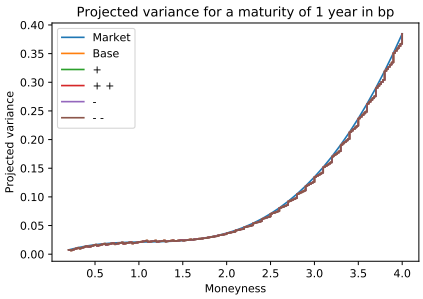
5.9.2 Parameter Dependence

Considering the numerical examples in Section 5.6, we put the focus therein on the calibration algorithm. In this section we provide additional insight on the general model behaviour. In particular, we focus on the dependencies of the model towards the underlying parameters and demonstrate the impact of each individual parameter. We work with the slv-Heston model as done in Section 5.6.4 and again use the parametrization provided in Table 5.5 to generate a market curve and establish a first reference curve for the slv-Heston model. Starting from this reference parametrization, we modify the underlying parameters one by one and compare the resulting projected volatility curves. Here the parameters are modified in accordance to Table 5.6. In each case we apply two shocks in the upward as well as the downward direction. The resulting projected volatility functions, when calibrating the slv-Heston model to the market curve as described in Section 5.6.4, are presented in Figure 5.6.

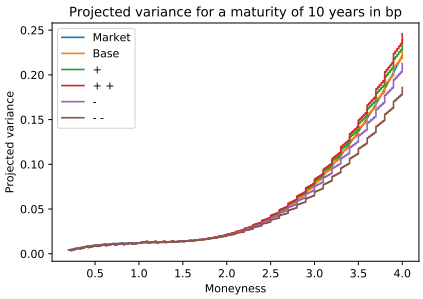
Parameter	Base	$+$	$++$	$-$	$--$
α	0.08	0.1	0.12	0.06	0.04
ν	0.1	0.2	0.4	0.05	0.025
ρ	-0.5	0	0.5	-0.7	-0.9
κ	0.01	0.1	0.5	0.005	0.001

Table 5.6: Parameters of the slv-Heston model resulting from a shock applied to the individual parameter.

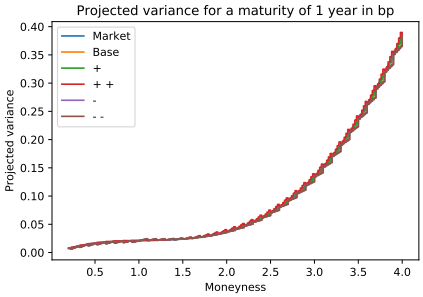
Considering, for the moment, a maturity of 1 year, the calibration algorithm becomes clearly visible. Along the specified calibration points, all computed curves



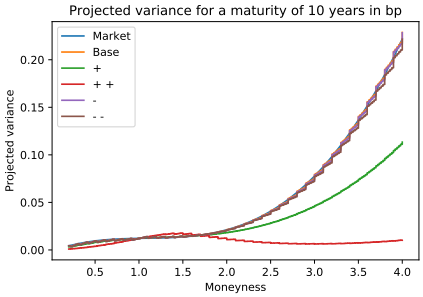
(a) Year 1 when α changes.



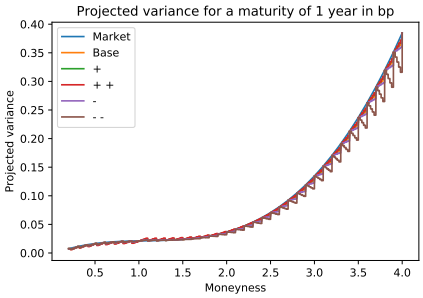
(b) Year 10 when α changes.



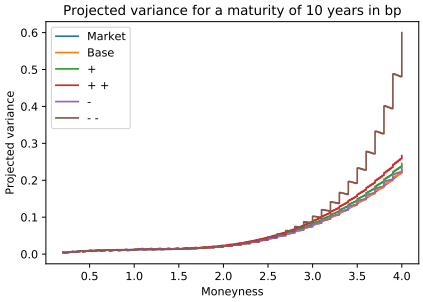
(c) Year 1 when ν changes.



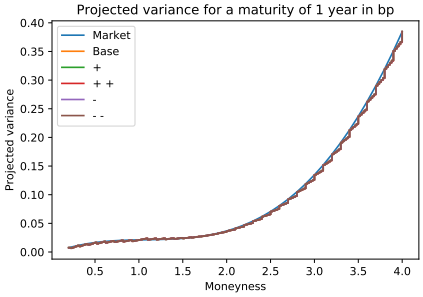
(d) Year 10 when ν changes.



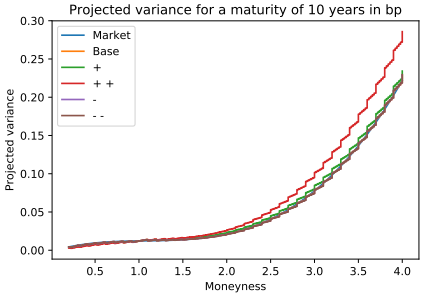
(e) Year 1 when ρ changes.



(f) Year 10 when ρ changes.



(g) Year 1 when κ changes.



(h) Year 10 when κ changes.

Figure 5.6: Projected volatility function with changing parameters.

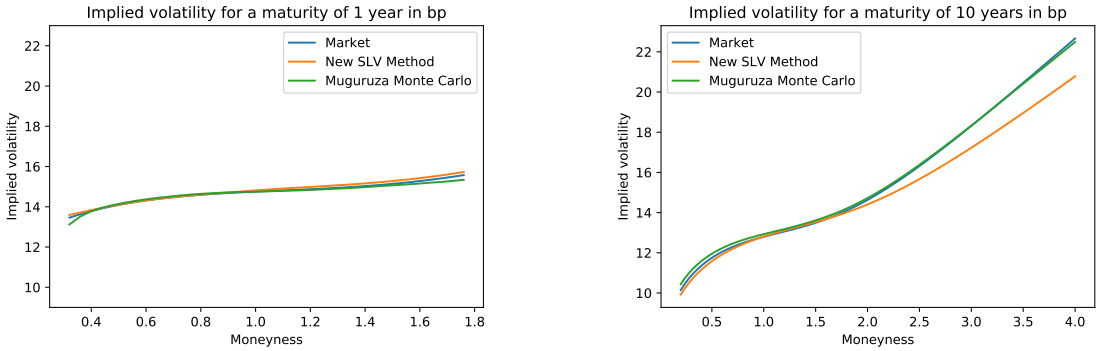
are indistinguishable due to the calibrated leverage function. Hence, the impact due to the changing parameters can only be observed on the points in between these calibration points. For the parameters α and κ the differences are only minimal. Changing the parameters ν and ρ, we can already observe some changes in the projected volatility curve. In particular, considering the case of $\rho = -0.9$ we can observe that the projected volatility curve induced by the underlying base model exhibits a different behaviour than our market curve. This in turn induces a divergence of the curves for larger values of moneyness. The observed differences are in some sense controlled by the grid size of the calibration grid, do however also show the importance, to tailor the underlying base model to the observed market behaviour.

These observations are only enhanced when considering the larger maturity of 10 years. Considering the larger values of moneyness starting from around 2, we can observe an impact of all modified parameters on the projected volatility curve. For the parameters α and κ, we observe a moderate fluctuation around the market behaviour. In the case of the parameter ρ where we already observed a divergence in year 1, this behaviour is further increased and the difference between the curves is vastly increasing depending on the moneyness. For the parameter ν we can observe a large sensitivity towards changes. Here it also possible to change the shape of the curve, suggesting ν to be one of the main handles in controlling the model behaviour.

These insights can be translated towards the implied volatility curve. In our algorithm, we use the projected volatility function in a PDE algorithm. For this algorithm a space grid is derived and the projected volatility function has to be evaluated along these grid points. These grid points do not have to correspond to the calibration grid points and may be placed in between those or far above the largest calibration point. This results in inter- and extrapolation which is driven by the underlying base model and follows the behaviour shown in Figure 5.6. This suggests two efficient handles to control the behaviour of the projected volatility function and in turn also the implied volatility function. The first possibility is to use the underlying base parameters to control and steer the behaviour. As demonstrated, this approach provides much more flexibility for larger maturities compared to fitting maturity placed at 1 year. The second option, is to control the differences through the calibration grid. As presented in Figure 5.6 this enables to greatly reduce the differences for the fitting maturity, is however not the optimal solution to control the impact on larger maturities.

To demonstrate how we can apply this knowledge, let us reconsider the example in Section 5.6.4. We assume that for a maturity of 10 years, the implied volatility curve is not satisfactory enough, e.g. induced by a questionable data quality of the

input data, and we would rather like to observe a curve which is not as steep as the current one. Therefore, we proceed and tune the underlying base parameters to our new requirements. Taking Figure 5.6 into account, we see that an increase of the parameter ν results in much smaller projected volatilities, translating in turn into smaller implied volatilities. Hence, we increase the parameter from a value of 0.1 to 0.2. The results are presented in Figure 5.7 and indeed show the desired improvements. For a maturity of 10 years and a moneyness greater then 2, the steepness is reduced while in all other areas of the curve, the quality of the approximation is still comparable to the original results deduced using the Monte Carlo scheme from Section 5.4.3.2. This demonstrates the control over the model behaviour through its parameters and provides an intuitive way to adjust the model to the desired requirements.



(a) Implied volatility for maturity 1.

(b) Implied volatility for maturity 10.

Figure 5.7: Implied volatility functions for the slv-Heston model in comparison to a Monte Carlo approach.

6 Conclusion and Future Research

Providing a suitable foundation for practical application, our techniques also exhibit enough possibilities for further improvements, investigations and future research. The direction in which further investigations may progress are manifold and before concluding this thesis we desire to outline a few possibilities.

6.1 Future Research

First of all, further research can be done to increase the class of viable models to which our techniques are applicable. Considering for example SLV models as done in Section 5.3, it can be analysed if it is possible to relax some of the assumptions imposed on the model. In particular, it would be quite useful to be able to avoid Assumption XIV since this would further increase the accuracy when modelling time dependencies. From a mathematical perspective this is very interesting since the full space and time dependence of the leverage function introduces additional terms into the perturbation analysis. Hence, it would be interesting to see if it is possible to derive results similar to that of Theorem 5.2.1 and if the leading order analysis is still applicable as before using a heat equation or if a different equation has to be found to allow the analysis. In turn, this could also affect the structure of the final approximation formula which would further yield to a generalization in the EMP algorithms of Chapter 4.

Second, the class of applicability can also be increased by adding additional components. For example, the inclusion of a drift term into the dynamics of the forward process can be considered. While we have shown that it is possible to price e.g. the spreads of two rates, where at least one rate has a drift term in a common measure, it would still be advantageous to be able to directly approximate a model formulated with a drift term in the forward process. In particular, when considering stocks instead of interest rates this provides additional flexibility in a practical

application. As before, the modification would induce additional terms into the perturbation analysis which have to be analysed carefully.

Alternatively, it can also be analysed if similar methods may be applicable to other types of models. Instead of considering stochastic volatility models, jump-diffusion models or even rough volatility models, see Section 2.3.5, could be analysed. This could open a new perspective to approach these type of models.

While these suggestions all aim to increase the scope of applicability, further investigations may also progress to deduce additional analysis on the presented methodology. This could include areas such as approximation quality or possible model limitations. Considering e.g. the results for the ZABR model in Section 3.2.3.1, our technique induced a parameter bound to enforce reasonable results. Here it would be interesting to see if such a bound is also reasonable when considering the model from a purely theoretical perspective and if it is possible to enforce general conditions to guarantee satisfying numerical results. Such analysis could also be extended with the goal to assess the approximation quality in terms of error bounds. Using the EMP-techniques of Chapter 4 as an example, we derived the approximation order for the singular perturbation approximation. The matching algorithms to project one model onto another do, however, not have such an explicit result. Here it would be interesting to see if it is possible to derive suitable bounds. These could have a theoretical or numerical nature and will possibly allow to deduce explicit conditions in which cases the techniques perform well and where special care has to be taken. In turn, this could also yield to possible improvements in the matching algorithms.

Finally, as a last suggestion for future research, we consider the numerical implementation of our presented techniques. While we have mainly used a one-dimensional finite difference scheme in accordance to Section 2.5.3, it can be analysed if the performance may be increased using different numerical implementations. This could yield to an increased stability when considering full time dependent coefficients or provide a measure of accuracy when comparing the results to a direct two-dimensional PDE solution using ADI methodologies. Moreover, new numerical techniques could be derived. Considering e.g. the proof of Theorem 3.2.3, a variety of transformations are used to reformulate the PDE in a way that is in some sense easier to analyse. A similar approach could also be used to e.g. transform the original stochastic volatility model into a formulation where the correlation between the two components vanishes. This would yield an orthogonal projection scheme and could allow a direct application of classical ADI schemes without the need to incorporate adjustments due to the correlation.

6.2 Conclusion

In summary, we have established a variety of techniques applicable to a large class of generalized stochastic volatility models. This includes classical stochastic volatility models as well as certain types of stochastic local volatility models. Our methods use singular perturbation techniques to analyse the model induced Kolmogorov equations for the probability density. Using a variety of transformation functions, we can perform a leading order analysis to accurately approximate these equations to a given order. This allows an accurate approximation of the generalized stochastic volatility model using one-dimensional differential equations which can correspondingly be seen as one-dimensional local volatility models. This approach provides a unified computational framework and yields to a multitude of explicit computation formulas. In particular, we derived explicit pricing formulas for vanilla options in terms on implied volatilities, demonstrated the pricing of basket options such as CMS spreads and analysed the calibration procedure for stochastic volatility as well as stochastic local volatility models. Explicit applications in form of numerical studies are provided for all derived techniques and show that these results may already be used in a practical application. This provides a solid foundation for future extensions and we are looking forward to observe which extension may arise and how our methods are applied in practice.

Bibliography

H. Albrecher, P. Mayer, W. Schoutens, and J. Tistaert. The little Heston trap. *Wilmott*, pages 83–92, 2006.

C. Alexander and L. M. Nogueira. Stochastic local volatility. *Available at SSRN*, 2008. doi: 10.2139/ssrn.1107685.

E. Alòs, O. Mazet, and D. Nualart. Stochastic calculus with respect to fractional Brownian motion with Hurst parameter lesser than $\frac{1}{2}$. *Stochastic Process. Appl.*, 86(1):121–139, 2000. ISSN 0304-4149. doi: 10.1016/S0304-4149(99)00089-7.

L. B. G. Andersen and V. Piterbarg. *Interest Rate Modeling – Volume I: Foundations and Vanilla Models*. Atlantic Financial Press, London New York, 2010a.

L. B. G. Andersen and V. Piterbarg. *Interest Rate Modeling – Volume II: Term Structure Models*. Atlantic Financial Press, London New York, 2010b.

L. B. G. Andersen and V. Piterbarg. *Interest Rate Modeling – Volume III: Products and Risk Management*. Atlantic Financial Press, London New York, 2010c.

J. Andreasen and B. Huge. Random grids. *Risk*, 24(7):62–67, 2011a.

J. Andreasen and B. Huge. ZABR–Expansions for the masses. *Available at SSRN*, 2011b. doi: 10.2139/ssrn.1980726.

J. Andreasen and B. Huge. Expanded forward volatility. *Risk*, 26(1):101–107, 2013.

A. Antonov. Black basket analytics for mid-curves and spread-options. *Available at SSRN*, 2020. doi: 10.2139/ssrn.3633709.

A. Antonov and T. Misirpashaev. Markovian projection onto a displaced diffusion: Generic formulas with applications. *International Journal of Theoretical and Applied Finance*, 12(4):507–522, 2009. doi: 10.1142/S0219024909005300.

A. Antonov, M. Arneguy, and N. Audet. Markovian projection to a displaced volatility Heston model. *Available at SSRN*, 2008. doi: 10.2139/ssrn.1106223.

A. Antonov, M. Konikov, and M. Spector. Mixing SABR models for negative rates. *Available at SSRN*, 2015a. doi: 10.2139/ssrn.2653682.

A. Antonov, M. Konikov, and M. Spector. The free boundary SABR: natural extension to negative rates. *Available at SSRN*, 2015b. doi: 10.2139/ssrn.2557046.

L. Bachelier. Théorie de la spéculation. *Ann. Sci. École Norm. Sup. (3)*, 17:21–86, 1900. doi: 10.24033/asens.476.

P. Balland and Q. Tran. SABR goes normal. *Risk*, 26(6):72–77, 2013.

D. Bang. Local-stochastic volatility models for vanilla modeling: A tractable and arbitrage free approach to option pricing. *Available at SSRN*, 2019. doi: 10.2139/ssrn.3171877.

D. S. Bates. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *The Review of Financial Studies*, 9(1):69–107, 1996. doi: 10.1093/rfs/9.1.69.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81(3):637–654, 1973. ISSN 0022-3808. doi: 10.1086/260062.

P. Bourgade and O. Croissant. Heat kernel expansion for a family of stochastic volatility models: delta-geometry. *Available at SSRN*, 2005. doi: 10.2139/ssrn.844229.

A. Brace, D. Gątarek, and M. Musiela. The market model of interest rate dynamics. *Math. Finance*, 7(2):127–155, 1997. ISSN 0960-1627. doi: 10.1111/1467-9965.00028.

D. Brigo and F. Mercurio. *Interest rate models—theory and practice*. Springer Finance. Springer-Verlag, Berlin, second edition, 2006. ISBN 978-3-540-22149-4; 3-540-22149-2. With smile, inflation and credit.

R. Brown. Xxvii. a brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *The Philosophical Magazine*, 4(21):161–173, 1828. doi: 10.1080/14786442808674769.

P. Carr and D. B. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4):61–73, 1999. doi: http://dx.doi.org/10.21314/JCF.1999.043.

J. Chan and M. S. Joshi. Fast and accurate long stepping simulation of the Heston stochastic volatility model. *Available at SSRN*, 2010. doi: 10.2139/ssrn.1617187.

X. Charvet and Y. Ticot. Pricing with a smile: An approach using normal inverse Gaussian distributions with a SABR-like parameterisation. *Available at SSRN*, 2011. doi: 10.2139/ssrn.1968453.

J. Choi, C. Liu, and B. K. Seo. Hyperbolic normal stochastic volatility model. *Journal of Futures Markets*, 39(2):186–204, 2019. doi: 10.1002/fut.21967.

V. Chuni. *On SDE systems with non-Lipschitz diffusion coeffcients*. PhD thesis, Alma Mater Studiorum Università di Bologna, 2020.

V. Costeanu and D. Pirjol. Asymptotic expansion for the normal implied volatility in local volatility models. *Available at arXiv*, 2011. doi: 10.48550/ARXIV.1105.3359.

Z. Cui, J. Kirkby, and D. Nguyen. A general valuation framework for SABR and stochastic local volatility models. *SIAM Journal on Financial Mathematics*, 9 (2):520–563, 2018. doi: 10.1137/16M1106572.

L. Decreusefond and A. S. Üstünel. Stochastic analysis of the fractional Brownian motion. *Potential Anal.*, 10(2):177–214, 1999. ISSN 0926-2601. doi: 10.1023/A:1008634027843.

F. Delbaen and W. Schachermayer. Arbitrage and free lunch with bounded risk for unbounded continuous processes. *Mathematical Finance*, 4(4):343–348, 1994a. doi: 10.1111/j.1467-9965.1994.tb00063.x.

F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3):463–520, 1994b. ISSN 0025-5831. doi: 10.1007/BF01450498.

F. Delbaen and W. Schachermayer. The fundamental theorem of asset pricing for unbounded stochastic processes. *Math. Ann.*, 312(2):215–250, 1998. ISSN 0025-5831. doi: 10.1007/s002080050220.

E. Derman and I. Kani. Riding on a smile. *Risk*, 7(2):32–39, 1994.

D. J. Duffy. *Finite difference methods in financial engineering: A partial differential equation approach*. Wiley Finance Series. John Wiley & Sons, Ltd., Chichester, 2006. ISBN 978-0-470-85882-0; 0-470-85882-6. doi: 10.1002/9781118673447.

B. Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, Cambridge, fourth edition, 2010. ISBN 978-0-521-76539-8.

A. Einstein. On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. *Annalen der Physik*, 17: 549–560, 1905.

A. Eriksson, E. Ghysels, and F. Wang. The normal inverse Gaussian distribution and the pricing of derivatives. *The Journal of Derivatives*, 16(3):23–37, 2009. doi: 10.3905/JOD.2009.16.3.023.

G. Evans, J. Blackledge, and P. Yardley. *Numerical methods for partial differential equations*. Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, 2000. ISBN 3-540-76125-X. doi: 10.1007/978-1-4471-0377-6.

F. Fang and B. Janssens. Characteristic function of the hybrid Heston–Hull–White model. *European Study Group Mathematics with Industry*, page 107, 2007.

K. Feldman. Change of measure in midcurve pricing. *Wilmott*, 2020(106):76–81, 2020. doi: 10.1002/wilm.10833.

M. Felpel, J. Kienitz, and T. A. McWalter. Effective stochastic volatility: Applications to ZABR-type models. *Quantitative Finance*, 21(5):837–852, 2021. doi: 10.1080/14697688.2020.1814396.

M. Felpel, J. Kienitz, and T. A. McWalter. Effective Markovian projection: Application to CMS spread options and mid-curve swaptions. *Quantitative Finance*, 0(0):1–24, 2022a. doi: 10.1080/14697688.2022.2043558.

M. Felpel, J. Kienitz, and T. A. McWalter. Effective stochastic local volatility models. *Available at SSRN*, 2022b. doi: 10.2139/ssrn.4016334.

J. Gatheral. Efficient simulation of affine forward variance models. *Available at SSRN*, 2021. doi: 10.2139/ssrn.3876680.

J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quant. Finance*, 18(6):933–949, 2018. ISSN 1469-7688. doi: 10.1080/14697688.2017.1393551.

J. Gatheral, P. Jusselin, and M. Rosenbaum. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. *Available at arXiv*, 2020. doi: 10.48550/ARXIV.2001.01789.

P. Glasserman. *Monte Carlo methods in financial engineering*. Springer-Verlag, New York, 2004. ISBN 0-387-00451-3.

J. Göttker-Schnetmann and K. Spanderen. Heston stochastic local volatility, 2015. Quantlib Report.

L. A. Grzelak and C. W. Oosterlee. On the Heston model with stochastic interest rates. *SIAM J. Financial Math.*, 2(1):255–286, 2011. doi: 10.1137/090756119.

H. Guerreiro and J. Guerra. Least squares Monte Carlo methods in stochastic Volterra rough volatility models. *Available at arXiv*, 2021. doi: 10.48550/ARXIV.2105.04511.

J. Guyon and P. Henry-Labordère. The smile calibration problem solved. *Available at SSRN*, 2011. doi: 10.2139/ssrn.1885032.

J. Guyon and P. Henry-Labordère. Being particular about calibration. *Risk*, 25 (1):88, 2012.

I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probability Theory and Related Fields*, 71(4):501–516, 1986. ISSN 0178-8051. doi: 10.1007/BF00699039.

P. S. Hagan. Change of variables and conservative numerical schemes. *Available at Researchgate*, 2013. doi: 10.13140/RG.2.1.2521.2320.

P. S. Hagan. Conservative schemes for solving 1D PDEs. *Available at Researchgate*, 2015.

P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.

P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Arbitrage-free SABR. *Wilmott*, 2014(69):60–75, 2014. doi: 10.1002/wilm.10290.

P. S. Hagan, A. S. Lesniewski, and D. E. Woodward. Probability distribution in the SABR model of stochastic volatility. In *Large deviations and asymptotic methods in finance*, volume 110 of *Springer Proc. Math. Stat.*, pages 1–35. Springer, Cham, 2015. doi: 10.1007/978-3-319-11605-1_1.

P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Universal smiles. *Wilmott*, 2016(84):40–55, 2016. doi: 10.1002/wilm.10523.

P. S. Hagan, A. S. Lesniewski, and D. E. Woodward. Implied volatility formulas for Heston models. *Wilmott*, 2018(98):44–57, 2018a. doi: 10.1002/wilm.10723.

P. S. Hagan, A. S. Lesniewski, and D. E. Woodward. Effective media analysis for stochastic volatility models. *Wilmott*, 2018(93):46–55, 2018b. doi: 10.1002/wilm.10645.

P. S. Hagan, A. S. Lesniewski, G. Skoufis, and D. E. Woodward. Convexity without replication. *Wilmott*, 2020(105):58–69, 2020a. doi: 10.1002/wilm.10820.

P. S. Hagan, A. S. Lesniewski, and D. E. Woodward. Implied volatilities for mean reverting SABR models. *Wilmott*, 2020(108):62–77, 2020b. doi: 10.1002/wilm.10859.

P. S. Hagan, A. S. Lesniewski, G. E. Skoufis, and D. E. Woodward. SABR for baskets. *Wilmott*, 2021(112):50–61, 2021a. doi: 10.1002/wilm.10917.

P. S. Hagan, A. S. Lesniewski, G. E. Skoufis, and D. E. Woodward. CMS spread options. *Quant. Finance*, 21(11):1809–1824, 2021b. ISSN 1469-7688. doi: 10.1080/14697688.2021.1912379.

P. Henry-Labordère. A general asymptotic implied volatility for stochastic volatility models. *Available at SSRN*, 2005. doi: 10.2139/ssrn.698601.

P. Henry-Labordère. Calibration of local stochastic volatility models to market smiles: A Monte Carlo approach. *Risk*, 22(9):112–117, 2009.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6 (2):327–343, 1993. ISSN 0893-9454. doi: 10.1093/rfs/6.2.327.

M. H. Holmes. *Introduction to perturbation methods*, volume 20 of *Texts in Applied Mathematics*. Springer, New York, second edition, 2013. ISBN 978-1-4614-5476-2; 978-1-4614-5477-9. doi: 10.1007/978-1-4614-5477-9.

B. Horvath and O. Reichmann. Dirichlet forms and finite element methods for the SABR model. *SIAM J. Financial Math.*, 9(2):716–754, 2018. doi: 10.1137/16M1066117.

J. Hunter. Asymptotic analysis and singular perturbation theory. *Department of Mathematics, University of California at Davis*, pages 1–3, 2004.

K. J. In 't Hout and S. Foulon. ADI finite difference schemes for option pricing in the Heston model with correlation. *Int. J. Numer. Anal. Model.*, 7(2):303–320, 2010. ISSN 1705-5105.

K. Itô. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20:519–524, 1944.

P. Jäckel. *Monte Carlo Methods in Finance*. Wiley, 2002. ISBN 047149741X.

P. Jäckel and C. Kahl. Hyp hyp hooray. *Wilmott Magazine*, 34:70–81, 2008.

M. John and Y. Wu. Calculating infinitesimal generators. *Journal of Stochastic Analysis*, 2(4):4, 2021. doi: 10.31390/josa.2.4.04.

N. L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36:149–176, 1949. ISSN 0006-3444. doi: 10.1093/biomet/36.1-2.149.

B. Jourdain and A. Zhou. Existence of a calibrated regime switching local volatility model. *Mathematical Finance*, 30(2):501–546, 2020. doi: 10.1111/mafi.12231.

M. Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4):1–23, 1966. ISSN 00029890, 19300972. doi: 10.2307/2313748.

H. Kammeyer and J. Kienitz. The Heston–Hull–White model part I: Finance and analytics. *Wilmott*, 2012(57):46–53, 2012a. doi: 10.1002/wilm.10074.

H. Kammeyer and J. Kienitz. The Heston–Hull–White model part II: Numerics and examples. *Wilmott*, 2012(58):34–45, 2012b. doi: 10.1002/wilm.10091.

H. Kammeyer and J. Kienitz. The Heston-Hull–White model part III: Design and implementation. *Wilmott*, 2012(59):44–49, 2012c. doi: 10.1002/wilm.10111.

I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991. ISBN 0-387-97655-8. doi: 10.1007/978-1-4612-0949-2.

M. Karlsmark. *Four essays in quantitative finance*. PhD thesis, University of Copenhagen, 2013.

J. Kevorkian and J. D. Cole. *Multiple scale and singular perturbation methods*, volume 114 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996. ISBN 0-387-94202-5. doi: 10.1007/978-1-4612-3968-0.

J. Kienitz and D. Wetterau. *Financial modelling: Theory, implementation and practice with MATLAB source*. John Wiley & Sons, 2013. doi: 10.1002/9781118818565.

J. Kienitz, T. A. McWalter, and R. Sheppard. PDE methods for SABR. In *Novel methods in computational finance*, pages 265–291. Springer, Cham, 2017. doi: 10.1007/978-3-319-61282-9_15.

S. G. Kou. A jump-diffusion model for option pricing. *Management science*, 48 (8):1086–1101, 2002. doi: 10.1287/mnsc.48.8.1086.166.

J. H. Lambert. Observations variae in Mathesin Puram. *Acta Helvitica, physico-mathematico-anatomico-botanico-medica*, 3:128–168, 1758.

J. D. Lawson and D. A. Swayne. A simple efficient algorithm for the solution of heat conduction problems. In *Proceedings of the Sixth Manitoba Conference on Numerical Mathematics (Univ. Manitoba, Winnipeg, Man., 1976)*, pages 239–250. Congress. Numer., XVIII, 1977.

F. Le Floc'h and G. Kennedy. Finite difference techniques for arbitrage-free SABR.

Journal of Computational Finance, 20(3):51–79, 2017. doi: 10.21314/JCF.2016.320.

A. E. Lindsay and D. Brecher. Simulation of the CEV process and the local martingale property. *Mathematics and Computers in Simulation*, 82(5):868–878, 2012. doi: 10.1016/j.matcom.2011.12.006.

R. Lord and A. Farebrother. Fifty shades of SABR simulation. In *Presentation at 10th Fixed Income Conference, WBS Training, Barcelona, Spain*, 2014.

R. Lord and C. Kahl. Complex logarithms in Heston-like models. *Math. Finance*, 20(4):671–694, 2010. ISSN 0960-1627. doi: 10.1111/j.1467-9965.2010.00416.x.

D. Madan and E. Seneta. The variance gamma (VG) model for share market returns. *Journal of business*, pages 511–524, 1990. doi: 10.1086/296519.

D. B. Madan, P. P. Carr, and E. C. Chang. The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105, 1998. doi: 10.1023/A:1009703431535.

K. Matsuda. Introduction to Merton jump diffusion model. *Department of Economics. The Graduate Center, The City University of New York*, 2004.

W. A. McGhee. An artificial neural network representation of the SABR stochastic volatility model. *Journal of Computational Finance*, 25(2), 2021. doi: 10.21314/JCF.2021.007.

R. C. Merton. Theory of rational option pricing. *Bell J. Econom. and Management Sci.*, 4:141–183, 1973. ISSN 0005-8556. doi: 10.2307/3003143.

R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144, 1976. doi: 10.1016/0304-405X(76)90022-2.

A. Muguruza. Not so particular about calibration: Smile problem resolved. *Available at SSRN*, 2019. doi: 10.2139/ssrn.3461545.

M. Musiela and M. Rutkowski. *Martingale methods in financial modelling*, volume 36 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, second edition, 2005. ISBN 3-540-20966-2.

B. Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. ISBN 3-540-04758-1. doi: 10.1007/978-3-642-14394-6. An introduction with applications.

V. Piterbarg. Markovian projection method for volatility calibration. *Available at SSRN*, 2006. doi: 10.2139/ssrn.906473.

Y. Ren, D. Madan, and M. Q. Qian. Calibrating and pricing with embedded local volatility models. *Risk*, 20(9):138, 2007.

F. Rouah. *The Heston Model and its Extensions in Matlab and C#*. John Wiley & Sons, Ltd, 2013. ISBN 9781118656471. doi: 10.1002/9781118656471.

P. A. Samuelson. Rational theory of warrant pricing. *Industrial Management Review*, 3:13–31, 1965.

Y. F. Saporito, X. Yang, and J. P. Zubelli. The calibration of stochastic local-volatility models: An inverse problem perspective. *Computers & Mathematics with Applications*, 77(12):3054–3067, 2019. ISSN 0898-1221. doi: 10.1016/j.camwa.2019.01.029.

R. Schöbel and J. Zhu. Stochastic volatility with an Ornstein–Uhlenbeck process: An extension. *Review of Finance*, 3(1):23–46, 1999. doi: 10.1023/A:1009803506170.

W. Schoutens. *Lévy Processes in Finance*. John Wiley & Sons, Ltd, 2003. ISBN 9780470870235. doi: 10.1002/0470870230.fmatter.

R. Sheppard. *Pricing equity derivatives under stochastic volatility: A partial differential equation approach*. PhD thesis, University of the Witwatersrand, 2007.

G. D. Smith. *Numerical solution of partial differential equations*. The Clarendon Press, Oxford University Press, New York, third edition, 1985. ISBN 0-19-859641-3; 0-19-859650-2.

M. Staunton. Monte Carlo for Heston. *Wilmott Magazine*, 29, 2007.

E. M. Stein and J. C. Stein. Stock price distributions with stochastic volatility: An analytic approach. *The review of financial studies*, 4(4):727–752, 1991. doi: 10.1093/rfs/4.4.727.

K. Suresh Kumar. A class of degenerate stochastic differential equations with non-Lipschitz coefficients. *Proc. Indian Acad. Sci. Math. Sci.*, 123(3):443–454, 2013. ISSN 0253-4142. doi: 10.1007/s12044-013-0141-8.

P. Tankov and E. Voltchkova. Jump-diffusion models: A practitioner's guide, 2009.

D. Tavella and C. Randall. *Pricing financial instruments: The finite difference method*. John Wiley & Sons, 2000.

B. Tavin. Implied distribution as a function of the volatility smile. *Bankers Markets and Investors*, 119:31–42, 2012.

S. Taylor. *Perturbation and symmetry techniques applied to finance*. PhD thesis, Frankfurt School of Finance & Management gGmbH, 2010.

J. W. Thomas. *Numerical partial differential equations: Finite difference methods*, volume 22 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1995. ISBN 0-387-97999-9. doi: 10.1007/978-1-4899-7278-1.

O. Tsuchiya. Markovian projection for the local stochastic volatility LIBOR market model. *Available at SSRN*, 2015. doi: 10.2139/ssrn.2610110.

H. J. H. Tuenter. An algorithm to determine the parameters of S_U-curves in the Johnson system of probability distributions by moment matching. *J. Statist. Comput. Simulation*, 70(4):325–347, 2001. ISSN 0094-9655. doi: 10.1080/00949650108812126.

A. Van der Stoep, L. A. Grzelak, and C. W. Oosterlee. The Heston stochastic-local volatility model: Efficient Monte Carlo simulation. *International Journal of Theoretical and Applied Finance*, 17(7):1450045, 30, 2014. ISSN 0219-0249. doi: 10.1142/S0219024914500459.

F. Verhulst. *Methods and applications of singular perturbations*, volume 50 of *Texts in Applied Mathematics*. Springer, New York, 2005. ISBN 978-0387-22966-9; 0-387-22966-3. doi: 10.1007/0-387-28313-7. Boundary layers and multiple timescale dynamics.

S. Wang. *Jump Diffusion Model*, pages 525–534. Springer US, Boston, MA, 2013. ISBN 978-1-4614-5360-4. doi: 10.1007/978-1-4614-5360-4_44.

N. Wiener. The Dirichlet problem. *Journal of Mathematics and Physics*, 3(3): 127–146, 1924. doi: 10.1002/sapm192433127.